



**HAL**  
open science

# Visualization to Support Multi-Criteria Decision-making in Agronomy

Mehdi Chakhchoukh

► **To cite this version:**

Mehdi Chakhchoukh. Visualization to Support Multi-Criteria Decision-making in Agronomy. Human-Computer Interaction [cs.HC]. Université Paris-Saclay, 2024. English. NNT : 2024UPASG085 . tel-04926672

**HAL Id: tel-04926672**

**<https://theses.hal.science/tel-04926672v1>**

Submitted on 3 Feb 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Visualization to Support Multi-Criteria  
Decision-making in Agronomy  
*Visualisation pour Soutenir la Prise de Décision  
Multicritère en Agronomie*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 580, Sciences et Technologies de l'Information et de la  
Communication (STIC)

Spécialité de doctorat: Informatique

Graduate School : Informatique et Sciences du Numérique. Référent : Faculté des  
sciences d'Orsay

Thèse préparée dans l'unité de recherche **Laboratoire interdisciplinaire des sciences  
du numérique (Université Paris-Saclay, CNRS, Inria)**, sous la direction  
de **Anastasia BEZERIANOS**, Professeure et le co-encadrement  
de **Nadia Boukhelifa**, Chargée de Recherche CRCN

Thèse soutenue à Gif-sur-Yvette, le 4 décembre 2024, par

**Mehdi R. Chakhchoukh**

**Composition du jury**

Membres du jury avec voie délibérative:

<b>Wendy MACKAY</b> Directrice de recherche, INRIA, Université Paris-Saclay, France	Présidente
<b>Michael J. MCGUFFIN</b> Professor, École de Technologie Supérieure (ÉTS), Canada	Rapporteur et Examineur
<b>Michael SEDLMAIR</b> Professor, University of Stuttgart, Germany	Rapporteur et Examineur
<b>Samuel HURON</b> Maître de conférences, Telecom Paris Tech, France	Examineur



**Titre:** Visualisation pour Soutenir la Prise de Décision Multicritère en Agronomie

**Mots clés:** Interaction Humain-Machine, Visualisation des Données, Analyse Visuelle, Prise de Décision Multicritère, Analyse des Compromis, Provenance, Comparaison.

**Résumé:**

L'augmentation de la complexité des systèmes agricoles nécessite des outils d'aide à la prise de décision sophistiqués capables de gérer plusieurs critères et d'accommoder des tâches complexes d'analyse des compromis. Cette thèse développe des visualisations qui facilitent les processus de prise de décision en agronomie. Ce travail apporte trois contributions principales : (i) comprendre comment la provenance peut soutenir l'analyse des compromis, (ii) articuler des besoins de conception et de visualisation de haut niveau pour soutenir la comparaison de groupes dans des scénarios de compromis, et enfin (iii) comprendre comment différentes visualisations peuvent affecter les comparaisons et la prise de décision dans l'analyse des compromis. Après un chapitre d'introduction et un chapitre sur les travaux connexes, la thèse détaille ces trois contributions principales.

Le 3ème chapitre de la thèse examine comment les mécanismes de provenance analytique peuvent aider les experts à se souvenir et à suivre des analyses de compromis complexes. Nous avons développé VisProm, un outil web intégrant des vues de provenance dans les visualisations pour aider les experts à suivre les compromis et leurs objectifs lors de l'exploration de résultats de simulation complexes. Des sessions d'observation avec des groupes d'experts ont révélé huit tâches clés soutenues par nos conceptions, mettant en lumière de nouvelles opportunités pour l'analyse des compromis guidée par la provenance, telles que le suivi de la couverture de l'espace de compromis et l'exploration de scénarios alternatifs. Un résultat clé a été la nécessité de considérer les objectifs conflictuels et de comparer comment différentes solutions ou espaces de compromis se comportent face à ces objectifs.

En s'appuyant sur cela, le 4ème chapitre explore les besoins et les défis auxquels sont confrontés les experts lorsqu'ils comparent des espaces de compromis (souvent exprimés sous forme de groupes de points de données, par exemple, des

groupes de résultats de simulation) qui optimisent différents objectifs. À travers des ateliers avec des experts du domaine et des concepteurs de visualisations, nous avons identifié des besoins de conception et de visualisation de haut niveau pour soutenir la comparaison de groupes dans des scénarios de compromis. Ce chapitre met les bases pour développer des techniques de visualisation capables de comparer des groupes représentant différents compromis en fonction des objectifs qu'ils optimisent. Cela a mené à la mise en œuvre d'un prototype de visualisation qui encode visuellement diverses métriques de compromis, en communiquant les priorités des experts, la notion de solutions idéales, et la distance des groupes de solutions par rapport à ces idéaux.

Le 5ème chapitre se concentre sur l'évaluation des techniques de visualisation pour comparer des groupes de points (solutions) lorsqu'ils représentent différents compromis. S'inspirant des besoins et des exigences de conception identifiés dans le chapitre précédent, nous avons sélectionné trois techniques prometteuses de visualisation basées sur des tableaux pour les étudier. Ces techniques encodent visuellement les priorités des compromis et les solutions idéales de différentes manières, soit en couplant, soit en découplant les métriques de compromis. Nous avons mené une étude utilisateur pour comprendre comment ces visualisations influencent les décisions de comparaison et la qualité des explications de ces décisions. Les résultats de cette étude montrent que les techniques qui séparent visuellement l'encodage des priorités et des solutions idéales entraînent une charge cognitive plus élevée et une confiance auto-évaluée plus faible, mais qu'elles soutiennent des stratégies de décision plus variées par rapport aux visualisations intégrées. Cependant, elles étaient toujours préférées par rapport à la visualisation de référence.

Nous concluons la thèse par une série de discussions et de perspectives sur les directions futures issues de ce travail.

**Title:** Visualization to Support Multi-Criteria Decision-making in Agronomy

**Keywords:** Human-Computer Interaction; Data Visualization; Visual Analytics; Multi-Criteria Decision-making; Trade-off Analysis; Provenance; Comparison

**Summary:**

The increasing complexity of agricultural systems necessitates sophisticated decision-making tools that can handle multiple criteria and accommodate complex trade-off analysis tasks. This thesis develops visualizations that facilitate decision-making processes in agronomy. This work has three main contributions: (i) understanding how provenance can support trade-off analysis, (ii) articulating high-level design and visualization needs to support group comparison in trade-off scenarios, and finally, (iii) understanding how different visualizations can affect comparisons and decision-making in trade-off analysis. After an introductory chapter and a chapter on related work, the thesis details these three main contributions.

The 3rd chapter of the thesis investigates how analytic provenance mechanisms can assist experts in recalling and tracking complex trade-off analyses. We developed VisProm, a web-based system integrating in-visualization provenance views to help experts track trade-offs and their objectives when exploring complex simulation results. Observation sessions with groups of experts revealed eight key tasks supported by our designs, highlighting new opportunities for provenance-driven trade-off analysis, such as monitoring trade-off space coverage and tracking alternative scenarios. One key outcome was the need to consider conflicting objectives and compare how different solutions or trade-off spaces fare under these objectives.

Building on this, the 4th chapter explores the needs and challenges experts face when comparing trade-off spaces (that are often expressed as groups of data points, e.g., groups of simulation results) that optimize different objectives. Through workshops with domain experts and visualization

designers, we identified high-level design and visualization needs to support group comparison in trade-off scenarios. This chapter lays the groundwork for developing effective visualization techniques for comparing groups that represent different trade-offs in terms of what objectives they optimize. They led to the implementation of a visualization prototype that visually encodes a variety of trade-off metrics. These encode and visually communicate experts' priorities in terms of objectives, the notion of ideal solutions, and how far current groups of solutions are from those ideals.

The 5th chapter focuses on the evaluation of visualization techniques for comparing groups of points (solutions) when they represent different trade-offs. Motivated by the visualization needs and design requirements of the previous chapter, we selected three promising tabular-based visualization techniques to study. These techniques encode trade-off priorities and ideal solutions in different ways: coupling or decoupling the trade-off metrics and presenting them visually. We conducted a user study to understand how visualizations affected comparison decisions and the quality of decision explanations. The findings from this study highlight that techniques that visually separate the encoding of priorities and ideal solutions lead to higher mental load and lower self-reported trust but may support more varied decision strategies than integrated visualizations. But, they were always preferred over a baseline visualization.

We conclude the thesis with a list of discussions and perspectives for future directions stemming from the results of this work.

# Acknowledgement

I would like to express my deep and sincere gratitude to my advisors, Anastasia Bez-erianos and Nadia Boukhelifa, for providing me with invaluable guidance and for their patience and continuous support. I would also like to thank them for inspiring me to do my best in research with their optimism, dynamism, and unwavering motivation.

I would also like to extend my gratitude to the members of my jury: Michael J. McGuffin, Michael Sedlmair, Samuel Huron, and Wendy Mackay. And a special thanks to the two reporters, Michael J. McGuffin and Michael Sedlmair.

I would like to express my heartfelt appreciation to my LISN colleagues, particularly my fellow ILDA team members.

To the permanents: Emmanuel Pietriga, Caroline Appert, Vanessa Peña-Araya, and Olivier Chapuis – thank you for your unwavering guidance and expertise (and board game nights) that have profoundly shaped my research journey.

To my fellow Ph.D. students Vincent Cavez, Julien Berry, and Camille Dupré, your camaraderie and insightful discussions have made this challenging journey a joy to make. To the new ones, Mengfei Gao and Xinpei Zheng, I wish you good luck on your journey!

To our amazing engineers, Ludovic David, Olivier Gladin, and Alexandre Kabil – your technical support, positivity, and dedication have been instrumental in bringing many of my ideas to life.

Thank you, Katia Evrat, for your constant help with administrative matters and for ensuring that everything ran smoothly when it was most needed.

Finally, to those who have moved on: Arthur Fages, Raphael James, Emmanuel Courtoux, Eugénie Brasier, and Dylan Lebout – your contributions and friendship continue to resonate in the work we do.

I would like to also thank the MIA and ISC-PIF for their warm welcome at the labs. It was a pleasure working with you.

Special thanks go out to all the experts from INRAe and affiliated institutes who participated in my experiments and all of the other study participants.

I would like to thank my friends and my loved ones who supported me and made me feel at home throughout my stay in France, especially Samuel Diop, Pierre and Annie Charpentier, Luca Jourdan, Quentin Lemasson, Carole, Alain and Juliette Botreau, and Lisa Pigniard. And the Bebou team, which contains too many people to cite everyone, but I would like to specifically thank Marion Gandolfini, Nicolas Barthelemy, Alban Petit, and Julien Guyot for their constant support.

Finally, I am deeply grateful to my parents and my brothers for their love, encouragement, and constant support. This achievement would not have been possible without them.



# Abstract

The increasing complexity of agricultural systems necessitates sophisticated decision-making tools that can handle multiple criteria and accommodate complex trade-off analysis tasks. This thesis develops visualizations that facilitate decision-making processes in agronomy. This work has three main contributions: (i) understanding how provenance can support trade-off analysis, (ii) articulating high-level design and visualization needs to support group comparison in trade-off scenarios, and finally (iii) understanding how different visualizations can affect comparisons and decision-making in trade-off analysis. After an introductory chapter and a chapter on related work, the thesis details these three main contributions.

The 3rd chapter of the thesis investigates how analytic provenance mechanisms can assist experts in recalling and tracking complex trade-off analyses. We developed VisProm, a web-based system integrating in-visualization provenance views to help experts track trade-offs and their objectives when exploring complex simulation results. Observation sessions with groups of experts revealed eight key tasks supported by our designs, highlighting new opportunities for provenance-driven trade-off analysis, such as monitoring trade-off space coverage and tracking alternative scenarios. One key outcome was the need to consider conflicting objectives and compare how different solutions or trade-off spaces fare under these objectives.

Building on this, the 4th chapter explores the needs and challenges experts face when comparing trade-off spaces (that are often expressed as groups of data points, e.g., groups of simulation results) that optimize different objectives. Through workshops with domain experts and visualization designers we identified high-level design and visualization needs to support group comparison in trade-off scenarios. This chapter lays the groundwork for developing effective visualization techniques for comparing groups that represent different trade-offs in terms of what objectives they optimize. They led to the implementation of a visualization prototype that visually encodes a variety of trade-off metrics. These encode and visually communicate experts' priorities in terms of objectives, the notion of ideal solutions, and how far current groups of solutions are from those ideals.

The 5th chapter focuses on the evaluation of visualization techniques for comparing groups of points (solutions) when they represent different trade-offs. Motivated by the visualization needs and design requirements of the previous chapter, we selected three promising tabular-based visualization techniques to study. These techniques encode trade-off priorities and ideal solutions in different ways: coupling or decoupling the trade-off metrics and presenting them visually. We conducted a user study to understand how visualizations affected comparison decisions and quality of decision explanations. The findings from this study highlight that techniques that visually separate the encoding of priorities and ideal solutions lead to higher mental load and lower



self-reported trust but may support more varied decision strategies than integrated visualizations. But they were always preferred over a baseline visualization.

We conclude the thesis with a list of discussions and perspectives for future directions stemming from the results of this work.

# Synthèse

L'augmentation de la complexité des systèmes agricoles nécessite des outils d'aide à la prise de décision sophistiqués capables de gérer plusieurs critères et d'accommoder des tâches complexes d'analyse des compromis. Cette thèse développe des visualisations qui facilitent les processus de prise de décision en agronomie. Ce travail apporte trois contributions principales : (i) comprendre comment la provenance peut soutenir l'analyse des compromis, (ii) articuler des besoins de conception et de visualisation de haut niveau pour soutenir la comparaison de groupes dans des scénarios de compromis, et enfin (iii) comprendre comment différentes visualisations peuvent affecter les comparaisons et la prise de décision dans l'analyse des compromis. Après un chapitre d'introduction et un chapitre sur les travaux connexes, la thèse détaille ces trois contributions principales.

Le 3ème chapitre de la thèse examine comment les mécanismes de provenance analytique peuvent aider les experts à se souvenir et à suivre des analyses de compromis complexes. Nous avons développé VisProm, un outil web intégrant des vues de provenance dans les visualisations pour aider les experts à suivre les compromis et leurs objectifs lors de l'exploration de résultats de simulation complexes. Des sessions d'observation avec des groupes d'experts ont révélé huit tâches clés soutenues par nos conceptions, mettant en lumière de nouvelles opportunités pour l'analyse des compromis guidée par la provenance, telles que le suivi de la couverture de l'espace de compromis et l'exploration de scénarios alternatifs. Un résultat clé a été la nécessité de considérer les objectifs conflictuels et de comparer comment différentes solutions ou espaces de compromis se comportent face à ces objectifs.

En s'appuyant sur cela, le 4ème chapitre explore les besoins et les défis auxquels sont confrontés les experts lorsqu'ils comparent des espaces de compromis (souvent exprimés sous forme de groupes de points de données, par exemple, des groupes de résultats de simulation) qui optimisent différents objectifs. À travers des ateliers avec des experts du domaine et des concepteurs de visualisations, nous avons identifié des besoins de conception et de visualisation de haut niveau pour soutenir la comparaison de groupes dans des scénarios de compromis. Ce chapitre met les bases pour développer des techniques de visualisation capables de comparer des groupes représentant différents compromis en fonction des objectifs qu'ils optimisent. Cela a mené à la mise en œuvre d'un prototype de visualisation qui encode visuellement diverses métriques de compromis, en communiquant les priorités des experts, la notion de solutions idéales, et la distance des groupes de solutions par rapport à ces idéaux.

Le 5ème chapitre se concentre sur l'évaluation des techniques de visualisation pour comparer des groupes de points (solutions) lorsqu'ils représentent différents compromis. S'inspirant des besoins et des exigences de conception identifiés dans le chapitre précédent, nous avons sélectionné trois techniques prometteuses de visualisation basées

sur des tableaux pour les étudier. Ces techniques encodent visuellement les priorités des compromis et les solutions idéales de différentes manières, soit en couplant, soit en découplant les métriques de compromis. Nous avons mené une étude utilisateur pour comprendre comment ces visualisations influencent les décisions de comparaison et la qualité des explications de ces décisions. Les résultats de cette étude montrent que les techniques qui séparent visuellement l'encodage des priorités et des solutions idéales entraînent une charge cognitive plus élevée et une confiance auto-évaluée plus faible, mais qu'elles soutiennent des stratégies de décision plus variées par rapport aux visualisations intégrées. Cependant, elles étaient toujours préférées par rapport à la visualisation de référence.

Nous concluons la thèse par une série de discussions et de perspectives sur les directions futures issues de ce travail.

# Contents

Acknowledgements	i
Abstract	iii
Synthèse	v
Contents	vii
List of Figures	xi
<b>1 Introduction</b>	<b>1</b>
1.1 Thesis Statement and Approach . . . . .	5
1.1.1 In-Visualization Provenance for Trade-off Analysis . . . . .	5
1.1.2 Trade-off Comparisons Between Groups . . . . .	6
1.1.3 Approach . . . . .	7
1.2 Contributions . . . . .	8
1.3 Thesis Overview . . . . .	8
1.4 Publications and Papers Submitted . . . . .	10
<b>2 Related Work</b>	<b>11</b>
2.1 Visual Analytics and Sensemaking. A focus on Agronomy. . . . .	11
2.1.1 Visual Analytics . . . . .	11
2.1.2 Sensemaking in Visual Analytics . . . . .	12
2.1.3 Visual Analytics and Decision Support in Agronomy . . . . .	15
2.2 Trade-off Analysis . . . . .	17
2.2.1 Trade-off Analysis and MCDM . . . . .	17
2.2.2 Trade-off Analysis Tasks . . . . .	18
2.2.3 Summary for Trade-off Analysis . . . . .	21
2.3 Multidimensional Data Visualization . . . . .	22
2.3.1 Visual Mapping in Multidimensional Data Visualization . . . . .	23
2.3.2 Summary on Multi-Dimensional Visualization . . . . .	27
2.4 Visualization tools for Trade-off Analysis . . . . .	27
2.4.1 Metrics to Quantify Trade-offs . . . . .	31
2.4.2 Summary on Vis tools for Trade-off Analysis . . . . .	31
2.5 Provenance and Trade-off Analysis . . . . .	32
2.5.1 Provenance Visualization for Sensemaking . . . . .	33
2.5.2 Purposes of Provenance . . . . .	34
2.5.3 Summary for Provenance in Trade-off Analysis . . . . .	35

2.6	Comparisons and Trade-off Analysis . . . . .	35
2.6.1	Comparison Visualization for Sensemaking . . . . .	36
2.6.2	Comparisons in Trade-off Analysis . . . . .	37
2.6.3	Evaluating the Quality of Trade-off Comparisons . . . . .	37
2.6.4	Summary for Comparisons in Trade-off Analysis . . . . .	38
2.7	Placement of this thesis . . . . .	39
<b>3</b>	<b>Understanding How In-Visualization Provenance Can Support Trade-off Analysis</b>	<b>41</b>
3.1	VisProm: a Provenance Visualization tool for Trade-off Analysis . . . . .	41
3.1.1	Design process . . . . .	43
3.1.2	Additional <i>in-visualization</i> Provenance Designs . . . . .	44
3.1.3	Implementation . . . . .	48
3.2	User Study Design . . . . .	48
3.2.1	Participants and Use Cases . . . . .	49
3.2.2	Study Procedure and Setup . . . . .	49
3.2.3	Data Collection and Video Coding . . . . .	51
3.3	Results . . . . .	52
3.3.1	Trade-off analysis tasks [RQ1] . . . . .	52
3.3.2	Refined provenance purposes and objects . . . . .	54
3.3.3	Provenance supports trade-off analysis [RQ2, RQ3] . . . . .	56
3.3.4	Opportunities for provenance visualization [RQ4] . . . . .	60
3.4	Discussion on Provenance for Trade-offs . . . . .	64
3.4.1	Trade-off analysis tasks . . . . .	65
3.4.2	Provenance objects in trade-off analysis . . . . .	66
3.4.3	How in-visualization views support trade-off tasks and provenance objects: during <i>a-posteriori</i> analysis and <i>ongoing</i> exploration . . . . .	67
3.4.4	Trade-off specific opportunities . . . . .	69
3.4.5	Limitations . . . . .	70
3.5	Conclusion . . . . .	70
<b>4</b>	<b>Designing for Trade-off Comparison</b>	<b>73</b>
4.1	Expert Needs and Requirements [RQ5a] . . . . .	74
4.1.1	Videos of Expert Exploration . . . . .	74
4.1.2	Workshops with Experts - Comparison of Discrete Solutions . . . . .	76
4.1.3	Workshops with Experts - Comparison of Composite Solutions . . . . .	77
4.1.4	Summary of Trade-off Comparisons by Experts . . . . .	79
4.2	Design Workshops [RQ5b] . . . . .	80
4.2.1	Participants . . . . .	80
4.2.2	Scenario and Data . . . . .	80

---

4.2.3	Design Task and Procedure . . . . .	81
4.2.4	Analysis and Results . . . . .	82
4.2.5	Summary of Design & Visualization Needs for Comparison [RQ5]	84
4.3	Visualizations for trade-off comparisons . . . . .	85
4.3.1	Visualizations and Comparison Layout . . . . .	86
4.3.2	Expressing Trade-off Metrics . . . . .	87
4.3.3	Bringing it all together: Tool & Interactions . . . . .	89
4.4	Conclusion . . . . .	91
<b>5</b>	<b>Evaluating Visualizations for trade-off comparisons between groups</b>	<b>93</b>
5.1	User Study Design . . . . .	93
5.1.1	Visualization Conditions . . . . .	94
5.1.2	Hypotheses . . . . .	97
5.1.3	Tasks . . . . .	97
5.1.4	Datasets . . . . .	98
5.1.5	Study Measures . . . . .	99
5.1.6	Procedure . . . . .	101
5.1.7	Participants . . . . .	101
5.1.8	Apparatus . . . . .	102
5.2	Results . . . . .	102
5.2.1	Decision Justifications: Information Considered, Decision Strategies & Quality of Explanations [RQ6,7]	103
5.2.2	Perceived load - NASA TLX [RQ6,7]	109
5.2.3	Visualization Preferences and Qualitative Feedback . . . . .	111
5.3	Discussion and Limitations . . . . .	112
5.4	Conclusion . . . . .	114
<b>6</b>	<b>Discussion, Future Work and Conclusion</b>	<b>117</b>
6.1	Summary . . . . .	117
6.2	Future perspectives . . . . .	119
6.2.1	Tracking Conflicting Objectives and Evolving User Preferences in Extended Exploration Sessions . . . . .	119
6.2.2	In-situ, Ex-situ or Hybrid Visualizations of Trade-off Analysis Provenance . . . . .	120
6.2.3	Evaluating the Effectiveness of Trade-off Analysis Tools . . . . .	122
6.2.4	Visualization of Gain and Loss . . . . .	123
6.2.5	Integrating Provenance and Comparison to Support Trade-off Analysis . . . . .	125
6.2.6	Collaboration in Trade-off Analysis . . . . .	125
6.3	General Conclusion . . . . .	127

<b>Bibliography</b>	<b>129</b>
<b>A Supplementary Material for Chapter 5</b>	<b>145</b>

# List of Figures

1.1	Wheat field being fertilized with ammonium nitrate fertilizer . . . . .	2
1.2	Wine fermentation in stainless steel tanks . . . . .	2
1.3	Agronomy experts using an interactive, SPLOM-based visual analytics tool to explore wine fermentation strategies . . . . .	3
2.1	Pirolli and Card’s model of sensemaking loop for intelligence analysis . . . . .	13
2.2	The Jigsaw system for supporting investigative analysis . . . . .	14
2.3	The Vismon interface . . . . .	19
2.4	Pareto Front across two dimensions (quantity of Item 1 and quantity of Item 2) . . . . .	20
2.5	The Lineup system for ranking Multi-Attribute alternatives and comparing the rankings of these alternatives . . . . .	21
2.6	The Weightlifter view integrated with other visualization tools . . . . .	22
2.7	VisDB tool with pixel-based visual encodings . . . . .	24
2.8	EvoGraphDice tool interface . . . . .	28
2.9	Different configurations of the LineUp tool bars . . . . .	29
2.10	The PAVED tool visualizes Pareto fronts from engineering . . . . .	30
2.11	HindSight provenance visualizations showing interaction history . . . . .	33
2.12	The different comparison layouts introduced by Gleicher et al. . . . .	36
3.1	The VisProm technology probe includes several in-visualization provenance views to aid trade-off analysis . . . . .	42
3.2	Visprom: ViewCount . . . . .	44
3.3	Visprom: query selection panel . . . . .	46
3.4	Visprom: Datatable . . . . .	47
3.5	Our identified objects and how they are distributed across Provenance purposes . . . . .	56
4.1	Sketches examples from our design workshops . . . . .	83
4.2	Prototype for trade-off comparisons between groups of points, each represented as a table. . . . .	86
4.3	Interaction using the comparison tool. The visualization in this example is Coupled Bars. . . . .	90
5.1	Our comparison study setup . . . . .	94
5.2	The three different tabular visualization setups used in our comparison study . . . . .	95
5.3	Types of data-related information considered in our coding . . . . .	104
5.4	Decision Strategy across visualizations . . . . .	106
5.5	Mean values for the different categories of the NASA TLX questionnaires . . . . .	110
5.6	Participants’ ranking of the visualizations . . . . .	111



## Introduction

Domains such as agronomy and manufacturing (Figure 1.1, Figure 1.2) often use simulation models that represent entities and parameters of complex biological or physical processes and their relationships. Experts explore the results of these models, reach insights, and make decisions about how to optimize these processes. For example, an agronomic engineer who wants to propose sustainable but robust wheat fertilization strategies to farmers needs to account for the wheat growth process and how it is affected by soil and weather conditions, as well as the impact of the chosen fertilization strategy on the environment [114]. Thus, experts must deal with *trade-offs*, attempting to consider and reconcile multiple competing objectives in a single investigation [106]. For instance, the agronomic engineer needs to find fertilization strategies that, on the one hand, maximize yield and, on the other hand, reduce the amount of supplied fertilizers and nitrogen loss to the plant.

This type of trade-off analysis can be complex as it captures experts' requirements and preferences across several objectives and often involves multiple experts working together to understand different aspects of the data [112]. Depending on the domain, datasets can include tens of dimensions and hundreds or thousands of datapoints. Using their domain knowledge, experts analyze these large multi-dimensional spaces, prioritizing requirements and balancing subjective preferences across several competing objectives. This complex analysis is conducted over long periods of time and several exploration sessions [112], as multiple experts try to explore alternatives and reach common ground, making this type of analysis hard to resume and revisit.

Research in the field of trade-off analysis has its roots in multi-criteria decision-making (MCDM). MCDM is a discipline that studies procedures to aid decision-making when dealing with a large number of alternatives. Often, MCDM problems involve the ordering or classifying of alternatives [39] by imposing a procedure or strategy, like explicitly ranking or weighting criteria [14]. Nevertheless, criteria may not always be known a priori or may be hard to express. In such cases, visualization can play an important role by offering visual and interactive tools that support open and iterative exploration processes. This approach enables domain experts to bridge the knowledge





Figure 1.1: Wheat field being fertilized with ammonium nitrate fertilizer [108].



Figure 1.2: Wine fermentation in stainless steel tanks [53].

gap they may have regarding the data or the (simulation) models they are working with [29, 112]. We, therefore, will use the term trade-off analysis to describe a process that is less structured than the general MCDM process:

*Trade-off analysis* is an exploration process, aided by visualizations, of multiple alternatives in order to understand the different criteria domain experts optimize without any imposed procedures or strategies.

In this thesis, we investigate the use of visual analytics to help with such exploratory trade-off analysis. In particular, we set out to study if and how interactive visualizations can aid trade-off analysis in such contexts. Indeed, it has already been shown that visualizations help with understanding complex data and making decisions [34, 78, 89]. We study how visualizations can help with trade-off analysis problems specifically, drawing from various domains (including machine learning and ecosystem services) but focusing on agronomy applications.

**Provenance.** As a first step, we consider how to explore the use of analytic provenance visualizations [20, 31, 80, 84] to aid experts keep track, resume, and revisit their trade-off analysis sessions. Provenance is "the history of the data and reasoning involved during complex sensemaking tasks and the context within which sensemaking was performed" [85]. In our work, we generally refer to analytical provenance related to interaction logs.

How provenance can support trade-off analysis tasks is a question that has not been considered before and poses unique challenges as analysts need to keep track of the objectives they have already considered, their importance, and the trade-offs they represent [59]. Inspired by past observations on trade-off analysis challenges [112], we design and develop a web-based trade-off analysis tool (VisProm) that embeds prove-

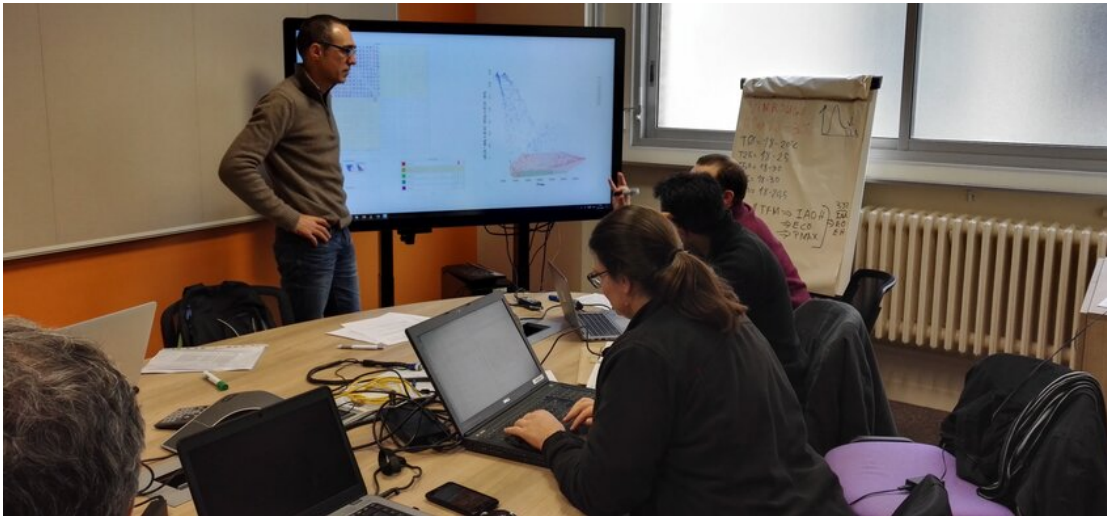


Figure 1.3: Agronomy experts using an interactive, SPLOM-based visual analytics tool to explore wine fermentation strategies [112].

nance visualizations directly inside the analysis environment, i.e., uses *in-visualization provenance*. The tool is designed specifically with trade-off analysis in mind and simplicity when it comes to the proposed provenance views. Since no previous study looked at the role of in-visualization provenance in supporting trade-off analysis, our hypothesis here is that if such simple views are helpful for trade-off analysis, then richer provenance views may be investigated in the future.

We use the VisProm system as a technology probe [48]: a prototype system designed to be field-tested in order to inspire users and designers to think about the benefits and limits of provenance support in trade-off analysis rather than a fully fledged trade-off analysis system. Through observation sessions with three groups of experts analyzing their own data, we bring new insights on analytical provenance usage in real-world analysis scenarios, both during and after data exploration, and we identify new opportunities for future trade-off provenance designs.

Our results of this initial investigation identify eight high-level tasks that experts engaged in during trade-off analysis, such as locating and characterizing interest zones in the trade-off space, and show how these tasks can be supported by provenance visualization. Additionally, we refine findings from previous work on provenance purposes, such as recall and reproduce [80], by identifying specific objects of these purposes related to trade-off analysis, such as interest zones and exploration structure (e.g., exploration of alternatives and branches). We distill insights on how the identified provenance objects and our designs support these trade-off analysis tasks, both when revisiting past analyses and while actively exploring. And finally, we identify new opportunities for provenance-driven trade-off analysis, for example related to monitoring

the coverage of the trade-off space and tracking and comparing alternative trade-off scenarios.

**Comparisons.** The next step in this research is largely inspired by our own observations of domain experts using visualization tools to conduct trade-off comparisons for real-world multi-criteria decisions in agronomy and other domains, as well as observations from prior work [5, 112]). Comparison is a fundamental task in information visualization, often considered a higher-level meta-operation [95, 102, 125, 137]. Gleicher identifies two key elements in visual comparison: a set of related *items* and an *action* that the user wants to perform on the relationship between these items, such as quantifying the relationship or dissecting it in detail to enhance understanding [101]. Comparisons are also essential for Multi-Criteria Decision-Making (MCDM), where trade-off analysis is used to evaluate alternative options with multiple, often conflicting goals [59, 112]. For example, in wheat fertilization, agronomists must balance competing criteria such as reducing fertilizer use and environmental impact while increasing crop yield. They achieve this by comparing the pros and cons of different strategies, such as one that minimizes environmental impact versus another that maximizes crop yield [97]. This trade-off comparison process is an extension of the information visualization comparison as described by Gleicher [101]. It involves evaluating a set of related items, such as the fertilization strategies mentioned earlier, through user actions that assess the advantages (pros) and disadvantages (cons) of these items within a specific trade-off context. This context includes the criteria to be optimized, such as fertilizer use, crop yield, and environmental impact, as well as user priorities related to these goals, such as prioritizing the reduction of environmental impact over maximizing yield.

Trade-off comparisons can become complex, as they involve multiple, often conflicting optimization criteria and many items, either individual or grouped. Previous research in trade-off analysis has shown that domain experts often work with groups of points (or items) when exploring large trade-off spaces [97, 112]. This approach is motivated by two primary factors. First, in many trade-off datasets, especially those derived from biological processes, there is inherent variability in measurements. As a result, experts tend to use ranges instead of exact thresholds when identifying and comparing groups. For instance, during an interactive session involving one thousand wine recipes, experts compared groups rather than individual recipes to determine the optimal fermentation strategy [97]. They selected two groups with differing temperature management strategies, using ranges instead of precise thresholds, allowing for a variety of options within each group. Second, trade-off datasets often include categorical variables that experts use to explore groups of points. In the wheat example, experts compared fertilization strategies aimed at achieving high-protein content with those that resulted in lower protein levels.

Designing effective visualizations for comparison, particularly for groups of items, remains challenging. As Gleicher noted, the complexity of comparisons increases with the number of items, the complexity of those items, and the complexity of their relationships [102]. Existing visualization tools for MCDM, such as Lineup [132] and Weightlifter [135], typically focus on ranking systems to compare and assess *individual items*. However, there has been limited research focused on comparing *groups of items*, especially in complex scenarios like trade-off analysis, where multiple criteria need to be considered simultaneously.

Part of this thesis work is to explore how to support the visual comparison of groups of data-points (or items) within the context of trade-off analysis. Our primary research question focuses on understanding the impact of different types of visualization on this group comparison task in terms of balanced consideration of multiple criteria, user trust, and mental load. To achieve this, we follow a multi-phase approach. First, we conduct a workshop with domain experts to understand the specific needs and challenges involved with comparing groups of data-points in trade-off analysis. This provided us with valuable insights into the types of comparisons users frequently make. Building on these insights, we organize two design workshops with visualization and Human-Computer Interaction (HCI) experts to iterate on and refine visualization designs that could effectively support the comparison of trade-off groups. We then develop a prototype visualization system based on the outcomes of the design workshops. Finally, we evaluate the effectiveness of two promising visualization techniques against a baseline in a user study involving 18 participants.

## 1.1 Thesis Statement and Approach

This dissertation makes a case for the following statement:

*Visualizations can aid in trade-off analysis.*

To support this statement, we study two situations where visualizations can impact trade-off analysis, attempting to answer more specific research questions described next.

### 1.1.1 In-Visualization Provenance for Trade-off Analysis

We first study the impact of using a specific type of visualization, in-visualization provenance, on trade-off analysis. We posit that if simpler provenance views that are integrated within the main visualization environment are helpful for trade-off analysis, then richer provenance views may be investigated in the future. We study if and

how in-visualization provenance can support trade-off analysis, aiming to answer the following four research questions:

- [RQ1] What tasks do experts engage in during trade-off analysis? How do provenance purposes manifest in trade-off analysis?
- [RQ2] Does in-visualization provenance support *a-posteriori* analysis, such as by helping recall of past exploration?
- [RQ3] Is in-visualization provenance taken into account by experts *during* trade-off analysis and how?
- [RQ4] What are untapped opportunities for using provenance to support trade-off analysis?

The approach we take to answer these questions involves observing experts, analyzing their tasks, and designing visualizations that incorporate provenance information to support their analysis. We revisit these questions in [Chapter 3](#).

We show that domain experts engage in different high-level tasks that are key in trade-off analysis, such as observing coverage, defining focus and priorities, branching and comparing alternatives, and looking for interest zones that maximize/minimize specific objectives. These tasks are supported by our in-visualization provenance views, which help our experts recall, replicate, recover, reflect upon, or present & communicate different aspects of their exploration that we identify as trade-off provenance objects. From these, keeping track of coverage and priorities, and to a lesser extent, exploration strategy, are the most important provenance aspects (or objects) for our experts when revisiting past trade-off analysis, whereas interest zones and exploration structure & steps are crucial when starting new trade-off explorations.

### 1.1.2 Trade-off Comparisons Between Groups

Our work in the previous chapter highlighted the need to analyze and compare interest zones or branches in the exploration across different objectives and priorities. Motivated by these results, our work next focuses on comparison trade-off tasks, in particular of groups. We investigate how visualizations can aid group comparison in trade-off analysis, aiming to answer the following three questions:

- [RQ5] What are the trade-off comparison needs of domain experts, and how can we design to support them?
- [RQ6] Does the visualization of trade-off information (metrics) aid comparison more effectively than a simple representation (e.g., tabular)?

- **[RQ7]** Is a coupled or decoupled visualization of trade-off measures better at supporting this comparison?

The approach we take to answer these questions involves understanding user needs and designing and evaluating different visualization techniques to determine their effectiveness in aiding comparison during trade-off analysis.

We first revisit past videos of trade-off analysis conducted with experts and organize an additional three workshops with experts to understand their comparison needs. Through a second set of workshops with visualization designers, we identify visualization requirements to support such comparison tasks. Informed by these workshops, we implement a prototype tool that visualizes in different ways a set of trade-off metrics, added on top of a tabular visualization. We explain this process and results in [Chapter 4](#).

Through a user study, we evaluate the two most promising visualizations we identified in the workshops and added to our prototype. One uses color, and the other bars. We compare these against a text-based baseline. We hypothesize that visualizing trade-off metrics, such as priority and distance from the idea, improves group comparison by reducing mental load and enhancing user understanding of the trade-off, resulting in more user trust in their decisions; and that a decoupled visualization of trade-off metrics enhances user understanding of the trade-offs and increases user trust in decisions compared to a coupled visualization, but it leads to higher mental load. Our findings show that visualizations that separate different trade-off metrics lead to higher mental load and lower self-reported trust but may support more varied decision strategies than integrated visualizations. But either way, all visualizations were better received than no visualization at all (our baseline). We expand on this study and results in [Chapter 5](#).

### 1.1.3 Approach

To support our thesis statement and answer our specific research questions, we followed a User-Centred Design (UCD) [\[68\]](#) approach. The two major investigations carried out in this work, pertaining to in-visualization provenance and comparison of groups to support trade-off analysis, involved: identifying end-user needs through observational studies, interviews, and design workshops; design and implementation of a technology probe [\[48\]](#) to test our hypotheses and designs; and validation in user studies that were either more open-ended and observational with domain experts using in-visualization provenance, or in the form of a controlled experiment approach on group comparison.

## 1.2 Contributions

The contributions of this research are the following:

1. the delineation of eight trade-off analysis tasks extending previous work in operations research [59];
2. the specification of the general provenance purposes [31], through eight provenance objects, that our participants appear to focus on during trade-off analysis;
3. a technology probe to support in-visualization provenance for trade-off analysis, used by various experts in agronomy;
4. insights on how the identified provenance objects and our implemented designs support these trade-off analysis tasks during a-posteriori exploration and active analysis;
5. the identification of specific needs and opportunities for the use of provenance during trade-off analysis, that refine or go beyond those identified in previous provenance work [31, 84];
6. a characterization of comparison needs in trade-off analysis processes;
7. high-level design and visualization needs to support group comparison in trade-off scenarios;
8. a prototype tool to study trade-off comparison of groups;
9. a user study comparing three visualization techniques for comparing groups and insights on: what information is considered during the comparison and the impact of visualization on decision-making strategy, quality of explanations, user trust and mental load.

## 1.3 Thesis Overview

We provide here an overview of the rest of the thesis.

**Chapter 2- Related work:** first provides background on visual analytics and sensemaking, and then trade-off analysis and decision-making, before focusing more specifically on multidimensional data visualization and relevant decision-making support visualization tools. It then provides more detailed related work on the specific topics that we address in the thesis, namely work on *provenance* visualization as we use it to support trade-off analysis, as well as work on *comparison* visualizations.



**Chapter 3- Understanding how in-visualization provenance can support trade-off analysis** presents a first user study (observational) with three groups of domain experts who used a provenance-augmented probe to explore their own trade-off datasets. Findings from this study are fourfold. First, we identify eight high-level tasks that experts engaged in during trade-off analysis, such as locating and characterizing interest zones in the trade-off space, and show how these tasks can be supported by provenance visualization. Second, we refine findings from previous work on provenance purposes such as recall and replicate by identifying specific objects of these purposes related to trade-off analysis, such as interest zones and exploration structure (e.g., exploration of alternatives and branches). Third, we discuss insights on how the identified provenance objects and our designs support these trade-off analysis tasks, both when revisiting past analyses and while actively exploring. Finally, we identify new opportunities for provenance-driven trade-off analysis, for example, related to monitoring the coverage of the trade-off space and tracking alternative trade-off scenarios.

**Chapter 4- Designing for comparisons: Experts requirements and generative workshops** motivates the need for trade-off comparison of groups through observations of domain experts conducting comparisons in real-world settings, and an additional workshop with domain experts to clarify these requirements. The chapter then describes potential visualization solutions to support these needs, based on findings from design workshops. Findings from these workshops (with domain experts and designers) showed that domain experts not only require access to aggregated information and metrics, such as means and variance, as well as detailed information about the individual points, but importantly they often engaged in trade-off comparisons of *groups* of datapoints, which motivated our work on trade-off comparison, described in the 5th chapter. We end the chapter by presenting the prototype system inspired by these results.

**Chapter 5- Visualizations to support trade-off comparisons between groups** presents a second user study (control experiment) to evaluate the two most promising visualizations to show trade-off metrics, one using color and the other bars, together with a text-based baseline, as identified from the workshops in chapter 4. The findings of the study show that visualizations separating different trade-off metrics lead to higher mental load and lower self-reported trust, but may support more varied decision strategies than integrated visualizations.

**Chapter 6- Discussion and conclusion** summarizes the contributions of this thesis and discusses possible directions for future work.

## 1.4 Publications and Papers Submitted

### PAPERS IN INTERNATIONAL JOURNALS

1. **M. Chakhchoukh**, N. Boukhelifa, and A. Bezerianos. 2022. Understanding How In-Visualization Provenance Can Support Trade-off Analysis. *IEEE Transactions on Visualization and Computer Graphics* (May 2022). <https://doi.org/10.1109/TVCG.2022.16> pages.

### WORKSHOP PAPERS AND POSTERS

2. **M. Chakhchoukh**, N. Boukhelifa, and A. Bezerianos. Visualizations to support trade-off comparisons. *Journée Visu*, 2023. 3 pages. <https://hal.science/hal-04719331>

3. J. Rogers, M. Anastacio, J. Bernard, **M. Chakhchoukh**, R. Faust, A. Kerren, S. Koch, L. Kotthoff, C. Turkey, and E. Wall. Visualization and Automation in Data Science: Exploring the Paradox of Humans-in-the-Loop. *Int. Workshop on Visualization in Data Science (VDS, at VIS'24)*, 2024. 5 pages.

### SUBMITTED PAPERS

4. **Mehdi Chakhchoukh**, Anastasia Bezerianos, and Nadia Boukhelifa. Visualizations to support trade-off comparisons between groups. [*submitted*] to the 2025 CHI Conference on Human Factors in Computing Systems. 30 pages

## Related Work

This thesis studies how to best support trade-off analysis using visualizations. Our aim is to support domain experts, such as ones that come from the agronomy domain, in their task to explore trade-offs in high-dimensional datasets, generated from simulations and optimized by multi-objective algorithms. Their goal is to understand the relationships between the different dimensions and to identify interesting patterns in their data.

As such, we first review general research on visual analytics and sensemaking, and then trade-off analysis and decision-making, before focusing more specifically on multidimensional data visualization and relevant decision-making support visualization tools. We then provide more detailed related work on the specific topics that we address in the thesis, namely work on *provenance* visualization as we use it to support trade-off analysis, as well as work on *comparison* visualizations<sup>1</sup>.

### 2.1 Visual Analytics and Sensemaking. A focus on Agronomy.

We next introduce the basic notion of Visual Analytics as our own work is positioned within it and its relation to Sensemaking. We additionally introduce visual analysis work in the domain of agronomy, whose experts we want to support.

#### 2.1.1 Visual Analytics

In their seminal work “Illuminating the Path,” Thomas and Cook [64] define Visual Analytics (VA) as *“the science of analytical reasoning facilitated by interactive visual*

---

<sup>1</sup>For all images used in this chapter, we include references to the original publication the image appeared in. Under French law, research documents may reproduce up to 10% of a work, provided it is properly referenced, without requiring explicit authorization, see **Protocole d'accord sur l'utilisation et la reproduction des livres, des œuvres musicales éditées, des publications périodiques et des œuvres des arts visuels à des fins d'illustration des activités d'enseignement et de recherche** From Bulletin officiel n35 du 29 septembre 2016 article Article 3.2.2.2 & 4.2.1 ([111])

*interfaces*". Visual Analytics as a domain is vast, including topics that range from understanding how to create and study visual tools that support reasoning and sense-making, to what low-level interactions or data transformation best support the analysis process. Later definitions by Keim et al. [28] are more specific, discussing how VA can combine automated analysis with interactive visualizations in order to support understanding, reasoning, and decision-making on complex data sets. They go on to explain that visual analysis is more than just visualization, but rather it is "*an integral approach to decision-making, combining visualization, human factors and data analysis*".

Our own work falls under the visual analytics umbrella, focusing on how to support a specific type of analysis and decision-making, that of *trade-offs* (discussed in more detail in [section 2.2](#) on prior work on trade-off analysis). In our work, we focus on the choice of visual representations for trade-off analysis (rather than automated analysis) and the impact these have on the analysis process. Nevertheless, as we will discuss in the section on trade-off analysis, this type of analysis often assumes an automated step to reduce the size of the trade-off simulation datasets used by domain experts, specifically through the creation of Pareto front datasets (see [section 2.4](#)).

### 2.1.2 Sensemaking in Visual Analytics

Since the goal of Visual Analytics (VA) is to support analytical reasoning facilitated by interactive visual interfaces [64], a lot of research in this domain has focused on understanding and characterizing the sensemaking process, drawing on theories from cognitive psychology and cognitive systems engineering. In this section, we will briefly present two such theories that are widely used in visualization and VA research, each focusing on different aspects of sensemaking: the information foraging model by Pirolli and Card [121], and the data-frame model by Klein and Card, before discussing the top-down perspective on sensemaking proposed by Liu and Stasko [156].

Very widely used in visualization and VA research, the sensemaking model proposed by Pirolli et al. [121] (see [Figure 2.1](#)) focuses on visual sensemaking and is derived from the cognitive task analysis of intelligence analysts. This model organizes the sensemaking process into two main loops: a *foraging loop* where analysts seek information, filter it, and gather relevant data. Analysts search through vast amounts of raw data, filter it for relevance, and collect pieces of information (called "nuggets") for further use. And a *sensemaking loop*: The goal of the analysts in this loop is to make sense of the information they gathered by developing a conceptual schema, generating hypotheses, and transforming raw data into actionable insights. Analysts iteratively build mental models, or schemas, to represent the information in ways that aid decision-making.

Drawing on the iterative nature of sensemaking but with a greater emphasis on the role of mental models, Klein et al. [45, 46, 47] propose the Data-Frame Theory of Sensemaking, which outlines two key cycles involved in interpreting complex data: the elaboration cycle and the reframing cycle. The *elaboration cycle* occurs when a

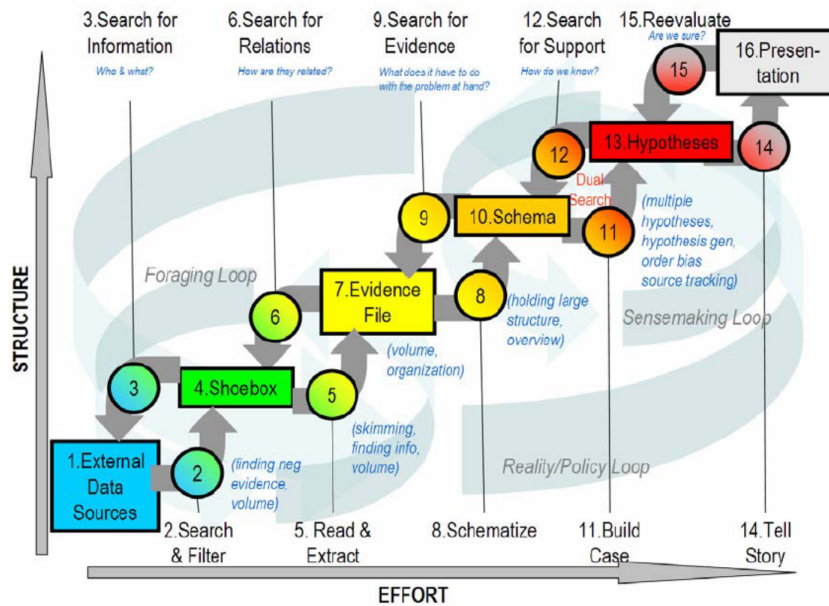


Figure 2.1: Pirolli and Card's model of sensemaking loop for intelligence analysis (Reproduced from [121]) where we notice the iterative nature of the process and the correlation of structure with effort.

frame is confirmed by the data, prompting individuals to seek additional information that fits and enriches the existing frame. On the other hand, the *reframing cycle* happens when new data challenges the current frame, leading individuals to question and potentially discard it in favor of an alternative frame that better explains the situation. This dynamic interaction between data and frames underpins the iterative nature of sensemaking, where people continually adjust their mental models in response to evolving information.

Liu and Stasko [156] further elaborate on the role of mental models in sensemaking, and the dynamic interaction and interplay between mental models and external visualizations in the context of information visualization. They argue that mental models are internal representations that help users make sense of data and external visualizations. Mental models are functional analogs of external visual systems, preserving structural, behavioral, and functional properties that allow individuals to simulate and reason about data in their minds. The authors propose four key relationships between mental models and external visualizations: internalization, processing, augmentation, and creation, emphasizing the importance of interaction in model-based reasoning. Interaction, according to the authors, serves purposes such as external anchoring, information foraging, and cognitive offloading, which help users reduce cognitive load

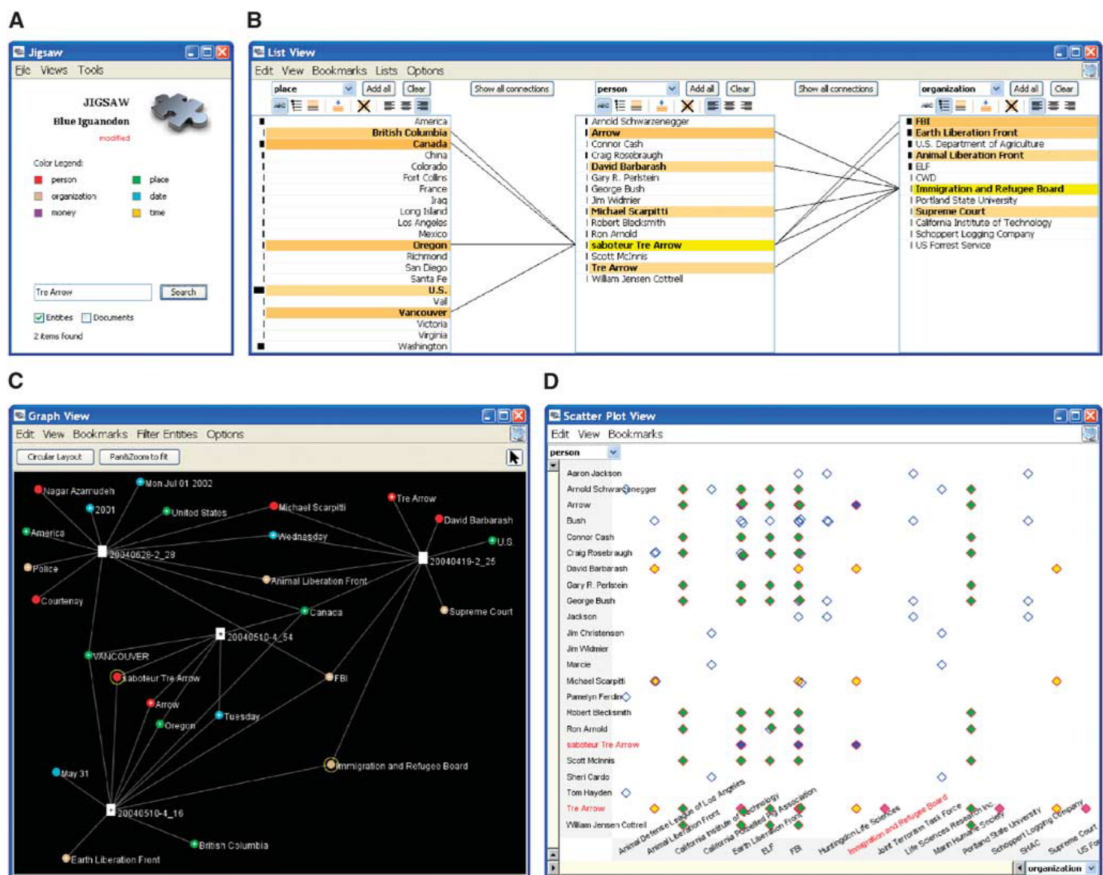


Figure 2.2: The Jigsaw system for supporting investigative analysis (Reproduced from [72]): In the list view (B) the yellow shading and the diagonal lines indicate the selected elements. The darker the color is the more related the element highlighted to the element selected. This list view is displayed along a (A) Control Panel, (C) Graph view and (D) a Scatter Plot View.

while engaging in data analysis.

An example of a VA tool whose design was inspired by these sensemaking theories and models is Jigsaw [72] (Figure 2.2), a visual analytics tool for investigative analysts when exploring and connecting evidence across large collections of textual documents. It is grounded in the cognitive task model of foraging and sensemaking, as its main purpose is to help analysts make sense of large amounts of data by providing a visual interface to support their cognitive processes.

We note that more generally when sensemaking is studied in VA, it is usually with one of two goals in mind. First, to use sensemaking theories to gather insights on user behavior, which can then inform the design of future VA and interactive tools

that better support the sensemaking process. For example, Dumais et al. [136] study how logs can be used to understand user behavior on a large scale without direct interference; this even includes behaviors that users might not recall or report correctly. Second, improve the performance of the data analysis processes by informing users about their own behavior and decision-making processes. For example, Kadivar et al. [116] introduce CzSaw, a visual analytics tool developed to support the analysis process in situations where analysts need to make sense of large, complex document collections. To do so, this tool enables analysts to review, edit, and replay their analysis steps through a visual history. CzSaw is designed to capture and visualize the steps analysts take during their investigations. Doing so provides a clearer understanding of the logic and tracking of the progression and alternatives explored in the analysis, helping analysts refine their conclusions and make better-informed decisions.

In our own work, we do not explore these sensemaking models in depth. However, similar to previous studies, our goal, when trying to understand the sensemaking process of domain experts during trade-off analysis, is to inform the future design of decision support tools. Additionally, prior research on sensemaking in visual trade-off analysis [112] demonstrated that experts seek to align external evidence (such as data and visual representations) with their internal knowledge of the studied systems—in other words, their expertise and mental models of those systems.

We will elaborate on how we studied expert sensemaking during trade-off analysis in [Chapter 3](#), through the thematic analysis of participants' verbalizations from a think-aloud protocol, and in [chapter Chapter 5](#), where we analyze participants' written descriptions of their reasoning and decision-making processes. The analytical processes and expert sensemaking we observed, particularly in the provenance project, confirm the iterative nature of sensemaking during trade-off analysis.

### 2.1.3 Visual Analytics and Decision Support in Agronomy

Some work in the field of data visualization and decision-making has studied the biases that influence decision-making processes. For example, in their work, Dimara et al. [36] provide a taxonomy for cognitive biases based on different user tasks and explore how these biases can influence decision-making even when the information is well-visualized. They highlight that decisions can sometimes be irrational despite a proper understanding of the data by the decision-maker due to the influence of irrelevant information. Their work also examines ways to mitigate these biases through design-based debiasing techniques in visualizations [35].

Although it is important to take into account such biases, our work is more related to another large body of work in decision-making and visualization that surrounds the creation of **sensemaking and decision-support tools**. Prior work [33] has a detailed review on visual decision support tools. We focus in our related work on visualizations and tools relevant to trade-off analysis ([section 2.3](#) and [section 2.4](#)).

As we aim to support the specific needs in the domains of agronomy, we will briefly review decision-making support tools in agriculture and agronomy. Agronomy is a subfield of agriculture. Agriculture is often referred to as the "art and science" of cultivating land, rearing animals, and management of livestock in order to produce plants and livestock for food, fiber, and other products. As such, the needs and shareholders are varied. In a recent review, Gutierrez et al. discuss [43] how stakeholders can be farmers, advisers, and policymakers who need to use software tools to facilitate tasks as diverse as farm management of resources, monitoring environmental sensors on vineyards, identifying and treating pests, etc. Thus, the decision support and visual analysis tools for that domain are equally broad. The survey explains that most visualizations used in these systems are 2D maps (for agricultural areas), time series (for sensor data in farms and vineyards), heatmaps (representing different types of information, such as affected areas, temperature, etc.), and bar charts (e.g., presenting histograms of different metrics).

While our own work is in Agronomy (discussed next), we are inspired by a recommendation from this literature review work in agriculture. Gutierrez et al. [43] explain that tools are often poorly received by the end-users, as they lack input from them during the design process. For example, only 11 of the 61 tools reviewed benefited from participatory design. Thus, they advise for human-centered design approaches to be used in the design of such tools. In our work, we took that into account and used a user-centered design approach to design our visualizations based on multiple interviews and workshops with domain experts and potential end-users.

Agronomy was originally defined as the study of crop production, with an emphasis on the interaction between climate, soil, agricultural techniques, and the resulting crop yield and quality [38]. This discipline draws from a variety of scientific fields, including biology, chemistry, soil science, ecology, and genetics. Over time, agronomists shifted their focus to the farmers responsible for implementing agricultural practices, as they investigated how these practices affect crop outcomes. As Lichtfousse et al. explain, this change brought new challenges in modeling farmer behavior and understanding the consequences of their decisions [38].

In our work, we collaborated with researchers in agronomy, primarily from INRAe (Institut national de recherche pour l'agriculture, l'alimentation et l'environnement) and affiliated institutes. These agronomists often analyze results collected in a structured form from field studies and communicate their results using simple multi-dimensional visualizations. For example, the results of agronomic field trials [8] can be expressed as scatterplots, bar charts, histograms, etc. Or they are confronted with models that simulate biological processes. They may manipulate existing wheat growth models [114] to explore how, for example, a late fertilization impacts wheat yield and quality, and if the outcome is impacted by weather conditions. These simulation models can generate thousands or millions of datapoints that are often presented in the form of large



multi-dimensional tables [112] and have been visualized in the past using scatterplot matrices.

No matter where their data originate from, their work attempts to reconcile multiple competing objectives [44] often in a single investigation. For instance, agronomists may search for fertilization strategies that on the one hand maximize yield, and on the other hand reduce the amount of supplied fertilizers and the nitrogen loss to the plants [155]. The approach they use to understand these trade-offs falls under the domain of Multi-Criteria Decision-making (MCDM), discussed in the next section. Our domain experts come from this field.

## 2.2 Trade-off Analysis

A specific type of analysis is that of trade-offs. In domains such as agronomy or engineering design, experts often explore the results of large multi-dimensional search spaces derived from model simulations. Their goal is to identify solutions (or simulations) that fit a set of criteria they have, that can be competing - representing *trade-offs*. This goal is related to multi-attribute choice tasks that focus on finding the best option out of a set of alternatives defined across multiple attributes [34, 54]. Nevertheless, trade-off analysis does not consist of finding one single best option, but rather of identifying a set of "good" options that may differ greatly in what criteria they optimize. In such cases, the decision-maker must balance conflicting choices, weighing each by its strengths and weaknesses.

### 2.2.1 Trade-off Analysis and MCDM

Trade-off analysis is a process that falls under Multi-Criteria Decision-making (MCDM), a discipline that studies procedures to aid decision-making when dealing with a large number of alternatives. MCDM problems generally involve the ordering or classifying of alternatives [39], by imposing a procedure or strategy, like explicitly ordering or weighting criteria [14]. Typically, MCDM problems are addressed by ranking and classifying alternatives using mathematical functions or heuristics [14, 39] to calculate the best possible solutions, referred to as "metrics" in this thesis and discussed next.

Trinka et al. [50] distinguish between three types of multicriteria decision tasks: black-box, white-box, and hybrid. They explain that white-box multicriteria problems have well-defined descriptions and are completely specified. For example, multicriteria optimization methods fall into this category as their goal is to find the best solutions out of all possible options, according to known criteria (objectives) and constraints. However, if nothing is known about the structure or dependencies within the multicriteria problem, it is classified as a black-box problem. In this case, solutions are typically discovered through trial and error, or by employing heuristics. Trinka et al.

conclude that most real-life decision-making falls between the extremes of black and white, leading to what they call hybrid multicriteria problems.

As mentioned in the introduction, in our work, the term trade-off analysis describes a process that is less structured than general MCDM, one that is more exploratory, where analysts consider multiple alternatives in order to understand the different criteria they optimize, without any imposed procedures or strategies. This less structured process assumes that the description of the potential decision that the experts are trying to make is not completely known or defined a priori: the criteria and priorities experts have are not fixed and may shift over time as they explore their data. In this thesis, our focus is thus on hybrid decision problems, which reflect the nature of the real-life applications that inspired our research, namely from agronomy. As such, we need interactive visualizations to allow experts to explore the trade-offs in their data.

## 2.2.2 Trade-off Analysis Tasks

We discuss next work that has identified tasks conducted during trade-off analysis, as part of our own work in this thesis is to characterize these tasks so as to better support them with visualization solutions ([Chapter 3](#)).

Trade-off tasks feature in many visualization and visual analytics systems, in particular where users, via an interactive interface, have to make decisions according to multiple criteria. Some systems treat trade-off analysis as a single coarse task. For example, Sedlmair et al. [106] analyze tasks that users engage in when conducting visual parameter space analysis. They describe an optimization analysis task where the goal is to find the best parameter combination given some objectives.

Booshehrian et al. [94] make use of visualization to understand trade-offs of a set of precomputed simulations. They describe a task within the analysis workflow of scientists working in fisheries management, where the goal is to quantify trade-offs between selected options, avoiding sensitive regions of the parameter space and those with high uncertainty. Their work distinguishes between trade-offs that are globally distributed in the output space and those that are focused on a local neighborhood.

Matkovic et al. [81] describe typical tasks and propose an interactive analysis workflow for complex engineering systems. The identified tasks are grouped into three main categories: exploration and analysis, regression model computation, and data generation. Many of these tasks relate to automatic workflows, rather than human analysis. However, some are mirrored in human analysis. For example, in the category of exploration and analysis, we find the task of comparing results across different regions of the parameter space, a subtask we also identified with our domain experts.

We note that Matkovic et al. [81] in the identified tasks include several steps of generating the data, one being automatic optimization. This is of relevance to our work, as domain experts often explored optimized datasets. Some of their simulation experiments need to sample a very large solution space; hence, very frequently, a multi-

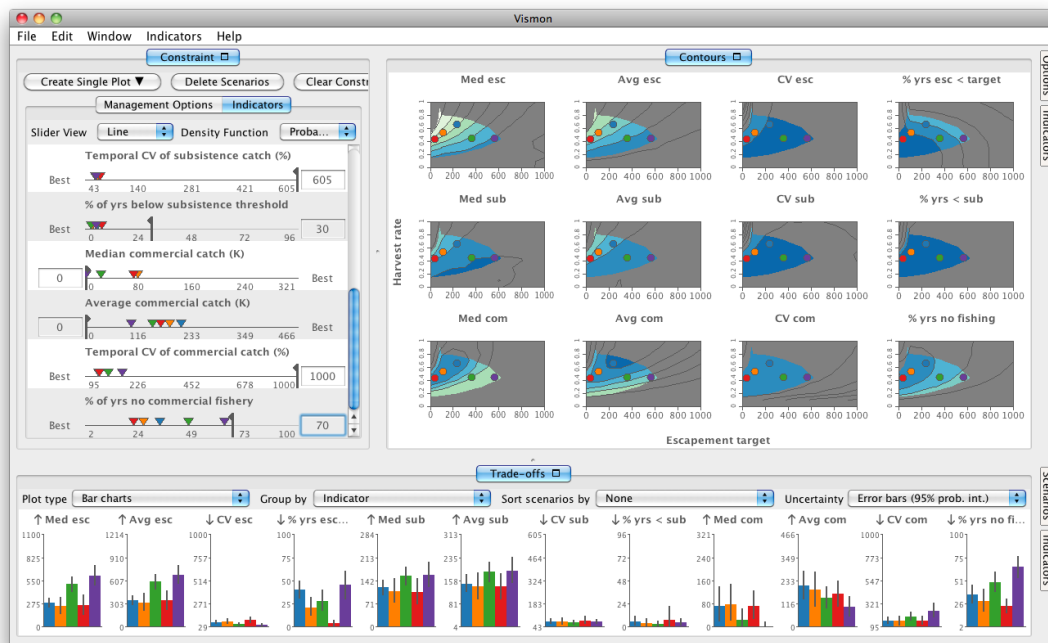


Figure 2.3: The Vismon interface [94] (Reproduced from [93]) consists of three key components. The top-left Constraint pane, which organizes management options and indicators into separate tabs. It also provides individual sliders, designated for adjusting management options and indicator. The top-right Contours plot matrix pane, displays contour plots of indicators as they relate to two management options, and allowing for scenario selection. And the bottom Trade-offs pane, provides detailed information on indicators for the selected scenarios.

objective optimization algorithm (e.g., NSGA-II [77]) is used to explore this solution space, generating Pareto Front datasets, which are then analyzed by domain experts.

**Pareto front datasets** are a subset of an original dataset computed from multi-objective optimization. The resulting, smaller dataset typically indicates that in the remaining points no objective can be improved without sacrificing at least one other objective [52]. Pareto front points (see Figure 2.4) are often called non-dominated solutions, as no other solution is strictly better. All points from the original dataset that are not part of the Pareto front are sometimes called dominated, as there is at least one point in the Pareto front that dominates them (i.e., is better in one or more criteria).

Other works break down trade-off analysis into smaller sub-tasks, often focusing on particular exploration strategies such as ranking of alternatives or weights to express preferences. For instance, in the LineUp system [132] (seen in Figure 2.5) users can

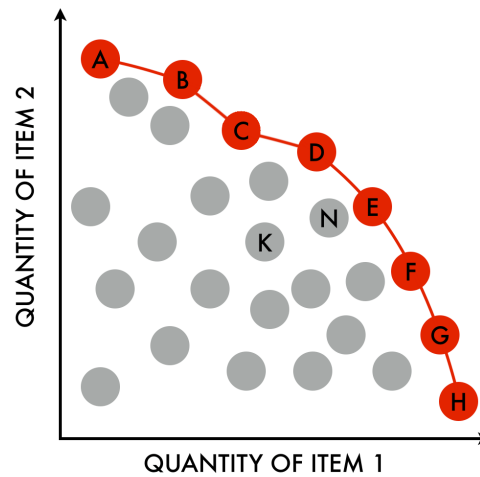


Figure 2.4: Pareto Front across two dimensions (quantity of Item 1 and quantity of Item 2) (Reproduced from [149]). Each red point (A-H) represents optimal combinations of two items, while gray points (K, N, etc.) represent less efficient choices. The frontier shows the trade-offs between the quantities of Item 1 and Item 2, where moving along the curve improves one dimension at the expense of the other.

rank solutions and evaluate their performance relative to each other. The tool also provides means to combine dimensions (or attributes) and weights to convey priorities for the different attributes. Similarly, Weightlifter [135] (seen in Figure 2.6) helps analysts understand the effects of different weights for multi-criteria decision spaces with up to ten criteria. We revisit these systems later in our thesis.

Work that breaks trade-off analysis tasks into finer components or sub-tasks can also be found outside the visualization and visual analytics community. The work of Hakanen et al. [59] from operations research is the closest to ours, as they assume that the priorities of experts may shift. They propose an interactive visual analytics system where a decision-maker explores Pareto front solutions that are computed iteratively, based on user preferences. As such, the authors also discuss trade-off analysis in the context of interactive multi-objective optimization and identify seven high-level tasks: compare Pareto optimal solutions, specify preferences, check feasibility of preferences, determine the most preferred solution, learn about problem characteristics, detect correlations, and post-process the most preferred solution.

In our work, we consider both general trade-off datasets and those generated by an a-posteriori multi-objective optimization method, which tends to produce larger sets of pre-computed solutions. We hypothesize that most of their tasks [59] are still relevant for both general trade-off datasets (not just Pareto front solutions) and for a-posteriori methods such as ours, with the exception of the tasks “specify preferences” and “check

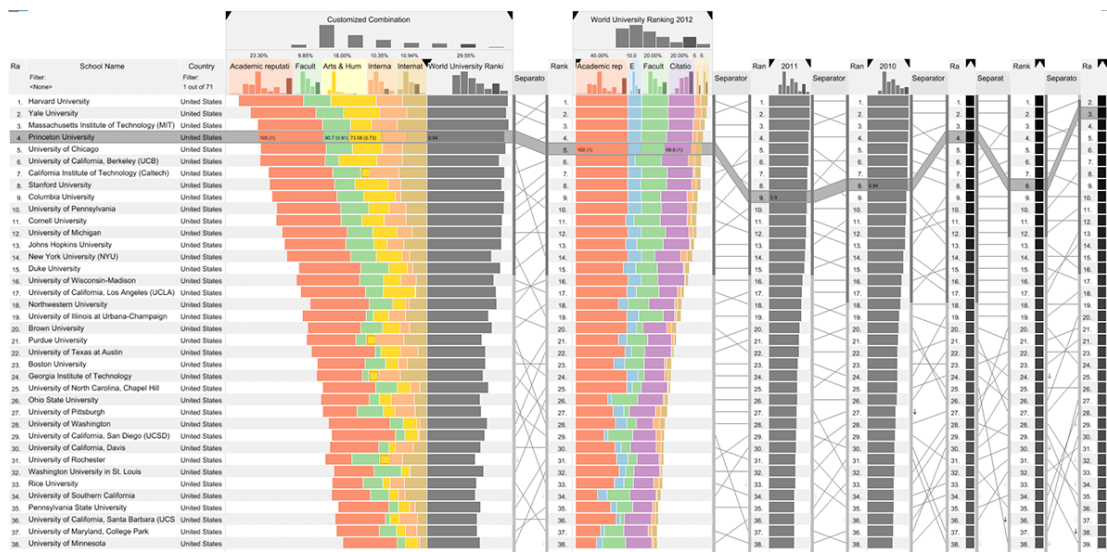


Figure 2.5: The Lineup system (Reproduced from [132]) for ranking Multi-Attribute alternatives and comparing the rankings of these alternatives. Each line represents a different alternative and colors indicate the weighted dimension. In this image, we see multiple iterations of rankings (as different columns), created from different weight configurations.

feasibility of preferences,” which are specific to interactive multi-objective optimization. To the best of our knowledge, trade-off analysis tasks have not been described in visual analytics context specifically. Our goal in this thesis is to shed light on what trade-off analysis tasks domain experts are interested in, and how visualization can assist those tasks. We revisit trade-off tasks in Chapter 3.

### 2.2.3 Summary for Trade-off Analysis

We were inspired by prior work on defining trade-off analysis needs and tasks, attempting to study how visualizations can support trade-off analysis. To the best of our knowledge, no prior work has looked at the effect of *provenance* visualization on trade-off exploration (Chapter 3), nor of *group comparison* visualization during trade-off analysis (Chapter 4 and Chapter 5). But before we expand on these two topics that relate to our main contributions, we present general work on visual analytics that relates to trade-off analysis.

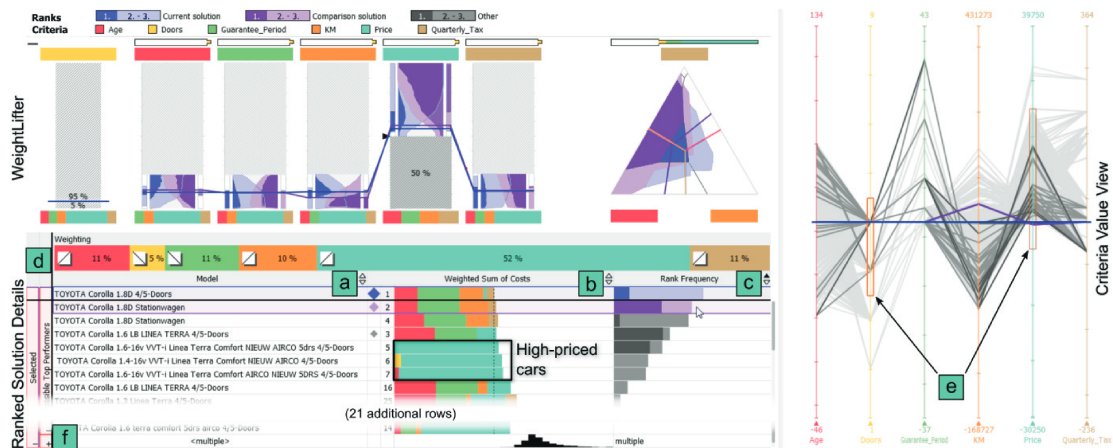


Figure 2.6: The Weightlifter (Reproduced from [135]) view (top left) is integrated with other visualization tools to aid ranking. In this case, we see at the bottom a ranking view with details on multiple data points, spread over multiple dimensions (columns a, b, c), and a row (d) of the current weights. On the right (e) we see a parallel coordinates plot (e) of the same datapoints. All views are linked. Excluded solutions are shown as well for context (f).

## 2.3 Multidimensional Data Visualization

The type of data used in trade-off analysis in domains such as agronomy is often the result of simulations in the form of high-dimensional (or multi-dimensional) datasets. These are usually long tabular datasets of input and output parameter values. Given this type of data, we focus next specifically on work in visual analysis for Multidimensional Data.

In their literature review on visualizing high-dimensional data, Liu et al. [134] present the pipeline for visual analytics that describes the process of going from raw data to views. The pipeline consists of the following stages: Data transformation (techniques to manipulate the underlying data), Visual mapping (encoding data derived from the data transformation to visual representations on the screen), and View Transformation (methods that focus on screen space and rendering) A crucial part of the pipeline is human interaction that is integrated at each stage. Among others, they discuss how it is challenging to visualize high-dimensional datasets that contain more than three dimensions (in practice often tens or hundreds of dimensions) and hundreds or thousands of data points. Past research has introduced different solutions within this pipeline to address this and other challenges.

For example, during the *Data transformation* phase, one approach is to apply dimensionality reduction techniques [75, 99, 107] such as Principal Component Analysis

(PCA) [83], Multidimensional Scaling (MDS) [133], and t-distributed Stochastic Neighbor Embedding (t-SNE) that project high-dimensional data into lower-dimensional spaces while preserving the structure of the data. In our own work, we preserve the data in its original form, with the dimensions experts are familiar with, and we allow them to create their own combined dimensions. Nevertheless, our experts also often conduct other types of data transformations before engaging in the visual analysis of their data, that of calculating Pareto front datasets (discussed in trade-offs, see [section 2.4](#)).

Work on *View transformation* focus on the rendering that generates images in the screen space. For example, some approaches introduce visual representations that are either illustrative, adopting specific rendering, e.g., to highlight specific structures in the data (like the TableLens [127]). Others provide rendering abstractions to handle clutter (like using contours instead of rendering dense regions in a scatterplot [10]). Finally, some work under view transformation considers image space metrics to analyze the visual structure and automatically identify interesting patterns. A classic example is Scagnostics [90], a set of measures for identifying interesting plots in a scatterplot matrix (discussed next). In our own work, we do not consider View manipulations, although our visualizations could be combined with such techniques.

More relevant to our work is the *Visual mapping* stage, in other words, the choice of how to visually encode data. We thus review it next in more detail, motivating the visualizations we use as basis for our provenance technology probe ([Chapter 3](#)) and comparison prototype tool ([Chapter 4](#)).

### 2.3.1 Visual Mapping in Multidimensional Data Visualization

In the survey by Liu et al. [134], we see under the Visual mapping stage a set of visualization techniques focusing on different visual representations of multi-dimensional data. These include: Axis-based, Pixel-oriented, Glyph, Hierarchical, Animation, and Evaluation techniques. An older organization of this work also appears in Keim and Kriegel's VisDB work [26]. Here we follow more closely the categorization by Liu et al., and we additionally discuss past literature that has compared different visualization approaches. We note that the most relevant category for us is *Axis-Based techniques* which we expand on in more detail at the end of this section, but we briefly mention the remaining categories.

*Pixel-oriented techniques*, like those used in their [26], represent data points as pixels with color indicating relevance, allowing for the visualization of millions of items simultaneously (see [Figure 2.7](#)).

As we work with domain experts, we decided to work with more commonly used visual mappings.

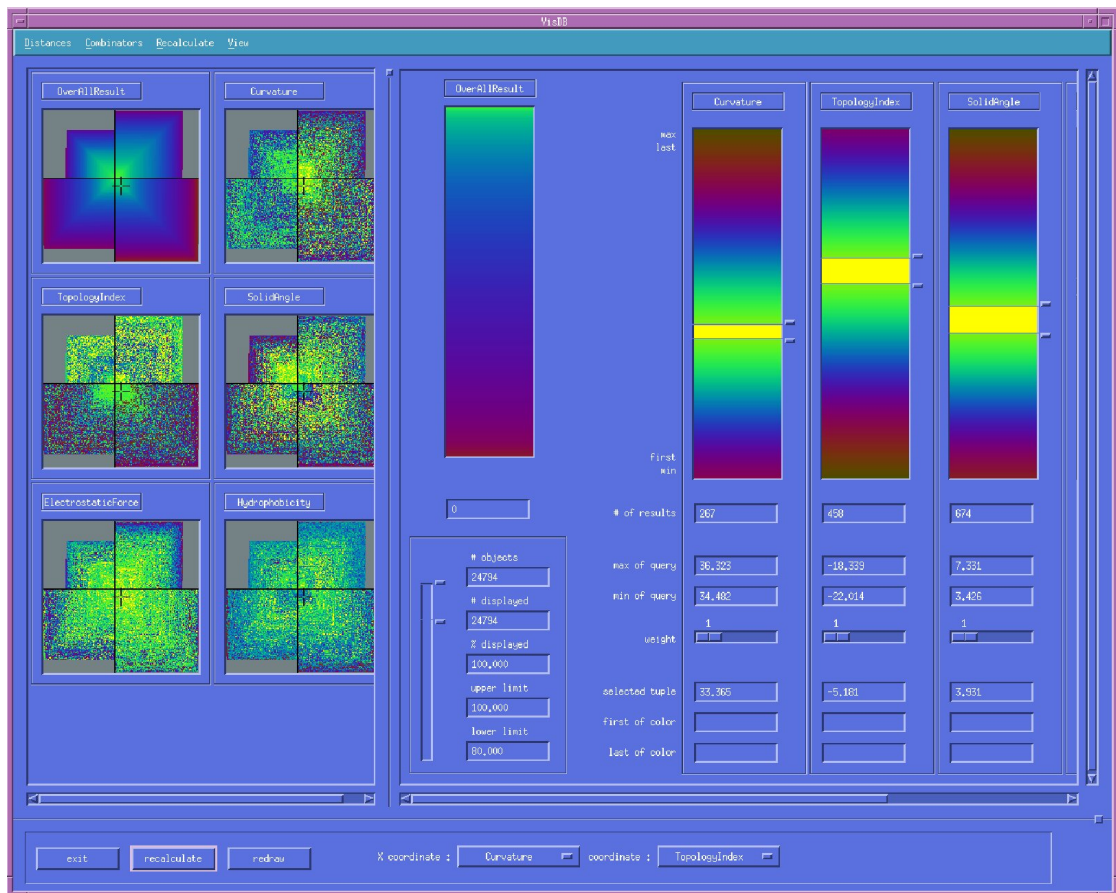


Figure 2.7: VisDB (Reproduced from [26]): uses pixel-based visual encodings to represent high-dimensional data (on left). Each small pixel corresponds to a data point, with different views such as Curvature, Topology Index, and Solid Angle mapped onto heatmaps.

*Glyphs* (that are also known as Icon-based methods [26]) include Star glyphs or Chernoff faces [49]. Glyphs map dimensions to visual properties such as shape or facial features, helping users intuitively identify patterns. For a review on glyphs and their applications, see [126] and for their evaluations, see [58]. Similarly to pixel-oriented techniques, these approaches are less familiar to domain experts.

*Hierarchical techniques*, including dimensional stacking and treemaps [27], organize data into nested structures for better exploration of hierarchical datasets. As the simulation data we are considering in our work is not inherently hierarchical, these are less relevant to us.

The *Animation* category discusses work that uses animation to aid understanding, such as Rolling the Dice work [115] that uses animation to transition between



scatterplots in a scatterplot matrix (discussed next).

The most relevant category to us is the collection of *Axis-Based techniques*, as well as relevant results from the *Evaluation* category of literature that explores a variety of prior studies that have compared different visual mappings.

### Axis-Based Visualization techniques

*Axis-based methods* are visual mappings where datapoint relationships are shown through axes representing the data dimensions [134]. These can be considered as being lossless geometric projection approaches according to Keim and Kriegel's grouping [26]. Axis-based methods are the most numerous approaches of multi-dimensional visualizations and the ones that are more ubiquitous in their use. This is why they have inspired the designs used in our own work. We report next on the most common and close this section with results from studies that have compared them.

**Parallel coordinates and Radial graphs:** Parallel coordinates plots visualize multidimensional data by mapping simultaneously multiple dimensions onto a 2D plane (for a review, see [62, 65]). Each dimension is visualized as a parallel axis. Data points are represented as polylines that intersect the axes at corresponding values, making it easier to identify patterns, correlations, and clusters across multiple variables (for a review of the topic see [62]). Parallel coordinates enable users to visualize relationships between dimensions and the distribution of items on the different dimensions, but do not scale well because of the visual clutter induced by a high number of data points [65]. Radial graphs [37] are analogs of parallel coordinates where the axes are arranged circularly, in a radial layout.

**Scatterplot matrices (SPLOMs):** A scatterplot encodes two dimensions as orthogonal axes, and datapoints are rendered as points on them. As a scatterplot can only show 2 (or 3) dimensions, a common way to extend them to multi-dimensional datasets is using a scatterplot matrix, or SPLOM. A SPLOM is a grid of scatterplots where each plot represents two dimensions of the dataset [74]. SPLOMs allow users to visualize pairwise interactions between different pairs of dimensions by plotting multiple scatterplots side by side. The number of scatterplots in a SPLOM increases quadratically as the number of dimensions grows [134]. To address this, several solutions have been proposed, such as methods for automatically or semi-automatically identifying interesting plots within the SPLOM, such as incorporating Scagnostics [90] or with a metric like the trade-off index [109] that we will discuss in the Trade-off analysis section 2.4. Finally, some tools improve on SPLOM visualization with animations that either provide visual transitions across the different scatterplots in the matrix [115], or that combine them with parallel coordinates [25].

**Tabular visualization and stacked bars:** A set of visualizations that is not discussed extensively in past surveys are tabular visualizations. They are used to represent data in a table format, where rows represent items and columns represent attributes. These tables can be enhanced with representations of data attributes in the cells using encodings such as color [104] to construct heatmaps, size [54], or shape [21] to improve the readability and/or richness of the visualization. They can be considered as axis visualizations, as each column of the table represents an axis (a dimension). In systems that incorporate tabular visualizations, many [54, 104, 132, 135] display the values with bars inside table cells. Bar lengths are easier to compare than color or size [146]. In these systems, colors are used as a secondary encoding (e.g., for categorical information). For more information on the perception of bar charts, please see [73, 91, 130].

No matter the visual mapping, we note the visualization techniques, including the axis-based ones we mentioned here, are accompanied by **appropriate interaction techniques**. These include highlighting individual points and linking them across views (e.g., [54, 112, 115, 132]); selections of multiple datapoints through visual querying and sculpting or range selection across dimensions (e.g., [112, 122, 132]), filtering (e.g., [112, 115]), and axis reorganization (e.g., [115, 132]). We provide such support in our own visualizations in this thesis.

**Evaluation** We summarize here previous work that has compared these visualization (visual mapping) techniques under different tasks, explaining why we integrated them in our own work. In their work Kuang et al. [151], have compared parallel coordinates with SPLOMs and found that SPLOMs are better at identifying correlations than parallel coordinates and they were better suited for a higher number of dimensions (4 or over) and more than 30 datapoints (30). Later work on decision-making [34] has also highlighted that SPLOMs performed better than parallel coordinates in high-level tasks such as identifying correlations and relationships between dimensions, which is key when attempting to understand trade-offs. Thus, in our work that focuses on trade-off exploration, we use scatterplot matrices (SPLOMs) as the basis for our exploration tool (Chapter 3). The work of Dimara et al. [34] additionally highlighted that tabular visualizations are well suited for decision-making tasks, in their case for choosing one among several multi-dimensional datapoints. Following their recommendation, in our work studying comparison for trade-off analysis and decision-making, we adopt tabular visualizations (Chapter 4).

### 2.3.2 Summary on Multi-Dimensional Visualization

In our work, we focus on the use of visualizations to support trade-off analysis tasks, a form of multi-criteria decision-making, i.e., decision-making that considers a large number of dimensions (as discussed in the Trade-off [section 2.2](#)). The extensive body of work on visual analytics and multi-dimensional visualization proposes several tools and visual mapping approaches. Among them, we chose two for our own purposes. We use SPLOMs to visualize and help explore the trade-off datasets as they allow experts to see relationships between different dimensions and identify interesting patterns in the data, such as correlations. And for our work on comparison and decision-making, we use tabular visualizations enhanced with colors (similar to heatmaps) and bars, as related work has suggested that tables are appropriate visual representations for decision-making.

## 2.4 Visualization tools for Trade-off Analysis

Trade-off datasets are almost always multi-dimensional, covering objective dimensions to be optimized, and whose values vary in relation to other dimensions (or parameters). Thus, it is not surprising that many existing visualization techniques (visual mappings) discussed before have been used in tools for trade-off analysis. These include tabular visualizations [71], parallel coordinates [92], scatterplot matrices [9], and self-organizing maps [129]. For an overview of visualization techniques for trade-off sets, see [82, 140]. We report next on the ones that use some of the most common multi-dimensional visualizations (described before) and are thus relevant to our work.

We report first on three of these tools that have been used to either construct or explore Pareto fronts ([subsection 2.2.2](#)).

In their work, Gratzl et al. [92] compare two techniques to visualize Pareto fronts used during the process of designing the optimization procedure for generating the Pareto front. They use parallel coordinates and the projection of the Pareto front over a 2D space as a scatterplot. Their results show that parallel coordinates are better than scatterplots, but only if the number of items in the Pareto front is small.

Chen et al. [129] use self-organizing 2D maps to represent the balance of objectives in solutions in Pareto front datasets. To do so, they represent each solution with a radial glyph. These glyphs are then organized on top of a 2D self-organizing map of the different objectives, where the size of the objective maps to a weight defined by the user. This visualization attempts to support both the adjacency of solutions, where similar solutions are placed close to each other, but assumes a fairly small number of solutions and dimensions.

EvoGraphDice [6, 113, 115, 145] is a SPLOM-based multidimensional visual exploration tool. It includes several interactive tools such as brushing and linking across

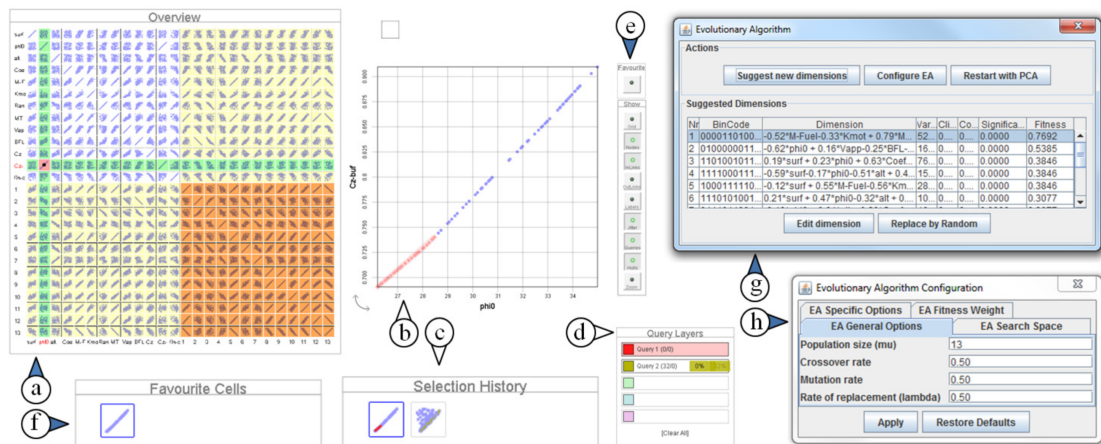


Figure 2.8: EvoGraphDice (Reproduced from [145]): Users can select groups of points using a lasso tool on the scatter plot (b). Selecting points assigns them a color so that users can observe their positions within the SPLOM(a) visualization. This allows for easy tracking of selected points across multiple dimensions. Other features include a toolbar (e), a history (c), a favorites bookmarking location (f) and a visual query panel (f). The tool includes an evolutionary algorithm configuration panel (h). We notice the tabular view of the data (g), where users can also create new composite dimensions.

views, zooming, and visual query sculpting across different dimensions (see Figure 2.8). It provides views of the underlying data in the form of a table and ways to combine multiple dimensions. In addition, it recommends interesting views (based on Scagnostics), as well as possible dimension combinations to aid experts reach views that exhibit a particular pattern. It has been successfully used by domain experts in agronomy to explore large trade-off spaces that are Pareto fronts of simulation datasets [112].

We next report on more general-purpose tools that allow users to provide weights on dimensions and rank datapoints accordingly.

LiteVis [71] is a visualization to represent simulation data for lighting design. Apart from a 3D scene of the environment, their tool also provides a stacked barchart view. Each simulated lighting alternative is a stacked bar, with stacks representing different lighting parameterizations and abstracted result indicators. Users can provide weights to these indicators that are used to calculate a score for each simulation. And can rank simulation results (in the stacked bar view) based on these scores. According to their results, their tool was efficient in helping users to navigate the solution space that would have overwhelmed them otherwise. They also explain that the linking of both representations led to a shorter feedback loop for the users.

Another example, also using tabular and stacked-bar visualizations, is the LineUp

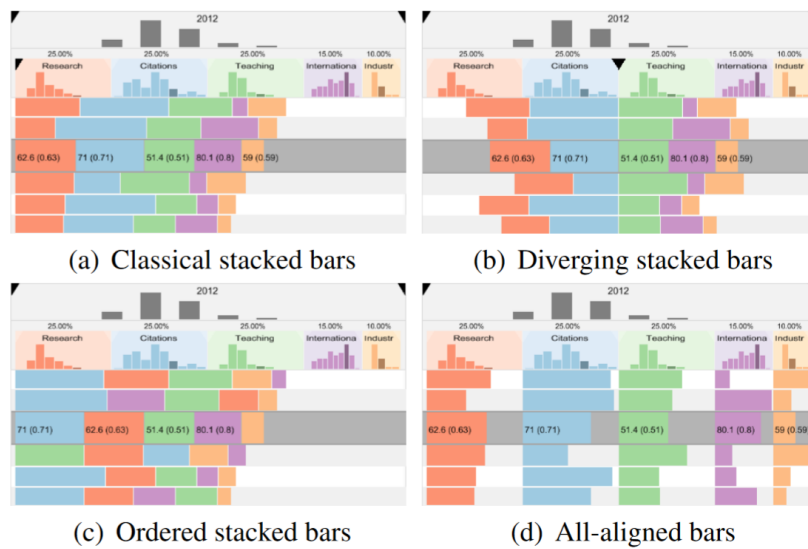


Figure 2.9: Different configurations of the LineUp tool bars (Reproduced from [132]). Notice how some configurations (a,b,c) are stacked bars, while others are tabular views (d). The full system can be seen in Figure 2.5.

[132] decision-making tool (see Figure 2.5 for system and Figure 2.9 for examples of tabular and bars configurations). It is an example of a visual analytics system designed to create, analyze, and explore rankings of items based on multiple heterogeneous attributes. But it has functionalities that are relevant for trade-off analysis. It uses weighted scores to compute an overall ranking by combining different attributes, which are then visualized in a stacked-bar representation by default. LineUp provides several views of the data. Some views are in tabular format, i.e., dimensions are aligned as columns and rows are datapoints, and internally, cells include bars of different lengths. And as the other views represent datapoints as stacked bars, i.e., each datapoint is a bar divided into segments of different lengths (one segment per dimension). In LineUp, color is used to differentiate dimensions in the stacked bars. They use these two views for different purposes; the tabular visualization with bars simplifies the comparison of cell-level values since the bars for each dimension have a shared baseline.

Conversely, the stacked bars allow for the comparison of the sum of the values for each item while still having an indication of the composition of the whole by stacking the bars on top of each other. LineUp [132] can be used in the context of simple trade-off analysis, as it allows users to give weights to different dimensions and to visualize a ranking between all points according to these weights.

Weightlifter [135] (Figure 2.6) assists analysts in understanding the impact of different weights for multi-criteria decision spaces with up to ten criteria. Weightlifter is

based on the core concept of ranking solutions using weights. It provides an interactive visualization that highlights the stability and change of rankings depending on how weights are adjusted and combined. The tool visualizes the weight space, which is a geometric representation of all possible combinations of criteria weights, with lines that represent changes in ranking, enabling users to see how changes in these weights affect the overall ranking of solutions. We note that each solution in their case, similarly to LineUp, is a single point.

Finally, a tool specifically built for trade-off analysis is PAVED [86]. PAVED (Figure 2.10) uses parallel coordinates to depict a large number of design alternatives and criteria simultaneously, allowing engineers to visually explore trade-offs and conflicts between different criteria. The tool supports tasks such as filtering alternatives, identifying superior compromises, and confirming design choices. This tool works best for contexts with a small number of criteria and alternatives, as it can become cluttered with larger datasets.

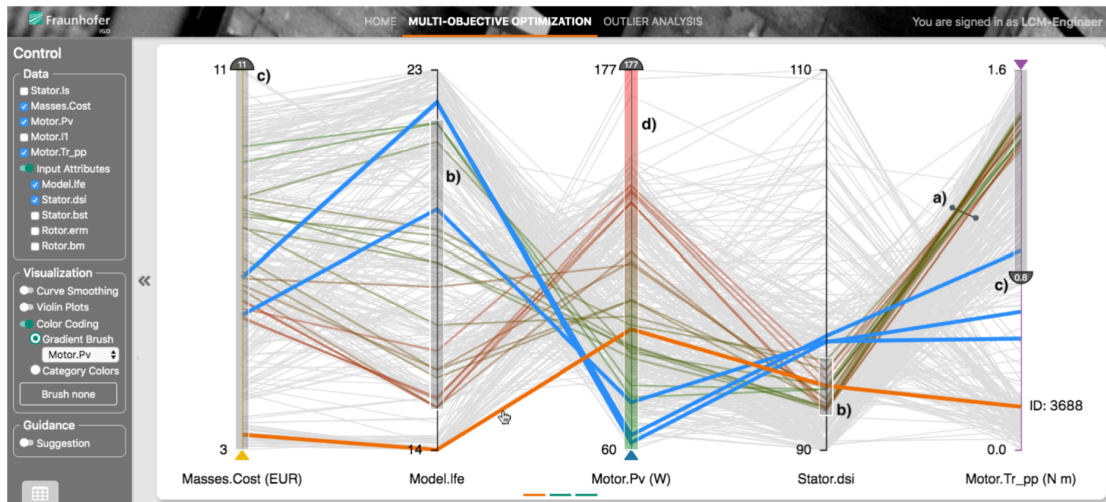


Figure 2.10: PAVED (Reproduced from [86]): Visualizes Pareto fronts from engineering, using parallel coordinates visualizations. The users can select the lines (datapoints) to highlight by brushing over the ranges on the axis.

Because trade-off exploration is a complex process [112], existing work looked at how to better support decision-makers during the analysis, such as in terms of powerful interactive visualizations as we discussed above, but also with robust ranking and optimization algorithms [9, 71, 135, 140, 152]. We will discuss some of these ranking metrics next. Tušar et al. [140], however, argue that trade-off sets (such as those found by evolutionary multi-objective optimization algorithms) require further support that traditional multidimensional visualizations do not necessarily cater for. Our work

investigates these additional needs and identifies relevant metrics, and hypothesizes that visualization of these can aid trade-off analysis.

### 2.4.1 Metrics to Quantify Trade-offs

Since the method for solving trade-off problems varies based on the context and criteria defined by the decision-maker, the majority of the tools mentioned above integrate some metrics to characterize trade-offs. These metrics often integrate user preferences, such as weighted means or other scoring methods.

Some of these metrics come from research from MCDM on quantitative methods for characterizing trade-offs, outside the visualization domain. For example, Daniels et al. [57] propose a trade-off analysis method for systems engineering that simplifies evaluating and comparing different designs. The method uses weighted mean aggregation to assign numerical scores, allowing easier comparison of options based on stakeholder preferences and priorities. Another example is Shih et al. [51] who present a method for ranking and evaluating MCDM alternatives based on the similarity of each alternative to an ideal solution.

Visualization researchers have also argued for using these quantitative metrics and mathematical functions when designing visual MCDM support tools. Amar et al. [125] explain that problems that are marked by uncertain criteria, of which trade-off analysis is a prime example, can be solved either through a black-box aggregation function or by employing distribution or clustering methods.

We can see such metrics included in visualization decision support tools. LineUp [132], uses weighted scores to compute an overall ranking by combining different attributes, which are then visualized in a stacked-bar representation. This allows users to view the rankings but also adjust the importance (weights) of individual attributes and observe how these adjustments affect the overall ranking in real-time. LineUp also supports the comparison of multiple rankings. For example, users can visualize how rankings change over time or between different attribute configurations by displaying multiple rankings side-by-side. This comparison is also supported by slope graphs that show the change in the rank position of items between the two rankings.

Building on these works, we developed our own trade-off metrics, in particular when considering *comparisons* (see Chapter 4).

### 2.4.2 Summary on Vis tools for Trade-off Analysis

Several of the tools presented in this section (with few exceptions) were not designed to support trade-off analysis per se. In these tools, analysts can indicate preferences and prioritization of the different dimensions in the form of weights and get a ranking in relation to a score calculated from these weights. We are inspired by the tools discussed in this section for creating our own exploration tool for trade-off analysis

(Chapter 3), noting in particular the good performance of 2D projection solutions (scatterplots) when Pareto front datasets are large. In addition, we consider the prevalence of tabular and barchart visualizations when the goal is ranking of alternatives (Chapter 4, Chapter 5). We go beyond these ranking approaches by studying how to visually incorporate more trade-off metrics and how to consider trade-off solutions that may be made up of multiple datapoints, rather than a single datapoint.

## 2.5 Provenance and Trade-off Analysis

One of the characteristics of trade-off analysis as used in our work is that trade-off analysis is conducted over long periods of time and several exploration sessions [112], as multiple experts try to explore alternatives and reach common ground, making this type of analysis hard to resume and revisit. Moreover, during this analysis, the criteria and priorities experts have are not fixed and may shift over time. This motivated the first part of our own work: how to best keep track of trade-off analysis, using provenance (Chapter 3).

Xu et al. define provenance information as the *“history of the data and reasoning involved [during complex sensemaking tasks] and the context within which sensemaking was performed”* [85].

It is invaluable as a means to help analysts retrace their exploration and how it led to insights. As Ragan et al. explain [31], the types of provenance information that may be of interest to analysts can vary greatly. It can range from keeping track of low-level information such as data transformations, view changes, and user interactions; to high-level such as capturing insights reached during the analysis, or the rationale behind decisions and hypotheses.

In this thesis, we consider provenance visualization for sensemaking more generally. Although trade-off analysis and provenance visualization are prominent areas of research separately, little work exists that looked at exploiting provenance information to specifically assist trade-off analysis. An exception is recent work by Cibulski et al. [86], who developed an interactive Pareto front visualization tool to explore multi-criteria alternatives in engineering design. Since their focus is on trade-off analysis, they describe 13 design requirements for this domain, such as showing criteria ranges for simulation steering, supporting selections, and highlighting conflicting criteria; but they also describe one additional requirement for provenance pertaining to storing favorite alternatives for future comparison. Additionally, Trinkaus et al. [50] presented a generic multi-criteria decision support system called knowCube. They use provenance to support trade-off analysis by logging navigation paths, then clustering, storing, and viewing “good” decision alternatives. As such, past navigation paths and good candi-



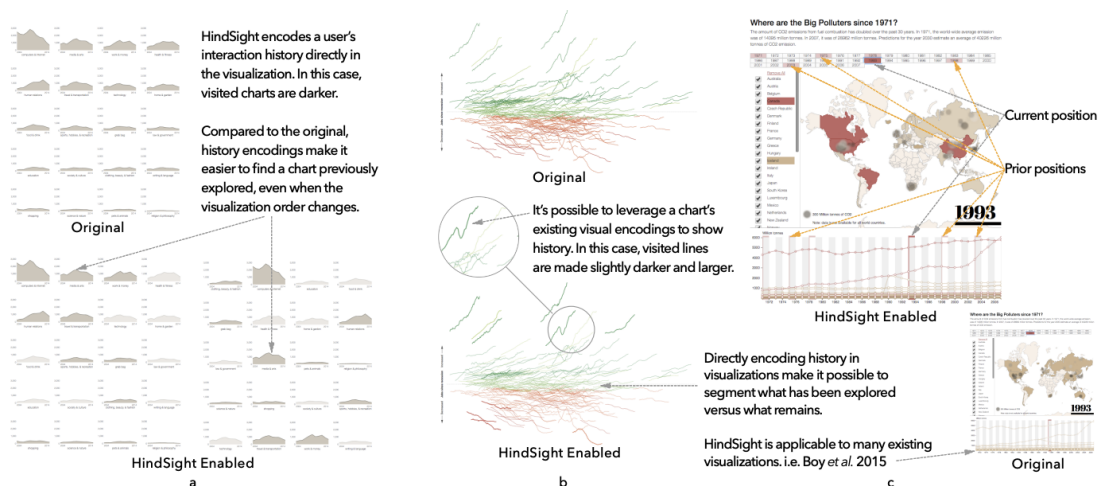


Figure 2.11: HindSight (Reproduced from [100]) encodes interaction history directly on visualizations, differentiating what has been visited from what remains.

date solutions are stored in a log file, and the user can later replay them in an animated movie-like fashion. KnowCube provides a simple visualization in the form of a radar chart to show the different decision alternatives and is demonstrated through real-life applications, but no user study or formal evaluation were conducted.

### 2.5.1 Provenance Visualization for Sensemaking

Despite the few works on combining provenance and trade-off analysis, there is considerable work conducted on provenance visualization more generally.

Visual analysis tools often capture basic view and interaction events that analysts can access in the form of a history or action log. But several systems go beyond this basic provenance capture and visualization. Past work has explored how to revisit, operate on (search, filter, annotate) [61] and more generally how to visualize and curate past history [118]. It also studied ways to help connect exploration steps [17] and aid analysts summarize and hand-off their analysis to colleagues [76]. A recent survey on the topic in interaction provenance provides further reading [84].

More relevant to our work are tools that provide *in-situ* provenance visualizations (or in-visualization provenance), augmenting existing parts of the interface with simple visual marks to encode provenance information. The first such instance is the idea of “visit ware” [11] that applies a fish-eye lens on a graph to magnify the nodes that are more visited. Later, BookVoyager [110] attempted, on the contrary, to de-emphasize the most visited timeseries in a line chart by graying them out in order to encourage users to visit new parts of the visualization (“road-less-traveled navigation”).

Scented Widgets [147] introduce a framework for enhancing interface components (such as buttons or sliders) with information that can aid navigation. While the concept is general, the work suggests several examples of visualizing provenance information, such as changing the color of a button based on the frequency of use or keeping track of datasets visited in a drop-down list.

HindSight [100] encoded interaction history directly on visualizations (in-visualization history), for example by changing the opacity of the most visited chart in a small multiples list, or the color and width of the most visited line in a line chart (Figure 2.11). In an online study, they observed that adding such interaction history increased user interaction with the interface and may have led users toward typically unexplored areas of the visualization.

Most recently Lumos [40, 41] added to scatterplots (among other views) in-situ visualizations of interaction traces: real-time ones, such as coloring individual data-points or attributes to encode frequency of interaction; and summative ones to show the distribution of data-points that users interacted with compared to a target distribution on a dimension. Their goal was to enhance awareness of biases [40] and mitigate them [41] during exploration and decision-making. In a series of laboratory and crowdsource experiments, they found that these traces increased awareness of biases [40], and somewhat mitigated unintentional bias but enabled intentional bias [41].

The provenance designs introduced in our technology probe (section 3.1) follow this *in-situ* approach, augmenting existing components of the interface with simple visual indicators of provenance information. Nevertheless, we contribute a study that uses in-situ visualizations to observe real analysts: our experts revisit their own past analysis and data, days or even months later; and also use our provenance visualizations during active exploration. Moreover, we introduce in-situ visualizations in an environment targeting a specific type of visual analysis: trade-off exploration.

## 2.5.2 Purposes of Provenance

While provenance support has not been studied in the context of trade-off analysis, past work has highlighted what we should expect provenance to be used for. Beyond provenance information types, Ragan et al. [31] have also identified the following reasons why analysts use provenance (purposes):

**Recall:** maintaining and recovering awareness and memory of the current and past states of the analysis;

**Replication:** reproducing aspects of a previous analysis (steps or workflow);

**Action recovery:** undoing/redoin past actions during the analysis;

**Collaborative communication:** sharing data, information, and ideas with colleagues;

**Presentation:** summarizing and sharing the results or processes of the analysis with someone not involved in the analysis; and

**Meta-analysis:** reviewing the analysis process itself.

These general provenance purposes serve as the starting point for our own work; nevertheless, we contribute a refinement of these purposes based on specific trade-off analysis needs (subsection 3.3.2).

### 2.5.3 Summary for Provenance in Trade-off Analysis

Provenance is important for visual exploration. Nevertheless, the few tools that are developed for trade-off analysis as well as the studies attached to them do not use nor study provenance or its impact on the analysis process. Thus, they are not tailored to capture aspects specific to trade-off analysis, such as the balance of trade-offs, the coverage of objectives and their respective importance, etc. Our work aims to cover this gap, but we are inspired by the recommendations of past provenance studies to design our own provenance-driven trade-off analysis probe and user study (Chapter 3). Moreover, we use this probe as a means to study trade-off analysis more generally, in particular in characterizing trade-off tasks for visual analysis. To the best of our knowledge, no prior work has looked at the effect of provenance visualization on trade-off exploration, nor at the characterization of trade-off tasks in the context of visual analysis.

## 2.6 Comparisons and Trade-off Analysis

Comparisons are essential in trade-off analysis where experts evaluate alternative options with multiple, often conflicting goals [59, 112]. For example, in wine production, experts must balance competing criteria such as low temperature profiles for wine production vs. produced aromas. Or in wheat production, reducing fertilizer use and environmental impact while increasing crop yield. They achieve this by comparing the pros and cons of different strategies, such as different temperature management strategies for wine production that either minimize energy or maximize aromas, or wheat production strategies—one that minimizes environmental impact versus another that maximizes crop yield [97].

Previous research in trade-off analysis has shown that domain experts often work with groups of points (or items) when exploring large trade-off spaces [97, 112]. Datasets derived from biological processes have inherent variability in measurements, so experts tend to use ranges instead of exact thresholds when identifying and comparing groups. In the wheat example, experts compared fertilization strategies aimed at achieving high-protein content with those that resulted in lower protein levels. In the wine example, they compared groups of wines that were produced under different temperature ranges (high and low).

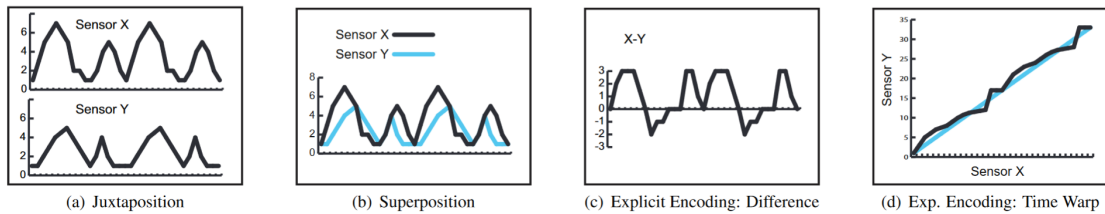


Figure 2.12: The different comparison layouts introduced by Gleicher et al. (Reproduced from [102])

Our goal is to design visualizations that support *comparisons of groups of points* in the context of trade-off analysis (topics covered in [Chapter 4](#) and [Chapter 5](#)). We thus review here existing methods for supporting comparison tasks both in visualization and sensemaking and then more specifically in trade-off analysis. Finally, we review methods for evaluating the quality of trade-off comparisons.

## 2.6.1 Comparison Visualization for Sensemaking

Comparison is important in sensemaking with data as analysts generate and compare multiple schemas or alternatives [121]. Li et al. characterized these alternatives in data analysis as multiples, options, and choices [67], identifying five high-level processes around alternatives. Comparison was one of these processes, regarded as a high-level reasoning task alongside inspect, interpret, and evaluate.

In terms of the presentation of these alternatives visually, Gleicher et al. [101] introduced a taxonomy for comparisons which includes three approaches: juxtaposition, superposition, and explicit encoding of relationships (see [Figure 2.12](#)). Juxtaposition involves placing multiple datasets side by side to facilitate direct visual comparison. Superposition overlays datasets within the same visual space, enabling immediate comparison through shared spatial coordinates. Explicit encoding directly represents the relationships or differences between datasets using distinct visual elements, abstracting the raw data into comparative metrics. This work notes that one of the main challenges with comparisons is scalability, as comparison problems grow in complexity (e.g., large number of criteria) and the number of objects being compared.

A similar approach to the juxtaposition and superposition method was proposed by Tominski et al. [23] to compare data visually by placing them close together. They introduce an interaction concept inspired by the tangible actions individuals take when comparing physical documents, such as arranging papers side by side or overlapping them to see through. This concept was validated by a qualitative user study in terms of effectiveness, intuitiveness, and user satisfaction of their novel interaction techniques.

In their work [22], Schmid et al. suggest comparative visualization, essentially

allowing analysts to interactively visualize the same data over multiple visualizations at the same time. The goal of this work is more about understanding one set of data and less about comparison between two (or more) groups of datapoints.

## 2.6.2 Comparisons in Trade-off Analysis

Hakanen et al. [59] discuss trade-off analysis in the context of interactive multi-objective optimization and identify seven high-level tasks, including comparison. They view comparison as a high-level process where the input is a set of alternative solutions, and the output is insights about trade-offs. They further break down the comparison task into five low-level tasks derived from taxonomies of information visualization tasks [7, 60, 70]: filter, abstract/elaborate, encode, reconfigure, and connect.

In our own work on using provenance, we also identified comparison as a key task in an observational study with domain experts, focusing on the use of provenance to support trade-off analysis. We discuss these more in the next chapter of this thesis, but our findings revealed that domain experts create alternative exploration paths leading to different solution groups, which they then compare.

Although not related to trade-off analysis across datasets, in the field of aircraft design that deals with trade-offs such as trade-offs between the geometry (space/aesthetics) and the aerodynamics (performance), Trapp et al. [56] talk about a new way to look at aircraft design using data comparison. They move from comparing pictures (like how airflow looks in experiments versus computer models) to comparing the actual data when the pictures are too difficult to compare. To do that, they rely on a side-by-side visualization of the data from the two designs they are comparing, with explicitly encoded metrics visualized on top of the designs.

Finally, although the trade-off metrics (see subsection 2.4.1) and visualizations to support trade-off analysis have been proposed, especially for analyzing individual points (e.g., [86, 132, 135]), their impact on comparison tasks in trade-off analysis and on the decision-making process is still unclear. Next, we review the main methods for evaluating the quality of trade-off comparisons.

## 2.6.3 Evaluating the Quality of Trade-off Comparisons

Evaluating the effectiveness of visualizations in complex decision-making scenarios, such as those involving uncertainty or trade-off analysis, where ground truth is either unavailable or hard to assess, presents significant challenges. Hullman et al. [63] developed a taxonomy for evaluating uncertainty visualizations and highlighted a common bias towards measuring performance based on accuracy rather than decision quality. In our research, we *indirectly* assess decision quality by examining the reasoning processes of decision-makers during trade-off comparisons (Chapter 5). We hypothesize that improved sensemaking and reasoning is likely to lead to better decision outcomes.

Following methods used in other sensemaking studies [128], we employ think-aloud protocols and self-reported decision justifications to assess the quality of reasoning in trade-off comparison tasks.

To analyze these decision justifications, as a proxy for their decision quality, we were inspired by existing literature in eXplainable Artificial Intelligence (XAI), a subdomain of AI concerned with developing methods to interrogate black box AI systems and extract insightful information about their inner workings [18]. Miller et al. describe explanations of AI systems as *transfers of knowledge about the behavior of AI systems* [139]. Specifically, we looked at work by Nauta et al. [105] who conducted a systematic review of 300 papers on evaluating explainable AI systems. The reviewed papers were published at major AI and ML conferences between 2014 and 2020. They identified twelve conceptual properties, termed Co-12, to assess the quality of AI explanations. These properties encompass dimensions such as correctness, completeness, contrastivity, compactness, and confidence, offering a multifaceted approach to evaluate the interpretability of AI models (Table 5.1). We will elaborate on these properties in section 5.1 and describe how we use them as *transfers of knowledge about the human decision-making process*, to then characterize comparisons in trade-off analysis.

To the best of our knowledge, no previous user study has evaluated the impact of visualization and trade-off metrics on comparison tasks involving groups of data points. The most closely related work is by Dy et al. [16], who evaluated the impact of visualization types (scatterplot matrices, parallel coordinates, heat maps, radar charts) and data complexity (number of criteria dimensions and points) on decision time and accuracy when selecting the ‘optimal’ option. In their work, accuracy refers to the consistency of choices with self-reported preferences, following the suggestion of Dimara et al. [34]. Their results show that accuracy remained similar across all four visualizations but improved when users dealt with fewer options and dimensions. This insight supports our design choice for the user study (section 5.1), where we limit the number of dimensions and data points. In contrast to Dy et al., and besides focusing on group comparison, we adopt a more nuanced approach, assessing decision-making rationale and user strategies by considering a wider range of qualitative factors including properties of explanations, user trust, and cognitive load.

#### 2.6.4 Summary for Comparisons in Trade-off Analysis

Comparisons play a key role in both trade-off and visual analysis, with extensive work both in identifying comparisons in trade-off tasks and in supporting visual comparisons in contexts other than trade-off analysis. Nevertheless, the two have not been considered together. The existing approaches introduced in visual comparisons do not take into account the need to compare groups of points, nor the context of trade-offs where analysts need to balance objectives and priorities that may shift.

## 2.7 Placement of this thesis

Trade-off analysis is a process that involves making decisions using criteria and priorities that are often conflicting and evolving. Our work falls at the intersection of trade-off analysis and visual sensemaking. More specifically, we consider how visualizing analytic provenance and comparisons can aid the trade-off analysis process. While existing work has considered separately visualization of provenance and visual support of comparisons, none has studied them in the context of trade-off analysis.





# Understanding How In-Visualization Provenance Can Support Trade-off Analysis

To start our exploration of how visualization can aid trade-off analysis, we focus on the challenge experts face to keep track, resume, and revisit their trade-off analysis sessions and objectives [59]. To aid them in this task, we propose using analytic provenance visualization [20, 31, 80, 84]. In particular, we consider embedding provenance visualizations directly inside the analysis environment, i.e., *in-visualization provenance*, explained more in this chapter.

Our first goal is to explore the impact of provenance visualization and how existing purposes of provenance defined by Ragan et al. [31] manifest in the context of trade-off analysis. Our second goal is to use this exploration as an opportunity to more broadly understand trade-off analysis tasks, analyst needs when it comes to supporting trade-off analysis (beyond provenance), and opportunities for further research.

This work was published in IEEE TVCG 2022 [97]. Supplementary material and video are available at <https://github.com/tradeoff-analysis/provenance>.

## 3.1 VisProm: a Provenance Visualization tool for Trade-off Analysis

We designed VisProm (Visualization Provenance master) to act as a technology probe [48]. Technology probes are systems traditionally used to field-test the usage of a technology in real-world settings and inspire ideas for new technologies to support user needs. As such, VisProm had to be robust, but also effective for field use. Our goal was to see how provenance visualizations can aid trade-off analysis, focusing in particular on *in-visualization* views that are embedded in the exploration environment (rather than dedicated views). To this end, we needed as a starting point a tool that has

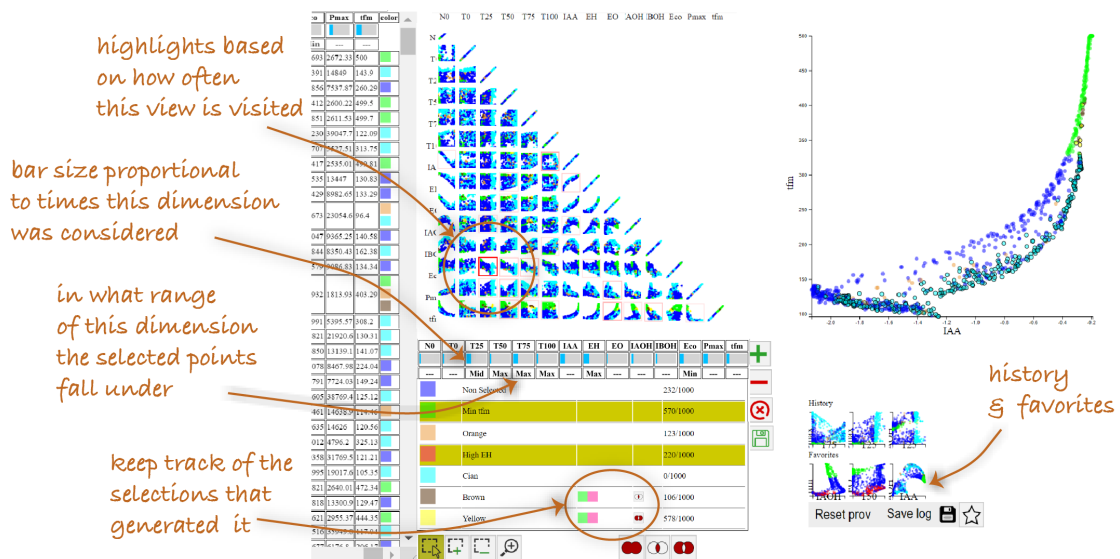


Figure 3.1: The VisProm technology probe includes several in-visualization provenance views to aid trade-off analysis.

been shown to be appropriate for trade-off analysis, that we could then augment with in-visualization provenance views.

**Choice of Scatterplot Matrix:** The inspiration for the basic trade-off support (before adding provenance) comes from EvoGraphDice [6, 113, 115]. This tool relies on a scatterplot matrix (SPLOM) that is well suited to communicate relationships between dimensions, such as correlations, and distributions of data points [34, 124, 151], it is thus appropriate for visualizing relationships between objectives to show trade-offs. The SPLOM includes multiple visual query selections, differentiated by color, to help experts narrow their search space to important parameters [140]. Importantly, EvoGraphDice has been used effectively in the past for trade-off analysis conducted by groups of agronomy experts [112].

While we did not use EvoGraphDice, in our new VisProm probe we replicated the following functionality from EvoGraphDice: (i) a scatterplot matrix of all dimensions from the dataset; (ii) a main scatterplot that is seen in larger detail; (iii) a query panel where users can choose colors in order to perform selections on the main scatterplot (visual query sculpting); (iv) a datatable available on demand, with points colored depending on the visual queries they belong to; and (v) a means to create new dimensions by combining existing ones using arithmetic operators and weights. This basic version also includes two components that can be considered as simple forms of provenance support: a history of actions such as view changes and visual query selections, as well

as a list of favorite views (or bookmarks) that is populated by the user.

We implemented the above functionality in VisProm, and also provided additional support designed explicitly for provenance tracking. Thus, although the trade-off analysis part of VisProm is inspired by EvoGraphDice, they are distinct: EvoGraphDice is a desktop application that uses an interactive evolutionary algorithm to guide user exploration, and does not include in-visualization provenance aside from a basic history and favorite views.

**Choice of in-visualization Provenance:** In the past, EvoGraphDice was used to study the role of expertise in collaborative exploration of complex model simulations (and how domain experts structure their exploration) [112]. Our current work goes deeper into the analysis tasks domain experts engage in when they explore trade-off datasets, whether they consult provenance views and how they use them while they are engaged in the exploration.

An informal test of EvoGraphdice [5] which included a provenance view as a separate window was not always consulted by users, who were often too absorbed by their main task. Thus in our approach we provided instead simple *in-visualization views* embedded in the analysis environment, in an attempt to make them always visible. Past work has also introduced *in-situ* provenance visualizations (or in-visualization provenance), augmenting existing parts of the interface with simple visual marks to encode provenance information (see subsection 2.5.1), but none of them were designed specifically with trade-off analysis in mind, nor were they studied under this context. We chose simple designs to not clutter and not compete with the main analysis views, but this begs the question of whether these simple designs are consulted and when, and if they carry enough information to support provenance needs. A contribution of our work is studying if the same, simple provenance designs can do both.

The basic trade-off support functionality and our provenance additions can be seen in Figure 3.1. The additional provenance designs were motivated by workshops with expert users, described next.

### 3.1.1 Design process

To inform the design of VisProm we are inspired by prior work, reporting on findings from brainstorming and design workshops [5], and insights from related work in trade-off analysis (section 2.2). Past work [5] reported on three workshops carried out with nine participants in total, including experts in HCI, visualization, and agronomy. The goal of these workshops was to gather user requirements for visualizing exploration histories (analytic provenance). Participants were asked to brainstorm functionalities to better support sensemaking of exploration history generated for a SPLOM-based tool [113]. Inspired by these findings, we report next our recommendations (or user requirements) for the design of VisProm :

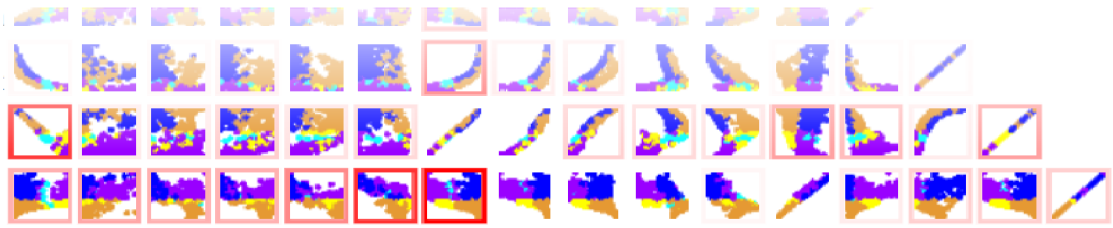


Figure 3.2: ViewCount: in this partial view of the SPLOM we see the ViewCounts as red borders around the thumbnails, indicating how frequently they have been visited (more saturated red for more visits).

[R1] History and Favorites (existing features) are important for reverting to past steps when errors occur and for storing important views. Thus, they should be maintained.

[R2] Key objective dimensions that have been visited (or ones that have not) are important to know. This is confirmed by past work that observed that when experts have a priori hypothesis, they use objectives to prioritize their exploration [112].

[R3] Views already visited, i.e., what relationships between dimensions have been considered, was a common suggestion as a type of information the experts would like to track.

[R4] Combined dimensions they created and with what weights were identified by some experts as important to track. Previous work has also discussed the importance of weights for expressing preferences [135].

[R5] What objective dimensions were maximized or minimized in their decisions, or more generally what range of the dimensions their decisions fell in, were also identified as important to keep track of. Past work also highlights the need to identify and “lock” part of decision ranges [50].

[R6] Alternative search spaces considered should be tracked. Experts mentioned this in the workshop, and past work on trade-off analysis has identified that comparisons of alternatives are important [59].

[R7] A way to summarize the entire exploration would be desirable. This is an important topic in provenance visualization (section 2.5), not only for trade-offs.

### 3.1.2 Additional *in-visualization* Provenance Designs

We discuss next the different in-visualization components of VisProm that we added to communicate provenance and how these support the requirements identified before.

### 3.1.2.1 ViewCount on Scatterplot matrix (Figure 3.2)

Our tool relies on a SPLOM matrix that displays thumbnails of scatterplots for all pairs of dimensions, coupled with a main view of one selected scatterplot shown in large scale that users can interact with. The SPLOM acts as a navigation tool, letting the user choose which pairs of dimensions they want to interact with in the main scatterplot view. Users can select points on the main scatterplot using colored queries and see the impact of these selections in all thumbnail views. This ensures the user knows exactly where the selected points fall under different dimensions that they may be trying to optimize.

Each scatterplot in the SPLOM corresponds to a relationship between two dimensions, for example, correlations or competing objectives that represent trade-offs - when one increases, the other decreases. To help analysts keep track of what pairs of trade-offs they have already considered [R3], we log how often they visited each scatterplot. We augment the SPLOM thumbnails with a colored edge, whose opacity changes in real time in proportion to the number of times the corresponding scatterplot was visited, relative to the most accessed scatterplot.

### 3.1.2.2 DimensionCount table (Figure 3.3a)

Participants in the workshop expressed the need to track the important objective dimensions they have already tackled in their analysis [R2]. To aid this task, we introduced a table with one cell per dimension, where each cell contains a bar that represents the number of times each dimension has been accessed (all bars are relative to the total access counts in the SPLOM). The length of the bars indicates which dimensions have been already considered through visits of scatterplots that include them, and which dimensions may have been overlooked. This representation differs from the ViewCount on the scatterplot matrix, as it focuses on the access history of individual dimensions rather than pairs. This design resembles past work, like the Lumos attributes panel [40] that shows frequency of interactions with different dimensions, to provide in their case awareness of biases in the exploration.

We initially planned to only add the DimensionCount table on top of our collapsible datatable (Figure 3.4). In the datatable, each column is a dimension and each row a datapoint with corresponding numerical values, as well as the colored selections they belong to (colored rectangles). Given the existing organization by dimension columns, we added here the DimensionCount for each column (dimension). Feedback from experts indicated that while the datatable is extremely useful to confirm surprising findings and double-check values (also observed in previous work [112]), it nevertheless remains collapsed and not visible during the majority of the exploration. We thus decided to replicate the DimensionCount table on top of the visual query selection panel as well, so as to ensure it is visible at all times (Figure 3.3a).

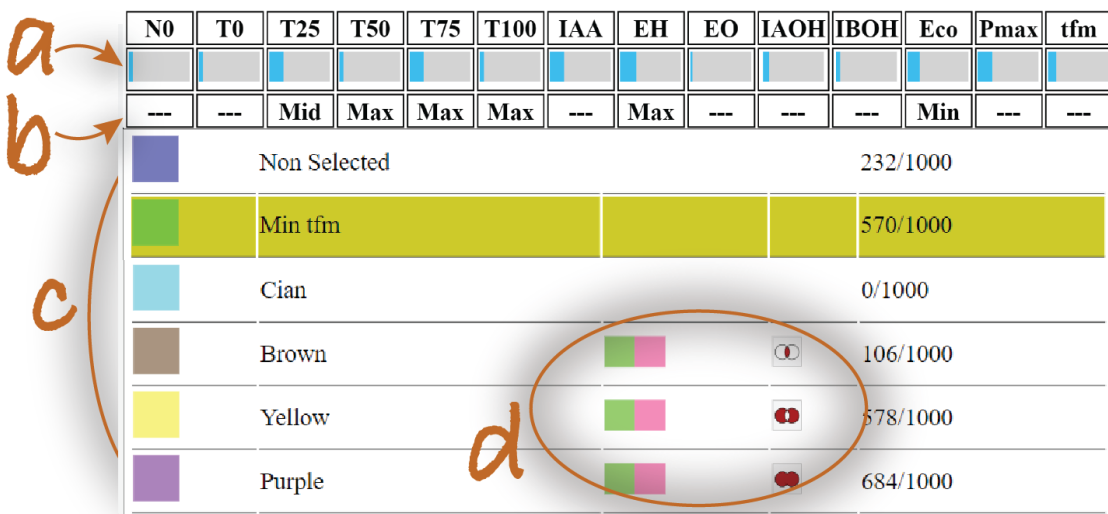


Figure 3.3: The query selection panel includes: the dimension names and their DimensionCounts (a) as a bar that is filled depending on frequency; the SelectionRange tool (b) that indicates with text in what range (if any) the current selection falls under for the particular dimension; and finally (c) all queries/selections. Selections can be renamed (like the active selection *Min tfm*) and their colors changed. For selections that were created by group operations between other selections, their SelectionOrigin (original selections and operation) can be seen next to their name (d).

### 3.1.2.3 SelectionRange table (Figure 3.3b)

Experts commented that it is often challenging to keep track of what key objective dimensions are optimized (maximized or minimized) in the choices they make [R5]. In VisProm, choices are generally expressed as colored queries/selections of interesting points. To help analysts understand where their selections fall under the different dimensions, we added the SectionRange table. Each cell of the table corresponds to one data dimension and indicates with text (Min/Max/Mid or —) two key pieces of information: (i) if the points in the current selection are mostly clustered under that dimension; and (ii) if this cluster tends to be in the max range of values of the dimension, the min, the middle range, or if no pattern is detected. We consider as clustered, datapoints that have a low value for the ratio  $\text{variance}/(\text{value range of dimension})$ . Then we determine the approximate position of the cluster in the range of values for the dimension (higher, middle, lower third). The SelectionRange table is placed below the DimensionCount table, and thus added to both the datatable and selection query panel.

Other work [40] visualizes where individual interactions with data-points take place along a dimension, compared to an ideal distribution for that particular dimension. In

custom dim																	
No	T0	T25	T50	T75	T100	IAA	EH	EO	IAOH	IBOH	Eco	Pmax	tfm	$((0.33 \text{ IAA}) + (0.33 \text{ EO})) + (0.33 * \text{EH})$	$(\text{Eco } 0.50) + (0.50 \text{ Pmax})$	color	
---	---	Mid	Max	Max	Max	---	Max	---	---	---	Min	---	---	---	---	---	
0.07	18.13	18	18	18	18.38	-0.2	-0.35	-0.37	178.45	18.74	546693	2672.33	500	-0.3	274682.66		
0.22	18.06	24.05	29.84	30	20.32	-0.94	-0.39	-0.51	248.15	68.71	420391	14849	143.9	-0.61	217620		
0.17	19.94	18.05	18	18	18	-0.57	-0.42	-0.54	200.05	28.31	568856	7537.87	260.29	-0.51	288196.93		
0.07	18.13	18	18	18	18	-0.19	-0.35	-0.37	177.1	18.45	555412	2600.22	499.5	-0.3	279006.11		
0.07	18.25	18.14	18	18	18.01	18	-0.19	-0.35	-0.37	177.8	18.62	556851	2611.53	499.7	-0.3	279731.27	

Figure 3.4: Top of the Datatable, including colored indications of points belonging to queries (right column). On the top just under the dimension names, are the DimensionCount and the SelectionRange tools.

our case, our experts are more interested in larger zones of data-points (expressed as selections) and where these fall across all the different dimensions they are trying to optimize.

### 3.1.2.4 Selection Origin (Figure 3.3d)

Analysts can interact with the main scatterplot view and perform visual query selections (Figure 3.3c). The visual query panel helps them choose a color for each of their selections and provides additional information regarding the number of selected data points. As in previous work [112, 115], we provide tools to sculpt the visual selections (adding and subtracting datapoints in different views). Past work [112] and our workshops indicated that experts express their search and trade-off criteria with these colored selections. They often compare them when they represent different decision alternatives [R6], observing where they fall within the multi-dimensional space. One aspect that the experts requested was to also be able to perform grouping operations between multiple important selections (intersection, union, difference) in order to combine their criteria more effectively. In an effort to help them keep track of the provenance of such combinations, we added information about their origin in the form of two icons next to the query name: one with the grouping operation that created it and one with the colors of the original queries used in the grouping.

### 3.1.2.5 History and Favorites (Figure 3.1)

While not unique to VisProm, we mention History and Favorites as they represent basic provenance visualizations that are important for our participants [R1]. The basic history displays thumbnails of the last states of the tool—both selections and navigation changes. The Favorites act as a bookmarking tool, where analysts can explicitly store a view for easy access. Both History and Favorites are interactive; clicking on a thumbnail restores the saved view.

Our goal is to create in-visualization provenance views rather than summaries that can be considered outside the main exploration view. We thus do not provide elaborate

Table 3.1: Datasets and participants (P1-5) for the three case studies. Each was conducted in two sessions, with a time interval in between (Weeks).

CS Domain		#Dimensions	#Points	P#	Weeks
ml	Benchmark meta-analysis	10	3000	P1,2	1
eco	Sustainable farming	9	100	P2,3	17
w	Wine fermentation	14	1000	P4,5	156

summarizations of past exploration history [R7]. Nevertheless, our History provides a snapshot of the final workspace that is valuable for remembering past analysis [32].

### 3.1.3 Implementation

As our work took place during the COVID-19 pandemic, we opted for a web tool that could be operated easily without any installation and intervention from our part. We implemented VisProm using D3 and Node.js for the rendering aspects, and the Google Firebase database service to log the user actions during the exploration. For the individual thumbnails on the SPLOM matrix, we used Canvas drawing; as they are quite small, we optimized their rendering using sampling [154] that reduced the number of points drawn inside the thumbnails, but ensured the larger patterns are still visible. The prototype has been used on the Chrome browser with datasets of up to 20 dimensions and 1000 datapoints without any performance issues. The provenance visualizations on the SPLOM update every time a selection is made or a combined dimension is created, but will only update for every three accesses to the SPLOM thumbnails (this makes the exploration of the SPLOM smoother).

## 3.2 User Study Design

We conducted an observational study to understand how in-visualization provenance is used during trade-off analysis. We focused on the following research questions on *task support* and *usefulness* of provenance during and after the exploration:

[RQ1] What tasks do experts engage in during trade-off analysis? How do provenance purposes manifest in trade-off analysis?

[RQ2] Does in-visualization provenance support *a-posteriori* analysis, such as by helping recall of past exploration?

[RQ3] Is in-visualization provenance taken into account by experts *during* trade-off analysis and how?

[RQ4] What are untapped opportunities for using provenance to support trade-off analysis?



### 3.2.1 Participants and Use Cases

Five domain experts (P1–5) from research institutions in France and the UK participated in this study (5 male, mean age 36.6). Participants were researchers (4 domain experts) and one first-year PhD student. The recruitment procedure drew on previous collaborations between the authors and participants, or between participants, but none took part in the design brainstorming sessions described in section 3.1.1. At the time of the study, all participants were involved in different research projects, we call *Case Studies* (CS), where trade-off analysis is an important aspect of their work. Different domain experts participated in each case study, with the exception of participant P2 who was involved in two case studies (CS-ml and CS-eco).

**Case Study 1: ML Benchmark Meta-Analysis (CS-ml):** The goal of this exploration is to characterize possible trade-offs and relationships between different features of a number of benchmark Machine Learning (ML) datasets. An example *trade-off* concerns the number of features of a benchmark dataset, its intrinsic dimensionality (i.e., dimensions that exclude noise) and the average correlation between features. As the number of features grows, the intrinsic dimensionality is likely to grow. However, the feature average correlation is likely to decrease.

**Case Study 2: Sustainable Framing (CS-eco):** The goal of this exploration was to assess the trade-offs between different aspects of sustainability in farming practices in Europe. These sustainability aspects or indicators cover environmental, economic, and social factors that are often in conflict, such as between fodder consumption, animal production, and farmers' perceived happiness. An example *trade-off* pertains to animal production, which is likely to increase with fodder consumption, but in many cases has a negative environmental impact.

**Case Study 3: Wine Fermentation (CS-w):** The goal of this exploration is to compare different wine fermentation strategies and their characteristics. In terms of *trade-offs*, experts are searching for fermentation strategies that maximize wine aroma (three esters) and minimize undesired compounds (higher alcohols) and the energy required to control fermentation.

All datasets were prepared by the domain experts themselves (see Table 3.1), who also brought their own research questions. We note that all domain experts performed basic statistical analysis of their data prior to this study, using software tools such as MS Excel, R, or Matlab, but they have not used interactive visualization.

### 3.2.2 Study Procedure and Setup

Similar to Ragan et al.'s method [32] where participants conducted analysis in two steps one week apart, we run our study in two sessions separated by a time interval of one week for CS-ml, 17 weeks for CS-eco, and 156 weeks for CS-w. Since our focus is on studying provenance through the history of user interactions, in the first session

we only collected history data, and we only activated provenance and analyzed the second session of each case study. We also note that for CS-w, whose first exploration session took place three years earlier, we used a similar trade-off analysis tool (Evo-GraphDice [113]), which had all the basic functions provided by VisProm but lacked the new provenance visualization functionalities described in [section 3.1](#).

In the first session, participants reached *new insights* in the form of: interesting relationships between aspects of the data (e.g., between animal production system typology and sustainability indicators, CS-eco); new findings about clusters of data points and how they are situated with regard to a decision boundary (CS-ml); or alternative exploration paths to reach the same (aroma) objectives (CS-w). To reach these insights, participants consulted multiple scatterplots, created numerous selections, and in some cases composed new combined dimensions.

All second sessions of the three case studies were run online using a video conference platform. Participants used their own machines with either a 13" or 15" monitor. For each case study, one participant led the exploration by interacting with the online tool and sharing their screen with the second participant (and the study facilitators) when they were not co-located with them (CS-ml and CS-w). Both domain experts equally participated in the trade-off exploration by proposing directions for the exploration, advancing new hypotheses, and discussing findings.

Each session lasted  $\sim 1.5$  hours (mean 108 mins). The second session (analyzed) was structured in the following Parts:

- 1- *Introduction* [ $\approx 10$  mins]: Welcome and tool setup.
- 2- *Reproduce Insight Without Log* [ $\approx 15$  mins]: Our experts first saw the stripped-down version of VisProm, *without* log data from their previous session. In this part, experts attempted to recreate the results they reached in their previous exploration, relying on their memory. Our goal was to identify potential challenges in reproducing insights when no provenance history data is visualized.
- 3- *Reproduce Insight With Log* [ $\approx 15$  mins]: Our experts then saw the full version of VisProm with log data from their previous session visualized. We explained the additional provenance visualizations then participants were asked to reflect on their past exploration, in particular, any aspects that they were unable to reproduce by memory. The goal of this study part was to observe how experts used the new provenance tools to revisit past findings.
- 4- *Open Explore* [ $\approx 30$  mins]: Participants continued their exploration with the full VisProm functionality, without any further instructions. The goal of this study part was to see if and how experts used the in-visualization provenance during their active exploration.
- 5- *Discussion* [ $\approx 20$  mins]: Finally, we conducted an open discussion session to elicit the experts' feedback about their experience, in particular related to how they used the provenance aspects of VisProm and what more they would like added or improved.

We chose to study provenance both when revisiting past analysis (days and even months later) and during active exploration. We suspect that the provenance needs differ: revisiting past analysis may require more support to acquire high-level understanding and reproduction of steps when memory may be faulty, whereas it may not be needed for reproduction of steps during active exploration and may even not be used in on-going analysis.

### 3.2.3 Data Collection and Video Coding

We collected video recordings of the three case studies (324 minutes), user interaction log files, and observational notes. Two different coders independently coded each video using discrete events, then met to discuss annotation codes and resolved conflicts. We use the term 'event' to refer to a chunk of video for which we attach a code pertaining to our research questions, mainly about trade-off tasks, provenance, and opportunities for provenance supported trade-off analysis.

We followed standard thematic analysis [142] in coding the videos and analyzing the collected data, using both inductive and deductive coding. We coded separately trade-off tasks and provenance events. For trade-off tasks, we did not rely solely on specific terms to find trade-off tasks (such as trade-off, objectives, criteria), we also observed what participants did and what they said to enrich the descriptions of those tasks. For example, one participant's quote: "*this is NitOut it would be nice to combine with NilInput*" (CS-ml) was coded as "Combine interest zones & dimensions". Another event was based on our observation that the participant was checking where a selection fell with regards to the bisector (and later on confirmed in the Discussion). This was coded as the trade-off task "Locate & characterize". For provenance events and associated opportunities, an initial annotation scheme was decided before coding, based on the general provenance purposes proposed in Ragan et al. [31]. We coded events in the study parts where experts engaged with their data (Parts 2-4, 84.26%) as well as their comments in the discussion (Part 5, 15.73%).

In addition to coding for trade-off tasks and observed cases of provenance use, we also coded opportunities for provenance. We identified these when observing domain experts express hesitance in remembering important aspects of their exploration (current or previous), or when we thought the challenges encountered by the exploration may be overcome using provenance visualization. Our corpus contained 178 events in total (CS-w: 39.32%, CS-ml: 34.83%, CS-eco: 25.84%), related to trade-off tasks (25.84%), provenance tool use (23.03%), and opportunities (51.12%). A thematic saturation analysis [12] found conceptual stability approximately halfway through the data collection.

The results of the analysis are described in the following section. Participants' comments are in *italics* and we indicate their case study, as well as whether they were

translated using the † sign.

## 3.3 Results

We first discuss the trade-off tasks that our study participants engaged in [RQ1]. We report next a summary of the provenance purposes we identified that refine ones from previous work [31] by adding subcategories and concrete specifications from real-world trade-off analysis. We then show how in-visualization provenance was used both to revisit past analysis [RQ2] and during the actual exploration [RQ3]. We conclude with the opportunities that our probe helped us collect that can inform the design of future trade-off analysis systems that include provenance [RQ4].

Further details about our study, including the video codes, participants' comments and analysis notebook, can all be found in <https://github.com/tradeoff-analysis/provenance>.

### 3.3.1 Trade-off analysis tasks [RQ1]

We coded key events related to the different tasks participants performed during trade-off analysis. Our goal was not to code each one of those events. Rather, it was to first showcase the variety of tasks decision-makers engage in during trade-off analysis. And second, to investigate if there are new tasks not reported in past literature. We qualify these tasks as high-level because we focus on user needs for trade-off support rather than the usage of any implemented feature of our system. As such, some of the trade-off tasks described below are not currently fully supported by VisProm.

Overall, there were 46 task events (CS-ml: 47.82%, CS-w: 32.60%, CS-eco: 19.56%), most of which were observed during the Reproduce Insight Without Log part where participants tried to reproduce an old insight without log data loaded into the tool (41.30%), and the *Open Explore* part where they conducted open exploration using the additional provenance support tools (52.17%). We only observed one trade-off task event in Reproduce Insight With Log as participants were mostly concerned with interpreting the loaded history data, and reflecting on how the previous insight was reached, rather than actively conducting trade-off analysis.

We coded these events and categorized them into the following *eight* high-level trade-off tasks:

**[T1] Minimize & maximize (6.52%):** When aspects of the multi-criteria problem were clearly described, participants took a direct approach to explore trade-offs by minimizing or maximizing known objective dimensions or criteria. For example, participants of CS-w made various colored selections to maximize the amount of aroma and

to minimize undesired wine compounds, and the energy and time required to control the fermentation process.

**[T2] Locate & characterize (28.26%):** When little or no information is known about the structure or dependencies within the multi-criteria problem, participants tried first to locate interesting areas of the trade-off space using known thresholds, decision boundaries, or ranges that are important for the problem domain. For example, experts in CS-ml examined the trade-off between the number of features and the number of samples in their dataset, focusing on a cluster of points having 2-16 features. They then tried to characterize this cluster by checking its spread and distribution in other views. As the number of selections grew, participants gave meaningful names to their selections such as “grains high subs” (CS-eco) or “optimum” at the end of the exploration (CS-w).

**[T3] Cascade & refine (15.21%):** Participants deepened their understanding of the multi-criteria problem and its characteristics by *cascading* selections across multiple views, using brushing and linking. Oftentimes, they *refined* and narrowed down the initial selections in order to find the best possible solutions. For example, in CS-w, starting from an initial large selection of fermentation recipes requiring a minimum level of initial nitrogen (149 recipes), participants found a much reduced set of “optimal” solutions (9 recipes), based on successive cascade and refine operations, using three criteria: aromas, fermentation time, and energy.

**[T4] Rank & prioritize (4.34%):** A key decision participants often made when exploring trade-offs was to decide the order in which to examine the different objective dimensions. This order or ranking may have a different meaning depending on the use case or participants’ research questions. For CS-w, the order determined the size of selections (i.e., interest zones). Initial criteria were less strict, and participants tended to make larger and more generous selections (e.g., on initial nitrogen amount), whereas selections based on subsequent criteria were more strict, and thus their sizes were smaller (e.g., aromas). For CS-ml, the rank did not determine the size of selections. Instead, participants first visited features of their dataset that were most relevant to their research questions.

**[T5] Create branches & compare alternatives (13.04%):** When resuming previous analysis, participants wanted to branch out with a different but related goal, or create an alternative exploration path, often with shared objective dimensions. New branches and alternative exploration paths helped domain experts discuss possible explanations and reasons behind certain trade-off choices.

**[T6] Combine interest zones & dimensions (10.86%):** Our participants often used a divide and conquer strategy, looking at selected zones of the trade-off space across different dimensions, and then combining them to explore more complex trade-offs. For example, participants created multiple selections to express different interest zones such as farms with high subsidies, and those who support organic practices (CS-

eco). They then combined these selections to see farms that fell under both zones. Our experts were also interested in combining objectives, and dimensions more generally. In particular, they selected a subset of objective dimensions and assigned weights to reflect their relative importance.

**[T7] Observe coverage (4.34%):** Participants wanted to keep track of values explored within a specific data dimension or across many. They commented on whether they visited most dimensions, but sometimes they had difficulty remembering the dimensions they visited and the part of the exploration where this happened.

**[T8] Find clusters, correlations & outliers (17.39%):** Like in multidimensional data exploration, participants looked for clusters, correlations, and outliers in their datasets, including between pairs or multiple conflicting objective dimensions. They explored groups of data points forming a cluster in one of the views or meeting a specific criterion, then looked at how these relate to other dimensions.

In section 3.4.1 we discuss the differences between our identified tasks and these reported in past work on trade-off analysis in operations research [59], and reflect on the generalizability of these tasks. In the next section we will see how the additional provenance tools aided specific trade-off tasks, like keeping track of coverage, and what provenance needs they met.

### 3.3.2 Refined provenance purposes and objects

We next looked at how provenance purposes manifest within trade-off analysis. We report on aspects of provenance that we observed or were reported to us by our domain experts (74.15% of coded events). When analyzing these events, not only did we observe all of the provenance purposes discussed in Ragan et al. [31] (see [subsection 2.5.2](#)), but we also identified aspects central to the trade-off analysis process or the data records that participants were interested in. We call these aspects *objects of provenance purposes*, and in our coding, they were identified by asking ourselves this question: “**What** do our experts want to X here to support trade-off analysis?” Where X is a provenance purpose [31], either recall, replicate, recover, meta-analyze, or present. We categorize these provenance objects into the following eight types:

**[O1] Interest zones (27.27%):** Constraints that decision-makers are willing to accept for one or multiple dimensions, such as a threshold, range, boundary, or distribution.

**[O2] Exploration structure & steps (25%):** The analytics steps taken during trade-off analysis, or the general phases and scenarios that constitute the exploration, which can include refinement, branching, and alternative paths.

**[O3] Focus & priority (15.90%):** The dimensions that the experts focus on, either as single, multiple, or combined dimensions.

**[O4] Coverage (11.36%):** The extent to which the trade-off space has been explored, such as in terms of data dimensions and views visited, or the data files loaded into the tool.

**[O5] Data semantics & origin (8.33%):** The meaning conveyed by the data dimensions, specific values or interest zones, and the context and data collection methods used to interpret them.

**[O6] Exploration approach (7.57%):** High-level reflections on the exploration strategy, which often includes a motivation, a rationale, or an evaluation of the approach.

**[O7] Verification & validation (3.03%):** Confirmation of known insights or findings, such as to verify or validate a dataset.

**[O8] Annotation (1.51%):** Notes and labels to document important aspects of the exploration, such as key views, useful insights, or the rationale behind the exploration.

Interest zones **[O1]** appears to be the most important provenance object in trade-off analysis as experts were interested in their recall, replication, recovery, meta-analysis, and presentation (subsection 2.5.2). This is followed by exploration structure & steps **[O2]** (e.g., alternatives or branches), where experts were also interested in using this information for all purposes except for presentation, presumably because the trade-offs and insights expressed as interest zones were more valuable to share with others than the exact steps to get there. Other objects were tied to specific purposes, such as recall for both coverage **[O4]** and data semantics & origin **[O5]**, meta-analysis for the exploration approach **[O6]**, and presentation for annotation objects **[O8]**.

Our experts were interested in remembering all provenance objects apart from the exploration approach. This could be because experts did not experiment with different approaches to explore trade-offs given their brief exposure to our tool, and so they did not feel the need to be reminded. However, they still reflected on their choice of exploration approach, and sometimes also on their exploration structure & steps, and interest zones. They were mostly interested in *reproducing* past exploration structure & steps, and, to a lesser extent, past interest zones, the dimensions they focused on **[O3]**, and the results of any verification & validation they carried out **[O7]**. Finally, they mostly wanted to *recover* interest zones and their past exploration steps. Coincidentally, these are the two aspects they struggled with the most, either by accidentally losing selections, or because the exploration had too many steps.

It is interesting to note that some objects have a direct mapping to the identified tasks. We discuss the importance of these identified objects for trade-off analysis design in section 3.4.2.

	Recall	Replicate	Recover	Meta	Present	#Events
Interest zones %	78	3	11	3	6	36
Exploration structure & steps %	48	27	15	9	-	33
Focus & priority %	95	5	-	-	-	21
Coverage %	100	-	-	-	-	15
Data semantics & origin %	100	-	-	-	-	11
Exploration approach %	-	-	-	100	-	10
Verification & validation %	75	25	-	-	-	4
Annotation %	-	-	-	-	100	2

Figure 3.5: Our identified objects and how they are distributed across Provenance purposes [31] ("Present" purpose also includes Collaboration).

### 3.3.3 Provenance supports trade-off analysis [RQ2, RQ3]

We report next observed events of experts' using provenance views in trade-off analysis. These events are either direct observations of the experts using the provided tools based on pointing gestures or direct references, or cases where experts mention explicitly having used the tools. We rely on deictic and verbal references, as most of our provenance tools are non-interactive and thus we cannot log their use. The 41 events collected were distributed across case studies (CS-eco: 8, CS-ml: 13, CS-w: 20).

#### 3.3.3.1 How provenance tools were utilized

We first report on how experts made use of the VisProm probe. Our goal in this section is to identify when the provenance views met the experts' provenance needs, sometimes in unexpected ways.

We explain how each provenance tool supports the tasks identified in section 3.3.1 (in **bold**) and the provenance purposes (in *italics*). A contribution of our work is introducing objects of provenance purposes, i.e., what the experts want to recall/replicate/annotate ... (section 3.3.2), so we report these explicitly (in underlined). For example, in the sentence "X helped experts *Recall* details of their dimension Coverage [T7: **ObserveCoverage**]", the word *Recall* refers to the original provenance purpose [31], the notation [T7] is a trade-off task identified in section 3.3.1 related to dimension coverage, and the word Coverage refers to a refined object for provenance (Recall Coverage) related to trade-off analysis.

**ViewCount** that colors the SPLOM thumbnails based on use frequency, was the most used provenance tool (16/41). It aided *Recall* (13/16) and *Action Recovery* (3/16). In half the cases, ViewCount helped experts keep track of Coverage [T7] in their exploration, i.e., objective dimensions they had already considered and ones they



had not (7/16). They also used it as a reminder of the objective dimensions that were a Priority ([T4:Rank&prioritize] - 4/16). As one participant mentioned “*Because for example this [unvisited SPLOM thumbnail], I ignored completely.*”, to which their colleague responds about priorities and unvisited regions “*Yeah but it’s also a matter of priorities right? Maybe after you get all you can from these visualizations you might in the end go there [unvisited thumbnail] just to check.*” (CS-eco). Although ViewCount does not provide temporal information, it sometimes (4/16) triggered experts’ memory and helped retrace their Exploration Steps. When the red borders appeared, one expert that was struggling to remember their past analysis looked at the SPLOM and exclaimed “*Uh, now I remember something, we probably started from here*” (CS-ml).

**History & Favorites** followed in use (8/41). History was surprisingly used rarely (3/41), but always for Action Recovery. Participants used History thumbnails to recreate lost selections that represented Interest zones and trade-offs between objectives they had considered (1/3), and to retrace Exploration Steps (2/3). Favorites (5/41) were sometimes used for storing selections that corresponded to Interest zones (4/5) [T2 & T3: Locate&characterize; Cascade&refine] that participants wanted to share and discuss further, serving the broader purpose of Presentation/Collaboration (2/5); or for Action Recovery instead of the History in a few cases (2/5) to recover important lost selections. Experts also used them to reflect on the effectiveness of their past exploration (1/5) (Meta-Analysis).

**SelectionRange** presents in text the range selected points fall under each dimension (Min/Mid/Max/-), and came next in terms of use (4/41). Participants used it mainly to Recall (3/4) if the datapoints they were considering were in specific zones of the objective dimensions (Interest zone [T1:Minimize&maximize]). For example, when two experts were discussing a selection, one showed the SelectionRange bar to point out the solutions considered: “*Since we did not want to use a lot of energy, we have Eco in the mid-range [points to the SelectionRange]*” (CS-w†). In one event (1/4), our experts used the range bar to Replicate a past action, such as when they remembered their past strategy of giving Priority [T4:Rank&prioritize] to some criteria: “*So we try to find sub-groups of datasets that are like, small number of features [referring to SelectionRange]*” (CS-ml).

**DimensionCount** bars indicate how often a dimension was viewed. It was used less (4/41), in all cases to Recall information. Participants used it to keep track of Coverage [T7:ObserveCoverage] of the important objective dimensions they wanted to optimize (3/4), and to Validate that they remembered correctly their past exploration (1/4).

**Queries** allowed experts to make selections that express choices about datapoints that fit specific Interest zones [T2:Locate&characterize]. It was used extensively during exploration as experts explored alternatives [T5:CreateBranches&Compare Alternatives], but we observed a few events where it was explicitly used to support provenance (4/41),

in particular to retrace *Data Semantics* (3/4). The visuals in the **SelectionOrigin** of whether they were created from group operations on other selections helped in one event experts *Recall* the meaning of a selection (1/4). In another, changing the name and color of their query (e.g., call a selection “*optimum*”, CS-w) helped them at a later stage to *Recall* what selections represented a good subset and what they were optimizing (2/4). In one event (1/4) seeing the number of datapoints in a selection allowed them to reflect on their exploration (*Meta Analysis*) and comment “yes, here instead of selecting 73 points for the green selection [...] as we did in the past exploration we just selected 3 or 4 points, it’s the same idea in today’s exploration we drew a line on which we aligned points [...]” (CS-w†).

Our participants also mentioned verbally (5/41) that loading and viewing their past history with the different provenance tools helped their analysis. They did not single out specific tools, but commented in the Discussion (Part 5 of the study) on how the visualizations holistically helped them retrace their past *Exploration Steps* (1/5) for *Action Recovery*; allowed them to *Recall* what they *Focused* on in past explorations in terms of dimension coverage [**T7:ObserveCoverage**] and also the combined dimensions they created [**T6:CombineInterestZones&dimensions**] (2/5); helped them *Recall* selections that represented *Interest zones* [**T1,T2: Minimize&maximize; Locate&characterize**] (1/5); and enabled them to reflect (*Meta-Analysis*) on the effectiveness of their *Exploration Approach* (1/5).

*Overall*, we observed that our tools were used to help experts:

- (i) recall important areas in the trade-off space in the form of selections (*Interest zones* [**T2:Locate&characterize**]), and what these selections represented and optimized (*Semantics* [**T1:Minimize&maximize**]). In total, 13/41 events (SelectionRange, History, Favorites, Query);
- (ii) keep track of the *Coverage* [**T7:ObserveCoverage**] of objective dimensions they had already visited and what remained. In total, 10/41 events (ViewCount, DimensionCount); and
- (iii) recall what objectives were of high *Priority* [**T4:Rank&prioritize**] (7/41 events - ViewCount, SelectionRange);
- (iv) retrace past *Exploration Steps* (7/41 - ViewCount, History);
- (v) or to *Validate* and reflect on the effectiveness of their *Exploration Approach* (4/41 - DimensionCount, Query).

### 3.3.3.2 Provenance at different exploration parts [RQ2,RQ3]

To understand if the provenance tools were used to recall past exploration [RQ2] or to aid the current analysis [RQ3], we focus respectively on events that happened during the Reproduce Insight With Log Part and the Open Explore Part (28/41). The remaining events come either from the Discussion or before the logs were loaded in

the provenance tools (when only Favorites and History were available) and are not discussed further in this section.

Interestingly, the provenance tools were used actively both during the Reproduce Insight With Log when they were revisiting past explorations (17/41 events) and during the active Open Explore (11/41 events). The longer Open Explore (30 min) had fewer events than the Reproduce Insight With Log (15 min), but this can be expected, as in the latter the participant's goal was to reflect and reproduce their past analysis—so provenance was at the forefront; whereas in Open Explore their primary goal was to explore new aspects of the dataset. Thus, we are more interested in trends of use between the two parts, rather than a direct comparison of numbers. To that end, we provide percentages of events within each exploration Part.

**[RQ2]** In Reproduce Insight With Log where participants were trying to reconstruct their past analysis, they mainly used the ViewCount tool (41.2%), followed by DimensionCount (23.5%) and Favorites (11.8%). We also observed use of the Query tool to reflect on the size of a past selection (5.9%). The remaining events are general comments about how the tools helped them to revisit past findings, but without specifying which tools. We did not observe use of the SelectionRange, the SelectionOrigin, nor of the History tool.

Unsurprisingly, most events relate to the general provenance purpose [31] of *Recall*-ing the past exploration (70.6%), with a few events where experts reflected on the quality of past exploration (*Meta Analysis*) (17.6%), or attempted *Action Recovery* (11.8%). When looking at the more specific provenance objects, participants mainly used the tools to track the *Coverage* of dimensions they had visited in the past exploration (35.3%) and to remember which ones were their *Focus & Priorities* (23.5%). The other provenance objects were less represented in this study part, with only a few events where participants tried to *Validate* (5.9%) or reflect on their *Exploration approach* (17.6%), retrace detailed *Exploration Structure & Steps* (11.8%) or remember specific *Interest zones* (5.9%).

**Overall**, we confirmed that participants used the in-visualization provenance when revisiting past findings. Based on the provenance objects they focused on, high-level summaries of what was covered and what were the priorities in the past exploration, are key.

**[RQ3]** The use patterns differ in Open Explore, with tool use being more distributed. ViewCount and SelectionRange were the most used (27.3% each), followed by History (18.2%), SelectionOrigin and Favorites (9.1% each). The few remaining events are general comments about their usefulness without specifying which tools. We did not observe any use of the DimensionCount tool.

These events point out to a different distribution of Ragan et al. [31] provenance purposes. While *Recall* remains important (45.5%), the main purpose is now *Action Recovery* (54.5%). The reasons why participants used the tools also differed. More

than half the time it was to keep track of *Interest zones* (54.5%), followed by retracing their *Exploration Structure & Steps* (27.3%). Less frequently, they used it to verify *Coverage* and *Data semantics* (9.1% each).

*Overall*, we confirmed that experts also used in-visualization provenance while conducting a traditional exploration analysis. Based on the provenance objects they focused on, here participants were mainly interested in tracking more detailed information, such as keeping a trace of specific areas in the trade-off space that were of interest to them and their characteristics, as well as their analysis steps. Nevertheless, we did observe events where higher-level goals, such as keeping track of dimension coverage, were met.

### 3.3.4 Opportunities for provenance visualization [RQ4]

Our goal in using VisProm as a technology probe was to identify situations where new designs could aid users in their tasks. We coded several events across use cases that presented opportunities for provenance (CS-w: 35, CS-eco: 29, CS-ml: 27). We chose to report such events as they provide concrete and actionable opportunities for provenance in trade-off analysis.

We found opportunities in Study Part 2 where experts were tasked to replicate past events **without** the additional aids of log data and provenance visualization (31.86%), but also in Reproduce Insight With Log (27.47%), Open Explore (24.17%), and the Discussion (16.48%). These provenance opportunities are analytical in nature (93.4%), although we also describe opportunities for data provenance related to data semantics and origin (6.6%). Below, we report on these opportunities using the general provenance purposes in [31]: recall (**73.62%**), replication (**12.08%**), meta-analysis (**12.08%**), and presentation & collaboration (**2.19%**).

#### 3.3.4.1 Recall Opportunities

We observed more opportunities for provenance *Recall* than any other type, as often-times experts struggled to remember previous interest zones (35.82%), focus & priority (20.89%), exploration structure & steps (20.89%), data semantics & origin (11.94%), coverage (7.46%), and verification & validation (2.98%).

**Recall *Interest zones***: When trying to recall previous interest zones, expressed in VisProm as colored selections, experts discussed but often struggled to fully remember what they saw as important trade-offs and their characteristics. We observed opportunities for tracking and visualizing provenance of: **(i)** the view/s in which the interest zone (or selection) was defined; **(ii)** selection size, range and distribution; **(iii)** the thresholds and cut-off points they used to constrain the selection; **(iv)** when relevant, the decision boundary they used to interpret the selection, for instance, whether the data points are above or below the bisector; **(v)** the nature of the trade-off selection,

e.g., minimization versus maximization. In addition, experts also raised the need for tracking contextual information (**vi**): for example, to differentiate between past selections that express an insight from the intermediate ones that led to it; sometimes even remembering the research questions or hypotheses that motivated them to create selections can be hard. Interestingly, experts used unexpected cues to help them recall the reasons behind selections, such as green color for preferable solutions and red for undesirable or costly ones.

**Opportunities.** *Creating and refining interest zones (e.g., via selections) is crucial in trade-off analysis, as it helps analysts progressively learn the multi-criteria problem characteristics, determine most preferred solutions and compare alternative ones. However, experts struggled to remember key aspects of those selections not only when revisiting them after a period of time, but also during the same exploration session. We need to enrich trade-off selections with summary information and visual cues to help analysts quickly understand or recall the nature of the trade-off, and the rationale that led to create those selections (e.g., to minimize or maximize certain objectives) [T2:Locate&characterize]. Color is an interesting visual encoding for provenance as it appears experts assign meaning to hue and are sometimes able to remember this meaning without access to labels. While in some cases experts were able to reconstruct what objectives were maximized, what selections were important and fell within specific boundaries or had specific patterns, we believe there is an opportunity to help them track this more explicitly. For example, we can allow them to set thresholds that the system can keep track of (beyond min and max), and let them express what patterns are of interest (data above bisector, or of a specific distribution) and why, and allow them to differentiate between selections that are intermediate steps in their analysis and ones that express important findings [T1,T2: Minimize&maximize, Locate&characterize].*

**Recall Focus & priority:** Domain experts often mentioned some objectives are more important than others, and this drove their exploration strategy. They sometimes reported they could not remember all of the objectives that they tried to optimize, but VisProm (DimensionCount, ViewCount) helped them with recall. We observed that participants were likely to remember plots they found interesting, such as in terms of their dimensions and shape, but not so accurately or easily the steps they took to get there. Moreover, sometimes participants were also able to guess why they did not focus on a particular dimension or view.

**Opportunities.** *During trade-off exploration, analysts need to recall what objectives drove the exploration and their priority [T4:Rank&prioritize]. Showing the individual or combined criteria they created, and the interesting views they visited (even highlighting those that are not interesting) can help them infer their overall trade-off exploration approach. Besides communicating focus and priority, analysts can themselves express these, possibly visually to indicate objectives that have not been explored*

yet [T7:ObserveCoverage]. *It is interesting to note that we observed participants attempting to keep track of priorities both when revisiting past exploration, but also during the time they were actively exploring to help strategize and plan their next actions.*

**Recall *Exploration structure & steps*:** Participants had difficulty remembering the successive steps they took during the analytical process. Experts from CS-w also struggled to recall the first step that triggered a particular exploration path, and this was important as the order in which they visited the trade-offs mattered to them. When reasoning in terms of stages of exploration [112] rather than individual steps, participants were able to remember more, such as the part of the exploration where they analyzed an important trade-off or when they found an important insight.

**Opportunities.** *Overall participants were able to remember phases of the exploration more than they did for steps, but not all stages were remembered indicating that different parts of the exploration may require different support. Besides showing phases of exploration as part of provenance, trade-off exploration might benefit from explicitly showing alternative or branching paths taken during the analysis, and emphasizing similarities and differences, as analysts are interested in making comparisons [T5:CreateBranches&CompareAlternatives].*

**Recall *Data semantics & origin*:** We had several events where participants failed to remember the semantics of the data they were exploring. This happened at three different granularities: **(i) Dataset:** we had events where participants mentioned explicitly the origin of data and relations that were not captured in the dataset; **(ii) Dimension:** In some cases experts struggled to remember what each dimension of their dataset captured. This is particularly important for trade-off analysis, as often experts combine dimensions that express objectives that are related (e.g., in CS-eco one expert had to reflect before confirming that a dimension in the dataset combined two others, and in CS-w another expert mentioned how remembering this is important); **(iii) Value:** We also observed events where experts were attempting to optimize dimensions, but could not remember if the objective needed to be maximized or minimized. There were also cases where they had trouble remembering the meaning of specific values or ranges.

**Opportunities.** *We need to provide data provenance at different levels - not only to enable experts to keep track of the origin of data, but also to capture and be able to retrieve the semantics of dimensions and values. This is particularly important for trade-off analysis, where some dimensions may capture more than one objective, and where the experts need to track the direction (range) that optimizes specific objectives [T6:CombineInterestZones&dimensions].*

**Recall Coverage:** Participants had difficulties keeping track of data dimension coverage when they had limited provenance support (Reproduce Insight Without Log). However, when we provided them with provenance visualization tools, they commented on how these aided them in their task. When recalling past exploration, they also wanted to know whether the dimensions they visited came from single or multiple sources (data files).

**Opportunities.** *Given the high dimensional nature of trade-off spaces, there are opportunities to improve the exploration by keeping track, as the analysis progresses, of what data dimensions or sources have been explored, and highlighting areas that are under-explored. This can potentially show an implicit exploration bias or an intended prioritisation in analysing trade-offs [T7:ObserveCoverage].*

**Recall Verification & validation:** Our experts were also interested in finding out whether important known insights were confirmed in their previous (and current) analysis, thus validating the approach that was used to generate the trade-off dataset.

**Opportunities.** *Provenance visualization could make it clear whether the explored dataset has gone through a validation procedure (whether manual or automatic). This can give analysts confidence in the data and avoids them repeating the verification & validation procedure in any future explorations.*

#### 3.3.4.2 Replication Opportunities

The provenance opportunities we observed to support *Replication* during trade-off analysis are mostly related to *Exploration structure & steps* (81.81%). Experts attempted to reproduce sequences of events in order to replicate an insight or a lost selection. The replication of previous analysis could be an important step when resuming work carried out a long time ago. We found that experts might want to repeat previous analyses from the beginning to refresh their memory, or to gain more confidence in their findings, but often with variations, such as by examining additional dimensions or a new dataset. This could occur in a branching or refining manner. Another sub-goal of replication is to compare alternatives. For example, experts were interested in comparing alternative solution sets in terms of differences and similarities.

**Opportunities.** *There are opportunities to support replication during trade-off analysis by assisting analysts in resuming major or key phases of their exploration, including important selections and dimension groupings. Different types of replications should be supported, whether to continue the same analysis, or to create a new branch or a refinement from an old one [T3,T5:Cascade&refine, CreateBranches&CompareAlternatives].*

### 3.3.4.3 Meta-Analysis Opportunities

Participants reviewed the analysis process, often to understand their *Exploration approach* and structure (90.90%), or to reflect on their choice of *Interest zones*. The majority of those opportunities relate to events that occurred during Parts 3 and 4 of the study (72.72%), both with provenance visualization tools enabled. For example, participants in CS-ml wondered whether the next step would be a fresh start or a new branching session extending from the previous analysis. In CS-eco, when we asked participants to reflect on why they wanted to keep the provenance history, even though they were starting the open exploration (Part 4), they stated that previous analysis could provide new inspiration or good starting points.

***Opportunities.** Automatic provenance tools can support meta-analysis, for instance, by showing similarities and differences between exploration sessions. They may provide useful information and metrics, to help analysts reflect more deeply on their trade-off analysis choices. This in turn, may provide inspiration and valuable starting points for the following steps of the exploration.*

### 3.3.4.4 Presentation & Collaboration Opportunities

The views experts visit can become complex over time as they create combined dimensions and new selections (e.g., using intersection). They might want to *Annotate* such views or take notes to facilitate revisiting and sharing. Whereas one domain expert noted that it could be useful to directly annotate interesting views, another expert found annotations a burden and add complexity to the tool. The question of annotation was also raised in CS-w where domain experts came to our study prepared with notes describing the research questions and hypotheses they wanted to test.

***Opportunities.** The role of annotations as a provenance tool for trade-off analysis merits further investigation, as our observations show that annotation support could help analysts address the growing complexity of views as the analysis progresses, but with caution as to not distract from the main trade-off task.*

In section 3.4.4, we will discuss further these opportunities and highlight those that are generic and concur with existing findings, and those that are new and more specific.

## 3.4 Discussion on Provenance for Trade-offs

The use of our provenance-augmented probe by three groups of domain experts allowed us to study trade-off analysis and to distill new findings with regard to the tasks people engage in during trade-off analysis and how provenance can support these tasks



both when conducting a-posteriori and active analysis. Moreover, it helped us identify opportunities on how to better support such analysis tasks.

### 3.4.1 Trade-off analysis tasks

In subsection 2.2.2 we theorized that five of the tasks presented in Hakanen et al. [59] are relevant to our work, even though, unlike their approach, we do not integrate an interactive multi-objective optimization procedure. While this is generally true, we went beyond this past work. First, we found a case where our tasks are more refined and detailed, and second, we identified additional trade-off analysis tasks not reported in [59]. We contributed a set of refined tasks that fall under their *determine most preferred solution: minimize & maximize, rank & prioritize, and combine interest zones & dimensions*, which are all tasks our participants followed to narrow down and converge to their preferred solutions. We also further contribute two new tasks: *cascade & refine* and *observe coverage*, that our participants found to be important for trade-off analysis.

Their other tasks have a direct correspondence to the ones we identified. Their task *compare Pareto optimal solutions* corresponds to our more general task *create branches & compare alternatives*. Their *learn about problem characteristics* and *detect correlations* tasks directly map to our *locate & characterize* and *find clusters, correlations & outliers* tasks respectively. Thus, together with the refinement of the *determine most preferred solution* task, we independently identified several tasks reported in operations research work [59]. This replication indicates our technology probe was effective in aiding us to identify high-level tasks that are generalizable across tools and research domains. The *post-processing* task is beyond the scope of our study, as we did not do a follow-up with our participants to investigate whether the insights gained from the exploration session helped them conduct new experiments or generate new trade-off datasets.

Our trade-off tasks of finding clusters, correlations & outliers, also correspond directly to high-level tasks in traditional analytic task taxonomies, such as the tasks identified by Amar et al. [125] on finding anomalies, clustering and correlating. Other high-level trade-off tasks like observing coverage and defining focus & priorities (that are key in trade-off analysis), do not have a direct correspondence. This is also the case for trade-off tasks such as creating branches & comparing alternatives, looking for interest zones that maximize/minimize objectives, and characterizing interest zones. These different tasks remain fairly high-level and do not have a direct equivalent in analytic tasks; nevertheless, in order to achieve these trade-off tasks, our experts engage in a combination of low-level analytic activities [125], such as deriving threshold values, filtering, sorting, characterizing distributions, finding extremum, etc.

While scatterplot matrix tools are popular in trade-off analysis visualization[82, 140], it is natural to consider if these refined and new tasks may be specific to the

scatterplot matrix design of our probe that favors sculptured selections. We believe the majority of our identified tasks are generalizable and not specific to our probe design. A scatterplot matrix view favors the identification of relationships between two dimensions at a time (e.g., correlation[34]). Nevertheless, our tasks relate either to tracking progress and prioritizing (rank and prioritize dimensions, track coverage of multiple dimensions) or to understanding possible solutions in the search space across multiple dimensions (e.g., identifying possible solutions that minimize or maximize a dimension of interest and characterize zones of interest across multiple dimensions).

It is possible that other tasks, such as cascading & refining of interest zones, may indeed be influenced by how analysts select regions of interest in our tool, using sculptured queries; or the combine multiple dimensions task may be due to the fact that a single scatterplot can only show two dimensions at a time. For the latter, we feel combining dimensions makes sense irrespective of visualization, our experts often had dimensions (like aromas) that they semantically grouped together and wanted to treat as a group.

### 3.4.2 Provenance objects in trade-off analysis

We identified eight provenance objects that characterize provenance purposes from previous work [31]. These are objects that our experts wanted to recall, replicate, recover, reflect upon, or present & communicate. Out of these objects, we found that interest zones and exploration structure & steps are the most important, as our domain experts frequently used or mentioned them during their exploration. They should thus be supported by provenance visualizations in future trade-off analysis tools.

Our provenance objects do not all reside on the same analytical or abstraction layer [30, 84]. Some are directly related to exploration actions [30] and thus can be tracked using low-level user interactions such as view and query selections (interest zones, exploration structure & steps, focus & priority, coverage). Other provenance objects are situated at a higher level pertaining to insight actions (data semantics & origin, annotation, verification & validation), or meta-actions (exploration approach). Tracking and visualizing provenance data related to exploration actions is more straightforward and is easily supported in interactive systems, whereas provenance for higher abstraction layers is more challenging to capture and visualize, although progress is being made, for example, in automated data insights [120]. This highlights the need for further research in detecting and visually summarizing insight actions and exploration structure to support trade-off analysis.

Provenance objects express 'what' aspects of the analysis experts need support with (to recall/replicate/etc), and were observed either through the use of the provenance tools and experts' discussions (section 3.3.3.1), or observed opportunities (section 3.3.4). Some of these provenance objects directly map to, and support, identified trade-off tasks (section 3.4.1). For example the direct mapping between the 'Focus &

priority' object and 'Rank & prioritize' task, and between the 'Coverage' object and the 'Observe coverage' trade-off task. We also observed analysts recall/replicate/reflect upon other objects, namely interest zones, especially for trade-off tasks such as 'Minimize & maximize' and 'Create branches & compare alternatives'. Or the 'Exploration structure & steps' object that often appeared as an opportunity, expressing the need to keep track of past branches and alternatives for the task 'Create branches & compare alternatives'. These tasks of 'Minimize & maximize' and 'Create branches & compare alternatives' relate to trade-off tasks from previous work [59]: 'determine most preferred solution' and 'Compare Pareto optimal solutions' respectively.

We thus feel the objects 'Focus & Priority', 'Coverage', 'Interest zones' and 'Exploration structure & steps' (to create branches and compare alternatives) are central to trade-off analysis and require further exploration in order to design provenance views that are well-adapted to such analysis. The remaining objects we observed (Exploration structure & steps for action recovery rather than branches and alternatives, Data semantics & origin, Exploration approach, Verification & Validation, and Annotation) could be relevant to any type of visual analysis exploration. Future work may also identify more trade-off tasks and related objects.

The nature of our study (revisiting past findings and active exploration), naturally relies on our participants remembering their past analysis, and thus focuses on replication or recovery of past results or choices. Thus our findings often link provenance objects with the provenance goals of Recall or Replication (subsection 3.3.2). We see these same objects (e.g., Interest zones) also appear with other purposes such as 'Recover' and 'Present, so we expect these objects are applicable across provenance goals. Nevertheless, further studies are needed to investigate whether different types of analysis, focusing less on recall and replication, and more on reflection and meta-analysis, could clarify which objects are important for which provenance purposes, or even reveal other types of provenance objects associated with other purposes.

### 3.4.3 How in-visualization views support trade-off tasks and provenance objects: during *a-posteriori* analysis and *ongoing* exploration

Most work evaluating provenance considers provenance itself as the main user goal, for example to understand and recall past exploration. Fewer consider the impact of provenance views during the exploration itself (see section 2.5). We consider both together, in a structured evaluation of experts conducting real-world analysis.

This poses a challenging question of design balance: as data exploration is the focus of our experts, can we use provenance to support both *a-posteriori* analysis and active exploration? Motivated by past informal observations that experts largely ignored separate provenance views, we decided to integrate in the analysis environment

simple designs that do not add visual complexity but are always accessible. We set out to see if and how these designs were used, both a-posteriori and during the analysis, and how they tied to trade-off needs. Our findings show that our domain experts found in-visualization provenance views useful both when revisiting old analyses and when starting new ones.

In *a-posteriori exploration*, when revisiting old analysis, domain experts were able to first confirm past findings in the use of traditional views, such as History and Favorites. The use of a stripped down History that only showed a final snapshot of the last exploration, and Favorites were a great aid for recall. Our experts explained that they preferred these few snapshots rather than detailed step-by-step recollection of previous analyses. Other studies investigated what granularity or level of detail of provenance information is appropriate for different tasks [32, 103, 157]. Most relevant to our work is a user study by Ragan et al. [32] who found that even lightweight provenance visualizations (such as ours) including simple in-visualization cues and low-fidelity snapshots of final workspaces or important views that contain an insight can be beneficial for recall and helping the process memory.

We were also able to extract patterns of use from our new in-visualization views during this a-posteriori exploration. Our experts used provenance views that mainly communicated high-level information, such as views with counts of dimensions explored or scatterplots visited. In particular, we identified that these views aided experts mainly to recall or reflect on: their coverage of the objective dimensions, their focus and priorities in terms of objective dimensions in their past analysis, and to a lesser extend their past exploration strategy. As mentioned, understanding coverage and focus & priority were identified both by us, but also previous work [59], as critical for trade-off analysis irrespective of the tool used. As such, this finding highlights the importance of providing provenance support for them when a-posteriori analysis is expected.

Whereas in the *ongoing* analysis, experts used our in-visualization views focusing on more detailed information. For example, they used them to keep a trace of specific areas in the trade-off space that were of interest to them and their characteristics, as well as their analysis steps. Our experts often consulted what range their selections fall in in the trade-off space, or what selections were grouped to generate new ones. These point to the need to support recall exploration steps, but most importantly the recall of aspects related to interest zones (one of the trade-off provenance objects identified in our work). Aspects that may be interesting to recall relate to several trade-off tasks that are associated with characterizing preferred solutions or interest zones: for example if the interest zones maximize or minimize objectives, if they fall under specific thresholds, if and how they have been refined, if they represent an alternative branch, etc.

As our observations of ongoing analysis stem from a scatterplot matrix probe, it is possible that the importance of interest zones may be influenced by the existence of

visual sculptured query selection. However, as discussed earlier, the task of ‘determine preferred solution’ that optimizes some dimensions is very general in trade-off analysis. And the characterization of these possible solutions or areas of interest (e.g., where they fall across other important dimensions) includes simultaneously maintaining awareness of multiple dimensions and thus is very likely applicable to any multi-dimensional visualization tool.

Provenance is a secondary task supplementing data exploration, but we observed many cases where our visually simple and in-visualization provenance tools were used effectively, both when reconstructing past trade-off analyses and during an open-ended analysis. More work is needed to investigate whether provenance support needs to adapt to more specific and refined stages of these analysis. For example, during an active exploration, we can expect that the start of the analysis will be similar to a-posteriori analysis: verifying and validating past knowledge, checking coverage of dimensions and setting priorities for the exploration, in order to ground the upcoming exploration. Whereas the middle of a trade-off analysis exploration, it is possible that the main considerations are keeping track of alternatives or of characterizing the identified interest zones. And towards the end of an exploration, when experts are close to converging and sharing of results, it is possible that tracking coverage and reflecting on the exploration strategy may once again become important. More studies are needed to compare different design choices for provenance as a secondary task, including understanding better potential exploration phases [5], what provenance objects are important in these more refined phases, and how to draw users’ attention and awareness [153] to the visual cues when the information is needed.

#### 3.4.4 Trade-off specific opportunities

The use of provenance tools by our study participants helped us identify different types of opportunities for recall, replication, meta-analysis, representation and collaboration. These opportunities are directly related to actual challenges our study participants faced during trade-off analysis, for which we thought provenance could help overcome. The general provenance opportunities that are discussed in the literature confirm some of the provenance needs we identified for trade-off analysis, such as the recommendation for capturing provenance information with varying levels of detail [31] and the associated opportunities for multi-layer provenance [84]. The overlap between our recommendations and those in the provenance literature is expected, as trade-off analysis is a subset of and a specific type of data exploration.

Moreover, we covered new ground with regards to the provenance objects associated with the identified trade-off tasks. Our opportunities for helping analysis recall/replicate/reflect indicate these act upon objects that can inspire future designs, focusing on: interest zones, exploration structure and steps (like refining and compar-

ing alternatives), focus & priority, and coverage. In particular, our call for supporting provenance capture, recall and replication of focus & priority argues for storing and exploiting information highly relevant to trade-off analysis such as the objectives users focus on, and how these are prioritized, which is intrinsic to trade-off analysis, but less pertinent in general data exploration and provenance.

### 3.4.5 Limitations

While the use of the provenance probe helped us identify opportunities and challenges, it may have also influenced our findings. Our participants had access to a specific visual representation (SPLOM) and in-visualization provenance views, which may have focused their feedback on tool-specific issues. We believe the high-level findings, such as many of the identified trade-off tasks, provenance objects and opportunities hold irrespective of the underlying analysis tool; nevertheless, it remains future work to replicate them in different trade-off analysis environments. Our probe was also used by experts from different domains, and we report our findings collectively. This we hope provides a broad coverage of user needs when it comes to trade-off analysis, but it does not help us understand if the nature of provenance requirements is common across experts. Future work should consider if domain plays a role in both trade-off analysis tasks and provenance needs. Finally, we looked at pairs of analysts exploring trade-offs simultaneously. It is possible different provenance objects and even tasks emerge when analysts work asynchronously, for example during hand-off or when establishing common ground.

## 3.5 Conclusion

In this chapter, we reported on an observational study with three groups of domain experts who used a provenance-augmented probe to explore their own trade-off datasets. Our aim was to investigate whether and how in-visualization provenance can support real-world trade-off analysis. Our findings show that domain experts engage in different high-level tasks that are key in trade-off analysis, such as observing coverage, defining focus & priorities, branching alternatives to compare them, and looking for interest zones that maximize/minimize specific objectives. These tasks were supported by our in-visualization provenance views, which helped our experts recall, replicate, recover, reflect upon, or present & communicate different aspects of their exploration, that we identified as trade-off provenance objects. From these, keeping track of coverage and priorities, and to a lesser extent exploration strategy, was the most important provenance aspect (or object) for our experts when revisiting past trade-off analysis; whereas interest zones and exploration structure & steps were crucial when starting new trade-off explorations.

---

These findings highlighted several needs that are specific to trade-off analysis that we felt required further investigation, namely the need to understand interest zones (i.e., data-point selections) that maximize/minimize specific objectives, and the observation that experts often branch alternatives to compare them, while their focus and priorities may be shifting. This led us to our following investigation: how to aid analysts understand the pros/cons of different interest zones, which may be branches or independent solution spaces, with regard to their focus and priorities.





## Designing for Trade-off Comparison

The results of our previous chapter highlighted the need to identify and characterize interest zones (groups of points, often in specific ranges), or alternative branches (i.e., alternative groups), with respect to what objectives or criteria they maximize/minimize and how to do this characterization under potentially shifting focus and priorities. To explore these concepts in more detail, we study a more explicit task - instead of the task of characterizing different zones (groups of points, alternatives) we consider how to explicitly *compare* them with respect to what they optimize and the expert's priorities. We make the assumption that if analysts can characterize the zones they can also effectively compare them and vice versa. Importantly, in our observations of our experts we have instances of explicit comparisons of zones of interest that we will discuss in the next section.

We note that it is not surprising that our experts are analyzing and comparing interest zones (groups of points). As we mentioned in the Introduction of this thesis, experts often analyze simulation results from biological processes and models that have inherent variability in measurements. So they are used to using ranges when identifying zones of interest, like our example of the wine experts in the previous chapter that considered the benefits of groups of wines with different temperature management strategies - their temperatures were defined as ranges. It is also often the case that trade-off datasets include categorical variables that are semantically important; for example, our ecosystem experts compared sustainable farming practices across different countries, defining their interest zones by country.

This chapter answers [RQ5], i.e., what are the trade-off comparison needs of domain experts and how can we design to support them. To understand comparison needs during trade-off analysis, we first analyze how comparisons are currently carried out in real-world settings by revisiting past trade-off exploration sessions and conducting three new workshops with domain experts. This answers the first part of our research question [RQ5a]. We then describe the results of two design workshops with visualization and HCI researchers that explore design and visualization needs to

support trade-off comparisons. This answers the second part of our research question about what designs we can provide to support them [RQ5b]. These findings together inspired the prototype system we developed to support trade-off comparisons. The evaluation of the designs generated here is reported in Chapter 5.

This work was published in part in *Journée Visu* [98] and supplementary material is available at [96].

## 4.1 Expert Needs and Requirements [RQ5a]

In the previous chapter, we explained how we worked closely with eight domain experts to understand their comparison needs as they carried out trade-off analysis in real-world settings (i.e. working with their own data and/or models, and domain-specific questions). The experts came from different domains: two from agronomy, two from ecosystem services, two from ML, and two from nutrition. To distill general requirements for trade-off comparisons, we first revisited three past trade-off exploration sessions that we carried out for the provenance research work in chapter 3.

To ensure we understood the subtleties of comparisons in trade-off analysis, we further organized three trade-off exploration workshops with domain experts. For the first workshop, we invited back the two Machine Learning (ML) experts whose videos we analyzed in section 4.1.1.

Furthermore, in an attempt to cover as many types of possible comparisons during trade-off analysis tasks, we considered two types of trade-off datasets characterized by the cardinality of trade-off solutions they contain: *discrete* (non-additive) trade-off solutions, or *composite* (additive) solutions. The former concerns cases where valid trade-off solutions are discrete individual options, as was the case for the wine fermentation case study where each individual data point represents an alternative wine recipe (or more precisely, a valid fermentation trajectory). The latter refers to cases where a *group of points* together constitutes a coherent trade-off solution, such as when creating a diet plan out of a dataset of food items; or putting together a sports team starting from individual athletes, each having different strengths and weaknesses.

### 4.1.1 Videos of Expert Exploration

We first reviewed  $\approx 8$  hours of video recordings and notes from three past trade-off analysis sessions [97, 112], each involving a pair of domain experts from sustainable farming, agronomy, and machine learning. One pair of experts focused on ecosystem services (eco), explored a 9-D (dimensional) dataset with 100 data points; another with expertise in wine fermentation analyzed a 14-D dataset with 1k data points (wine); and a third pair, specializing in ML models, worked with a 10-D benchmark dataset [13] of 3k data points.

We identified six instances of explicit comparisons, with two in each group. It is important to note that previous analysis sessions did not prompt participants to make explicit comparisons between data groups, as this was not the main focus of that research. Consequently, many comparisons likely occurred implicitly while participants familiarized themselves with the data, without being verbally articulated. The six instances we identified are those where participants explicitly mentioned comparisons, usually in response to particularly surprising findings or insights that warranted direct discussion.

Across all sessions, we observed that experts worked with *groups* of data points, treating them as a coherent *solution* to their multi-criteria decision-making problem. For instance, the ecosystem services experts identified farms from specific countries that represent regional farming practices and traditions. For wine experts, a cluster of wine recipes corresponds to a fermentation strategy characterized by a particular temperature profile. The ML experts partitioned their dataset into three distinct groups of points, each corresponding to a particular ML model, searching for qualities inherent to each model. Within each group, data points represent various trade-offs between the number of features in a model and the model's accuracy. Overall, experts aimed to *rank* alternative solutions in these comparisons to select the most promising option to implement.

While the identified instances of explicit comparisons in the videos were few (2 per group), they were crucial for the experts, as they often led to an important decision, like choosing which costly biological experiments to run in a controlled laboratory study or even in a real-world setting. These comparisons involved groups of points. The groups were few in number (2-6 groups per comparison) and their size varied across sessions but often tended to be small (4-5 for wine, 8-43 for eco), except for the ML case (17-1090), likely because the dataset itself was large and the number of points in a group was inherent to each benchmark dataset, with some having significantly more data points than others.

These groups represented a selection or a set of points that shared similar values over one or more data dimensions (e.g., a specific fertilization trajectory or a temperature profile). They compared these groups in terms of their value ranges, the overlap between these ranges, and their relative position in the data (e.g., if the centroid of one group tended to be towards the max or min compared to another group). Experts also looked at the number of points of each group and explored relationships within the groups, such as trends or correlations.

### 4.1.2 Workshops with Experts - Comparison of Discrete Solutions

We invited back the two Machine Learning (ML) experts whose videos we analyzed in section 4.1.1, and their actual exploration described in chapter 3. As a reminder, these experts explored an ML benchmark suite called Openml-cc18 [13], focusing on trade-offs such as those involving the number of features of a benchmark dataset, its intrinsic dimensionality (i.e., dimensions that exclude noise) and the average correlation between features. As the number of features grows, the intrinsic dimensionality is likely to grow, while the feature average correlation is likely to decrease.

**Workshop Procedure and Tasks:** The workshop lasted 90 minutes. We first welcomed participants and explained the goal of the workshop, which was to understand their comparison needs. To ground the discussion, we presented the instances of comparisons identified in their exploration video and asked them to expand on the nature of these comparisons. We ended the workshop with a debrief.

**Workshop Observations:** The machine learning experts reported that they usually compare groups of points such as datasets and clusters, confirming our video observations. As one explained, “*What normally happens is that we try to compare groups inside the same [benchmark] dataset. So maybe clusters, maybe classes, stuff like that... In the last few works, we were mainly looking at specific trade-offs for the models’ accuracy and complexity*”. When asked about their focus during these comparisons, they reported examining the relative positions of the groups: “*You might be interested in assessing whether the two groups overlap for one [ML] metric or another for one dimension over another, or how much they overlap*”.

This claim is further supported by our observation from the videos that all expert groups worked with ranges and noted the overlap between them. When inquiring further about the criteria they use to compare groups, such as ranges or specific values (e.g., minimum/maximum), they indicated that the choice of metrics is problem-dependent. They did, however, provide some examples: in some cases, they may need to identify and remove outliers from the analysis or focus on the relative best points in each group. This suggests a need for ranking the points within a group. In other cases, they compared the group ranges or checked if they fell above or below specific threshold values.

Additionally, they mentioned instances where they compared the means and variances of the groups. These observations are also reflected in the videos of other experts conducting trade-off analysis.

### 4.1.3 Workshops with Experts - Comparison of Composite Solutions

In two separate workshops we call here *Diet I* and *Diet II*, we invited two participants with extensive knowledge in nutrition and dietetics to explore a food trade-off dataset. The participants themselves had specific dietary needs due to health constraints or targets, which they communicated to the author of this manuscript before the workshop. We describe their trade-off needs below, before explaining the workshop procedure and task, followed by key observations.

*Diet I:* This workshop session was conducted with a participant who, due to a medical condition, needs an anti-inflammatory diet with a focus on weight loss. The participant follows a flexitarian approach but aims to minimize or completely avoid red meat. Their diet prioritizes key nutrients for minimizing the symptoms of their condition, like magnesium, zinc, iron, and omega-3 fatty acids. At the same time, they work to limit their intake of omega-6, high glycemic index carbohydrates, glucose/fructose, and gluten, all of which can interfere with their anti-inflammatory and weight loss objectives. All of these constraints are hard to balance, and trade-offs need to be made to ensure a diet that is both nutritionally complete and aligned with their health goals.

*Diet II:* This participant followed a specific diet due to particular dietary needs. As a vegetarian who is also lactose and soy intolerant, they faced challenges in maintaining balanced nutrition, particularly when trying to meet an adequate protein intake. For example, they compensated for their inability to consume dairy products and sufficient amounts of soy for protein intake through legumes. Additionally, the absence of dairy products in their diet made it more difficult to meet their calcium needs, which they addressed by incorporating almonds and tahini. However, relying on these foods poses its own set of challenges since they are high in fat and calories. Thus, the participant was interested in finding a subtle balance between the different nutrients they needed to consume to maximize the coverage of their nutritional needs while minimizing the excessive intake of certain nutrients.

**Workshop Procedure and Tasks:** The first workshop (*Diet I*) lasted 1.5 hours, while the second (*Diet II*) lasted 2 hours. We loaded the USDA Foundation foods dataset <sup>1</sup> in the visual data exploration tool VisProm presented in chapter 3. This dataset has 20 data dimensions and 209 data points. Each data point corresponds to a food item, and the dimensions represent macro and micro nutrients. The dataset was prepared to include as much as possible the objectives and diet requirements of the two participants.

Inspired by the participants' self-reported dietary constraints, we designed the following three trade-off tasks that we hypothesized would require making comparisons.

---

<sup>1</sup>Dataset publicly available from <https://fdc.nal.usda.gov/download-datasets.html>

Tasks 1 & 2 were performed by the two participants, and the more challenging Task 3 was only performed by the second participant:

- T1 (For Diet I&II) Create two diets, with 40 items each, and having at least 50% dissimilarity between them: One that maximizes energy (calories) while minimizing sugar and fat. This diet should not have more than 5 carbohydrate-heavy foods (>5g); And another diet that Maximizes energy (calories) while minimizing sugar and carbohydrates. This diet should not have more than 5 fat-heavy foods (>5g).
- T2 (For Diet I&II) Compare these two diets in terms of nutrients (iron, magnesium, fibers, vitamins,... ) and choose a diet that best fits your needs.
- T3 (For Diet II) Adapt the diet chosen in T1&2 to the requirements for Diet II, i.e. maintain a similar nutritional balance as Diet I, but include additional constraints related to vegetarian and dairy-free products.

Participants were asked to articulate their thoughts as they progressed, using a think-aloud protocol.

#### **Workshop Observations:**

*Observations from Diet I workshop:* The participant in that session worked with acceptable ranges and thus compared the groups of points generated from the ranges. As the participant commented “*I looked at the data [...] on the ranges. I'm thinking that those who go up very high [shows points in a high range of values in a dimension], could be interesting, negative or positive. But of those who were medium, there are meh [so-and-so], there was a whole group that was more or less average*”. Then they refined their selections little by little: They removed groups of points strategically by comparing subgroups of the same group and removing the ones deemed *less desirable*. Then, they skimmed through the individual points and selected the ones to remove. To do that, they compared points within the same group with one another. These comparisons were conducted by looking at the differences between the nutritional values on each dimension.

*Observations from Diet II workshop:* The participant worked mainly with percentile ranges to identify foods that correspond to their criteria. For example, the participant selected the top 5% elements rich in Omega-3. This meant that they were constantly comparing the points in their selection to the points outside the selection. Once they had a selection that was compliant with the requirements, they skimmed through the individual points to select the ones they liked the most. “*What I lost the most [with changes to their selection] isn't fat or protein, what I lost the most is "enjoyment" [of diet]. So I tried to make up for it, [...] So I went for things that are quite fatty*”. In other words, experts were comparing their original selection to a revised selection that included more points. To do that, they examined points within a single group: “*I*

looked through that list, checking which ones interested me based on the diet, what I like to eat, etc. I picked very few, because I saw it was mostly cheese and meat, which I couldn't eat".

We noticed that comparisons of groups were conducted by looking at the *difference between the values* on each dimension. Even when they reached a good trade-off balance, they reported looking for diversity and complementary food items to constitute a diet that is pleasant to eat: "A real problem is that my diet lost its variety when I cut out chicken, and other stuff..., and it's true that it's missing something. Before, there was chicken and beans on the plate, and now it's just beans, and it's true that it's less diverse. So, I had to find some "replacements," so to speak. That's what I went looking for".

#### 4.1.4 Summary of Trade-off Comparisons by Experts

To summarize, our workshop and video analysis reveal that while the specifics of "what" is being compared and "how" is problem-dependent, there are recurring elements in the trade-off analysis problems we studied. Experts often compare groups of points, examining aspects such as the number of points, the ranges of the groups, and the best and worst points in relation to the trade-off criteria. While means and variances of groups are considered in the comparison, experts also rank points within a group and look at individual points, such as outliers or best/worst values in the group.

We distill a summary of these needs [RQ5a] into the following set of expert requirements for trade-off comparisons between groups:

- Groups (eR0): compare groups of points;
- Metrics (eR1): see best and worst points, means, variance, differences and complementarity across points;
- Detail (eR2): see number of points, access to individual points of groups;
- Ranges (eR3): see group value ranges;
- Rank (eR4): be able to rank items inside groups, and to potentially add or remove them.

This preliminary work was the basis for the tasks used in the design workshops we describe next, and the training we provide to participants for the controlled experiment (chapter 5).

## 4.2 Design Workshops [RQ5b]

After confirming the importance of comparing groups of points (eR0), and deriving expert requirements for such comparisons, we set out to design appropriate visualizations to support them. We conducted two design workshops where participants generated hand-drawn visualizations for comparing groups of data points, which we then analyzed for inspiration. We describe next our workshop considerations, procedures, and the analysis of the created designs.

### 4.2.1 Participants

We conducted two 2-hour design workshops, with three and five participants each. Of the eight participants, four self-identified as women and four as men. To ensure a breadth and depth of designs, participants were practitioners or researchers in HCI and visualization (4.8 mean years of experience, min:2-max:14). The author of this manuscript and his supervisors also participated in the generation of designs.

### 4.2.2 Scenario and Data

To provide participants with a realistic design task that did not require domain expertise, we used a hypothetical scenario. Participants were asked to create designs to compare multiple groups of second-hand vehicles from different manufacturers in order to choose a group of cars for a new car-on-demand taxi service, as this is a dataset that can act as both a comparison of discrete solutions (choosing one bundle from one manufacturer) and as a comparison of composite solutions (creating a fleet by mixing cars from multiple manufacturers).

The dataset we used in the workshop is composed of 60 used cars and was generated as follows. We first merged two real-life car datasets [3, 4]. Next, we generated the Pareto front [123] optimizing four objective dimensions: Miles per Gallon (MPG), horsepower, price, and odometer readings. From the Pareto front, we randomly drew 60 used cars.

The cars were then grouped by manufacturer, resulting in ten groups: five from the United States, one from Japan, and four from Europe. The number of cars per manufacturer varied, ranging from a single car to fourteen cars. The choice of the groups presented a real trade-off, i.e., no one group was strictly better than the other. For example, if one group outperformed the other in terms of MPG or price, it necessarily lagged behind in terms of horsepower or usage.



### 4.2.3 Design Task and Procedure

**Procedure:** We began the workshops with an introduction to trade-off analysis to ensure that participants understood the concepts, especially the need to balance competing criteria. We then introduced the design tasks and described the criteria they need to balance: minimize cost and odometer while maximizing mpg (fuel efficiency) and horsepower. Informed by past work on multi-dimensional visualization [34], we provided inspirational materials in the form of different visualizations for showing the ten groups. To avoid biasing participants towards any particular design, we included three different data representations: radar graphs, parallel coordinates, and tabular visualizations.

We decided against using scatterplots as inspirational material. While effective for visualizing relationships between two variables, scatterplots may not be ideal for representing statically multi-dimensional data involving more than two or three variables, especially when comparing multiple groups<sup>2</sup>.

Following the sketching sessions, we held a roundtable where participants explained their designs and their views on which designs best addressed the needs of the hypothetical scenario. We also discussed the high-level ideas that emerged across all designs. Each design workshop lasted two hours including two ten-minute breaks and the discussions.

**Design Tasks:** We asked participants to perform three sketching tasks. In the first, they had to sketch a visualization to compare the ten groups; in the second, to consider additional constraints (brands from a specific country), and in the last to consider how they could combine cars from two groups to create an optimal fleet.

The first design task involved a general group comparison: comparing groups of points representing manufacturers, similar to how wine experts compare groups of recipes corresponding to different fermentation strategies (comparison of discrete solutions).

The motivation behind the second task was to assess how participants would design a visualization to support comparisons, considering that decision criteria might change during the trade-off analysis process. This reflects a behavior observed in past work [34, 97], where users occasionally adjust constraints and shift priorities in decision tasks based on the characteristics of the data. For example, if two groups are satisfactory on a given criterion, the expert may choose to exclude that dimension from subsequent analysis.

The third design task focused on situations where groups can be refined by adding or removing points (comparison of composite solutions), as opposed to adjusting di-

---

<sup>2</sup>We note that scatterplots have worked well in previous work in decision-making [34] and trade-off analysis [97] because they consisted of interactive tools. The materials we provided however in this workshop were static images.

mensions as in the second task. This behavior of adding or removing points, observed in our workshop (eR4) and previous work with experts, is a primary method they use to explore data with certain visualization tools [97, 113]. This refinement process may involve comparing groups of points against each other, as well as comparing the group's previous state with its new state after points have been added or removed.

## 4.2.4 Analysis and Results

### Analysis Methodology

The three authors of the work involved in this chapter [97] conducted a thematic analysis [144] of the sketches from the first workshop collectively. Each stand-alone idea or system was isolated as an individual sketch, either from the designers indicating the grouping of their drawings or by using a recording of the roundtable that we held at the end of the workshop to understand the designers' intentions.

From these sketches, we identified different aspects or themes related to comparison (discussed next). After this, one of the authors then applied the coding to the sketches from the second workshop. To analyze the sketches, we adopted a mixed coding approach that combined both closed and open coding methods. For the comparison layout, we applied a closed coding schema based on the taxonomy introduced by [101]. Similarly, we used a closed coding schema to categorize the types of graphs inspired by [1] which was also used during the workshops as inspiration material for the designers. In contrast, the level of granularity, displayed data, and visual encodings were analyzed through open coding, which allowed us to identify emerging patterns and themes. This open coding process was refined iteratively over three cycles.

Coding the sketches was carried out using Obsidian with the dataview add-on [2] where we linked the sketches with the thematic analysis. This provided a framework for faceted browsing, where codes could be tracked and reused, as well as for providing quick overviews and summaries of all sketches with specific codes (see sup.material of generated HTML structure).

### Results

The workshops produced 51 sketches (35 were created by participants and the remaining 16 by the authors). [Figure 4.1](#) shows examples of the generated sketches (see sup.material for a full list [96])

We identified six main themes across the generated sketches for group comparison visualizations. Our coding scheme was thus organized into six themes and 64 unique codes (a total of 498 instances). We present next these themes and list the most common representative codes for each (the full list exists in sup.material [96]).

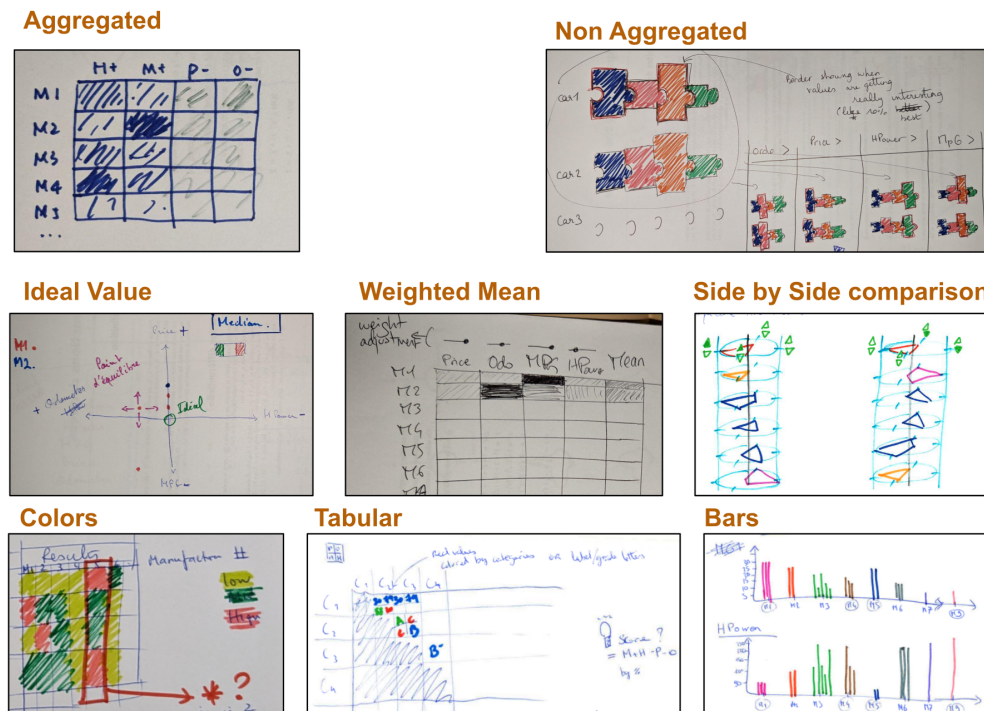


Figure 4.1: Sketches examples from our design workshops exploring different visualization techniques for comparing car bundles in trade-off analysis, such as Aggregated and Non-Aggregated views, Ideal Value plots, Weighted Mean tables, Side-by-Side comparisons, use of Colors, Tabular formats, and Bar charts. These sketches come from different participants (but not from the authors).

**Level of granularity:** (3 codes, 88 instances) This theme expresses at what level the groups are presented: if the data displayed is at the level of individual points/vehicles (26), at the level of group/manufacturer (31), or if the group is aggregated/abstracted in a different way (31), for example, using the weighted mean of the group or a score. Aggregated was the most common level of granularity view.

**Type of graph:** (9 codes, 49 instances) This theme encodes the visualization type used for the comparison. The most common types were tabular visualizations (16), bar-charts horizontal or vertical (6), radio-charts (10) and scatterplots (5). Surprisingly, parallel coordinates, even though presented as inspiration material, appeared in only 4 sketches.

**Comparison layout:** (3 codes, 54 instances) We were inspired by the layouts described by Gleicher et al. for point (not group) comparisons [101, 102] when categorizing how groups were positioned with respect to each other. Most commonly, the two groups were placed side by side (32), followed by superimposed designs (18), and very few explicit encoding of relationships (4).

**Visual Encoding:** (12 codes, 103 instances) This theme describes how the designer encoded the displayed data visually. The most common encoding was color (37) with red used often to indicate "bad values" - values that do not align with the optimization criteria, and green to signify "good values" - values that do align with the criteria. Color was also used as a gradient (often green and red) to represent how good or bad a value is. Another common encoding was direction arrow (17), followed by shapes (10) and lines (11).

**Displayed Data:** (26 codes, 144 instances) This theme is used to describe the data, information, and metrics displayed in each design. Most designs included some visual reference of the direction of optimization (23) and raw data values for individual data points (19). Other common data presented were weighted means (9) and means (9), as well as other forms of scoring (8). An interesting metric displayed in some sketches was the notion of an "ideal solution," i.e., a fictional data point with ideal values to aid in the comparison (10).

**Interactions:** (11 codes, 60 instances) This theme describes any possible interactions indicated in the designs. Several interactions were suggested for changing the display data (17): sorting, numerical transformations (e.g., to create averages), etc. The next most common action was to rank objectives to signify priorities in an interactive manner (12). Several designs also allowed the ordering of the objectives (3) also as a means to indicate priorities. The remaining interactions were more general selection and navigation actions.

#### 4.2.5 Summary of Design & Visualization Needs for Comparison [RQ5]

We summarize next, high-level design and visualization needs for supporting group comparison in trade-off scenarios ([RQ5]):

- R1-Metrics: Simplify the comparison task by providing aggregated views and metrics (means and scores), or grouping of points by filtering. This aligns with expert requirement eR1.

- R2-Detail: Maintain flexibility in visualization granularity: although aggregations are helpful, it is important to access the details of the individual points. This aligns with eR2.
- R3-Context: Communicate the position of the groups within the bigger context (all available data). For example their range of values within the population, their place in the general population distribution, their own distribution, the number of solutions they include, etc. This relates also to how experts consider value ranges of groups and their overlap (eR3).
- R4-Rank: Allow ranking, for both individual points within a group (sorting), but also for the criteria considered in the trade-off (and their priorities). Ensure the direction of priority (to maximize or minimize) is clear. This relates to how experts want to rank items in groups and identify their best / worst items (eR4).
- R5-Reference: Introduce a reference point for comparison of trade-offs. The comparison of groups is aided by considering them with respect to a theoretical ideal point, an equilibrium point, or a reference to the minimum and maximum values in the data. This is a new outcome that we did not observe with experts, likely because it is a visual way to indicate the users' goal (ideal) and the distance from it.

We additionally identified aspects that are more directly related to visual encoding choices made by our participants.

- R6: Tabular visualizations are appropriate for group comparisons.
- R7: Side-by-side presentation of groups is most common.
- R8: Color is the most prevalent indicator of trade-off.

### 4.3 Visualizations for trade-off comparisons

Following the outcomes of the design workshop and expert needs, we implemented a visualization system for the comparison of groups in trade-offs (Figure 4.2). The prototype was built using HTML, JavaScript, CSS for representing the tables and all interactive components. We only use the D3.js library for the histograms. This ensures that the prototype can scale to very large tables. We describe next our design choices.

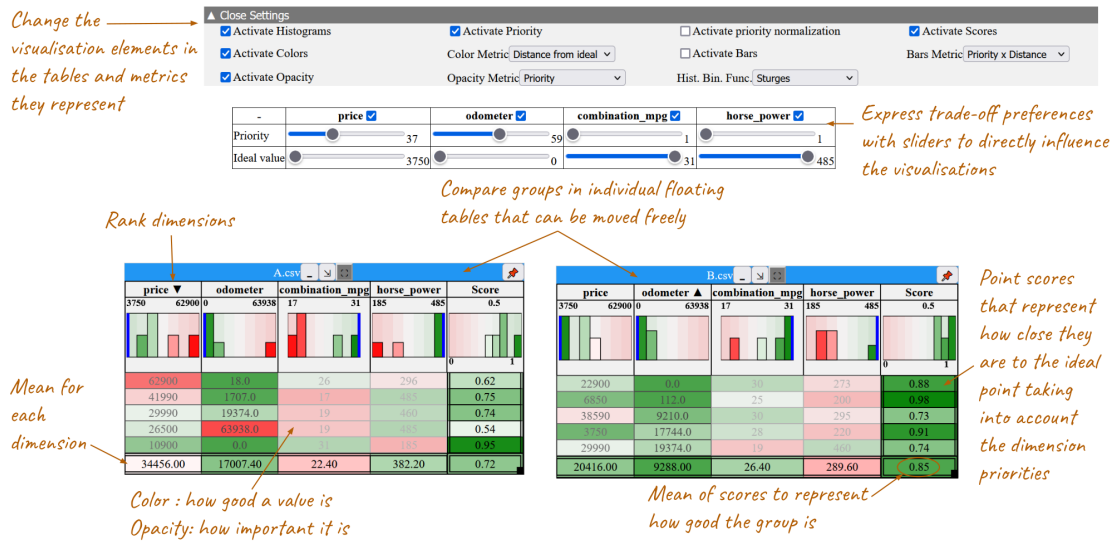


Figure 4.2: Prototype for trade-off comparisons between groups of points, each represented as a table. (Bottom) In this example we see two groups of used-cars datasets from different manufacturers, placed side-by-side. Rows represent used cars and columns their properties (Price, odometer, MPG, horse power). To facilitate comparisons across multiple criteria, the tables feature histograms for each dimension, sorting options, a column of weighted scores, and a row of means per dimension. In addition, table cells are colored to encode trade-off metrics: transparency indicates the priority of a dimension and color hue how good the value is (distance from an ideal value). (Middle) A set of sliders is used to adjust the priority and ideal value of a dimension. (Top) Finally, a control panel can be accessed to change how the trade-off metrics are represented (coupled or decoupled) and how they are visually encoded (using bars, colors or just text).

### 4.3.1 Visualizations and Comparison Layout

We opted for a *tabular* visualization as the basis of our system, as it was both the most popular in our design workshop (**R6-Tab**) and has been shown to be an effective visualization for decision-making [34]. Thus, columns represented dimensions and rows data points. We note that our experts in the previous chapter often referred to their data tables when they made point selections (i.e., identified interest zones to compare) and some even downloaded the CSVs for their selections to compare them in a table viewer, an observation that further supports this choice. Following the majority of designs in the workshop, we adopted a side-by-side comparison (**R7-Layout**), with one table per group that analysts can move around and align horizontally or vertically.

Based on inspiration from the design workshops, we implemented several different visual encodings for *cells* of the tabular visualization:

- **Textual:** This encoding is a basic data representation showing in each cell the value of a data point for that dimension (e.g., the car's price, MPG, etc). It is familiar to domain experts who are accustomed to working with tabular data in CSV files without specialized visualization tools.
- **Colors:** Cells can be additionally colored. This encoding was popular in the design workshops (**R8-Color**), where participants adopted common codifications like coloring "good" values with green and "bad" values with red. We explain in the next section on "Trade-off Metrics" ways of defining "good" and "bad" values. This encoding can be seen in [Figure 4.2](#).
- **Bars:** Cells can be additionally filled with a bar - the less filled the cell the "worst" the data value and the more filled the cell the "better" the value. While bars are not as prominent as colors in the generated designs, this visualization type is notably used in past decision-making tools, such as the LineUp ranking tool [132]. This encoding can be seen in [Figure 4.3](#).

### 4.3.2 Expressing Trade-off Metrics

Participants of our design workshops expressed information about the trade-offs present in their comparison in several ways for the groups. We integrated two that were often identified in the workshop: the notion of a weight or priority to help calculate a weighted mean or score (**R1-Metrics**); and the distance from an ideal point (**R5-Reference**). We describe next the intuition behind these metrics, as well as how they can be applied at the level of data values, data points, and groups (**R2-Detail**).

- **Priorities:** Analysts can express the importance of the different dimensions by assigning *weights* to them. This declaration of weights is common practice in decision-making tools [132, 135]. Priority weights can help participants focus on what matters most to them by emphasizing the dimensions that align with their goals or preferences. This can also facilitate comparisons between different options or scenarios, allowing users to see how changes in priorities impact the overall assessment in various 'what-if' scenarios. In our case participants can express weights between 1-100 (to evoke percentages of importance). Weighted means have been used in visualization tools such as Lineup [132] and weightlifter [135].
- **Ideal Values:** Analysts can also express an ideal value that they are searching for each of the dimensions. In our tool we restrict the ideal value between the minimum and maximum values that exist in a dataset. This restriction ensures that the ideal value remains realistic and grounded within the range of observed data. Ideal values can also express the direction of optimization, for example an

ideal value at the minimum of a dimension such as Price indicates that analysts want to minimize price in their selections. Ideal values help us define a numeric distance that indicates how "good" a value is. For example, the smaller the distance of the actual Price of a car from the ideal value, the better that car is for the Price dimension. Ideal values distances are used to calculate several metrics: distance per point from a particular dimension, average distance per point across all dimensions, and average distance per group from the ideal value per dimension. These distances can be assigned to visual encodings of a cell (example a color scale or a bar length) to indicate how good or bad a point is with regards to the ideal for that dimension. Ideal value driven analysis is a practice that already exists in MCDM literature for ranking alternatives [51].

#### 4.3.2.1 Displayed Data & Granularity

We noted that most of the generated designs provided aggregated views (**R1-Metrics**) but for the experts, individual points and values are important. We thus used the two trade-off metrics to provide scores for values, for individual data points, and for groups (**R2-Detail**).

**Metric per point: Score** The score represents how close the values in a row (data-point) are to their corresponding ideal values, considering the importance (priority) of each column. A high weighted mean indicates that the corresponding data-point aligns well with the analyst's prioritized dimensions and ideal value. It is defined as follows:

$$\text{Score} = \frac{\sum_i P_i \times \left(1 - \frac{|V_i - I_i|}{|\text{globalMax}_i - \text{globalMin}_i|}\right)}{\sum_i P_i}$$

Where:

- $V_i$  is the value in the  $i$ -th column for the current row.
- $I_i$  is the ideal value for the  $i$ -th column.
- $P_i$  is the priority of the  $i$ -th column.
- $\text{globalMax}_i$  and  $\text{globalMin}_i$  are the global maximum and minimum values for the  $i$ -th column across the groups compared. This is used to normalize the distance to the ideal.

**Metric per group: Mean of Scores** To represent each group in an aggregated way, we calculate the mean of the Scores of all points in the group, effectively treating the Score column as any other data column for this calculation.



**Metric per cell: Coupled vs Decoupled Metrics** In examining the sketches from our workshops, we observed that each metric is typically represented by a distinct visual encoding (29 recorded instances from participant sketches), a method we refer to as **decoupled representation**. For example, in some sketches, weights can be shown as color transparency, and distance to ideal values is shown as color hue (red for far and green close to the ideal). While this separation clearly shows each of the metrics, it requires more visual information than the coupled metric discussed next.

Conversely, some weighted mean or score-based tools such as LineUp [132] encode both the priority and distance from an ideal variable in the same visual representation (the length of the bars in the case of LineUp). In other words, what they are representing visually is the explicit encoding of these metrics into a unique score or weighted mean. **Coupling** multiple values into the same visual encoding simplifies the visualization, reducing clutter but potentially obscuring the relationship between weights and values.

**decoupled metrics per cell:** The decoupled metrics we use are the priorities that the participants input for each dimension (1-100) and the distance from the ideal value for each dimension normalized across the range of values in all groups compared.

$$\text{distance}_i = \frac{\text{globalMax}_d - \text{globalMin}_d - |V_{d_i} - I_d|}{\text{globalMax}_d - \text{globalMin}_d}$$

Where:  $\text{globalMax}_d$  and  $\text{globalMin}_d$  are the global maximum and minimum values for the  $d$ -th column across the groups compared.  $V_{d_i}$  is the value in the  $d$ -th column for the  $i$ -th point of the group;  $I_d$  is the ideal value for the  $d$ -th column.

**coupled metrics per cell:** The coupled metric we used is a weighted distance from the ideal value:

$$\text{coupled}_i = \frac{\text{priority}_d \times \text{distance}_i}{100}$$

where the distance and the priority are the same as in the decoupled case. 100 is the maximum possible priority value for each dimension by design.

In our system, we allow analysts to choose what metrics to map on the Colors or Bar lengths of our cells. As this design choice is not obvious, in our study (next section) we consider the impact of coupling/decoupling the priority and ideal value distance metrics.

### 4.3.3 Bringing it all together: Tool & Interactions

To communicate the position of the groups within the bigger context of all available data (R3-Context), we included at the very top of each table (group) histograms that

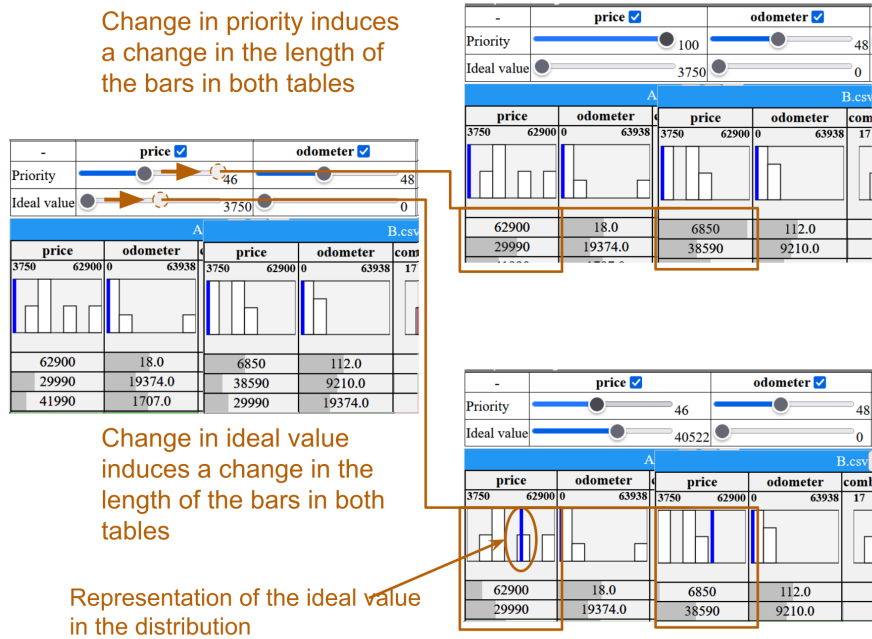


Figure 4.3: Interaction using the comparison tool. The visualization in this example is Coupled Bars. The user can change the priority or ideal values using the provided sliders (left), which then updates the length of the bars in both group tables (right), as well as the position of the ideal value on the histogram

visualize the distribution of all data for each dimension, where the points of the group fall into that histogram, as well as the ideal values themselves (discussed in "Trade-off Metrics," subsection 4.3.2). The last column and row provide aggregated information (**R2-Detail**). The last column is the "Score" column, showing the weighted score for the data point calculated based on priorities and ideal values. The last line of the table (group) shows the mean of each dimension (column) including the score column (i.e. the mean of all data-point scores), which gives an overall assessment of the group's alignment with the user preferences across all dimensions.

Analysts can sort data by clicking on the column headers (descending, ascending, and back to initial order), as this was deemed important (**R4-Rank**). Users can rearrange the column order (as it is possible in a spreadsheet) and this is reflected across all tables/groups. Finally, tables can be moved to align horizontally (aligning points across groups) and vertically (aligning dimensions). Beyond these basic interactions, the tool includes an interaction drawer (opened on demand). Here, users can load datasets to compare, but also customize the visualization for trade-off analysis.

- **Priority Sliders:** These sliders adjust the significance of each column in the

overall analysis (a value between 1 and 100). For example, if one cares about the price dimension above anything else, one could assign a high value to the priority of the price dimension. Modifying a slider's position alters the weight of that dimension in the score or weighted mean calculations.

- **Ideal Value Sliders:** These sliders allow setting a target value for each dimension, which the tool uses to calculate the deviation of each data point from this target. It helps in assessing which entries are closest to or farthest from the desired value.
- **Metric Mapping:** Both the bars and colors can represent any of the following metrics: the priority assigned to each dimension, the distance from the ideal for each value, or a combination of these two factors (deviation multiplied by priority).

While our tool does not support the movement of points across datasets (to construct new composite datasets) this is an easy extension. We focused on the main components needed for the visual comparison of two groups, irrespective of whether they are composite or discrete.

## 4.4 Conclusion

Our observations from working with experts from agronomy and other domains have shown that they work primarily with solution spaces (i.e. groups of items as discrete or composite solutions). To better identify user needs for comparisons of groups, we conducted workshops with three groups of participants. These provided us with valuable insights into how trade-off comparisons are conducted, confirming the importance of comparing groups of points. We also derived other expert requirements for such comparisons, such as the need for aggregate information, such as in the form of metrics (e.g., mean, variance, etc.), access to individual point details and range information, and finally the ability to rank items inside groups, with the possibility to add or remove options.

Building on these insights, we organized two design workshops with visualization and Human-Computer Interaction (HCI) experts to iterate on and refine visualization designs that could effectively support the comparison of trade-off groups. Findings from these workshops include concrete visual encodings that inspired the design and implementation of a prototype system. In summary, we noticed how the majority of designs presented groups in the form of data tables, with side-by-side views of the different groups to compare.

The workshops additionally inspired us to create visualizations to express more tailored, trade-off specific metrics related to the analyst's objectives, such as weighted

means and distance to ideal values. The prototype we built to support trade-off comparisons has several possible alternatives to express these trade-off metrics, both visually (color, bars) and in terms of mapping (coupling both metrics on one visual encoding or decoupling them). As these design choices have themselves well trade-offs(!), we decided to compare them. This is the topic of the following chapter.

# Evaluating Visualizations for trade-off comparisons between groups

In this chapter, we explore how to support the visual comparison of groups of data-points (or items) within the context of trade-off analysis. Our primary research question focuses on *understanding the impact of different types of visualization* on this group comparison task, in terms of balanced consideration of multiple criteria, user trust, and mental load.

In the previous chapter, we followed a multi-phase approach to understand expert needs and design choices. First, we conducted three workshops with domain experts to understand the specific needs and challenges involved with comparing groups of data-points in trade-off analysis. This provided valuable insights into the types of comparisons users frequently make. Building on these insights, we organized two design workshops with visualization and Human-Computer Interaction (HCI) experts to iterate on and refine visualization designs that could effectively support the comparison of trade-off groups. We then developed a prototype visualization system based on the outcomes of the design workshops.

In what follows, we evaluate the effectiveness of two promising visualization techniques against a baseline in a user study involving 18 participants.

This work was submitted to ACM CHI 2025. Supplementary material and video are available at [96] and statistical results in Appendix A.

## 5.1 User Study Design

Our design workshops (section 4.2) and resulting comparison system (section 4.3) led us to several design choices that we believe warrant further investigation. First, it remains unclear if the additional trade-off metrics, such as weighted means and distance to ideal values, do indeed aid in comparing groups during trade-off analysis. And if they do so,

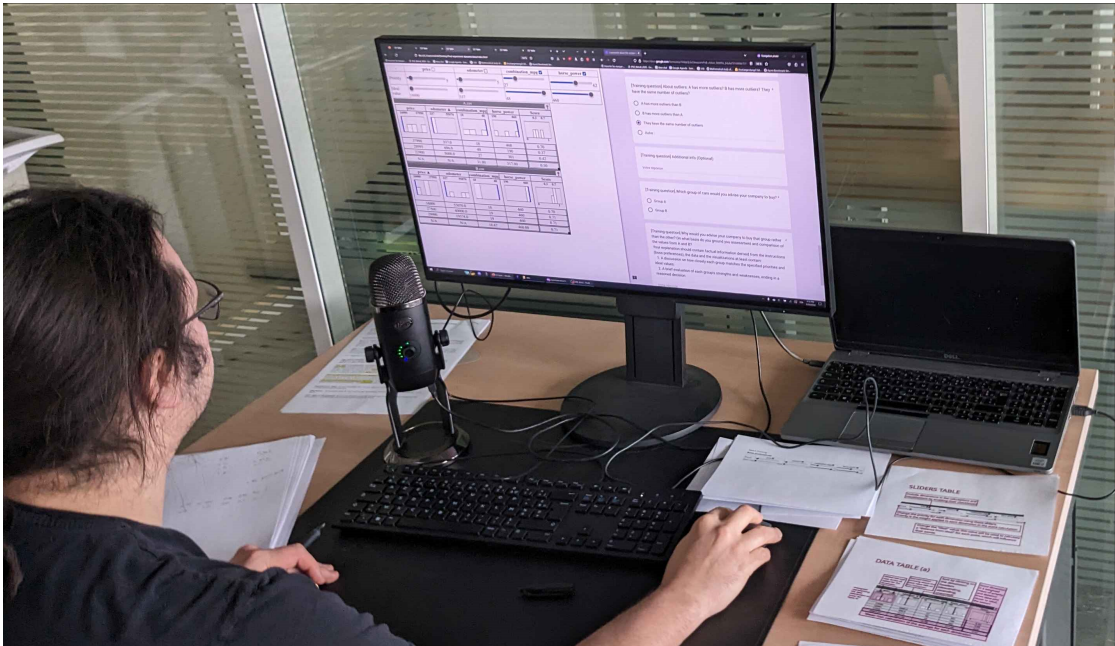


Figure 5.1: Our study setup with a split screen between the visualization and the tasks and questions. Papers, with memory aids, as well as a blank paper for note-taking, are seen on the side.

how do they contribute, by enhancing understanding or increasing trust in decisions? Moreover, the choice to couple or decouple these metrics visually introduces both costs and benefits, e.g., potential visual clutter versus improved access to detailed metric information. We thus set out to understand the impact of these design choices, and more generally the impact of using visualization support in the comparison of trade-offs across groups. Our investigation centered around the following two research questions:

- RQ6: Does the visualization of trade-off metrics aid comparison more effectively than a simple tabular representation?
- RQ7: Is a coupled or decoupled visualization of trade-off measures better at supporting this comparison?

### 5.1.1 Visualization Conditions

Testing all possible combinations of visual encodings and visualization layouts, data granularities, trade-off metrics, and how they are coupled—each possible and adjustable

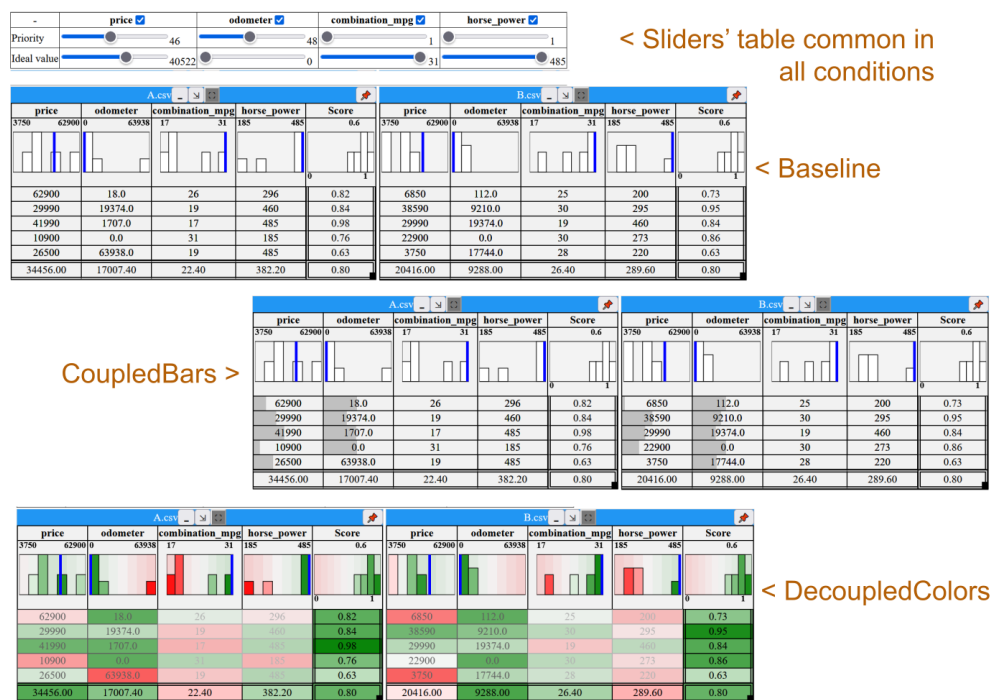


Figure 5.2: The three different tabular visualization setups used in our study: Baseline with simple histograms and no additional visualizations in the tables, CoupledBars with added bars, and DecoupledColors with a color gradient, all linked to a table of sliders for adjusting priorities and ideal values.

within our tool—is beyond the scope of a single user study. Guided by our two research questions (RQ6, RQ7), we selected three visualization conditions for evaluation (Figure 5.2) that we further motivate in the remainder of this section:

1. **Baseline (Textual Data)**: This condition involves the use of textual data as the baseline for comparison, augmented with additional information to facilitate comparison of trade-offs: the weighted scores per data point (last cell in each row), the mean data values per dimension (bottom row), and the data distribution in the form of a histogram.
2. **Colors (Decoupled)**: This condition adds color to the baseline (i.e., each tabular cell is assigned a fill color using a green/white/red color map). Hue represents the distance from the ideal value, while opacity indicates the priority, thus visually de-coupling the two trade-off metrics (e.g., a dark green cell indicates high priority and low distance to ideal, and a light red indicates low priority and high distance).

3. **Bars (Coupled)**: This condition adds bars to the baseline (i.e., each tabular cell includes a horizontal bar). The bar length represents the *weighted distance* from the ideal value for that dimension, serving as a type of trade-off score, similar to the approach used by LineUp [132].

The *baseline* condition (textual data) was chosen to study if the more complex trade-off metric visualizations that add more information on trade-off analysis do indeed aid in the comparison of groups. We note that even though we call this condition a baseline, it does include more trade-off related information than a simple table in the form of histograms and weighted scores.

When it comes to the more complex trade-off metric visualizations, we focused on representing two main metrics, the priority of a dimension (weight) and distance from an ideal point for that dimension (discussed in [subsection 4.3.2](#)). As we discussed in that section, the two metrics can be represented coupled (merged and mapped on a single visual encoding) or decoupled (each metric mapped on a different visual encoding). It is unclear which approach works best: a decoupled representation clearly shows each of the metrics, but is visually more dense, while a coupling of the two metrics into the same visual encoding simplifies the visualization, reducing clutter but potentially obscuring the relationship between weights and distance to ideal values. We chose to test different methods in representing the trade-off metrics (coupled vs. decoupled): a bar-based visualization (similar to LineUp) for the coupled metric representation, and a color-based method inspired by the design workshops for the decoupled metric representation.

For the *decoupled* condition (Colors), we decided against using a bar visualization that assigns each variable to a different visual property of the bars, such as height and width or the stroke weight when drawing the bars because we found that the visual change of the bars was too subtle and easily overlooked. As color-based visualizations were very prominent in our sketches, we used them to decouple the two metrics using two different visual properties: color hue to denote the distance from ideal and opacity to denote the priority.

For the *coupled* condition (Bars), we chose to map the single combined metric to bar length, as it is perceptually easy to read and aligns with established practices in similar tools, such as LineUp [132].

**Other modifications to the tool for the user study:** For the purposes of the study, we also made small modifications to the tool presented in [Chapter 4](#) to ensure a consistent environment across participants and conditions. We deactivated the choice of switching between visual encodings (bars / colors / just text) and metrics mapped to these visual encodings, as these were compared conditions. We additionally deactivated the option to choose a different aggregation metric; thus, all participants had consistent visual encodings and metrics for each condition. Furthermore, the aggregation metric



was chosen to be the mean for the dimensions and the weighted mean (score) for the data-points in all the conditions, as this was the most common aggregation metric in our sketches.

### 5.1.2 Hypotheses

To address our two main research questions pertaining to how different types of visualization impact group comparison tasks in trade-off analysis [RQ6,RQ7], we formulated the following two hypotheses :

- **Hypothesis 1:** Visualizing trade-off metrics, such as priority and distance from the ideal, improves group comparison by reducing mental load and enhancing user understanding of the trade-off, resulting in more user trust in their decisions.
- **Hypothesis 2:** A decoupled visualization of trade-off metrics enhances user understanding of the trade-offs and increases user trust in decisions compared to a coupled visualization, but it leads to a higher mental load.

### 5.1.3 Tasks

We designed three trade-off comparison tasks (T1–3) for each of the three visualization techniques (Baseline, Color, Bars). Each task involved evaluating two different groups of used cars and deciding which group would be most suitable for forming the initial fleet of a hypothetical taxi company. Participants were required to balance four conflicting criteria—minimizing price and odometer readings while maximizing MPG and horsepower—based on a set of priorities and ideal values provided by the company’s boss. At the end of their analysis, participants indicated which group of cars they would recommend for purchase, providing a written rationale for their choice, discussing how each group aligned with the boss’s specified priorities and ideal values, and evaluating each group’s strengths and weaknesses.

We designed a different dataset for each task, each having four dimensions (odometer, price, MPG, horsepower) and seven points (section 5.1.4), but a different set of priorities.

This comparison task emulates how our experts made comparisons of groups in order to make a decision (e.g., choose a set of wine recipes). The fact that we gave the specific set of priorities and initial weights for each dimension ensures that participants do not need domain knowledge in order to decide which priorities are more important or what values dimensions should have. But it emulates the a priori knowledge experts have about what they want to optimize and how important they feel it is, even if their criteria may shift as they explore the available data. Asking participants to provide data-driven justifications to explain their choices with evidence

from the datasets ensures we can capture the information they consider important in their analysis. We note that the task datasets (section 5.1.4) are chosen in a way that decisions are not trivial and there is no right answer, as is often the case in real-world decision-making that requires human analysis. These choices allow us to explore how participants engage in a challenging trade-off comparison of groups (as there is no ground truth), what evidence they consider important for their decisions, and whether visualizations used may affect the nature of this evidence and their understanding of the trade-offs present in the datasets.

The setup of the visualization tool used for this task is described in section 4.3. The tool presented the data for each group in separate tables, facilitating side-by-side comparison. Although participants could adjust the provided sliders to explore different scenarios, their final decision needed to align with the initial requirements (i.e., priorities and ideal values) set by the boss.

### 5.1.4 Datasets

We started with the 4D-PF dataset created for the design workshops (section 4.2). To ensure similar levels of difficulty across tasks, the datasets we generated shared the same number of dimensions and data-points, the same weighted mean, and had a similar “amount of trade-off”. To achieve this, we calculated the trade-off index introduced by Unal et al. [109], which quantifies the degree of trade-off within any dataset. This index measures how much a solution’s ranking shifts between each combination of two criteria, indicating the ‘tension’ between competing objectives.

More precisely, the calculation of the Trade-off Index [109], – denoted as  $\lambda_{ikl}$ , for the  $i^{th}$  solution between objectives  $k$  and  $l$  –, employs a binary factor  $\beta_{ijk}$  that indicates whether two solutions,  $i$  and  $j$ , intersect across these objectives. By aggregating these intersections across all solutions, the Trade-off Index reflects the extent of trade-off between two objectives. In other words, every time a solution changes ranking between two dimensions, this index goes up by the difference of rankings. As a result, a high trade-off index implies changes in ranking, highlighting strong trade-offs, while a low index indicates consistent rankings and minimal trade-offs.

The groups’ participants had to compare in our experiment had the same trade-off index, the same weighted mean for a specific set of dimension weights, and the same number of points. Thus, the choice between them could not be done automatically and requires human judgment and decision-making.

In total, we generated nine datasets, three for participant training, and six for the different tasks (2 datasets per task). The training datasets had three points (cars) and the task datasets had seven data points. All our generated datasets have an average trade-off index of around 0.6.

We chose to limit our groups to seven data-points as in pilot tests we found that this size allowed for groups that were not trivial to scan visually, but also not overwhelming,

especially in the case of the Baseline condition. The number falls in the range of sizes of groups our experts compared (section 4.1).

### 5.1.5 Study Measures

Measuring the quality of a decision or the effectiveness of the decision-making process is challenging, especially in trade-off analysis where there is no ground truth or an objective 'best' solution. We therefore designed four types of study measures that directly relate to our hypotheses:

- **Explanation Quality (H1 & H2):** We adapted the Co-12 Explanation Quality Properties from the AI domain [105] to assess the quality of participants' explanations (explained in more detail in section 5.1.5.1). We consider the quality of explanations as a proxy for measuring the quality of comparisons and reasoning during the trade-off analysis process, based on the assumption that better explanations indicate the extraction of more or more varied correct information from the comparisons, leading to more informed decisions.
- **Trust (H1 & H2):** We evaluated participants' *trust* in their conclusions by including self-report questions that assessed their confidence in the decisions they made. Since trust is a key element in both Hypotheses 1 and 2, gauging participants' confidence in their choices provides insights into how effectively the visualization tools supported their decision-making process.
- **Cognitive Load and Satisfaction (H1 & H2):** We incorporated the NASA-TLX Task Load Index [131] questionnaire after each task to capture participants' cognitive load and satisfaction with the decision-making process. This measure relates both to H1, which posits that additional visual encodings help compared to the baseline; and to H2, which suggests that while decoupling visual representations may increase cognitive demand, it should lead to better understanding and trust.

#### 5.1.5.1 Quality of Explanation as a proxy for understanding

We drew inspiration from the Co-12 explanation quality properties [105], originally introduced to evaluate the quality of AI-generated explanations. To apply these metrics to our context, we adapted five of these properties for evaluating the explanations provided by our participants during trade-off comparisons. The remaining seven properties we considered either more AI-specific, difficult to extract from our participants' textual explanations, or beyond the scope of our study (e.g., controllability, consistency, continuity, composition), and are thus excluded. The following table (Table 5.1) presents the adapted metrics we will use to determine the quality of an explanation.

Property	Description for AI context	Application in Trade-off Analysis
Correctness	How faithful the explanation is w.r.t. the black box	Evaluate whether the justification accurately reflects the reasoning process and the decision-making criteria
Completeness	How much of the black box behavior is described in the explanation	Check if the justification is comprehensive and includes all relevant aspects of the decision, incl. the contextual data, the trade-off criteria & priorities
Contrastivity	How discriminative the explanation is w.r.t. other events or targets	Determine if the justification effectively contrasts the chosen option with other alternatives, clearly explaining why one was preferred over others
Compactness	Size of the explanation	We use the inverse of compactness to check if the justifications is detailed and information-rich
Confidence	Presence and accuracy of probability information in the explanation	Look for indications of confidence in the decision, such as probabilities, risk assessments, or certainty levels expressed in the justification. We use self-reported confidence

Table 5.1: Explanation Quality Properties from [105] and their Application in Trade-off Analysis

We explain next how we used these explanation properties to code the quality of explanations from our user study:

- **Correctness:** We checked for factual errors in our participants justifications, whether related to the data or to the trade-off criteria and priorities. All but one (P15) of our participants' decisions were aligned with their explanations.
- **Completeness:** We represent this with the number of unique codes corresponding to data-driven justifications in a response. As we assume that more variety in codes indicates more variety in comparison approaches which in turn leads to a more complete explanation that considers all aspects of the data, the trade-offs, and task goals.
- **Contrastivity:** We count the number of pairs of arguments in favor of each group the participant has expressed. The assumption here is that each pair of arguments in favor of each group shows a consideration for arguments for each group and thus a contrast between the two groups. Note that we also looked at

arguments against groups, but to facilitate counting these were inverted: e.g., an argument against group B, counts as an argument for group A.

- **Compactness:** We count the number of assessments or pieces of information expressed during the explanation. Since these explanations are generated by humans, we count the number of observations and pieces of information and we assume that the more there are in the explanation the ‘better’ it is in terms of details.
- **Confidence:** We asked participants explicitly for their confidence level using a 7-point Likert scale.

### 5.1.6 Procedure

Before starting the tasks, we asked participants to self-assess their familiarity with data analysis and histograms using a 7-point Likert scale. We also inquired about color blindness to provide them with an alternative color palette if needed.

**Task Execution:** For each task, participants completed a 15-minute training session to familiarize themselves with the tool’s functionalities. Following training, they proceeded with the main task of evaluating the two car bundles using the visualization tool (15 minutes). Throughout the experiment, participants had access to a reference sheet that served as a memory aid, containing essential information about the task (goals and priorities of their boss), as well as details about the tool’s functionality and visualization features (see sup.material [96]). They then had to justify their choice in a written statement, providing a detailed assessment of each group’s strengths and weaknesses, explicitly referencing the data and visualizations.

To capture the reasoning behind participants’ decision-making, they were encouraged to verbalize their thought processes using a think-aloud protocol, with their comments audio recorded. The recordings served as a backup for the experimenters in case the written responses were unclear.

After each task, participants completed the NASA Task Load Questionnaire [131]. At the end of the experiment, we asked participants to rank the three visualizations from most to least preferred and explain their choices. While explaining their rankings, participants were given access to the development version of the tool with access to all techniques and metrics to help them recall the conditions of the experiment.

### 5.1.7 Participants

We recruited 18 participants aged between 21 and 36 (mean 26.3). All participants had normal or corrected-to-normal vision, and no participant reported being color blind. Their mean familiarity with data analysis and histograms were 5.52 and 6.35 respectively (both on a 7-point Likert scale). Our study lasted around 2 hours per participant.

### 5.1.8 Apparatus

Participants conducted the experiment on a 27-inch screen with a resolution of 3840 x 2160. The screen was split into two windows; on the right was the comparison tool, and on the left was a form with the instructions, questions, and answer fields. Sessions were audio recorded and screen captured.

## 5.2 Results

We report next on (1) the qualitative analysis of participants' open-ended responses as a means to shed light on the data they considered for their comparisons, the quality of their reasoning as reflected in their explanations, and their self-reported trust in the decisions they made using the different visualizations; (2) the perceived load associated with using each of the visualizations; and (3) the tool aspects they found useful for their comparisons.

**Analysis Methodology:** We first conducted a qualitative analysis of the open-ended responses, for example, to understand the types of information used, quality of explanation properties, the participant strategies, etc. To systematically analyze the open-ended qualitative data, we used a reflexive Thematic Analysis (TA) approach [144] in three steps. In a first pass, one author assigned labels (codes) to text segments (snippets), identifying statements, observations, insights, or subjective assessments about the data related to the comparison of groups or points. In a second pass, all coders determined which snippets referred to information used and which to decision strategies. They also grouped codes to the higher-level semantic categories reported in sections 5.2.1.1 and 5.2.1.2. Finally, in a third pass, all authors considered how our codes relate to quality of explanation metrics (reported in section 5.2.1.3). All our snippets and codes (and the iterative coding passes) can be found in sup.material [96].

We then conducted statistical analysis to compare the means of the different conditions for some of these qualitative codes, as well as self-reporting ratings. Statistical analysis for mean comparison was conducted using interval estimation [117]. Sample means of 95% confidence intervals (CIs) are constructed using BCa bootstrapping (10,000 bootstrap iterations). When interpreting results, a CI of a mean difference that does not overlap with 0 provides evidence of a difference, corresponding to statistically significant results in traditional p-value tests. Here we only report on evidence of a difference (but not the detailed CI values). Analysis scripts and detailed CIs can be found in sup.material [96] and detailed CIs are additionally seen in Appendix A. For the qualitative analysis, translated quotations are indicated with a † notation.

### 5.2.1 Decision Justifications: Information Considered, Decision Strategies & Quality of Explanations [RQ6,7]

We analyzed the open-ended responses that participants provided to justify their decisions in order to characterize their explanations across visualization techniques. Specifically, we identified (i) what **information** from the visualizations participants refer to (and thus use) to compare groups of points and justify their decision; and (ii) their self-reported **decision strategies** for choosing one of the two data groups being compared. These two aspects will shed light on how group comparisons are made and what information is crucial that we provide as visualization designers. We then use this information to (iii) determine explanation **quality** (see section 5.1.5.1) as a means to understand how effective participants' reasoning strategies were with the different visualizations (H1).

#### 5.2.1.1 Information considered in Comparisons

This first group of codes relates to the types of information participants reported about the groups they compared and thus used to conduct or justify their comparisons. We note here that user preferences, in the form of priorities they have and criteria they try to optimize (max/min) are also part of the information considered, but we report these separately in "Justification Quality" (section 5.1.5.1), as these are explicitly given in our tasks as targets for participants. Rather, in this section we coded types of data-related information participants reported analyzing during comparisons or using in their justifications. These can inform design guidelines for future trade-off comparison visualizations.

We coded in total 248 instances (BASELINE: 65, DECOUPLED COLORS: 82, COUPLED BARS: 101) of data-related information mentioned by participants (see Figure 5.3), grouped in three major types. These include direct references to data points or their values at different levels of aggregation, to elements in the visualization that encode data, or to references to how closely the data aligns with the priorities and preferences provided in the tasks as goals.

**Data References.** The most common references to information analyzed and considered in the comparisons were related to data points or their values at different aggregation levels:

(C1) - *Individual Point analysis*: Specific cars/points, outliers, or numerical values of individual points were commonly considered and/or influenced decisions, 88/248 instances coded (45 for COUPLED BARS - 27 from one participant, 27 for DECOUPLED COLORS, and 16 for BASELINE).

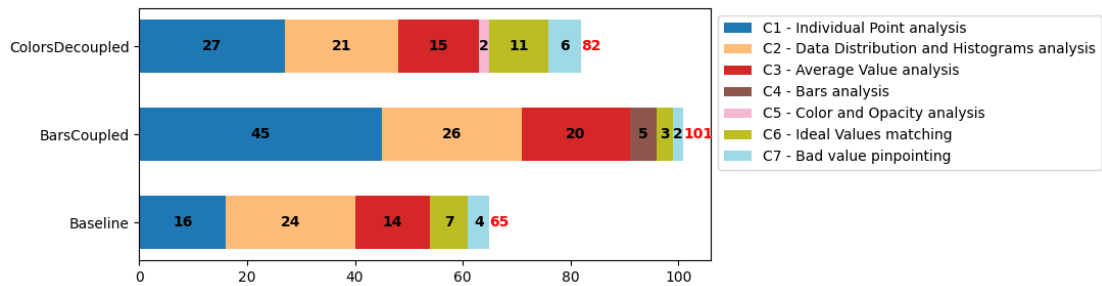


Figure 5.3: Types of data-related information considered (C1-C7).

When considering individual participants, in the `DECOUPLED COLORS` and `BASELINE` more than half of the participants referred at least once to individual points (11/18  for both). But less than half in the `COUPLED BARS` (8/18 ).

(C2) - Data Distribution and Histograms analysis: Observations on the overall distribution of values for each group or explicit mentions of using the histogram charts in their evaluation or of observations on the histograms were the second most common information mentioned, 71/248 instances coded (26 for `COUPLED BARS`, 21 for `DECOUPLED COLORS`, and 24 for `BASELINE`)).

Most participants stated at least once distributions in their decision justifications, more frequently for `DECOUPLED COLORS` (14/18 ), followed by `BASELINE` (13/18 ) and `COUPLED BARS` (11/18 ).

(C3) - Average Value analysis: Explicit mentions of average values of the dimensions being considered and/or influencing decisions were also used, but not as common, 49/248 instances coded (20 for `COUPLED BARS`, 15 for `DECOUPLED COLORS`, and 14 for `BASELINE`).

In all visualizations, most participants stated at least once average values as information influencing their decisions: more frequently for `BASELINE` (14/18 ) , followed by `COUPLED BARS` (12/18 ), and finally `DECOUPLED COLORS` (10/18 ).

**Trade-off Visualization Elements.** Participants also often made reference to visualization components that encode data.

(C4) - Bars analysis: In the `COUPLED BARS` condition, several participants referred to the size of the gray bars in the cell backgrounds (5/101 instances) to justify their choices. We note that while the bar visual encoding is only available in this condition, it was only mentioned by a small number of participants (5/18 participants ).

(C5) - Color and Opacity analysis: In the `DECOUPLED COLORS` condition, participants rarely referred to either the green/white/red color (distance from ideal) or the



opacity/transparency (weight) to justify their choices (only 2/82 instances). Only two participants mentioned explicitly comparison of colors (2/18 ).

**Relation to task Goals.** This category of codes is explicit mentions by participants about why one group was better than another in terms of how well it aligns with the preferences given in the task (i.e., the priorities and ideal values).

(C6) -Ideal Values matching: Reference to elements matching the ideal values for one of the stated priority dimensions was not very common, 21/248 instances (11 for DECOUPLED COLORS, 7 for BASELINE and more rare, 3 in COUPLED BARS). Even though these references were used only once or twice per participant, several participants used this justification in DECOUPLED COLORS and in BASELINE (7/18 participants on both, ), but very few in COUPLED BARS (2/18 participants, .

(C7) - Bad value pinpointing: Explicit reference to an element being a bad value for one of the stated priority dimensions was also used, but rarely as a justification (12/248 instances). In DECOUPLED COLORS there were 6 instances, 4 in BASELINE and only 2 in COUPLED BARS. In DECOUPLED COLORS it was mentioned by 4/18 participants () , and only 2/18 participants used this justification in BASELINE and in COUPLED BARS().

**Other considerations.** We noted a few cases (7 instances, not added in the total) where participants incorporated additional factors or interpretations beyond the stated task priorities or the data alone. For example, one participant changed the task goal as they reasoned that a vehicle with a high odometer might soon break down and thus likely to be replaced, which would ultimately improve the average miles per gallon (mpg) of its group.

**Summary on Information considered:** Participants considered a variety of types of data in their analysis (Figure 5.3). For the BASELINE the data type used in the justification of most participants was average values and data distributions, but there were also instances of individual point analysis (close to half of the participants) and references to values that were close to or far from ideal values. For COUPLED BARS the data types used by most participants were distribution analysis, followed by average values, but less than half used individual point analysis and very few referenced ideal values. Finally, for DECOUPLED COLORS most participants referred to data distributions, followed by individual points and average value analysis, as well as references to values that were close to or far from ideal values. We note that references to visualization-specific elements (bar length and color) were less frequent in the justifications participants provided. Participants' responses provide more information for

DECOUPLEDCOLORS and COUPLED BARS over the BASELINE, which can be explained by the fact that in the latter, access to information was more challenging. There is also surprisingly more access to information for COUPLED BARS than for DECOUPLEDCOLORS, but this is due to one outlier participant - we discuss this more in our quality of explanation measures.

### 5.2.1.2 Decision Strategy

This category includes justifications from participants related to the comparison strategy they used, in other words how they made their decision with respect to goals and priorities of the group comparison task (Figure 5.4). We note here that most participants reported one such strategy. In rare cases where participants mentioned two strategies in their justifications, it generally indicated that they considered two different strategies but ultimately chose one. Here we provide numbers for the final decision strategy, but report cases where participants mentioned more than one.

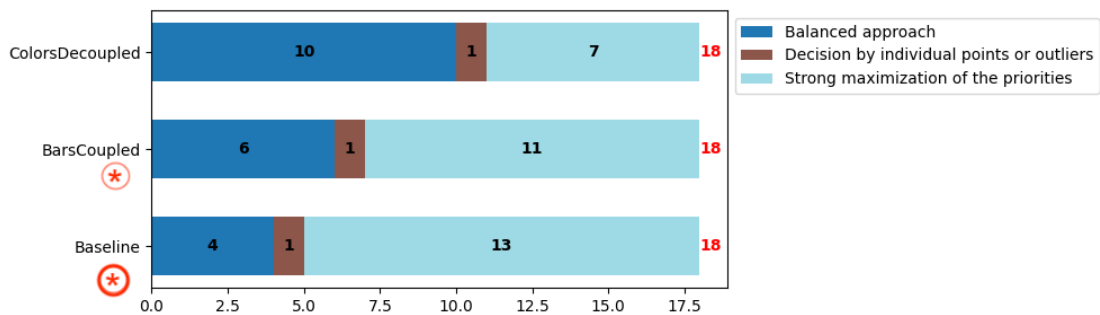


Figure 5.4: Decision Strategy across visualizations. A thick red star (\*) indicates strong statistical evidence that one strategy was more common for BASELINE, a thin red star indicates that this is possible trend also for the case of COUPLED BARS. All analysis was done using CI estimation (see Appendix A for detailed CIs).

(S1) - Strong maximization of the priorities: Choosing a group based on its optimal performance in one or two dimensions, particularly those with the highest priorities. For example, P11 mentioned during the COUPLED BARS condition: “I make my comparison taking into account only the mpg and horse power(main priorities)”. Many participants explained that their strategy was to ignore objectives with lower priorities and thus exclude them from the comparison. P7 further explains why this exclusion strategy works well for BASELINE: “Much more difficult to make a decision here [BASELINE], having to go through all the data in text”. Beyond excluding low-priority objectives, participant P17 adds: “The choice in BASELINE was made more on the whole using averages”, highlighting a tendency to simplify the decision-making process due to the limitations of the visualization.

In BASELINE, the maximization strategy was used by 13/18 participants (■□) and in COUPLED BARS by 11/18 participants (■□). But was less common in DECOUPLED COLORS and used by 7/18 participants (■□). We note that one participant in BASELINE and two in COUPLED BARS considered a balanced approach (see next) but ultimately made a decision that strongly maximized 1-2 top priorities. There is statistical evidence that the maximization strategy is more frequent in BASELINE (and to a lesser degree in COUPLED BARS) than it is in DECOUPLED COLORS.

(S2) - Balanced approach: Deciding on a group that is not the best for the highest priorities but a very good choice for several lower priority dimensions and not a bad choice for the strongest priorities. In other words, this is a strategy that considers trade-offs across many objectives. For example, P10 mentioned while in DECOUPLED COLORS: *“Even if group B has a slightly better mpg than group A (based on the histogram), it doesn't seem significant and we look at the other priorities. which are all three low indeed, but the odometer comes first: it's better in group A. then we look at the price, better in group A. Then we look at the horsepower, bad in both. Group A seems preferable”*. The balanced approach was most common in DECOUPLED COLORS used by 10/18 participants (■□), but less frequent in COUPLED BARS used by 6/18 participants (■□) and BASELINE by 4/18 participants (■□). As above, there is statistical evidence that the balanced approach is less frequent in BASELINE (and to a lesser degree COUPLED BARS) than it is in DECOUPLED COLORS.

(S3) - Decision by Individual Points or Outliers: Justifying a decision based on the specific details and anomalies within the data points rather than the overall trends or averages as a main justification. This was a rare strategy, used by 3 participants in total, 1/18 in each of the three visualizations (□). For example, P17 declared in DECOUPLED COLORS: *“The taxi that penalizes the mpg in group A already has a high odometer and may need to be replaced, which will improve the average mpg. In group B, replacing taxis with a high odometer will penalize the average mpg making it less attractive. In the long term, group A will, therefore, be better on all criteria”*.

**Summary on Decision Strategies:** For decision strategies (Figure 5.4) there is statistical evidence that most participants using BASELINE tend to adopt strategies that optimize one or two dimensions with the highest priority in the task. To a lesser extent, this is true when they use COUPLED BARS (more than half of participants). On the other hand, in DECOUPLED COLORS their strategies are more mixed, with more than half adopting a balanced approach.

### 5.2.1.3 Justification Quality

We now report on the quality of participants' reasoning as reflected in their explanations using the five quality measures discussed in section 5.1.5.1.

**1. Correctness: alignment with task goals:** We note that ‘Correctness’ [105] is too strong a term for our comparison tasks, as they have no ground truth and thus no basis to objectively judge them for correctness. Instead, we captured if participants explicitly mentioned the task objectives in their justifications, and whether the final choice they made is aligned with task objectives (similar to the accuracy measure in [16] discussed in section 2.6.3). By objectives, we mean the priorities (weighted objectives) and ideal values (min or max of an objective dimension) given as targets in the task. The low-level codes of this category consider the individual priorities and ideal values mentioned by participants, but here we report on two higher-level categories:

- Task Alignment: We encoded mentions of how closely decisions align with task goals (priorities and/or ideal values given as goals) and tagged as aligned choices that have a justification that follows these goals. Overall participants (18) and for all visualizations (3), participants understood well the priorities and direction of maximization/minimization (53/54 codes), with a single exception, discussed next.

- Task Misalignment: We found only one case (1/54) of task misunderstanding, in other words, statements that indicate that participants wrongly interpreted weights of objectives (priorities) or ideal values (minimization/maximization). Participant P15 with visualization `COUPLED BARS` misunderstood the meaning of the size of the bars, thus thinking that smaller bars mean better values when it is, in fact, the opposite.

**2. Completeness: number of unique context codes in a response:** The higher the completeness, the more unique codes (i.e. codes related to the position of a group within the bigger context, such as the range, position in a distribution, distribution of the group, number of solutions they include). This reflects a richer explanation as participants considered more types of information. In `BASELINE`, the completeness of participants’ explanations ranged from 0 to 4, with an average of 2.22. In `COUPLED BARS`, the completeness ranged from 0 to 4 unique context codes, with an average of 2.67. In `DECOUPLED COLORS`, the completeness ranged from 0 to 4 unique context codes, with an average of 2.61. We do not have strong evidence of a difference across techniques.

**3. Compactness: total number of context tags in a response:** In our case, high compactness is equivalent to a lower amount of data-driven information provided as justifications. `BASELINE` mean was the lowest / most compact; it measured a total number of context tags that ranged from 0 to 9, with an average of 3.61. In `COUPLED BARS`, the compactness ranged from 0 to 31 context tags (most between 1 and 5, with 3 outliers between 14 and 31), with an average of 5.61. In `DECOUPLED-COLORS`, compactness ranged from 0 to 12, with an average of 4.56. We do not have evidence of a difference across techniques.

**4. Contrastivity: number of pairs of arguments pro group A and pro group B:** In `BASELINE`, the contrastivity, represented by the number of pairs of arguments for Group A and Group B, ranged from 0 to 5 pairs, with an average of

1.94. In COUPLED BARS, contrastivity ranged from 0 to 4 pairs, with an average of 1.44. In DECOUPLED COLORS, contrastivity ranged from 0 to 5 pairs, with an average of 1.33. There is strong evidence that BASELINE had indeed the highest average contrastivity compared to the other two, indicating that participants were more likely to present arguments for both groups.

**5. Confidence:** In BASELINE, participants' confidence (or trust), in their decisions ranged from 4 to 7, with an average of 5.6. In COUPLED BARS, confidence ranged from 5 to 7, with an average of 5.94. In DECOUPLED COLORS, confidence ranged from 2 to 7, with an average of 5.17.

We have strong evidence that participants were less confident with DECOUPLED COLORS than COUPLED BARS.

We were surprised to see that the self-reported confidence scores are in favor of COUPLED BARS (5.93/7 on average) against BASELINE (5.6/7 on average). DECOUPLED COLORS is the setting where participants reported the least confidence in their choice (5.1/7).

**Summary on Quality of Explanations:** In terms of quality of explanation metrics, there is statistical evidence of few differences across visualizations. The BASELINE had more contrastivity than the other two visualizations. In terms of confidence and trust, contrary to **H2** participants were less confident with DECOUPLED COLORS than COUPLED BARS. This is contrary to our hypothesis **H2**, where we had hypothesized that decoupling of priorities and ideal values would lead to a higher 'understanding' and awareness of how each metric impacts the trade-off solutions and thus more confidence in one's decisions.

### 5.2.2 Perceived load - NASA TLX [RQ6,7]

We report next on the NASA TLX questionnaire results filled out by participants, as a measure of how demanding they found the tasks using the three visualizations. All questions are Likert scale items out of 21 (see [Figure 5.5](#)).

**Mental Demand.** We see that COUPLED BARS has the lowest mean (10.72), and there is strong evidence that it was perceived as less mentally demanding compared to both BASELINE (13.5) and DECOUPLED COLORS (13.28).

**Physical Demand.** As expected, physical demand was low for all tasks as they do not require any particular physical effort.

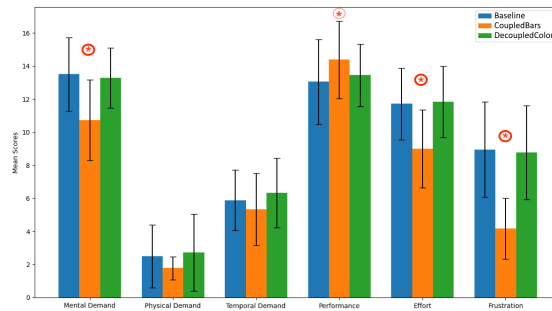


Figure 5.5: Mean values for the different categories of the NASA TLX questionnaires. Thick circled stars (\*) indicate strong statistical evidence of a difference, and thin circles a trend. All analysis was done using CI estimation (see Appendix A for detailed CIs).

**Temporal Demand.** No evidence of a difference, with mean temporal demand being slightly lower for COUPLED BARS (5.33), followed by BASELINE (5.89) and DECOUPLED COLORS (6.33).

**Performance.** There is possibly a trend for mean performance in COUPLED BARS (14.39) to be higher than BASELINE (13.06) but no other evidence of a difference. The mean for DECOUPLED COLORS (13.44) was between the two.

**Effort.** The COUPLED BARS has the lowest mean (9), and there is strong evidence that participants felt it required less effort compared to both BASELINE (11.72) and DECOUPLED COLORS (11.83).

**Frustration.** Once again COUPLED BARS shows the lowest mean (4.17), and there is strong evidence that it caused less frustration compared to both DECOUPLED COLORS (8.78) and BASELINE (8.94).

**Summary on Perceived Load:** When considering the results on perceived load (Figure 5.5), COUPLED BARS consistently exhibit lower mean values for mental demand, as well as requiring less effort and causing less frustration. BASELINE and DECOUPLED COLORS exhibit similar means across all metrics. Our results partially validate hypothesis **H1**: although indeed COUPLED BARS reduced mental load compared to BASELINE, DECOUPLED COLORS did not. The results also validate **H2** that decoupling priorities and ideal values would lead to a higher mental load, as DECOUPLED COLORS was perceived as more mentally demanding compared to COUPLED BARS.

### 5.2.3 Visualization Preferences and Qualitative Feedback

After the completion of the tasks, we asked the participants to rank the three visualizations based on preference (see figure 5.6) and explain their ranking. The results show that COUPLED BARS was the most preferred visualization, with 10 participants ranking it as their most preferred choice. DECOUPLED COLORS was the second most preferred visualization, with 8 participants ranking it as their most preferred choice. BASELINE was by far the least preferred visualization, with 0 participants ranking it as their most preferred choice.

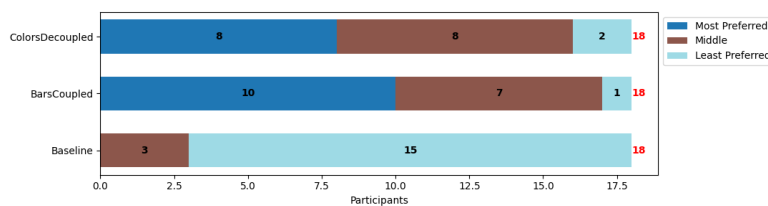


Figure 5.6: Participants' ranking of the visualizations

These results show that participants clearly prefer the visualization of trade-off metrics compared to the BASELINE that only includes textual data (and histograms). As P7 mentioned BASELINE: *"Much more difficult to make a decision here [BASELINE], having to go through all the data in text"*. And P17 adds that: *"The choice in BASELINE was made more on the whole using averages"*, highlighting a tendency to simplify the decision-making process due to the limitations of the visualization.

Slightly more participants preferred the coupled COUPLED BARS over the decoupled DECOUPLED COLORS (10/18). The participants who preferred COUPLED BARS generally found them easier to read. One such participant (P11) explained: *"The color representation greatly overloaded the reading of the data. The fact of having two distinct variables (color and opacity) to represent the priority and the ideal value did not facilitate handling. The representation of the relationship between priority and ideal value in the form of a single variable (the gray bar) was easy to use for me"*.

Nevertheless, there were several participants who preferred DECOUPLED COLORS (8/18). They felt it made outliers and discrepancies with their goal stand out. For example, (P10) mentioned *"The DECOUPLED COLORS is ideal for me with the colors that greatly facilitate comparison and allow you to see the big differences directly..."*. Another (P1) commented on the intuitiveness of the representation: *"I found the color and opacity combination very helpful in the analysis –the green to red scale is very intuitive, as is the scale from very solid color to nearly transparent. That aligned well with how I was already interpreting the decisions"*. Some of these participants commented on how (compared to colors) with COUPLED BARS they had *"difficulty assessing what the maximum of the bars corresponds to"* (P9).

### 5.3 Discussion and Limitations

Gleicher's et al. foundational work and our own interactions with domain experts allowed us to reflect on how their definition can be extended for comparisons for trade-off analysis. Our work underscores the relational and contextual facets highlighted in [101], but requires the addition of the trade-off context, represented by the user's goals and expressed as *criteria and priorities* (and the direction of their optimization). It also underlines the need to consider *groups of points* as the elements to assess, in terms of their differences and similarities over a multidimensional space (relative assessment between elements) and their position in the dataset (in a data-derived context).

Understanding what are the crucial elements that constitute the contextual facets that support this comparison is one of the contributions of our work. In our video analysis and workshop with expert users, we were able to identify, captured as key requirements, high-level elements in the form of metrics, levels of data detail, data ranges, and ranking. We explored how to visualize a set of these contextual factors in two design workshops, and our subsequent user study allowed us to investigate how concrete representations of these elements are used.

**User study results and discussion.** Next, we revisit our user study results with respect to our hypotheses. We hypothesized (**H1**) that trade-off specific visualizations would aid comparisons. This hypothesis was only partially confirmed. We chose as our baseline a tabular visualization that already provides elements to help comparisons, such as means, weighted scores, and dimension histograms. It was clear from participant preferences that such a visualization is not preferred; it was never their top choice for their task, and their comments highlighted how it made their comparison task difficult. Nevertheless, when looking at other measures captured in our study, the difference between the baseline and the trade-off visualization does not come across so strongly. In quality of justifications, the baseline had more contrastivity than other visualizations; in other words, participants tended to discuss pros/cons of their decisions more. It is possible this is because participants had a hard time making a choice. And when it comes to mental effort, it was only considered worse than only one of the trade-off visualizations (bars). Finally, confidence and trust in participants' decisions were not lower than the trade-off specific visualizations. This indicates that while dedicated trade-off visualizations are preferable, simpler tabular visualizations are still an effective representation as long as they provide context (histograms) and metrics (weighted scores and means).

We also expected **H2** that decoupled visualizations (colors in our case) would provide access to more information and lead to higher quality and trust, but also higher mental load than coupled visualizations (bars in our case). One part of our hypothesis was confirmed; decoupled colors were indeed more mentally demanding



than bars. Nevertheless, other aspects of our hypothesis were disproved. We found no evidence of higher quality of justifications, and most surprisingly confidence in the decisions was, in fact, lower with colors. It is hard to explain this lower trust. Given our original hypothesis, it is possible that some participants captured in colors the subtleties of trade-offs better, thus understanding that there was no clear "best" choice in our datasets and were thus less confident in their answers. Participant comments (from the ones that preferred colors) often highlighted that they had difficulty interpreting the coupled bar, and it was in this condition that we encountered the only case of a participant that misunderstood the meaning of a visualization. As our tasks do not have a ground truth, we rely on the quality of justification measures as a proxy to capture understanding. Nevertheless, we do not have evidence for differences between techniques for most metrics. This may be either because our introduced measures are not sensitive enough to capture understanding, or because the effect does not exist or is small.

Our initial results found that bar visualizations with a coupled metric (similar to that proposed in prior systems [132]) are best in terms of mental load and not worse than any of the other visualizations in other measures. Nevertheless, participant preferences were almost divided between bars and colors (10 vs. 8) do not shed light on this question. It was interesting to note mixed reactions to the color visualization; a big chunk of participants who preferred it explicitly commented on how it was very intuitive to understand the different metrics. Whereas participants who preferred the bars explained how colors were hard to understand and occasionally overwhelming, and bars were easy to use.

The justifications for our participants' final decision also hint at differences in the types of strategies the visualizations support. We see a statistical trend in the baseline tabular visualization (and to a lesser extent, the coupled bars) for most participants to adopt decision strategies that optimize one or two criteria only (what we refer to as strong maximization of priorities). Whereas in the decoupled colors, we see more varied strategies from the same participants, where balanced optimizations across all criteria are more frequent. It is possible that colors may better support balanced decision strategies, but this remains an open question. A more interesting question that emerges is whether specific visualizations promote some decision strategies more than others and can thus influence decision-makers. Our results suggest this is indeed the case for the baseline, where most decisions were strong maximization of few priorities and should thus be avoided if balanced decision strategies are important.

**Limitations and future work.** We acknowledge that our experts deal with complex datasets often derived from models of biological processes. There are clearly everyday trade-off comparison tasks that involve fewer criteria and data points, possibly fewer needs in terms of contextual facets, or are simple choice tasks between elements

(rather than groups). In our workshops, we were careful in our choice of datasets and tasks to ensure they reflect what we observed from domain experts. More challenging comparison tasks are possible, such as when comparing then merging groups of points to create a new one, and when decision-makers need to observe the complementarity of individual points within a group (section 4.1.4). However, our user study focuses on comparison tasks that are perhaps less challenging but more common as we saw from our expert and design workshops. It is also possible the datasets, tasks and even the inspirational material influenced the generated visualization designs that inspired our prototype. Nonetheless, given how the design requirements align with those that experts expressed, we feel confident in the identification of what our visualizations need to provide.

Our visualization prototype supports several possible ways to combine trade-off metrics and visual encodings. In an effort to keep our study duration reasonable (trade-off analysis is complex; each session lasted  $\approx 2$ h), we limited the number of variations to test in practice to three: one driven by frequent designs appearing in our workshop (decoupled colors), one from the state of the art (coupled bars), and one that we consider a usable baseline. This resulted in a limitation of our user study: we compared a mix of metric coupling and visualizations (bars with coupled metrics and colors with decoupled). We are thus not able to study in isolation the effect of coupling. It is thus possible some of our results, e.g., higher mental load in decoupled colors, are more due to the visual encoding than the metric coupling. Clarifying this question remains future work.

## 5.4 Conclusion

Comparison is an important task in multi-criteria decision-making. Although prior work has considered the comparison of individual points, comparing *groups of points* remains an open challenge. To better identify user needs for comparisons of groups, we conducted a workshop with experts that provided valuable insights into how trade-off comparisons are conducted and what support visualizations need to provide. As existing visualization tools do not support comparisons of groups, in the previous chapter we conducted two design workshops that in turn inspired the implementation of a prototype to support group comparisons.

As we had to consider several design decisions about how to present trade-off metrics, we evaluated two trade-off metric visualizations against a baseline. One of the two visualizations used color and decoupled the trade-off metrics, mapping one (weights) onto opacity and another (distance to ideal) onto color hue. The other visualization used bars to encode trade-off metrics and coupled both metrics, mapping them onto one single visual encoding - bar length. Our findings show that the visualization separating different trade-off metrics (colors) leads to higher mental load and lower

self-reported trust but may support more varied decision strategies than more integrated visualizations like bars. But both visualizations encoding trade-off metrics were preferred over a baseline that provided simple distribution visualizations (histograms) and weighted scores in textual format.

In our work, we focused on explicitly designing a prototype for conducting group comparisons and tested variations of trade-off metric visualizations. It is clear that such support needs to be integrated within a broader trade-off analysis workflow where analysts do not only compare but also explore their data to understand what trade-offs exist and set or refine their priorities. Our comparison prototype supports the refinement of priorities and the adjustment of importance (weights) but it is not a full exploration system. Rather, we see its use as an on-demand feature available within a larger visual analysis system (e.g., our tool from [Chapter 3 \[97\]](#)) the moment analysts need to compare specific solution spaces. Our analysis of expert needs suggests these moments of explicit comparisons exist and are often involved in making crucial decisions. Nevertheless, how to integrate these features without disrupting the analysis flow, and understanding how they will be used by domain experts in longer analysis sessions, remain future work.



# Discussion, Future Work and Conclusion

We next summarize this thesis and thesis statement, discuss future work and conclude with some closing remarks.

## 6.1 Summary

Trade-off analysis is a complex process, particularly in domains such as agronomy, where decision-makers need to balance multiple, often conflicting objectives such as environmental sustainability, crop yield, and resource efficiency. Experts need effective tools to support them as they navigate these multi-criteria decision-making scenarios. This thesis addresses that need, focusing on both provenance visualization and comparison visualization techniques to enhance decision-making in these contexts.

This work has three main contributions: (i) understanding how provenance can support visual trade-off analysis conducted by domain experts in agronomy, (ii) articulating high-level design and visualization needs to support group comparison in trade-off scenarios in this context, and finally (iii) understanding how different visualizations can affect comparisons and decision-making in trade-off analysis. After an introductory chapter and a chapter on related work, the thesis details these three main contributions.

We first conducted a study on analytic provenance visualization to better support trade-off analysis. Provenance gives analysts the ability to track and recall the steps taken during an analysis which is crucial in trade-off tasks where experts need to revisit their decisions and the paths they followed over time. The 3rd chapter of the thesis investigates how analytic provenance mechanisms can assist experts in recalling and tracking complex trade-off analysis. We developed VisProm, a web-based system integrating in-visualization provenance views to help experts track trade-offs and their objectives when exploring complex simulation results. Through observation sessions with three groups of experts analyzing their own data, we make the following contributions. We first, identify eight high-level tasks that experts engaged in during trade-off

analysis, such as locating and characterizing interest zones in the trade-off space, and show how these tasks can be supported by provenance visualization. Second, we refine findings from previous work on provenance purposes such as recall and reproduce, by identifying specific objects of these purposes related to trade-off analysis, such as interest zones, and exploration structure (e.g., exploration of alternatives and branches). These objects answer the question, "What does the analyst want to X?" (where X corresponds to one of the purposes for provenance as defined by Ragan et al. [31]). Thus provenance objects give a deeper understanding of the analysts' intent during exploration and should be considered when designing provenance views for trade-off analysis. By incorporating these objects, the design can move beyond just recalling steps and help support more sophisticated decision-making tasks. Third, we discuss insights on how the identified provenance objects and our designs support these trade-off analysis tasks, both when revisiting past analysis and while actively exploring. Finally, we identify new opportunities for provenance-driven trade-off analysis, for example related to monitoring the coverage of the trade-off space, and tracking alternative trade-off scenarios.

Building on the outcomes of our provenance work, we conducted a follow-up study on comparison visualization to further support trade-off analysis. Comparison is a fundamental element of trade-off decision-making, especially when experts are faced with groups of solutions or simulation results representing different outcomes. To better understand the needs for comparison visualization in this context, we organized three workshops with domain experts and expert users. These workshops revealed the critical role of visualizing relationships between solution sets and how comparisons help prioritize different trade-offs. Then, using insights from these workshops, we held two additional design workshops with HCI and visualization experts, where participants developed designs for comparison tools. From these explorations we contribute high-level design and visualization needs to support group comparison in trade-off scenarios which lay the groundwork for developing effective visualization techniques for comparing groups that represent different trade-offs in terms of what objectives they optimize. These interactive visualizations encode and visually communicate experts' priorities in terms of objectives, the notion of ideal solutions, and how far current groups of solutions are from those ideals.

From these designs, we implemented comparison visualizations in a prototype tool. We chose two to evaluate more closely: *DecoupledColors* and *CoupledBars*. *DecoupledColors* separates trade-off metrics visually, allowing users to see how solutions perform across distinct objectives, while *CoupledBars* integrates these metrics into a single, holistic view that simplifies comparisons. In a user study, we evaluated the effectiveness of both visualization approaches in terms of several metrics, such as quality of explanations to justify decisions, trust, and mental load. We first characterized the types of information participants used for making comparisons. In addition, our

results showed that DecoupledColors, while it increases the cognitive load on users, enables diverse decision-making strategies. In contrast, CoupledBars reduces mental load and fosters greater trust by simplifying the comparison process, though it may lead to decision strategies that favor fewer criteria. Both visualizations, however, were preferred over scenarios without any visualization of trade-off metrics, emphasizing the importance of providing visual representations of trade-offs to support decision-making.

Collectively, our work shows that visualizations can aid trade-off analysis, our *thesis statement*. In particular, we show that additions to traditional visualizations (like scatterplots or tables) in the form of in-visualization provenance or visual indicators of trade-off metrics can support several trade-off tasks.

In the following section, we conclude this thesis by discussing the implications of these previous findings for the use of provenance and comparison visualizations in the context of trade-off analysis. We also highlight future opportunities for enhancing trade-off analysis through visualization and the potential to further improve experts' decision-making in trade-off analysis.

## 6.2 Future perspectives

In this thesis, we explored the design and implementation of visualizations to support trade-off analysis, a specific type of data analysis characterized by complex solution spaces, conflicting optimization objectives and changing user preferences. This section reflects on the key challenges we encountered during this research, how some of these challenges were addressed, and the potential avenues for future work.

### 6.2.1 Tracking Conflicting Objectives and Evolving User Preferences in Extended Exploration Sessions

Trade-off analysis requires balancing user preferences across dimensions to minimize or maximize, that may be conflicting. The priority of these preferences may evolve throughout the exploration process as the analyst gains deeper insights into their data, considers alternatives, and may re-calibrate their vision of ideal solutions, becoming more realistic, conservative, or liberal with their constraints. In our experience and in reports from past work with domain experts [5, 97, 112], trade-off exploration sessions have varied in length from minutes to hours, and sometimes extended over multiple consecutive days or spread out over time, leading to the generation of large, complex log files.

These observations guided our decision on which aspects of the trade-off analysis to track, referred to as provenance objects in [Chapter 3](#). Provenance objects are defined as key elements of the trade-off analysis process or the data records that participants were primarily interested in during their analysis. Although we identified eight such

objects at different abstraction levels, in our work we only implemented those at a low analytical level, such as view and query selections (e.g., interest zones, exploration structure and steps, focus and priority, coverage). While this approach sufficed for shorter exploration sessions and aligned with our goal of creating simple in-visualization provenance views rather than external rich summaries, there remains a need to define and track higher-level provenance objects for longer, more extended sessions. These could include provenance objects related to insight actions (e.g., data semantics and origin, annotation, verification, and validation), preferences (e.g., shifting of priorities over time), or meta-actions (e.g., exploration approach).

Tracking and visualizing provenance data related to low level exploration actions is relatively straightforward and currently well-supported in interactive systems. However, capturing provenance at higher abstraction layers presents more challenges, despite advances like automated data insights [120]. This highlights the need for further research into detecting insight actions, exploration structures, and higher-level tasks such as group comparisons, and exploring how to visually summarize them, whether in-situ, ex-situ, or through a combination of both.

### 6.2.2 In-situ, Ex-situ or Hybrid Visualizations of Trade-off Analysis Provenance

In our work, we chose simple in-situ visualizations of provenance for two main reasons. First, to ensure provenance information was not overlooked during exploration (and to gather data on provenance usage in our observational study). And second, we assume that if these simple views prove useful for trade-off analysis, more advanced provenance views could be explored in the future. We confirmed the usefulness of simple in-situ visualization in our observational study described in [Chapter 3](#), as it helped our experts recall, replicate, recover, reflect upon, and present different aspects of their exploration.

Furthermore, more user studies are needed to compare the effectiveness of different provenance layout configurations. Not only as integrated in-situ views within the main trade-off analysis visualization (that we have studied so far), but also as separate views or a combination of both (hybrid).

**In-visualization Provenance:** Future work could investigate richer provenance views that incorporate more aspects of provenance, particularly those related to higher abstraction levels, such as insight and meta-actions. The challenge that we are faced is balancing what information is crucial to add, given that in-visualization views add clutter and compete with the main visualization the experts are focusing on.

Research on visual saliency for data visualizations (e.g., [87]) can inform future work on how to render in-situ provenance views. We may follow their recommendation to enhance the visibility of provenance views when significant information emerges,



such as major shifts in optimization priorities or indications of bias [41]. In contrast, less salient visualizations can be used by default avoiding distraction from the main analysis tasks and ensuring they do not break the workflow [19, 42].

**Separate Provenance Views:** A separate view layout can consist of a main trade-off analysis view where primary trade-off analysis tasks occur, such as minimizing and maximizing, ranking and prioritizing, and combining interest zones and dimensions. And secondary provenance views to display provenance information at both low and high levels, that are either seen at a side window or available on demand. Existing work on visualizing branching and decision-making in sensemaking tasks, as shown by Heer et al. [61], could be used to display provenance information in separate views dedicated to, for example, assisting experts in tracking the evolution of their preferences, such as in a tree structure [138] or a timeline visualization [116]. Insights from storytelling [24] could make summary visualizations of long trade-off exploration session data more engaging.

**Hybrid Layout:** A hybrid view layout can include both in-situ and separate views of provenance. In-situ visualizations could be used to display coverage, as our work revealed that coverage provenance information was frequently consulted *during* the analysis for recall. In contrast, visualizations for branching could be moved or offloaded to a separate view, as they are likely to be consulted less frequently. This separate view would also be suitable for exploratory analysis with high-level provenance purposes in mind, such as for meta-analysis or when reflecting on the exploration approach prior to handover. Our findings suggest what information can be shown on what view (during vs. a-posteriori). But it remains future work to empirically confirm the choice of what provenance information needs to be provided in-situ and what should be available on-demand. And to study how analysts integrate the different sources of provenance views.

Existing work from various domains can guide the choice of provenance layout configuration, whether in-situ, separate, or hybrid, as well as inform the design and evaluation of provenance views: For example, the field of Multiple Coordinated Views (CMV) [55, 150] can provide insights into how to effectively link the main trade-off analysis view with the provenance views, possibly through a specialized form of brushing and linking. It can also help evaluate the impact of multiple-view patterns, which Chen et al. [150] define as *composition* (which quantifies the types and numbers of views), and *configuration* (which characterizes the spatial arrangement of view layouts in the display space). Applying though these approaches to provenance views needs to be done with caution taking into account the context of user interactions: e.g., when a

user makes a selection, should this be reflected to other views of the main visualization, to provenance views, or both.

### 6.2.3 Evaluating the Effectiveness of Trade-off Analysis Tools

Trade-off analysis is characterized by open-ended exploration, multiple paths and branches, and the subjective nature of expert preferences. Ideal solutions or decision choices are typically not known in advance. Even final choices (ex., in comparison) may only be relevant or preferable to the specific analyst, reflecting their expertise along with their subjective preferences and priorities. Evaluating the effectiveness of trade-off analysis tools is therefore challenging, as there is no definitive “ground truth” to assess against.

Additionally, extra visualizations or metrics, such as those for provenance and trade-off metrics (used in our work on group comparisons), can aid users in tracking conflicting objectives and shifting preferences or quickly comparing groups of solutions. However, these additions may also increase display clutter and the associated mental load for users.

In our work, we employed a combination of qualitative and quantitative methods to assess the effectiveness of our visualization technology probes. Given that our research questions were open-ended and exploratory for the first work investigating the role of provenance in trade-off analysis, we adopted a qualitative study methodology. We conducted thematic analysis [142, 143, 144] of video recordings from exploration sessions with domain experts, and utilized qualitative feedback and quotes from participants to characterize their use of provenance and support our findings. Coding provenance events from video recordings, even with a think-aloud protocol, proved challenging, as experts did not always articulate their use of provenance explicitly; they were more focused on the primary task of trade-off analysis than on the secondary task of provenance tracking.

Even though it was challenging to conduct and analyze, this type of user study is extremely fruitful and informative. Alongside the rich findings on trade-off analysis tasks and the provenance objects identified in our study (Chapter 3), we were also able to highlight several trade-off specific opportunities for various provenance purposes, including recall, replication, meta-analysis, representation, and collaboration (see Section 3.3.4). Importantly, the results of this work informed our subsequent project on visualizations designed to support trade-off comparisons within groups (Chapter 5). We are certain these broad findings would not have been possible in a more controlled user-study setting.

For our study of how to support trade-off group comparisons, our research questions focused on evaluating three visualization methods. Thus, we employed a controlled user study method, measuring the effectiveness of the compared visualization through

user understanding, trust in decisions, and mental load. In the absence of ground truth, we used the quality of explanations participants provided for their choices as a proxy for their understanding of the data and trade-offs. To evaluate the quality of explanations, we adapted metrics from the field of explainable AI [105], including correctness, completeness, contrastivity, compactness, and confidence, while excluding metrics relevant only to AI. We assessed trust and mental load using the NASA TLX questionnaire [131].

Nevertheless, we were unable to identify differences across visualizations in terms of quality of explanations (even though there were differences in strategy and mental load). As we discussed in Chapter 5 it is likely these measures are not sensitive enough to measure differences in quality between different trade-off visualizations. We considered other effectiveness measures, such as response time, but ultimately decided against it, as quick decisions do not necessarily correlate with sound choices. For instance, a participant might randomly select a response, and since each solution is a valid alternative in a trade-off dataset, we cannot distinguish this behavior from that of someone who makes quick but informed decisions.

Another possible evaluation measure is the consistency of answers with respect to original preferences [34]. However, this is harder to evaluate in our context, as again, our trade-off datasets lack a ground truth. Consequently, any choice can be valid as long as the participant can justify it and the justification aligns with their *interpretation* of the indicated trade-off preferences. From our observations, when given a set of trade-off preferences, participants tend to either strongly maximize the few objectives with the highest priority or consider less prioritized dimensions, striving for a balance among all objectives. Both strategies are valid as long as the participant makes a purposeful choice and can justify it.

In our work, we consider that human expertise and judgment are crucial for decision-making, and thus other evaluation metrics, such as user satisfaction with decision outcomes, could also be used. Longitudinal studies could be conducted to investigate the long-term outcomes of trade-off analysis, and the impact visualizations can have on them. For example, our study in comparisons indicated that different visualizations may favor different decision strategies: a simple table with histograms and scores (and to a lesser degree bars) favors a min/maxing strategy. Future work could investigate if analysts can choose visualizations according to the type of strategy they want to adopt.

#### 6.2.4 Visualization of Gain and Loss

If we consider that comparison in general data exploration is a way to characterize the relationships between datapoints [101], then in trade-off analysis it is also a way to assign value to the datapoints according to the trade-off preferences of the analyst. In other words, instead of just finding differences and similarities, we also find pros and cons between datapoints. Pros and cons for the solution entail a gain or a loss when

selecting one solution over another. In the context of the optimized (Pareto front) dataset our experts explore, each solution represents a gain in at least some aspect compared to all the other solutions and a loss in other aspects.

**Gain and Loss:** In fields such as behavioral economics, this idea of weighing gain versus loss to make decisions has been discussed in terms of a mathematical metric for decision-making in the context of mixed gambles. A mixed gamble is a decision or bet that involves both potential gains and potential losses. It is often used in the context of decision-making under uncertainty and prospect theory [15, 79].

Given that each trade-off solution represents a gain and a loss in some aspect compared to all the other solutions, we could argue that mixed gambles are a good way to represent trade-off analysis. The only difference is that in trade-off analysis, the odds are not necessarily random. For example, there may be an inverse correlation between say fertilization amount and crop yield, thus a set of decisions that lead to a high gain in one dimension will necessarily lead to high loss in another.

In our work on the comparison of groups of trade-off solutions, we have touched on the idea of gain and loss by visualizing how each solution relates to the ideal point of the experts. But we could imagine a system where the comparison would not be in relation to a theoretical ideal point, but between a selected point from the data and all the other points, or between different alternative compositions of a solution group that the analyst composes by adding or removing points.

Future work could investigate how such real-time interactive quantification and visualization of gain and loss associated with each alternative affects the process of trade-off analysis. In addition, such visualizations could be studied in tasks where analysts are trying to compose a group of solutions, merging datapoints across different groups (a comparison task we identified but did not study). When composing a solution group, gain and loss indicators could be used to help experts make better groups, showing them in real-time how the composition of the group affects the strengths and weaknesses of the group as a whole.

**Ranking Systems:** Another avenue for future work is to test different ways to rank the alternatives other than just the weighted means used typically. For example, the Elo rating system, traditionally used in ranking players in competitive games [88], could offer a novel way to model and quantify the relative strengths of different trade-off solutions in multi-criteria decision analysis, similar to what has been done in other applications [66, 69]. In the Elo system, each solution can be assigned an initial score, which adjusts dynamically as solutions are compared to one another, based on their strengths in different criteria. For instance, when two solutions are compared, the one with a perceived advantage in a particular criterion (e.g., higher MPG) gains points, while the other loses points, reflecting their relative performance on that criterion.

This could even take into account the analyst's priorities in the dimensions and assign a different amount of gained or lost points depending on how important the dimension is. Over time, this approach could provide a ranking of solutions based on how often they "win" in specific comparisons across various criteria.

Applying the Elo system to trade-off analysis would allow decision-makers to iteratively compare solutions and see real-time adjustments in rankings, providing insight into which solutions consistently perform better across multiple dimensions and trade-off priorities. This could also help quantify the degree of gain or loss in each decision, aligning with the concept of mixed gambles from behavioral economics.

### 6.2.5 Integrating Provenance and Comparison to Support Trade-off Analysis

In this thesis we designed two prototypes in an attempt to study how visualizations for provenance and group comparison can improve trade-off analysis. They were designed and implemented as two separate tools. In this section we discuss possible future work to integrate both tools in a single trade-off analysis workflow.

We can imagine such a workflow to be SPLOM based, like VisProm (described in [section 3.1](#)). As such users can select groups of points from a scatterplot, and navigate through the SPLOM for an overview of the data context. Then, when important solution groups are identified (as we observed in our study with experts), users can decide to view these selections in a dedicated comparison view, such as the prototype described in [section 4.3](#). Each selection will then be presented in a floating table together with the control sliders to express the users' preferences and priorities per dimension. We note that our experts often revisited the tabular view of their data in the SPLOM based exploration VisProm (some keeping it always visible on the side). It thus seems natural to combine them. But it also raises the question if this switch of visual representation between comparison and exploration breaks the analysis flow.

Conversely, the comparison of trade-off solutions could benefit from provenance visualization, as analysts need to track the trade-off preferences they used in their comparisons and how those preferences evolved over time. Tracking the solutions that were compared is also important to assess data coverage (an important task as we identified in [Chapter 3](#)) and ensure that experts have sufficiently explored potential solutions that align with their trade-off preferences. Visualizing the trade-off preferences behind each comparison could help inform experts of the goal of each selection and provide insight into how their priorities have shifted throughout the process.

### 6.2.6 Collaboration in Trade-off Analysis

In agronomy, many trade-off tasks require collaboration between experts from different fields [[112](#)], each contributing to the analysis with their own expertise and priorities. In

our work, even though most of the exploration sessions we conducted were with pairs of experts, we did not study how the experts collaborated in their exploration process, nor how we could design tools to support this collaboration [119].

We have, however, identified a number of opportunities for future work in this area. For example, we have noticed that when in pairs, experts would naturally fall into one of two roles: one that leads the interaction directly manipulating the tool and the other that takes notes and asks questions or makes suggestions. While this may be an artifact of our study setup (single display and mouse), prior work using a shared larger display identified similar behavior [112].

Literature on asynchronous collaboration based on note-taking and knowledge graphs [76, 148] has commented on how colleagues report on their findings and explain during hand-off the sources they have already explored. Since this literature aims to support hand-off situations, the reports made with the tools and techniques they propose often contain the insights from the exploration and their provenance. This could help even in synchronous collaboration for complex subjects and situations, like trade-off analysis, where the exploration sessions are separated by long spans of time, and analysts need to get back up to speed at the start of each session on what has been done.

Future work could also take advantage of the advances in the field of generative AI and research on Large Language Model (LLMs), and use such models to support the exploration by summarizing for example. In their work, Zhang et al. [141] test various models for the summarization of news against human freelance writers and find that the performance of the models are comparable to the human writers and better on some aspects like objectivity. It would be worthwhile to see if such summarization could help analysts recover key information before resuming past explorations.

We could imagine, for example, a system that uses the auto transcription of the discussions between the experts and LLM models to extract information about the trade-off preferences of the experts and how they evolve over time, as well as the different strategies they use to make comparisons, in a non-intrusive way. During the exploration, this approach could provide experts with real-time notes and summaries of their discussions and decisions, potentially helping them keep track of their exploration steps. And after the exploration, it could provide a summary of the analysis process and the decisions that were made, which the experts could revisit or present to collaborators in a hand-off or reporting situation for example. While this is a promising avenue of research it also poses the question of whether the summaries provided, especially if not accurately reflecting the choices made, may themselves influence the outcome of the analysis.

## 6.3 General Conclusion

In this dissertation, we have presented a series of studies aimed at understanding the visualization needs of experts engaged in trade-off analysis, a complex data analysis process focused on balancing and prioritizing objectives. Our findings demonstrate that provenance visualization aids experts in navigating complex and long trade-off analysis sessions and in recalling past findings and exploration processes. We have also identified key objects related to provenance purposes that we hope will help future visualization designers make more informed designs. Experts also expressed their appreciation of the additional support that our provenance visualization provided in their trade-off analysis tasks. Additionally, our research shows that in the context of group comparison, visualizations that highlight how each solution relates to trade-off preferences, aids the comparison of groups and can reduce mental load. And are always preferred over a visualization that does not provide such visual trade-off indicators. It also indicated that the choice of group visualization could influence experts' decision strategies, favoring strategies that focus on optimizing fewer objectives, or conversely supporting a variety of strategies. This underscores the importance of carefully considering design choices in comparison visualizations to effectively support trade-off analysis and decision-making. These results enable us to confidently say that, in accordance with our thesis statement, visualization can indeed support experts in their trade-off analysis tasks. We hope our work inspires further research in trade-off analysis with its unique challenges and proves valuable in designing tools that assist experts in their decision-making, especially when the path to a good solution is unclear.





# Bibliography

- [1] 2023. Data to Viz. <https://www.data-to-viz.com> Accessed: 2023-03-10. 82
- [2] 2023. Obsidian Dataview. <https://blacksmithgu.github.io/obsidian-dataview/> Downloaded: 2023-03-10. 82
- [3] 2024. Cars API Documentation. <https://api-ninjas.com/api/cars> Accessed: 2024-02-23. 80
- [4] 2024. Craigslist Cars and Trucks Data. <https://www.kaggle.com/datasets/prena0808/craigslist-cars-and-trucks-data> Accessed: 2024-02-23. 80
- [5] A. Barczewski, A. Bezerianos, and N. Boukhelifa. 2020. How Domain Experts Structure Their Exploratory Data Analysis: Towards a Machine-Learned Storyline. In *Extended Abstracts of the Annual ACM Conference on Human Factors in Computing Systems (CHI 2020)*, Honolulu, HI, USA. ACM, 18. <https://doi.org/10.1145/3334480.3382845> 4, 43, 69, 119
- [6] A. Bezerianos, F. Chevalier, P. Dragicevic, N. Elmqvist, and J. D. Fekete. 2010. Graphdice: A System for Exploring Multivariate Social Networks. In *Proceedings of the 12th Joint Eurographics - IEEE VGTC Conference on Visualization (EuroVis 2010)*, Bordeaux, France. The Eurographs Association & John Wiley & Sons, Ltd., 863–872. <https://doi.org/10.1111/j.1467-8659.2009.01687.x> 27, 42
- [7] A. Kerren and F. Schreiber. 2012. Toward the Role of Interaction in Visual Analytics. In *Proceedings of the Winter Simulation Conference (WSC 2012)*. 1–13. <https://doi.org/10.1109/WSC.2012.6465208> 37
- [8] A. Laurent, P. Kyveryga, D. Makowski, and F. Miguez. 2019. A Framework for Visualization and Analysis of Agronomic Field Trials From On-Farm Research Networks. *Agronomy Journal* 111, 6 (2019), 2712–2723. <https://doi.org/10.2134/agronj2019.02.0135> 16
- [9] A. Marsaw, B. German, P. Hollingsworth, D. Mavris, and S. Csonka. 2007. An Interactive Visualization Environment for Decision Making in Aircraft Engine Preliminary Design. In *Proceedings of the 45th AIAA Aerospace Sciences Meeting and Exhibit*. 1333. <https://doi.org/10.2514/6.2007-1333> 27, 30
- [10] A. Mayorga and M. Gleicher. 2013. Splatterplots: Overcoming Overdraw in Scatter Plots. *IEEE Transactions on Visualization and Computer Graphics* 19, 9 (2013), 1526–1538. <https://doi.org/10.1109/TVCG.2013.65> 23
- [11] A. Skopik and C. Gutwin. 2005. Improving Revisitation in Fisheye Views with Visit Wear. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems (CHI 2005)*, Portland, Oregon, USA. ACM, 771–780. <https://doi.org/10.1145/1054972.1055079> 33

- [12] A. Strauss and J. Corbin. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Sage publications. <https://doi.org/10.4135/9781452230153> 51
- [13] B. Bischl, G. Casalicchio, M. Feurer, P. Gijbbers, F. Hutter, M. Lang, R. G. Mantovani, J. N. van Rijn, and J. Vanschoren. 2021. OpenML Benchmarking Suites. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks 2021, December 2021, Virtual*. <https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/hash/c7e1249ffc03eb9ded908c236bd1996d-Abstract-round2.html> 74, 76
- [14] B. Ceballos, M. T. Lamata, and D. A. Pelta. 2016. A Comparative Analysis of Multi-Criteria Decision-Making Methods. In *Progress in Artificial Intelligence*, Vol. 5. 315–322. <https://doi.org/10.1007/s13748-016-0093-1> 1, 17
- [15] B. de Langhe and S. Puntoni. 2015. Bang for the Buck: Gain-Loss Ratio As a Driver of Judgment and Choice. *Management Science* 61, 5 (2015), 1137–1163. <https://doi.org/10.1287/mnsc.2014.2045> 124
- [16] B. Dy, N. Ibrahim, A. Poorthuis, and S. Joyce. 2022. Improving Visualization Design for Effective Multi-Objective Decision Making. *IEEE Transactions on Visualization and Computer Graphics* 28, 10 (2022), 3405–3416. <https://doi.org/10.1109/TVCG.2021.3065126> 38, 108
- [17] C. Dunne, N. H. Riche, B. Lee, R. A. Metoyer, and G. G. Robertson. 2012. Graphtrail: Analyzing Large Multivariate, Heterogeneous Networks While Supporting Exploration History. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems (CHI 2012), Austin, TX, USA*. <https://doi.org/10.1145/2207676.2208293> 33
- [18] C. Herzog. 2022. On the Risk of Confusing Interpretability with Explicability. *AI Ethics* 2, 1 (2022), 219–225. <https://doi.org/10.1007/s43681-021-00121-9> 38
- [19] C. M. Chewar, D. S. McCrickard, A. Ndiwalana, C. North, J. Pryor, and D. Tessendorf. 2002. Secondary Task Display Attributes - Optimizing Visualizations for Cognitive Task Suitability and Interference Avoidance. In *Proceedings of the 4th Joint Eurographics - IEEE TCVG Symposium on Visualization (VisSym 2002), Barcelona, Spain*. Eurographics Association, 165–171. <https://doi.org/10.2312/VISSYM/VISSYM02/165-171> 121
- [20] C. North, R. Chang, A. Endert, W. Dou, R. May, B. Pike, and G. A. Fink. 2011. Analytic Provenance: Process+interaction+insight. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems (CHI 2011), Extended Abstracts, Vancouver, BC, Canada*. ACM, 33–36. <https://doi.org/10.1145/1979742.1979570> 2, 41

- [21] C. Perin, P. Dragicevic, and J. D. Fekete. 2014. Revisiting Bertin Matrices: New Interactions for Crafting Tabular Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2082–2091. <https://doi.org/10.1109/TVCG.2014.2346279> 26
- [22] C. Schmid and H. Hinterberger. 1994. Comparative Multivariate Visualization Across Conceptually Different Graphic Displays. In *Proceedings of the Seventh International Working Conference on Scientific and Statistical Database Management*. 42–51. <https://doi.org/10.1109/SSDM.1994.336963> 36
- [23] C. Tominski, C. Forsell, and J. Johansson. 2012. Interaction Support for Visual Comparison Inspired by Natural Behavior. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2719–2728. <https://doi.org/10.1109/TVCG.2012.23736>
- [24] C. Tong, R. C. Roberts, R. Borgo, S. P. Walton, R. S. Laramée, K. Wegba, A. Lu, Y. Wang, H. Qu, Q. Luo, and X. Ma. 2018. Storytelling and Visualization: An Extended Survey. *Information* 9, 3 (2018), 65. <https://doi.org/10.3390/INF09030065> 121
- [25] C. Viau, M. J. McGuffin, Y. Chiricota, and I. Jurisica. 2010. The FlowVizMenu and Parallel Scatterplot Matrix: Hybrid Multidimensional Visualizations for Network Exploration. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1100–1108. <https://doi.org/10.1109/TVCG.2010.20525>
- [26] D. A. Keim and H.-P. Kriegel. 1994. VisDB: Database Exploration Using Multidimensional Visualization. *IEEE Computer Graphics and Applications* 14, 5 (1994), 40–49. <https://doi.org/10.1109/38.310723> 23, 24, 25
- [27] D. A. Keim and H.-P. Kriegel. 1996. Visualization Techniques for Mining Large Databases: a Comparison. *IEEE Transactions on Knowledge and Data Engineering* 8, 6 (1996), 923–938. <https://doi.org/10.1109/69.553159> 24
- [28] D. A. Keim, G. L. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. 2008. Visual Analytics: Definition, Process, and Challenges. In *Information Visualization - Human-Centered Issues and Perspectives*, J.-D. Fekete A. Kerren, J. T. Stasko and C. North (Eds.). Lecture Notes in Computer Science, Vol. 4950. Springer, 154–175. [https://doi.org/10.1007/978-3-540-70956-5\\_7](https://doi.org/10.1007/978-3-540-70956-5_7) 12
- [29] D. Ceneda, T. Gschwandtner, T. May, S. Miksch, H.-J. Schulz, M. Streit, and C. Tominski. 2017. Characterizing Guidance in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 111–120. <https://doi.org/10.1109/TVCG.2016.2598468> 2
- [30] D. Gotz and M. X. Zhou. 2009. Characterizing Users' Visual Analytic Activity for Insight Provenance. *Information Visualization* 8, 1 (2009), 42–55. <https://doi.org/10.1057/ivs.2008.31> 66

- [31] E. D. Ragan, A. Endert, J. Sanyal, and J. Chen. 2016. Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 31–40. <https://doi.org/10.1109/tvcg.2015.2467551> 2, 8, 32, 34, 41, 51, 52, 54, 56, 59, 60, 66, 69, 118
- [32] E. D. Ragan, J. R. Goodall, and A. Tung. 2015. Evaluating How Level of Detail of Visual History Affects Process Memory. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems (CHI 2015)*, Seoul, Republic of Korea. ACM, 2711–2720. <https://doi.org/10.1145/2702123.2702376> 48, 49, 68
- [33] E. Dimara. 2017. *Information Visualization for Decision Making : Identifying Biases and Moving Beyond the Visual Analysis Paradigm*. Ph.D. Dissertation. Université Paris Saclay (COMUE). <https://theses.hal.science/tel-01968033> 15
- [34] E. Dimara, A. Bezerianos, and P. Dragicevic. 2018. Conceptual and Methodological Issues in Evaluating Multidimensional Visualizations for Decision Support. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 749–759. <https://doi.org/10.1109/TVCG.2017.2745138> 2, 17, 26, 38, 42, 66, 81, 86, 123
- [35] E. Dimara, G. Bailly, A. Bezerianos, and S. Franconeri. 2019. Mitigating the Attraction Effect with Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 850–860. <https://doi.org/10.1109/TVCG.2018.2865233> 15
- [36] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. 2020. A Task-Based Taxonomy of Cognitive Biases for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics* 26, 2 (2020), 1413–1432. <https://doi.org/10.1109/TVCG.2018.2872577> 15
- [37] E. Kandogan. 2000. Star Coordinates: A Multi-Dimensional Visualization Technique with Uniform Treatment of Dimensions. In *Proceedings of the IEEE Information Visualization Symposium, Late Breaking Hot Topics*. <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.4.8909> 25
- [38] E. Lichtfouse, M. Navarrete, P. Debaeke, V. Souchère, C. Alberola, and J. Ménassieu. 2009. Agronomy for Sustainable Agriculture: a Review. *Sustainable Agriculture* (2009), 1–7. [https://doi.org/10.1007/978-90-481-2666-8\\_1](https://doi.org/10.1007/978-90-481-2666-8_1) 16
- [39] E. Triantaphyllou. 2013. *Multi-Criteria Decision Making Methods: a Comparative Study*. Springer Science & Business Media. [https://doi.org/10.1007/978-1-4757-3157-6\\_2](https://doi.org/10.1007/978-1-4757-3157-6_2) 1, 17
- [40] E. Wall, A. Narechania, A. Coscia, J. Paden, and A. Endert. 2021. Lumos: Increasing Awareness of Analytic Behavior During Visual Data Analysis. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 1009–1018. <https://doi.org/10.1109/tvcg.2021.3114827> 34, 45, 46

- [41] E. Wall, A. Narechania, A. Coscia, J. Paden, and A. Endert. 2022. Left, Right, and Gender: Exploring Interaction Traces to Mitigate Human Biases. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 966–975. <https://doi.org/10.1109/TVCG.2021.3114862> 34, 121
- [42] F. Alt, S. Torma, and D. Buschek. 2016. Don't Disturb Me: Understanding Secondary Tasks on Public Displays. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays (PerDis 2016), June 20-22, 2016, Oulu, Finland*. ACM, 1–12. <https://doi.org/10.1145/2914920.2915023> 121
- [43] F. Gutiérrez, N. N. Htun, F. Schlenz, A. Kasimati, and K. Verbert. 2019. A Review of Visualisations in Agricultural Decision Support Systems: An HCI Perspective. *Computers and Electronics in Agriculture* 163 (2019). <https://doi.org/10.1016/j.compag.2019.05.053> 16
- [44] G. Gésan-Guiziou, A. Alaphilippe, J. Aubin, C. Bockstaller, R. Boutrou, P. Buche, C. Collet, A. Girard, V. Martinet, J.-M. Membré, R. Sabbadin, M. Thiollet-Scholtus, and H. van Der Werf. 2020. Diversity and Potentiality of Multi-Criteria Decision Analysis Methods for Agri-Food Research. *Agronomy for Sustainable Development* 40, 6 (2020), 44. <https://doi.org/10.1007/s13593-020-00650-3> 17
- [45] G. Klein, B. M. Moon, and R. R. Hoffman. 2006. Making Sense of Sensemaking 1: Alternative Perspectives. *IEEE Intelligent Systems* 21, 4 (2006), 70–73. <https://doi.org/10.1109/MIS.2006.75> 12
- [46] G. Klein, B. Moon, and R. R. Hoffman. 2006. Making Sense of Sensemaking 2: A Macrocognitive Model. *IEEE Intelligent Systems* 21, 5 (2006), 88–92. <https://doi.org/10.1109/MIS.2006.100> 12
- [47] G. Klein, J. K. Phillips, E. Rall, and D. A. Peluso. 2007. A Data-frame Theory of Sensemaking. In *Expertise out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*. 113–155. <https://doi.org/10.4324/9780203810088> 12
- [48] H. B. Hutchinson, W. E. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, and B. Eiderbäck. 2003. Technology Probes: Inspiring Design for and with Families. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems (CHI 2003), Ft. Lauderdale, Florida, USA*. ACM, 17–24. <https://doi.org/10.1145/642611.642616> 3, 7, 41
- [49] H. Chernoff. 1973. The Use of Faces to Represent Points in K-Dimensional Space Graphically. *J. Amer. Statist. Assoc.* 68, 342 (1973), 361–368. <https://doi.org/10.1080/01621459.1973.10482434> 24

- [50] H. L. Trinkaus and T. Hanne. 2005. Knowcube: a Visual and Interactive Support for Multicriteria Decision Making. *Computers & Operations Research* 32 (2005), 1289–1309. <https://doi.org/10.1016/j.cor.2003.11.010> 17, 32, 44
- [51] H.-S. Shih, H.-J. Shyr, and E. S. Lee. 2007. An Extension of TOPSIS for Group Decision Making. *Mathematical and Computer Modelling* 45, 7-8 (2007), 801–813. <https://doi.org/10.1016/j.mcm.2006.03.023> 31, 88
- [52] H. T. Kung, F. Luccio, and F. P. Preparata. 1975. On Finding the Maxima of a Set of Vectors. *J. ACM* 22, 4 (1975), 469–476. <https://doi.org/10.1145/321906.321910> 19
- [53] I. Gratton. 2015. Wine Fermenting, Konoba Toni, Dol. <https://www.flickr.com/photos/60911558atN04/21963547533>. Accessed: 2024-10-02. 2
- [54] J. Bautista and G. Carenini. 2008. An Empirical Evaluation of Interactive Visualizations for Preferential Choice. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI 2008), May 28-30, 2008, Napoli, Italy*. ACM Press, 207–214. <https://doi.org/10.1145/1385569.1385603> 17, 26
- [55] J. C. Roberts. 2007. State of the Art: Coordinated & Multiple Views in Exploratory Visualization. In *Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*. IEEE, 61–71. <https://doi.org/10.1109/cm.2007.20> 121
- [56] J. C. Trapp and H.-G. Pagendam. 1996. Data Level Comparative Visualization in Aircraft Design. In *Proceedings of the 7th IEEE Visualization Conference (VIS 1996), October 27 - November 1, 1996, San Francisco, CA, USA*. IEEE Computer Society and ACM, 393–396. <https://doi.org/10.1109/VISUAL.1996.568140> 37
- [57] J. Daniels, P. W. Werner, and A. T. Bahill. 2001. Quantitative Methods for Tradeoff Analyses: Regular Papers. *Systems Engineering* 4, 3 (2001), 190–212. 31
- [58] J. Fuchs, P. Isenberg, A. Bezerianos, and D. A. Keim. 2017. A Systematic Review of Experimental Studies on Data Glyphs. *IEEE Transactions on Visualization and Computer Graphics* 23, 7 (2017), 1863–1879. <https://doi.org/10.1109/TVCG.2016.2549018> 24
- [59] J. Hakanen, K. Miettinen, and K. Matkovic. 2021. Task-Based Visual Analytics for Interactive Multiobjective Optimization. *Journal of the Operational Research Society* 72, 9 (2021), 20732090. <https://doi.org/10.1080/01605682.2020.1768809> 2, 4, 8, 20, 35, 37, 41, 44, 54, 65, 67, 68
- [60] J. Heer and B. Shneiderman. 2012. Interactive Dynamics for Visual Analysis. *Commun. ACM* 55, 4 (2012), 45–54. <https://doi.org/10.1145/2133806.2133821> 37

- [61] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. 2008. Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1189–1196. <https://doi.org/10.1109/TVCG.2008.137> 33, 121
- [62] J. Heinrich and D. Weiskopf. 2013. State of the Art of Parallel Coordinates. In *Proceedings of the 34th Annual Conference of the European Association for Computer Graphics (Eurographics 2013), Girona, Spain*. Eurographics Association, 95–116. <https://doi.org/10.2312/CONF/EG2013/STARS/095-116> 25
- [63] J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay. 2019. In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 903–913. <https://doi.org/10.1109/TVCG.2018.2864889> 37
- [64] J. J. Thomas and K. A. Cook. 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center. 11, 12
- [65] J. Johansson and C. Forsell. 2016. Evaluation of Parallel Coordinates: Overview, Categorization and Guidelines for Future Research. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 579–588. <https://doi.org/10.1109/TVCG.2015.2466992> 25
- [66] N. Loiseau D. Mouillot J. Langlois, F. Guilhaumon, F. Baletaud, N. Casajus, C. De Almeida Braga, V. Fleuré, M. Kulbicki and J. P. Renoult. 2022. The Aesthetic Value of Reef Fishes Is Globally Mismatched to Their Conservation Priorities. *Public Library of Science Biology* 20, 6 (2022), e3001640. 124
- [67] J. Liu, N. Boukhelifa, and J. R. Eagan. 2020. Understanding the Role of Alternatives in Data Analysis Practices. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 66–76. <https://doi.org/10.1109/TVCG.2019.2934593> 36
- [68] J. Rubin and D. Chisnell. 2011. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. John Wiley & Sons. 7
- [69] J. S. Evans-Cowley and G. Akar. 2014. Streetseen Visual Survey Tool for Determining Factors That Make a Street Attractive for Bicycling. *Transportation Research Record* 2468, 1 (2014), 19–27. <https://doi.org/10.3141/2468-03> 124
- [70] J. S. Yi, Y. A. Kang, J. T. Stasko, and J. A. Jacko. 2007. Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1224–1231. <https://doi.org/10.1109/TVCG.2007.70515> 37

- [71] J. Sorger, T. Ortner, C. Luksch, M. Schwärzler, M. E. Gröller, and H. Piringer. 2016. LiteVis: Integrated Visualization for Simulation-Based Decision Support in Lighting Design. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 290–299. <https://doi.org/10.1109/TVCG.2015.2468011> 27, 28, 30
- [72] J. T. Stasko, C. Görg, Z. Liu, and K. Singhal. 2007. Jigsaw: Supporting Investigative Analysis Through Interactive Visualization. In *Proceedings of the 2nd IEEE Symposium on Visual Analytics Science and Technology (VAST 2007), October 30 - November 1, 2007, Sacramento, CA, USA*. IEEE Computer Society, 131–138. <https://doi.org/10.1109/VAST.2007.4389006> 14
- [73] J. Talbot, V. Setlur, and A. Anand. 2014. Four Experiments on the Perception of Bar Charts. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2152–2160. <https://doi.org/10.1109/TVCG.2014.2346320> 26
- [74] J. W. Emerson, W. A. Green, B. Schloerke, J. Crowley, D. Cook, H. Hofmann, and H. Wickham. 2012. The Generalized Pairs Plot. *Journal of Computational and Graphical Statistics* 22, 1 (2012), 79–91. <https://doi.org/10.1080/10618600.2012.694762> 25
- [75] J. Wenskovitch, I. Crandell, N. Ramakrishnan, L. House, S. Leman, and C. North. 2018. Towards a Systematic Combination of Dimension Reduction and Clustering in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 131–141. <https://doi.org/10.1109/TVCG.2017.2745258> 22
- [76] J. Zhao, M. Glueck, P. Isenberg, F. Chevalier, and A. Khan. 2018. Supporting Handoff in Asynchronous Collaborative Sensemaking Using Knowledge-Transfer Graphs. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 340–350. <https://doi.org/10.1109/TVCG.2017.2745279> 33, 126
- [77] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. 2000. A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II. In *Proceedings of the 6th International Conference on Parallel Problem Solving from Nature (PPSN VI)*. Springer, 849–858. [https://doi.org/10.1007/3-540-45356-3\\_83](https://doi.org/10.1007/3-540-45356-3_83) 19
- [78] K. Eberhard. 2023. The Effects of Visualization on Judgment and Decision-Making: a Systematic Literature Review. *Management Review Quarterly* 73, 1 (2023), 167–214. <https://doi.org/10.1007/s11301-021-00235-8> 2
- [79] K. Hussinger and W. B. Issah. 2019. Firm Acquisitions by Family Firms: A Mixed Gamble Approach. *SSRN Electronic Journal* (2019). <https://doi.org/10.2139/ssrn.3476535> 124
- [80] K. Madanagopal, E. D. Ragan, and P. Benjamin. 2019. Analytic Provenance in Practice: The Role of Provenance in Real-World Visualization and Data Analysis Environments. *IEEE Computer Graphics and Applications* 39, 6 (2019), 30–45. <https://doi.org/10.1109/mcg.2019.2933419> 2, 3, 41



- [81] K. Matkovic, D. Gracanin, R. Splechna, M. Jelovic, B. Stehno, H. Hauser, and W. Purghofer. 2014. Visual Analytics for Complex Engineering Systems: Hybrid Visual Steering of Simulation Ensembles. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1803–1812. <https://doi.org/10.1109/TVCG.2014.2346744> 18
- [82] K. Miettinen. 2014. Survey of Methods to Visualize Alternatives in Multiple Criteria Decision Making Problems. *OR Spectrum* 36, 1 (2014), 3–37. <https://doi.org/10.1007/s00291-012-0297-0> 27, 65
- [83] K. Pearson. 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2, 11 (1901), 559–572. <https://doi.org/10.1080/14786440109462720> 23
- [84] K. Xu, A. Ottley, C. Walchshofer, M. Streit, R. Chang, and J. E. Wenskovitch. 2020. Survey on the Analysis of User Interactions and Visualization Provenance. *Computer Graphics Forum* 39, 3 (2020), 757–783. <https://doi.org/10.1111/CGF.14035> 2, 8, 33, 41, 66, 69
- [85] K. Xu, S. Attfield, T. J. Jankun-Kelly, A. Wheat, P. H. Nguyen, and N. Selvaraj. 2015. Analytic Provenance for Sensemaking: A Research Agenda. *IEEE Computer Graphics and Applications* 35, 3 (2015), 56–64. <https://doi.org/10.1109/MCG.2015.50> 2, 32
- [86] L. Cibulski, H. Mitterhofer, T. May, and J. Kohlhammer. 2020. PAVED: Pareto Front Visualization for Engineering Design. *Computer Graphics Forum* 39, 3 (2020), 405–416. <https://doi.org/10.1111/cgf.13990> 30, 32, 37
- [87] L. E. Matzen, M. J. Haass, K. M. Divis, Z. Wang, and A. T. Wilson. 2018. Data Visualization Saliency Model: A Tool for Evaluating Abstract Data Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 563–573. <https://doi.org/10.1109/TVCG.2017.2743939> 120
- [88] L. M. Hvattum and H. Arntzen. 2010. Using ELO Ratings for Match Result Prediction in Association Football. *International Journal of Forecasting* 26 (2010), 460–470. <https://doi.org/10.1016/j.ijforecast.2009.10.002> 124
- [89] L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. 2018. Decision Making with Visualizations: a Cognitive Framework Across Disciplines. *Cognitive Research: Principles and Implications* 3 (2018), 125. <https://doi.org/10.1186/s41235-018-0120-9> 2
- [90] L. Wilkinson, A. Anand, and R. Grossman. 2005. Graph-Theoretic Scagnostics. In *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS 2005)*. 157–164. <https://doi.org/10.1109/INFVIS.2005.1532142> 23, 25

- [91] M. A. Yalçın, N. Elmqvist, and B. B. Bederson. 2017. Raising the Bars: Evaluating Treemaps vs. Wrapped Bars for Dense Visualization of Sorted Numeric Data. In *Proceedings of the 43rd Graphics Interface Conference*. Canadian Human-Computer Communications Society, 41–49. <https://doi.org/10.20380/GI2017.06> 26
- [92] M. Bagajewicz and E. Cabrera. 2003. Pareto Optimal Solutions Visualization Techniques for Multiobjective Design and Upgrade of Instrumentation Networks. *Industrial & Engineering Chemistry Research* 42, 21 (2003), 5195–5203. <https://doi.org/10.1021/ie020865g> 27
- [93] M. Booshehrian, T. Möller, and T. Munzner. 2010. Vismon Screenshot. <https://vismon.org/index.html> Accessed: 2024-10-10. 19
- [94] M. Booshehrian, T. Möller, R. M. Peterman, and T. Munzner. 2012. Vismon: Facilitating Analysis of Trade-Offs, Uncertainty, and Sensitivity in Fisheries Management Decision Making. *Computer Graphics Forum* 31, 3pt3 (2012), 1235–1244. <https://doi.org/10.1111/j.1467-8659.2012.03116.x> 18, 19
- [95] M. Brehmer and T. Munzner. 2013. A Multi-Level Typology of Abstract Visualization Tasks. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2376–2385. <https://doi.org/10.1109/TVCG.2013.124> 4
- [96] M. Chakhchoukh. 2024. Comparison Project Supplementary Material. [https://osf.io/h3zdr/?view\\_only=7c66f9f17e8b4f91b8cd25224b87d83f](https://osf.io/h3zdr/?view_only=7c66f9f17e8b4f91b8cd25224b87d83f). Accessed: 2024-12-31. 74, 82, 93, 101, 102
- [97] M. Chakhchoukh, N. Boukhelifa, and A. Bezerianos. 2023. Understanding How In-Visualization Provenance Can Support Trade-Off Analysis. *IEEE Transactions on Visualization and Computer Graphics* 29, 9 (2023), 3758–3774. <https://doi.org/10.1109/TVCG.2022.3171074> 4, 35, 41, 74, 81, 82, 115, 119
- [98] M. Chakhchoukh, N. Boukhelifa, and A. Bezerianos. 2023. Visualizations to Support Trade-Off Comparisons. *Journée Visu* (2023), 3. 74
- [99] M. Espadoto, R. M. Martins, A. Kerren, N. S. T. Hirata, and A. C. Telea. 2021. Toward a Quantitative Survey of Dimension Reduction Techniques. *IEEE Transactions on Visualization and Computer Graphics* 27, 3 (2021), 2153–2173. <https://doi.org/10.1109/TVCG.2019.2944182> 22
- [100] M. Feng, C. Deng, E. M. Peck, and L. Harrison. 2017. HindSight: Encouraging Exploration Through Direct Encoding of Personal Interaction History. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 351–360. <https://doi.org/10.1109/TVCG.2016.2599058> 33, 34
- [101] M. Gleicher. 2018. Considerations for Visualizing Comparison. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 413–423. <https://doi.org/10.1109/TVCG.2017.2744199> 4, 36, 82, 84, 112, 123

- [102] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. 2011. Visual Comparison for Information Visualization. *Information Visualization* 10, 4 (2011), 289–309. <https://doi.org/10.1177/1473871611416549> 4, 5, 36, 84
- [103] M. H. Loorak, M. Tory, and S. Carpendale. 2018. ChangeCatcher: Increasing Inter-Author Awareness for Visualization Development. *Computer Graphics Forum* 37, 3 (2018), 5162. <https://doi.org/10.1111/CGF.13400> 68
- [104] M. Luboschik, M. Röhlig, G. Kundt, O. Stachs, S. Peschel, A. Zhivov, R. F. Guthoff, K. Winter, and H. Schumann. 2014. Supporting an Early Detection of Diabetic Neuropathy by Visual Analytics. In *Proceedings of the 5th International EuroVis Workshop on Visual Analytics (EuroVA 2014)*, Swansea, UK. <https://doi.org/10.2312/EUROVA.20141145> 26
- [105] M. Nauta, J. Trienes, S. Pathak, E. Nguyen, M. Peters, Y. Schmitt, J. Schlötterer, M. van Keulen, and C. Seifert. 2022. From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable Ai. *Computing Research Repository* abs/2201.08164 (2022). <https://doi.org/10.1145/3583558> 38, 99, 100, 108, 123
- [106] M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer, and T. Möller. 2014. Visual Parameter Space Analysis: A Conceptual Framework. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2161–2170. <https://doi.org/10.1109/TVCG.2014.2346321> 1, 18
- [107] M. Sedlmair, M. Brehmer, S. Ingram, and T. Munzner. 2012. Dimensionality Reduction in the Wild: Gaps and Guidance. *Department of Computer Science, University of British Columbia, Vancouver, BC, Canada* (2012). <https://www.vis.uni-stuttgart.de/documentcenter/staff/sedlmaml/papers/sedlmair2013infovis.pdf> 22
- [108] M. Trolove. 2010. Top Dressing Winter Wheat. <https://commons.wikimedia.org/w/index.php?curid=14477239>. Accessed: 2024-10-02. 2
- [109] M. Unal, G. P. Warn, and T. W. Simpson. 2016. Quantifying Tradeoffs to Reduce the Dimensionality of Complex Design Optimization Problems and Expedite Trade Space Exploration. *Structural and Multidisciplinary Optimization* 54, 2 (2016), 233–248. <https://doi.org/10.1007/s00158-015-1389-7> 25, 98
- [110] M. Wattenberg and J. Kriss. 2006. Designing for Social Data Analysis. *IEEE Transactions on Visualization and Computer Graphics* 12, 4 (2006), 549–557. <https://doi.org/10.1109/TVCG.2006.65> 33
- [111] Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche. 2016. Protocole d'accord sur l'utilisation et la reproduction des livres, des œuvres musicales éditées, des publications périodiques et des œuvres des arts visuels à des fins d'illustration des activités d'enseignement et de recherche. <https://www.education.gouv.fr/bo/16/Hebdo35/MENE1600684X.htm> Accessed: 2024-12-01. 11

- [112] N. Boukhelifa, A. Bezerianos, I. C. Trelea, N. Perrot, and E. Lutton. 2019. An Exploratory Study on Visual Exploration of Model Simulations by Multiple Types of Experts. *Computing Research Repository* abs/1902.01721 (2019). <https://doi.org/10.1145/3290605.3300874> 1, 2, 3, 4, 15, 17, 26, 28, 30, 32, 35, 42, 43, 44, 45, 47, 62, 74, 119, 125, 126
- [113] N. Boukhelifa, W. Cancino Ticona, A. Bezerianos, and E. Lutton. 2013. Evolutionary Visual Exploration: Evaluation With Expert Users. *Computer Graphics Forum* 32, 3 (2013), 31–40. <https://doi.org/10.1111/cgf.12090> 27, 42, 43, 50, 82
- [114] N. Brisson, B. Mary, D. Ripoché, M.-H. Jeuffroy, F. Ruget, B. Nicoulaud, P. Gate, F. Devienne-Barret, R. Antonioletti, C. Dürr, and others. 1998. STICS: a Generic Model for the Simulation of Crops and Their Water and Nitrogen Balances. I. Theory and Parameterization Applied to Wheat and Corn. *Agronomie* 18, 5-6 (1998), 311346. 1, 16
- [115] N. Elmqvist, P. Dragicevic, and J. Fekete. 2008. Rolling the Dice: Multidimensional Visual Exploration Using Scatterplot Matrix Navigation. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1539–1148. <https://doi.org/10.1109/tvcg.2008.153> 24, 25, 26, 27, 42, 47
- [116] N. Kadivar, V. Y. Chen, D. T. Dunsmuir, E. Lee, C. Z. Qian, J. Dill, C. D. Shaw, and R. F. Woodbury. 2009. Capturing and Supporting the Analysis Process. In *Proceedings of the 4th IEEE Symposium on Visual Analytics Science and Technology (VAST 2009), October 11-16, 2009, Atlantic City, NJ, USA*. <https://doi.org/10.1109/VAST.2009.5333020> 15, 121
- [117] P. Dragicevic. 2016. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*. Springer, 291–330. [https://doi.org/10.1007/978-3-319-26633-6\\_13](https://doi.org/10.1007/978-3-319-26633-6_13) 102
- [118] P. H. Nguyen, K. Xu, A. Bardill, B. Salman, K. Herd, and B. L. W. Wong. 2016. SenseMap: Supporting Browser-Based Online Sensemaking Through Analytic Provenance. In *Proceedings of the 11th IEEE Conference on Visual Analytics Science and Technology (VAST 2016), October 23-28, 2016, Baltimore, MD, USA*. 91–100. <https://doi.org/10.1109/VAST.2016.7883515> 33
- [119] P. Isenberg, N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen. 2011. Collaborative Visualization: Definition, Challenges, and Research Agenda. *Information Visualization* 10, 4 (2011), 310–326. <https://doi.org/10.1177/1473871611412817> 126
- [120] P.-M. Law, A. Endert, and J. T. Stasko. 2020. Characterizing Automated Data Insights. *Computing Research Repository* abs/2008.13060 (2020). <https://doi.org/10.1109/vis47514.2020.00041> 66, 120

- [121] P. Pirolli and S. Card. 2005. The Sensemaking Process and Leverage Points for Analyst Technology As Identified Through Cognitive Task Analysis. In *Proceedings of the International Conference on Intelligence Analysis*. 2–4. 12, 13, 36
- [122] P. Riehmman, J. Opolka, and B. Froehlich. 2012. The Product Explorer: Decision Making with Ease. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI 2012), May 22-25, 2012, Capri Island, Naples, Italy*. ACM, 423–432. <https://doi.org/10.1145/2254556.2254638> 26
- [123] P.T. Eendebak, A.R. Vazquez. 2019. OApkg: A Python Package for Generation and Analysis of Orthogonal Arrays, Optimal Designs and Conference Designs, *Journal of Open Source Software*. [https://oapackage.readthedocs.io/en/latest/examples/example\\_pareto.html](https://oapackage.readthedocs.io/en/latest/examples/example_pareto.html) Accessed: 2024-01-01. 80
- [124] R. A. Rensink and G. Baldridge. 2010. The Perception of Correlation in Scatterplots. *Computer Graphics Forum* 29, 3 (2010), 1203–1210. <https://doi.org/10.1111/j.1467-8659.2009.01694.x> 42
- [125] R. Amar, J. Eagan, and J. Stasko. 2005. Low-Level Components of Analytic Activity in Information Visualization. In *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS 2005)*. 111–117. <https://doi.org/10.1109/INFVIS.2005.1532136> 4, 31, 65
- [126] R. Borgo, J. Kehrer, D. H. S. Chung, E. Maguire, R. S. Laramée, H. Hauser, M. Ward, and M. Chen. 2013. Glyph-Based Visualization: Foundations, Design Guidelines, Techniques and Applications. In *Proceedings of the 34th Annual Conference of the European Association for Computer Graphics (Eurographics 2013), Girona, Spain*. Eurographics Association, 39–63. <https://doi.org/10.2312/CONF/EG2013/STARS/039-063> 24
- [127] R. Rao and S. K. Card. 1994. The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+context Visualization for Tabular Information. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems (CHI 1994), Boston, Massachusetts, USA*. ACM, 222. <https://doi.org/10.1145/259963.260391> 23
- [128] S. Carpendale. 2008. Evaluating Information Visualizations. In *Information Visualization - Human-Centered Issues and Perspectives*, J.-D. Fekete A. Kerren, J. T. Stasko and C. North (Eds.). *Lecture Notes in Computer Science*, Vol. 4950. Springer, 19–45. [https://doi.org/10.1007/978-3-540-70956-5\\_2](https://doi.org/10.1007/978-3-540-70956-5_2) 38
- [129] S. Chen, D. Amid, O. M. Shir, L. Limonad, D. Boaz, A. Anaby-Tavor, and T. Schreck. 2013. Self-Organizing Maps for Multi-Objective Pareto Frontiers. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis 2013), February 27 2013-March 1, 2013, Sydney, NSW, Australia*. <https://doi.org/10.1109/PacificVis.2013.6596140> 27

- [130] S. Few. 2016. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press. 26
- [131] S. G. Hart and L. E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, P. A. Hancock and N. Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9) 99, 101, 123
- [132] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. 2013. LineUp: Visual Analysis of Multi-Attribute Rankings. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2277–2286. <https://doi.org/10.1109/TVCG.2013.173> 5, 19, 21, 26, 29, 31, 37, 87, 89, 96, 113
- [133] S. Ingram, T. Munzner, and M. Olano. 2009. Glimmer: Multilevel MDS on the GPU. *IEEE Transactions on Visualization and Computer Graphics* 15, 2 (2009), 249–261. <https://doi.org/10.1109/TVCG.2008.85> 23
- [134] S. Liu, D. Maljovec, B. Wang, P.-T. Bremer, and V. Pascucci. 2017. Visualizing High-Dimensional Data: Advances in the Past Decade. *IEEE Transactions on Visualization and Computer Graphics* 23, 3 (2017), 1249–1268. <https://doi.org/10.1109/TVCG.2016.2640960> 22, 23, 25
- [135] S. Pajer, M. Streit, T. Torsney-Weir, F. Spechtenhauser, T. Möller, and H. Piringer. 2017. Weightlifter: Visual Weight Space Exploration for Multi-Criteria Decision Making. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 611–620. <https://doi.org/10.1109/TVCG.2016.2598589> 5, 20, 22, 26, 29, 30, 37, 44, 87
- [136] S. T. Dumais, R. Jeffries, D. M. Russell, D. Tang, and J. Teevan. 2014. Understanding User Behavior Through Log Data and Analysis. In *Ways of Knowing in HCI*, J. S. Olson and W. A. Kellogg (Eds.). Springer, 349–372. [https://doi.org/10.1007/978-1-4939-0378-8\\_14](https://doi.org/10.1007/978-1-4939-0378-8_14) 15
- [137] S. Wehrend and C. Lewis. 1990. A Problem-Oriented Classification of Visualization Techniques. In *Proceedings of the 1st IEEE Visualization Conference (VIS 1990), October 23-26, 1990, San Francisco, CA, USA*. IEEE Computer Society Press, 139–143. <https://doi.org/10.1109/VISUAL.1990.146375> 4
- [138] T. Grossman, J. Matejka, and G. W. Fitzmaurice. 2010. Chronicle: Capture, Exploration, and Playback of Document Workflow Histories. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology, October 3-6, 2010, New York, NY, USA*. ACM, 143–152. <https://doi.org/10.1145/1866029.1866054> 121
- [139] T. Miller. 2017. Explanation in Artificial Intelligence: Insights From the Social Sciences. *Computing Research Repository* abs/1706.07269 (2017). <https://doi.org/10.1016/j.artint.2018.07.007> 38

- [140] T. Tušar and B. Filipič. 2015. Visualization of Pareto Front Approximations in Evolutionary Multiobjective Optimization: A Critical Review and the Prosection Method. *IEEE Transactions on Visualization and Computer Graphics* 19, 2 (2015), 225–245. <https://doi.org/10.1109/tevc.2014.2313407> 27, 30, 42, 65
- [141] T. Zhang, F. Ladhak, E. Durmus, P. Liang, K. R. McKeown, and T. B. Hashimoto. 2023. Benchmarking Large Language Models for News Summarization. *Computing Research Repository* abs/2301.13848 (2023). [https://doi.org/10.1162/tacl\\_a\\_00632](https://doi.org/10.1162/tacl_a_00632) 126
- [142] V. Braun and V. Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. <https://doi.org/10.1191/1478088706qp063oa> 51, 122
- [143] V. Braun, V. Clarke, N. Hayfield, and G. Terry. 2019. *Thematic Analysis*. Springer Singapore, 843–860. [https://doi.org/10.1007/978-981-10-5251-4\\_103](https://doi.org/10.1007/978-981-10-5251-4_103) 122
- [144] V. Clarke and V. Braun. 2021. *Thematic Analysis: a Practical Guide*. SAGE Publications Ltd. <https://doi.org/10.1177/1035719X211058251> 82, 102, 122
- [145] W. Cancino, N. Boukhelifa, and E. Lutton. 2012. EvoGraphDice: Interactive evolution for visual analytics. In *Proceedings of the 2012 IEEE Congress on Evolutionary Computation*. 1–8. <https://doi.org/10.1109/CEC.2012.6256553> 27, 28
- [146] W. S. Cleveland and R. McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79 (1984), 531–554. <https://doi.org/10.1080/01621459.1984.10478080> 26
- [147] W. Willett, J. Heer, and M. Agrawala. 2007. Scented Widgets: Improving Navigation Cues with Embedded Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1129–1136. <https://doi.org/10.1109/TVCG.2007.7058934>
- [148] W. Willett, P. Goffin, and P. Isenberg. 2015. Understanding Digital Note-Taking Practice for Visualization. *IEEE Computer Graphics and Applications* 35, 4 (2015), 38–51. <https://doi.org/10.1109/MCG.2015.52> 126
- [149] Wikipedia username: Njr00. 2014. A Pareto Optimal Front, Illustrated Using a Production-Possibility Frontier. [https://commons.wikimedia.org/wiki/File:Pareto\\_Efficient\\_Frontier\\_1024x1024.png](https://commons.wikimedia.org/wiki/File:Pareto_Efficient_Frontier_1024x1024.png). Accessed: 2024-10-02. 20
- [150] X. Chen, W. Zeng, Y. Lin, H. M. Al-Maneea, J. C. Roberts, and R. Chang. 2020. Composition and Configuration Patterns in Multiple-View Visualizations. *Computing Research Repository* 27, 2 (2020), 1514–1524. <https://doi.org/10.1109/tvcg.2020.3030338> 121

- [151] X. Kuang, H. Zhang, S. Zhao, and M. J. McGuffin. 2012. Tracing Tuples Across Dimensions: A Comparison of Scatterplots and Parallel Coordinate Plots. *Computer Graphics Forum* 31, 3pt4 (2012), 1365–1374. <https://doi.org/10.1111/j.1467-8659.2012.03129.x> 26, 42
- [152] X. Zhao, Y. Wu, W. Cui, X. Du, Y. Chen, Y. Wang, D. L. Lee, and H. Qu. 2017. SkyLens: Visual Analysis of Skyline on Multi-Dimensional Data. *Computing Research Repository* 24, 1 (2017), 246–255. <https://doi.org/10.1109/tvcg.2017.2744738> 30
- [153] Y. B. Shrinivasan and J. J. van Wijk. 2009. Supporting Exploration Awareness in Information Visualization. *IEEE Computer Graphics and Applications* 29, 5 (2009), 34–43. <https://doi.org/10.1109/MCG.2009.87> 69
- [154] Y. Park, M. Cafarella, and B. Mozafari. 2016. Visualization-Aware Sampling for Very Large Databases. In *Proceedings of the 2016 IEEE 32nd International Conference on Data Engineering (ICDE)*. 755–766. <https://doi.org/10.1109/icde.2016.7498287> 48
- [155] Z. Cui, F. Zhang, X. Chen, Z. Dou, and J. Li. 2010. In-Season Nitrogen Management Strategy for Winter Wheat: Maximizing Yields, Minimizing Environmental Impact in an Over-Fertilization Context. *Field Crops Research* 116, 1 (2010), 140–146. <https://doi.org/10.1016/j.fcr.2009.12.004> 17
- [156] Z. Liu and J. T. Stasko. 2010. Mental Models, Visual Reasoning and Interaction in Information Visualization: A Top-Down Perspective. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 999–1008. <https://doi.org/10.1109/TVCG.2010.177> 12, 13
- [157] Z. Liu, Y. Wang, M. Dontcheva, M. Hoffman, S. Walker, and A. Wilson. 2017. Patterns and Sequences: Interactive Exploration of Clickstreams to Understand Common Visitor Paths. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 321–330. <https://doi.org/10.1109/TVCG.2016.2598797> 68



APPENDIX



# Supplementary Material for Chapter 5

# Detailed Statistical Results - Chapter 5

## Visualizations to support trade-off comparisons between groups

This document presents detailed inference results (CI), organized in relation to the sections of the paper. We report CIs as *mean [lowerBound : upperBound]*

Statistical analysis for mean comparison was conducted using interval estimation. We report sample means and 95% confidence intervals (CIs), which means we are 95% confident that this interval includes the population mean. All CIs were constructed using BCa bootstrapping (10,000 bootstrap iterations). We additionally report the CIs of *mean differences* to compare different conditions.

When interpreting results, a CI of a mean *difference* that does not overlap with 0 provides evidence of a difference, corresponding to statistically significant results in traditional p-value tests. Equivalent p-values can be obtained from CI results following [Krzywinski and Altma, 2013]. Nonetheless, CIs allow for more subtle interpretations: the farther from 0 and the tighter the CI is, the stronger the evidence.

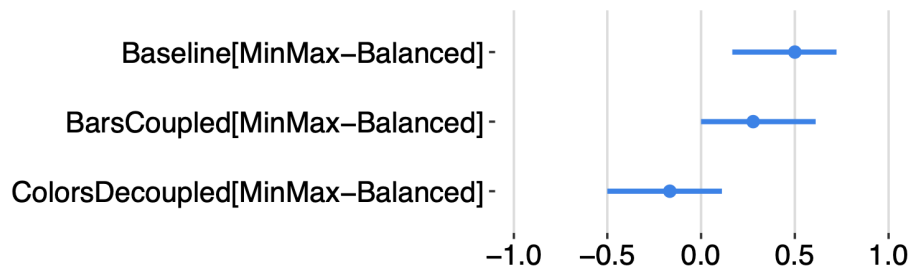
Analysis scripts in R can be found in [supp.material](#) for regenerating these graphs and results (see the link provided in Chapter 5).

## 7.1.2 Decision Strategy

We compare the number of different decision strategies (MinMax, Balanced, Other) across visualization conditions. All CIs are normalized between [0-1], showing the mean (and their differences) of participants adopting a strategy normalized over all participants (18). As these are not balanced samples, we used bootstrap CI calculations for independent samples.

### Differences for each technique (ie tendency for one strategy more than another)

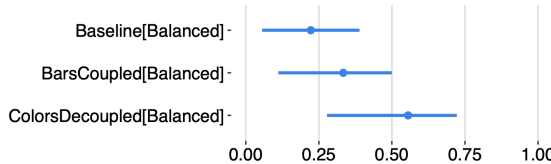
Mean differences



Baseline[MinMax-Balanced] = 0.5 [0.16 : 0.72] **strong evidence of favoring a MinMax strategy**  
 BarsCoupled[MinMax-Balanced] = 0.2 [0 : 0.61] **weak evidence to favor MinMax strategy**  
 ColorsDecoupled[MinMax-Balanced] = -0.16 [-0.5 : 0.1] no evidence of favoring a strategy over another

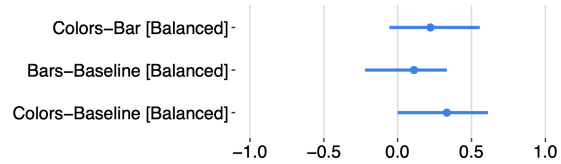
### Balanced Strategy Frequency per Visualization

Means



Baseline[Balanced] = 0.22 [0.05 : 0.38]  
 BarsCoupled[Balanced] = 0.33 [0.11 : 0.55]  
 ColorsDecoupled[Balanced] = 0.5 [0.27 : 0.72]

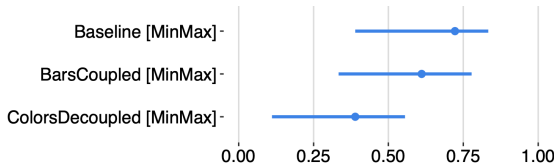
Mean differences



Colors-Bar [Balanced] = 0.2 [-0.05 : 0.5] **possible trend of diff**  
 Bars-Baseline [Balanced] = 0.1 [-0.2 : 0.3] no evidence of diff  
 Colors-Baseline [Balanced] = 0.3 [0 : 0.6] **weak evidence of diff**

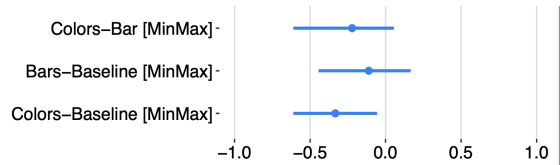
### MinMax Strategy Frequency per Visualization

Means



Baseline[MinMax] = 0.72 [0.38 : 0.83]  
 BarsCoupled [MinMax] = 0.6 [0.3 : 0.7]  
 ColorsDecoupled [MinMax] = 0.38 [0.11 : 0.55]

Mean differences



Colors-Bar [MinMax] = -0.2 [-0.6 : 0.05] **trend of diff**  
 Bars-Baseline [MinMax] = -0.1 [-0.4 : 0.16] no evidence of diff  
 Colors-Baseline [MinMax] = -0.3 [-0.6 : -0.05] **evidence of a diff**

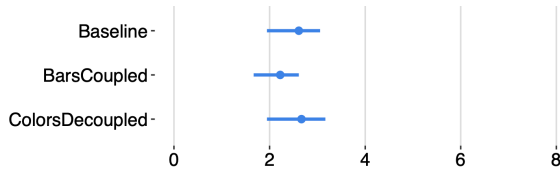
### 7.1.3 Justification Quality

We compare here the different Justification Quality measures across techniques. As these are balanced samples we use CI of mean-differences (within subjects).

#### 1. Correctness / Alignment : no test conducted as only 1 misalignment in the study

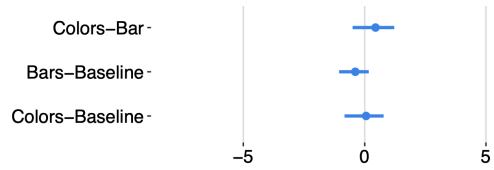
#### 2. Completeness: number of unique context codes in a response

Means



Baseline = 2.66 [1.94 : 3.16]  
 BarsCoupled = 2.22 [1.66 : 2.61]  
 ColorsDecoupled = 2.61 [1.94 : 3.05]

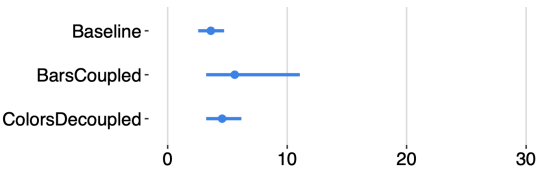
Mean differences



Colors-Bar = 0.44 [-0.5 : 1.22] no evidence of diff  
 Bars-Baseline = -0.38 [-1.05 : 0.16] no evidence of diff  
 Colors-Baseline = 0.05 [-0.83 : 0.7] no evidence of diff

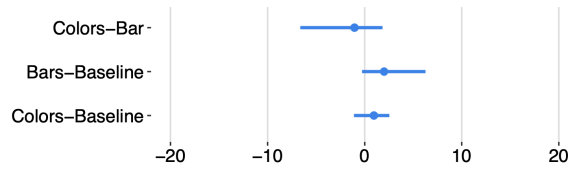
#### 3. Compactness: total number of context tags in a response

Means



Baseline = 3.6 [2.5 : 4.7]  
 BarsCoupled = 5.61 [3.22 : 11.05]  
 ColorsDecoupled = 4.55 [3.22 : 6.16]

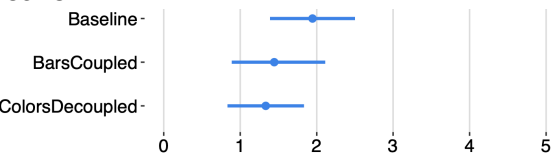
Mean differences



Colors-Bar = -1.05 [-6.63 : 1.83] no evidence of diff  
 Bars-Baseline = 2 [-0.27 : 6.27] no evidence of diff  
 Colors-Baseline = 0.94 [-1.11 : 2.55] no evidence of diff

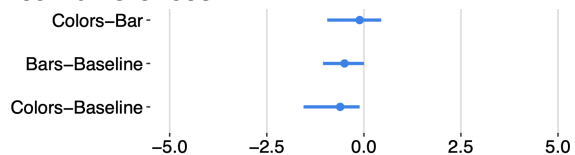
#### 4. Contrastivity: number of pairs of arguments pro group A and pro group B

Means



Baseline = 1.94 [1.38 : 2.5]  
 BarsCoupled = 1.44 [0.88 : 2.11]  
 ColorsDecoupled = 1.33 [0.83 : 1.83]

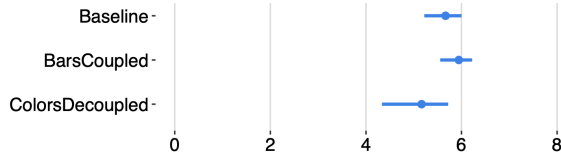
Mean differences



Colors-Bar = -0.11 [-0.94 : 0.44] no evidence of diff  
 Bars-Baseline = -0.5 [-1.05 : 0] weak evidence of diff  
 Colors-Baseline = 0.61 [-1.55 : -0.11] evidence of a diff

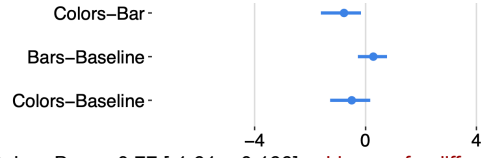
## 5. Confidence / Trust in decision (out of 7)

### Means



Baseline = 5.66 [5.22 : 6]  
 BarsCoupled = 5.93 [5.55 : 6.22]  
 ColorsDecoupled = 5.16 [4.33 : 5.72]

### Mean differences



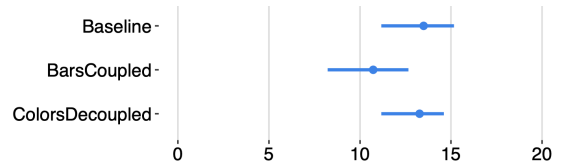
Colors-Bar = -0.77 [-1.61 : -0.166] **evidence of a diff**  
 Bars-Baseline = 0.27 [-0.27 : 0.77] no evidence of diff  
 Colors-Baseline = -0.5 [-1.27 : 0.16] no evidence of diff

## 7.2 Perceived load - NASA TLX

We compare here the different NASA TLX self-reported metrics techniques. As these are balanced samples we use CI of mean-differences (within subjects). All values out of 21.

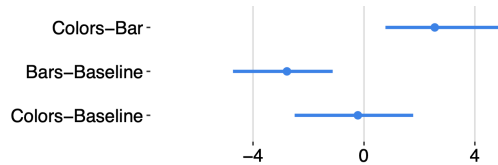
### 1. Mental Demand

#### Means



Baseline = 13.5 [11.16 : 15.16]  
 BarsCoupled = 10.72 [8.22 : 12.66]  
 ColorsDecoupled = 13.28 [11.16 : 14.61]

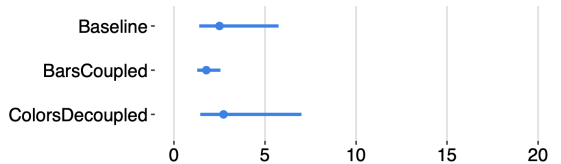
#### Mean differences



Colors-Bar = 2.55 [0.77 : 5] **evidence of a diff**  
 Bars-Baseline = -2.77 [-4.72 : -1.12] **evidence of a diff**  
 Colors-Baseline = -0.22 [-2.5 : 1.77] no evidence of diff

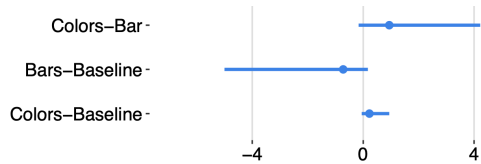
### 2. Physical Demand

#### Means



Baseline = 2.5 [1.38 : 5.74]  
 BarsCoupled = 1.77 [1.27 : 2.55]  
 ColorsDecoupled = 2.72 [1.44 : 7]

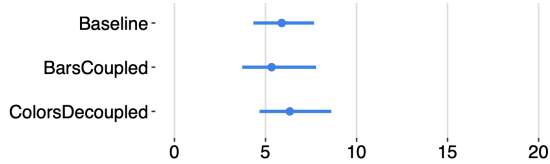
#### Mean differences



Colors-Bar = 0.94 [-0.16 : 4.22] no evidence of diff  
 Bars-Baseline = -0.72 [-5 : 0.16] no evidence of diff  
 Colors-Baseline = 0.22 [-0.05 : 0.94] **possible trend of diff**

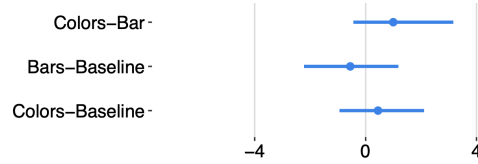
### 3. Temporal Demand

#### Means



Baseline = 5.89 [4.33 : 7.66]  
 BarsCoupled = 5.33 [3.72 : 7.77]  
 ColorsDecoupled = 6.33 [4.66 : 8.61]

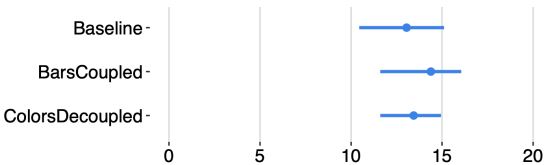
#### Mean differences



Colors-Bar = 1 [-0.44 : 3.16] no evidence of diff  
 Bars-Baseline = -0.55 [-2.22 : 1.18] no evidence of diff  
 Colors-Baseline = 0.44 [-0.94 : 2.11] no evidence of diff

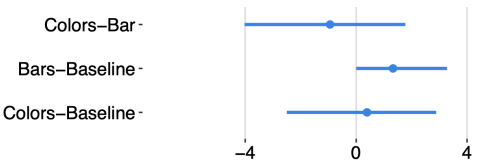
### 4. Performance

#### Means



Baseline = 13.06 [10.44 : 15.11]  
 BarsCoupled = 14.39 [11.61 : 16.05]  
 ColorsDecoupled = 13.44 [11.61 : 14.94]

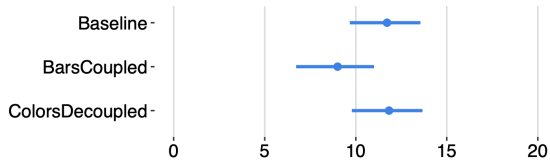
#### Mean differences



Colors-Bar = -0.94 [-4.02 : 1.77] no evidence of diff  
 Bars-Baseline = 1.33 [0 : 3.27] weak evidence of diff  
 Colors-Baseline = 0.38 [-2.5 : 2.88] no evidence of diff

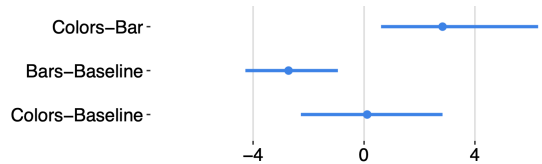
### 5. Effort

#### Means



Baseline = 11.72 [9.66 : 13.55]  
 BarsCoupled = 9 [6.72 : 11]  
 ColorsDecoupled = 11.83 [9.77 : 13.66]

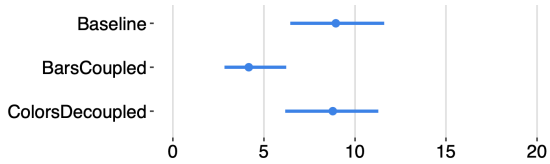
#### Mean differences



Colors-Bar = 2.83 [0.61 : 6.27] evidence of a diff  
 Bars-Baseline = -2.72 [-4.27 : -0.94] evidence of a diff  
 Colors-Baseline = 0.11 [-2.27 : 2.83] no evidence of diff

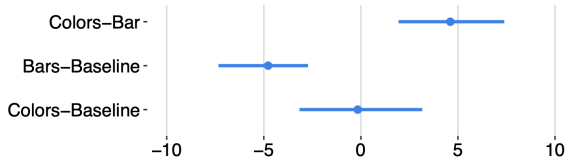
### 6. Frustration

#### Means



Baseline = 8.94 [6.44 : 11.61]  
 BarsCoupled = 4.17 [2.83 : 6.22]  
 ColorsDecoupled = 8.77 [6.16 : 11.27]

#### Mean differences



Colors-Bar = 4.61 [1.94 : 7.38] evidence of a diff  
 Bars-Baseline = -4.77 [-7.33 : -2.72] evidence of a diff  
 Colors-Baseline = -0.16 [-3.16 : 3.16] no evidence of diff