Deconstructing Categorization in Visualization Recommendation: A Taxonomy and Comparative Study

Doris Jung-Lin Lee, Vidya Setlur, Melanie Tory, Karrie Karahalios, Aditya Parameswaran

Abstract—Visualization recommendation (VisRec) systems provide users with suggestions for potentially interesting and useful next steps during exploratory data analysis. These recommendations are typically organized into categories based on their analytical actions, i.e., operations employed to transition from the current exploration state to a recommended visualization. However, despite the emergence of a plethora of VisRec systems in recent work, the utility of the categories employed by these systems in analytical workflows has not been systematically investigated. Our paper explores the efficacy of recommendation categories by formalizing a taxonomy of common categories and developing a system, *Frontier*, that implements these categories. Using *Frontier*, we evaluate workflow strategies adopted by users and how categories influence those strategies. Participants found recommendations that add attributes to enhance the current visualization and recommendations that filter to sub-populations to be comparatively most useful during data exploration. Our findings pave the way for next-generation VisRec systems that are adaptive and personalized via carefully chosen, effective recommendation categories.

Index Terms—Visual analysis; analytical workflow; discovery-driven analysis; visualization recommendations.

1 Introduction

Exploratory visual analysis is an iterative process of asking and answering questions about data through visualizations, where new questions often arise from unexpected observations. Challenges arise when the current analysis path does not yield interesting observations; this common pain point can cause users to feel stuck or overwhelmed, unsure of what question to ask next [1], [2]. Visualization recommendation (VisRec) systems guide users along their exploration journey by suggesting effective visual encodings [3], [4], [5] or potentially interesting visualizations [6], [7], [8], [9], [10], [11].

Recommendations are often organized into categories based on the *analytical actions* they embody. *Analytical actions* can be thought of as transitions between visualization states, corresponding to the operations performed to generate recommendations given the current visualization state. Example categories include `Filter`, displaying recommendations of sub-populations of the data, derived by adding filters to the current visualization, and `Enhance`, displaying recommendations of an additional attribute added to the current visualization. Figure 1 illustrates how recommendation categories can support the analysis of a college dataset. A scatterplot of `SATAverage` and `AverageCost` can be *filtered* to specific `HighestDegrees` or `Regions` (right side, top) or *enhanced* by adding the `FundingModel` attribute to the color channel (right side, bottom). Users can explore their data via moves in the visualization space, selecting visualizations based on analytical actions that represent potential "next steps" in their analysis.

Most VisRec systems display a small set of bespoke analytical-action-based recommendation categories without a clear understanding of why the set was selected. This limited selection of categories in existing systems stems from challenges in both *development* and *evaluation*. From an evaluation standpoint, determining the value of a given recommendation for a specific user goal is, in general, a challenge in recommender system design [12], but doing so for visual analysis tools is even harder. Unlike web search, where the typical goal is to find a single item (e.g. a movie to watch), insights in visual analytics often arise from multiple visualizations. This design is further complicated by the variety of user goals, ranging from specific inquiries to more complex open-ended objectives such as understanding relationships across attributes [13]. From a development standpoint, there is an interface design and performance cost for dynamically computing large numbers of recommendations [14]. As a result, most systems rely on a small set of fixed recommendation categories.

While prior work has certainly demonstrated the benefits of VisRec in supporting exploratory analysis [4], [5], [6], the design space of VisRec categories has not been thoroughly explored and evaluated. With new VisRec systems being introduced time and again in the visualization and HCI literature [15], there is a pressing need to take a step back to organize and make sense of the design space of analytical-action-based categories in VisRec, and to further evaluate the benefits of various types of recommendation categories as a whole, and relative to each other. This evaluation is crucial for distilling design guidelines for next-generation VisRec systems and for enabling past, present, and future VisRec systems to be understood and compared in the context of an organization.

In this paper, we deconstruct categorization in VisRec systems by comparing and evaluating the value of different recommendation categories in visual analytic workflows. We further investigate how analysis strategies are influenced by employing recommendation categories, as well as

- *D. Lee and A. Parameswaran are with the University of California, Berkeley. V. Setlur and M. Tory are with Tableau Research. K. Karahalios is with the University of Illinois, Urbana-Champaign.*

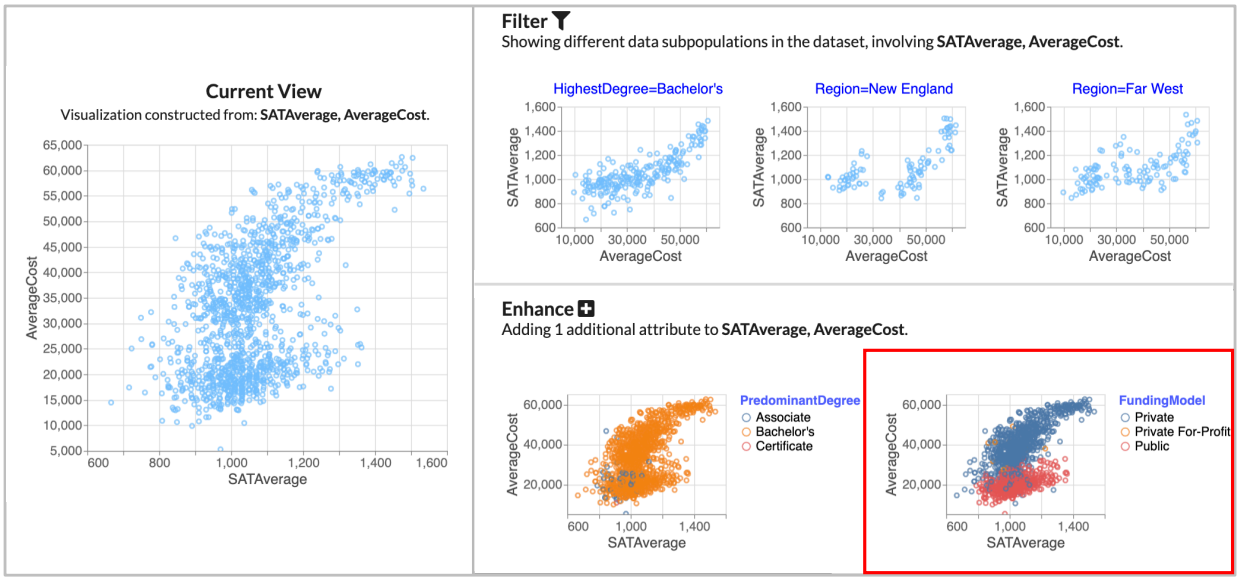

Fig. 1: A screenshot of *Frontier* with a dataset containing college information. Starting from the Current View displaying a scatterplot of AverageCost versus SATAverage on the left, the user finds an interesting visualization recommended through the `Enhance` category highlighting the two distinct clusters for Private and Public FundingModels (shown with a red border). This recommendation is generated from the Current View, further "enhanced" by adding FundingModel to the color channel.

the efficacy of various recommendation categories for different task and dataset characteristics. While recommendation categories such as `Filter` and `Enhance` are in fact present in prior systems [2], [4], [5], [7], [16], [17], there has been no systematic organization or comparison of recommendation categories and their underlying analytical actions. Another challenge is that no existing VisRec system comprehensively implements the space of possible categories to compare their effects on analytical workflows. This crucial gap in existing literature motivated our system, *Frontier*[1]. We developed *Frontier* as an apparatus for investigating the merits and pitfalls of various recommendation categories in a single system.

Our contributions are summarized as follows:

- We present a taxonomy of common analytical action-based recommendation categories employed in visual analysis, synthesizing existing literature from VisRec and online analytical processing (OLAP). The taxonomy enables us to map out the design space of existing VisRec systems as well as future ones. (Section 3)
- We develop a design probe, *Frontier*, implementing ten recommendation categories from the taxonomy to explore the usage and impact of these categories in a visual analysis workflow. (Section 4)
- We present a mixed-methods user study to understand how recommendation categories support visual analysis and the relative efficacy of various recommendation categories. (Section 5, 6)

As part of this work, one of our goals was to take stock of and systematize research in the rapidly-evolving area of VisRec systems. Our findings validate prior conjectures about the value of categorizing recommendations [4], [10] and further reveal how there are substantial differences in the usage of various categories. For instance, participants indicated that recommendation categories `Enhance` (adding one attribute) and `Filter` (displaying data subpopulations) were most useful, while `Pivot` (swapping an attribute) was one of the least useful. Such findings point to the importance of comparative evaluation across categories and guidelines for improving category design in future VisRec systems presented in Section 8.

1. The name *Frontier* is inspired by how the application of analytical actions offer next steps along potential exploration paths—enabling an explorer to expand the *frontier* of discovery. In the context of graph search algorithms, the term *frontier* refers to the nodes that lie between what has been discovered and those as yet undiscovered [18].

2 Related Work

This work is draws from work on VisRec systems, faceted search interfaces, and traditional recommender systems.

2.1 Visualization recommendation systems

Manual visualization specification tools [19], [20] require the user to specify the exact data attributes, data subset of interest, and the visual encodings for creating a visualization. This process is often tedious and overwhelming during the early, exploratory stages of analysis, especially for users without visualization design or data exploration experience [2], [5], [7]. To alleviate this issue, VisRec systems suggest visualizations to assist users with visual analysis.

VisRec systems can be classified based on whether they suggest visual encodings (i.e., encoding recommenders) or aspects of the data to visualize (i.e., data-based recommenders) [21]. The earliest VisRec systems focused on recommending visual encodings, assuming that the data attributes were already identified by the user [3], [22]. Subsequent work automatically recommended graphical encodings based on perceptual effectiveness [23], [24]. Our focus

in this work is on data-based recommenders, as in many recent papers [8], [9], [10], [11], [16], [25], [26], that suggest interesting visualizations based on statistical properties of the data. To narrow the design space of recommendation categories, we leverage visualization best practices for determining visual encodings. Throughout the rest of this paper, we use the term VisRec system to refer to those that employ data-based recommendations.

Some VisRec systems are entirely automatic [9], [27], whereas other, more recent, mixed-initiative systems support some user interaction to guide the recommendations [5], [10], [26]. Mixed-initiative VisRec systems combine manual specification with recommendations [7], [17], [28], [29], [30], [31], [32], [33]. For instance, Voyager [4], [5], [21] suggests visualizations based on user-selected fields and wildcards to iterate through possible data attributes or encodings. Many of these systems organize the resulting recommendations into categories based on their analytical actions [4], [5], [7], [10], [29]. For example, DIVE [7] and Voyager display a set of univariate distributions to help users get an overview of the distributions that exist in the data as well as recommendations that add an additional attribute to the user-selected visualizations. On the other hand, Zenvisage [26] searches for visualizations based on their similarity with a user-specified pattern over a space of possible filter values. While intuitively valuable, it remains unclear what the effects of different categories are on users and which categories are most helpful. Our work specifically investigates analytical action-based recommendation categories and how they impact mixed-initiative visual analytics workflows.

2.2 Faceted search interfaces

Facets help users quickly refine their search options by applying multiple filters based on a categorized classification of the information elements in web and product search [34], [35] and text corpora querying [36], [37], [38] interfaces. Faceted search interfaces bear similarities to VisRec systems in that both support progressive disclosure and incremental construction of queries where users can formulate the equivalent of a sophisticated data query through a series of small, exploratory steps [35], [39]. One key problem in building a faceted search interface is selecting the facets. Some systems simply present the first few alphabetized facets [40]; others present a subset of facets ranked by frequency of use [41], [42], [43]. Prior VisRec systems have drawn an analogy between faceted browsing and exploratory visual analysis [4], [6]. Similarly, our work draws inspiration from principles of facets as recommendation categories to address the complexity of analytical tasks [13]. While facets are constructed based on aspects of items in the corpus (e.g., size, brand for products; topics for articles), in VisRec, categories can be organized based on their relationship with the current visualization or visual characteristics as discussed in more detail in Section 3.

2.3 Recommender Systems

Unlike search systems that aim to maximize the relevance of retrieved items, the goal of a recommender system is to help users discover items of interest. As a result, an ideal recommender system must strike the right balance between suggesting items that are relevant as measured by recommender accuracy, and those that are diverse, and are therefore surprising and unexpected.

This diversity-accuracy tradeoff has been well-studied in the recommender system literature and has led to metrics beyond accuracy such as serendipity, novelty, coverage, and diversity [48], [49], [50]. The tradeoff manifests itself in visual data exploration as well, where users often want to discover non-obvious, unexpected insights, but still want to see visualizations relevant to the attributes or values that they are interested in. The context-dependent actions in the taxonomy described next, are examples of visualization recommendations relevant to the user's context; context-independent actions can reveal more surprising aspects about the data. Akin to diversification in traditional recommender systems, our paper highlights the importance of surfacing a relevant, yet diverse set of potential next steps that satisfies the user's information needs in a visual analytic workflow.

Category	Related Work
Distribution	[4], [5], [7], [9], [10], [11], [44]
Correlation	[5], [6], [7], [9], [10], [45], [46]
Enhance	[4], [5], [7], [17], [33]
Generalize (attribute)	[7], [11], [33]
Generalize (filter)	[7]
Pivot	[7], [17], [33]
Filter (add)	[16], [17], [27]
Filter (swap)	[2]
Difference	[8], [27], [30]
Similarity	[2], [47]

TABLE 1: Survey of recommendation categories from 20 VisRec systems. We included systems that recommend visualization(s) based on the result of some analytical action.

3 Taxonomy of Recommendation Categories

Analytical actions correspond to transitions through the visualization space to generate categories of recommendations given a user's current visualization state. While various taxonomies [51], [52], [53] exist for describing the types of tasks (or *actions*) employed during visual analysis, as stated in Law et al. [47], we are not aware of any taxonomy that encompasses the types of data-based visualization recommendations that can be generated via analytical actions and are reflected in present-day VisRec system. This section defines such a taxonomy, providing a common vocabulary for the organizing principles behind recommendation categories. The taxonomy arose through a systematic review of 20 VisRec systems detailed in the supplementary material.

Through open and axial coding [54], we uncovered ten major types of analytical actions used to generate and group visualization recommendations in existing systems, summarized in Table 1. To keep the design space of recommendation categories tractable, we focused on data-based recommendations driven by the operational and characteristic transitions in the visualization design space. We codified these actions into a taxonomy as seen in Figure 2. The table is not intended to capture a comprehensive set of all analytical action types, but rather to synthesize the most

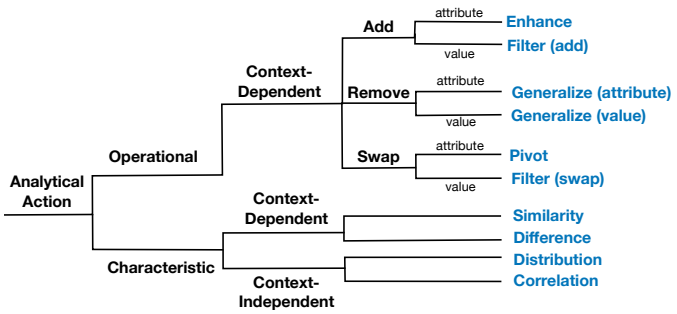

Fig. 2: A taxonomy of common analytical actions used in recommending visualizations for visual analysis. The analytical actions are indicated in blue.

common ones used to categorize recommendations so that we can explore the interaction design space more deeply.

At its highest level, the taxonomy defines two main categories: *operational* and *characteristic*. The *operational* category describes analytical actions that navigate users through the visualization space via operations such as add, remove, and swap. The *characteristic* category describes actions that reveal certain characteristic patterns in the data, such as skewness and correlation. The taxonomy is further broken down into *context-dependent* and *context-independent* categories. Actions are *context-dependent* if they depend on the *current view* or visualization (i.e., the selected attributes, values, and visual encodings); they are *context-independent* if they do not depend on the *current view*.

Note that the operational actions described above overlaps with some of the categories in the chart transition model in GraphScape [24]. However, GraphScape provides a chart transition model that describes visualization edits, whereas our taxonomy describes actions in a VisRec context drawn from existing VisRec systems. Since our focus is less on encoding-based recommendations, we do not consider such aspects of the GraphScape taxonomy, such as Scale and Mark. Likewise, GraphScape does not include characteristic actions described in this taxonomy.

3.1 Operational Analytical Actions

Operational actions apply data-oriented operations that transition the current view to a related, neighboring part of the visualization space. By definition, analytical actions that are operational must also be *context-dependent* as they operate on the current view. As seen in Figure 2, there are three broad categories of operational actions based on whether an attribute or value is added, removed, or swapped, leading to six (3×2) individual categories.

The example in Figure 3 demonstrates how operational actions can be thought of as moving along different paths in the *attribute* or *value* hierarchy. Every node in the attribute or value hierarchy represents a set of selected attributes or values. A user's current visualization is composed of their position on the attribute hierarchy (i.e., the space of all attribute combinations) and their position on the value hierarchy (i.e., the space of all filter value combinations). Movements through these hierarchies defines the set of possible operational actions. This conceptual model formalizes the space of possible visualizations that are one move away from the current visualization. Our model draws on Online Analytical Processing (OLAP), a sub-field of data management that targets analytical querying of multi-dimensional data. However, unlike OLAP, which only considers the value hierarchy [55], we introduced the analogous attribute hierarchy to help capture common operations in visual analytics. Here are the six operational analytical action categories:

- **`Enhance`:** adds an additional attribute to the current view. If the user selects attributes A and B, `Enhance` displays visualizations involving attributes A, B, and C. This action corresponds to moving down the attribute hierarchy (Figure 3 red).
- **`Filter (add)`:** adds an additional filter to the current view. If the user selects attributes A and B, `Filter (add)` displays visualizations involving A, B, and a filter F. In OLAP [55], this is known as a *drill-down* on the value hierarchy (Figure 3 orange).
- **`Filter (swap)`:** switches out the filter value to a different value, while keeping the filter attribute fixed. If the user selects attributes A and filter $F = V$, `Filter (swap)` displays visualizations involving A and an alternative filter $F = V'$. This action corresponds to moving horizontally across the value hierarchy to a node with the same filter attribute (Figure 3 purple).
- **`Generalize (attribute)`:** removes one attribute from the current view to display the more general trend. If the user selects attributes A and B, visualizations involving either A or B are displayed. This action corresponds to moving up the attribute hierarchy (Figure 3 turquoise).
- **`Generalize (value)`:** removes one filter from the current view to display the more general trend. If the user selects attributes A and a filter F, visualizations involving only A are displayed. In OLAP, the removal of a filter is known as a *roll-up* on the value hierarchy (Figure 3 pink).
- **`Pivot`:** displays visualizations that can be constructed if one of the attributes from the current view is replaced with another attribute. If the user selects attributes A and B, `Pivot` displays visualizations involving either A and another attribute B', or B and another attribute A'. This action corresponds to moving horizontally along the attribute hierarchy (Figure 3 green).

3.2 Characteristic Analytical Actions

Characteristic analytical actions are designed to surface salient visual and statistical characteristics of the data, sorted based on an interestingess metric that is described in Section 3.3. Characteristic actions that are *context-independent* are designed with an overview intent, similar to "breadth-first" exploration strategies in web search [14] that are independent of the user's search query. We describe two types of independent actions that highlight patterns that may be of interest to the user:

- **`Correlation`:** highlights bivariate relationships between quantitative fields in the data through scatterplots of different combinations of quantitative attributes.
- **`Distribution`:** displays the possible univariate distributions in the dataset, with `COUNT` as the default measure. The visualization can either be a histogram, bar chart, or line chart depending on the data type of the attribute.

Characteristic actions that are *context-dependent* showcase salient visual characteristics based on the current view.

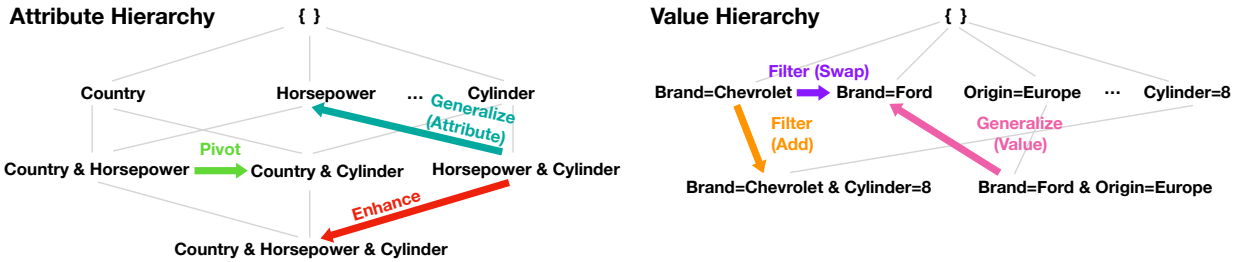

Fig. 3: Operational actions represent transitions through the attribute and value hierarchies.

- **`Similarity/Difference`:** highlights data patterns that are visually similar or different from the current view.

3.3 Ranking Objectives

Within each analytical action category, visualizations are often ranked using some interestingness objective. Given the different chart characteristics for different types of visualizations, the interestingness objective, even for a given action, may be different for visualization types. For example, a user may be interested in the degree of correlation in a scatterplot, while they may be interested in differences between the bar values in a bar chart. In this paper, we consider commonly occurring basic chart types, including bar charts, histograms, line charts, and scatterplots, typically employed by existing VisRec systems. Even this set results in a considerable number of choices corresponding to every combination of action and visualization type. We identified a small number of classes of objectives that have been used in prior work for these combinations, which we catalog below.

For the *characteristic* actions, the objective typically captures the salient visual characteristics expressed by a visualization, such as the degree of correlation or skew. Visualizations are ranked from the `Correlation` action based on monotonicity [25], typically most to least correlated, while those from the `Distribution` action are ranked from most to least skewed. For the `Similarity` action, bar and line chart visualizations are ranked based on similarity to the *current view*, computed via the Euclidean distance between the measure values of the visualizations [2], [26].

For the *operational* actions, the objective used is typically determined by the visualization type of the recommended visualizations. These objectives capture perceptual characteristics generally associated with something unexpected or insightful in the visualization, including:

- **Non-uniformity**: For *bar/line charts and histograms without a filter*, visualizations are ranked highly if they are highly uneven, indicating the presence of outlying categories or shifts in distributions [6], [10].
- **Deviation**: For *bar/line charts and histograms with a filter*, the ranking is based on the deviation between the filtered and unfiltered (overall) distributions, based on the intuition that a visualization is potentially interesting if it differs greatly from some expected reference [8], [16].
- **Correlation**: For *uncolored scatterplots*, a visualization is ranked higher if it displays a high degree of dependence between the two measures, as measured by mutual information [56], [57] or Spearman's correlation [25].
- **Separability:** For *colored scatterplots*, a visualization is ranked higher if the colors for each category distinctly separate clusters of data points in the scatterplot [6], [58].

Supplementary materials provide details as to how these objectives were implemented in our design probe, *Frontier*.

4 The Frontier System

We introduce a system, *Frontier*, that provides visualization recommendations across multiple analytical action-based categories. *Frontier* is a design probe that enables us to systematically explore and compare these categories. For brevity, we refer to the analytical action-based categories displayed in *Frontier* simply as *recommendation categories* henceforth. The *Frontier* interface is composed of four areas as illustrated in Figure 4. Starting from the left (Figure 4A), we have the Control Panel, a manual specification interface for specifying the visualization in the Current View (Figure 4B). The Control Panel lists measures and dimensions and allows users to add or remove attributes and values. The Specification Panel (Figure 4A top) allows users to fine-tune their visualization by arranging attributes across specific encoding channels. Users can toggle on and off specific recommendation categories. If a category is not applicable for the given Current View, the cursor icon changes to a forbidden sign upon hover in the Category Menu (Figure 4C). The recommendations are displayed row-by-row on the right (Figure 4D) analogous to faceted web search results to encourage browsing [35]. We drew

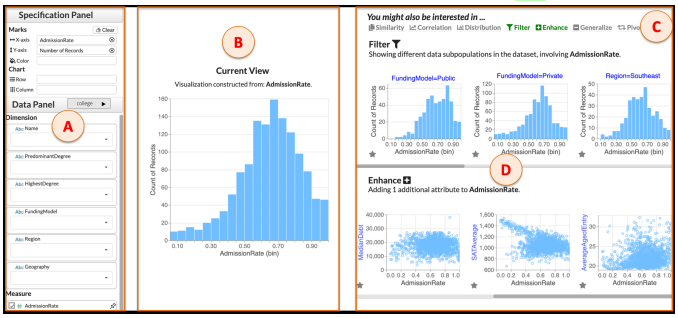

Fig. 4: *Frontier* consists of four areas: Control Panel (A), Current View (B), Category Menu (C), and Recommendations Panel (D).

inspiration from existing VisRec interfaces [5], [7], [10] and employed guidelines from mixed-initiative interfaces [59], [60] to balance interface usability with comprehensiveness in the display of recommendation categories. We iterated on the design with feedback from an interaction designer and made significant changes to the interface over a period of six months.

4.1 Design considerations

The following design considerations emerged while iteratively designing *Frontier*:

- C1: *Concise and informative.* Recommendation categories should provide a *manageable* set of options as "next-steps" in a user's analytical workflow. Users should never be shown an empty category nor should they be shown categories with overlapping recommendations.
- C2: *Coordinated and actionable.* Recommendation categories should be coordinated and consistent with other parts of the system, such as in the Category Menu and Current View. The user should be able to bring a recommended visualization into the Current View.
- C3: *Interpretable and visually discernible.* Recommendation categories should be self-explanatory and display visual indicators that convey their key characteristics or highlight how they differ from the Current View.

These requirements echo design considerations from prior work in mixed-initiative visual analytics systems [5], [45].

4.2 System Overview

Frontier is a web-based system with components described as follows. First, the *Data Manager* loads the dataset and metadata (i.e., the data type, data model, and default aggregation) and computes statistics (i.e., cardinality, correlation, minimum, maximum). The *Context Manager* maintains information about the attributes and values that the user has selected. Then the *Category Manager* determines which recommendation categories to display for a given Current View and maintains a list of categories. Finally, each *Category* contains information about specific recommendation categories, a sorted list of top-k recommended visualizations, and their associated scores. Details of the system architecture and implementation along with source code can be found in the supplementary materials.

4.3 Category Generation and Visualization Interaction

To select a manageable set of recommendation categories (C1) to display to users, we designed the following workflow. These rules are similar to the ones adopted in Voyager [4] and DIVE [7], which first provide an overview via univariate distributions, followed by more relevant visualizations based on subsequent user selection.

At the start of the analysis, when no attributes are selected, the `Correlation` and `Distribution` actions display univariate and bivariate visualizations, enabling users to get an overview of the dataset (Figure 5A, 5B). Operational categories evolve based on the Current View and come into play once any attribute is selected. To avoid redundancy across categories (C1), *Frontier* only shows context-independent categories when there are no attributes selected. For example, when a user selects a single quantitative attribute, `Enhance` (Figure 5F) generates a collection of scatterplots, similar to what is shown for `Correlation`. Similarly, when a categorical attribute is in the Current View, only `Pivot` recommendations (Figure 5E) are displayed, to avoid operating over the same collection of visualizations as the ones in `Distribution`.

To make the recommendation categories more succinct and manageable (C1), from Table 1, *Frontier* consolidates filter (add) and filter (swap) to give `Filter`, generalize (attribute) and generalize (value) to give `Generalize`, similarity and difference to give `Similarity`. *Frontier*'s `Filter` adds an additional filter to the Current View when there is no filter in the Current View (Figure 5G). When a filter is in place, `Filter` keeps the specified filter attribute, while swapping out one of the attribute values to showcase alternative data subsets for comparison. Note that any applied filter is always retained in all actions except in `Filter`. In `Generalize`, we display all possible visualizations by removing either one filter or one attribute that is in the Current View (Figure 5C). In `Similarity`, visualizations that look most similar to the Current View are ranked highest, but users can reverse the sort order to look at the most dissimilar visualizations (Figure 5D).

Users can double click any recommendation to bring the visualization into the Current View; this sets elements in the Control Panel to be consistent with that of the selected visualization (C2). We display the axis label of any element that differs between the Current View and the recommendation in blue, to ease comparison and highlight differences (C3).

5 Study Design

We conducted a mixed-methods study to explore how various recommendation categories impact visual analysis workflows. Our primary goal was to study the relative usefulness of various categories as well as how categorization influences the analysis workflow in general. To study these goals, our design probe, *Frontier*, implements the categories described in the previous section. Further, to understand the effect of categorization in mixed-initiative VisRec workflows, we included a mixed-initiative VisRec baseline, mirrored off of *Frontier*, that featured the same recommendations, but without any categorization. We describe this baseline later on. We opted against comparing with manual specification tools without recommendations, given the general benefits of VisRec in prior studies [4], [5], [6]. We also opted against comparing with existing VisRec systems that implement a subset of categories, as this would not allow us to tease apart the impact of categorization on analytic workflows.

Overall, our exploratory study aimed to address the following research questions:

- RQ1: How do recommendation categories support and influence analytical workflows? What problem-solving and exploration strategies do users adopt when using recommendation categories in a mixed-initiative context?
- RQ2: What are the differences in user behavior across recommendation categories? What is the value and impact of individual recommendation categories and how does this vary across tasks and datasets?

5.1 Participants

We recruited 24 participants (10 female, 14 male) from within a software company. Nine were experienced users of a popular, commercial charting tool, 13 had limited proficiency, and two had no experience. In a between-subjects design, participants were randomly assigned to use *Frontier* or *Baseline* with either the College [61] or Olympic

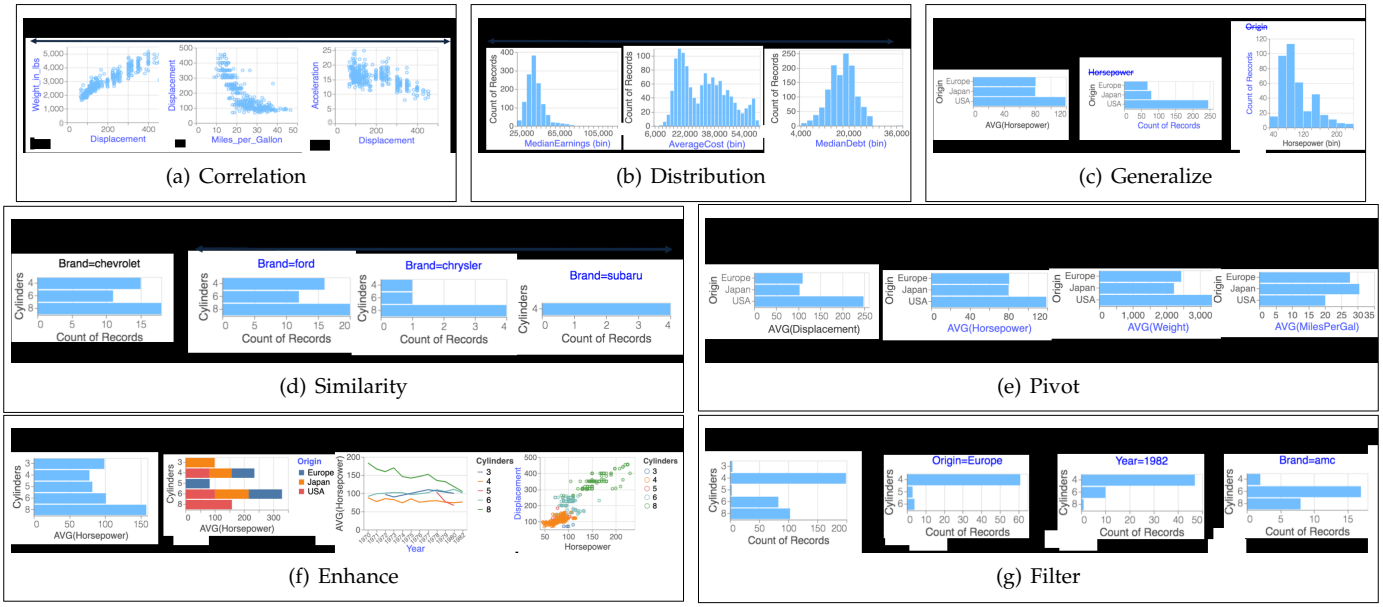

Fig. 5: Examples of various recommendation categories implemented in *Frontier*. (A) `Correlation` generates scatterplots with bivariate relationships between quantitative fields ranging from high to low correlation. (B) `Distribution` shows the possible univariate distributions from the dataset ranging from skewed to normal distributions. In the following examples, the current view is shown on the left, with the corresponding recommendations shown on the right. (C) `Generalize` shows possible visualizations when one attribute or filter from the current view is removed (removed attributes shown with strikethroughs). (D) `Similarity` highlights data patterns ranging from most to least similar to the current view. (E) `Pivot` shows possible visualizations that can be constructed if one of the current attributes is changed to another (changed attributes shown in blue). (F) `Enhance` shows possible visualizations when an additional attribute is added to the current view (additional attributes shown in blue). (G) `Filter` displays the data subsets that can be constructed from the current view when a filter is applied.

Medals [62] dataset, with six participants per condition-dataset combination. Henceforth, we suffix .F or .B in the identifier to display whether the participant used *Frontier* or the *Baseline* condition.

5.2 Tasks and Data

There were two main parts to the study: closed-ended tasks and open-ended exploration.

Part 1: Closed-ended tasks

Closed-ended tasks were mainly intended to familiarize participants with the system while also providing some consistent objectives for task comparison. Participants completed four closed-ended questions that included common visual analytic tasks, including:

- Q1 (*Correlate*): find other measures that are linearly correlated with a selected attribute.
- Q2 (*Filter Compare*): compare bar charts across different data sub-populations.
- Q3 (*One v.s. All*): compare a filtered distribution with the overall distribution.
- Q4 (*Pattern*): compare the temporal trend across different measures.

For each task, participants answered a multiple choice question on a paper worksheet. They were instructed to use *Frontier* to answer the question, but were not told how to do so. All participants used the Cars dataset [63] for closed-ended tasks. This dataset was chosen because of its simple schema (five measures and five dimensions), clean insight patterns, and because it is commonly used for demonstrating visualization systems [5], [6], [20], thus enabling comparisons.

Part 2: Open-ended exploration

Following the closed-ended tasks, participants completed an open-ended exploration task. This task enabled us to observe how people would choose to use (or not use) the recommendations in a natural analysis flow. Participants explored either the College or Olympic Medals dataset. Instructions were: *"We'd like you to explore this data to look for interesting insights. As you work, please let us know what questions or hypotheses you're trying to answer as well as any insights that you are learning about the data."* Participants were instructed to talk aloud and star recommendations they found useful.

The two datasets for open-ended exploration were chosen due to their real-world and accessible nature. We chose datasets with different characteristics, enabling us to study a wider range of analytical inquiries: the College dataset has ten measures and six dimensions with low to medium cardinality, while the Medals dataset contains only three measures and twelve dimensions with medium to high cardinality.

5.3 Apparatus

In the *Frontier* condition, participants used the full version of *Frontier* with all of the recommendation categories. The ordering of recommendation categories within the interface

was randomized for each user to minimize the preference for recommendations displayed at the top of the page.

To study the impact of categorization as a whole, we introduced a *Baseline* condition. This condition displayed the same set of recommendations except that the recommendation categories were removed so that all the recommendations appeared in a single, grid layout[2]. The goal of this *Baseline* is to establish a vanilla VisRec system that *(i)* eliminates the effects of recommendation categories, while preserving certain characteristics for a controlled comparison in that *(ii)* it is mixed-initiative and *(iii)* displays the same overall set of recommendations.

To understand this condition better, note that the organization of VisRec into categories is a result of both the labeling (i.e., interface elements such as dividers and textual descriptions of the categories) as well as the ranking within each category. To remove the effects of categorization *(i)*, we not only had to remove the category labels, but also had to shuffle the display order of the recommendations. In both conditions, participants can browse for more visualizations via horizontal scrolling (single scroll bar for the *Baseline*, one scroll bar per recommendation category for *Frontier*). To prevent preferential bias towards top-ranked visualizations, we further ensured that the exact same set of visualizations appeared with and without scrolling across both conditions.

While we acknowledge that our baseline is not perfect, we considered alternative designs, including a no-recommendations baseline or no baseline at all. However, a baseline that only removes the category labels, but does not alter the display order would only evaluate the effects of explicit category labels and is thus not a meaningfully different baseline. Since the goal of the study is not to demonstrate performance difference between *Frontier* and *Baseline*, we opted for a baseline with recommendations for investigating the research questions around recommendation categories.

5.4 Procedure

Sessions lasted approximately one hour, consisting of approximately five minutes of introduction and tutorial, 15 minutes of closed-ended tasks, 30 minutes of open-ended exploration, and 10 minutes of semi-structured interviewing. The tutorial video introduced the interface using the Cars dataset and stated that recommendations were selected based on an interestingness ranking and that the blue text indicated changes from the Current View. For *Frontier*, the video additionally described each recommendation category. The post-study interview included 7-point Likert scale questions (e.g., overall usability, recommendation usefulness) and open-ended questions on the system design and recommendations. Study scripts and protocols can be found in the supplementary material.

5.5 Analysis Approach

We employed a mixed-methods approach involving both qualitative and quantitative analyses. The primary focus of our work was a qualitative analysis of how recommendations of different categories influenced people's analytical workflows. We conducted a thematic analysis through open-coding of session videos, focusing on strategies participants took to answer their questions.

We thematically classified each participant based on how frequently they engaged with manual controls versus the recommendation panels. To obtain these classifications, we assigned separate labels for characterizing each participant's usage of the Control Panel and the recommendations (1: Majority of the time, 2: Sometimes, 3: Not often). Based on these labels, we grouped the participants by their relative frequency of use, where participants employed a *manual-oriented strategy* if they exhibit a higher usage of the Control Panel than recommendations, *balanced* if they had comparable usage of both, and *recommendation-oriented* if they exhibit a higher usage of recommendations than the Control Panel. Additionally, we define a visualization as *useful* if one or more of the following occurred: (a) the participant verbally described an insight, (b) the visualization was brought into view, (c) the visualization was starred, or (d) the participant expressed that it was useful or interesting. We coded insights from the video recordings, reusing the definition of an insight from prior work [44], [64].

The quantitative analysis consisted of Likert question results from the interview as well as counts of expressed hypotheses, data insights, and recommendations participants found useful. We employed statistical testing where appropriate, but considered the quantitative analysis mainly as a complement to our qualitative findings. We adopt a 95% confidence interval for all statistical analyses. Our analysis approach is similar to other studies that employed mixed-methods to investigate analytical workflows [44], [65].

6 Study Findings

6.1 RQ1: How do recommendation categories support mixed-initiative analysis workflows?

To understand how recommendation categories support analytical workflows, we first examine the strategies participants adopted and understand their motivations for switching between different modes of exploration. Then we delve deeper to examine the specific benefits of recommendation categories and their affordances. Finally, we highlight how user's perceptions regarding the recommendation categories can evolve over the course of an analysis workflow.

Strategies in mixed-initiative recommendation workflows

Based on thematic analysis of how frequently participants engaged with manual control versus recommendation panels, we observe three major strategies that they employed across both the *Frontier* and *Baseline* conditions. We generally observe that participants were more inclined to use recommendations in their workflow when using *Frontier* than in the *Baseline*. We sought to better understand participants' motivations for opting for different analysis options.

Participants employed a recommendation-oriented strategy for exploring unfamiliar attributes ($N_{F,B} = 6,3$ participants[3]) during preliminary analysis ($N_{F,B} = 5,4$) or

2. See screenshot in supplementary material. The *Baseline* interface has a similar layout to several existing VisRec systems [9], [28]

3. We use the notation $N_{F,B}$ to report measurements for *Frontier* and *Baseline* respectively. In the example above, $N_{F,B} = 6,3$ means that six participants using *Frontier* and three participants using *Baseline* used recommendations to explore unfamiliar attributes.

when they were out of ideas on what to pursue further ($N_{F,B} = 5, 1$). We found a small group of participants ($N = 3$ for *Frontier*; $N = 1$ for *Baseline*) who relied almost entirely on the recommendations to drive their analyses and used the Control Panel only sparingly. Most of these participants either expressed that they had limited experience with creating visualizations or were unsure what to expect from the dataset. The sentiment expressed by these participants largely corresponded to the challenges that visualization novices face in translating abstract questions about their data to visualization specifications [1]. As $P6.F$ explained *"...the recommendations gave me a jumpstart [...] because if I didn't have the recommendations to begin with, I wouldn't even know where to start."*

Participants also adopted a *balanced* strategy intermixing the use of the Control Panel and recommendations in unexpected ways. Three participants ($P7.F$, $P8.F$, $P17.B$) selected recommended visualizations that were "close enough" to what they wanted, then made minor tweaks using the Control Panel to attain their desired visualization. Participants also created familiar visualizations to trigger desired recommendations. For example, $P9.F$ wanted to look for linear trends in the data. They recalled seeing a clean linear trend between ACT and SAT scores previously, so they first created the same visualization via the Control Panel. Then they browsed through recommendations resulting from `Similarity` in order to find similar visualizations. Participants were able to leverage recommendations effectively in their workflow since the recommendation categories were transparent and interpretable, leading to predictable behavior.

Participants followed a *manual-oriented* strategy when the perceived cost of engaging in manual specification was lower than the effort it took to interact with the recommendations. This occurs when participants had a specific hypothesis in mind ($P12.F$, $P15.B$, $P16.B$, $P17.B$) or when participants expressed a preference for manual specification due to their familiarity with existing charting interfaces ($P10.F$, $P16.B$, $P22.B$). $P17.B$ explained the reason why they adopted a manual-oriented approach:

> *If the question that I want to answer is very clear, then I will go do it myself. There are two scenarios that I will switch from the left panel to recommendations. One thing is, I don't know what the next step is and I want insights. Second thing is, I don't know how to do it.*

As shown in Figure 6, a similar pattern is also observed for the strategies taken to solve the closed-ended task. In tasks where manual specification required significantly more work than simply browsing the recommendations for answers (*Correlate*, *Filter Compare*), participants were more likely to adopt a recommendation-oriented strategy. On the other hand, in the *One vs. All* task where participants had to compare a filter and unfiltered visualization, participants opted for the manual-oriented strategy as it was fairly easy to remove a filter.

Participants also adopted a manual-oriented strategy when the perceived effort to interact with recommendations was higher than usual, such as when they are overwhelmed by the large, unorganized panel of recommendations in the *Baseline*. This is supported by the post-study Likert ratings, where participants reported recommendations in *Baseline* to be less useful ($\mu_{F,B}$=4.58, 5.50; $\sigma_{F,B}$=0.90, 0.79; U=33.5, p¡0.05 via Mann-Whitney test) and more overwhelming ($\mu_{F,B}$=2.58, 1.76; $\sigma_{F,B}$=1.44, 1.76; U=58.0, p=0.21) than *Frontier*.

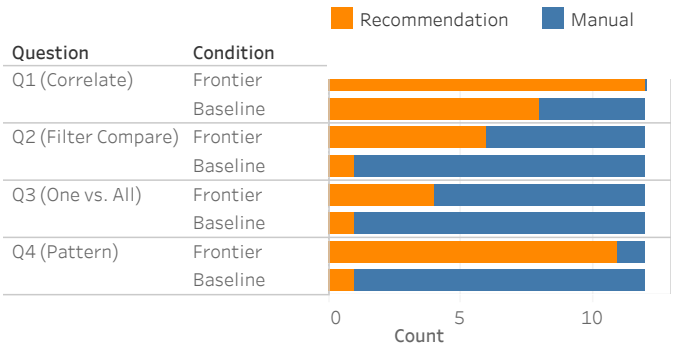

Fig. 6: The number of participants who took a manual-oriented approach in solving the closed-ended question versus a recommendation-oriented approach.

Value and impact of recommendation categories

We find that the presence of recommendation categories leads to richer and higher-utility exploration. During open-ended exploration, there were more insights generated via recommendations in *Frontier* than in the *Baseline* ($N_{F,B} = 171, 96$; t=2.66, $p< 0.05$). A similar trend was observed for the total number of useful visualizations generated via recommendations ($N_{F,B} = 149, 82$; t=2.47, $p< 0.05$), also shown in Figure 7 (top).

Our observations suggest that recommendation categories reduced overhead associated with interpreting the visualization recommendations. While we did not measure the visualization read-time directly due to the exploratory nature of the tasks, several qualitative observations support this idea. First, six out of 12 participants using *Frontier* expressed that they appreciated the organization. $P5$ noted: *"I like being able to really quickly visually inspect a bunch of things, because I can just slide a bunch of stuff past my eyes and be able to pick the stuff that jumped out."* In contrast, many participants in the *Baseline* condition went back and forth multiple times between visualizations to make comparisons and ensure that they had the right answer when solving the closed-ended questions ($P18$, 15, 25, 26), which, at times, led to mistakes.

During the study, we noticed that some participants appeared to be "stuck" in their analyses if they either: a) verbally expressed that they were out of ideas, b) implicitly when they had hypothesizing time of greater than one minute, or c) expressed reluctance to explore further. In particular, only one out of 12 participants using *Frontier* got "stuck" (once) compared to three participants getting "stuck" (total of five occurrences) in the *Baseline* condition. This is partly attributed to how *Frontier* participants repurposed and adapted their workflows to take advantage of the diverse set of actions available through various recommendation categories.

Participants often leveraged categories with the same axes such as `Enhance` and `Filter` to attain insights involving comparisons across multiple visualizations. For exam-

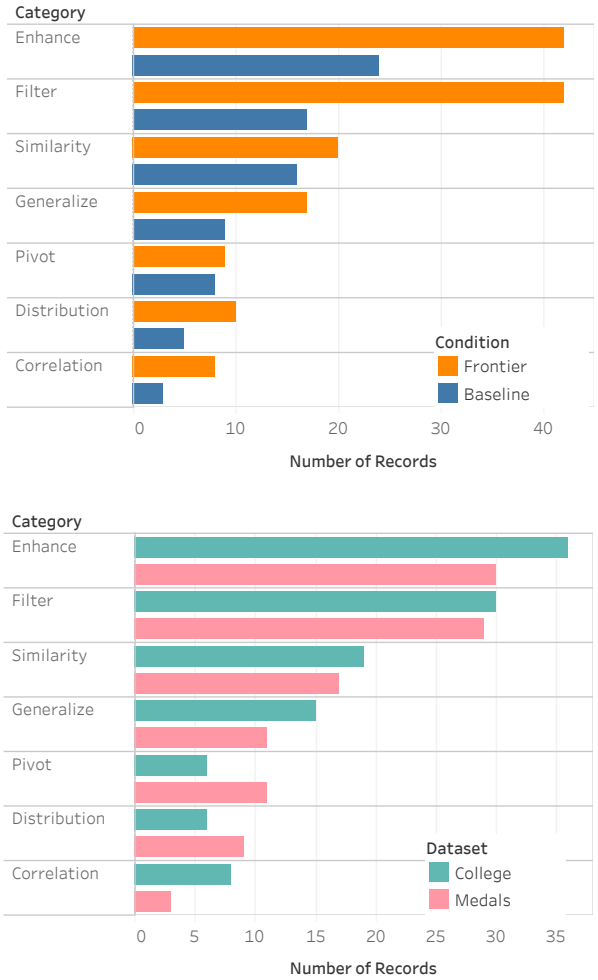

Fig. 7: Number of useful visualizations for each recommendation category by condition (top) and by dataset (bottom) for open-ended tasks. Top: `Enhance` and `Filter` are most useful in *Frontier*, but to a lesser extent in *Baseline*. Bottom: `Pivot` and `Distribution` are more useful in dimension-heavy datasets (e.g., Medals), whereas `Correlation` is more useful in measure-heavy datasets (e.g., College).

ple, $P10.F$ was interested in the age distribution of Russian athletes because of their highest medal count. They created a histogram distribution of age for Russia and browsed through the `Filter` action to see distributions for other countries. They exclaimed: "*Oh wow! Italy has some really old people for their medalists*". Seeing the Italy age distribution in the context of other age distributions highlighted its uniqueness; the visualization in isolation would have been uninteresting. Such comparisons across visualizations within an axes-consistent category are prevalent and often lead to better distributional awareness and understanding of the general patterns and trends in the dataset.

Evolving perceptions around recommendation categories

We found that participants came into the study with a diverse set of perceptions and expectations about recommendations that evolved throughout the course of the open-ended exploration session. For example, $P8.F$ explained that "*I feel like these suggestions require a lot more thought process in my head. So for the suggestions, because the one or two times that it doesn't seem a lot useful, I probably disregard it afterwards.*" $P4.F$ echoed a similar sentiment; several uninteresting visualizations early on deteriorated their confidence in `Correlation`: "*I thought that `Correlation` would be interesting. And it showed me like height and weight correlation and you heard me say I wasn't really interested in that. So it kind of made me nullify the entire `Correlation` panel together.*"

We also observed the reverse where participants with a negative initial impression of recommendations gained more trust and understanding over time. $P24.B$ expressed that they had a bias against recommendation systems and was reluctant to look at it. However, finding useful things from the recommendation encouraged them to adopt more of the recommendations in their workflow later on.

> *Before I even started the study, I have a bias about recommendation panels. Because most of the time recommendation panels do not show you what you want. So I'm already kind of wanting to do my own thing and do it myself, because my bias is that is more reliable than using a recommendation. [...] Once it started showing things, I was like, 'Oh, that is kind of interesting', or 'Oh, that is kind of relevant'. I started paying a little more attention to it. I had to keep reminding myself like, 'Oh yeah, this is the part of the screen I'm supposed to be looking at, it can actually be useful'.*

We also observed similar effects on a per-category level. For example, $P11.F$ discovered interesting insights based on `Filter` and noted in the subsequent analysis that they explicitly focused on the `Filter` category because they knew it would likely give something interesting. 11 out of the 24 participants also expressed that there was a learning curve in familiarizing themselves with the recommendation categories. A more longitudinal follow-up study is required to understand how users would interact with the recommendation categories when they become more familiar.

6.2 RQ2: What are the differences in usage and utility across recommendation categories?

Enhance and Filter were most useful, while Pivot least

As shown in Figure 7, some categories of recommendations were more useful than others. Somewhat surprisingly, while participants reported significantly more useful visualizations in *Frontier* than in *Baseline* especially for `Enhance` and `Filter`, we found that the relative ordering of usefulness for different recommendation types was largely the same independent of the condition. Note that while the categories were not explicitly shown in the *Baseline* interface, they were logged on the system side for the purpose of this analysis. As shown in Figure 7, `Enhance` and `Filter` are significantly more useful than `Pivot` (t=4.12, p¡0.05; t=3.24, p¡0.05 respectively via t-test).

In both conditions, we observed that participants manually performed analytical sequences that were similar to the analytical actions that produced the recommendations. We saw repeated patterns of participants manually performing the pivot operation on 14 separate occasions and the filtering operation on seven occasions. Given that `Pivot` was not actually regarded as very useful compared to the other categories (Figure 7), we suspect there may be a difference between the types of operation a user would like to perform

manually, versus ones that they would prefer being recommended to them. Regardless, these interaction patterns indicate that `Filter` and `Pivot` do indeed resemble the natural, intuitive "next-steps" in users' workflows.

Participants' post-study ratings of different categories in *Frontier* largely corresponded to the usefulness counts during the session in Figure 7. One exception is that `Correlation` and `Distribution` were perceived as easily understandable and useful by more than five out of the eight participants who provided a rating, but had low ranks relative to the usage frequencies captured in Figure 7. The discrepancy likely stems from the fact that these two context-independent categories are only displayed when the Current View is empty, so participants did not see them as frequently as the other categories.

Utility of recommendation categories across datasets

As shown in Figure 7 (bottom), the number of useful visualizations from certain categories depended on the dataset. While the usefulness of each category largely followed the trend observed in Figure 7 (left), the usage of `Pivot`, `Distribution`, and `Correlation` differed across datasets. Given that the College dataset contained large numbers of measures, while the Medals datasets had few measures and more dimensions, it was no surprise that there were more uses of `Correlation` in the College dataset ($N = 8$) than in the Medals dataset ($N = 3$).

On the other hand, `Distribution` was used more frequently with the Medals dataset ($N = 9$) than in the College dataset ($N = 6$), possibly because it also contained bar charts of the count distributions of dimensions (showing a surprising trend that Europe won significantly more medals than any other continent), whereas `Distribution` for the College dataset showed mostly histograms of measures. `Pivot` was also used more frequently in the Medals dataset ($N = 11$) than in the College dataset ($N = 6$), although the we were unable to determine the reason.

7 Study Limitations

While the analytical workflows that participants chose may be influenced by the category selection logic, we tried to minimize the effect of the display order on category preferences by randomizing the ordering of the various categories across users. We did not investigate the effects of recommendation ranking functions, but instead adopted standard data interestingness metrics from the literature [6], [10], [16], [46], [66]. Future work should explore how the interplay between ranking functions, visualization types, and category labels influences the usefulness of recommendations.

While we employed two different datasets to study task effects, future studies with more realistic dataset properties (large, higher-dimensional), larger sample sizes, different problem domains, and varying user expertise would be helpful. We have not explicitly controlled for visualization expertise, leading to more participants with limited proficiency. We also acknowledge potential novelty and unfamiliarity effects in our short one-hour study: most participants did not become fully fluent with all the categories and features provided in *Frontier*. The correlation between the 'warm-up' closed-ended task and the participant behavior may also be a potential confound. As a result, participants exhibited a strong affinity towards the Control Panel due to preconceptions of and familiarity with existing charting tools. A longitudinal study that examines how categorized recommendations are used in practice is important future work.

8 Design Implications

From our study findings, we first describe the guiding principles for the design of VisRec categories. We then discuss interface considerations for recommendation categories and their potential pitfalls. Finally, we identify opportunities for supporting analytical actions in visual analysis.

8.1 Design guidelines for recommendation categories

Evidence from our study shows how recommendation categories can be powerful constructs that help situate users by establishing a mental framework for reasoning about recommendation results. The semantic grouping and visual affordances of recommendation categories "lift" the visual analysis to operate at the level of analytical actions. By observing how participants switched between manual specification and recommendations, we find that predictable categories reduce users' perceived cost of employing recommendations — crucial for establishing an effective mixed-initiative workflow where recommendations can be used seamlessly in conjunction with manual specification.

Furthermore, the success of `Enhance` and `Filter` lends an important lesson for future VisRec systems in designing simple and readily-accessible recommendation categories. In particular, our study suggests that transparency and interpretability are essential characteristics that lead to recommendation categories that are predictably useful.

One heuristic for evaluating the complexity of a recommendation is to check whether the underlying action addresses a question involving a single element, which can either be a descriptor of the visualization's characteristics or an element that differs from the current view. For example, `Correlation` answers the question: "*Which attributes are correlated?*" and `Enhance` answers: "*What visualizations can be generated by adding one additional attribute?*" Participants' failure to adopt `Pivot` may be partially due to multiple degrees of freedom in which attributes could be swapped. Further, there may difficulty in articulating the analytical question that `Pivot` affords. This led to additional cognitive effort to reason about what was retained versus varied. Another contributor might be the drastic encoding change that can occur during swapping. This inconsistent behavior can be jarring and inhibits one's ability to compare across the collection. It remains an open question whether "anchoring" techniques that recommends appropriate encodings based on prior context [33] can be applied to recommended collections to offer a more consistent `Pivot`. Encoding inconsistency across collections is never an issue when swapping out values in `Filter` as the visualized attributes are unchanged when we move across the value hierarchy. This may explain why `Filter` was useful in providing complementary views on sub-populations of data, often leading to unexpected insights.

While our category selection algorithm takes an overview-first [67] approach in showing context-independent categories at the beginning and context-dependent recommendations once a specified view exists, several users explicitly cleared their selection in order to get the overview from time-to-time. Additionally, two participants indicated that they hoped to find visualization recommendations that were more unrelated and surprising rather than simple alternatives to their Current View. The diversity-accuracy tradeoff is a classical problem in recommender system design [68]. Supporting a blend of both types of visualization recommendations is a first step towards assisting users with different information preferences and needs. Further research is needed to develop and evaluate these recommendations as well as to validate our proposed taxonomy.

8.2 Pitfalls of categorized recommendations

While recommendations are at most an annoyance when they are not interesting to the user, they can be potentially detrimental if they are used to draw conclusions without deeper examination. $P3.F$ summarized this tradeoff between exploration and exploitation: "*For recommendation, sometimes you get completely irrelevant things, things that are kind of normal. But then on the other side, you get this serendipitous discovery, which is very cool. [...] I mean, it's also dangerous, because you maybe see something where you should further investigate it if it's really an effect.*"

Over-reliance on recommendations could be problematic. For example, $P17.B$ said that they had built trust that the system would show something interesting. When the interface did not show anything interesting, they quickly moved onto the next hypothesis instead of digging deeper because they inferred that the system was telling them not to look there. We observed a similar effect during a closed-ended task, where participants were asked to select the data sub-populations that had more 8-cylinder cars (Q2). When *Frontier* displayed only visualizations for three out of the four multiple choice answer options in `Filter`, many participants who employed recommendations simply drew their conclusions based on the three visualizations included in the category, without verifying the remaining one.

The potential for erosion of creativity and critical thinking when interacting with an intelligent system is well-known [6], [69]. While this issue is not particular to recommendations based on analytical actions, but a more general phenomenon with recommendations [70], the ease of use and the apparent trust that users perceive from recommendation categories may exacerbate these issues. This challenge points to a need for future research in designing recommendations that provide some notion of coverage or inform users about what has or has not been examined.

8.3 Towards personalized, adaptive recommendations

Even though categorized visualization recommendations provide a means of organization, limited screen space makes it impossible to show all categories at once. Furthermore, users typically only peruse the first few items of a recommended list [71]. The diversity of preferences and individual strategies observed in our study suggests that personalized selection of recommendations may be worthwhile For instance, while ten out of the 24 participants believed that there should be fewer recommendations, two participants (one from each condition) wanted to see more.

In post-study interviews, participants indicated that they wanted a more user-driven approach in creating their own organization. Three wanted the ability to extract selected recommendations into a separate dashboard, tab, or page and rearrange them freely into their own groups; eight wanted to hide some or all parts of the recommendation categories and retrieve them on demand. This points to an interesting future direction towards a hybrid human-recommender workflow. Similar to the variability of people's web search behavior [72], recommendations could be adaptive and personalized to tailor to users' preferences.

Personalization yields potential benefits beyond providing adaptive interfaces, namely, in providing optimization opportunities for system scalability. One of the challenges for visualization recommendations systems is the high computational cost associated with searching over a large search space of possible visualizations [6], [8], [73]. Given that there are preferences for certain categories over others for different users, datasets, and tasks, there is an opportunity to reduce the computational cost of an exhaustive search by pruning the recommendation search space.

9 Beyond The Final Frontier

The goal of recommender systems is in some sense to anticipate future user needs. Recommendation categories help organize these possible futures as readily-available options to drive analytical workflows. We introduced a taxonomy based on prior literature to examine the usefulness of various recommendation categories based on the underlying analytical actions. We implemented *Frontier* as a design probe to better understand how users push towards the frontier, taking next steps in their exploration. Our user study confirmed that recommendation categories are indeed useful for facilitating data exploration, helping users understand and interpret the visualizations. While the general utility of categorization was not surprising, we more deeply explored *how* various categories of visualization recommendations were employed and the diverse workflow strategies that users adopted. Design implications stemming from this study provide unique opportunities for supporting delightful user experiences with next-generation VisRec systems.

References

[1] L. Grammel, M. Tory, and M. Storey, "How information visualization novices construct visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 943–952, Nov 2010.

[2] D. J.-L. Lee, J. Lee, T. Siddiqui, J. Kim, K. Karahalios, and A. Parameswaran, "You can't always sketch what you want: Understanding sensemaking in visual query systems," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2019.

[3] J. D. Mackinlay, P. Hanrahan, and C. Stolte, "Show Me: Automatic presentation for visual analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1137–1144, 2007.

[4] K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe, and J. Heer, "Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 649–658, 2016.

[5] K. Wongsuphasawat, Z. Qu, D. Moritz, R. Chang, F. Ouk, A. Anand, J. Mackinlay, B. Howe, and J. Heer, "Voyager 2 : Augmenting Visual Analysis with Partial View Specifications," 2017.

[6] Z. Cui, S. K. Badam, A. Yalçin, and N. Elmqvist, "Datasite: Proactive visual data exploration with computation of insight-based recommendations," *CoRR*, vol. abs/1802.08621, 2018. [Online]. Available: http://arxiv.org/abs/1802.08621

[7] K. Hu, D. Orghian, and C. Hidalgo, "Dive: A mixed-initiative system supporting integrated data exploration workflows," in *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*. ACM, 2018, p. 5.

[8] M. Vartak, S. Madden, and A. N. Parmeswaran, "SEEDB : Supporting Visual Analytics with Data-Driven Recommendations," 2015.

[9] G. Wills and L. Wilkinson, "Autovis: automatic visualization," *Information Visualization*, vol. 9, no. 1, pp. 47–69, 2010.

[10] Ç. Demiralp, P. J. Haas, S. Parthasarathy, and T. Pedapati, "Foresight: Rapid data exploration through guideposts," *arXiv preprint arXiv:1709.10513*, 2017.

[11] J. Seo and B. Shneiderman, "A rank-by-feature framework for interactive exploration of multidimensional data," *Information visualization*, vol. 4, no. 2, pp. 96–113, 2005.

[12] S. M. McNee, J. Riedl, and J. A. Konstan, "Making recommendations better: An analytic model for human-recommender interaction," in *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 1103–1108. [Online]. Available: https://doi.org/10.1145/1125451.1125660

[13] T. Munzner, *Visualization Analysis and Design*, ser. AK Peters Visualization Series. CRC Press, 2015. [Online]. Available: https://books.google.de/books?id=NfkYCwAAQBAJ

[14] D. Tunkelang, *Faceted Search*. Morgan and Claypool Publishers, 2009.

[15] D. J.-L. Lee, "Insight machines: The past, present, and future of visualization recommendation," *Medium*, Feb. 2020.

[16] D. J.-L. Lee, H. Dev, H. Hu, H. Elmeleegy, and A. Parameswaran, "Avoiding drill-down fallacies with vispilot: Assisted exploration of data subsets," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, ser. IUI '19. New York, NY, USA: ACM, 2019, pp. 186–196. [Online]. Available: http://doi.acm.org/10.1145/3301275.3302307

[17] S. van den Elzen and J. J. van Wijk, "Small multiples, large singles: A new approach for visual data exploration," in *Computer Graphics Forum*, vol. 32, no. 3pt2. Wiley Online Library, 2013, pp. 191–200.

[18] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, Third Edition*, 3rd ed. The MIT Press, 2009.

[19] C. Stolte, D. Tang, and P. Hanrahan, "Polaris: A system for query, analysis, and visualization of multidimensional relational databases," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 1, pp. 52–65, 2002.

[20] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, "Vega-Lite: A High-Level Grammar of Interactive Graphics," 2017. [Online]. Available: https://vega.github.io/vega-lite/

[21] K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe, and J. Heer, "Towards a general-purpose query language for visualization recommendation," in *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*. ACM, 2016, p. 4.

[22] J. Mackinlay, "Automating the design of graphical presentations of relational information," *ACM Transactions on Graphics*, vol. 5, no. 2, pp. 110–141, 1986. [Online]. Available: http://portal.acm.org/citation.cfm?doid=22949.22950

[23] D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer, "Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 438–448, 2019.

[24] Y. Kim, K. Wongsuphasawat, J. Hullman, and J. Heer, "GraphScape: A Model for Automated Reasoning about Visualization Similarity and Sequencing," *Proc. of ACM CHI 2017*, 2017.

[25] L. Wilkinson, A. Anand, and R. Grossman, "Graph-theoretic scagnostics," in *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* IEEE, 2005, pp. 157–164.

[26] T. Siddiqui, A. Kim, J. Lee, K. Karahalios, and A. Parameswaran, "Effortless data exploration with zenvisage: an expressive and interactive visual analytics system," *Proceedings of the VLDB Endowment*, vol. 10, no. 4, pp. 457–468, 2016.

[27] A. Anand and J. Talbot, "Automatic Selection of Partitioning Variables for Small Multiple Displays," vol. 2626, no. c, 2015.

[28] A. Key, B. Howe, D. Perry, and C. Aragon, "VizDeck," *Proceedings of the 2012 international conference on Management of Data - SIGMOD '12*, p. 681, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2213836.2213931

[29] M. A. Yalçın, N. Elmqvist, and B. B. Bederson, "Keshif: Rapid and expressive tabular data exploration for novices," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 8, pp. 2339–2352, 2018.

[30] P.-M. Law, R. C. Basole, and Y. Wu, "Duet: Helping data analysis novices conduct pairwise comparisons by minimal specification," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 427–437, 2019.

[31] Google, "Explore in google sheets," 2015. [Online]. Available: https://www.youtube.com/watch?v=9TiXR5wwqPs

[32] F. Viegas, M. Wattenberg, D. Smilkov, J. Wexler, and D. Gundrum, "Generating charts from data in a data table," *US 20180088753 A1*, 2018.

[33] H. Lin, D. Moritz, and J. Heer, "Dziban : Balancing Agency & Automation in Visualization Design via Anchored Recommendations," *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems - CHI '20*, 2020.

[34] M. A. Hearst, "Design recommendations for hierarchical faceted search interfaces," in *Proc. SIGIR 2006, Workshop on Faceted Search*, August 2006, pp. 26–30.

[35] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst, "Faceted metadata for image search and browsing," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '03. New York, NY, USA: ACM, 2003, pp. 401–408. [Online]. Available: http://doi.acm.org/10.1145/642611.642681

[36] N. Cao, J. Sun, Y. Lin, D. Gotz, S. Liu, and H. Qu, "Facetatlas: Multifaceted visualization for rich text corpora," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1172–1181, Nov 2010.

[37] C. Collins, F. B. Viégas, and M. Wattenberg, "Parallel tag clouds to explore and analyze faceted text corpora," *VAST 09 - IEEE Symposium on Visual Analytics Science and Technology, Proceedings*, pp. 91–98, 2009.

[38] M. Dork, N. Henry Riche, G. Ramos, and S. Dumais, "Pivotpaths: Strolling through faceted information spaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2709–2718, Dec 2012.

[39] J. English, M. Hearst, R. Sinha, K. Swearingen, and K.-p. Yee, "Flexible Search and Navigation using Faceted Metadata," 2002.

[40] M. Hearst, A. Elliott, J. English, R. Sinha, K. Swearingen, and K.-P. Yee, "Finding the flow in web site search," *Commun. ACM*, vol. 45, no. 9, pp. 42–49, Sep. 2002. [Online]. Available: http://doi.acm.org/10.1145/567498.567525

[41] "eBay," https://www.ebay.com/, 2019, accessed: 2019-09-11.

[42] "Amazon," https://www.amazon.com, 2019, accessed: 2019-09-11.

[43] "Netflix," https://www.netflix.com, 2019, accessed: 2019-09-11.

[44] A. Sarvghad, M. Tory, and N. Mahyar, "Visualizing Dimension Coverage to Support Exploratory Analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 21–30, 2017.

[45] A. Srinivasan, S. M. Drucker, A. Endert, and J. Stasko, "Augmenting Visualizations with Interactive Data Facts to Facilitate Interpretation and Communication," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 672 – 681, 2019.

[46] T. N. Dang and L. Wilkinson, "Scagexplorer: Exploring scatterplots by their scagnostics," in *Proceedings of the 2014 IEEE Pacific Visualization Symposium*, ser. PACIFICVIS '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 73–80. [Online]. Available: http://dx.doi.org/10.1109/PacificVis.2014.42

[47] P.-M. Law, A. Endert, and J. Stasko, "Characterizing Automated Data Insights," *IEEE VIS*, p. arXiv:2008.13060, Aug. 2020.

[48] S. Vargas and P. Castells, "Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems," *RecSys*, 2011.

[49] M. Kaminskas and D. Bridge, "Diversity, Serendipity, Novelty, and Coverage : A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems," *ACM Transactions of Intelligent System*, vol. 7, no. 1, pp. 1–42, 2016.

[50] C. L. A. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, H. I. Storage, and R. Information, "Novelty and Diversity in Information Retrieval Evaluation," *SIGIR*, pp. 659–666, 2008.

[51] T. Munzner, "Chapter 3. Why: Task Abstraction," in *Visualization Analysis and Design*, 2014, pp. 42–65. [Online]. Available: http://dx.doi.org/10.1201/b17511-4

[52] M. Brehmer and T. Munzner, "A Multi-Level Typology of Abstract Visualization Tasks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2376–2385, 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6634168/

[53] R. Amar, J. Eagan, and J. Stasko, "Low-level components of analytic activity in information visualization," *Proceedings - IEEE Symposium on Information Visualization*, pp. 111–117, 2005.

[54] A. Strauss and J. Corbin, *Grounded theory methodology: An overview.*, ser. Handbook of qualitative research. Thousand Oaks, CA, US: Sage Publications, Inc, 1994, pp. 273–285.

[55] J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, D. Reichart, M. Venkatrao, F. Pellow, and H. Pirahesh, "Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals," *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 29–53, Mar 1997. [Online]. Available: doi.org/10.1023/A:1009726021843

[56] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[57] S. Kandel, R. Parikh, A. Paepcke, J. M. Hellerstein, and J. Heer, "Profiler: Integrated statistical analysis and visualization for data quality assessment," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2012, pp. 547–554.

[58] M. Sedlmair, A. Tatu, T. Munzner, and M. Tory, "A taxonomy of visual cluster separation factors," *Computer Graphics Forum*, vol. 31, no. 3pt4, pp. 1335–1344, 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2012.03125.x

[59] E. Horvitz, "Principles of mixed-initiative user interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '99. New York, NY, USA: ACM, 1999, pp. 159–166. [Online]. Available: http://doi.acm.org/10.1145/302979.303030

[60] J. Nielsen and R. Molich, "Heuristic Evaluation of user interfaces," *CHI '90 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, no. April, pp. 249–256, 1990.

[61] "Us department of education: College scorecard data," https://collegescorecard.ed.gov/data/documentation/, 2019, accessed: 2019-09-16.

[62] "120 years of olympic history: athletes and results," https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results, 2019, accessed: 2019-09-16.

[63] "Vega-dataset: Cars," https://vega.github.io/vega-datasets/data/cars.json, 2019, accessed: 2019-09-16.

[64] Z. Liu and J. Heer, "The effects of interactive latency on exploratory visual analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2122–2131, 2014.

[65] N. Mahyar and M. Tory, "Supporting communication and coordination in collaborative sensemaking," *IEEE transactions on visualization and computer graphics*, vol. 20, no. 12, pp. 1633–1642, 2014.

[66] D. J.-L. Lee, J. Kim, R. Wang, and A. Parameswaran, "Scattersearch: Visual querying of scatterplot visualizations," in *IEEE Symposium on Information Visualization, 2019 (Poster)*, ser. InfoVis '19, 2019. [Online]. Available: https://arxiv.org/abs/1907.11743

[67] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Proceedings of the 1996 IEEE Symposium on Visual Languages*, ser. VL '96. Washington, DC, USA: IEEE Computer Society, 1996, pp. 336–. [Online]. Available: http://dl.acm.org/citation.cfm?id=832277.834354

[68] G. Adomavicius and Y. Kwon, "Overcoming accuracy-diversity tradeoff in recommender systems: A variance-based approach," 1 2008, pp. 151–156, 2008 Workshop on Information Technologies and Systems, WITS 2008 ; Conference date: 13-12-2008 Through 14-12-2008.

[69] E. Fast, B. Chen, J. Mendelsohn, J. Bassen, and M. Bernstein, "Iris: A Conversational Agent for Complex Tasks," *CHI 2018*, 2018. [Online]. Available: http://arxiv.org/abs/1707.05015

[70] M. Correll, "Ethical dimensions of visualization research," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–13. [Online]. Available: https://doi.org/10.1145/3290605.3300418

[71] T. Zhou, Z. Kuscsik, J.-g. Liu, J. Rushton, and Y.-c. Zhang, "Solving the apparent diversity-accuracy dilemma of recommender systems," *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4511–4515, 2010.

[72] R. W. White and S. M. Drucker, "Investigating behavioral variability in web search," in *Proceedings of the 16th International Conference on World Wide Web*, ser. WWW '07. New York, NY, USA: ACM, 2007, pp. 21–30. [Online]. Available: http://doi.acm.org/10.1145/1242572.1242576

[73] M. Vartak, S. Huang, T. Siddiqui, S. Madden, and A. Parameswaran, "Towards Visualization Recommendation Systems," *ACM SIGMOD Record*, vol. 45, no. 4, pp. 34–39, 2017. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3092931.3092937