Evolutionary Design of a Visual Analytics Interface to Study Predictive Patterns in High Dimensional Data[★]

Darius Coelho[a,b,*], Eric Papenhausen[a,b] and Klaus Mueller[a,b]

[a]Stony Brook University, Stony Brook, New York, United States
[b]Akai Kaeru, Stony Brook, New York, United States

ARTICLE INFO

Keywords:
High-Dimensional Data
Multivariate Data
Pattern Analysis
Data Mining
Visual Analytics
Predictive Analytics

ABSTRACT

Organizations increasingly rely on data mining to discover predictive patterns that inform decision-making. However, the vast number of patterns produced by mining algorithms, especially in high-dimensional datasets, can overwhelm analysts. To address this challenge, we present the evolutionary design of a visual analytics interface that enables users to explore and interpret mined predictive patterns effectively. Starting from a projection-based prototype rooted in research tools, we iteratively refined the interface through feedback from data analysts and real-world usage. Each design iteration – projection maps, bubble plots, and finally a card-based layout – was shaped by insights into user comprehension, trust, and pattern comparison needs. The final interface supports overview-to-detail exploration, pattern organization, and contextual analysis, as demonstrated in a case study on student attrition in Computer Science. We reflect on the lessons learned across iterations and offer guidance for designing interpretable pattern exploration tools for high-dimensional data.

1. Introduction

Many industries increasingly rely on the ability to extract and act on descriptive or predictive patterns in complex datasets. A *pattern* refers to a subgroup of data points that share similar characteristics and, on average, exhibit similar values for a predictive outcome variable – such as elevated customer churn, at-risk students, financial loss, or increased revenue. While algorithmic advances have automated pattern discovery, these methods often generate an overwhelming number of patterns, making it difficult for analysts to discern the most meaningful or actionable ones. Moreover, high-dimensional datasets introduce additional interpretability challenges, particularly when a pattern's relevance depends on subtle feature combinations.

To address these challenges, researchers have developed visual analytics systems that integrate computational analysis with interactive visualization to facilitate multidimensional data exploration Huber (1985); Stahnke, Dörk, Müller, and Thom (2016); Sivaraman, Li, and Perer (2025). Over time, these systems have become increasingly sophisticated, incorporating advanced dimensionality reduction (DR) techniques and complex visual encodings. However, such sophistication often comes at the cost of interpretability and accessibility – particularly in DR-based interfaces, where users must understand which projection methods to apply based on data characteristics Espadoto, Martins, Kerren, Hirata, and Telea (2021); Cashman, Keller, Jeon, Kwon, and Wang (2025), and where inherent distortions may obscure the original data relationships Nonato and Aupetit (2019); Jeon, Ko, Jo, Kim, and Seo (2022). As a result, these systems can be difficult for broader audiences without specialized visualization expertise. This paper describes how our design process evolved from a complex, projection-based interface toward a simpler, more transparent, and interpretable system.

Our motivation stems from the development of a pattern mining engine that automatically identifies statistically significant and interpretable subspaces in multivariate data. While algorithmically robust, this approach tends to produce dozens to hundreds of candidate patterns per dataset, leaving our collaborators – including data scientists and analysts – struggling to interpret the results. To address this challenge, we designed a visual analytics interface for pattern exploration, emphasizing clarity, comparability, and interpretability.

This paper describes the evolutionary design of our interface through three major iterations. Each stage – beginning with a projection-based interface, advancing to an axis-aligned bubble plot view, and culminating in a card-based layout – was shaped by observations of user behavior in real-world use cases. We found that specialized visualization methods, such as high-dimensional projections, suffer from interpretablity issues Seo and Shneiderman (2004); Faust, Glickenstein, and Scheidegger (2019); Cheng and Mueller (2016), that confused users, whereas simpler axis-aligned and list-based views improved comprehension and facilitated analysis.

Our final card-based design enables users to sort, filter, and organize patterns, examine detailed subpopulations, and explore relationships across patterns. We demonstrate its utility in a case study at an educational institution investigating student attrition from the Computer Science major, where analysts applied the tool to uncover and interpret meaningful predictive patterns. Finally, we reflect on design lessons learned from this iterative process and propose

[★]This document is the results of the research project funded by the National Science Foundation.

[*]Corresponding author
dcoelho@cs.stonybrook.edu (D. Coelho); epapenhausen@cs.stonybrook.edu (E. Papenhausen); mueller@cs.stonybrook.edu (K. Mueller)
ORCID(s):

generalizable recommendations for developing interpretable visual interfaces for data mining outputs.

Our contributions include:

- A comparative analysis of three visual interface designs for pattern exploration in high-dimensional data.
- Design insights on how visualization techniques influence pattern interpretability and user trust.
- A production-ready interface demonstrated through a real-world case study.
- Design guidance for building effective pattern exploration tools that bridge machine learning and human sensemaking.

By documenting the iterative process that led to the final design and implementation of our tool, we aim to provide generalizable recommendations that can guide developers in building exploration tools for high-dimensional data.

2. Related Work

Visualizing and interpreting patterns in high-dimensional data has long been a core challenge in visual analytics. Liu et al. Liu, Maljovec, Wang, Bremer, and Pascucci (2017) provide a comprehensive review of techniques in this space. Traditional approaches such as scatterplot matrices (SPLOMs) Becker, Cleveland, and Wilks (1987) and parallel coordinates Inselberg (1985) are intuitive but scale poorly with increasing dimensionality. Wilkinson et al. Wilkinson, Anand, and Grossman (2005) addressed these scalability challenges by introducing scagnostics, a set of measures designed to help users identify interesting relationships within large collections of scatterplots. Radial layouts such as Star Coordinates Kandogan (2000) and Radviz Hoffman, Grinstein, Marx, Grosse, and Stanley (1997) aimed to provide global overviews of all attributes but often proved difficult to interpret accurately.

To address scalability more directly, modern techniques employ dimensionality reduction (DR) and projection methods such as PCA Jolliffe (2005), MDS Torgerson (1952), t-SNE Maaten and Hinton (2008), UMAP McInnes, Healy, Saul, and Großberger (2018), and ULCA Fujiwara, Wei, Zhao, and Ma (2022) to generate two-dimensional representations of high-dimensional data. While these methods effectively reveal clusters and structure, they introduce information loss and are often ill-suited for conveying attribute-level explanations. To improve interpretability, Cheng et al. Cheng and Mueller (2016) proposed the Data Context Map (DCM), which jointly embeds data points and features within a single projection space. Similarly, Faust et al. Faust et al. (2019) introduced DimReader, which improves the interpretability of nonlinear DR visualizations by perturbing input features to model scalar fields expressed as isocontours that reveal how individual dimensions influence the projection. Glyph-based methods Ward (2008) offer an alternative by encoding multiple dimensions within compact symbols, but they struggle with scalability and can be cognitively demanding to interpret.

To enable more comprehensive exploration, many of these techniques have been embedded in visual analytics systems that help users identify and analyze subspaces of interest in high-dimensional data. Seo and Shneiderman Seo and Shneiderman (2004) proposed a rank-by-feature framework that allows users to rank one- and two-dimensional axis-parallel projections of multidimensional data, making it easier to identify interesting feature relationships and behaviors. Guo et al. Guo (2003) combined parallel coordinates with linked views to support subspace exploration and cluster detection. They noted, however, that while MDS layouts helped users observe structural patterns, they lacked clarity around the underlying attribute relationships.

Several efforts have sought to improve the interpretability of projections. Early work by Huber Huber (1985) introduced *projection pursuit* techniques to optimize projection views, increasing the likelihood of revealing interesting structures in high-dimensional data. Stahnke et al. Stahnke et al. (2016) introduced a suite of interactive and visual methods for interpreting dimensionality-reduced layouts: users can probe projection errors, question point placements, and examine how high-dimensional attributes influence the 2D embedding. Li et al. Li and Zhou (2023) proposed incorporating human-provided class labels into the embedding loss function, enabling users to iteratively refine projections for semantic clarity. Zhang et al. Zhang, Li, and Xu (2024) applied rule-based embeddings (REs) to facilitate exploration of multi-dimensional data in a projection-based interface. Other systems, such as those by Yang et al. Yang, Patro, Huang, Mehta, Ward, and Rundensteiner (2004) and Kammer et al. Kammer, Keck, Gründer, Maasch, Thom, Kleinsteuber, and Groh (2020), replaced points in MDS layouts with feature-encoded glyphs to convey additional dimensional context. More recently, Sivaraman et al. Sivaraman et al. (2025) introduced *Divisi*, a notebook-embedded visual analytics tool that integrates fast approximate subgroup discovery with dynamic filtering, ranking, and a novel "Subgroup Map" visualization to help users explore and evaluate meaningful subpopulations.

These systems often rely on users to visually detect patterns or interactively filter the data to uncover them. In contrast, our approach starts with a pattern mining engine that automatically identifies statistically meaningful subgroups. We then apply a sequence of visual design iterations, many inspired by prior work, to build interpretable interfaces that support pattern exploration and refinement. A similar direction was explored by Knittel et al. Knittel, Lalama, Koch, and Ertl (2021), who used neural networks to extract rule-based patterns, and by Neto et al. Neto and Paulovich (2022), who proposed a visual interface to explore jumping emerging patterns. However, these systems either focus on expert users or require prior modeling steps. In contrast, our work is designed for broader accessibility, enabling domain experts and decision-makers to visually explore and understand data patterns without requiring deep machine learning expertise.

Table 1
Comparison of Interface Design Iterations.

Iteration	Visualization Method	Key Strengths	Key Limitations	Design Lessons Learned
Projection-Based	Data Context Map (DCM)	Rich contextual layout of patterns & features	Difficult to interpret; low trust; unfamiliar projection concepts	Avoid complex, niche visualizations for wide audiences
Bubble Plot	Axis-aligned scatterplot	Familiar axes; clearer target effect encoding	Limited context; attribute switching burdens memory	Users require multi-attribute overviews and summaries with minimal interaction and view changes
Card-Based	Pattern cards in lists	High readability; supports sorting, grouping	Space constraints; overview requires scrolling	Overview-to-detail, familiar visualizations, and organization tools improve data comprehension and trust

3. Background: Pattern Mining

Our approach is grounded in pattern analysis, a well-established area in data science and AI Kriegel, Kröger, and Zimek (2009). A pattern is typically defined as a subgroup of data points that share similar characteristics or features Atzmueller (2015) and that, on average, exhibit higher or lower values for a predictive outcome variable compared to the overall dataset. Such patterns are often referred to as *interesting* groups of data points, but since real-world datasets are often high-dimensional and sparse, manual pattern discovery can become impractical.

Visualization tools such as the Data Context Map (DCM) Cheng and Mueller (2016) enable users to explore high-dimensional data on a 2D map, aiding them in identifying patterns. However, scanning for patterns manually in such tools is often time-consuming and cognitively demanding, and assessing their statistical significance through visual inspection alone is unreliable. In contrast, the visual components of these tools are well suited for presenting and analyzing patterns that have already been identified.

To bridge this gap, we developed a pattern mining engine that automatically discovers patterns ready to be visually explored. Our mining algorithm extends well-established techniques to identify subgroups in high-dimensional data that show statistically significant differences in a target variable. These patterns are expressed as combinations of feature constraints and resemble subspace clusters or slices.

We extend the FP-Growth algorithm Han, Pei, and Yin (2000) to efficiently mine these patterns, integrating statistical tests such as the Mann-Whitney U test for continuous targets and the χ^2 test for binary targets. To quantify impact, we incorporate measures like effect size McGraw and Wong (1992) and odds ratios. The resulting patterns are interpretable, defined over constrained ranges of a few features strongly associated with the target outcome.

For example, in a student dataset with dropout rate as the target outcome, the algorithm might identify a subgroup with significantly higher dropout rates among students who live off campus and have lower SAT scores. Conversely, it might also reveal low-risk subgroups, such as students with high SAT scores and strong GPAs. This technique has been applied in diverse domains, including the identification of socioeconomic risk factors for COVID-19 spread Coelho, Gupta, Papenhausen, and Mueller (2022) and the detection of promising donors in advancement and fundraising Mueller and Papenhausen (2022).

Our work aligns with growing interest in data slicing and pattern-based model validation Chung, Kraska, Polyzotis, Tae, and Whang (2019); Knittel et al. (2021); Zhang, Ono, Song, Gou, Ma, and Ren (2023), which aim to surface coherent subgroups where models behave differently. While those systems often focus on debugging and typically target expert users, our interface is designed for broader accessibility. It enables analysts and decision-makers to interactively explore and refine mined patterns through a visual interface, without requiring machine learning expertise.

4. Evolutionary Visual Interface Design

Designing a visual interface for exploring mined patterns in high-dimensional data poses unique challenges. While the underlying patterns are statistically robust and defined by compact feature constraints, their sheer volume and complexity can overwhelm users. Furthermore, users vary widely in their visualization literacy, ranging from experienced data scientists to domain experts with limited exposure to advanced visual encodings.

To address these issues, we developed our visual pattern exploration interface through three major design iterations, each informed by user feedback and real-world usage. While all versions used the same underlying pattern mining engine, they differed in how patterns were visually represented, interacted with, and compared. This section presents the three major design iterations - projection-based, bubble-plot-based, and card-based - and discusses their strengths and weaknesses based on user feedback. A summary of this section is provided in table 1.

4.1. Design Goals

Our interface design was guided by four key goals based on discussions with users:

DG1: Efficient Pattern Identification. Large multidimensional datasets often generate many patterns, even after applying statistical filters such as thresholds on *p*-values or effect sizes. While these filters remove weak or insignificant results, users are still left with a sizable set to interpret. Our first design goal is to provide an interactive technique that helps users efficiently identify patterns of interest from this reduced but still complex set of patterns.

DG2: Interpreting Patterns. Mined patterns are defined by combinations of feature constraints such as ranges for numeric attributes or specific values for categorical ones. However, users often struggle to understand how these constraints work together to define a subpopulation. Our second goal is to present each pattern's definition in a clear, compact format that makes its meaning immediately interpretable without requiring users to inspect raw data tables.

DG3: Pattern Comparison and Organization. Analysts often want to compare patterns to detect redundancies, relationships, or conflicts-such as nested patterns, overlapping features, or opposing effects. They may also wish to group patterns based on domain relevance (e.g., "transfer students" or "female stroke victims"). Our third goal is to support flexible comparison and organization of patterns.

DG4: Expand or Customize Patterns Domain experts often want to refine patterns based on their knowledge or intuition by adding, removing, or adjusting constraints. Our fourth goal is to support interactive "what-if" exploration, enabling users to customize pattern definitions and immediately visualize how such changes affect the target variable (e.g., dropout rate or risk of stroke). This bridges automated discovery with expert-driven hypothesis generation.

4.2. Iteration 1: Projection-Based Interface

The first version of the interface aimed to support pattern exploration through a multidimensional scaling (MDS)-based projection that jointly embedded patterns, features, and data points into a shared 2D layout we call the *Pattern Map*. As shown in Figure 1(a), each pattern was visualized as a bubble, positioned in high-dimensional space based on the average values of its defining attributes and projected on the 2D layout. Bubble size reflected the percentage of the data it encompassed, and color indicated the pattern's effect on the target variable (green for positive, red for negative), offering a dense yet compact encoding of statistical properties. Features were represented as labeled triangles, with their proximity to pattern bubbles indicating contextual relevance. This view provided a high-level overview that allowed users to scan for interesting patterns (**DG1**) and make coarse comparisons (**DG3**) based on size, target impact, and feature proximity.

Selecting a pattern opened a detail panel displaying its feature constraints, effect size, p-value, and confidence level (**DG2**). To support further inspection, a secondary view - the Pattern Inspector - used the Data Context Map (DCM) Cheng and Mueller (2016) to project the data points associated with the selected pattern, as shown in Figure1(c). In this view, the selected pattern was visualized as a contour whose perimeter was defined by the conjunction of its attribute constraints, effectively outlining the subset of data points within the pattern. Additionally, the dataset attributes were visualized as a list of brushable distribution plots to the left of the DCM. The brush ranges were set according to the ranges of the attributes that define the selected pattern. Users could adjust the brushes and observe how changes impacted the visual boundary i.e., the patterns and associated metrics in real time (**DG4**).

User Feedback and Limitations: Despite its expressive and context-rich design, the projection-based interface was ultimately inaccessible to most users:

- *Projection Comprehension:* Users unfamiliar with multidimensional scaling (MDS) found the layout difficult to interpret. The absence of labeled axes led to confusion, with several asking, "What do the axes mean?" – revealing a gap between the abstract projection and users' visualization knowledge.
- *Unintuitive spatial encoding:* Users often assumed nearby elements shared similar attribute values, unaware that MDS could distort such relationships. This mismatch led to misinterpretation and reduced trust in the visualization.
- *Over-encoded visuals:* The projection embedded patterns, data points, and attributes in a single view, using multiple encodings (position, color, shape). While rich, the dense layout caused visual overload, making it hard for users to focus or understand individual elements.
- *Unexplained feature relevance:* Feature labels were shown, but users had no way to assess which attributes were most important or how they contributed to patterns. This lack of clarity hindered meaningful interpretation.

Takeaway: While projection maps can provide rich spatial context, they are poorly suited for general-purpose pattern exploration by non-experts. Users strongly preferred interfaces with explicit axes, interpretable summaries, and concrete visual encodings. These limitations drove a pivot toward a broadly readable, structured, and task-aligned design in later iterations.

4.3. Iteration 2: Bubble Plot Interface

The second design iteration replaced the abstract projection view with a more familiar and axis-aligned bubble plot, aimed at improving interpretability and visual clarity. In this version, the earlier *Pattern Map* was replaced by a bubble plot at the center, as shown in Figure 2. Similar to the previous design, each pattern was represented as a bubble, where size indicated the number of data points encompassed by the pattern, and color reflected its effect on the target

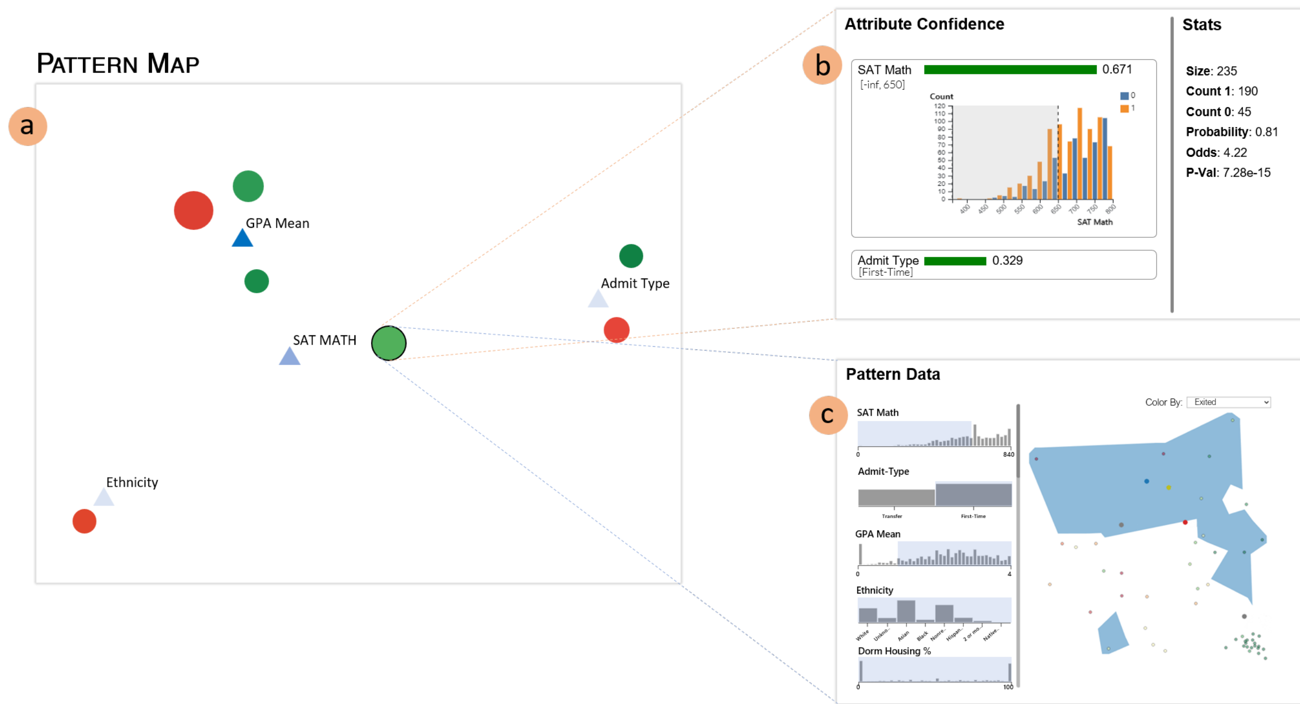


Figure 1: The projection-based interface for exploring patterns. In this case, patterns of student dropout. (a) A multidimensional projection maps patterns (as bubbles) alongside their defining attributes (as labeled triangles). The selected pattern is positioned closest to *SAT Math*, indicating this attribute is a strong constraint. Its green color signifies a higher-than-average dropout rate. (b) The pattern detail view summarizes its defining features, showing that *SAT Math* contributes 67% to the pattern, while *Admit Type* contributes 33%. The pattern includes 170 students who exited the major and 45 who did not. (c) The Data Context Map (DCM) view projects the individual data points within the selected pattern. The contour outlines the pattern boundary, while brushable attribute distributions on the left allow users to adjust feature ranges and explore their impact.

variable (green for high exit risk, red for low). However, unlike the projection, bubbles were now plotted along two explicit axes: the y-axis represented the target variable (e.g., CS exit probability), and the x-axis showed a user-selected feature (e.g., SAT score). Bubbles were positioned based on the mean value of the selected feature within the pattern's data points. This layout provided a more grounded and intuitive spatial representation, enabling users to quickly identify patterns based on their effect and coverage (DG1) and to compare them along specific attributes (DG3).

To the left of the bubble chart, a vertical panel presents both a global overview and pattern-specific information. At the top, a *feature importance bar chart* displays aggregated Shapley scores for each attribute across all patterns (gray bars), addressing users' need to understand global feature relevance. When a pattern is selected (clicking on a bubble), its individual feature contributions are shown as blue bars (**DG2**). Clicking on any feature updates the x-axis of the bubble plot accordingly, enabling focused comparisons.

Below this, a *target variable distribution chart* compares the full dataset (gray) with the selected pattern (blue), allowing users to see how the pattern shifts the outcome distribution. At the bottom, a *summary statistics panel* displays key metrics for the entire dataset as well as those for the selected pattern, including coverage, class counts (e.g., enrolled vs. exited), exit probability, odds ratio, and common statistics (min, max, mean, median, and standard deviation) of the target variable (**DG2**).

In addition to the overview panel, the interface includes two components for detailed pattern inspection positioned below the bubble plot (**DG2**). Building on the detail panel from the projection interface (Figure 1(b)), these components present patterns in a clearer, more structured layout. The first is a list of pattern-defining attributes with their associated value ranges (shown in bracketed text) and corresponding Shapley values, displayed as green horizontal bars. Shapley values were chosen over confidence scores in this version because they are more familiar and interpretable for many users. To the right, a histogram allows inspection of any selected attribute (e.g., Math SAT score), showing how many data points in each value bin fulfill the target condition (exiting the major, orange bars) versus those that do not (staying in the major, blue bars). The gray box highlights the value range of the selected attribute that defines the pattern. Users can interactively choose other attributes to see how constraints affect pattern membership, supporting exploratory "what-if" analysis (**DG4**). To further aid comprehension, the interface also generates a natural language description of each pattern, summarizing its constraints and

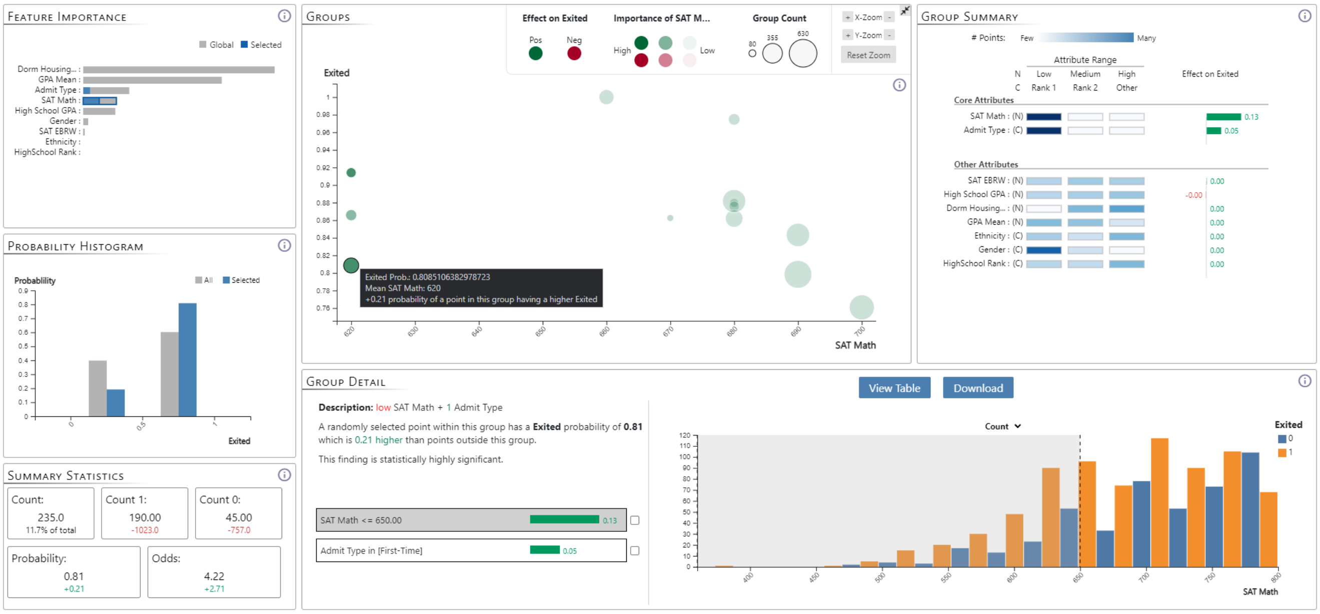


Figure 2: The second iteration of our interface adopts a bubble-based layout to support more interpretable exploration of patterns, illustrated here using the student dropout dataset. The central bubble plot displays mined patterns with the dropout probability (target variable) along the y-axis and the mean value of a selected attribute-here, *SAT Math* along the x-axis. Bubbles are sized by data coverage and colored by effect on the target: green for higher dropout probability, and red for lower. The selected bubble represents a pattern characterized by lower *SAT Math* scores and an 80% dropout risk. To the left, the feature importance chart shows that *SAT Math* has a medium global and high pattern importance. The probability distribution chart compares the dropout rates of the selected pattern (blue) against the full dataset (gray), and the statistics panel summarizes the number of students and their outcomes. Below the bubble plot, the core attributes - *SAT Math* and *Admit Type* - are shown with Shapley contributions and a value histogram for a selectable variable shows how many data points in each value bin meet the pattern target variable condition (orange bars) and how many do not (blue bars). On the top right, the group summary panel shows how all the dataset features are distributed within this pattern.

its impact on the target variable.[1] We replaced p-value reporting with a simple statement indicating significance, as all patterns were statistically significant and users showed little interest in their exact values.

The final component, the pattern summary panel, appears to the right of the bubble chart. This view addresses the user need to examine attributes beyond the core defining features. It presents a list of all dataset attributes, each accompanied by a Shapley value bar (green for positive, red for negative) and three bins representing the attribute's distribution within the pattern. For numerical attributes, the full value range is divided into three bins, colored on a white-to-blue scale based on data density (from none to all). For categorical attributes, the first two bins show the top two categories, and the third bin groups the remaining values. The summary is split into two sections: core attributes (those that define the pattern) and non-core attributes, allowing users to quickly distinguish between driving factors and contextual variables.

User Feedback and Limitations: The bubble plot interface was seen as more intuitive and accessible, especially by domain experts unfamiliar with advanced visualization techniques. Most users were able to navigate and interpret the interface effectively. Like the projection-based design, it supported comparison across patterns based on size and target effect, but offered greater interpretability through explicit axes. Users appreciated the feature importance chart for guiding attribute selection and found the summary statistics informative. However, still several key limitations emerged:

- *Single-feature limitation:* The x-axis could only show one feature at a time, making it difficult for users to reason about patterns defined by multiple attributes or to see how attributes interact. This constrained exploration to a narrow slice of the feature space.
- *Summary distortion:* Bubbles were positioned based on the mean value of a selected feature within the pattern, which often failed to reflect the full constraint range. This led to misinterpretation – users sometimes assumed a pattern was centered around that value, when its range was broader or skewed.
- *Clutter in dense views:* In datasets with many patterns, bubbles frequently overlapped or crowded together along common feature values, making individual selection and comparison challenging. This reduced readability and made fine-grained pattern exploration difficult.

[1] For a pattern, not all data points must fulfill the target condition; only the *average* of all points in the pattern must meet the condition at statistical significance.

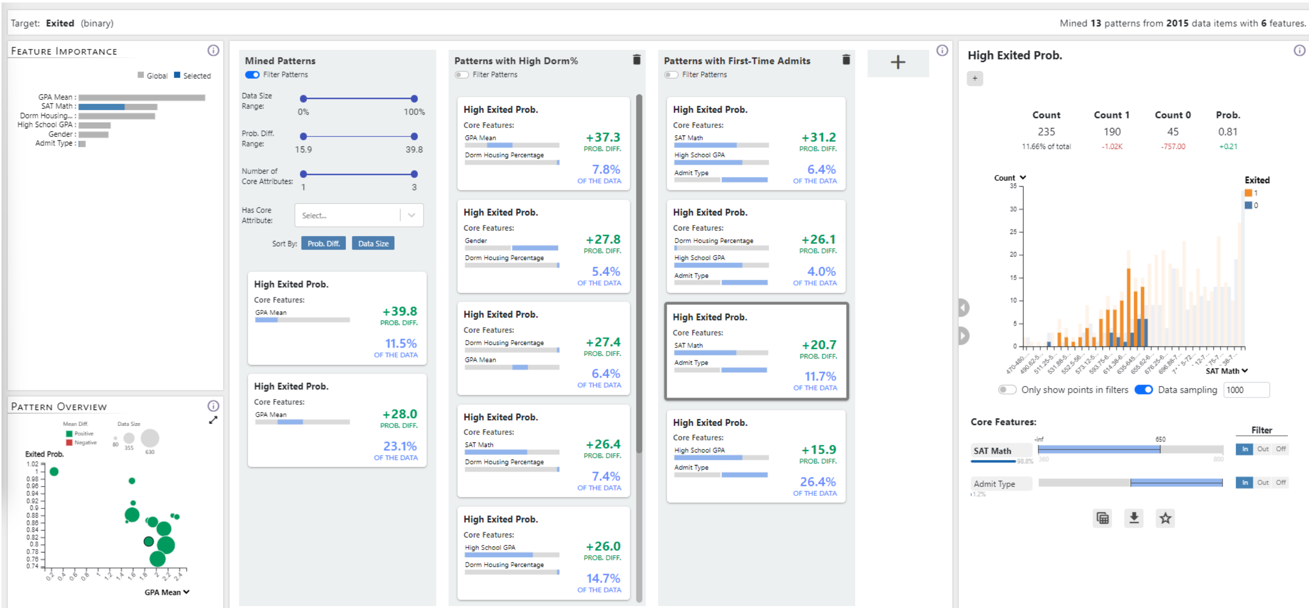


Figure 3: The third design iteration is a card-based pattern browser interface, shown here for the student dropout dataset. Here, patterns are represented as compact cards arranged within customizable, scrollable lists at the center of the interface. In this example, the default "Mined Patterns" list displays multiple pattern cards, each summarizing their effect on dropout rate, data coverage, and the top three defining attributes. The user has also set up two additional lists "Patterns with High Dorm %" and "Patterns with First-Time Admits". The selected card highlights a pattern involving low *SAT Math* scores and transfer *Admit Type*, together associated with a 20.7% increase in dropout rate and covering 11.7% of the student population. To the right, the pattern detail panel reveals further information: a comparison of dropout statistics between the pattern and the full dataset, a plot comparing distributions (similar to the histogram in Figure 2 bottom right), and an editable list of attribute constraints with contribution scores. On the left, the feature importance plot and bubble chart are carried over from the previous interface.

- *Unstructured comparison:* While the axis-aligned layout improved interpretability, the interface lacked tools to sort, group, or cluster patterns meaningfully. Users had to rely on memory or manual scanning, limiting their ability to identify relationships or organize patterns around domain-relevant themes.

Takeaway: Switching from abstract projections to a familiar axis-based layout improved comprehension and usability by aligning with familiar visual representations. However, limiting users to one attribute view at a time constrains interpretability for multi-dimensional patterns and overloads the user's memory. These challenges highlighted the need for an interface that combined compact summaries, multi-attribute clarity, and user-driven organization, all of which informed the next design iteration.

4.4. Iteration 3: Card-Based Interface

The final design iteration introduced a card-based layout, inspired by prior approaches that present data subsets or patterns using card-like visual representations Knittel et al. (2021); Strobelt, Oelke, Rohrdantz, Stoffel, Keim, and Deussen (2009); Wu, Liu, Guo, and Wu (2023). In this interface (Figure 3), the spatial encoding is replaced by a Kanban-style layout of *pattern cards* arranged in columns. This redesign was driven by user feedback seeking improved interpretability, reduced cognitive load, and support to organize large sets of patterns. The card-based design offers an accessible and interpretable summary-based approach that does not rely on dimensionality reduction or advanced visualization literacy.

Pattern cards are easily readable modular infographic-like summaries of individual patterns (Figure 4) (**DG2**). Cards display the pattern's effect on the target variable, the percentage of data it covers, and its top three constraining attributes. Target effect and coverage are prominently displayed as textual values on the right-hand side of the card for quick reference. The top three constraints are visualized as horizontal range bars, offering a glanceable view of the pattern's core definition. To preserve compactness, additional constraints (if present) are summarized as a count below the range bars. Clicking on a card reveals a detailed pattern view with full constraint information and statistical summaries.

Multiple patterns are displayed as individual pattern cards organized within a scrollable column in a Kanban-style interface. By default, all mined patterns are placed in a single column labeled *Mined Patterns*, but to support custom organization (**DG3**), users can create additional columns with custom labels and drag-and-drop pattern cards across them. This flexible structure allows users to group, prioritize, and arrange patterns according to their analytical goals.

To aid in pattern exploration and the identification of high-interest subgroups (**DG1**), each column includes built-in tools for sorting, filtering, and tagging. Users can sort cards based on effect size, data coverage, p-value, or number

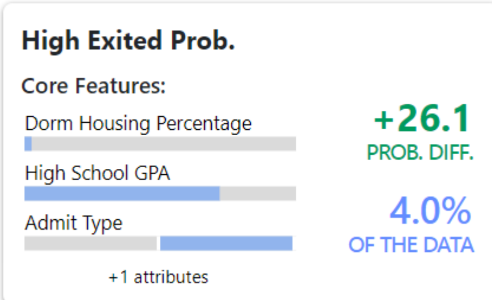


Figure 4: A card representation of a pattern. Here, the pattern covers 4% of the data or students who have an increase of 26.1% in the probability of exiting the CS major as compared to the overall probability. It is defined by 4 attributes with the top 3 attributes and their ranges shown in blue.

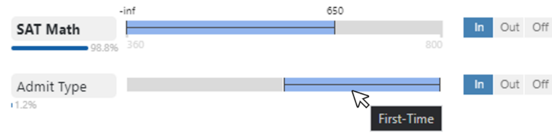


Figure 5: We visualize a pattern's attribute constraints using bullet-style charts Coelho et al. (2020). Black whiskers indicate the constraint range defining the pattern (e.g., *First-time admits* and *SAT Math* below 650), while the blue highlight shows a user-defined filter. Here, the filter is set to show data *in* the pattern, so the blue range and black whiskers overlap.

of constraints. Filtering options allow narrowing the visible set based on attribute inclusion, statistical thresholds, or pattern categories (e.g., generalizations, specializations). Users can mark patterns as "Interesting" by clicking a star icon on the card, or assign custom text tags for labeling. Tags can also be used as filters to focus on subsets of patterns across the workspace.

Selecting a pattern card opens a detail panel on the right side of the interface, allowing users to inspect the pattern (**DG2**) and interactively edit its defining constraints (**DG4**). The panel is composed of three main components that support exploration, interpretation, and refinement.

The first component presents summary statistics for the target attribute within the selected pattern, including values such as mean, median, and support. Below these, the panel displays differences from the global dataset statistics, using color cues to enhance interpretability: green for positive deviations and red for negative ones. We omitted the natural language description and significance statement from the previous iteration, as the card design effectively conveyed this information and all patterns were inherently significant.

The second component visualizes the distribution of the target attribute, enabling users to compare local and global patterns. For numerical targets, a scatterplot shows the spread of values inside and outside the pattern, with opacity used to distinguish the selection. For nominal targets, bar charts or histograms are employed. Data points within the pattern are shown with full opacity, while points outside appear faded. For binary target variables, a consistent color scheme is used – blue for class *0* and orange for class *1*, similar to the histogram in Figure 2 bottom right.

The third component presents a concise, interpretable summary of the pattern's attribute constraints using a bullet-chart–based design introduced by Coelho et al. Coelho et al. (2020). As shown in Figures 3 and 5, this view supports both numerical and nominal attributes. Each attribute is shown with its name, a blue contribution bar, a detailed range bar, and a toggle control for interactive editing. The blue bar beneath the attribute name represents the attribute's contribution to the pattern's effect on the target variable, computed using Shapley values, enabling users to quickly identify influential features. The range bar builds on the simplified encoding used in pattern cards by incorporating additional visual cues for the global distribution of values and the subset of data points that satisfy the constraint. The toggle control allows users to disable the constraint or invert its range, supporting exploratory "what-if" analysis and enabling fine-tuned customization of patterns (**DG4**).

The detail panel can be expanded for deeper analysis, as shown in Figure 6. The expanded view retains the core elements of the standard detail panel but introduces a reorganized layout and two key additions. First, the attribute range list now includes *other features* – attributes not involved in defining the pattern. While these attributes do not constrain the pattern, their value ranges across data points within the pattern are visualized and provide a broader context.

Second, the interface introduces three horizontal pattern lists – *Enhance*, *Generalize*, and *Related Patterns* – to support pattern exploration and comparison (**DG3**). The *Enhance* list contains more specific variants of the current pattern that retain the same core features, but with narrower value ranges or additional constraints. The *Generalize* list presents broader patterns that match a subset of the core features, with value ranges that encompass those of the current pattern. Finally, the *Related Patterns* list surfaces patterns that share a significant number of data points with the selected pattern, even if their defining features differ.

In addition to the card-based list and detail views, the interface preserves the bubble plot and feature importance chart from the previous design (Figure 2). These components function as before, allowing users to view and sort patterns by statistical attributes or visual position. Selected patterns are highlighted in both the card view and the bubble plot, supporting bi-directional interaction. This continuity in design ensures a smooth transition for users familiar with the earlier interface, while enriching the experience with improved organization and pattern discovery tools.

User Feedback and Limitations: Users responded positively to the card-based interface, describing it as significantly more readable, visually simple, and easier to navigate than the projection and bubble plot designs. They reported that identifying relevant patterns was faster, and that the layout felt more natural for reviewing and organizing results.

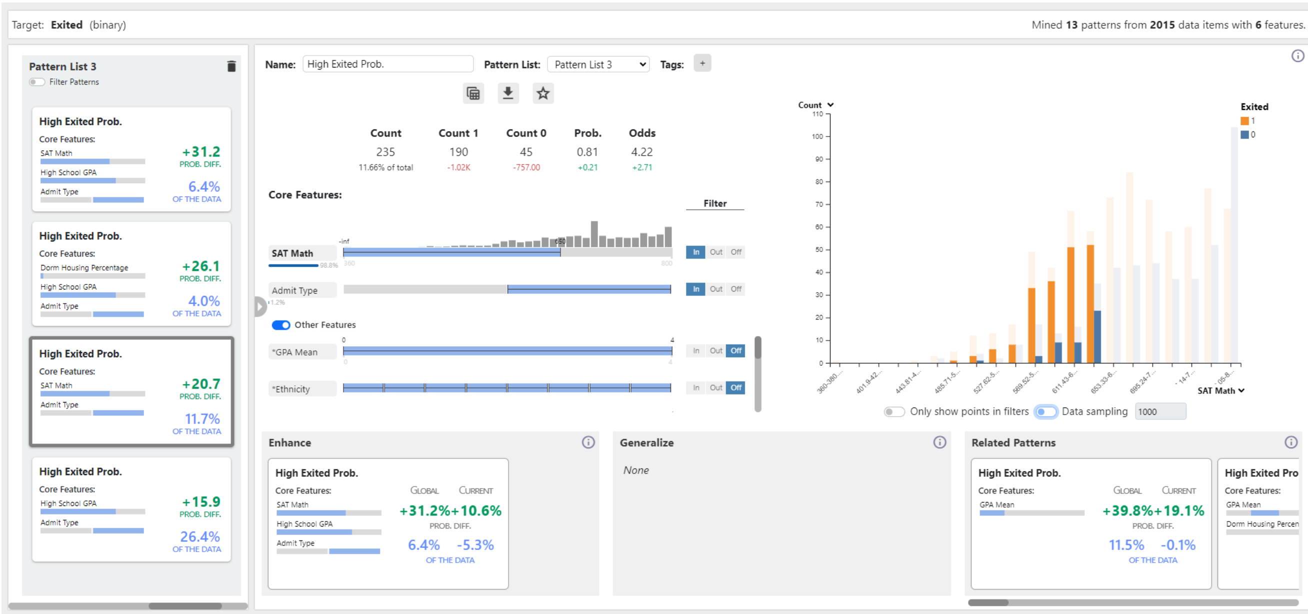


Figure 6: Expanded view of a pattern's details. This panel presents summary statistics, detailed attribute ranges for both core and non-core features, a histogram showing the distribution of the selected attribute with pattern data points highlighted, and three horizontally arranged lists of related patterns that enhance, generalize, or share significant data overlap with the current pattern. In this example, the selected pattern is defined by two constraints: *First-time admits* and *SAT Math* below 650. The *Enhance* list includes a more specific variant of this pattern that adds a third constraint: *High School GPA*.

- *Improved interpretability:* The card format presented information in a familiar, scannable structure. Users appreciated the plain-language pattern descriptions and compact visual summaries, which reduced the need for mentally decoding complex visuals or statistical jargon.
- *Multi-pattern organization:* Sorting, filtering, and tagging tools gave users control over pattern grouping and prioritization. These features helped reduce information overload and allowed users to focus on subsets aligned with analytic goals or domain hypotheses.
- *Effective customization:* The ability to interactively edit attribute constraints within each card enabled users to experiment with modifications and instantly assess how changes affected key metrics. This supported exploratory, hypothesis-driven workflows and fostered trust in the results.

Takeaway: The card-based interface improved usability and interpretability over earlier designs by replacing abstract visual encodings with structured, user-centric summaries. This shift supported flexible pattern review, precise customization, and meaningful comparisons across results, making it especially effective for real-world data analysis. While the design requires extensive scrolling to navigate large sets of patterns, this tradeoff proved worthwhile as users reported higher satisfaction and demonstrated a clearer understanding of the design.

4.5. Summary and Lessons Learned

Through three design iterations, our interface evolved from abstract projections to axis-aligned plots, and finally to card-based summaries. Each step improved usability and interpretability by addressing user feedback and aligning more closely with real-world analytic needs. The projection-based view provided rich context but was inaccessible; the bubble plot improved interpretability but was limited in multi-attribute reasoning and organization; and the card-based interface offered the most effective balance of readability, customization, and comparison. While it requires more scrolling, this tradeoff proved worthwhile as users reported greater satisfaction and clearer understanding. Overall, the progression highlights the value of moving from complex visual encodings toward structured, user-centric summaries for pattern exploration.

5. Case Study: Patterns of Student Retention

To evaluate the usability and effectiveness of our evolving interface, we conducted a longitudinal case study in collaboration with a large public university. Over several months, a team of analysts – four Computer Science (CS) faculty members and institutional researchers – used successive iterations of our tool to investigate a pressing institutional question: Why do students leave the CS major?

The case study served as both a real-world application testbed and a comparative evaluation, allowing us to observe how different visual representations influenced users' ability to interpret, compare, and act on patterns mined from high-dimensional student data.

5.1. Dataset and Pattern Mining Setup

The dataset comprised over 17,000 student records spanning more than a decade. Each record had the attributes *SAT Math Score, SAT EBRW Score, High School GPA, High School, Percentage of time in dorm housing, Mean GPA across all courses completed, Admit Type (first-time or transfer), Ethnicity, Gender,* and *Dropout*. As these are real administrative data, we have slightly modified images showing them for this paper to protect privacy, while maintaining the outcomes and effectiveness of our method.

University analysts were interested in finding patterns related to the *Dropout* attribute - a binary variable indicating whether the student continued with the major (1) or exited (0). Using our mining engine, analysts identified 13 statistically significant patterns associated with a higher-than-average CS major exit rate. The analysts then explored and analyzed these patterns using each version of the interface.

5.2. Observations Across Interface Iterations

5.2.1. Projection-Based Interface

Users began by exploring the patterns using the projection-based interface shown in Figure 1, starting with the *Pattern Map* to identify interesting patterns. However, most users - novice and experienced - struggled to interpret the projection. They frequently asked questions like, “What do the axes represent?” and “Why are these patterns close together if they have different dropout effects?” Even CS professors with theoretical knowledge of projection maps expressed confusion and skepticism.

Due to the difficulty in interpreting the map, users resorted to selecting patterns based on bubble size and color alone, then inspected pattern details in the accompanying views. When examining the pattern in the Data Context Map (DCM), they understood that the contour enclosed the relevant data points, but were confused about the placement and meaning of individual points. They asked questions such as, “I get that the contour shows the pattern, but what are the GPA values of the points in this area?” or “Why are students who live in different dorms plotted so close together?”

Overall, the abstract nature of the projection led to limited exploration and few actionable insights. Users often disengaged after initial confusion, relying instead on the summary statistics and constraint distributions in the pattern detail panel to make sense of each pattern.

5.2.2. Bubble Plot Interface

The bubble plot interface marked a noticeable improvement in usability and comprehension. Users quickly understood the axis-aligned layout and appreciated the ability to select features, inspect distributions, and toggle attribute views. They responded positively to the feature importance plot, which helped them identify influential variables more easily. Several users explicitly noted that it helped structure pattern navigation early in their exploration.

However, limitations persisted. One key issue was the inability to view multi-attribute constraints simultaneously, which made interpreting patterns with multiple core features cognitively demanding. As one analyst noted: “I can see that the SAT score is low, but what about the dorm variable? I have to flip back and forth and remember.” Users also questioned the reliance on mean values for positioning bubbles. While the x-axis helped clarify feature trends, it obscured important distributional context. One user remarked: “This bubble says SAT Math is 620, but the range is broader. Is there a way to see whether the range is broad or narrow in the bubble plot itself?” We experimented with showing whiskers on bubble hover to indicate range, but users found the interaction cumbersome and preferred seeing constraint ranges directly without extra effort.

Despite these issues, users began mentally clustering patterns and noting recurring themes - such as low high school GPA, dorm location, and transfer status - as risk factors for dropout. They requested more advanced filtering tools to support this reasoning, asking for the ability to filter patterns based on specific attribute values and then reorient the view. For example: “Can we filter to only low GPA scores and then see how those patterns are distributed by dorm housing?” Although the bubble plot improved clarity, it still placed a high cognitive load on users to manually organize and compare patterns.

Users continued to rely heavily on the pattern detail view, particularly the summary statistics that contrasted pattern-level metrics with those of the entire dataset. They appreciated the clarity this provided: “It’s good to know exactly how much of the dataset is covered by this pattern and what the baseline dropout rate is.”

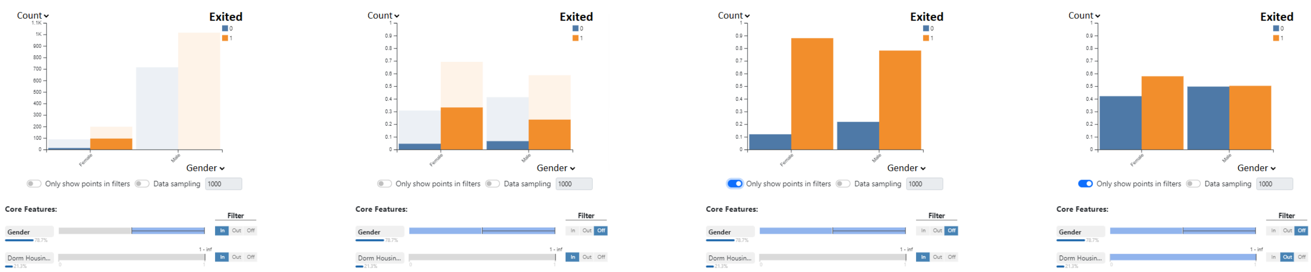
However, the pattern summary panel that shows how all dataset features (pattern constraints and other attributes) behaved in the pattern was not well-received. Binning attribute distributions into just three groups was perceived as overly reductive. Users reported that this abstraction obscured actual distributions, especially skewed distributions, and made them less confident in the visual representation.

5.2.3. Card-Based Interface

With the final interface, user engagement improved dramatically. The card-based representation enabled analysts to quickly scan patterns, view their top three constraints, and understand the associated ranges at a glance. Combined with the Kanban-style layout, users were able to organize patterns into meaningful lists such as “High Dorm %,” “First-Time Admits,” and “Low Academic Performance.” The ability to rename lists and tag cards made it easy to mark, group, and revisit patterns of interest. This flexibility significantly reduced users’ cognitive load during exploration.

Initially, patterns were presented in a single scrollable list, which introduced some interaction burden. However, integrated sorting and filtering tools mitigated this issue by allowing users to explore patterns based on constraint attributes, statistical thresholds, or target effects. These tools helped users navigate large pattern sets more efficiently and focus their attention on specific areas of interest.

Notably, the nature of user feedback shifted. Questions were no longer about understanding the interface but rather



(a) (b) (c) (d)

Figure 7: Interactive filtering applied for one specific pattern of two attributes - Gender and Dorm Housing Percentage - to study the effect of time spent in the dorm on the rate of exiting the CS major for male and female students. (a) shows the distribution of this pattern - female students with high (full-time) Dorm Housing Percentage - as visualized by the blue bars in the respective two bullet charts. (b) changes the y-axis from counts to probability and sets the Gender filter to 'Off', showing both genders staying in the dorm full-time; the noticeable discrepancy of the pairs of bars shows that in general females in full-time dorm housing are considerably more likely to exit the CS major than males. (c) shows only the points within the filter which confirms this assessment. (d) inverts the filter for Dorm Housing Percentage, now showing the exit rate for both genders of students who spent only some time on campus. Here the exit vs. non-exit rate is more balanced, although the females still have a somewhat higher probability of exiting.

about interpreting the data and extracting insights. Less than two minutes into using the interface, one analyst noted: "I noticed that there are a lot of patterns with a high range of Dorm Housing Percentage." Another quickly responded, "Let's put them in their own list and compare them later." They proceeded to create a new list, apply filters, and sort the patterns before analyzing them in detail. A similar workflow emerged when users observed a split between First-Time Admits and Transfer Students again leading patterns being grouped based on Admit Type. These behaviors indicated that the interface was intuitive, lowered the barrier to entry, and supported productive, insight-driven exploration of complex pattern spaces.

Using the attribute range toggles and histograms (Figure 7), they discovered: Female students who spent 100% of their time in dorms had an exit rate of 90%, compared to 80% for males. Students living off-campus had substantially lower exit rates across all demographics (details in caption).

These insights led to new institutional questions about dorm-specific student experiences, prompting the team to conduct a follow-up mining session focused solely on female students. That secondary analysis revealed that students in certain dorm buildings had significantly better retention outcomes, likely due to community or support factors.

5.3. Summary of Impact

The evolutionary design process of our visual analytics interface significantly improved the interpretability and usability of pattern mining and exploration for high-dimensional data. Each design iteration, from projection-based to bubble plots to card-based layouts, brought tangible benefits and revealed new limitations, which were systematically addressed. As interface iterations progressed, we observed a clear shift in the type and depth of user inquiry. With the card-based interface, users stopped questioning the visual encoding and instead focused on domain-relevant

Table 2
Comparison of Interface Design Iterations.

Iteration	User Questions
Projection-Based	"What does it mean if two patterns are close to each other?", "Why are both green and red patterns near this attribute?"
Bubble Plot	"How do I see the effect of gender on patterns?", "How do I check if these patterns have similar constraints?"
Card-Based	"Can we see how this pattern behaves only on females?" "Can we turn off admit type for this pattern and look at the effect?"

questions and comparative pattern reasoning, a key indicator of interface success. Examples of these questions are shown in table 2.

The final card-based interface enabled domain experts to meaningfully explore and organize complex pattern spaces, uncover non-obvious insights, and formulate new domain-specific hypotheses without requiring deep expertise in data mining or visualization techniques. The case study with university analysts highlighted the practical utility of the interface. Users rapidly identified key factors associated with student attrition and grouped patterns in intuitive ways, facilitating collaborative analysis and institutional action. Importantly, the interface not only supported interpretation but also inspired deeper inquiry and additional mining efforts, demonstrating its role in enabling iterative, insight-driven analytics.

6. Discussion

The iterative development of our system, driven by user feedback and practical analytic needs, progressed through three major design stages: from projection-based plots to axis-aligned plots, and finally to a card-based summary layout. While the initial projection-based approach offered rich context, it was difficult for users to interpret. Subsequent designs systematically simplified the visual encodings, culminating in the card-based layout, which users preferred for its clarity, customization, and accessibility despite requiring more scrolling. The design progression revealed not only strengths and limitations of specific visualization techniques but also broader lessons about aligning system capabilities with users' cognitive needs and analytic workflows.

Balancing complexity and interpretability. Projection-based views offered rich spatial context but required a level of visualization literacy that many users lacked; even those familiar with dimensionality reduction found them difficult to interpret. Axis-aligned plots improved interpretability but constrained multi-attribute reasoning, as users could only examine a limited number of metrics at once. The card-based design ultimately struck a balance by prioritizing readability and organizational flexibility over dense encodings, despite occupying more screen space and requiring scrolling.
This progression highlights a recurring challenge in visual analytics: sophisticated visual abstractions may be powerful for experts, but accessible, structured views often better serve broader audiences.

The role of trust and cognitive load. A key insight from our study is that user trust in mined patterns is closely tied to clarity of presentation. Overly complex encodings eroded confidence, while scannable summaries and explicit comparisons fostered engagement and deeper exploration. Reducing cognitive load through filtering, tagging, and decomposition proved critical in shifting analysts' attention from deciphering the interface to reasoning about the data.

Domain-driven insights. The case study on student attrition demonstrated that our interface not only facilitated pattern discovery but also supported collaborative sensemaking and hypothesis generation. The ability to group patterns into domain-relevant categories (e.g., dorm housing, gender) enabled analysts to align computational results with institutional knowledge. This interplay suggests that effective interfaces for pattern mining should not only present results but also provide mechanisms for domain experts to contextualize and reframe them.

Generalizability: Our initial motivation for developing this tool was to provide an intuitive user interface for interacting with the pattern mining algorithm described in Section 3. However, the final interface is model-agnostic and can be adapted to any pattern mining or subgroup discovery method. Fundamentally, it provides a way to visualize and compare different sets of attribute constraints within a dataset, regardless of how those subgroups are generated. For instance, it can be integrated with tree-based models such as decision trees or random forests, where each pattern corresponds to a set of conditions along a branch or subtree. In such cases, feature importance could be derived using measures like the Gini index rather than Shapley values, which are more suitable for model-agnostic contexts.

The card-based interface is also inherently more generalizable across data domains than projection-based methods, which often require selecting and tuning specific dimensionality reduction techniques Cashman et al. (2025). By emphasizing structured, interpretable summaries rather than abstract spatial encodings, the design principles demonstrated here can extend to diverse analytical settings. For example, in a behavioral analytics scenario, the interface could support studies of user activity on shopping websites, where each card represents a set of filters or conditions applied to product searches by individual users or user groups. In this context, cards would summarize behavioral subgroups—such as users filtering by price range, brand, or category—allowing analysts to compare and interpret usage patterns at scale.

Limitations. Despite these advances, several challenges remain. As with most visualization interfaces, there are trade-offs between clarity, scalability, and information density. The scatterplot and bar chart representations used across all three iterations are intuitive and widely recognized but do not scale well as the number of attributes increases. Scatterplots, in particular, can suffer from overplotting when many data points share similar values, reducing their effectiveness for large datasets.

The current interface was designed primarily for data where the target or dependent variable is numeric or nominal. Likewise, the attribute range bars were optimized for these data types. While this approach supports a wide range of analytical scenarios, it is less suitable for geographic or temporal attributes, where alternative visual forms—such as maps or timeline-based views—may convey relationships more effectively. Exploring such data-specific representations remains an avenue for future work.

Pattern scalability also presents ongoing challenges. The projection- and bubble-based views in the first two iterations can lead to visual clutter and cognitive overload as the number of patterns increased. The card-based interface mitigates these issues by presenting patterns as individual, interpretable cards; however, this approach trades overplotting for increased scrolling. When the number of patterns grows into the hundreds or thousands, navigation and organization can still become difficult, even with sorting and filtering.

Future work should therefore explore more scalable and adaptive solutions – such as pattern aggregation, semantic summarization, and intelligent recommendation mechanisms – to help users focus on the most relevant and meaningful patterns. These directions could enhance efficiency and scalability while preserving the accessibility and trust that guided our design evolution.

Future directions. We see opportunities to extend this work in several ways. First, tighter integration with machine learning pipelines could allow analysts to not only interpret mined patterns but also use them for model validation and inspection, fairness assessment, or feature engineering. Second, incorporating collaborative features such as shared workspaces could better support group analysis and decision-making. Finally, exploring multimodal explanations – combining text, geomaps, visuals, and interaction – could further lower barriers for domain experts unfamiliar with statistical reasoning.

7. Conclusion

In conclusion, our study underscores that designing effective tools for high-dimensional pattern exploration requires balancing statistical rigor with interpretability, usability, and audience awareness.As our design evolution showed, complex visualization techniques may not suit broad user groups, whereas clear and cognitively lightweight representations can foster engagement, trust, and faster adoption. By documenting this evolutionary design process, we not only deliver a production-ready interface but also offer design principles that can inform the development of future visual analytics tools bridging machine learning outputs and human sensemaking.

References

P. J. Huber, Projection pursuit, The Annals of Statistics 13 (1985) 435–475.

J. Stahnke, M. Dörk, B. Müller, A. Thom, Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions, IEEE Transactions on Visualization and Computer Graphics 22 (2016) 629–638.

V. Sivaraman, Z. Li, A. Perer, Divisi: Interactive search and visualization for scalable exploratory subgroup analysis, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25, Association for Computing Machinery, New York, NY, USA, 2025. URL: https://doi.org/10.1145/3706598.3713103. doi:10.1145/3706598.3713103.

M. Espadoto, R. M. Martins, A. Kerren, N. S. T. Hirata, A. C. Telea, Toward a quantitative survey of dimension reduction techniques, IEEE Transactions on Visualization and Computer Graphics 27 (2021) 2153–2173.

D. Cashman, M. Keller, H. Jeon, B. C. Kwon, Q. Wang, A critical analysis of the usage of dimensionality reduction in four domains, IEEE Transactions on Visualization and Computer Graphics 31 (2025) 9405–9423.

L. G. Nonato, M. Aupetit, Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment, IEEE Transactions on Visualization and Computer Graphics 25 (2019) 2650–2673.

H. Jeon, H.-K. Ko, J. Jo, Y. Kim, J. Seo, Measuring and explaining the inter-cluster reliability of multidimensional projections, IEEE Transactions on Visualization and Computer Graphics 28 (2022) 551–561.

J. Seo, B. Shneiderman, A rank-by-feature framework for unsupervised multidimensional data exploration using low dimensional projections, in: IEEE Symposium on Information Visualization, 2004, pp. 65–72. doi:10.1109/INFVIS.2004.3.

R. Faust, D. Glickenstein, C. Scheidegger, Dimreader: Axis lines that explain non-linear projections, IEEE Transactions on Visualization and Computer Graphics 25 (2019) 481–490.

S. Cheng, K. Mueller, The data context map: Fusing data and attributes into a unified display, IEEE Transactions on Visualization and Computer Graphics 22 (2016) 121–130.

S. Liu, D. Maljovec, B. Wang, P.-T. Bremer, V. Pascucci, Visualizing high-dimensional data: Advances in the past decade, IEEE Transactions on Visualization and Computer Graphics 23 (2017) 1249–1268.

R. A. Becker, W. S. Cleveland, A. R. Wilks, Dynamic Graphics for Data Analysis, Statistical Science 2 (1987) 355 – 383.

A. Inselberg, The plane with parallel coordinates, The Visual Computer 1 (1985) 69–91.

L. Wilkinson, A. Anand, R. Grossman, Graph-theoretic scagnostics, in: IEEE Symposium on Information Visualization, 2005. INFOVIS 2005., 2005, pp. 157–164. doi:10.1109/INFVIS.2005.1532142.

E. Kandogan, Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions, in: IEEE Information Visualization Symposium, volume 650, 2000, p. 22.

P. Hoffman, G. Grinstein, K. Marx, I. Grosse, E. Stanley, Dna visual and analytic data mining, in: IEEE Visualization, 1997, pp. 437–441.

I. Jolliffe, Principal component analysis, Encyclopedia of Statistics in Behavioral Science (2005).

W. S. Torgerson, Multidimensional scaling: I. theory and method, Psychometrika 17 (1952) 401–419.

L. v. d. Maaten, G. Hinton, Visualizing data using t-sne, Journal of machine learning research 9 (2008) 2579–2605.

L. McInnes, J. Healy, N. Saul, L. Großberger, Umap: Uniform manifold approximation and projection, Journal of Open Source Software 3 (2018) 861.

T. Fujiwara, X. Wei, J. Zhao, K.-L. Ma, Interactive dimensionality reduction for comparative analysis, IEEE Transactions on Visualization and Computer Graphics 28 (2022) 758–768.

M. O. Ward, Multivariate Data Glyphs: Principles and Practice, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 179–198.

D. Guo, Coordinating computational and visual approaches for interactive feature selection and multivariate clustering, Information Visualization 2 (2003) 232–246.

J. Li, C.-q. Zhou, Incorporation of human knowledge into data embeddings to improve pattern significance and interpretability, IEEE Transactions on Visualization and Computer Graphics 29 (2023) 723–733.

T. Zhang, J. Li, C. Xu, Visual exploration of multi-dimensional data via rule-based sample embedding, Visual Informatics 8 (2024) 53–56.

J. Yang, A. Patro, S. Huang, N. Mehta, M. O. Ward, E. A. Rundensteiner, Value and relation display for interactive exploration of high dimensional datasets, in: IEEE Symposium on Information Visualization, 2004, pp. 73–80.

D. Kammer, M. Keck, T. Gründer, A. Maasch, T. Thom, M. Kleinsteuber, R. Groh, Glyphboard: Visual exploration of high-dimensional data combining glyphs with dimensionality reduction, IEEE Transactions on Visualization and Computer Graphics 26 (2020) 1661–1671.

J. Knittel, A. Lalama, S. Koch, T. Ertl, Visual neural decomposition to explain multivariate data sets, IEEE Transactions on Visualization and Computer Graphics 27 (2021) 1374–1384.

M. P. Neto, F. V. Paulovich, Multivariate data explanation by jumping emerging patterns visualization, IEEE Transactions on Visualization and Computer Graphics (2022) 1–16.

H.-P. Kriegel, P. Kröger, A. Zimek, Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering, ACM Transactions on Knowledge Discovery from Data (TKKD) 3 (2009) 1–58.

M. Atzmueller, Subgroup discovery, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 5 (2015) 35–49.

J. Han, J. Pei, Y. Yin, Mining frequent patterns without candidate generation, ACM SIGMOD Record 29 (2000) 1–12.

K. O. McGraw, S. P. Wong, A common language effect size statistic., Psychological bulletin 111 (1992) 361.

D. Coelho, N. Gupta, E. Papenhausen, K. Mueller, Patterns of social vulnerability - an interactive dashboard to explore risks to public health on the us county level, in: Workshop on Visual Analytics in Healthcare (VAHC), 2022, pp. 01–09.

K. Mueller, E. Papenhausen, Patterns of philanthropy: Using pattern mining for predictive analysis in advancement and fundraising, Journal of Advancement Analytics 2 (2022) 101–115.

Y. Chung, T. Kraska, N. Polyzotis, K. H. Tae, S. E. Whang, Slice finder: Automated data slicing for model validation, in: 2019 IEEE 35th International Conference on Data Engineering (ICDE), 2019, pp. 1550–1553. doi:10.1109/ICDE.2019.00139.

X. Zhang, J. P. Ono, H. Song, L. Gou, K.-L. Ma, L. Ren, Sliceteller: A data slice-driven approach for machine learning model validation, IEEE

Transactions on Visualization and Computer Graphics 29 (2023) 842–852.
H. Strobelt, D. Oelke, C. Rohrdantz, A. Stoffel, D. A. Keim, O. Deussen, Document cards: A top trumps visualization for documents, IEEE Transactions on Visualization and Computer Graphics 15 (2009) 1145–1152.
J. Wu, D. Liu, Z. Guo, Y. Wu, Rasipam: Interactive pattern mining of multivariate event sequences in racket sports, IEEE Transactions on Visualization and Computer Graphics 29 (2023) 940–950.
D. Coelho, H. He, M. Baduk, K. Mueller, Eating with a conscience: Toward a visual and contextual nutrition facts label, VisComm: Workshop on Visualization for Communication (2020).

Darius Coelho holds the Ph.D. degree in computer science from Stony Brook University (SUNY). His research interests include visual analytics, human-computer interaction, data science, and medical imaging at the intersection of AI. He has over 4 years of industry experience and has served in founding roles at two startups. He has published over 15 peer-reviewed research articles with over 90 citations to date.

Eric Papenhausen holds the Ph.D. degree in computer science from Stony Brook University (SUNY). He has over 5 years of industry experience in software engineering at small and large companies alike and over 8 years cofounding multiple startups. He has authored 20 papers in performance optimization and visual analytics with over 170 citations to date.

Klaus Mueller is currently a professor of computer science at Stony Brook University and a senior scientist at Brookhaven National Lab. His research interests include explainable AI, visual analytics, data science, and medical imaging. He won the US NSF Early CAREER Award, the SUNY Chancellor's Award for Excellence in Scholarship and Creative Activity, and the IEEE CS Meritorious Service Certificate. To date, his 300+ papers have been cited over 14,500 times.