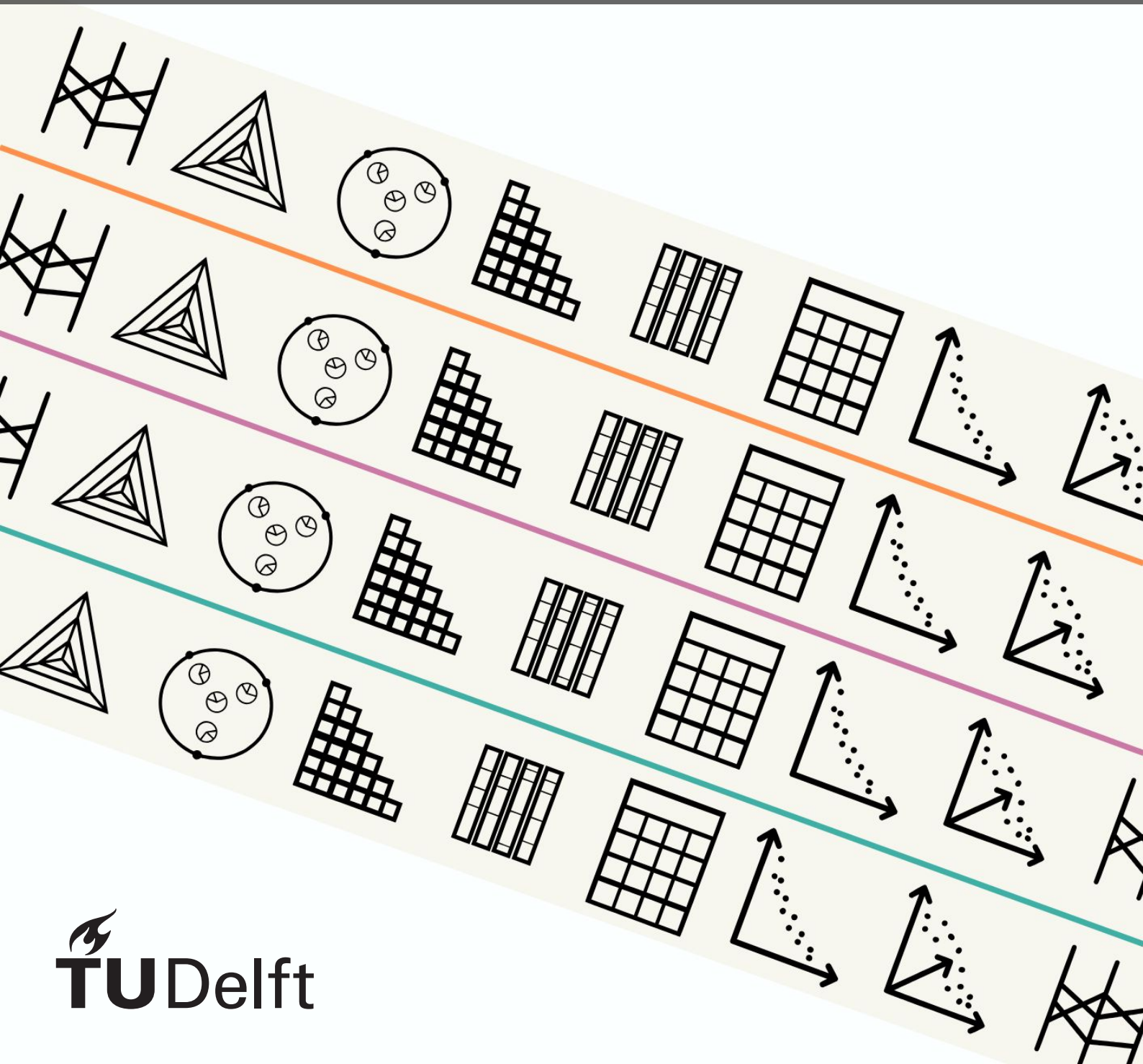
# An Analysis of Visualisation Techniques for High-Dimensional Pareto Frontiers

## A Case Study in Investment Portfolio Optimisation

Vlad-Gabriel Vrânceanu



TU Delft

# An Analysis of Visualisation Techniques for High-Dimensional Pareto Frontiers

## A Case Study in Investment Portfolio Optimisation

by

Vlad-Gabriel Vrânceanu

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on
25th of August 2026

Student number: 5467659
Project Duration: November, 2025 - August, 2026
Thesis committee: Dr. U.K. Gadiraju, TU Delft, supervisor
Dr. M. Skrodzki, TU Delft, daily supervisor
Afrasiab Kadhum, Ortec Finance, daily supervisor
Dr. Frans A. Oliehoek, TU Delft, thesis committee member

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

# Preface

*In this project, I faced the daunting task of self-managing my progress, as before the master's thesis, most of the projects that I had done were self-contained and clearly defined, never lasting more than a few months. With this project, I found it difficult to find the motivation to continue after certain milestones, as I was still in the mindset of the quarter system and kept expecting things to settle down. However, with the support of the people close to me, I always managed to get over this hurdle.*

*I would like to first thank my girlfriend, Adina, for always being there for me whenever I felt like I couldn't move forward. Thank you for always supporting me and helping me get through both my Bachelor's and Master's degrees.*

*I would like to thank my parents and my brother, Oana, George, and Teo, for always trying to understand my issues with the thesis and trying to provide comfort about them.*

*I would like to thank my supervisors, Martin and Afrasiab, for always being excited about my work and always pushing me to do better.*

*Lastly, I would like to thank all of my friends, whose frequent get-togethers helped me recharge and push through this project. Thanks to my bouldering buddies, Codrin Ogreanu, Alexandru Păcurar, Daniel Popovici, and Mihnea Bernevig, for always being up for a nice session. To everyone I often met in the Echo CS offices—Alexandru Dumitrache, Vlad Ștefănescu, Irina Marinescu, Alexandra Căruțașu, Alexandru Bolfă, Vlad Iftode, Călin Diacicov, Andra Alăzăroaie, Albert Sandu, Alexandru Ojică, and Matei Grigore—thank you for always making the study sessions enjoyable. I would also like to thank my international friend group, Konstantin Yordanov, Yiğit Çolakoğlu, Kaan Altınay, and Matej Tomášek, for all the amazing karaoke sessions where we sang like broken trumpets.*

*If I missed anyone in my thanks, you can always contact me to receive a drink as an apology. Thank you again, everyone!*

*Vlad-Gabriel Vrânceanu*
*Delft, August 2026*

# Contents

# 1

# Introduction

In the modern world, we increasingly delegate decision-making to systems that make use of vast amounts of data to deliver optimised solutions. Although these answers are usually reliable, certain situations carry profound consequences for the public. This emphasises the need for human decision-makers who can navigate multi-objective decision problems where the interests of each stakeholder group need to be balanced. Analysing such complex trade-offs requires a set of specialised tools that can process data into interpretable formats. To address this, the thesis investigates ways in which data visualisation can assist decision-making when considering multiple objectives and large solution sets. We ground this research in a concrete use case, investment portfolio optimisation, where analysts weigh objectives such as return, risk, sustainability, and liquidity across large sets of candidate portfolios.

Data visualisation is at the core of this interdisciplinary study, providing the integration framework through which everything is combined. In this role, visualisation focuses on the process of transforming data into graphical representations; however, as noted in *Readings in Information Visualization* [CMS99], "the purpose of visualization is insight, not pictures". Consequently, this research centres on the interpretability of multidimensional solution sets, ensuring that decision-makers can gain the required understanding for informed rulings. This process uses a diverse range of techniques, from spatial plots that preserve data structure to complex multivariate plots and dimensionality reduction algorithms. In order to determine which tools best assist the end-user, the technical design must be validated through user studies. This evaluation ensures that the design prioritises human perception, allowing decision-makers to gain the necessary insight to solve today's complex problems.

To establish the basis for addressing these complex problems, this chapter introduces the core concepts required to understand the visualisation of high-dimensional Pareto frontiers. We first look at the cognitive processes behind human decision-making, clarifying how people navigate trade-offs to determine the appropriate outcome, and at the optimisation methods that produce the solution sets these decisions are drawn from. We then present the principles of data visualisation, which stand at the core of this thesis, moving from the foundations of visual encoding to the techniques designed for high-dimensional data and the interactive systems that combine them. Finally, the project scope describes our primary use case and the specific practical problem we are tackling.

The remainder of the thesis is structured around a research paper that shows the main analysis of this problem domain. This section details the evaluation methodology used to answer our research questions and presents the results and findings from our experiments. Following this, we expand upon our methodology and results in the research paper, finishing with the conclusion to the whole thesis.

## 1.1 Intro to Decision Making

When faced with a complex decision, the challenge is in balancing multiple conflicting objectives affected by a given action. This leads to a trade-off, which is defined as "the balancing of factors all of which are not attainable at the same time". We come across these situations on a daily basis, from simple routines to scientific principles, such as the famous space-time complexity trade-off in computer science. Within the context of economics and multi-objective optimisation, trade-offs are frequently modelled mathematically through the Pareto frontier. Before getting to theoretical details, we first need to look at the context in which these decisions are made.

### 1.1.1 Data-Driven Decision Making

Data-Driven Decision Making (DDDM) is a shift from intuitive, heuristic-based choices towards evidence-based strategies. This is helped by the expansion of data collection technologies, which have made it possible to extract actionable, real-time insights. Consequently, systems can optimise performance and forecast outcomes with increasing accuracy, establishing a strong base for both organisational and technical methods [BHK11].

This concept is essential to the scope of this thesis, as the proposed methodology is part of a DDDM framework: transforming complex data into an interpretable format from which insights can be gathered. However, while these insights form the basis for action, decision-makers cannot fully rely on autonomous methods. These algorithms frequently lack the contextual awareness required to understand the "why" of a scenario, especially real-world constraints, ethical considerations, and domain-specific nuances. Therefore, integrating a Human-in-the-Loop (HITL) architecture is key in modern decision-making systems to bridge the gap between analytical power and real-world applicability [ACKK14].

### 1.1.2 Human-in-the-Loop

A HITL architecture integrates human oversight directly into the process, ensuring that machine-generated optimisations are validated by a specialist. By having a human at the centre of the pipeline, the system can use the computational speed and scale of DDDM while relying on the domain expert to provide evaluation of solutions. This human-computer collaboration is particularly important in high-stakes, multi-objective domains such as finance and medicine. In these environments, algorithmic outputs must be compared against complex variables, from unpredictable market volatility and macroeconomic shifts in economics, to patient-specific medical histories and ethical treatment considerations in healthcare [Hol16]. While algorithms can rapidly calculate optimal solutions, a human decision-maker is ultimately required to evaluate the subjective trade-offs between those solutions.

Therefore, the challenge shifts from computing optimal mathematical models to effectively communicating these trade-offs to the domain expert. To achieve this, outputs must be translated into interpretable visual representations, which leads directly to the primary mathematical and visual framework used to map these optimisations: the Pareto Frontier.

### 1.1.3 Pareto Frontier

The Pareto frontier represents the set of solutions that are considered "equally good" overall, where no single objective can be improved without degrading another. This frontier is most commonly visualised in two dimensions, as illustrated in Figure 1. The figure highlights the true boundary of dominance (the solid red line) and an approximation of the Pareto Front surface (the dashed line). It differentiates between optimal solutions that lie on the frontier and dominated points that fall behind it. Furthermore, it visualises the Ideal Point, which acts as a theoretical anchor guiding the optimisation direction.

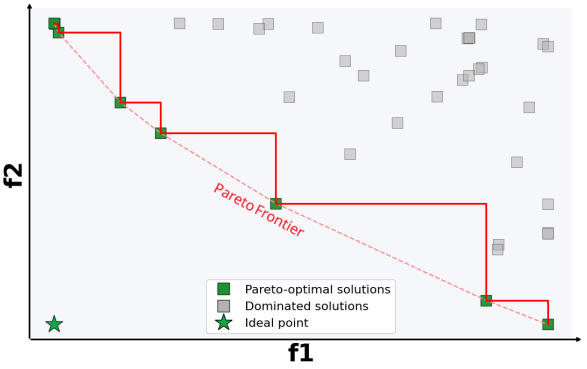



Figure 1: A two-dimensional Pareto frontier showing the ideal point, dominated solutions, and the non-dominated boundary.

More formally, the Pareto frontier, $P(Y)$, is described as follows [Mie98]. Consider an objective function $f : X \to \mathbb{R}^m$, where $X \subset \mathbb{R}^n$ represents the set of all feasible decision variables, and $Y \subset \mathbb{R}^m$ represents the corresponding set of attainable outcomes, such that:

$$Y = \{y \in \mathbb{R}^m : y = f(x), x \in X\} \tag{1}$$

Assuming the desired directions of criteria values are known, a point $y'' \in \mathbb{R}^m$ strictly dominates another point $y' \in \mathbb{R}^m$, denoted as $y'' \succ y'$, when it is at least as good in every objective and strictly better in at least one: $y''_i \leq y'_i$ for all $i \in \{1, \ldots, m\}$, with $y''_j < y'_j$ for at least one $j$. The Pareto frontier is defined as the set of non-dominated points:

$$P(Y) = \{y' \in Y : \{y'' \in Y : y'' \succ y', y' \neq y''\} = \emptyset\} \tag{2}$$

When analysing the Pareto frontier, several points of interest guide decision-makers. The Ideal Point ($z^*$) serves as the theoretical best solution. It is defined as the point containing the best value for each objective function. While typically unattainable in reality, it is the mathematical anchor used to calculate the "closeness" of any feasible Pareto solution. For a minimisation problem, it is defined as:

$$z^* = \left( \min_{x \in X} f_1(x), \min_{x \in X} f_2(x), \ldots, \min_{x \in X} f_m(x) \right) \tag{3}$$

Another feature is the Knee Point. Unlike the Ideal Point, the knee point is a feasible solution located at the sharpest bend of the frontier, where the exchange rate between objectives changes most drastically. Because this depends on the local curvature of the frontier, identifying it is subjective. As a result, there are several methods rather than one universal definition. Geometric approaches such as Normal Boundary Intersection select the solution farthest from the chord connecting the extremes of the front [DD98], maximum-curvature criteria locate the sharpest bend in continuous, differentiable frontiers [Das99], and angle-based methods identify the point of greatest directional deviation in discrete sets [DG11]. Also, on a linear frontier, all trade-offs are identical, and no knee exists.

The Pareto frontier defines the trade-offs a decision-maker must weigh when making a choice. How these frontiers are computed, and why they become difficult to inspect as the number of objectives grows, is discussed in the next section.

## 1.2 Intro to Multi-Objective Optimisation and High-Dimensional Data

This section covers the many options for where Pareto frontiers come from and the difficulties that appear when reading their outputs.

### 1.2.1 Multi-Objective Optimisation Algorithms

Classical approaches reach the frontier through single-solution optimisation, by solving a series of scalarised problems to map the front one point at a time. While these methods provide strong theoretical guarantees for simpler problems, they quickly fall apart when dealing with non-convex spaces or a high number of objectives. Therefore, the field is dominated by population-based metaheuristics [HM25]. Evolutionary algorithms maintain a population of candidate solutions and improve it over generations through selection, recombination, and mutation. NSGA-II [DPAM02], which ranks candidates by dominance and preserves diversity along the front, serves as a benchmark for the field and was used to generate the datasets for our study. Swarm-intelligence variants follow the same population principle with different update rules, and decomposition-based techniques scale to higher objective counts by splitting the problem into many single-objective subproblems. Recently, machine learning approaches have increased in popularity in the field, optimising conflicting objectives directly through gradient descent and automatic differentiation [OVLF25].

The different options are not that important to this thesis, as whichever algorithm runs, the output is a dense, high-dimensional set of non-dominated solutions, and it is that output that we are interested in. The analysis is therefore model-agnostic: the visualisation layer sits downstream of every method above, and the techniques we evaluate make no assumptions about the models.

### 1.2.2 Challenges of High-Dimensional Data

All of these algorithms deliver data whose difficulty grows with every added objective. On the data side, the curse of dimensionality sets in: as dimensions are added, the volume of the space grows exponentially, and the solution set becomes increasingly sparse within it [VF05]. As distances concentrate, the difference between near and far points disappears, destroying the 'closeness' that makes plots readable. For Pareto data specifically, raising the number of objectives leads to dominance resistance: fewer solutions dominate one another, so the frontier becomes large enough to contain most of the set, and no compact boundary remains to draw [HY16]. The frontier also grows as a geometric object, expanding from a curve between two objectives into a surface at three and a multi-dimensional manifold beyond.

The viewer's limits are just as strict. Human spatial comprehension stops at three dimensions, so no mental image of the manifold itself is available past that point. Furthermore, working memory holds only around seven items at a time [Mil56], which rules out mentally handling the values of many objectives across many solutions. And attention is bounded: when a display grows dense, features that are plainly visible go unnoticed if attention is elsewhere, a phenomenon known as inattentional blindness [SC99]. A high-dimensional Pareto

frontier therefore overwhelms both the plotting surface and the person. Bridging the gap between the high-dimensional data and the cognitive limits of the domain expert is the primary challenge of this thesis, and data visualisation provides the tools to do it.

## 1.3 Intro to Data Visualisation

In low-dimensional cases, visualising a solution set is straightforward. With two objectives, a scatterplot places every solution by its two values, and the Pareto frontier appears directly as the curve along the plot's optimal edge. The analyst sees the shape of the trade-off, the density of alternatives, and the outliers at a glance. This is why the two-objective risk-and-return scatterplot has been the go-to visualisation portfolio analysts have been using for decades.

The previous section established why this cannot be used in higher dimensions: the data grows sparse and dominance-resistant while the viewer's spatial understanding, working memory, and attention stay fixed. Data visualisation tries to solve this issue in multiple ways. Some techniques transform the data so that familiar 2D views still apply, others re-encode the dimensions so that all of them remain visible at once, and interactive systems combine several such views and let the analyst filter, highlight, and drill down. Before looking at these different options, we will first go through fundamentals that are needed to reason about the encodings described later.

### 1.3.1 Foundations of Visual Encoding

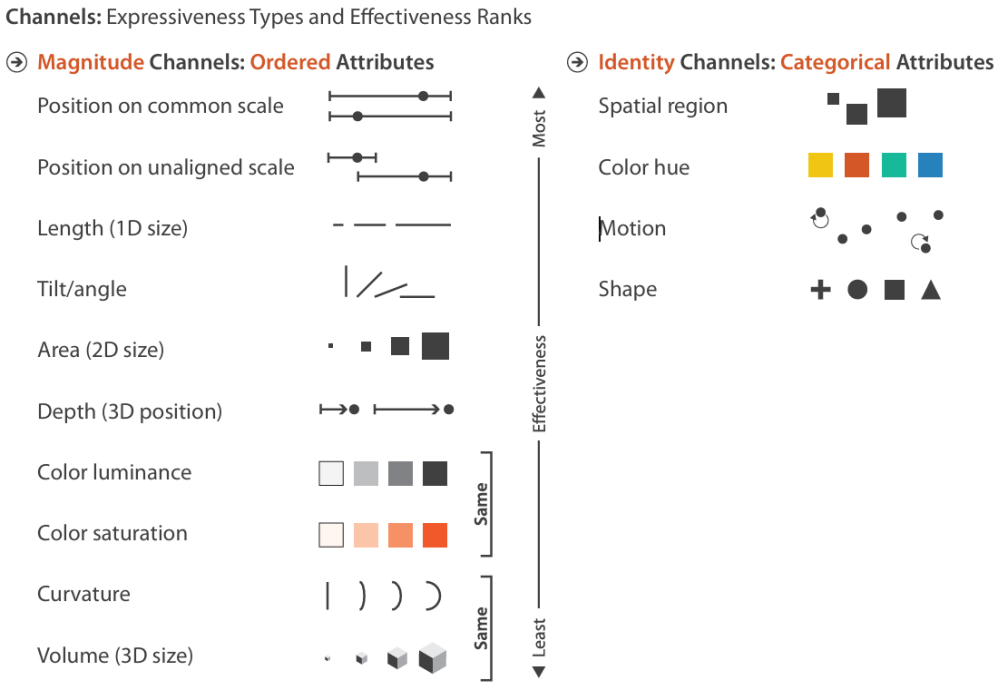



Figure 2: Visual channels ranked by effectiveness, separated into magnitude channels for ordered attributes and identity channels for categorical attributes. Reproduced from Munzner [Mun14, Fig. 5.6].

Visualisation theory decomposes a chart into *marks*, the geometric primitives that represent data items, and *channels*, the visual variables that control a mark's appearance [Mun14].

A complete mapping from data to marks and channels forms an *idiom*, the term we use throughout this thesis. A scatterplot, for instance, is an idiom that maps each item to a point mark and two attributes to the horizontal and vertical position channels.

Two principles guide the choice of mapping. *Expressiveness* requires the visual encoding to show exactly what is in the data, which necessitates matching the channel to the attribute's type. Ordered attributes call for *magnitude* channels that read as more or less, such as position or length, while categorical attributes call for *identity* channels that read as different, such as hue or shape. Using a magnitude channel for categorical data creates a false ranking, while using an identity channel for ordered data obscures a real one. *Effectiveness* demands that the most important attributes receive the best channels, because channels differ measurably in accuracy. The graphical perception experiments of Cleveland and McGill [CM84] established this, and Munzner [Mun14] gathers the results into the ranking seen in Figure 2.

With these foundations, the techniques for high-dimensional data fall into two families: projection-based methods that transform them, and axis-based methods that lay them all out.

### 1.3.2 Projection-Based Techniques

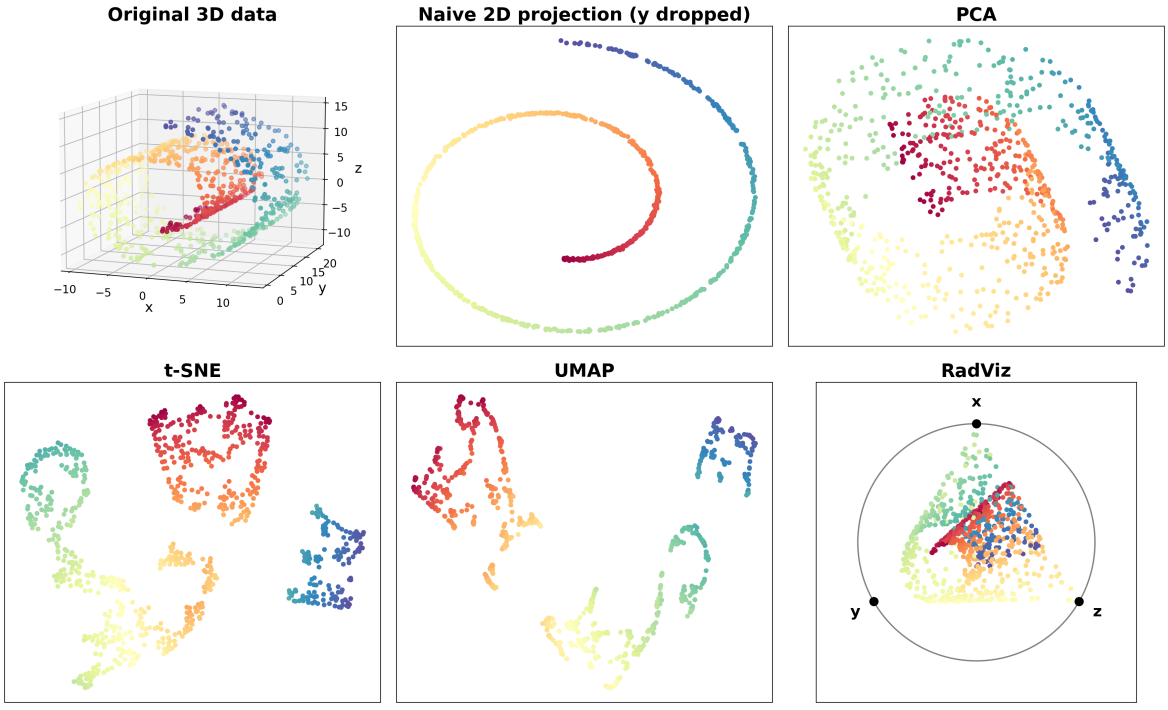



Figure 3: The same three-dimensional swiss-roll dataset flattened by different projection methods, with colour encoding the position along the rolled-up sheet. A naive projection and PCA fail to separate the roll, t-SNE and UMAP recover neighbourhoods but produce abstract axes and fragment the manifold, and RadViz keeps its axes as labelled anchors.

Projection-based techniques keep the familiar low-dimensional spatial encoding and in-

stead transform the data to fit it, computing a low-dimensional embedding of the high-dimensional points. This lets us keep using 2D position, the strongest visual channel. However, the trade-off is that the axes no longer map to the original data. Instead, spatial relationships only show what the specific projection algorithm chooses to preserve.

General-purpose dimensionality reduction methods make this trade-off explicit. *Principal Component Analysis* [Jol14] finds the linear axes of greatest variance, so distances in the plot approximate the dominant structure of the data, and the axes can at least be reported as linear combinations of the original objectives. Nonlinear methods such as *t-SNE* [vdMH08] and *UMAP* [MHSG18] preserve neighbourhood structure instead, producing compact cluster overviews in which neither the axes nor the distances between clusters carry a direct meaning.

Another group of techniques take a landmark-driven approach that keeps the objectives visible in the plot. *RadViz* [HGM*97] is a good example: it places one labelled anchor per objective evenly around a circle and positions each solution at the weighted average of the anchors, so that a solution sits closest to the objectives it scores highest on. The position remains a projection, and different value combinations can land on the same spot, but every anchor is a real, named objective, so the plot stays interpretable.

Figure 3 makes these differences concrete on the classic Swiss-roll dataset, a two-dimensional sheet rolled up in three dimensions, with colour encoding the position along the sheet. Each method flattens the same data differently, and none of them preserves everything.

### 1.3.3 Axis-Based Techniques

Axis-based techniques take a different route. Rather than transforming the data, they give every dimension its own explicit axis, so each visual element stays directly linked to a named objective. The *parallel coordinates plot* (PCP) [Ins85] draws a vertical axis per objective and each solution as a line crossing all of them, so every objective keeps an aligned positional channel and the slopes between adjacent axes reveal pairwise relationships. Its known weakness is that only neighbouring axes can be compared directly, which can be mitigated by interactive reordering. The *radar chart* bends the same layout into a circle, turning each solution into a closed polygon. The radial axes read less precisely than aligned ones and the polygon's shape depends on the axis order, but the closed shape is quickly matched by eye, making the idiom effective for comparing a handful of selected solutions. The *scatterplot matrix* (SPLOM) arranges pairwise scatterplots in a grid, giving each objective pair a precise view at the price of quadratic growth in panels. *Heatmaps* lay solutions and objectives out on a grid and encode values with colour, giving a dense overview in which large-scale patterns stand out. Finally, the *table*, while not a visualisation in the strict sense, presents exact values in a row-column format and remains ubiquitous in practice.

Figure 4 and Figure 5 show four of these idioms on the same small example, eight synthetic portfolios scored on five objectives, with the same two solutions highlighted throughout: an aggressive portfolio (P2, orange) and a defensive one (P7, blue).

### 1.3.4 Coordinated Views and Interactivity

If no single view is sufficient, the question is how several views work together, and the answer the field converged on is interactivity. Shneiderman's information-seeking mantra, *overview*

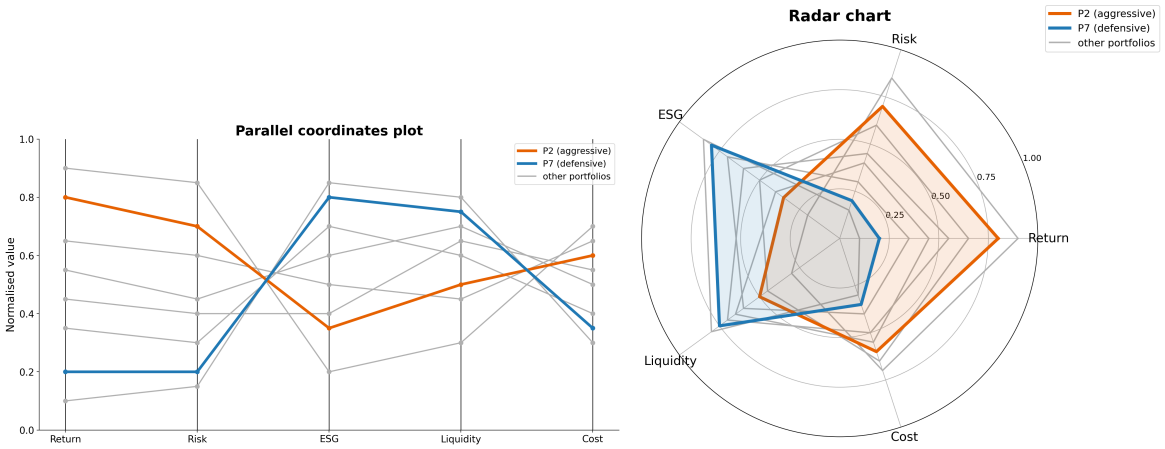



Figure 4: Axis-based idioms on a dataset of eight synthetic portfolios scored on five objectives, with the same two solutions highlighted in every view (P2, orange; P7, blue). The parallel coordinates plot (left) exposes trends and crossings between neighbouring axes, while the radar chart (right) turns the same profiles into comparable shapes.

*first, zoom and filter, then details on demand* [Shn96], reframes visualisation as a process. This counters the cognitive limits described earlier, since at every step the display holds only what the current question requires. What combines multiple views into one is brushing and linking [BCS96], where a selection made in one view is propagated to all others, letting the analyst see the same solutions across different representations at once.

Applied to multi-objective optimisation, these principles produce visual analytics dashboards that link views of the *objective space* with views of the *decision space*. The decision space holds the input variables the optimiser controls (the set $X$), while the objective space holds the outcome values those inputs produce (the set $Y$). Linking the two is what turns a picture of trade-offs into decision support, since the analyst sees not only which trade-offs exist but which decisions produce them. Most works produce dashboards specialised for one optimisation workflow, each fixing the set of views and interactions to its intended analysis [HMM21, HZJ*24, ZWC*18, MZC*25]. The alternative is a modular, general-purpose framework such as ManiVault [VKT*24], whose plug-in architecture supports many data types and view configurations. This field leaves room open for exploration, most visibly around the comparison of several Pareto frontiers [FT18].

## 1.4 Intro to Project Scope

This research was carried out in collaboration with a company (Ortec Finance) whose analysts construct and advise on investment portfolios, and their daily practice is the use case this thesis explores. A portfolio is an allocation of weights over assets that determines how capital is divided. For decades, the analysis of such allocations has been the classical two-objective case of risk against return [Mar52], and the analyst's tools reflect this history. They evaluate optimisation results in a dual-panel dashboard, a 2D scatterplot of the objective space with a tabular view of the decision space.

This workflow is breaking down on two sides. The models have grown to include more objectives, turning a simple curve into a multi-dimensional frontier the 2D scatterplot cannot

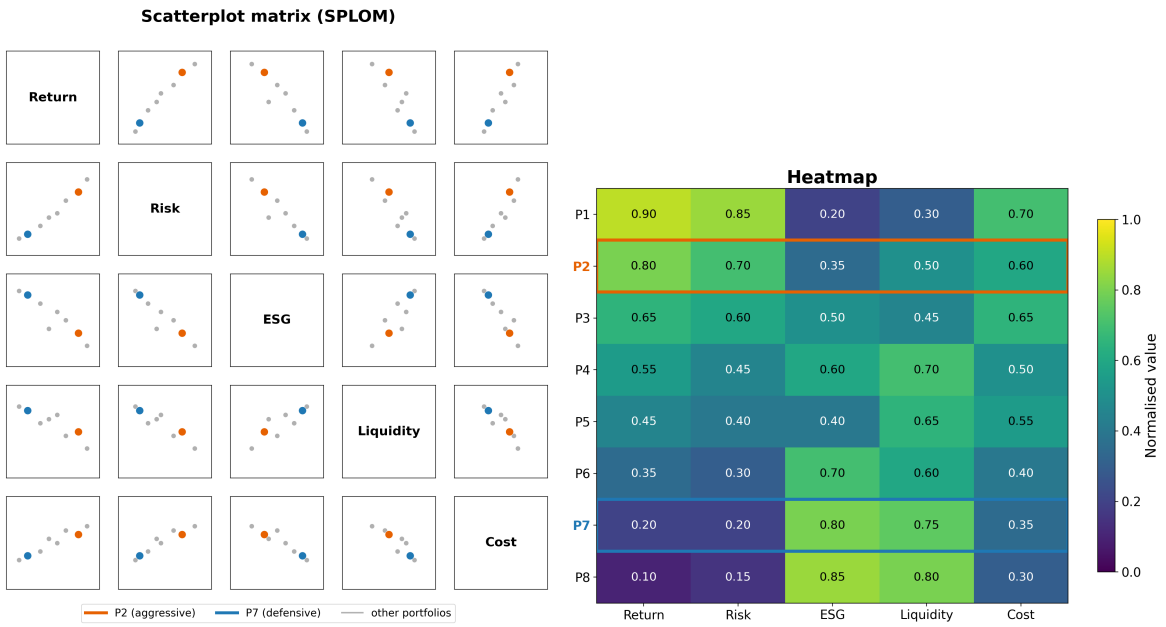



Figure 5: The same portfolios and highlights as Figure 4: the SPLOM (left) isolates every pairwise relationship at the cost of many small panels, and the heatmap (right) compresses all values into one dense overview in which the highlighted rows read as inverted colour patterns.

show. This has pushed analysts toward experimental views that they themselves describe as patches rather than solutions. At the same time, the table view clutters as the number of asset variables grows, and the absence of dynamic highlighting or cross-filtering makes it hard to map an asset composition to the trade-off it produces. The work is also iterative, as analysts re-run optimisations under new constraints and must understand how the results shift, which means comparing several frontiers at once rather than reading one.

This use case perfectly presents the challenges of the preceding sections. It combines high-dimensional frontiers with a human-in-the-loop environment. Because the output must be understood by both the consultants and the clients, interpretability is a must-have. The techniques are evaluated together with the analysts, to see how they put them in practice in real use cases.

The thesis couples this use case with the open question of which visualisation techniques actually support such work. We build a visual analytics framework that links the objective and decision spaces and supports multiple frontiers. The framework is then used as an instrument to study how real users, from domain experts to non-experts, perform analysis tasks with each technique. Two research questions guide the work: *(RQ1) Which visualisation techniques best support the analysis of decision variables and trade-offs within high-dimensional Pareto frontiers?*, and *(RQ2) Which visualisation techniques best facilitate the direct comparison of multiple high-dimensional Pareto frontiers to identify shifts in trade-offs?*. In this study, "best" is evaluated through empirical performance, rather than by any measure of visual quality. A technique supports a task well when users finish it accurately, in reasonable time, and with confidence. The paper presents this investigation, and the chapters after it expand its methodology, results, and societal impact.

# 2

# Paper

# An Analysis of Visualisation Techniques for High-Dimensional Pareto Frontiers: A Case Study in Investment Portfolio Optimisation

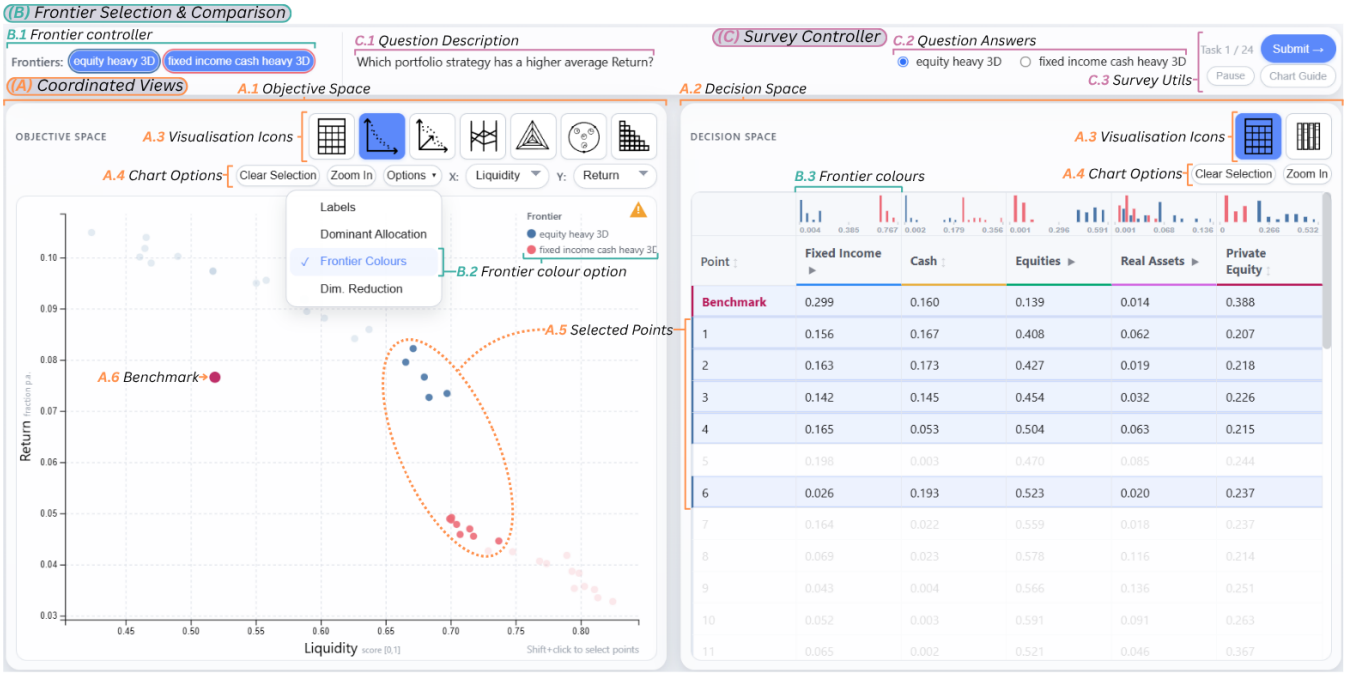



Figure 1: The interface of the framework comprises a set of components (A.1-A.6, B.1-B.3, C.1-C.3) for inspecting and comparing portfolios across three aspects: *(A)* within-frontier trade-off and decision analysis, *(B)* comparison of multiple frontiers, and *(C)* the survey controller.


**Abstract**

*To explore high-dimensional Pareto frontiers calculated using multi-objective optimisation, analysts rely on visualisation techniques ranging from general-purpose charts to purpose-built dashboards. Prior empirical comparisons were not able to select a single chart type as best practice. Additionally, dashboards are evaluated as complete systems, and no study covers investment portfolio optimisation. We address this gap through a study in which domain experts helped characterise the analytical tasks and served as an initial filter for the techniques we evaluate. We then assessed six visualisation techniques arranged in an interactive dashboard with 57 non-expert and 10 expert participants, across a single-frontier analysis and a frontier comparison on 3D and 6D datasets. The results show that the view choices depend on the task being solved, while no single technique dominates overall in usage or performance. Specialised views stand out on specific tasks, while the tables' role as a robust and easy-to-learn baseline is reinforced by the findings in this study. Our framework is available at:* https://github.com/Gunterinos/master_thesis



**CCS Concepts**

• ***Human-centered computing*** → ***Empirical studies in visualization;*** • ***Applied computing*** → *Multi-criterion optimization and decision-making;*


## 1. Introduction

Many real-world decisions must satisfy several conflicting objectives at once. Choosing an apartment, for example, means balancing rent, commute time, and living space, and these kinds of problems do not allow for a realistic single best answer. Gaining on one objective costs something on another, so instead of one optimum, these problems have a set of best possible trade-offs, each striking a different balance. In multi-objective optimisation, this set is called the Pareto frontier. Such problems appear across many fields. Engineers design car bodies that must be light yet crashworthy [XWC*18], and policymakers who price CO2 emissions to fight climate change must still keep fuel affordable [GRK16].

The use case we study focuses on the outputs of multi-objective optimisation for portfolio investment. A portfolio is an allocation of wealth among assets like stocks and bonds. Pension funds and insurers manage these investments for millions of people. Analysts who choose between these portfolios weigh several measures against each other, for example how much money the portfolio is expected to earn, how likely it is to lose value, and whether the investments meet environmental and social standards [PFP21]. Optimising over these measures yields a whole frontier of candidates, and the analyst must understand it well enough to reason over it.

Visualisation is a natural way to read a Pareto frontier. With two objectives, a scatter plot can clearly present the distribution of solutions and the trade-offs among them, so it has been a standard choice for analysts. Once the number of objectives goes beyond two, that frontier becomes hard to read, and a single static picture can no longer reliably represent the trade-offs on its own. Coordinated and interactive views do better: brushing a selection in one view and seeing it linked across the others, following the overview-first, details-on-demand pattern, lets an analyst trace the same solutions through several representations [Shn96, MZC*25]. Surveys of decision support for many-objective optimisation reach the same

conclusion: interactivity and interpretability are essential, and no single static view is enough to support the choice among candidate solutions [OZSR*23]. Moreover, analysts rarely reason about a single frontier in isolation, as they need to compare frontiers originating from different assumptions or constraints to understand how a change in the allocations shapes the trade-offs.

Two gaps remain, however. First, no single technique satisfies all analytical criteria, as the evidence points toward combining complementary views rather than just one [HY17]. Second, existing systems overwhelmingly address single frontiers, while comparing several at once is a secondary concern [FT18]. This leaves domain practitioners without interpretable, multi-frontier tooling appropriate for decision-making contexts.

This work addresses that need through a study of how analysts actually use visualisations to reason over multiple high-dimensional Pareto frontiers. For the study we conduct, we pool the visual techniques into a comparable and interpretable setting. Informed by interviews with domain experts done within the design phase, we build a visual analytics dashboard that links potential solutions to their performance metrics. The resulting dashboard supports the comparison of multiple frontiers and keeps every view tied to the original objectives. With it, we observe how both non-experts and experts in portfolio investment perform analysis tasks with the dashboard, and from their performance and spoken reasoning we gather insights on which techniques suit which situations. Specifically, we address two research questions: *(RQ1) Which visualisation techniques best support the analysis of decision variables and trade-offs within high-dimensional Pareto frontiers?* and *(RQ2) Which visualisation techniques best facilitate the direct comparison of multiple high-dimensional Pareto frontiers to identify shifts in trade-offs?* Here, "best" means empirical performance measured by whether users complete tasks accurately, quickly, and with confidence, rather than a specific measure of visual quality.

Our results show that on a single frontier (RQ1), offering a choice of views raises accuracy at no time cost, and specialised views beat the table on specific tasks, while the table is a robust default. For the comparison of frontiers (RQ2), this performance advantage disappears, and the added views cost extra time, partly because the comparison tasks call for statistics that no view displays. Our investigation of experts and non-experts showed no significant difference in usage, though experts rate the dashboard as more usable and use the complex multi-dimensional views more often than non-experts. In summary, our contributions are:

- a design study with portfolio investment experts and non-experts that defines user tasks and introduces a visual dashboard connecting objective and decision spaces used in solving said tasks;
- an evaluation of six visualisation techniques for Pareto frontiers across single and paired comparisons, along with a usability test of the dashboard as a precondition to our results.

## 2. Related Work

Multi-objective optimisation is applied across domains from engineering design to portfolio management, with visualisation emerging as one of the main strategies for helping analysts interpret the resulting Pareto frontiers [Mie14]. He and Yen [HY16] explain that as the number of objectives grows beyond three, no representation can at once exactly represent the original frontier, support decision-making, and scale to the dimensionality required. This is recognised by Hakanen et al. [HGMR23], who provide a review that focuses on more accessible, flexible, and interactive visualisation toolkits and on better support for the decision-making process, including the integration of analyst preferences and domain context. Osika et al. [OZSR*23] survey post-optimisation decision support visualisation as one of the primary methods, while highlighting interactivity and explainability as essential for new solutions. What these surveys agree on is not which view is best, but that progress depends on extracting *insights* by combining complementary views, rather than on building yet another dashboard. We take this as the starting point of our study. Instead of proposing a new framework, we pool views into one linked dashboard and evaluate how each contributes to the analyst's tasks.

To structure our work, we follow the taxonomy of Filipič and Tušar [FT18] between methods that visualise a single approximation set and those that visualise several sets, which mirrors our two research questions. Earlier work [TF15] showed that established single-frontier methods scale poorly as the number of objectives grows, so any technique carried into the multi-frontier setting must be assessed carefully at higher dimensionality. We use two terms throughout the study. *The decision space* is the set of variable values that make up a solution. *The objective space* is the scores those values produce. Within the single-frontier category, we introduce the grouping of projection-based and axis-based methods, which both address the objective space, before moving on to the interactive systems that link it to the decision space. This shows that few existing studies address how to compare multiple solution sets.

### 2.1. Projection-Based Techniques

A common approach to high-dimensional Pareto data is to project it into a lower-dimensional space for visual inspection. General-purpose dimensionality reduction methods such as Principal Component Analysis [Jol14], t-SNE [vdMH08], and UMAP [MHSG18] produce compact overviews of solutions by finding lower-dimensional embeddings that preserve certain properties of the original data, such as variance or neighbourhood structure.

A distinct family of projection methods takes a landmark-type approach, placing labelled anchors that correspond to objectives and projecting solutions relative to those fixed reference points. RadViz [HGM*97] is a prime example, mapping multi-dimensional objective values to a position within a circular layout anchored by each objective dimension. Further specialised methods have been proposed that build on or depart from this idea, including 3D-RadViz [IRMD16], PaletteViz [TD20], and iSOM [NRD23]. Based on expert interviews, our framework includes PCA and RadViz but excludes specialised variants.

### 2.2. Axis-Based Techniques

Axis-based representations maintain a direct link between visual elements and objective dimensions [Mun14]. The 2D scatterplot has each solution as a point placed by two chosen objectives on the horizontal and vertical axes. Both axes encode an objective through position along a common scale, which is the channel with the least perceptual error [CM84, Mun14]. 3D scatterplot adds a third orthogonal axis for a third objective, trading visual clarity and depth accuracy for higher-dimensional context. Scatterplot matrices (SPLOMs) arrange pairwise scatter plots in a grid for bivariate comparison between any two objectives. Parallel coordinate plots

(PCPs) draw each solution as a line across a set of labelled vertical axes. Radar charts radiate objectives from a central point and encode each solution as a closed polygon. Heatmaps lay solutions and objectives out on a grid and map each objective value to colour intensity, giving a dense overview in which patterns across many solutions become readable at a high level. Tabular representations, while not a visualisation in the strict sense, are widely used in practice, presenting objective values in a row-column format.

Several empirical studies have compared these axis-based representations in multi-objective decision-making contexts. Dy et al. [DIPJ22] evaluated PCPs, SPLOMs, radar charts, and heatmaps across different objective counts and solution set sizes, finding accuracy to be comparable and more sensitive to the number of options and dimensions shown than to the choice of chart. Dimara et al. [DBD18] compared PCPs, SPLOMs, and tabular views on multi-attribute choice tasks and found them comparable on most measures, with a slight advantage for tabular views. He and Yen [HY17] extended the comparison across both projection-based and axis-based families, stating that no technique fully satisfies all criteria and that a combination of complementary views is needed. Axis-based views clearly label every objective, making them central to our dashboard, while the table serves as our baseline due to its robustness in prior research. Since overall performance is comparable, our study includes all of these views.

### 2.3. Interactive Visual Analytics Dashboards

Prior work combines projection- and axis-based views into interactive dashboards that link the two spaces. Hakanen et al. [HMM21] proposed a task-based framework linking visualisation design to specific analyst tasks. This helps structure the interaction between decision-makers and the optimisation process. Related systems apply the same concept to different aspects of the problem. Huang et al. [HZJ*24] join objective-space views with population-dynamics panels to show how the search evolves across generations. Similarly, SkyLens [ZWC*18] combines projection, tabular, and pairwise-comparison views to reveal why individual solutions dominate in skyline queries. Finally, ParetoLens [MZC*25] is a complete system that links the objective and decision spaces with interactive filtering and framing visual analytics as a path to explainability, so that users explore the Pareto front through interaction. These systems help design our dashboard, as we adopt the same pattern of linked objective-space and decision-space views with interactive filtering. However, these systems are evaluated as complete units, not looking into which individual views support which tasks, and none deal with the financial portfolio setting.

### 2.4. Comparing Multiple Pareto Frontiers

While single Pareto frontier visualisation is well studied, prior work highlights that the comparison of multiple frontiers is essential for iterative decision-making [TF15]. When several solution sets are superimposed without proper differentiation, visual overlap and data density can reduce readability. Most of this work originates in the evolutionary optimisation literature, where comparing algorithm outputs across repeated runs is essential. Tušar and Filipič [TF15] treat the display of several approximation sets as a prerequisite for comparing them, and propose a prosection-based method that preserves dominance relations across sets. Von Lücken et al. [vPJDL24] combine shape-based clustering with clustering-defined filters and ranking methods, together with size and colour encodings, to separate frontiers while preserving their structure. In more practical settings, Shavazipour et al. [SLM21] adapt empirical attainment functions and heatmaps so that a decision-maker can compare solution sets across scenarios.

Despite this body of work, multi-frontier comparison has not yet been adopted as a primary design goal in visual analytics systems targeting domain practitioners, namely users who require interpretable, objective-labelled views and who must communicate findings to non-technical stakeholders. Together with the interpretability constraints discussed in the preceding sections and the absence of domain-specific tooling for financial optimisation, this gap defines the design space that the framework presented in this study directly addresses.

## 3. Domain Characterisation and Design Goals

To establish our domain characterisation and design goals, we followed the design study methodologies proposed by Munzner [Mun09] and Sedlmair et al. [SMM12]. During the study, we collaborated with two distinct groups. The company's R&D specialists served as an initial expert filter, refining the visual encodings for domain-specific analysis needs. In parallel, quantitative analysts and financial consultants were our front-line evaluators. Through constant collaboration with the analysts, we kept the framework grounded in their real-world applications. Furthermore, by holding exploratory discussions with both groups, we established the specific domain context and identified the current bottlenecks analysts face in their workflows. We then translated these findings into functional requirements, design goals, and system tasks.

### 3.1. Multi-Objective Portfolio Optimisation

In defining the domain context, our discussions showed a swift evolution within the field. Traditional portfolio optimisation has historically relied on a mean-variance approach, aiming to find the optimal trade-off between risk and return [Mar52]. However, the advancement of computational capabilities has expanded these early frameworks, allowing for the use of specialised metrics such as ESG (Environmental, Social, Governance) scores [Pol24], liquidity, and higher-order moments [SQH08, KTF14]. As the number of objectives grows beyond standard risk and return, the classical framework itself has been generalised into multi-objective models capable of constructing efficient surfaces across three or more dimensions [BCHL24, UWHS14]. For scenarios where these traditional solvers face scalability limitations, the field has adopted metaheuristics like Multi-Objective Evolutionary Algorithms (MOEAs) [HM25]. As more advanced models output dense, multidimensional datasets rather than simple two-dimensional curves, they shift the analytical bottleneck from computation to interpretation.

### 3.2. Current User Workflow

While the final validation will include a wide range of users, the design process must reflect the needs of domain experts. To understand the real-world bottlenecks that analysts face, we observed how they use their current frameworks during our discussions, identifying their needs, friction points, and practical limitations.

In their existing workflow, optimisation is typically a two-dimensional case of risk and return, featuring a small set of approximately ten portfolios. Analysts evaluate these results using a

dual-panel dashboard: the objective space is visualised on the left via a 2D scatterplot, while the decision space is displayed on the right in a tabular format. This dashboard includes a basic selection mechanism for portfolios of interest, and it allows users to load multiple, colour-coded frontiers alongside a benchmark portfolio.

During our exploratory discussions, analysts noted that the tabular representation of the decision space introduces significant visual clutter, particularly when evaluating portfolios with numerous asset variables. This issue is worsened by the lack of dynamic highlighting and cross-filtering, making it difficult to map complex asset compositions to their resulting trade-offs. Furthermore, for optimisation models that scale beyond two objectives, analysts have noted occasional use of experimental visualisations, such as 3D scatter plots and ternary grid plots incorporating pie glyphs. In response to their requests, we incorporated both a 3D scatter plot and an extension of the RadViz plot [HGM*97], as detailed in section 4. However, they noted that these methods were used as patches rather than permanent solutions for their high-dimensional bottlenecks.

### 3.3. Requirement Elicitation and Design Goals

To improve on the current workflows, we established system requirements through semi-structured exploratory interviews. These discussions included two R&D specialists and two quantitative analysts. The sessions were structured in four progressive phases: domain definition and experience, general visualisation experience and workflows, visualisation utility for investment portfolios, and evaluation of high-dimensional data representations. During the final phase, users were presented with a prototype containing visual idioms such as parallel coordinate plots (PCP), scatterplot matrices (SPLOM), radar charts, 3D scatter plots, and stacked bar charts. Based on the explicit and implicit needs gathered during these validation sessions, we summarised the interviewees' requirements into three design goals:

- **G1: Enable Cross-Space Interactivity:** All participants agreed that static representations are insufficient for high-dimensional analysis. The system must support dynamic actions, such as brushing, filtering, and coordinated highlighting. This helps to link the objective space and the decision space, which is essential for identifying patterns and mitigating the visual clutter that is present in high-dimensional datasets.
- **G2: Enable Multi-Frontier Comparison:** Analysts work iteratively, starting with unrefined frontiers and changing parameters to narrow down results. The system must support the comparison of multiple Pareto frontiers to help with this progression. Furthermore, because analysts rely on the expectations of their clients, the framework must have a distinct benchmark solution across all views to serve as an anchor for these comparisons.
- **G3: Manage High-Dimensional Complexity:** All participants stressed that interpretability must be prioritised. When evaluating high-dimensional data, analysts rejected abstract dimensionality reduction techniques (such as t-SNE [vdMH08] or PCA [Jol14]), noting that the artificial axes could not be linked to real portfolio values, making the results nearly impossible to explain to non-technical stakeholders. Therefore, the framework must show the original, explainable performance metrics, mitigating the resulting density through filtering, selection, and the option to toggle specific dimensions off. Meanwhile, density in the decision space can be naturally reduced by grouping individual asset allocations into higher-level asset classes.

Together, the goals of enabling interactivity, aiding comparison, and managing complexity form the blueprint for the proposed visual framework. We now transform these requirements into the specific tasks the system must support.

### 3.4. Tasks

We abstracted the analysts' workflow into two primary tasks. These tasks reflect the need to explore cross-space relationships within a single optimisation run (T1) and the iterative process of refining constraints across multiple runs (T2).

- **T1: Query Trade-offs and Analyse Decision Variables – RQ1:** When analysing a single Pareto frontier, the user first attempts to build an understanding of the dataset through a series of exploratory queries. Following the taxonomy outlined by Dimara et al. [DBD18], we decompose this workflow into fundamental tasks, such as "Retrieve Value", "Determine Range", and "Correlate", that support a final decision task. For the scope of our evaluation, value retrieval and range determination were merged to accommodate time constraints. Finally, these exploratory queries provide the trade-off context necessary for the user to select the portfolio that best fits their specific investment goals and constraints.
- **T2: Compare Multiple Pareto Frontiers – RQ2:** Given the iterative nature of portfolio optimisation, analysts must evaluate how changing constraints impact performance. This task centres on comparing multiple frontiers to support a choice task between different optimisation strategies. Analysts must identify areas of Pareto dominance within the objectives and investigate divergences across both the objective and decision spaces. By identifying where the frontiers shift, such as observing how a change in the asset allocation impacts the performance metrics, users gain the insight needed to choose the optimisation scenario best suited for their specific investment strategy.

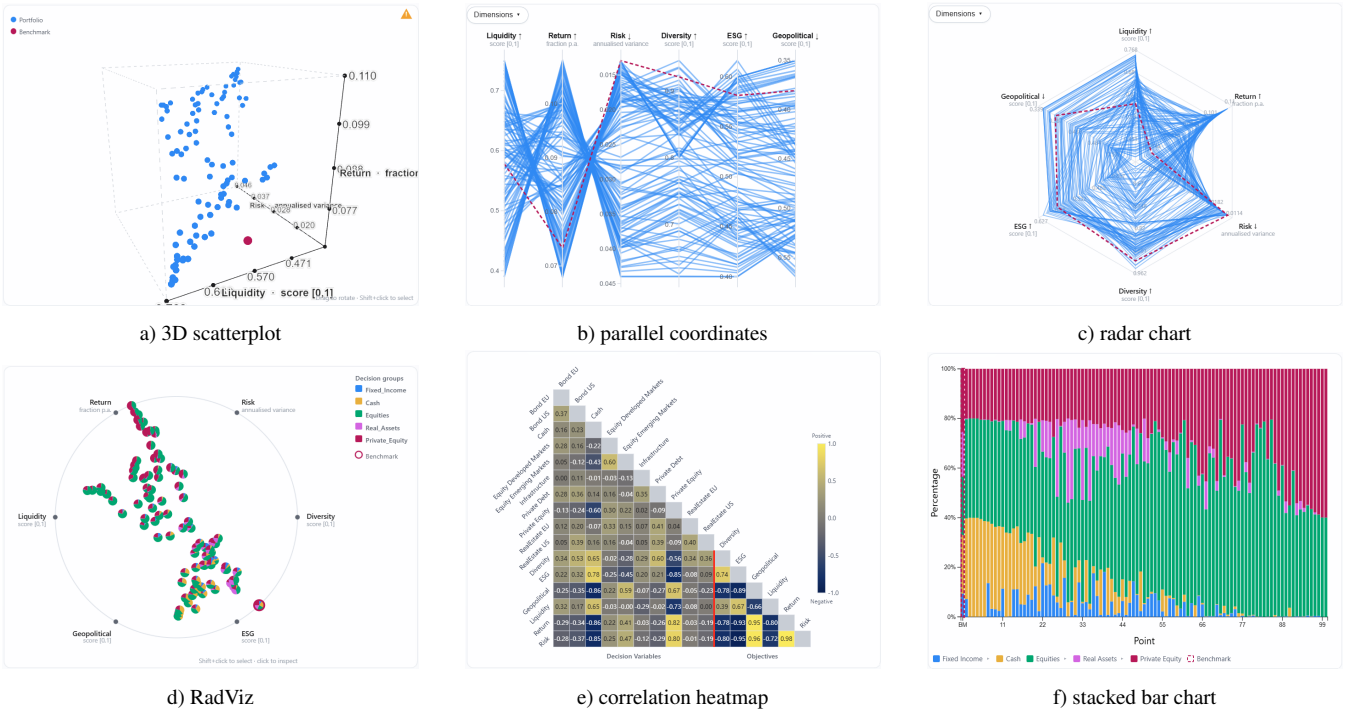
These two tasks of analysing individual frontier trade-offs and comparing multiple optimisation scenarios define the requirements for our proposed framework. By matching our visualisation design with these needs, we ensure that the system directly supports the decision-making process of analysts.

## 4. Visualisation Design

The framework turns the goals and tasks behind the analysts' work into concrete views. We start from the dual-panel dashboard the analysts already use and keep the same split for our interface of two linked panels, one for the *objective space* and one for the *decision space*. Figure 1 shows the resulting interface. Two coordinated panels sit at its centre, the objective space on the left and the decision space on the right, each showing the same frontiers through one idiom chosen from the row of icons above it, so that a shift in trade-offs is seen in both views, the insight at the core of task T1. A benchmark solution is always present as a reference, and several frontiers can be loaded and managed from the controller at the top left, with consistent colours across views to differentiate runs.

### 4.1. Coordinated Views

Both panels rest on the same foundation: the way an idiom is chosen and the interactions that keep the two spaces in sync. These realise goal G1 and guide every idiom described below.



a) 3D scatterplot b) parallel coordinates c) radar chart

d) RadViz e) correlation heatmap f) stacked bar chart

Figure 2: The objective-space idioms and the stacked bar chart shown on the same frontier.

The objective space panel (Figure 1, A.1) and the decision space panel (A.2) sit side by side, each showing the same solutions through one idiom at a time. The idiom is chosen from a row of icons above each panel (A.3), where every icon is a preview of the view it switches to, so a chart is recognised by its shape rather than its name. Neither panel opens on a default idiom. Until the user picks one, the panel shows the prompt "Select a chart type above to get started". We chose this so the layout does not steer the user toward any view, which lets the evaluation capture an unbiased preference. The cost is an extra step before any data is visible.

Within these panels, interaction is the main way the framework deals with clutter, which the interviews raised as essential and which we set as goals G1 & G2. Hovering or selecting a solution in one panel highlights or selects the same solution in the other (A.5), so the analyst can read the decision variables behind an interesting trade-off and the inverse. Solutions can be selected in every idiom, and the chart options (A.4) hold the shared actions around a selection. Zooming in keeps only the selected solutions so a crowded region can be studied on its own, zooming out restores the full set, and a clear action drops the selection. An options menu shows the per-idiom settings, described alongside each idiom. Together these realise the overview-first, details-on-demand pattern [Shn96] and give the cross-space reasoning that task T1 needs.

Every chart also shows a benchmark solution in a crimson colour (A.6). In a real environment, the benchmark is the solution an analyst already expects, generated from the assumptions of the user.

### 4.2. Objective Space

With the shared foundation in place, we turn to the idioms. The objective space panel is where analysts weigh solutions against each other, so it holds most of the idioms. We move from the familiar low-dimensional cases up to the views that show many objectives at once, all shown in Figure 2, where they depict the same frontier.

The *2D scatterplot*, seen in Figure 1 A.1, was the analysts' main view in the two-objective case, so it is the natural starting point. Only two objectives are shown at a time, so a warning label reminds the user that the rest are hidden. Labels can be turned on or off, and the dominant variable option colours each point by the decision variable that makes up a plurality of the solution. The dimensionality reduction option uses PCA to project all objectives into a two-dimensional view. We selected this approach because it maintains interpretability (G3), displaying the principal components and their linear combinations as axis labels. The *3D scatterplot*, seen in Figure 2a), extends the same idea into three axes. Although the literature discourages unjustified 3D [Mun14], it is justified here by the analysts' explicit demand for it. It adds options such as rotation controls, a surface coloured by its distance to the ideal point, and a cloud of dominated points that makes the shape easier to see. The *parallel coordinate plot*, seen in Figure 2b), was among the most liked views in the initial discussions and follows ParetoLens [MZC*25]. It gives control over the axes: they can be toggled, reordered by dragging, filtered together, and minimised axes are flipped, so the ideal solution would be a line along the top. The *radar chart*, shown in Figure 2c), displays the same information as the parallel coordinates plot in a radial layout. While it is less precise [CM84, Mun14], it forms a closed shape that the eye matches quickly and provides a familiar format seen in sports and video game statistics.

An extension of *RadViz*, seen in Figure 2d), originates from a ternary grid plot with pie glyphs that one of the analysts had built. A pie glyph at each position encodes the solution's decision variables, so a position stays interpretable (G3) and the two spaces are read together. A spread option pushes the overlapping glyphs apart into a polygon at the cost of positional accuracy, and the encoding itself is sensitive to anchor order, with distinct solutions able to be located at the same position. Finally, the *correlation heatmap*, seen in Figure 2e), uses the Spearman correlation [Spe04] for every pair of decision variables and objectives. Colour encodes the correlation through a gradient, with dark blue for the most negative and yellow for the most positive. Based on pilot feedback, a vertical red line and bottom labels clearly separate objectives from decisions.

### 4.3. Decision Space

The decision space panel shows the values each solution takes across the decision variables. The *stacked bar chart* , seen in Figure 2f), is the compact decision-space overview the analysts demanded in the interviews. Each solution is one bar split into its groups of decision variables, which can be expanded by clicking on them. The bars follow the order of the trade-offs, so the values shift gradually along the front, mitigating the issue of a missing baseline for the middle bars [CM84]. It also features a colourblind-safe palette for accessibility. The *table*, seen in Figure 1 A.2, is the most direct view and the closest to what the analysts already used. Dimara et al. [DBD18] note that a table is a strong all-round support for decision-making across many dimensions. Available in both panels, it shows the direction of each objective column and includes a distribution plot at the top for brush-filtering. It also lets users easily sort columns and expand or collapse variable groups.

### 4.4. Comparing Multiple Frontiers

Everything so far has described a single frontier. Comparing several optimisation runs is the focus of task T2 and goal G2, while keeping the clutter manageable for goal G3, and it is supported by a frontier controller along with a shared colour scheme used by the idioms.

The loaded frontiers are listed and managed from the controller at the top left of the interface (Figure 1, B.1). Each frontier can be turned on or off, with the constraint that at least one always stays visible. Hovering over a frontier shows the constraints that produced it, which lets the analyst know what separates one run from another without leaving the view.

Each frontier is given a colour, shown in the controller and its legend (B.2). The colours carry through the views wherever an idiom can show which frontier a solution belongs to, through a frontier colours option. In the objective space, the scatterplots, parallel coordinates, radar, and RadViz colour their points and lines by frontier, so several runs can be overlaid and told apart, as shown for the 2D scatter plot in Figure 1 A.5. Hue is used because it is the strongest remaining channel for a categorical attribute [Mun14]. The correlation heatmap cannot overlay runs in a single matrix, so it draws one matrix per frontier, placed side by side and each marked with its frontier colour and label. Splitting each cell between frontiers would trade away the readability of a single frontier, which juxtaposition preserves [Gle18]. In the decision space, the stacked bar chart faces a similar issue, since its colours already encode the groups of decision variables. Rather than recolour the bars, it keeps each frontier's solutions together and marks the runs with bracket annotations on top, since proximity groups more strongly than the palette and the contiguous blocks read as separate frontiers [BSLF24]. The table colour-codes its distributions and gives each row a colour band showing its frontier (B.3).

Together, these interactions help the user through task T2. Overlaying runs in the objective space shows where one frontier dominates another, while the decision-space views show where the runs differ in their variables, letting an analyst weigh both sides and pick the optimisation that best fits their goal. The benchmark stays fixed in its crimson colour as the shared reference point that fulfils the anchoring role goal G2 asked for.

## 5. Evaluation

Having presented the design of the dashboard, we now turn to whether it achieves its actual purpose, which is to let analysts *reason* about high-dimensional Pareto frontiers. Following the design study methodology of Sedlmair et al. [SMM12], this chapter forms the *validate* stage, returning to the problem characterisation set out in the domain analysis and testing it against evidence from real users. The aim is therefore not to evaluate the dashboard as a product, but to determine whether any of the individual visualisations generate valuable insights in this specific domain context.

The validation is based on tasks T1 and T2. For both, the starting question is the same: whether the freedom to choose among coordinated views lets a user perform analytical tasks more accurately and quickly than just the baseline, and whether that advantage survives when more frontiers are in view. This section breaks down the tasks into hypotheses.

### 5.1. Study Design

Users of this type of tool do not form a single audience. On one side are the experts, the consultants who know the underlying model well enough to explain results to clients. On the other side are the clients, who lack that model knowledge and prior experience. We mirror this imbalance in the evaluation by having the non-expert group do a quantitative study to capture what the average user perceives. Then, the smaller expert group does a qualitative study to see how a practised analyst reasons with the views.

A purely task-based, controlled study struggles to capture insights. North [Nor06] argues that abstracting analysis into benchmark tasks with definite answers strips away the very aspects that make a finding insightful. Open exploration by itself is also insufficient as a measurement, since it gives the participant no starting point. Chang et al. [CZGR09] integrate the two by describing insight as a loop, in which structured tasks build the domain knowledge that later makes a discovery possible. We therefore structure both studies around the layered insight- and task-based methodology of Gomez et al. [GGZL14], which interleaves blocks of low-level tasks with periods of free exploration. This also lets us use the terminology found in Lam et al. [LBI*12], measuring *user performance* through time and accuracy, gathering *user experience* through a closing questionnaire, and reserving the more demanding *visual data analysis and reasoning* for the expert interviews, where thinking aloud and asking questions make real insight observable. Combining several methods in this way follows the wider guidance that a single evaluation method is rarely sufficient [IAI*24], and that experimental visualisation work benefits from pairing quantitative and qualitative evidence tailored to each stage of the study [LWR*24].

The two datasets we use are synthetic rather than generated by the company's production model. Generating the data ourselves lets us shape cases with clear outliers that stress the views and ensure the evaluation remains fully reproducible. We generate the frontiers with the NSGA-II algorithm [DPAM02] on two structurally identical problems. As a fixed *benchmark*, we use the average across all generated solution sets, which gives the user a stable reference point to orient themselves against the results. To onboard users, each session opens with a twelve-step guided tutorial covering every visualisation and interaction in the dashboard, supported during the tasks by a built-in cheat sheet that summarises each view and

animates how its filtering works. A pause button, added after pilot feedback, lets participants stop the timer for a break without affecting recorded times. These two interactions can be seen in Figure 1, C.3. To capture the insights in the decision making process, we recorded spoken reasoning as it happened during the expert interviews. Finally, both studies were approved by the Human Research Ethics Committee of the university where this study was conducted.

### 5.2. Quantitative Study with Non-Experts

The quantitative study asks whether the dashboard, which offers a choice of coordinated views rather than a single fixed table, lets non-expert users complete tasks more accurately and quickly. Testing this at scale requires abstract, repeatable tasks, so for a single frontier we use a block of three drawn from the decision-support taxonomy of Dimara et al. [DBD18]. A *value retrieval* task asks the participant to select the solutions that meet a threshold on one objective or decision variable. A *correlation* task asks which variable/objective moves most strongly with a given objective/variable. A *decision* task asks for the solution that offers the best overall trade-off across every objective. The decision task is the most demanding and closest to real practice, since it forces the participant to weigh all objectives at once, and it forms a path to the open exploration found in the expert study.

The comparison of several frontiers at once follows the same three steps: a retrieval, a correlation, and a decision. A *simple comparison* asks which frontier has the higher average on an objective. A *most distinct objective* task asks where two frontiers differ the most, the kind of trade-off analysis identified as essential to comparing solution sets [HY17]. A *decide on frontier* task mirrors the single-frontier choices by asking for a final selection.

For each of the tasks, we used a structure that can be easily interchanged for different use cases:

- *value retrieval*: "Select all {solutions} where {objective or decision variable} is {above/below} {threshold}."
- *correlation*: "Which {objective or decision variable} is most strongly correlated with {target} ({positively/negatively})?"
- *decision*: "Which {solution} do you think offers the best overall trade-off across all objectives?"
- *simple comparison*: "Which {frontier} has a {higher/lower} average {objective or decision variable}?"
- *most distinct objective*: "In which {objective or variable group} do the two {frontiers} differ the most on average?"
- *decide on frontier*: "Which {frontier} performs better overall across all objectives?"

The non-expert participants are divided between the two research questions, with one part working on a single frontier and the other on the comparison of two frontiers. Each part is split again by dataset, the financially literate participants working with the *portfolio* data and those standing in for the general population with a *dietary* dataset of the same structure that swaps financial jargon for everyday quantities, and within every cell the order of the dashboard and the table baseline is counterbalanced to avoid order effects. Each participant then works through eight blocks of three tasks, four on a 3D frontier and four on a 6D one. Every block runs the same three task types, while the interface records the time on each task, the time spent in every chart, and a timestamped log of each interaction. This non-expert evaluation structure is summarised in Figure 3. Figure 1 shows the task controls, including the question description (C.1), the answer input (C.2), and the submission controls (C.3)

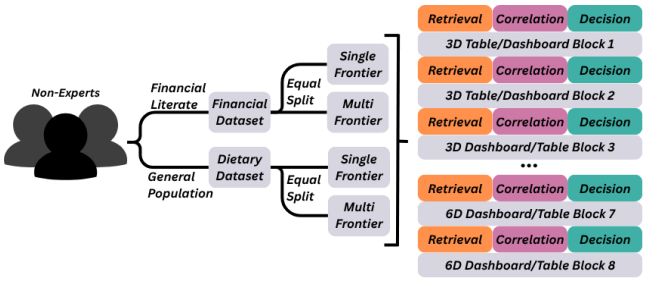



Figure 3: Summary of the non-expert evaluation setup.

Every task yields an accuracy score in $[\mathbf{0}, \mathbf{1}]$ alongside the time to answer, awarding partial credit where possible so that a near miss is not treated as a flat failure. Value retrieval is scored by the Jaccard overlap between the selected points and the correct set. The decision task scores a selection by how close its points sit to the normalised ideal point, so choosing only the closest point scores **1**, and choosing the most distant point or making no selection yields **0**. Correlation and most distinct objective share a rank-based score that rewards the strongest candidate, while the simple comparison and decide on frontier tasks are binary (**1** for correct, **0** for incorrect), scoring **1** for the frontier with the higher objective mean and the higher mean normalised objective value, respectively.

The performance comparisons are driven by a set of hypotheses. The first of which is a precondition on usability. We measure usability once per participant at the end with the four UMUX [Fin10] items, chosen as a short questionnaire focused on pragmatic quality rather than on enjoyment [SKT23], answered on a seven-point scale with two items worded positively and two negatively and combined into a single score on the $[\mathbf{0}, \mathbf{100}]$ range. A final open-ended question catches any feedback missed by the scales. The remaining hypotheses are listed below together with robustness checks.

- **H1 — Usability precondition.** As we have designed the dashboard with best practices in mind, we expect it to achieve a high usability score. More specifically, we expect a mean UMUX above the average of 68 on the SUS scale [LS18] with UMUX mapping onto the SUS scale [Fin10].
- **H2 — A choice of views beats the table baseline.** While the table is an established, robust support for multi-attribute decision making [DBD18], given the empirical support for interactive views [NS00], we expect that the table view as a separate, single view, performs worse than a combination of (user-chosen) views in task execution. The claim is guarded by two inline robustness checks: a *dimensionality* check that the advantage holds as the objective count rises, and a *practice* check to ensure it holds even when accounting for learning.
- **H3 — View Specialization.** As idiom effectiveness depends on the task [KARC15], we expect users to converge on specific idioms based on task type. A *popularity-versus-performance* check verifies that the most-used view is also the most accurate, and an *order-effect* check tests whether being exposed to the table baseline first leads to reliance on the table.
- **H4 — Interactions manage complexity (G3).** Focusing and linking make high-dimensional data manageable [BCS96], so we expect views to perform better when their interactions are used than when they are used statically.

Following the shift from significance testing to estimation [Dra16, DBD18], we report aggregate effects as an estimate with a 95% Bias-Corrected and accelerated (BCa) bootstrap interval over $B = 10{,}000$ participant resamples. BCa refines the percentile bootstrap by correcting for two effects: a bias correction that shifts the interval if the resampled estimates are not centred on the observed value, and an acceleration factor that adjusts for non-constant variance when the standard error depends on the true parameter.

For testing H3, the objective and decision panels are scored separately, and a view $\mathbf{v}$ is credited on a task when its viewing time makes up more than a quarter of the total time spent across all views. The table is counted once per panel, so it does not track usage across both. The most-used view scores highest on a task's usage only when it confidently beats the runner-up in a direct comparison, filtering out the noise of less-used views: with $u_1$ observations for the leader and $u_2$ for the runner-up, the leader wins if the Clopper–Pearson lower bound on its share exceeds one half,

$$B_{\alpha/2}(u_1,\, u_2+1) \;>\; 0.5, \tag{1}$$

where $B$ is the Beta quantile function and $\alpha = 0.05$. The accuracy of a view on a task is its mean score over credited tasks, judged by a Wilson score interval, used instead of the bootstrap because BCa cannot resample beyond the observed scores and becomes unreliable at small $n$. A view is reliably better than chance on a task only when its Wilson lower bound exceeds 0.5.

For testing H4, we adopt this same assignment: a task counts as with-interaction when at least one action belonging to $\mathbf{v}$ appeared and without-interaction otherwise, and we compare a view's accuracy between the two, estimated with a percentile bootstrap interval over tasks. The H3 and H4 patterns are read as associations, not causes, because participants chose which views to use and how much to interact with. Therefore, more interaction can indicate that a task is harder rather than produce a worse score.

### 5.3. Mixed-Methods Study with Experts

To see where expert and non-expert behaviour intersects, the experts in portfolio investment work through the same structured tasks, which here play a double role. They are scored with the same accuracy and time measures as in the quantitative study, but here those scores act only as signals to inform the discussion rather than as formal results, while the tasks also build familiarity before the exploratory part. At the end of each block, the decision task is reframed for the experts. Rather than aiming at the single best trade-off that the quantitative study scores, they are asked to explore the data freely while thinking aloud and to reach a choice on their intuition. Their reasoning and discoveries are recorded and later summarised. This open phase, supported by the interview format, is where the *visual data analysis and reasoning* [LBI*12] that the quantitative study could not reach actually takes place. The experts close with the same UMUX questionnaire and open-ended question used in the quantitative study.

## 6. Results

We first test the hypotheses quantitatively with the non-experts and then employ a mixed-methods approach with the experts to gain a deeper understanding of underlying procedures and motivations.

### 6.1. Quantitative Results

The quantitative study contributes 57 non-experts, split into 29 on a single frontier and 28 on the comparison of two. All results are reported as a mean followed by its 95% interval in brackets. Starting with H1, it is supported, as the mean UMUX of 73.8 [69.2, 77.3] clears the established average of 68.

H2 holds whether offering a choice of views beats the table baseline. This is true for the single frontier, as there is a small accuracy gain ($+0.061$ $[+0.015, +0.109]$), in score points on the $[0,1]$ scale, at no time cost ($\times 0.92$ $[0.75, 1.13]$ of baseline time, parity at $\times 1.00$), which supports H2. For the two frontiers, this gain disappears ($-0.023$ $[-0.075, +0.026]$), and the dashboard also costs a third more time ($\times 1.33$ $[1.16, 1.59]$). Looking at the robustness checks, the accuracy advantage does not scale with dimensionality, as the 6D check is inconclusive on both the single frontier ($+0.007$ $[-0.064, +0.082]$) and the comparison ($-0.036$ $[-0.138, +0.053]$). However, the practice check holds because the accuracy gap is order-invariant ($-0.019$ $[-0.114, +0.081]$ and $-0.018$ $[-0.119, +0.084]$), confirming that the performance boost is not a learning effect. The benefit of a dashboard with more choices is there, though it is present only on a single frontier.

Turning to H3, we look at which view best supports each task, read separately for the objective and decision panels. Figure 4 sets the two quantities against each other. The left panel counts, for each view, the number of tasks on which it cleared the 25% usage share. The right panel gives each view's mean accuracy per task, split by 3D, 6D and overall, with cells whose Wilson lower bound clears the 0.50 benchmark shaded and marked. On a single frontier, only the objective panel specialises, as different tasks lead participants to different views: the correlation heatmap dominates *correlation* and the table dominates *value retrieval*, while on *decision* no view dominates. Alignment holds on *correlation*, where the dominant heatmap also clears the benchmark, but fails on *value retrieval*, where the table dominates usage while the PCP is the more accurate view. The decision panel cannot separate the tasks this way, offering only two views of which one is the general-purpose table. It dominates all three tasks without an accuracy advantage over the stacked bar, and on *decision* neither view clears the benchmark.

On the comparison of frontiers, specialisation isn't really present in either panel: the objective panel confirms it only for *decide on frontier*, and the decision panel does not confirm it on any task. In the objective panel, the 2D scatter plot is the better and more used view for *simple comparison*, while the parallel coordinates plot and the radar chart are more accurate than the table for *most distinct* objective. For *decide on frontier*, the table is both the most used and accurate. In the decision panel, the table and the bar chart stay close on both usage and accuracy across all three tasks, with neither one a clear winner.

Participants who saw the table baseline before the dashboard then overused the table once all views were available, raising its usage share on the single frontier ($+16.6$ $[+3.7, +29.8]$ percentage points), though the same shift is inconclusive on the comparison ($+17.0$ $[-3.2, +36.7]$). A retrospective check shows the effect is seen only in the behaviour of participants, as the usability score was untouched by the order of exposure (table-first 73.9 vs dashboard-first 73.6, difference $+0.3$ $[-7.6, +8.6]$).

For H4, we investigate whether a view's own interactions pay

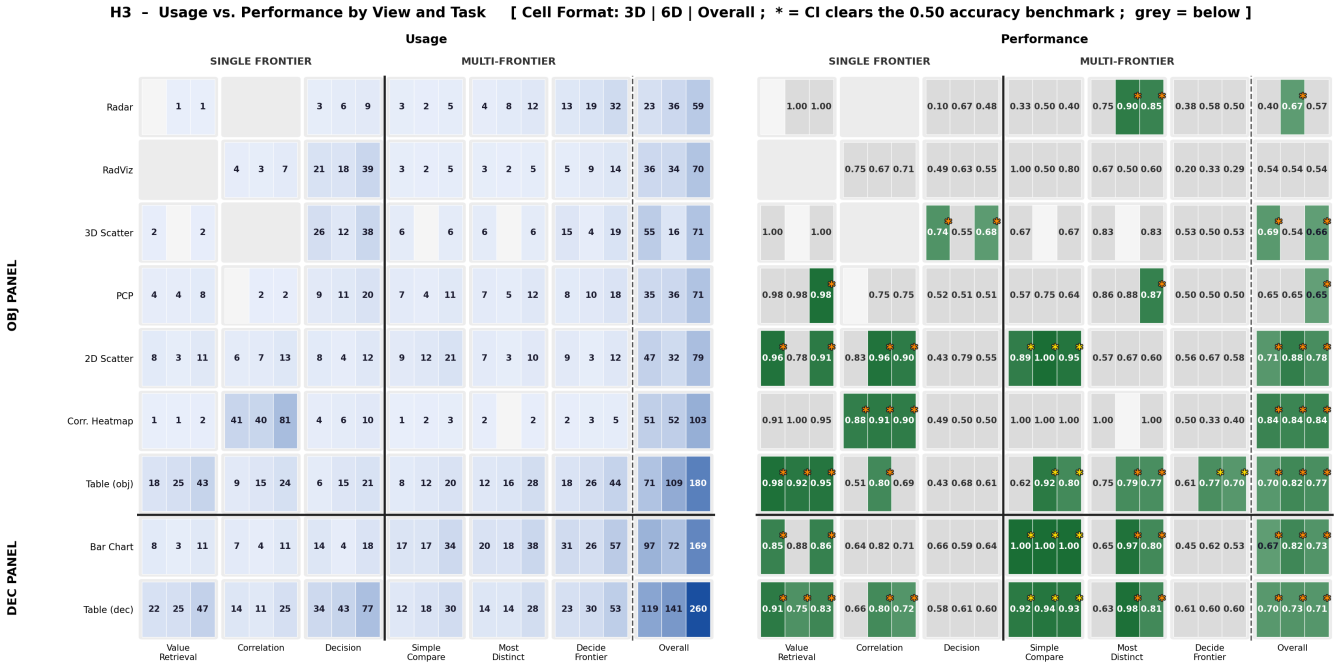



Figure 4: **H3 — usage vs. performance by view and task.** Left: the number of tasks where a view was used more than 25% of total panel usage. Right: mean accuracy per view and task. (95% Wilson interval), split by 3D, 6D, and Overall. Cells whose 95% Wilson lower bound clears the 0.50 accuracy benchmark are shaded green and marked with an asterisk; grey cells fall below it. Yellow asterisks mark binary tasks where 0.50 is chance level. Orange asterisks mark graded tasks where chance is lower and 0.50 serves as a conservative threshold.

off: tasks where the view's own interactions were used are compared to when the view was not interacted with. Interacting improves the accuracy for the table ($+0.130$ $[+0.026, +0.236]$ and $+0.184$ $[+0.093, +0.270]$ across the two panels) and the bar chart ($+0.199$ $[+0.088, +0.306]$ ), while for every other objective view this is not the case. The correlation heatmap has no interactive controls and wins its task passively.

### 6.2. Mixed-Methods Results

While the quantitative data captures what choices users made, the expert sessions provide deeper insight than just accuracy metrics. Ten experts, mostly insurance and pension professionals, worked through a compressed version of the study while thinking aloud, and their telemetry and remarks are read against the non-expert results (H1–H4) for consensus or contradiction.

On the usability metric, experts rated the dashboard higher than non-experts (mean UMUX 82.5 Experts E vs 73.8 Non-Experts NE, with one expert outlier at 54). A retrospective check shows a difference of +8.7, whose 95% interval [0.3,15.9] excludes zero, so the experts rate the dashboard reliably higher. Experts also ranked it higher on individual scoring items and voiced their enthusiasm verbally ("I really like it"), while calling it an "expert toolkit" that "not every client" could use. To ground this comparison, we ran a retrospective analysis of the experts' view usage and performance against the non-experts. Experts differ in how they work, not in how well, as the accuracy difference between the groups shows no reliable gap overall ($+0.04$ $[-0.03, +0.10]$) or on any task. Experts used the correlation heatmap on every correlation task (E 100% vs NE 74%), with one expert noting "we really wanted this to know about the correlation between objectives," and largely avoided the table on objective tasks (E 0–8% vs NE 18–23%), while mainly using the table in the decision space. The divergence between experts and non-experts is in the harder tasks, where experts used the views non-experts avoided: the parallel coordinates plot on the multi-frontier objective tasks, and the 3D scatter on the single-frontier decision, despite expert concerns about its depth cues. The non-experts' statistical friction is also shared with the experts, as the comparison tasks demand an average the dashboard never displays, so experts estimated it from the small distribution histograms in the table ("I'm not a human calculator"), and their strain at six objectives ("the data sort of fogged my mind at times") matches where the dashboard struggled to scale to 6D. Similar to the non-experts, increased interaction was a sign of task difficulty.

The insights come from the decision task, which the experts performed in its reframed, open-ended form. On a single frontier, no expert looked for a single best point. The task became curation, selecting a couple of candidates "for the storyline" to take to a client, and the choice was structured around client profiles, since which points survive "depends on what type of client you're dealing with". During the study, the decision space was a filter for points: points with attractive scores were discarded on their allocation ("I would never pick this one, it has 96% in fixed income"). The additional options views offered served as anchors, with one expert calling up the ideal point only to overrule it ("I'm not sure if I should go for the closest point, because it is very high in liquidity …I would trade some liquidity for some performance"). Several experts learned the unfamiliar views through the linked selection, marking points in the 2D scatter and watching where they appeared in the parallel plot or the 3D scatter. On the comparison, the decision question was never about which frontier scores better, but about which is the better starting point for the next optimisation run: experts proposed changes to the constraints to shift the frontier in a better direction ("it would be a good starting point, from there I would tweak it"), anticipated how a constraint would impact the allocations, and made the choice dependent on the stage we were in ("if you're exploring, unconstrained; if your client already knows what they need, constrained"). The answer to *decide on frontier* depends on client context rather than on anything the views themselves show. Averages were read from shape rather than

computed, judging skew ("return seems to be skewed for the red one, so it will be dragging averages down"), the same averaging gap behind the lost accuracy advantage.

Taken together, H1 holds, while the experts rated the dashboard even higher. H2 holds on the single frontier (RQ1) but not on the comparison (RQ2). H3 holds in the objective panel on the single frontier and fades on the comparison, and H4 holds only for the table and the bar chart. The experts follow the same pattern but pick different views: they match non-expert accuracy, while they use the correlation heatmap, the PCP and the 3D scatter, where non-experts fall back on the table. The table stays reliable, and the specialised views clear the benchmark on specific tasks. The open exploration shows that *decide on frontier* depends on client context rather than on anything the views display.

## 7. Discussion

On a single frontier (RQ1), giving analysts a choice of views pays off: accuracy rises against the table baseline at no time cost, and the robustness checks show the gain is not due to a practice effect of going over the same tasks multiple times. The experts had the same experience, enjoying the dashboard but calling it an "expert toolkit". Also, some non-experts said the same in their comments, calling it "useful if you know what you are doing". Both audiences agree that the views are worth having, but they become valuable through practice, and a single session is not enough time to learn them. This explains why participants often default to the table, our established baseline [DBD18]. The table is a familiar view with interaction attached to it, as part of the table's strength is that it was never static, as its accuracy was increased when using interaction compared to when not. This supports the position of Munzner [Mun14] that interactivity is what lets a view carry complexity. However, another part of the table's pull is exposure rather than preference, since seeing the table baseline first increased later table use. Although the table remains a solid general choice in both spaces, on the single frontier some view outperforms it on every task. The clearest is the correlation heatmap, which leads the *correlation* task on both usage and accuracy and holds up against the bias above. Moreover, in the *decision* task, the table itself falls below our benchmark, leaving the 3D scatter as the only view that consistently remains above it. This shows that a view tailored to a task will perform better and more consistently. Therefore, there is no single best view; the table is the safe default, and specialised views beat it when the task is specific.

For the comparison of multiple frontiers (RQ2), the accuracy advantage seen on the single frontier does not carry over, and the added views cost a third more time while returning nothing in accuracy. Part of the reason is that *simple comparison* and *most distinct* both rely on an average that the dashboard never shows, leaving analysts to answer it by eye. This appeared on both sides of the study: experts said that they were not a "human calculator" and that there was "no average metric," and several non-experts independently asked in their feedback why there was "no way to actually see the average of the portfolio". Where the task calls for a numerical answer rather than a visual one, a dedicated statistics view is the best solution, as it can do for the comparison tasks what the heatmap does for correlation. The third task resists this fix, since on *decide on frontier* the table is the only view that stays reliable. These results do not mean that no visualisation works for high-dimensional Pareto data. Two things made the bar hard to clear. First, the table is a robust tool for multi-attribute choice [DBD18], and it sits inside the dashboard as one of its own views, so any view has to beat a strong option that users can always fall back on. Second, with six views on offer and a single long session per participant, there was little room for habits to form, so several tasks ended in ties between views. The recommendation for RQ2 is therefore the same as for RQ1, and for the same reason: keep the table as the familiar default, and add views built for the specific task, with a statistics view being the gap this study leaves open.

### 7.1. Limitations and Future Work

Several limitations constrain this study. The clearest is the lack of an established set of tasks for multi-frontier comparison. Where the single-frontier tasks build on Dimara et al. [DBD18], the comparison tasks had no such basis. A second limitation is that the six views themselves add complexity. These are a means to an end for this work rather than a claim that they are the right set. That choice might have led to decision fatigue, something the baseline avoids, which could be a reason participants did sometimes better with it. The remaining limitations concern how far the results transfer. The study was long, as several participants said, which risks fatigue late in a session. More training might have given the unfamiliar views a fairer chance, though watching people reason with unfamiliar views was the goal of this study. Furthermore, the expert group is also small, so a more varied expert group could change the findings and improve the quality of our evidence. Finally, because participants chose their own views and how much to interact, the H3 and H4 patterns are associations rather than causes.

Each of these points to a next step. The most direct is to define a proper set of comparison tasks, as the literature did for single-frontier decisions, so that later studies rest on shared, well-defined tasks. Another way forward is visualisation recommendation systems tailored to high-dimensional Pareto frontiers: rather than offering every view at once, the system would suggest the ones suited to the current task and data, and a study could test whether this lowers the cognitive strain seen here [ZLQW24, WZW*23, ALF*25]. Finally, assigning views to participants would turn the H3 and H4 associations into causal claims, at the cost of the natural behaviour this study set out to observe.

## 8. Conclusion

We set out to understand which visualisations help analysts reason about high-dimensional Pareto frontiers, and which help compare several of them. To answer this, we built a visual analytics dashboard with multiple coordinated views and studied non-experts quantitatively and portfolio-investment experts both quantitatively and qualitatively as they worked through analysis tasks with the dashboard. For a single frontier (RQ1), a choice of coordinated views outperforms a plain table, and the benefits are task-specific rather than universal. For the comparison of several frontiers (RQ2), that advantage disappears and the added views cost time, largely because the comparison asks for statistics that no view displays. Across both, the table stands as a strong default for either audience, while the specialised views earn their place on the tasks they were built for, with the correlation heatmap as the clearest case. More views do not make a dashboard better, while matching views to tasks does. That match exists for a single frontier but not for the comparison, and the interactive table remains the most reliable tool for this.

## References


[ALF*25] AODENG G., LI G., FENG Y., CHEN Q., ZHANG Y., LIU C. H.: Inreactable: Llm-powered interactive visual data story construction from tabular data. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2025), UIST '25, Association for Computing Machinery. URL: https://doi.org/10.1145/3746059.3747719, doi:10.1145/3746059.3747719. 10

[BCHL24] BLITZ D., CHEN M., HOWARD C., LOHRE H.: 3d investing: Jointly optimizing return, risk, and sustainability. *Financial Analysts Journal 80*, 3 (2024), 59–75. doi:10.1080/0015198X.2024.2335142. 3

[BCS96] BUJA A., COOK D., SWAYNE D. F.: Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics 5*, 1 (1996), 78–99. URL: http://www.jstor.org/stable/1390754. 7

[BSLF24] BEARFIELD C. X., STOKES C., LOVETT A., FRANCONERI S.: What does the chart say? grouping cues guide viewer comparisons and conclusions in bar charts. *IEEE Transactions on Visualization and Computer Graphics 30*, 8 (2024), 5097–5110. doi:10.1109/TVCG.2023.3289292. 6

[CM84] CLEVELAND W. S., MCGILL R.: Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association 79*, 387 (1984), 531–554. URL: http://www.jstor.org/stable/2288400. 2, 5, 6

[CZGR09] CHANG R., ZIEMKIEWICZ C., GREEN T. M., RIBARSKY W.: Defining Insight for Visual Analytics. *IEEE Computer Graphics and Applications 29*, 2 (Mar. 2009), 14–17. doi:10.1109/MCG.2009.22. 6

[DBD18] DIMARA E., BEZERIANOS A., DRAGICEVIC P.: Conceptual and Methodological Issues in Evaluating Multidimensional Visualizations for Decision Support. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (Jan. 2018), 749–759. doi:10.1109/TVCG.2017.2745138. 3, 4, 6, 7, 8, 10

[DIPJ22] DY B., IBRAHIM N., POORTHUIS A., JOYCE S.: Improving Visualization Design for Effective Multi-Objective Decision Making. *IEEE Transactions on Visualization and Computer Graphics 28*, 10 (Oct. 2022), 3405–3416. doi:10.1109/TVCG.2021.3065126. 3

[DPAM02] DEB K., PRATAP A., AGARWAL S., MEYARIVAN T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation 6*, 2 (Apr. 2002), 182–197. doi:10.1109/4235.996017. 6

[Dra16] DRAGICEVIC P.: Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*, Robertson J., Kaptein M., (Eds.). Springer International Publishing, Cham, 2016, pp. 291–330. doi:10.1007/978-3-319-26633-6_13. 8

[Fin10] FINSTAD K.: The Usability Metric for User Experience. *Interacting with Computers 22*, 5 (Sept. 2010), 323–327. doi:10.1016/j.intcom.2010.04.004. 7

[FT18] FILIPIČ B., TUŠAR T.: A taxonomy of methods for visualizing pareto front approximations. In *Proceedings of the Genetic and Evolutionary Computation Conference* (Kyoto Japan, July 2018), ACM, pp. 649–656. doi:10.1145/3205455.3205607. 2

[GGZL14] GOMEZ S. R., GUO H., ZIEMKIEWICZ C., LAIDLAW D. H.: An insight- and task-based methodology for evaluating spatiotemporal visual analytics. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Oct. 2014), pp. 63–72. doi:10.1109/VAST.2014.7042482. 6

[Gle18] GLEICHER M.: Considerations for visualizing comparison. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (2018), 413–423. doi:10.1109/TVCG.2017.2744199. 6

[GRK16] GARNER G., REED P., KELLER K.: Climate risk management requires explicit representation of societal trade-offs. *Clim. Change 134*, 4 (Feb. 2016), 713–723. 1

[HGM*97] HOFFMAN P., GRINSTEIN G., MARX K., GROSSE I., STANLEY E.: DNA visual and analytic data mining. In *Proceedings. Visualization '97 (Cat. No. 97CB36155)* (Phoenix, AZ, USA, 1997), IEEE, pp. 437–441,. doi:10.1109/VISUAL.1997.663916. 2, 4

[HGMR23] HAKANEN J., GOLD D., MIETTINEN K., REED P. M.: *Visualisation for Decision Support in Many-Objective Optimisation: State-of-the-art, Guidance and Future Directions*. Springer International Publishing, Cham, 2023, pp. 181–212. doi:10.1007/978-3-031-25263-1_7. 2

[HM25] HANNE T., MOGHADDAM M. J.: A review of the evolution of multi-objective evolutionary algorithms. *Computers, Materials and Continua 85*, 3 (2025), 4203–4236. URL: https://www.sciencedirect.com/science/article/pii/S1546221825009889, doi:10.32604/cmc.2025.068087. 3

[HMM21] HAKANEN J., MIETTINEN K., MATKOVIĆ K.: Task-based visual analytics for interactive multiobjective optimization. *Journal of the Operational Research Society 72*, 9 (Sept. 2021), 2073–2090. doi:10.1080/01605682.2020.1768809. 3

[HY16] HE Z., YEN G. G.: Visualization and performance metric in many-objective optimization. *IEEE Transactions on Evolutionary Computation 20*, 3 (2016), 386–402. doi:10.1109/TEVC.2015.2472283. 2

[HY17] HE Z., YEN G. G.: Comparison of visualization approaches in many-objective optimization. In *2017 IEEE Congress on Evolutionary Computation (CEC)* (June 2017), pp. 357–363. doi:10.1109/CEC.2017.7969334. 2, 3, 7

[HZJ*24] HUANG Y., ZHANG Z., JIAO A., MA Y., CHENG R.: A comparative visual analytics framework for evaluating evolutionary processes in multi-objective optimization. *IEEE Transactions on Visualization and Computer Graphics 30*, 1 (2024), 661–671. doi:10.1109/TVCG.2023.3326921. 3

[IAI*24] ISLAM M. R., AKTER S., ISLAM L., RAZZAK I., WANG X., XU G.: Strategies for evaluating visual analytics systems: A systematic review and new perspectives. *Information Visualization 23*, 1 (Jan. 2024), 84–101. doi:10.1177/14738716231212568. 6

[IRMD16] IBRAHIM A., RAHNAMAYAN S., MARTIN M. V., DEB K.: 3d-radvis: Visualization of pareto front in many-objective optimization. In *2016 IEEE Congress on Evolutionary Computation (CEC)* (2016), pp. 736–745. doi:10.1109/CEC.2016.7743865. 2

[Jol14] JOLLIFFE I.: *Principal Component Analysis*. John Wiley Sons, Ltd, 2014. doi:10.1002/9781118445112.stat06472. 2, 4

[KARC15] KANJANABOSE R., ABDUL-RAHMAN A., CHEN M.: A multi-task comparative study on scatter plots and parallel coordinates plots. *Comput. Graph. Forum 34*, 3 (June 2015), 261–270. 7

[KTF14] KOLM P. N., TÜTÜNCÜ R., FABOZZI F. J.: 60 Years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research 234*, 2 (Apr. 2014), 356–371. doi:10.1016/j.ejor.2013.10.060. 3

[LBI*12] LAM H., BERTINI E., ISENBERG P., PLAISANT C., CARPENDALE S.: Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics 18*, 9 (Sept. 2012), 1520–1536. doi:10.1109/TVCG.2011.279. 6, 8

[LS18] LEWIS J. R., SAURO J.: Item benchmarks for the System Usability Scale. *Journal of Usability Studies 13*, 3 (2018), 158–167. URL: https://dl.acm.org/doi/10.5555/3294033.3294037. 7

[LWR*24] LIN F., WANG A. Z., RAHMAN M. D., SZAFIR D. A., QUADRI G. J.: Striking the Right Balance: Systematic Assessment of Evaluation Method Distribution Across Contribution Types. In *2024 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)* (Oct. 2024), pp. 129–135. arXiv:2408.16080, doi:10.1109/BELIV64461.2024.00020. 6

[Mar52] MARKOWITZ H.: Portfolio selection. *The Journal of Finance 7*, 1 (1952), 77–91. URL: http://www.jstor.org/stable/2975974, doi:10.2307/2975974. 3

[MHSG18] MCINNES L., HEALY J., SAUL N., GROSSBERGER L.: Umap: Uniform manifold approximation and projection. *Journal of Open Source Software 3*, 29 (2018), 861. URL: https://doi.org/10.21105/joss.00861, doi:10.21105/joss.00861. 2

[Mie14] MIETTINEN K.: Survey of methods to visualize alternatives in multiple criteria decision making problems. *OR Spectr. 36*, 1 (Jan. 2014), 3–37. 2

[Mun09] MUNZNER T.: A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (Nov. 2009), 921–928. doi:10.1109/TVCG.2009.111. 3

[Mun14] MUNZNER T.: *Visualization Analysis and Design*. A K Peters/CRC Press, Dec. 2014. 2, 5, 6, 10

[MZC*25] MA Y., ZHANG Z., CHENG R., JIN Y., TAN K. C.: Paretolens: A visual analytics framework for exploring solution sets of multi-objective evolutionary algorithms [application notes]. *IEEE Computational Intelligence Magazine 20*, 1 (2025), 78–94. doi:10.1109/MCI.2024.3487134. 1, 3, 5

[Nor06] NORTH C.: Toward measuring visualization insight. *IEEE Computer Graphics and Applications 26*, 3 (May 2006), 6–9. doi:10.1109/MCG.2006.70. 6

[NRD23] NAGAR D., RAMU P., DEB K.: Visualization and analysis of Pareto-optimal fronts using interpretable self-organizing map (iSOM). *Swarm and Evolutionary Computation 76* (Feb. 2023), 101202. doi:10.1016/j.swevo.2022.101202. 2

[NS00] NORTH C., SHNEIDERMAN B.: Snap-together visualization. In *Proceedings of the working conference on Advanced visual interfaces* (New York, NY, USA, May 2000), ACM, pp. 128–135. 7

[OZSR*23] OSIKA Z., ZATARAIN SALAZAR J., ROIJERS D. M., OLIEHOEK F. A., MURUKANNAIAH P. K.: What Lies beyond the Pareto Front? A Survey on Decision-Support Methods for Multi-Objective Optimization. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence* (Macau, SAR China, Aug. 2023), International Joint Conferences on Artificial Intelligence Organization, pp. 6741–6749. doi:10.24963/ijcai.2023/755. 2

[PFP21] PEDERSEN L. H., FITZGIBBONS S., POMORSKI L.: Responsible investing: The ESG-efficient frontier. *J. Financ. Econ. 142*, 2 (Nov. 2021), 572–597. 1

[Pol24] POLLMAN E.: The making and meaning of esg. *Harv. Bus. L. Rev. 14* (2024), 403. 3

[Shn96] SHNEIDERMAN B.: The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages* (1996), pp. 336–343. doi:10.1109/VL.1996.545307. 1, 5

[SKT23] SCHREPP M., KOLLMORGEN J., THOMASCHEWSKI J.: A Comparison of SUS, UMUX-LITE, and UEQ-S. *Journal of User Experience 18*, 2 (2023), 86–104. URL: https://dl.acm.org/doi/abs/10.5555/3604890.3604893. 7

[SLM21] SHAVAZIPOUR B., LÓPEZ-IBÁÑEZ M., MIETTINEN K.: Visualizations for decision support in scenario-based multiobjective optimization. *Information Sciences 578* (Nov. 2021), 1–21. doi:10.1016/j.ins.2021.07.025. 3

[SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (Dec. 2012), 2431–2440. doi:10.1109/TVCG.2012.213. 3, 6

[Spe04] SPEARMAN C.: The proof and measurement of association between two things. *The American Journal of Psychology 15*, 1 (1904), 72–101. URL: http://www.jstor.org/stable/1412159, doi:10.2307/1412159. 5

[SQH08] STEUER R. E., QI Y., HIRSCHBERGER M.: Portfolio selection in the presence of multiple criteria. In *Handbook of Financial Engineering*, Zopounidis C., Doumpos M., Pardalos P. M., (Eds.). Springer US, Boston, MA, 2008, pp. 3–24. doi:10.1007/978-0-387-76682-9_1. 3

[TD20] TALUKDER A. K. A., DEB K.: PaletteViz: A Visualization Method for Functional Understanding of High-Dimensional Pareto-Optimal Data-Sets to Aid Multi-Criteria Decision Making. *IEEE Computational Intelligence Magazine 15*, 2 (May 2020), 36–48. doi:10.1109/MCI.2020.2976184. 2

[TF15] TUŠAR T., FILIPIČ B.: Visualization of Pareto Front Approximations in Evolutionary Multiobjective Optimization: A Critical Review and the Prosection Method. *IEEE Transactions on Evolutionary Computation 19*, 2 (Apr. 2015), 225–245. doi:10.1109/TEVC.2014.2313407. 2, 3

[UWHS14] UTZ S., WIMMER M., HIRSCHBERGER M., STEUER R. E.: Tri-criterion inverse portfolio optimization with application to socially responsible mutual funds. *European Journal of Operational Research 234*, 2 (2014), 491–498. 60 years following Harry Markowitz's contribution to portfolio theory and operations research. URL: https://www.sciencedirect.com/science/article/pii/S037722171300605X, doi:10.1016/j.ejor.2013.07.024. 3

[vdMH08] VAN DER MAATEN L., HINTON G.: Visualizing data using t-sne. *Journal of Machine Learning Research 9*, 86 (2008), 2579–2605. URL: http://jmlr.org/papers/v9/vandermaaten08a.html. 2, 4

[vPJDL24] VON LÜCKEN C., PEREIRA U., JAVIER DÁVALOS E., LÓPEZ-PIRES F.: Improvement Strategies for Visualizing Solution Sets in Many-Objective Optimization Problems. *IEEE Access 12* (2024), 142406–142418. doi:10.1109/ACCESS.2024.3467997. 3

[WZW*23] WANG L., ZHANG S., WANG Y., LIM E.-P., WANG Y.: LLM4Vis: Explainable visualization recommendation using ChatGPT. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track* (Singapore, Dec. 2023), Wang M., Zitouni I., (Eds.), Association for Computational Linguistics, pp. 675–692. URL: https://aclanthology.org/2023.emnlp-industry.64/, doi:10.18653/v1/2023.emnlp-industry.64. 10

[XWC*18] XIONG F., WANG D., CHEN S., GAO Q., TIAN S.: Multi-objective lightweight and crashworthiness optimization for the side structure of an automobile body. *Struct. Multidiscipl. Optim. 58*, 4 (Oct. 2018), 1823–1843. 1

[ZLQW24] ZHANG S., LI H., QU H., WANG Y.: Adavis: Adaptive and explainable visualization recommendation for tabular data. *IEEE Transactions on Visualization and Computer Graphics 30*, 9 (2024), 5923–5938. doi:10.1109/TVCG.2023.3316469. 10

[ZWC*18] ZHAO X., WU Y., CUI W., DU X., CHEN Y., WANG Y., LEE D. L., QU H.: SkyLens: Visual Analysis of Skyline on Multi-Dimensional Data. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (Jan. 2018), 246–255. doi:10.1109/TVCG.2017.2744738. 3

# 3

# Detailed Methodology

This chapter expands the methodology of the paper, as Figure 6 gives the overview, showing the three groups that were involved in the study. Five collaborators shaped the framework through the initial discussions and the visual design (*discover & design*), a shared study design then carried the framework into a qualitative evaluation with ten experts and a quantitative evaluation with 57 non-experts (*evaluate*), and their combined results fed the discussion (*reflect*).

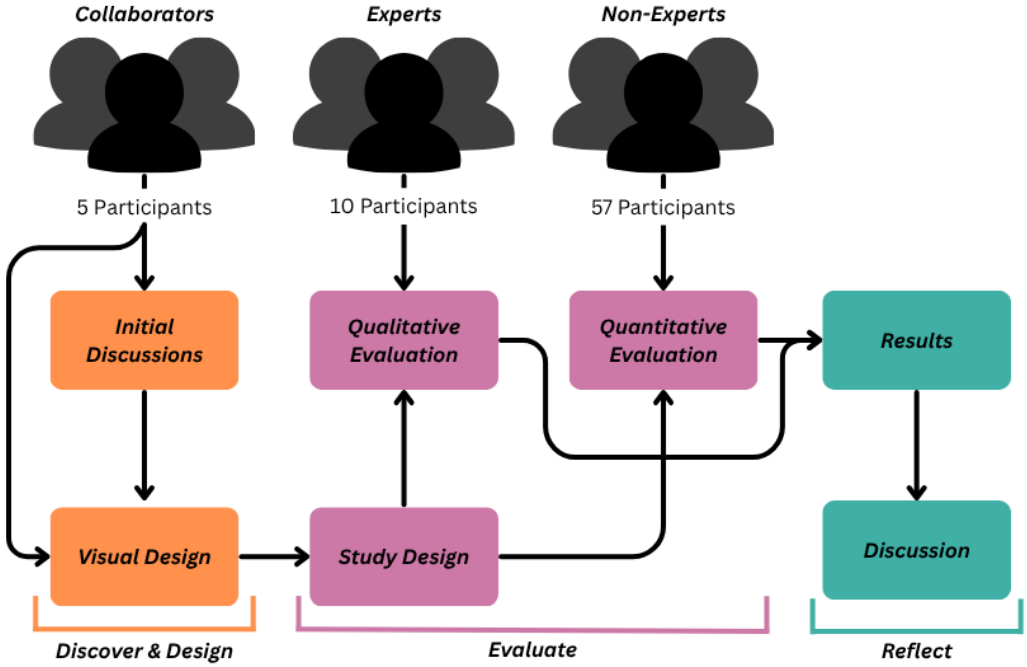



Figure 6: Overview of the study methodology. Three participant groups drive three phases: five collaborators inform the initial discussions and the visual design (*discover & design*); a shared study design leads into the qualitative evaluation with ten experts and the quantitative evaluation with 57 non-experts (*evaluate*); and the combined results feed the discussion (*reflect*).

## 3.1 Initial Interviews Pointing Directions

Before any design work, we ran a round of semi-structured exploratory interviews with members of both collaborator groups, the company's R&D specialists and its quantitative analysts and consultants, to characterise the domain and elicit requirements. The paper summarises their outcome into three design goals; this section presents the interviews themselves and the full feedback that shaped the design.

Each interview moved through four progressive phases. The first, an *introduction to the domain of investment portfolios*, outlined the participant's role and asked about their decision

workflow: what people look at most when making such decisions, which aspects are obvious to a professional but not to an outsider, and whether the workflow is similar across cases. The second, an *introduction to data visualisation*, asked about their existing exposure: how they had worked with visualisation before, how useful they find dashboards for justifying results, and their common frustrations. The third phase joined the two, asking where visualisation could serve portfolio analysis, how they currently deal with many variables at once, and how they inspect trade-offs. The final phase was a discussion about *high-dimensional data*: after asking what experience they had with it, we presented a prototype with several candidate idioms, a parallel coordinates plot, a scatterplot matrix, a radar chart, a 3D scatterplot, and a stacked bar chart.

The remainder of this section reports what came out of these sessions, starting with the general feedback and then with chart-specific notes.

### 3.1.1 General Feedback

Interactivity is essential, as there was a strong consensus in the first validation that actions such as selection and filtering are useful in discovering patterns between the objective space and decision space. Furthermore, some visualisations suffer from clutter, and these actions can help with showing only the desired data. Further decluttering can be done by offering the possibility of disabling other features, such as labels for data points or a simplified overview. Another action that can help with decluttering and was suggested is excluding points from charts.

All of the participants have mentioned that interpretability should be prioritised, and we should not focus on showing all the data we have gathered from the optimisation. Therefore, we should start with an overview from which we can dive deeper if the user demands more detail. This links directly to the idea of progressive disclosure, where we start with essential information, and we reveal more complex information when needed or requested. Furthermore, this is similar to the current workflow of users, where they start from a frontier that is unrefined and, through an iterative process, they apply constraints to the optimisation to arrive at results closer to what is desired. This leads to the request to compare multiple Pareto frontiers, to show progress from unrefined to refined results.

A small detail raised by all participants was the inclusion of a benchmark portfolio. Most users have a benchmark that they use to formalise their expectations for a portfolio, which can then be compared against the optimisation's results.

Participants rejected dimensionality reduction for the objective space, such as PCA and t-SNE. They argued that clusters would be too abstract for the users to use them in their decision-making. Therefore, after some discussion, we had the idea of predefined clusters in the decision space to better interpret the results in the objective space, so as to distinguish quickly between points that have very similar scores. Furthermore, other variants for enhancing the utility and interpretability of dimensionality reduction in the objective space can be explored while trying to use prior knowledge to reduce the abstraction of clusters.

Lastly, uncertainty was more valuable for the technical participants, while the non-technical participants rejected or did not know how to make use of it. Although this was only discussed and no supporting visualisations were shown.

### 3.1.2 Chart Specific Feedback

The stacked bar chart for showing the decision space was well received, and it solved the issue with clutter that was present in their current dashboard, which is a table view. Although they said that they would prefer to switch to the table view, in case users are more familiar with a numerical view. Furthermore, another point of discussion was around clustering categories in the decision space, as the overview showed the percentage of global equity and not regional. Therefore, a way to show different levels of this hierarchy should be done to build upon the details-on-demand aspect.

For the Parallel Coordinate Plot, participants found it easy to see trends between the different metrics. Furthermore, the added interactivity of disabling, moving, and filtering axes was another positive for the users. The ability to easily declutter and constrain the axes was something that was already requested by users, where they would seek to only look at points that were within a threshold.

For the Radar Chart, people preferred the representation of points compared to the previous plot, but the impact of clutter was stronger. Therefore, the consensus was that it was better used to compare a selection of a few points, as the comparison was more understandable than the PCP. It is also noted that most of the participants were already familiar with it, as it is used quite commonly in showing statistics in popular sports.

The 2D scatter plot was already used in their original visualisation framework, and most of the feedback was focused on specific details, such as labels, showing optimality direction, and horizontal/vertical lines to help visualise points. The 3D scatter plot was very contentious, as most participants had trouble with it. The principal issues were focused on its interpretability. Manipulating the plot to see the frontier better was useful, but it still caused a lot of confusion. Furthermore, showing the shape of the frontier using a surface was useful, but could be improved to be more interpretable, as was noted by the participants.

The SPLOM was received negatively, as it was an instant information overload for the participants. They mentioned that they understand its usefulness, but they would prefer to have less information shown at a time. Therefore, a solution was proposed by one of the participants, where you could have two scatter plots side by side to achieve a similar purpose to the SPLOM.

There were some proposals for charts made during the interviews. This included using a heat map to show the correlation between the objective and the decision space, and to also show the pressure of portfolios on constraints. Another proposal was a triangle plot using pie glyphs, where the location in the triangle shows the objective space, while the pie charts show the decision space. We later discovered that this can be expanded using RadViz, and we can easily expand the number of dimensions. This can be seen in Figure 7.

## 3.2 The Full Dashboard

This section introduces the complete interface shown in Figure 8. While the paper described the base visual encodings, here we detail the specific options for each idiom and the user feedback that shaped them.

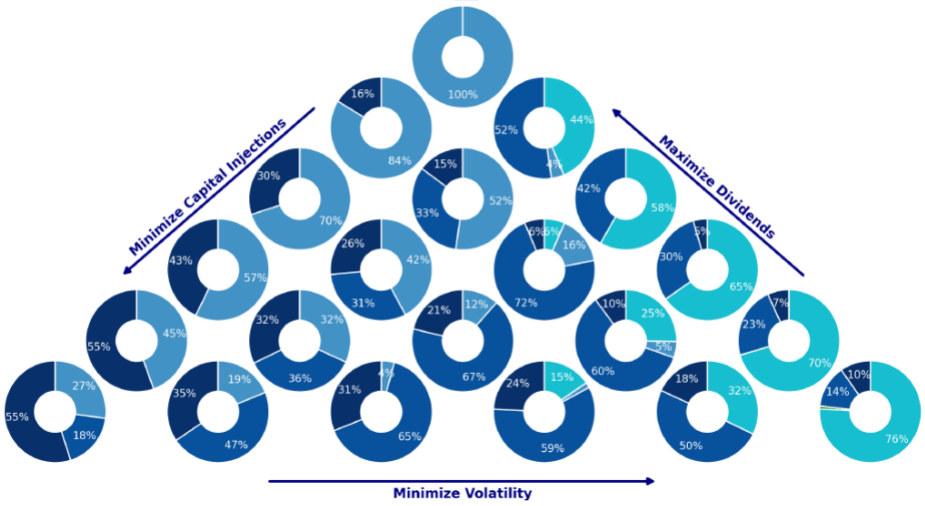



Figure 7: The ternary grid plot proposed by one of the analysts.

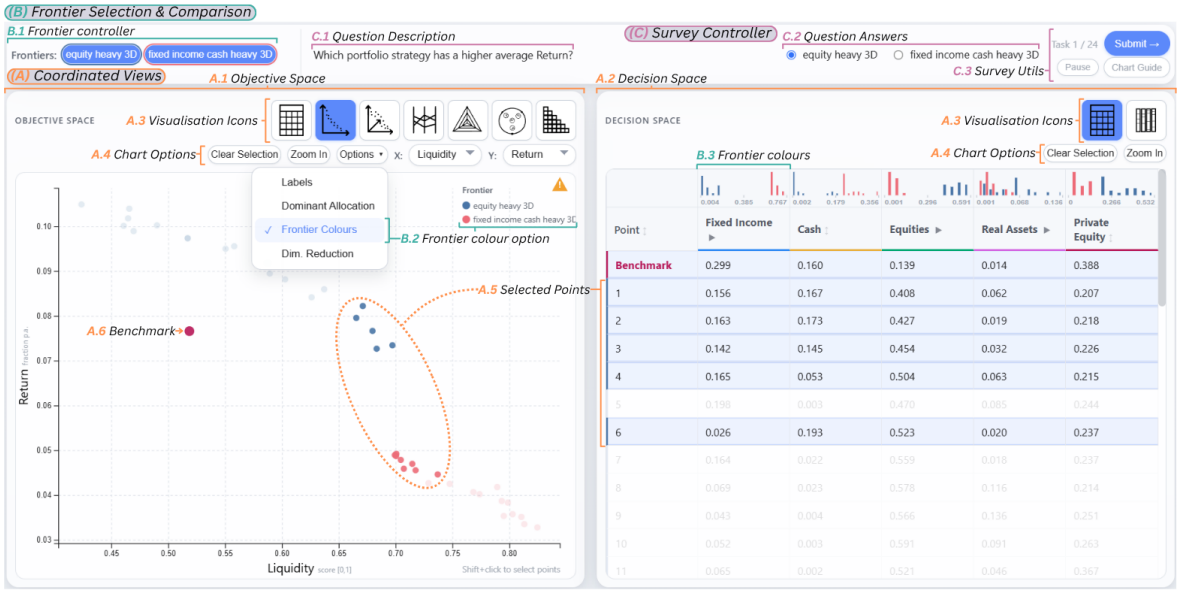



Figure 8: The full dashboard: the objective-space panel (left) and decision-space panel (right), the idiom selectors above each panel, the frontier controller and benchmark legend (top left), and the task controls used during the study.

### 3.2.1 Shared Foundation

Both panels rest on the same skeleton. The objective-space panel (A.1) and the decision-space panel (A.2) each show the same solutions through one idiom at a time, chosen from the row of icons above the panel (A.3), in which every icon is a miniature preview of its chart. Neither panel opens on a default idiom, the panel shows only a prompt until the user picks one. We chose this so the layout does not steer users toward any particular view, letting the evaluation record unbiased preferences, at the cost of an extra interaction before any data is visible.


Figure 9: The shared selection mechanics across both panels: solutions selected in the objective-space scatterplot are highlighted in the decision-space bar chart (top), and zooming in filters both panels down to the selected solutions (bottom).

Hovering or selecting a solution in one panel highlights it in the other (Figure 9), so a trade-off can be traced to the decision variables that produce it. Every idiom supports selecting individual solutions, and the shared chart options hold the actions around a selection: *zooming in* keeps only the selected solutions, so a crowded region can be studied alone, *zooming out* restores the full set, and *clear selection* drops the selection. A per-idiom *options* drop-down menu holds the settings described below.

Two persistent elements complete the foundation. The *benchmark* solution is drawn in crimson in every idiom (Figure 8 A.6), an anchor the analysts explicitly requested, representing the allocation a client already holds or expects. The *frontier controller* at the top left (Figure 8 B.1) lists the loaded frontiers with the colours of its legend (Figure 8 B.2); frontiers can be toggled on and off, and at least one always needs to stay visible. Hovering over one shows the constraints that produced it (Figure 10), so runs can be easily differentiated.

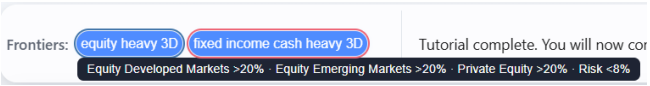



Figure 10: The frontier controller with its legend; hovering a frontier reveals the constraints that produced it.

### 3.2.2 Objective-Space Idioms

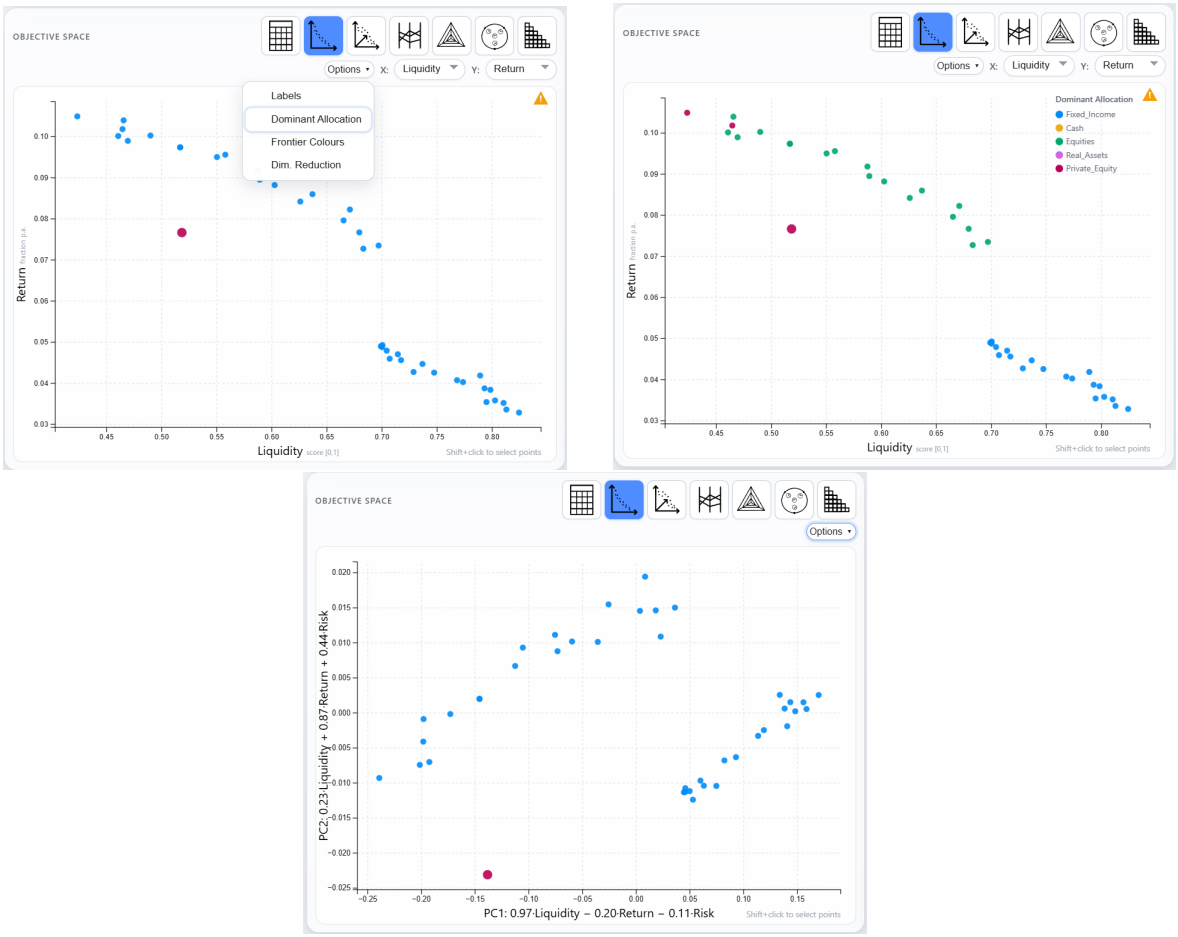



Figure 11: The 2D scatterplot options: the options menu (left), the dominant-variable colouring that reveals composition differences between nearby solutions (middle), and the PCA mode with the linear combination behind each axis displayed on its label (right).

**2D scatterplot.** Two dropdowns choose which objectives occupy the axes, and a warning label, added after interview feedback, reminds the user that the remaining objectives are hidden. Point labels can be toggled, a compromise between those who wanted them and those who found them noisy. The *dominant variable* option colours each point by the decision variable holding the largest share of that solution, a suggestion from the interviews, as it is a quick way to spot solutions that sit close in the objective space but differ in composition. A *dimensionality reduction* option projects all objectives onto two PCA axes. The initial interviews flagged such projections as too abstract; the linear combination behind each axis

is displayed so its meaning can be traced back to the original objectives. Selection works per point or with a lasso. Figure 11 shows the options menu, the dominant-variable colouring, and the PCA projection with the composition of each axis displayed.

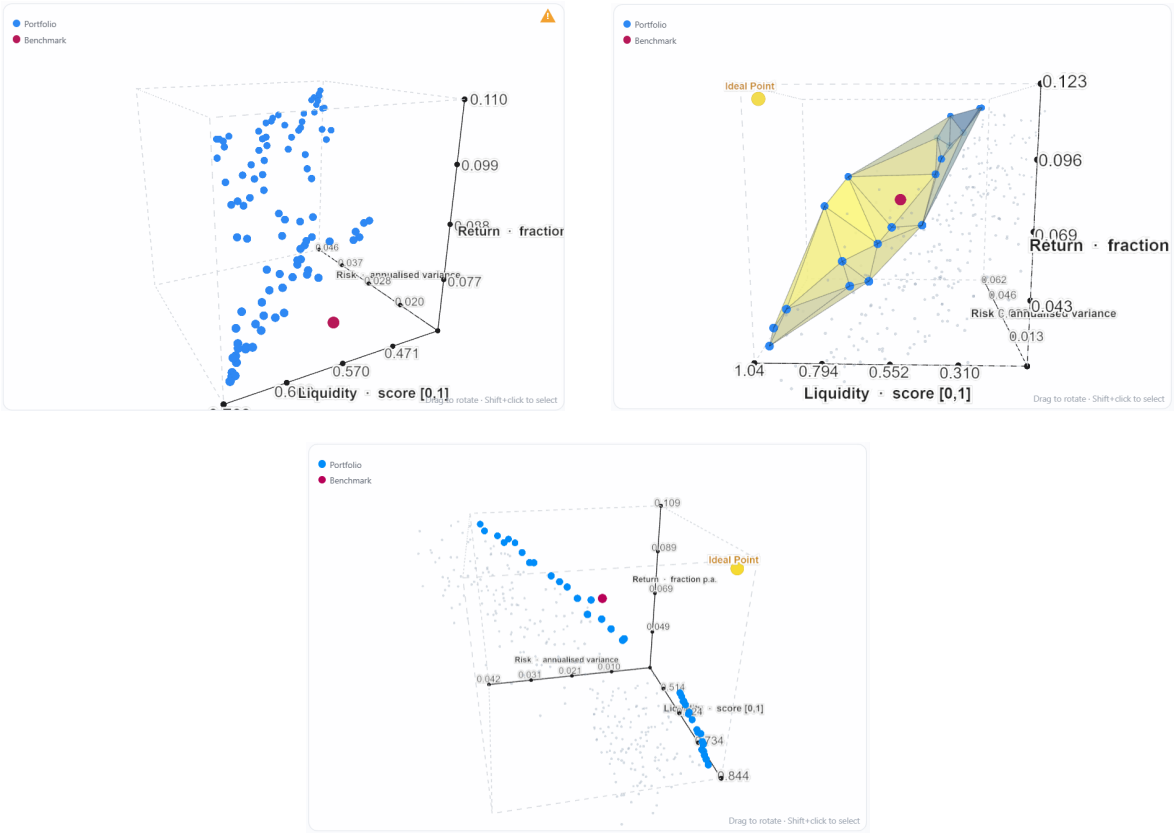



Figure 12: The 3D scatterplot plain, with the surface option colouring the front by proximity to the ideal point, and with the dominated cloud and ideal point displayed.

**3D scatterplot.** Requested by the analysts as the obvious extension to a third objective, as seen in Figure 12, and kept despite the theory that discourages unjustified 3D [Mun14]. The plot rotates by dragging, using an orbital rotation around the centre. A *surface* option draws a sheet over the front, built as a Delaunay triangulation of the points. With every displayed axis normalised to $[0, 1]$, the ideal point $\mathbf{z}^*$ takes the value 1 on maximised axes and 0 on minimised ones. Each triangle is coloured by the Euclidean distance from its centroid $\mathbf{c}$ to that point, normalised across the mesh and mapped onto the colourblind-safe Cividis gradient, so the regions nearest the ideal point are the brightest. A *dominated cloud* option scatters sample-dominated points below the front, an interviewee suggestion that makes the frontier's shape easier to read. The axis warning label and the options of the 2D scatterplot are carried over.

**Parallel coordinates.** One of the most liked views in the first-round interviews, as seen in Figure 13. A *dimensions* menu chooses which objectives are drawn; axis labels can be dragged to be reordered, the mitigation for the idiom's known weakness that only neighbouring axes compare easily. Every minimised axis is flipped so the best value sits at the top, making the ideal solution a straight line across the top of the chart. Filters can be brushed on

several axes at once, the behaviour analysts had already asked for in their own workflow. The dominant-variable and frontier-colour options are shared with the scatterplots.

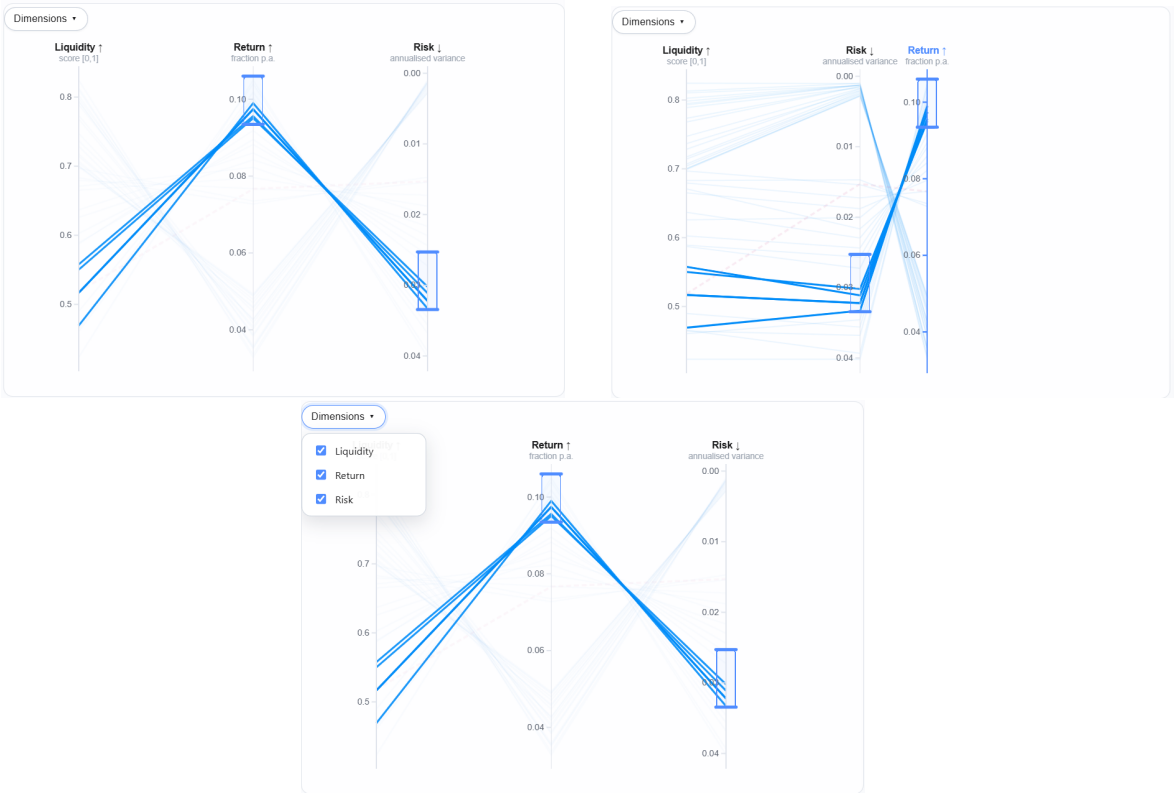



Figure 13: Parallel coordinates interactions: brushed axis filters constraining the visible set (left), dragging an axis label to reorder the axes (middle), and the dimensions menu choosing which objectives are drawn (right).

**Radar chart.** The radar chart, seen in Figure 14, shows the same information as the parallel coordinates plot but puts the axes into a circle, so each solution becomes a polygon instead of a line. The trade-off is deliberate, as a radial layout reads less precisely than aligned linear axes [CM84, Mun14], and the enclosed area of the polygon depends on the order of the axes, so the idiom is weaker than parallel coordinates for reading exact values. What results is the closed shape, which the eye matches quickly, so a few solutions can be compared. Analysts also noted that the shape is familiar from sports and game statistics, which lowers the entry barrier. Decision accuracy across these idioms is broadly comparable [DIPJ22], so it earns a place alongside the PCP. Because all solutions at once become cluttered, as the interviews warned, it works best on a small selection. It carries over the same options and interactions as the parallel coordinates plot.

**RadViz with pie glyphs.** Originating from the ternary grid plot with pie glyphs that one of the analysts had built themselves. Each objective anchors a position on a circle and each solution is placed at the weighted average of the anchors, so a solution sits nearest the objectives it scores well on. The pie glyph at that position shows the composition of the solution's decision variables, so the two spaces are read together in one mark. Formally, for

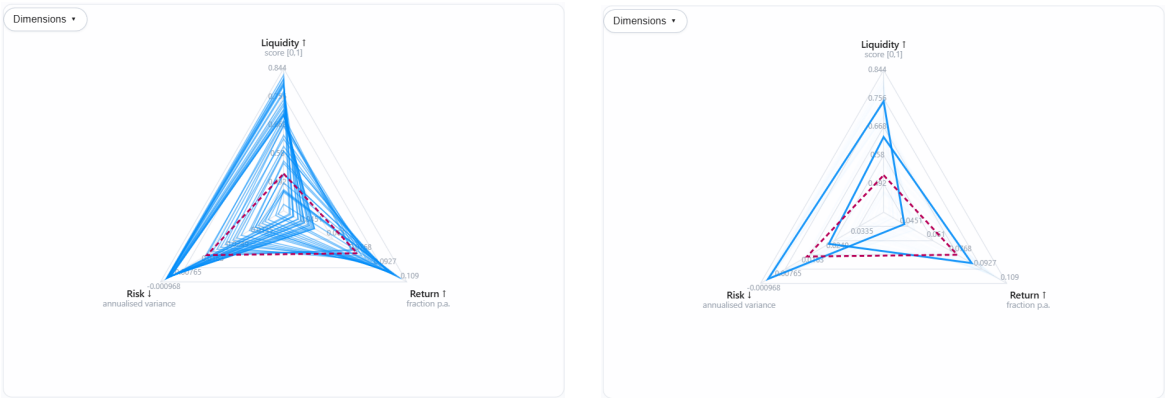



Figure 14: The radar chart with no filters, where every solution overlaps into clutter (left), and after filtering, where a handful of shapes can be compared directly (right).

$d$ objectives the anchors $\mathbf{a}_1, \ldots, \mathbf{a}_d$ are spaced evenly around the unit circle,

$$\mathbf{a}_j = \left( \cos \frac{2\pi(j-1)}{d},\ \sin \frac{2\pi(j-1)}{d} \right), \qquad j = 1, \ldots, d, \tag{4}$$

and every objective value is rescaled to $v_j \in [0, 1]$, so that all objectives pull with comparable strength. A solution is then placed at the weighted average of the anchors,

$$\mathbf{p} = \frac{\sum_{j=1}^{d} v_j \, \mathbf{a}_j}{\sum_{j=1}^{d} v_j}, \tag{5}$$

which draws it toward the anchors of the objectives it scores highest on. After placement, all points are scaled so the outermost one reaches the circle. Because the weighted average can stack glyphs on top of one another, a *spread* option runs a force-directed relaxation that pushes overlapping glyphs apart while keeping them inside the anchor polygon. Each glyph is treated as a node of fixed radius $r$, scaled so the glyphs fill the available area with small gaps. In every iteration of the simulation, a node $i$ at current position $\mathbf{x}_i$ is pulled back toward its true position $\mathbf{p}_i$ by a spring force acting on its velocity $\mathbf{v}_i$,

$$\mathbf{v}_i \leftarrow \mathbf{v}_i + k\,(\mathbf{p}_i - \mathbf{x}_i), \qquad k = 0.4, \tag{6}$$

while a collision force separates any two nodes closer than twice the glyph radius, enforcing $\|\mathbf{x}_i - \mathbf{x}_j\| \geq 2r$, and any node pushed outside the anchor polygon is projected back inside. The simulation runs for a fixed number of iterations, so the layout settles into spread positions that stay as close to the true ones as the collisions allow. Clicking a point enlarges its pie, while clicking a group inside the pie breaks it into its individual variables, and both single and lasso selection are available. Figure 15 shows the spread option and the glyph drill-down.

**Correlation heatmap.** A matrix of Spearman rank correlations between every pair of decision variables and objectives, chosen over Pearson because the values are not normally distributed and the relationships are not consistently linear. The colour gradient runs from dark blue (most negative) through to yellow (most positive), deliberately emphasising the two extremes the analysts said were the relevant ones. A vertical red separator with bottom labels splits the objective-to-objective block from the block relating decision variables to everything

else. A tooltip reports exact values, which matters in the multi-frontier case where one matrix is drawn per frontier and the cells become too small to label (Figure 16).

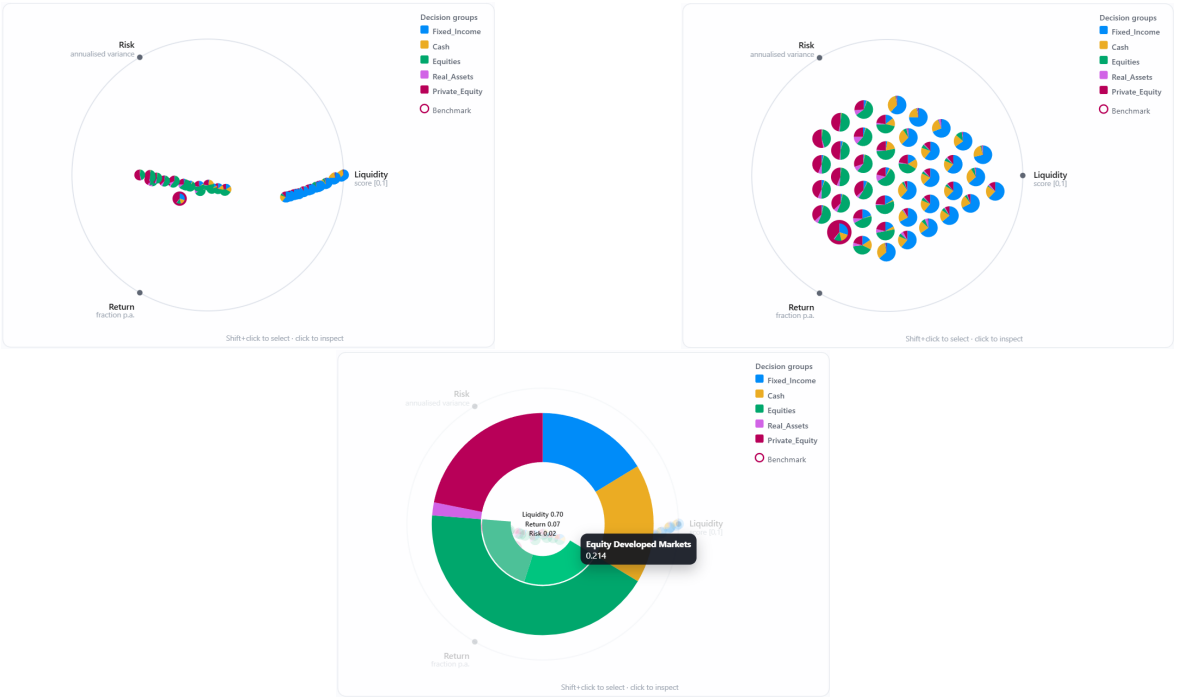



Figure 15: RadViz interactions: overlapping glyphs at their true positions (left), the spread option separating them with the force-directed relaxation (middle), and a clicked glyph enlarged with a group broken into its individual variables (right).

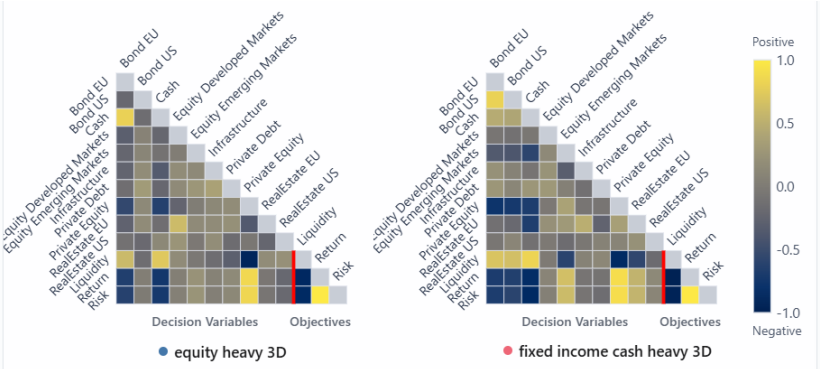



Figure 16: The correlation heatmap with two frontiers loaded: one matrix is drawn per frontier, side by side, each labelled with its frontier.

### 3.2.3 Decision-Space Idioms

**Stacked bar chart.** The compact overview of the decision space the interviews demanded. Each solution is one bar split into its groups of decision variables, drawn with a colourblind-safe palette of up to twelve colours. Bars follow the order in which the optimiser produces the solutions, which makes the allocation shift gradually along the front, so the analyst reads a transition rather than comparing distant bars. When a precise comparison is needed, zooming

brings the selected bars together. Clicking a group expands it into its variables (Figure 17), shown as well in the tooltip, and individual bars are selectable as everywhere else.

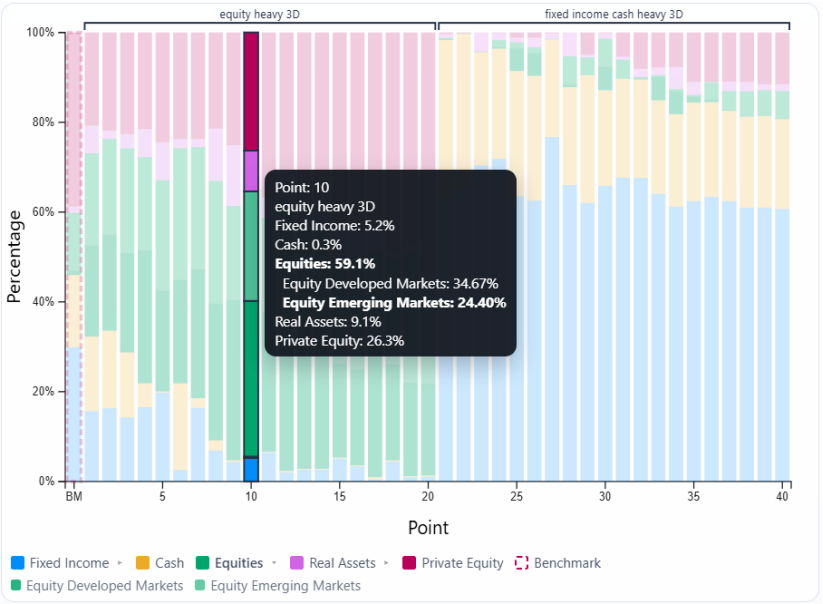



Figure 17: The stacked bar chart with a variable group expanded into its constituent variables.

**Table.** Closest to the analysts' existing tool, and available in both panels. Each objective column carries a green *max* or red *min* label for its optimisation direction (earlier arrow icons were replaced after feedback). A small distribution is drawn at the top of every column, and a range can be brushed directly on it to select the solutions inside. Rows are selectable, columns sortable, and in the decision space the variable groups expand to their members, keeping the same grouping used by the bar chart and the pie glyphs. Figure 18 shows several distribution filters active at once.

### 3.2.4 Comparing Multiple Frontiers

Where an idiom can colour its marks, the frontier colours from the controller show: the scatterplots, parallel coordinates, radar, and RadViz (Figure 19). The two idioms that cannot overlay solve it differently: the correlation heatmap draws one matrix per frontier side by side, as seen in Figure 16. The stacked bar chart, whose colours already encode the variable groups, keeps each frontier's bars together and marks the runs with bracket annotations above them, as seen in Figure 17. The table colour-codes its column distributions and adds a coloured band to each row, as seen in Figure 18.

### 3.2.5 Excluded Features

Not everything that was prototyped survived. The *SPLOM* was cut after the first-round interviews received it as an immediate information overload. The 3D scatterplot originally supported *brushing filters on its axes*, removed after pilot sessions showed that dragging to rotate clashed with dragging on an axis. The same view was first rendered as *SVG in a 2D setting*, which produced a hollow-face effect that made the front look inside out, and this is what moved the implementation to WebGL. Finally, the abstract *dimensionality reduction* of

the objective space survives only as the optional PCA mode of the 2D scatterplot described above, with its axis compositions displayed.

| Point | Fixed Income | Cash | Equities | Real Assets | Private Equity |
|---|---|---|---|---|---|
| Benchmark | 0.299 | 0.160 | 0.139 | 0.014 | 0.388 |
| 1 | 0.156 | 0.167 | 0.408 | 0.062 | 0.207 |
| 2 | 0.163 | 0.173 | 0.427 | 0.019 | 0.218 |
| 3 | 0.142 | 0.145 | 0.454 | 0.032 | 0.226 |
| 4 | 0.165 | 0.053 | 0.504 | 0.063 | 0.215 |
| 5 | 0.198 | 0.003 | 0.470 | 0.085 | 0.244 |
| 6 | 0.026 | 0.193 | 0.523 | 0.020 | 0.237 |
| 7 | 0.164 | 0.022 | 0.559 | 0.018 | 0.237 |
| 8 | 0.069 | 0.023 | 0.578 | 0.116 | 0.214 |
| 9 | 0.043 | 0.004 | 0.566 | 0.136 | 0.251 |
| 10 | 0.052 | 0.003 | 0.591 | 0.091 | 0.263 |

Figure 18: The table with ranges brushed on several column distributions at once, filtering the rows to the solutions inside every brushed interval.

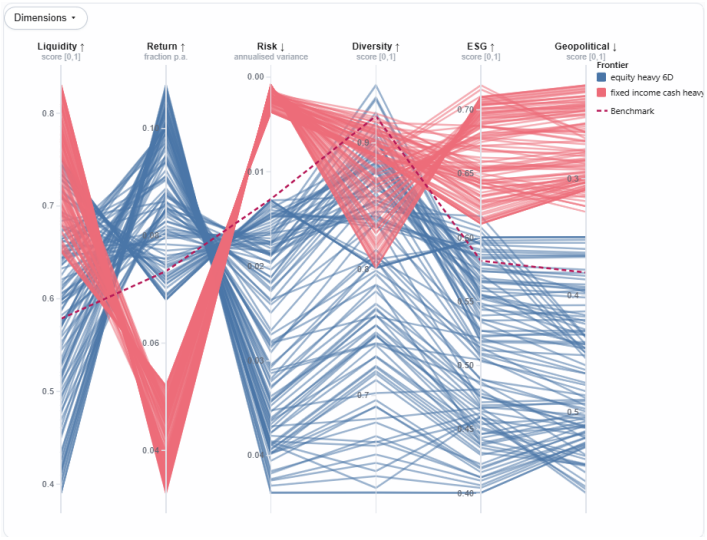



Figure 19: Multi-frontier comparison in the parallel coordinates plot: each run keeps its controller colour, with the benchmark in crimson.

## 3.3 Survey Setup

The methodological grounding of the evaluation is covered in the paper's study design. This section fills in the additional details behind that design.

### 3.3.1 Datasets and Their Generation

Both studies run on synthetic data, which lets us shape frontiers with specific outliers and edge behaviour. Frontiers are generated with the NSGA-II evolutionary algorithm [DPAM02], run with a population of 20 for 100 generations and a fixed seed. Candidates are sampled from a Dirichlet distribution so that every initial allocation $\mathbf{w}$ lies on the simplex. Constraints enter NSGA-II as inequality constraints, after which the final non-dominated set is exported alongside its constraint metadata.

The *portfolio* problem allocates capital over ten assets grouped into classes (developed- and emerging-market equity, US and EU bonds, US and EU real estate, cash, private debt, infrastructure, and private equity), each with set expected returns, volatilities, correlations, and liquidity, ESG, and geopolitical scores. Most objectives are linear in the allocation: Return, Liquidity, ESG, and Geopolitical risk each take the form

$$f_{\text{obj}}(\mathbf{w}) = \mathbf{w}^\top \mathbf{s}_{\text{obj}}, \tag{7}$$

where $\mathbf{s}_{\text{obj}}$ holds the per-asset values of that objective. Risk is

$$f_{\text{risk}}(\mathbf{w}) = \mathbf{w}^\top \Sigma\, \mathbf{w}, \tag{8}$$

with $\Sigma$ the asset covariance matrix built from the per-asset volatilities and their correlation matrix, and Diversity is the normalised inverse Herfindahl–Hirschman index of the allocation,

$$f_{\text{div}}(\mathbf{w}) = \frac{1 - \sum_i w_i^2}{1 - 1/n}, \tag{9}$$

which reads 1 for a perfectly even allocation and 0 for full concentration in one asset. Risk and Geopolitical risk are minimised and the rest maximised. The three-objective version uses Risk, Return, and Liquidity, and the six-objective version adds ESG, Diversity, and Geopolitical risk.

The *dietary* problem mirrors this structure while swapping financial vocabulary, so participants without financial literacy can work on tasks of identical form. It allocates the calories of a diet over twelve food items (chicken breast, tofu, eggs, sweet potato, white rice, oats, almonds, olive oil, avocado, spinach, berries, and broccoli), with linear per-calorie scores for Nutrient Density, Cost, Preparation Time, Carbon Footprint, and Satiety, and the same normalised inverse-HHI Diversity. The three-objective version scores Nutrient Density, Cost, and Preparation Time, and the six-objective version adds Carbon Footprint, Satiety, and Diversity.

For each problem and dimensionality, we generate several frontiers by varying the constraints, listed in Table 1. The fixed *benchmark* is built from these frontiers by averaging the decision weights across all of a problem's Pareto solutions and evaluating it with the full objective function.

Table 1: The generated scenario frontiers and their constraints, as recorded in each exported dataset. Constraints are identical across the 3D and 6D versions of a scenario, except where noted. Variable bounds are allocation fractions; objective bounds are in the objective's own units.

| Problem | Frontier | Constraints |
|---|---|---|
| Portfolio | Unconstrained | none |
| Portfolio | Equity heavy | Eq. Developed $\geq 0.20$; Eq. Emerging $\geq 0.20$; Private Equity $\geq 0.20$; Risk $\leq 0.08$ |
| Portfolio | Fixed income & cash heavy | Bond US $\geq 0.20$; Bond EU $\geq 0.20$; Cash $\geq 0.20$; Private Debt $\geq 0.20$ |
| Portfolio | Real assets heavy | Real Estate US $\geq 0.20$; Real Estate EU $\geq 0.20$; Infrastructure $\geq 0.20$ |
| Diet | Budget | Sweet Potato $\leq 0.35$; White Rice $\leq 0.30$; Oats $\leq 0.30$; Almonds $\leq 0.10$; Olive Oil $\leq 0.10$; Avocado $\leq 0.10$; Cost $\leq 0.25$ |
| Diet | Diverse | Tofu $\geq 0.05$; Sweet Potato $\leq 0.35$; Oats $\in [0.05, 0.30]$; Almonds $\leq 0.25$; Olive Oil $\leq 0.10$; Avocado $\in [0.05, 0.25]$; Berries $\geq 0.05$; Diversity $\geq 0.75$ (6D only) |
| Diet | High protein | Chicken Breast $\geq 0.25$; Sweet Potato $\leq 0.35$; White Rice $\leq 0.30$; Oats $\leq 0.30$; Almonds $\leq 0.25$; Olive Oil $\leq 0.10$; Avocado $\leq 0.25$; Nutrient Density $\geq 0.70$ |
| Diet | Low carb | Sweet Potato $\leq 0.05$; White Rice $\leq 0.05$; Oats $\leq 0.05$; Almonds $\leq 0.25$; Olive Oil $\leq 0.10$; Avocado $\leq 0.25$; Cost $\leq 0.50$ |
| Diet | Vegan | Chicken Breast $\leq 0.01$; Eggs $\leq 0.01$; Sweet Potato $\leq 0.35$; White Rice $\leq 0.30$; Oats $\leq 0.30$; Almonds $\leq 0.25$; Olive Oil $\leq 0.10$; Avocado $\leq 0.25$; Nutrient Density $\geq 0.65$ |

### 3.3.2 Task Blocks and Their Organisation

Both studies intertwine structured tasks with open exploration, following the layered insight- and task-based methodology of Gomez et al. [GGZL14]. Participants work through eight blocks of three tasks, as seen in Figure 20: four blocks on three-objective frontiers and four on six-objective ones, with half of the blocks completed in the dashboard condition (all views available) and half in the table-only baseline, in counterbalanced order so that neither condition benefits from always coming second.

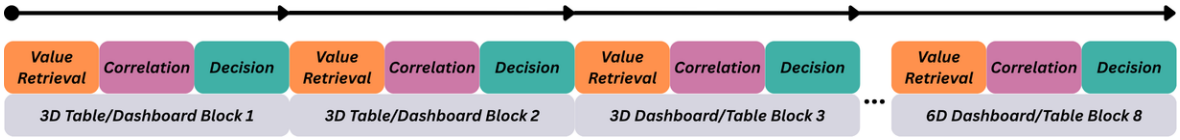



Figure 20: Task flow of the quantitative evaluation: eight blocks in sequence, each running the same three task types in fixed order (value retrieval, correlation, decision; their comparison counterparts in the multi-frontier study). The first four blocks are completed in 3D and the last four in 6D, with the table-first and dashboard-first orders counterbalanced between participants.

On a single frontier, each block runs the three task types drawn from the decision-support taxonomy of Dimara et al. [DBD18], in fixed roles: a *value retrieval* task ("Select all portfolios

where Return is above 0.09"), a *correlation* task ("Which decision variable is most strongly correlated with Return, negatively?"), and a *decision* task ("Which portfolio do you think offers the best overall trade-off across all objectives?"). The comparison blocks have the same structure of read–compare–choose progression across two frontiers: a *simple comparison* ("Which portfolio strategy has a higher average Return?"), a *most distinct objective* task ("In which objective do the two portfolio strategies differ the most on average?"), and a *decide on frontier* task ("Which portfolio strategy performs better overall across all objectives?"). Each question targets one of three types: the objective space, the decision space, or a cross-space relation. This is done so that both panels and the link between them are used. Every question of every setup can be seen in Appendix A.

### 3.3.3 Scoring and Telemetry

Every task yields a score in $[0, 1]$ alongside its completion time, with partial credit wherever the answer format allows. Value retrieval is scored as the Jaccard overlap between the selected and the correct set. The single-frontier decision task scores a selection by the proximity of its points to the normalised ideal point, so selecting the closest point scores one. Correlation and most-distinct-objective use a rank-based score that rewards near misses in proportion to the strength of the chosen candidate, while simple comparison and decide-on-frontier are binary, awarded for naming the frontier with the higher objective mean and the higher mean normalised value respectively. Throughout, the dashboard logs a timestamped record of every interaction and the time spent in every chart. A view is credited to a task when it accounts for at least 25% of total task time. This threshold is robust, as changing it from 15% to 50% yields the exact same verdicts.

### 3.3.4 Onboarding, Tutorial, and Session Control

Each session opens with a twelve-step guided tutorial that walks through the header controls, the two linked spaces, the shared interactions (hovering, selecting, zooming), and every idiom in turn, using highlighting and shadowing to direct attention and opening the relevant menus live (Figure 21, top). During the tasks, a built-in cheat sheet stays available (Figure 21, bottom), summarising each view with an example image; hovering over an image plays a short animated clip of interacting with that view, so a participant can see, for instance, how its filtering works without leaving the task. A pause button, added after pilot feedback, stops the task timer for breaks without affecting recorded times. Sessions close with the four-item UMUX questionnaire [Fin10], chosen over SUS and UEQ-S for its conciseness and its focus on pragmatic rather than hedonic quality [SKT23], plus an open-ended question inviting anything the scales missed.

### 3.3.5 Participants and the Expert Variant

The quantitative study contributes 57 non-expert participants, split between the research questions: 29 worked on a single frontier and 28 on the comparison of two. Each half is split again by dataset, with financially literate participants receiving the portfolio data and the remainder the dietary twin. The qualitative study adds ten domain experts, largely insurance and pension professionals, who work through a compressed version of the same survey while thinking aloud. For the experts, the structured tasks act as familiarisation. Also, at the end of each block, the decision task acts as open exploration. Instead of looking for a single 'best' trade-off, participants explore the data freely and 'think aloud' to make an intuitive choice. Experts finish with the same UMUX questionnaire and open question.

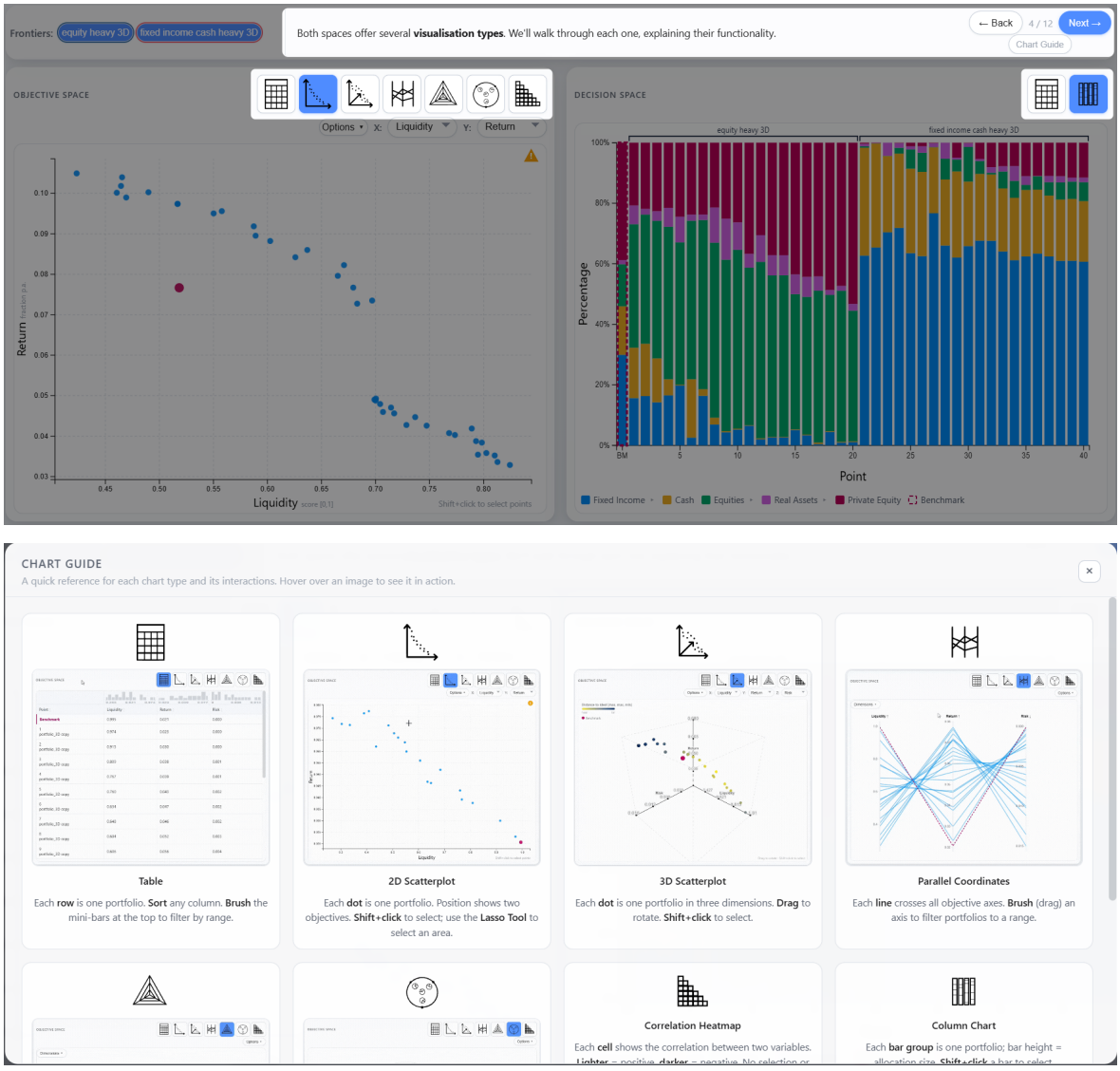



Figure 21: Onboarding support: a step of the guided tutorial, which spotlights the component being explained while shadowing the rest of the interface (top), and the built-in cheat sheet summarising every view (bottom), where hovering over an image plays an animated clip of interacting with that view.

# 4

# Detailed Results

## 4.1 Full Quantitative Results

This section reports every estimate behind the paper's results. Table 4 lists the complete set of estimates per hypothesis, scope, and metric.

### 4.1.1 Usability and the Toolbox Advantage (H1, H2)

The usability precondition holds overall, with a pooled UMUX of 73.8 [69.2, 77.3] across the 57 participants, clear of the established average of 68 [LS18]. For H2, the single-frontier study confirms the advantage of a choice of views: accuracy rises by +0.061 [+0.015, +0.109] over the table baseline at equal time (0.92× [0.75, 1.13]). On the comparison the advantage disappears, −0.023 [−0.075, +0.026], and costs a third more time (1.33× [1.16, 1.59]). The robustness checks are identical: the dimensionality check is inconclusive (+0.007 [−0.064, +0.082] single, −0.036 [−0.138, +0.053] multi), there is no block-to-block practice gain (−0.012 [−0.059, +0.045] and −0.008 [−0.058, +0.044]), and the advantage is order-invariant, so the dashboard-table gap is the same whether the dashboard came first or second (−0.019 [−0.114, +0.081] and −0.018 [−0.119, +0.084]).

Table 2: Per-view usage, accuracy, and time across all tasks and both studies, split by panel. Usage counts tasks in which a view dominated its panel's viewing time; accuracy is the mean task score in those tasks; time is the mean time to answer.

| Panel | View | Usage (tasks) | Accuracy | Time (s) | Ranked by use |
|---|---|---|---|---|---|
| Objective | Table | 180 | 0.77 | 90 | 1 |
| Objective | Corr. heatmap | 103 | 0.84 | 128 | 2 |
| Objective | 2D scatter | 79 | 0.78 | 192 | 3 |
| Objective | 3D scatter | 71 | 0.66 | 102 | 4 |
| Objective | PCP | 71 | 0.65 | 82 | 5 |
| Objective | RadViz | 70 | 0.54 | 108 | 6 |
| Objective | Radar | 59 | 0.57 | 99 | 7 |
| Decision | Table | 260 | 0.71 | 94 | 1 |
| Decision | Stacked bar | 169 | 0.73 | 110 | 2 |

### 4.1.2 View Specialisation in Full (H3)

Figure 22 shows usage against performance for every view, task, and dimensionality. Two tables unpack it. Table 2 pools across all tasks: in the objective panel the table is the most used view (180 tasks) but only the third most accurate (0.77), while the correlation heatmap has the second-highest usage with the highest accuracy (0.84). The per-view time column, shows the 2D scatter to be accurate but slow (192 s on average) and the parallel coordinates fast but inaccurate (82 s, 0.65). In the decision panel the table dominates usage (260 against 169 tasks) while the bar chart is marginally more accurate (0.73 against 0.71).

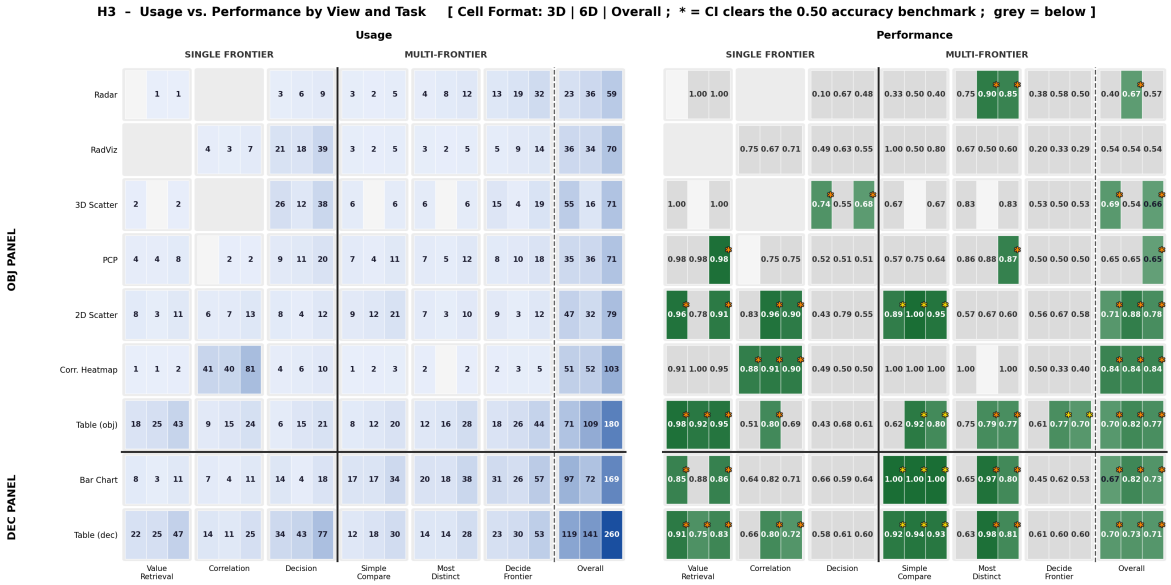



Figure 22: **H3 — usage vs. performance by view and task.** Left: the number of tasks where a view was used more than 25% of total panel usage. Right: mean accuracy per view and task. (95% Wilson interval), split by 3D, 6D, and Overall. Cells whose 95% Wilson lower bound clears the 0.50 accuracy benchmark are shaded green and marked with an asterisk; grey cells fall below it. Yellow asterisks mark binary tasks where 0.50 is chance level. Orange asterisks mark graded tasks where chance is lower and 0.50 serves as a conservative threshold.

Table 3 gives the per-task dominance shares behind the specialisation verdicts. The clearest specialisation is the correlation heatmap dominating 70% of single-frontier correlation tasks. The single-frontier decision task splits the objective panel nearly evenly between RadViz (26%) and the 3D scatter (26%), while the table leads every single-frontier decision-space task. On the comparison tasks, the shares are closer to ties, which is why specialisation there is inconclusive.

Part of the table's pull is exposure: participants who saw the table baseline before the dashboard over-used the table by +0.166 [+0.037, +0.298] of their panel time on a single frontier, with a comparable but inconclusive effect on the comparison (+0.170 [−0.032, +0.367]). The effect is behavioural, as exposure order left usability scores unchanged, with table-first participants at a UMUX of 73.9 against 73.6 for dashboard-first (+0.3 [−7.6, +8.6]). Multi-view use was rare, occurring in only a quarter of objective-panel tasks (Figure 23). Most activity was table-based verification, with the table being the primary second view. Furthermore, combining two charts without a table happened only on the decision tasks.

Table 3: View dominance per task type: for every task and panel, the views ranked by the number of tasks in which they dominated the panel's viewing time, with their share of that task's total.

| RQ | Task | Panel | View | Tasks | Share |
|---|---|---|---|---|---|
| RQ1 single | Correlation | obj | Corr. heatmap | 81 | 64% |
| | | | Table | 24 | 19% |
| | | | 2D scatter | 13 | 10% |
| | | | RadViz | 7 | 6% |
| | | | PCP | 2 | 2% |
| RQ1 single | Decision | obj | RadViz | 39 | 26% |
| | | | 3D scatter | 38 | 26% |
| | | | Table | 21 | 14% |
| | | | PCP | 20 | 13% |
| | | | 2D scatter | 12 | 8% |
| | | | Corr. heatmap | 10 | 7% |
| | | | Radar | 9 | 6% |
| RQ1 single | Value retrieval | obj | Table | 43 | 64% |
| | | | 2D scatter | 11 | 16% |
| | | | PCP | 8 | 12% |
| | | | 3D scatter | 2 | 3% |
| | | | Corr. heatmap | 2 | 3% |
| | | | Radar | 1 | 1% |
| RQ1 single | Correlation | dec | Table | 25 | 69% |
| | | | Stacked bar | 11 | 31% |
| RQ1 single | Decision | dec | Table | 77 | 81% |
| | | | Stacked bar | 18 | 19% |
| RQ1 single | Value retrieval | dec | Table | 47 | 81% |
| | | | Stacked bar | 11 | 19% |
| RQ2 multi | Decide on frontier | obj | Table | 44 | 31% |
| | | | Radar | 32 | 22% |
| | | | 3D scatter | 19 | 13% |
| | | | PCP | 18 | 12% |
| | | | RadViz | 14 | 10% |
| | | | 2D scatter | 12 | 8% |
| | | | Corr. heatmap | 5 | 3% |
| RQ2 multi | Most distinct obj. | obj | Table | 28 | 37% |
| | | | Radar | 12 | 16% |
| | | | PCP | 12 | 16% |
| | | | 2D scatter | 10 | 13% |
| | | | 3D scatter | 6 | 8% |
| | | | RadViz | 5 | 7% |
| | | | Corr. heatmap | 2 | 3% |
| RQ2 multi | Simple comparison | obj | 2D scatter | 21 | 30% |
| | | | Table | 20 | 28% |
| | | | PCP | 11 | 15% |
| | | | 3D scatter | 6 | 8% |
| | | | Radar | 5 | 7% |
| | | | RadViz | 5 | 7% |
| | | | Corr. heatmap | 3 | 4% |
| RQ2 multi | Decide on frontier | dec | Stacked bar | 57 | 52% |
| | | | Table | 53 | 48% |
| RQ2 multi | Most distinct obj. | dec | Stacked bar | 38 | 58% |
| | | | Table | 28 | 42% |
| RQ2 multi | Simple comparison | dec | Stacked bar | 34 | 53% |
| | | | Table | 30 | 47% |

Table 4: Evaluation results: precondition and hypotheses H1–H4 with 95% BCa bootstrap confidence intervals, read as predicted vs observed.

| Hyp | Scope | Metric | Estimate | 95% CI | $n$ | Predicted | Observed |
|---|---|---|---|---|---|---|---|
| H1 | single | UMUX (0-100) | 73.1 | [66.7, 78.3] | 29 | UMUX > 68 | inconclusive |
| H1 | multi | UMUX (0-100) | 74.4 | [66.2, 78.7] | 28 | UMUX > 68 | inconclusive |
| H1 | all | UMUX (0-100) | 73.8 | [69.2, 77.3] | 57 | UMUX > 68 | confirmed |
| H2 | RQ1 single | accuracy delta (views-table) | +0.061 | [+0.015, +0.109] | 29 | views > table | confirmed |
| H2 | RQ1 single | time ratio (views/table) | 0.92x | [0.75, 1.13] | 29 | – (descriptive) | n/a |
| H2 | RQ1 single | dim check: robustness gap | +0.007 | [-0.064, +0.082] | 29 | holds at 6D | inconclusive |
| H2 | RQ1 single | practice check: block gain | -0.012 | [-0.059, +0.045] | 29 | block-to-block gain | inconclusive |
| H2 | RQ1 single | practice check: order-invariance | -0.019 | [-0.114, +0.081] | 29 | gap survives order | confirmed (invariant) |
| H2 | RQ2 multi | accuracy delta (views-table) | -0.023 | [-0.075, +0.026] | 28 | views > table | inconclusive |
| H2 | RQ2 multi | time ratio (views/table) | 1.33x | [1.16, 1.59] | 28 | – (descriptive) | n/a |
| H2 | RQ2 multi | dim check: robustness gap | -0.036 | [-0.138, +0.053] | 28 | holds at 6D | inconclusive |
| H2 | RQ2 multi | practice check: block gain | -0.008 | [-0.058, +0.044] | 28 | block-to-block gain | inconclusive |
| H2 | RQ2 multi | practice check: order-invariance | -0.018 | [-0.119, +0.084] | 28 | gap survives order | confirmed (invariant) |
| H3 | RQ1 single [obj] | distinct dominant views / tasks | 2.0 | n/a | 2 | specialise by task | confirmed |
| H3 | RQ1 single [obj] | most-used view also best? (pool) | n/a | n/a | 343 | used ≠ best | contradicted |
| H3 | RQ1 single [dec] | distinct dominant views / tasks | 1.0 | n/a | 3 | specialise by task | confirmed |
| H3 | RQ1 single [dec] | most-used view also best? (pool) | n/a | n/a | 189 | used ≠ best | confirmed |
| H3 | RQ2 multi [obj] | distinct dominant views / tasks | 1.0 | n/a | 1 | specialise by task | inconclusive |
| H3 | RQ2 multi [obj] | most-used view also best? (pool) | n/a | n/a | 290 | used ≠ best | confirmed |
| H3 | RQ2 multi [dec] | distinct dominant views / tasks | 0.0 | n/a | 0 | specialise by task | inconclusive |
| H3 | RQ2 multi [dec] | most-used view also best? (pool) | n/a | n/a | 240 | used ≠ best | confirmed |
| H3 | RQ1 single table-skew | table-share(baseline-first) - table-share(dashboard-first) | +0.166 | [+0.037, +0.298] | 29 | – (descriptive) | n/a |
| H3 | RQ2 multi table-skew | table-share(baseline-first) - table-share(dashboard-first) | +0.170 | [-0.032, +0.367] | 28 | – (descriptive) | n/a |
| H4 | RQ1 single | actions/trial (dash - table) | -4.0 | n/a | 29 | – (descriptive) | n/a |
| H4 | RQ1 single | acc(high-int) - acc(low-int) | -0.046 | [-0.098, +0.005] | 306 | interactions help | inconclusive |
| H4 | RQ2 multi | actions/trial (dash - table) | 0.7 | n/a | 28 | – (descriptive) | n/a |
| H4 | RQ2 multi | acc(high-int) - acc(low-int) | +0.014 | [-0.049, +0.079] | 253 | interactions help | inconclusive |

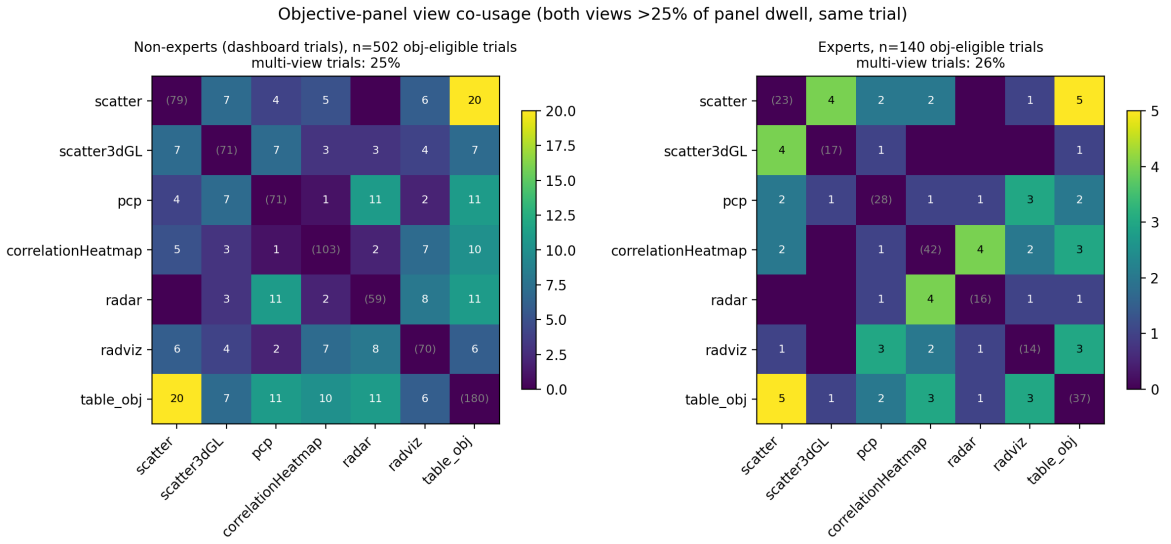



Figure 23: Within-panel view co-usage in the objective panel for non-experts (left) and experts (right).

### 4.1.3 Interactions in Full (H4)

H4 asks whether a view's own interactions pay off. This is answered with a with-without comparison per view: accuracy on the tasks where a view was interacted with against the tasks where it was not. That comparison finds three views where interacting leads to higher accuracy, all of them familiar: the table in the objective panel (+0.130 [+0.026, +0.236]) and the decision panel (+0.184 [+0.093, +0.270]), and the bar chart (+0.199 [+0.088, +0.306]). Every other view is inconclusive, and the heatmap, having no controls of its own, wins its task without any interactions.

We add a further interaction-level breakdown in Figure 24, which applies the same usage-versus-performance analysis as H3 to the individual interaction types. For each interaction and task type, how many tasks used it and how accurate those tasks were. The usage ranking mirrors the per-view result: the most-used interactions are the table's sorting (126 tasks overall), group expansion (83), and histogram filtering (70), together with axis changes (71) and the bar chart's group expansion (59). The specialised views' own manipulations, the RadViz glyph interactions, the PCP axis reordering, and the radar controls, are used far less and are inconclusive. Usage is more present where an interaction's purpose fits the task: sorting and histogram filtering are highest on value retrieval, while the lasso and the RadViz glyph interactions are highest on the decision task. Because participants chose when to interact, these are associations rather than causes.

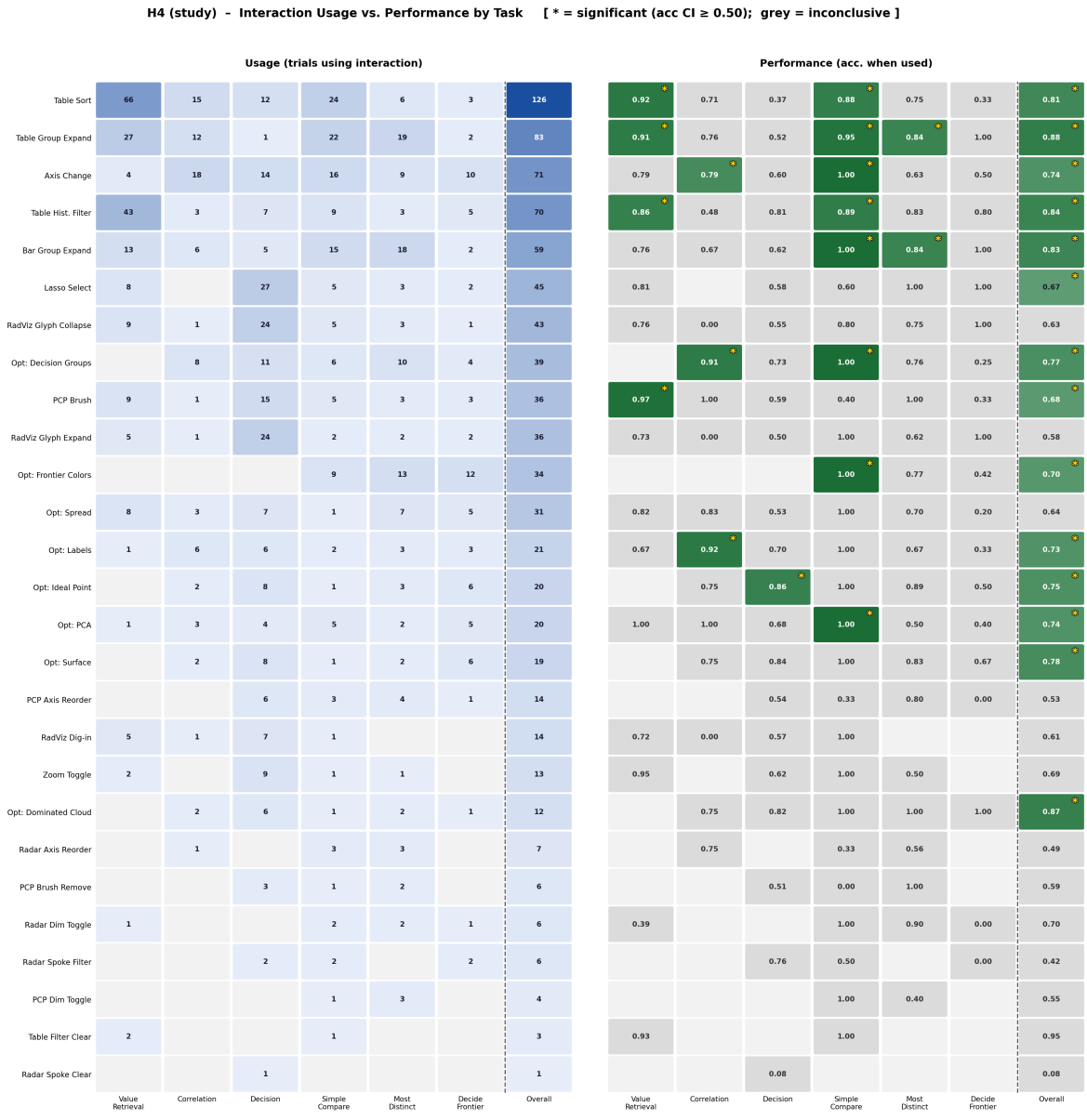



Figure 24: Interaction-level H4 analysis for the non-expert studies, in the same format as the H3 figure: for every interaction type and task, the number of tasks that used the interaction (left) and the accuracy of those tasks (right), with significant cells marked and inconclusive ones greyed.

## 4.2 Full Qualitative Results

Ten experts worked through the compressed survey while thinking aloud, and this section reports their sessions in greater detail, finishing with the open-ended questionnaire responses of both studies. The experts' UMUX mean of 82.5 sits above the non-experts' 73.8, a difference of +8.7 [+0.3, +15.9] whose interval excludes zero, so the experts rated usability reliably higher, with a single low outlier of 54 from the consultant who judged the dashboard an instrument "not every client" could use. Figure 25 places the experts' view usage against their accuracy in the same format as the non-expert figure, and Figure 26 does the same for their interaction types.


Figure 25: Expert usage versus performance by view and task, the expert counterpart of Figure 22.

### 4.2.1 Similarities Between Experts and Non-Experts

Expert telemetry closely tracks the non-expert specialisation. The correlation heatmap dominates the correlation tasks, credited on about 80% of them. In the decision space, experts fall back to the table just as non-experts do, relying on it for about 60–70% of single-frontier decision-panel tasks even where the bar chart is more accurate. The differences between the groups emerged during the harder tasks. When facing multi-frontier objective tasks, non-experts fell back on the table, whereas experts relied on the parallel coordinates plot. This plot actually became the experts' most-used view during the final decision task. Meanwhile, on the single-frontier task, experts reached for the PCP and Radar Chart, but they favoured the 3D scatterplot most of all, despite repeatedly complaining about its depth cues.

The action-level interaction analysis (Figure 26) matches the expert style seen in the recordings. Sorting the table was the experts' most frequent action, being used in 24% of tasks compared to 19% for non-experts. However, its negative point estimate ($-0.128$ [$-0.265$, $+0.004$]) suggests that sorting indicated a difficult task rather than a winning strategy. Furthermore, the two actions that significantly helped non-experts also proved helpful for experts.

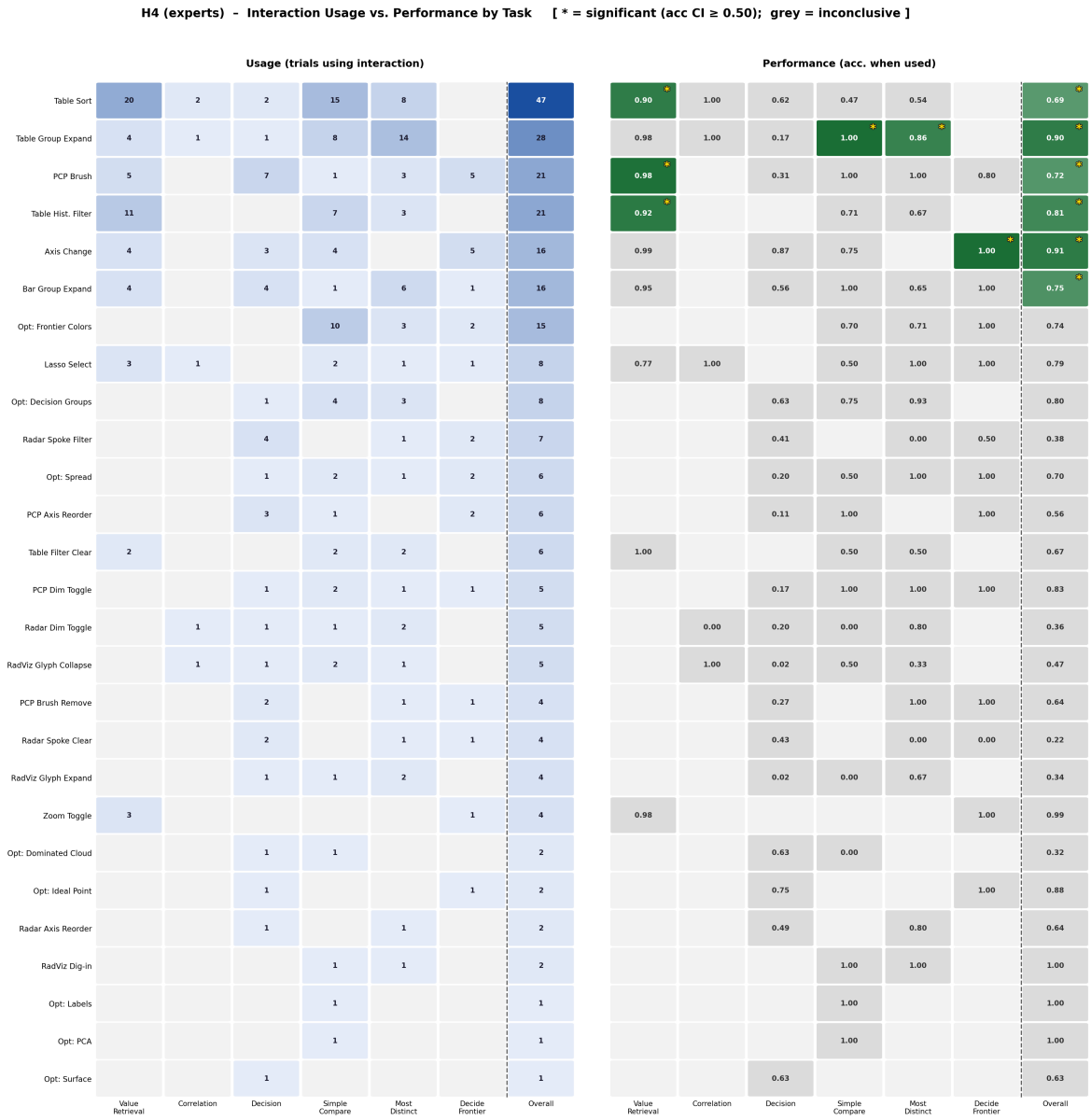



Figure 26: Action-level interaction analysis for the expert sessions.

In other words, the expert data is consistent with the non-expert finding that drill-down and filtering on the familiar table are the interactions that work the best.

A retrospective comparison of the two groups supports this reading. Accuracy shows no reliable difference, either overall ($+0.04\,[-0.03, +0.10]$) or on any individual task. On usage, we see a divergence in 8 cases: experts use the correlation heatmap on every correlation task (100% against 74%), and largely avoid the table on objective-space tasks (0–8% against 18–23%). Therefore, experts differ in how they work, not in how well they perform.

### 4.2.2 Common Patterns in Participant Reasoning

Three patterns recur across the transcripts.

**The averaging gap.** The comparison tasks ask for averages the dashboard never displays, so experts estimated them by eye from the small distribution histograms at the top of the table view, and reported this friction ("I'm not a human calculator"; "there is no average metric, right?"). This shows up throughout the recordings in other ways as well: experts pausing on a comparison task ("which portfolio has a higher average equity allocation? It's a bit difficult"), expressing uncertainty in their interpretations ("so on average, it's a bit difficult to judge"), and explaining errors afterwards by the missing statistics ("because I had to get the average"). Most recurring multi-frontier mistakes group on these comments.

**6D strain.** At six objectives, points shrink past legibility, labels drop out, and the parallel plot suffers from overplotting, the same place the quantitative accuracy degrades. One expert summarised it as "that's where it becomes complicated . . . this can be really hard to visualise," and another wrote after the session that the tool "becomes too abstract and user unfriendly" at six objectives and large score tables, adding that "the distribution charts were helpful in those cases."

**Depth cues.** The 3D scatter drew the most specific criticism even from the experts who used it most ("I don't see really the depth . . . it would be ideal to have two of these [2D scatters]"), and one described their workaround for 3D problems, "I usually put two of these," meaning two linked 2D scatterplots in place of one 3D view, linking back to the side-by-side proposal from the first-round interviews instead of the SPLOM option.

### 4.2.3 Open Exploration

On a single frontier, the reframed decision task turned into curation. Experts went through the frontier point by point to gather a short list for a client, weighing candidates aloud ("good liquidity, higher return, but the trade-off is a bit more risk"; "why would you need this much of cash?"). Participants quickly called out isolated points ("this one seems to be very isolated") and noted "flat frontiers that don't show the shape that well." The specific idioms proved effective: the radar chart was read directly against that benchmark ("the bigger the surface of the triangle, the better"), and the parallel plot demonstrated its value once users learned it ("a point good across all objectives would just be a line at the top").

For the two frontier case when asked which frontier had the higher average, experts read it off the distribution histograms ("this is close together, this is a bit wider out of each other . . . more centred to the right, so probably the return is a bit more"), and judged spread the

same way ("unconstrained is more spread out ... while equity heavy is more crowded in one area"). Filtering became the main comparison move, brushing a region on one frontier to see how many of the other's points were inside it, something that produced vocalised insights ("oh, it works ... this is a very nice one, I didn't realise this before"), and colouring the two frontiers let experts match a portfolio's composition back to its position ("a lot of influence from equity, so that's why it's more risky and more return"). The praise was specific to these moments, the filtered overview and the layered drill-down ("I like this smart way of adding more layers").

### 4.2.4 Open-Ended Questionnaire Responses

Both studies closed with a single open question inviting any feedback not captured by the questionnaire questions. Twenty-one non-experts and four experts answered. Their responses sort into three themes.

**The missing statistics view.** The most voiced request. One non-expert asked directly: "Graphs > table. Why is there no way to actually see the average of the portfolio (on objectives, allocation, etc)?" Another two suggested a potential solution: "it would be nice ... a facility to highlight performance on average,"; "some score metrics that would calculate directly how desirable a diet is would improve the experience for the high dimensional case". This is the averaging gap of the expert sessions, also present in the non-expert case.

**Learning curve and the expert-toolkit judgement.** One expert wrote that "it's a bit of a learning curve ... some charts I found to be a bit less useful in multi-dimension due to the amount of data," adding that "the data sort of fogged my mind at times." Non-experts said the same from below: "the system itself is fun to use and I guess quite useful if you know what you are doing. I didn't really know what I was doing to be honest," and "took me a bit to understand where to find the information and how I should interpret and use it. But I'd assume that if I would know better how to actually make use of all capabilities of the system, it would help me analyse the data." Several non-experts also mention the tutorial rather than the tool: "the tutorial was not really a tutorial, but more of a 'legend' behind all the icons and charts. I'd much rather have a walkthrough related to a simple example or two."

**What worked.** The positive answers are as specific as the complaints. "Very satisfied with how easy it is to see the right choice compared to only looking at tabular data," wrote one non-expert in the single-frontier study, and another said "liked the filtering functionality". One expert provided a summary that perfectly outlines this section: "overall, very neat visualisations that can definitely improve the analysis for a consultant and client alike. Some sections are more useful than others, but it is clear which visualisation is preferred in which situations". This captures the main takeaway of the specialisation result.

When viewed together, the qualitative and quantitative findings line up closely. The usability gap between the two audiences (H1) shows up in the qualitative data through the mention of a learning curve. Similarly, the interview feedback explains why the comparison view lost its advantage (H2). Experts complained about the use of averages, while non-experts requested a dedicated statistics view instead. Furthermore, the specialisation pattern (H3) was recognised by the users, with one expert summarising it in their answer to the open

question. The interaction results (H4) also match our observations in the expert interviews. Participants could achieve their goals when using effective moves, such as table drill-downs and distribution filtering, while actions linked to lower accuracy were usually accompanied by visible struggle from the experts.

# 5

# Reflection

In this chapter, we reflect on this work in its temporal and societal context. We first situate the research within the broader field and where it is heading, then reflect on the implications, limitations, and applicability of the techniques discussed for the stakeholders they affect.

## 5.1 Reflection on Research in the Broader DSAIT Field

This work sits where optimisation research meets its human side, and its findings also offer some insight about the field. Firstly, surveys put interactivity and interpretability at the centre of effective decision support [OZSR*23, HGMR23]. Our study makes explicit what previous work addressed implicitly: that when training time is short, interactivity pays off only on familiar views. Secondly, evaluating the dashboards as a whole [MZC*25, ZWC*18] does not show the impact of each view, and it does not explain which one was really improving the results. This is a common issue in human-centred AI: we know a system works, but we do not know which parts are actually doing the work. While recent advances try to isolate views per task with view recommendation [ZLQW24, WZW*23, ALF*25], they still need exactly the per-view, per-task evidence our work provides. Finally, comparing frontiers exposed a gap in the field: visualisation methods have taxonomies [FT18], but there is no shared task set for comparing solution sets like Dimara et al.[DBD18] built for single-set decisions. Each study must create its own tasks, so results do not accumulate as they do on shared benchmarks, and our mixed RQ2 findings show the cost of this.

## 5.2 Interpretability as Accountability

During the initial interviews, every participant highlighted that interpretability is of utmost importance to the decision-making process. Without it, the reasoning behind advice can not be explained to clients, which weakens accountability. For this reason, dimensionality reduction was rejected because data transformations change the real values of portfolios for a more favourable visual representation, making advice difficult to justify. When stakeholders depend on the literal interpretation of presented metrics, additional abstraction can only reduce explainability further. Interpretability also suffers when data is misread, due to either complex views or views that misrepresent the underlying data. The resulting advice can be improper, with severe consequences for clients downstream. In the context of regulation, pension and insurance asset allocation is a supervised activity across much of the world. In the EU, the law requires that allocation choices be explained to fund participants and that modelling decisions be recorded and their errors accounted for [ior16, sol09]. To a supervisor or auditor, "the optimiser said so" is not an acceptable answer. Furthermore, with the EU AI Act [Eur25] increasing transparency requirements for algorithmic decision support, interpretability is becoming a regulatory expectation rather than professional goodwill.

Ignoring interpretability leads to false confidence, and our study produced a small demonstration of it. When participants lacked a numerical statistic, they relied on visual estimation, which led to wrong task responses. In the real world, the same behaviour leads to issues that can propagate across the whole chain of stakeholders, from the analysts to the pension fund

participants and insurance policyholders. Furthermore, this chain amplifies consequences as they go down it, where a small error in interpreting the frontier can lead to personal damage rather than just a professional error. In our use case, these issues remain monetary, although they can be substantial: In JPMorgan's London Whale trading loss, a value-at-risk model that ran on spreadsheets with errors underestimated risk that was not displayed, leading to losses of over six billion dollars. In other domains, the failures can even cost lives. During the Three Mile Island nuclear accident, the control panel showed the command sent to the relief valve rather than its state, which had the accident misdiagnosed for hours. In the Challenger disaster, data linking temperature to structural integrity did not reach the launch decision in an interpretable format. These were not the direct result of a single bad chart, but show the cost incurred when decisions are made without critical data being properly displayed. Thus, interpretability cannot be overlooked, and this study helps mitigate issues related to explainability by analysing which visualisations offer the best support for high-dimensional Pareto frontiers.

### 5.3 The Limits to Adoption

However, another barrier to real deployment of these techniques is training cost. Participants who are unfamiliar with the domain or the visuals struggle to get used to them. In our study, the table was a familiar baseline to which users fell back to, while other more obscure views such as the PCP and RadViz were used mostly by experts. Therefore, a highly interpretable view that requires hours of practice presents a real trade-off for deployment.

Another limit is inertia. Once a user finds a view that works for their case, they are reluctant to learn another. A pattern made visible when offering the table first, as it led to increased table use. This adds friction to the adoption of any novel technique. It is also why we point, in the limitations and future work, to visualisation recommendation fitted to this setting [ZLQW24, WZW*23, ALF*25]. If the system suggests the right view, the user no longer has to overcome that inertia. Though the recommendations themselves could bias which views users reach for and thus shape the results.

The third limit is accessibility; for pension and insurance dashboards, this is a core fairness requirement. In our framework, interaction techniques reward people familiar with them, and several views can present too much raw data for non-expert stakeholders to navigate. These are design decisions that can favour trained analysts over inexperienced people. Furthermore, users with sensory or motor impairments face different barriers. While the visual design accounted for colour vision deficiency, the interface offers little screen reader support or keyboard navigation. Also, precise pointing and dragging require steady motor control.

Addressing these limitations prevents critical failures, such as unverifiable decisions and inaccessible tools. Our contribution provides insight into how analysts reason over high-dimensional Pareto frontiers. By identifying which visualisations support specific tasks, this study helps future dashboards to be designed around empirical evidence.

# 6

# Conclusion

This thesis explores the way data visualisation can be applied to the analysis of high-dimensional Pareto frontiers through a study that compares visualisation techniques, from a simple tabular view to a dimensionality-reduction-based view such as RadViz. The use case the work is centred around is Investment Portfolio Optimisation. This gives a narrower purpose for our visualisations, instead of trying to present results that are generalisable across all the cases of high-dimensional Pareto frontiers. Therefore, in this work we conduct a study with two audiences that actually use it: 10 domain experts in a semi-structured interview and 57 non-experts in a quantitative study.

Our findings show that the table's status as a reliable baseline is reinforced, in line with prior work [DBD18], while no single view dominates all the others. Each task has a specialised view that outperforms the rest, while interaction only paid off on views familiar to participants. Experts and non-experts had a similar performance and usage behaviour, diverging on the views they reached for when faced with the harder decision tasks.

Our results also reveal gaps in the tool, each pointing to a possible next step. One of the most recurring issues was the lack of an average, which shows the need for a dedicated statistics view. Furthermore, the learning curve and the users' inertia toward familiar views highlight the need for adaptive view recommendation, so the system suggests the right view instead of relying on users to find it.

What we set out from holds: visualisation does make high-dimensional Pareto frontiers easier to understand, but only when the view fits the task. Matching view to task, rather than searching for a general solution, is the principle this thesis puts forward.

# References


[ACKK14] Amershi S., Cakmak M., Knox W. B., Kulesza T.: Power to the people: The role of humans in interactive machine learning. *AI Magazine 35*, 4 (Dec. 2014), 105–120. URL: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/2513, doi:10.1609/aimag.v35i4.2513.

[ALF*25] Aodeng G., Li G., Feng Y., Chen Q., Zhang Y., Liu C. H.: Inreactable: Llm-powered interactive visual data story construction from tabular data. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2025), UIST '25, Association for Computing Machinery. doi:10.1145/3746059.3747719.

[BCS96] Buja A., Cook D., Swayne D. F.: Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics 5*, 1 (1996), 78–99. URL: http://www.jstor.org/stable/1390754.

[BHK11] Brynjolfsson E., Hitt L., Kim H.: Strength in numbers: How does data-driven decisionmaking affect firm performance? *SSRN Electronic Journal* (04 2011). URL: https://ssrn.com/abstract=1819486, doi:10.2139/ssrn.1819486.

[CM84] Cleveland W. S., McGill R.: Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association 79*, 387 (1984), 531–554. URL: http://www.jstor.org/stable/2288400.

[CMS99] Card S., Mackinlay J., Shneiderman B.: *Readings in Information Visualization: Using Vision to Think*. Interactive Technologies. Elsevier Science, 1999. URL: https://books.google.nl/books?id=wdh2gqWfQmgC.

[Das99] Das I.: On characterizing the "knee" of the pareto curve based on normal-boundary intersection. *Structural Optimization 18* (10 1999), 107–115. doi:10.1007/BF01195985.

[DBD18] Dimara E., Bezerianos A., Dragicevic P.: Conceptual and Methodological Issues in Evaluating Multidimensional Visualizations for Decision Support. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (Jan. 2018), 749–759. doi:10.1109/TVCG.2017.2745138.

[DD98] Das I., Dennis J. E.: Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization 8*, 3 (1998), 631–657. doi:10.1137/S1052623496307510.

[DG11] Deb K., Gupta S.: Understanding knee points in bicriteria problems and their implications as preferred solution principles. *Engineering Optimization 43*, 11 (08 2011), 1175–1204. doi:10.1080/0305215X.2010.548863.

[DIPJ22] Dy B., Ibrahim N., Poorthuis A., Joyce S.: Improving Visualization Design for Effective Multi-Objective Decision Making. *IEEE Transactions on Visualization and Computer Graphics 28*, 10 (Oct. 2022), 3405–3416. doi:10.1109/TVCG.2021.3065126.

[DPAM02] Deb K., Pratap A., Agarwal S., Meyarivan T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation 6*, 2 (Apr. 2002), 182–197. doi:10.1109/4235.996017.

[Eur25] European Data Protection Supervisor: *AI Act Regulation (EU) 2024/1689 – Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance).* Publications Office of the European Union, 2025. doi:10.2804/4225375.

[Fin10] Finstad K.: The Usability Metric for User Experience. *Interacting with Computers 22*, 5 (Sept. 2010), 323–327. doi:10.1016/j.intcom.2010.04.004.

[FT18] Filipič B., Tušar T.: A taxonomy of methods for visualizing pareto front approximations. In *Proceedings of the Genetic and Evolutionary Computation Conference* (Kyoto Japan, July 2018), ACM, pp. 649–656. doi:10.1145/3205455.3205607.

[GGZL14] Gomez S. R., Guo H., Ziemkiewicz C., Laidlaw D. H.: An insight- and task-based methodology for evaluating spatiotemporal visual analytics. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Oct. 2014), pp. 63–72. doi:10.1109/VAST.2014.7042482.

[HGM*97] Hoffman P., Grinstein G., Marx K., Grosse I., Stanley E.: DNA visual and analytic data mining. In *Proceedings. Visualization '97 (Cat. No. 97CB36155)* (Phoenix, AZ, USA, 1997), IEEE, pp. 437–441,. doi:10.1109/VISUAL.1997.663916.

[HGMR23] Hakanen J., Gold D., Miettinen K., Reed P. M.: *Visualisation for Decision Support in Many-Objective Optimisation: State-of-the-art, Guidance and Future Directions.* Springer International Publishing, Cham, 2023, pp. 181–212. doi:10.1007/978-3-031-25263-1_7.

[HM25] Hanne T., Moghaddam M. J.: A review of the evolution of multi-objective evolutionary algorithms. *Computers, Materials and Continua 85*, 3 (2025), 4203–4236. URL: https://www.sciencedirect.com/science/article/pii/S1546221825009889, doi:10.32604/cmc.2025.068087.

[HMM21] Hakanen J., Miettinen K., Matković K.: Task-based visual analytics for interactive multiobjective optimization. *Journal of the Operational Research Society 72*, 9 (Sept. 2021), 2073–2090. doi:10.1080/01605682.2020.1768809.

[Hol16] Holzinger A.: Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Informatics 3*, 2 (06 2016), 119–131. doi:10.1007/s40708-016-0042-6.

[HY16] He Z., Yen G. G.: Visualization and performance metric in many-objective optimization. *IEEE Transactions on Evolutionary Computation 20*, 3 (2016), 386–402. doi:10.1109/TEVC.2015.2472283.

[HZJ*24] Huang Y., Zhang Z., Jiao A., Ma Y., Cheng R.: A comparative visual analytics framework for evaluating evolutionary processes in multi-objective optimization. *IEEE Transactions on Visualization and Computer Graphics 30*, 1 (2024), 661–671. doi:10.1109/TVCG.2023.3326921.

[Ins85] Inselberg A.: The plane with parallel coordinates. *Vis. Comput. 1*, 2 (Aug. 1985), 69–91. doi:10.1007/BF01898350.

[ior16] Directive (EU) 2016/2341 of the European Parliament and of the Council of 14 December 2016 on the activities and supervision of institutions for occupational retirement provision (IORPs) (recast) (Text with EEA relevance), Dec. 2016.

[Jol14] Jolliffe I.: *Principal Component Analysis*. John Wiley & Sons, Ltd, 2014. doi:10.1002/9781118445112.stat06472.

[LS18] Lewis J. R., Sauro J.: Item benchmarks for the System Usability Scale. *Journal of Usability Studies 13*, 3 (2018), 158–167. URL: https://dl.acm.org/doi/10.5555/3294033.3294037.

[Mar52] Markowitz H.: Portfolio selection. *The Journal of Finance 7*, 1 (1952), 77–91. URL: http://www.jstor.org/stable/2975974, doi:10.2307/2975974.

[MHSG18] McInnes L., Healy J., Saul N., Grossberger L.: Umap: Uniform manifold approximation and projection. *Journal of Open Source Software 3*, 29 (2018), 861. doi:10.21105/joss.00861.

[Mie98] Miettinen K.: *Nonlinear Multiobjective Optimization*, vol. 12 of *International Series in Operations Research & Management Science*. Springer US, Boston, MA, 1998. doi:10.1007/978-1-4615-5563-6.

[Mil56] Miller G. A.: The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev. 63*, 2 (Mar. 1956), 81–97. doi:10.1037/h0043158.

[Mun14] Munzner T.: *Visualization Analysis and Design*. A K Peters/CRC Press, Dec. 2014.

[MZC*25] Ma Y., Zhang Z., Cheng R., Jin Y., Tan K. C.: Paretolens: A visual analytics framework for exploring solution sets of multi-objective evolutionary algorithms [application notes]. *IEEE Computational Intelligence Magazine 20*, 1 (2025), 78–94. doi:10.1109/MCI.2024.3487134.

[OVLF25] Oliva C., Ventura P. R., Lago-Fernández L. F.: Multi-objective portfolio optimization via gradient descent, 2025. URL: https://arxiv.org/abs/2507.16717, arXiv:2507.16717.

[OZSR*23] Osika Z., Zatarain Salazar J., Roijers D. M., Oliehoek F. A., Murukannaiah P. K.: What Lies beyond the Pareto Front? A Survey on Decision-Support Methods for Multi-Objective Optimization. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence* (Macau, SAR

China, Aug. 2023), International Joint Conferences on Artificial Intelligence Organization, pp. 6741–6749. doi:10.24963/ijcai.2023/755.

[SC99] Simons D. J., Chabris C. F.: Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception 28*, 9 (1999), 1059–1074. doi:10.1068/p281059.

[Shn96] Shneiderman B.: The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages* (1996), pp. 336–343. doi:10.1109/VL.1996.545307.

[SKT23] Schrepp M., Kollmorgen J., Thomaschewski J.: A Comparison of SUS, UMUX-LITE, and UEQ-S. *Journal of User Experience 18*, 2 (2023), 86–104. URL: https://dl.acm.org/doi/abs/10.5555/3604890.3604893.

[sol09] Directive - 2009/138 - EN - solvency ii directive - EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32009L0138, 2009.

[vdMH08] van der Maaten L., Hinton G.: Visualizing data using t-sne. *Journal of Machine Learning Research 9*, 86 (2008), 2579–2605. URL: http://jmlr.org/papers/v9/vandermaaten08a.html.

[VF05] Verleysen M., François D.: The curse of dimensionality in data mining and time series prediction. In *Computational Intelligence and Bioinspired Systems* (06 2005), vol. 3512 of *Lecture Notes in Computer Science*, Springer, pp. 758–770. doi:10.1007/11494669_93.

[VKT*24] Vieth A., Kroes T., Thijssen J., van Lew B., Eggermont J., Basu S., Eisemann E., Vilanova A., Höllt T., Lelieveldt B.: ManiVault: A Flexible and Extensible Visual Analytics Framework for High-Dimensional Data. *IEEE Transactions on Visualization and Computer Graphics 30*, 1 (Jan. 2024), 175–185. doi:10.1109/TVCG.2023.3326582.

[WZW*23] Wang L., Zhang S., Wang Y., Lim E.-P., Wang Y.: LLM4Vis: Explainable visualization recommendation using ChatGPT. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track* (Singapore, Dec. 2023), Wang M., Zitouni I., (Eds.), Association for Computational Linguistics, pp. 675–692. URL: https://aclanthology.org/2023.emnlp-industry.64/, doi:10.18653/v1/2023.emnlp-industry.64.

[ZLQW24] Zhang S., Li H., Qu H., Wang Y.: Adavis: Adaptive and explainable visualization recommendation for tabular data. *IEEE Transactions on Visualization and Computer Graphics 30*, 9 (2024), 5923–5938. doi:10.1109/TVCG.2023.3316469.

[ZWC*18] Zhao X., Wu Y., Cui W., Du X., Chen Y., Wang Y., Lee D. L., Qu H.: SkyLens: Visual Analysis of Skyline on Multi-Dimensional Data. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (Jan. 2018), 246–255. doi:10.1109/TVCG.2017.2744738.

# A

# Survey Setups

This appendix lists every task prompt of the study, exactly as shown to participants, per setup and block. Each participant received one setup; the dashboard-first and table-first variants of a setup contain identical questions and differ only in condition order.

## A.1 Single-Frontier, Portfolio Dataset

**Portfolio 3D – Block 1**

- **Value retrieval** *(equity heavy 3D)*: Select all portfolios where Return is above 0.09.
- **Correlation** *(equity heavy 3D)*: Which decision variable is most strongly correlated with Return (negatively)? Negative: as one variable increases, the other decreases.
- **Decision** *(equity heavy 3D)*: Which portfolio do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Portfolio 3D – Block 2**

- **Value retrieval** *(unconstrained 3D)*: Select all portfolios where Fixed Income allocation is above 25%.
- **Correlation** *(unconstrained 3D)*: Which objective is most strongly correlated with Liquidity (negatively)? Negative: as one variable increases, the other decreases.
- **Decision** *(unconstrained 3D)*: Which portfolio do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Portfolio 3D – Block 3**

- **Value retrieval** *(real assets heavy 3D)*: Select all portfolios where Risk is below 0.010.
- **Correlation** *(real assets heavy 3D)*: Which decision variable is most strongly correlated with Liquidity (positively)? Positive: both variables move in the same direction (both increase or decrease together).
- **Decision** *(real assets heavy 3D)*: Which portfolio do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Portfolio 3D – Block 4**

- **Value retrieval** *(fixed income cash heavy 3D)*: Select all portfolios where Cash allocation is above 25%.
- **Correlation** *(fixed income cash heavy 3D)*: Which objective is most strongly correlated with Liquidity (negatively)? Negative: as one variable increases, the other decreases.
- **Decision** *(fixed income cash heavy 3D)*: Which portfolio do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Portfolio 6D – Block 5**

- **Value retrieval** *(real assets heavy 6D)*: Select all portfolios where ESG score is above 0.60.
- **Correlation** *(real assets heavy 6D)*: Which objective is most strongly correlated with ESG score (negatively)? Negative: as one variable increases, the other decreases.
- **Decision** *(real assets heavy 6D)*: Which portfolio do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Portfolio 6D – Block 6**

- **Value retrieval** *(fixed income cash heavy 6D)*: Select all portfolios where Real Assets allocation is above 10%.
- **Correlation** *(fixed income cash heavy 6D)*: Which objective is most strongly correlated with Diversity (negatively)? Negative: as one variable increases, the other decreases.
- **Decision** *(fixed income cash heavy 6D)*: Which portfolio do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Portfolio 6D – Block 7**

- **Value retrieval** *(equity heavy 6D)*: Select all portfolios where Liquidity is above 0.60.
- **Correlation** *(equity heavy 6D)*: Which decision variable is most strongly correlated with Geopolitical Risk (positively)? Positive: both variables move in the same direction (both increase or decrease together).
- **Decision** *(equity heavy 6D)*: Which portfolio do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Portfolio 6D – Block 8**

- **Value retrieval** *(unconstrained 6D)*: Select all portfolios where EU Real Estate (Real Assets) allocation is above 10%.
- **Correlation** *(unconstrained 6D)*: Which objective is most strongly correlated with Geopolitical Risk (positively)? Positive: both variables move in the same direction (both increase or decrease together).
- **Decision** *(unconstrained 6D)*: Which portfolio do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

## A.2 Single-Frontier, Dietary Dataset

**Diet 3D – Block 1**

- **Value retrieval** *(Budget Diet 3D)*: Select all diets where Nutrient Density is above 0.75.

- **Correlation** *(Budget Diet 3D)*: Which decision variable is most strongly correlated with Nutrient Density (positively)? Positive: both variables move in the same direction (both increase or decrease together).
- **Decision** *(Budget Diet 3D)*: Which diet do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Diet 3D – Block 2**

- **Value retrieval** *(High Protein 3D)*: Select all diets where the total Fats allocation is above 20%.
- **Correlation** *(High Protein 3D)*: Which objective is most strongly correlated with Prep Time (negatively)? Negative: as one variable increases, the other decreases.
- **Decision** *(High Protein 3D)*: Which diet do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Diet 3D – Block 3**

- **Value retrieval** *(Diverse 3D)*: Select all diets where Cost is below 0.30.
- **Correlation** *(Diverse 3D)*: Which food group is most strongly correlated with Fats (negatively)? Negative: as one variable increases, the other decreases.
- **Decision** *(Diverse 3D)*: Which diet do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Diet 3D – Block 4**

- **Value retrieval** *(Vegan 3D)*: Select all diets where Tofu allocation is above 0.7.
- **Correlation** *(Vegan 3D)*: Which objective is most strongly correlated with Nutrient Density (negatively)? Negative: as one variable increases, the other decreases.
- **Decision** *(Vegan 3D)*: Which diet do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Diet 6D – Block 6**

- **Value retrieval** *(Vegan 6D)*: Select all diets where Satiety is above 0.65.
- **Correlation** *(Vegan 6D)*: Which decision variable is most strongly correlated with Satiety (positively)? Positive: both variables move in the same direction (both increase or decrease together).
- **Decision** *(Vegan 6D)*: Which diet do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Diet 6D – Block 6**

- **Value retrieval** *(Low Carb 6D)*: Select all diets where Chicken Breast (Proteins) allocation is above 0.4.
- **Correlation** *(Low Carb 6D)*: Which objective is most strongly correlated with Carbon Footprint (positively)? Positive: both variables move in the same direction (both increase or decrease together).
- **Decision** *(Low Carb 6D)*: Which diet do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Diet 6D – Block 7**

- **Value retrieval** *(High Protein 6D)*: Select all diets where Prep Time is below 10.
- **Correlation** *(High Protein 6D)*: Which decision variable is most strongly correlated with Carbon Footprint (positively)? Positive: both variables move in the same direction (both increase or decrease together).
- **Decision** *(High Protein 6D)*: Which diet do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Diet 6D – Block 8**

- **Value retrieval** *(Budget Diet 6D)*: Select all diets where White Rice (Carbohydrates) allocation is above 10%.
- **Correlation** *(Budget Diet 6D)*: Which objective is most strongly correlated with Diversity (negatively)? Negative: as one variable increases, the other decreases.
- **Decision** *(Budget Diet 6D)*: Which diet do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

### A.3 Multi-Frontier, Portfolio Dataset

**Portfolio 3D – Block 1**

- **Simple comparison** *(equity heavy 3D vs. fixed income cash heavy 3D)*: Which portfolio strategy has a higher average Return?
- **Most distinct objective** *(equity heavy 3D vs. fixed income cash heavy 3D)*: In which investment group do the two portfolio strategies differ the most on average?
- **Decide on frontier** *(equity heavy 3D vs. fixed income cash heavy 3D)*: Which portfolio strategy performs better overall across all objectives?

**Portfolio 3D – Block 2**

- **Simple comparison** *(real assets heavy 3D vs. unconstrained 3D)*: Which portfolio strategy allocates less to Real Assets on average?
- **Most distinct objective** *(real assets heavy 3D vs. unconstrained 3D)*: In which objective do the two portfolio strategies differ the most on average?

- **Decide on frontier** *(real assets heavy 3D vs. unconstrained 3D)*: Which portfolio strategy performs better overall across all objectives?

**Portfolio 3D – Block 3**

- **Simple comparison** *(equity heavy 3D vs. real assets heavy 3D)*: Which portfolio strategy has a lower average Risk?
- **Most distinct objective** *(equity heavy 3D vs. real assets heavy 3D)*: Among fixed income sources, in which do the two strategies differ the most on average?
- **Decide on frontier** *(equity heavy 3D vs. real assets heavy 3D)*: Which portfolio strategy performs better overall across all objectives?

**Portfolio 3D – Block 4**

- **Simple comparison** *(fixed income cash heavy 3D vs. unconstrained 3D)*: Which portfolio strategy allocates more to Private Equity on average?
- **Most distinct objective** *(fixed income cash heavy 3D vs. unconstrained 3D)*: In which objective do the two portfolio strategies differ the most on average?
- **Decide on frontier** *(fixed income cash heavy 3D vs. unconstrained 3D)*: Which portfolio strategy performs better overall across all objectives?

**Portfolio 6D – Block 5**

- **Simple comparison** *(equity heavy 6D vs. fixed income cash heavy 6D)*: Which portfolio strategy has a higher average ESG score?
- **Most distinct objective** *(equity heavy 6D vs. fixed income cash heavy 6D)*: In which investment group do the two portfolio strategies differ the most on average?
- **Decide on frontier** *(equity heavy 6D vs. fixed income cash heavy 6D)*: Which portfolio strategy performs better overall across all objectives?

**Portfolio 6D – Block 6**

- **Simple comparison** *(real assets heavy 6D vs. unconstrained 6D)*: Which portfolio strategy has a higher average Equity allocation?
- **Most distinct objective** *(real assets heavy 6D vs. unconstrained 6D)*: In which objective do the two portfolio strategies differ the most on average?
- **Decide on frontier** *(real assets heavy 6D vs. unconstrained 6D)*: Which portfolio strategy performs better overall across all objectives?

**Portfolio 6D – Block 7**

- **Simple comparison** *(equity heavy 6D vs. real assets heavy 6D)*: Which portfolio strategy has a higher average Return?
- **Most distinct objective** *(equity heavy 6D vs. real assets heavy 6D)*: In which type of equity do the two strategies differ the most on average?

- **Decide on frontier** *(equity heavy 6D vs. real assets heavy 6D)*: Which portfolio strategy performs better overall across all objectives?

**Portfolio 6D – Block 8**

- **Simple comparison** *(fixed income cash heavy 6D vs. unconstrained 6D)*: Which portfolio strategy has a higher average US Bond (Fixed Income) allocation?
- **Most distinct objective** *(fixed income cash heavy 6D vs. unconstrained 6D)*: In which objective do the two portfolio strategies differ the most on average?
- **Decide on frontier** *(fixed income cash heavy 6D vs. unconstrained 6D)*: Which portfolio strategy performs better overall across all objectives?

## A.4 Multi-Frontier, Dietary Dataset

**Diet 3D – Block 1**

- **Simple comparison** *(Budget Diet 3D vs. High Protein 3D)*: Which diet plan has a higher average allocation to Vegetables and Fruits?
- **Most distinct objective** *(Budget Diet 3D vs. High Protein 3D)*: In which objective do the two diet plans differ the most on average?
- **Decide on frontier** *(Budget Diet 3D vs. High Protein 3D)*: Which diet plan performs better overall across all objectives?

**Diet 3D – Block 2**

- **Simple comparison** *(Diverse 3D vs. Vegan 3D)*: Which diet plan has a lower average Cost?
- **Most distinct objective** *(Diverse 3D vs. Vegan 3D)*: In which food group do the two diet plans differ the most on average?
- **Decide on frontier** *(Diverse 3D vs. Vegan 3D)*: Which diet plan performs better overall across all objectives?

**Diet 3D – Block 3**

- **Simple comparison** *(Budget Diet 3D vs. Diverse 3D)*: Which diet plan has a lower average Oats (Carbohydrates) allocation?
- **Most distinct objective** *(Budget Diet 3D vs. Diverse 3D)*: In which objective do the two diet plans differ the most on average?
- **Decide on frontier** *(Budget Diet 3D vs. Diverse 3D)*: Which diet plan performs better overall across all objectives?

**Diet 3D – Block 4**

- **Simple comparison** *(High Protein 3D vs. Vegan 3D)*: Which diet plan has a higher average Nutrient Density?
- **Most distinct objective** *(High Protein 3D vs. Vegan 3D)*: Which carbohydrate source differs the most between the two diet plans?
- **Decide on frontier** *(High Protein 3D vs. Vegan 3D)*: Which diet plan performs better overall across all objectives?

**Diet 6D – Block 5**

- **Simple comparison** *(Low Carb 6D vs. Budget Diet 6D)*: Which diet plan has a lower average Prep Time?
- **Most distinct objective** *(Low Carb 6D vs. Budget Diet 6D)*: In which food group do the two diet plans differ the most on average?
- **Decide on frontier** *(Low Carb 6D vs. Budget Diet 6D)*: Which diet plan performs better overall across all objectives?

**Diet 6D – Block 6**

- **Simple comparison** *(High Protein 6D vs. Diverse Diet 6D)*: Which diet plan has a lower average allocation to Fats?
- **Most distinct objective** *(High Protein 6D vs. Diverse Diet 6D)*: In which objective do the two diet plans differ the most on average?
- **Decide on frontier** *(High Protein 6D vs. Diverse Diet 6D)*: Which diet plan performs better overall across all objectives?

**Diet 6D – Block 7**

- **Simple comparison** *(Low Carb 6D vs. High Protein 6D)*: Which diet plan has a lower average Carbon Footprint?
- **Most distinct objective** *(Low Carb 6D vs. High Protein 6D)*: In which protein source do the two diet plans differ the most on average?
- **Decide on frontier** *(Low Carb 6D vs. High Protein 6D)*: Which diet plan performs better overall across all objectives?

**Diet 6D – Block 8**

- **Simple comparison** *(Budget Diet 6D vs. Diverse Diet 6D)*: Which diet plan has a lower average Sweet Potato (Carbohydrates) allocation?
- **Most distinct objective** *(Budget Diet 6D vs. Diverse Diet 6D)*: In which objective do the two diet plans differ the most on average?
- **Decide on frontier** *(Budget Diet 6D vs. Diverse Diet 6D)*: Which diet plan performs better overall across all objectives?

## A.5 Expert Sessions

**Portfolio 3D – Block 1**

- **Value retrieval** *(unconstrained 3D)*: Select all portfolios where Return is above 0.09.
- **Correlation** *(unconstrained 3D)*: Which decision variable is most strongly correlated with Return (negatively)? Negative: as one variable increases, the other decreases.
- **Value retrieval** *(unconstrained 3D)*: Select all portfolios where Fixed Income allocation is above 25%.
- **Correlation** *(unconstrained 3D)*: Which objective is most strongly correlated with Liquidity (negatively)? Negative: as one variable increases, the other decreases.
- **Decision** *(unconstrained 3D)*: Which portfolio do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Portfolio 6D – Block 2**

- **Value retrieval** *(fixed income cash heavy 6D)*: Select all portfolios where ESG score is above 0.70.
- **Correlation** *(fixed income cash heavy 6D)*: Which objective is most strongly correlated with ESG score (negatively)? Negative: as one variable increases, the other decreases.
- **Value retrieval** *(fixed income cash heavy 6D)*: Select all portfolios where Cash allocation is above 10%.
- **Correlation** *(fixed income cash heavy 6D)*: Which decision variable is most strongly correlated with Diversity (negatively)? Negative: as one variable increases, the other decreases.
- **Decision** *(fixed income cash heavy 6D)*: Which portfolio do you think offers the best overall trade-off across all objectives? If unsure, multiple selections are also valid.

**Portfolio 3D – Block 3**

- **Simple comparison** *(equity heavy 3D vs. unconstrained 3D)*: Which portfolio strategy has a higher average Return?
- **Most distinct objective** *(equity heavy 3D vs. unconstrained 3D)*: In which decision variable do the two portfolio strategies differ the most on average?
- **Simple comparison** *(equity heavy 3D vs. unconstrained 3D)*: Which portfolio strategy allocates more to Private Equity on average?
- **Most distinct objective** *(equity heavy 3D vs. unconstrained 3D)*: In which objective do the two portfolio strategies differ the most on average?
- **Decide on frontier** *(equity heavy 3D vs. unconstrained 3D)*: Which portfolio strategy performs better overall across all objectives?

**Portfolio 6D – Block 4**

- **Simple comparison** *(real assets heavy 6D vs. unconstrained 6D)*: Which portfolio strategy has a higher average Equity allocation?
- **Most distinct objective** *(real assets heavy 6D vs. unconstrained 6D)*: In which objective do the two portfolio strategies differ the most on average?
- **Simple comparison** *(real assets heavy 6D vs. unconstrained 6D)*: Which portfolio strategy has a higher average Return?
- **Most distinct objective** *(real assets heavy 6D vs. unconstrained 6D)*: In which type of equity do the two strategies differ the most on average?
- **Decide on frontier** *(real assets heavy 6D vs. unconstrained 6D)*: Which portfolio strategy performs better overall across all objectives?

# B

# AI Tool Usage and Acknowledgments

Throughout this project, AI systems, such as **Gemini** and **Claude**, were used for the development workflow to handle routine tasks and allow me to focus on higher-level tasks. I generally tried to supply detailed context, iterate through prompts, and merge generated outputs. Applications within the research process included:

- **Understanding Math & Literature Review:** I used AI to translate complex math and jargon into simple terms, and to quickly summarise papers before diving into full reads. The potential issues I faced with this were due to potential hallucinations of complex terms and paper contents. To mitigate such issues, I did redundancy checks on my own for critical parts of the thesis, took time to understand critical concepts fully, and searched for alternative explanations of these concepts in related work papers.
- **Brainstorming & Ideation:** I bounced ideas back and forth with AI to explore research directions and frame questions. The main issue with this was that I could go on with a wrong idea reinforced by my conversation with AI, so to mitigate this I always discussed these ideas with my supervisors.
- **Targeted Search:** I built specific prompts to quickly find relevant papers and background material. The issue here is that references can be fabricated, so I retrieved and read every cited paper from its source and used this alongside conventional searches.
- **Code Generation:** I designed the system architecture manually, then used AI to write small, specific functions. The issue with generated code is that it can look correct while being wrong in the overall system, so I reviewed and debugged every line by hand.
- **Editing & Polishing:** I used AI to tighten up wordy explanations, fix sentence structure, and clean up rough drafts. The issue here is that rephrasing can shift or overstate a claim, so I checked every edit against what I originally meant.