Eurographics Conference on Visualization (EuroVis) 2018
J. Heer, H. Leitte, and T. Ropinski
(Guest Editors)

Volume 37 (2018), Number 3

Design Factors for Summary Visualization in Visual Analytics

A. Sarikaya[1], M. Gleicher[2], and D. A. Szafir[3]

[1] Microsoft Corporation
[2] University of Wisconsin-Madison
[3] University of Colorado Boulder

Abstract

Data summarization allows analysts to explore datasets that may be too complex or too large to visualize in detail. Designers face a number of design and implementation choices when using summarization in visual analytics systems. While these choices influence the utility of the resulting system, there are no clear guidelines for the use of these summarization techniques. In this paper, we codify summarization use in existing systems to identify key factors in the design of summary visualizations. We use quantitative content analysis to systematically survey examples of visual analytics systems and enumerate the use of these design factors in data summarization. Through this analysis, we expose the relationship between design considerations, strategies for data summarization in visualization systems, and how different summarization methods influence the analyses supported by systems. We use these results to synthesize common patterns in real-world use of summary visualizations and highlight open challenges and opportunities that these patterns offer for designing effective systems. This work provides a more principled understanding of design practices for summary visualization and offers insight into underutilized approaches.

Categories and Subject Descriptors (according to ACM CCS):

CCS Concepts

•Human-centered computing → Visualization theory, concepts and paradigms;

1. Introduction

Visual analytics systems help users navigate large and complex datasets. These datasets often have too much data or too many dimensions to display in one view, requiring designers to engineer systems that summarize available data. These *summary visualizations* use visual and statistical techniques to purposefully reduce (i.e., summarize) the amount of data shown to viewers such that systems can manage the scale and complexity of large datasets. Examples of summary visualizations include aggregating data across selected dimensions into histograms, projecting high-dimensional data into a two-dimensional scatterplot, and summarizing actor-network relationships between entities captured from a text corpora. Designers draw from a large body of techniques when constructing summaries. For example, they may heuristically filter data, compute statistical quantities, or visualize bounds of data series rather than individual points.

The choices made in constructing summary visualizations determine the kinds of analyses supported by the end system and may guide viewers towards specific characteristics or subsets of data. As a result, the analyses, tasks, and data that a system needs to support are all factors that must be considered when designing a summary visualization. However, there is little systematic guidance for reasoning about these factors. In this paper, we aid designers by identifying key factors that must be considered in the design of summary visualizations. We identify trends in how designers use these factors to provide a basis for a more structured consideration of the design of summary visualizations. These trends can advise designers in considering different aspects of design, indicate what kinds of design patterns are most common, and help transfer approaches across domains to inspire new designs.

We elicit these trends using a structured literature survey of summary visualization in visual analytics tools and quantify the use of different design factors to answer key questions about the design and use of summaries. Across these visualizations, we find that designers regularly considered four key factors: the narrative function of the visualization (its *purpose*), the *data types* used, the method for reducing that data (*data summarization*), and the operations supported by the visualization (its *tasks*). To simplify the broad range of options available for data summarization, we introduce a taxonomy of summarization approaches that captures the breadth of methods used in summary visualizations while allowing designers to weigh trade-offs between methods. This taxonomy consists of four categories: aggregation, subsampling, filtering, and projection. Our survey suggests that this taxonomy sufficiently captures common strategies for data summarization and helps to identify trade-offs between summary approaches.

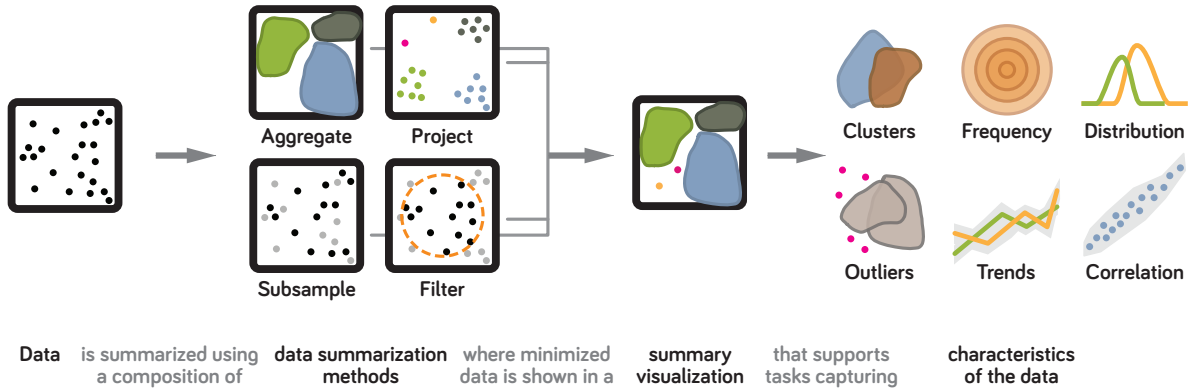

Figure 1: *A schematic of a generalized process for visual analytics with data summarization. A dataset (left) is reduced using data summarization techniques (center), comprised of four basic methods (aggregate, project, subsample, filter), and is presented visually to support judgments of high-level data characteristics (right). Both the summarization and visual presentation are factors that influence the efficacy of summary visualization to enable viewers to make high-level judgments.*

We use quantitative content analysis (QCA [RLF98]) to systematically characterize how the four design factors (purpose, data type, summarization method, and tasks) manifest in summary visualizations. Based on this characterization, we find that choices in these factors influence the types of analysis tasks supported by the resulting visualization. We identify common themes in how designers employ these factors in practice and how each factor contributes to the end utility of the system (Figure 1). These themes allow us to synthesize common practices in existing summary visualizations as well as identify underexplored design compositions that offer new possibilities for summary visualization.

Contributions: We provide a categorization of the factors in summary visualization design constructed through a systematic survey of the visualization literature conducted using QCA. Through this analysis, we

- provide a taxonomy of data summarization techniques used in summary visualization (§2–3);
- use this categorization and prior categorizations of purpose, data, and tasks to survey and analyze summary visualization design practices (§4);
- identify design patterns and trade-offs in summary design (Table 2), as well as potential opportunities for innovation (§4–5) grounded in existing practice.

This work provides a foundation for systematically reasoning about summarization in visual analytics and identifies gaps in our understanding of summarization in visualization design.

2. Background

In this work, we define a *summary visualization* as the result of an explicit set of design decisions that compress and/or simplify data for display, which includes choices of data reduction methods and visual representations. These visualizations communicate properties of a dataset using fewer marks than there are data entities, conveying the "gist" of critical high-level properties determined by the viewers' needs, data types, and necessary tasks. Summary visualizations provide analysts with a concise and focused representation they can use to navigate, sift, and winnow data [Shn96]. As an example, a scatterplot with points aggregated using KDE constitutes a summary visualization—it transforms individual points to spatial densities. In contrast, "zoomed-out" representations, such as a standard parallel coordinates plot with thousands of elements but no explicit data minimization, do not meet the criteria for a summary visualization: while individual relationships may be difficult to distinguish due to factors such as overdraw (see Fekete & Plaisant [FP02] and Cui, *et al.* [CWRY06]), the visualization does not intentionally summarize the data.

We draw on prior visualization taxonomies and design spaces as well as our own observations to identify four key factors in summary visualization design: *data summarization* method, *purpose*, *task*, and *data type*. We use these factors to understand how they can collectively guide the design and evaluation of summary visualizations. We look to related work in order to characterize these factors and use these characterizations to synthesize a codebook for a structured exploration of the visualization literature.

2.1. Factors of Summarization in Visualization

Data Summarization Methods: Data summarization reduces the scale and complexity of data for display in a summary. We specifically consider methods that summarize data in ways that provide a faithful representation of the underlying dataset. Prior work suggests general methods of re-organizing data for visualizations relevant to summary visualization. For example, Card & Mackinlay [CM97] offer a set of functions to process data for visualization: filtering, sorting, multidimensional scaling, and selection by slider. Ellis & Dix [ED07] taxonomize clutter reduction techniques for visualizations, including three techniques (sampling, filtering, and clustering) that explicitly reduce data. While clutter reduction is one goal of summarization, summary visualization must achieve a broader set of goals, including managing data scale, elic-

iting specific data characteristics, and guiding analysis. Elmqvist & Fekete [EF10] also survey aggregation techniques, focusing on hierarchical organization. In this work, we build on initial insights into aggregation introduced by Elmqvist & Fekete, and extend this analysis to a broader set of summarization methods.

We propose a taxonomy of data summarization methods that includes four categories: *aggregation*, *subsampling*, *filtering*, and *projection* (see §3.2 and §4.1 for details). We anticipate that the category of method used in a summary visualization influences the types of judgments that viewers can make from visualized data (e.g., Bertini et al.'s discussion of subsampling [BS06]). As noted in Ceneda *et al.* [CGM*17], the design choices made in a visual analytics system can guide the exploratory analysis process. By enumerating these four functional categories, we enable designers to consider the ramifications of each design choice focusing on how summaries might guide analysts towards certain tasks in order to inform the effective summarization targeting particular analyses.

Purpose: The *purpose* of a visualization describes its intended use. Bertin [Ber10] presents purpose as a dichotomy: the visualization either communicates previously understood information (presentation-oriented) or supports information processing to address new questions (exploratory). Schulz et al. [SNHS13] refine this division to consider the goals of an analysis: *exploratory* (undirected search), *confirmatory* (directed search), and *presentation* (communicating known results). We hypothesize that purpose of a summary visualization guides its design as purpose informs how viewers may wish to navigate the data. Specifically, we anticipate that summaries for presentation emphasize specific data characteristics more often than exploratory summaries. This is concordant with recent design guidelines proposed for presentation-oriented visualizations, which advise specificity and compactness over generalizability [Kos16].

Tasks: The summarization methods used to reduce a dataset influence the analysis tasks supported by a summary. For example, using kernel density estimation to spatially aggregate values in a scatterplot helps viewers find dense clusters, but obscures outliers. Understanding how the design of a summary can target different sets of tasks (e.g., presenting a few specific statistics versus permitting broad, flexible inference) allows designers to systematically reason about how well a summary visualization supports anticipated analysis goals (e.g., [AES05, BM13, JYSJ07]).

Task taxonomies provide perspectives on how viewers obtain information from visualizations (see Andrienko & Andrienko [AA06] and Shneiderman [Shn96] for canonical examples). Zhou & Feiner [ZF98] explore tasks related to high-level presentation intents and visual discourse, including several tasks relevant to summarization, such as *associate*, *compare*, *distinguish*, and *rank*. More recent work considers how tasks can drive visualization design (see Rind et al. [RAW*16] for a synthesis of this space). For example, Brehmer & Munzner [BM13] discuss how tasks can be abstracted and expressed to support design across different application domains. Schulz et al. [SNHS13] characterize tasks using "5 W's" (and one "H"): why is a task pursued (a task's *goal*), how is a task carried out (a task's *means*), what does a task seek (the *target* and *cardinality* of objects), when is a task performed, and who carries out the task? Schulz et al.'s hierarchical organization of tasks provides a comprehensive organization that we utilize in designing our codes for this work. Their questions allow us to systematically identify the role of different tasks in summary visualizations.

Data: The *data type* analyzed through a summary visualization may affect the summarization techniques summaries use and the features that analysts want to explore. As an example, hierarchical roll-up can work for high-dimensional data, but is not directly applicable to three-dimensional spatial data [EF10]. Exploratory database visualizations first summarize datapoints and their attributes using overviews of large amounts of high-dimensional data [KK96, Kei02]. Kehrer & Hauser [KH13] survey high-level design attributes of visual analytics overview approaches for multifaceted scientific data. They identify many techniques for summarizing particular data types, including spatial and high-dimensional data, but do not directly draw conclusions about the affordances of these techniques and the cross-applicability of summary designs for different data domains. Leung & Apperley [LA93] provide a framework for evaluating visualizations where there is too much data to display each datapoint clearly. This framework helps designers evaluate visual and computational representations of summaries based on their *effectiveness*, *expressiveness*, and *efficiency*; however, it provides no guidance for designing visualizations at these scales. While several surveys explore visualization for specific data types (e.g., Aigner et al. [AMM*07]), we instead look at higher-level relationships between data type and summary visualization design. This focus allows us to synthesize broad patterns across the design space and characterize common practices for summarizing data in different domains.

3. Methodology

Our goal is to understand how designers use the four target design factors to inform effective summary visualization and how these factors interact in different designs. We achieve this goal by conducting a structured survey of a range of examples in the visualization literature that enumerates how these factors lead to specific designs. We focus our analysis on four central research questions, each exploring the role of one factor in summary designs:

- **Q1** Does our taxonomy of *data summarization* cover the range of methods used in summary visualization design?
- **Q2** How does the *purpose* of a summary affect the design of the summary visualization?
- **Q3** How does the design of a summary visualization affect the *tasks* that it supports?
- **Q4** How does the type of *data* inform common design choices for summary visualization?

We also explore common correlations between factors to determine how designers combine these factors in visual analytics systems.

We use quantitative content analysis (QCA) [RLF98] to gather the necessary data to address these research questions. This methodology allows us to describe visualizations according to digestible, quantitative factors and uses statistical methods to identify trends between factors. We choose QCA over other methods such as grounded theory, which generates concepts from qualitative exploration, as such methods would likely be heavily biased by the sample of chosen summary visualizations. Instead, QCA depends

on a static codebook to quantify attributes, allowing us to draw on characterizations of our design factors identified in prior work. We derive our codebook from existing visualization taxonomies to mitigate bias from our chosen corpus and our own observations (see §3.2). We use the results from QCA to validate the organization of summarization methods and answer our research questions.

Two visualization researchers served as the coders for this survey. After a preliminary coding of ten papers, the two coders iterated on codebook definitions to clarify lingering ambiguities and to address emerging concerns regarding measure validity. Of 180 evaluated manuscripts, 54 randomly-selected papers (30%) were redundantly coded for validation—the Cohen's kappa measurement for intercoder reliability found substantial agreement between coders ($\kappa = 0.71$, 86% overall agreement). Section 4 presents the result of this process, identifying themes from our analysis.

3.1. Corpus Construction

We constructed a corpus of example visualizations from the data visualization research literature. These systems represent a collection of peer-reviewed visual analytics systems that discuss important components of design and intended use, minimizing the amount of inference required to apply our codes. We composed our corpus by collecting papers from the EuroVis, InfoVis, SciVis/Vis, and VAST conferences from 2009 to 2015 (1,158 papers). As coding every paper is intractable, we randomly sampled this larger corpus to create a representative sample as commonly done in traditional content analysis (e.g., [BKS*12]). This process generated a corpus of 180 papers (48 EuroVis, 53 InfoVis, 48 SciVis, and 31 VAST papers). Each paper was initially coded for whether or not they included an system or technique that used summary visualizations. Papers containing summaries were then coded according to protocol outlined below. We excluded theory, survey, toolkit, and evaluation papers as their focus was not a visualization design, making coding subjective as we had no explicit evidence of the designers' intents.

Using examples from the visualization research community allows us to focus on designs whose quality, effectiveness, and utility have been reviewed by external experts in the field and that are tailored for a wide variety of applications. Although visualization designs are also found in conferences outside of the immediate visualization community (e.g., NIPS, VLDB, KDD), specific visualization contributions in these fields are relatively rare and unlikely to appear in a random sample. Further, visualization research papers emphasize novel contributions and techniques that represent the state-of-the-art in visualization specifically, and these papers represent a vetted corpus of summary visualizations that contain explicit rationales for their design discussed within the article, increasing the validity of our coding practices. However, the choice of this corpus biases the results of this study toward exploratory visualizations that are used by researchers or domain experts (not the general public), which we discuss in Section 5.

3.2. Coding Protocol

Each example in our 180 paper sample was labeled using a predetermined codebook characterizing four factors of summary visualization design: the *data summarization* methods employed, the visualization's *purpose*, the *tasks* supported by the resulting summary visualization, and the type of *data* visualized. We constructed our codebook by collecting and abstracting categories across 15 existing typologies describing different aspects of these factors. Table 1 summarizes the coding scheme used in the survey. The final codes are as follows:

Category	Subcategory	Code
Data Summarization		Aggregation Subsampling Filtering Projection
Purpose		Exploratory Confirmatory Presentation
Task	Means: Navigation	Browsing Searching Elaborating Summarizing
	Means: Relation	Comparison Variations Relation-seeking
	Characteristics: High-level	Trends Outliers Clusters Frequency Distribution Correlation
Data		Data type Specific data
Other		Misc. observations

Table 1: *Two coders labeled 180 examples from the visualization literature according to 22 attributes describing the summary's purpose, data summarization, supported tasks, and data (§3.2).*

Data Summarization: We use our taxonomy of summarization methods (aggregation, filtering, subsampling, and projection) to characterize data summarization in visualization systems. Our four categories of methods are informed by our observations, coupled with categories from Schulz et al's *reorganization* task [SNHS13], the visualization design space [CM97], methods for clutter reduction [ED07], and methods of hierarchical abstraction [EF10].

Aggregation Computationally combining multiple elements (e.g., hierarchical aggregation [EF10]),
Subsampling Subsetting elements based on stochastic data selection (e.g., random subsampling [BS06]),
Filtering Subsetting elements based on properties of the data (e.g., selecting a representative set [ED07]), and
Projection Mapping data elements to a set of reduced or derived dimensions (e.g., principal component analysis [Jol02]).

We hypothesize that the data summarization methods used to construct a summary visualization heavily affect the analyses the summary supports (**Q3**). We coded high-level methods of data summarization using a combination of four binary codes (present or absent for each summarization category).

Purpose: We capture the purpose, or *goal*, of each visualization by considering whether it supports exploratory (undirected search), confirmatory (directed search) or presentation-oriented (exhibiting known results) analyses [Ber10, SNHS13]. These codes describe the high-level intent of the summarization and are treated as three binary (present/absent) codes.

Task: Our task codes were drawn from the *means* and *characteristics* of Schulz et al.'s taxonomy [SNHS13]. We chose this taxonomy as a general guide over other taxonomies as it comprehensively reflected most categories presented in other taxonomies. We use this taxonomy to code for three specific types of tasks: means of navigation, means of relation, and data characteristics.

Means of navigation describe how summary visualizations support analysis beyond the initial presentation. These tasks coincide with Springmeyer et al.'s concepts of maneuvering [SBM92], Casner's perceptual search operators [Cas91], Amar & Stasko's [AES05] and Yi et al.'s [JYSJ07] intent in interaction, Zhou & Feiner's [ZF98] modes of "enabling", and Heer & Shneiderman's "interactive dynamics" [HS12].

Means of object-object relations describe information foraging tasks, including *comparison* (seeking similarities; see [GAW*11, JYSJ07]), detecting *variation* (seeking dissimilarities; see [RM90, ZF98]), identifying *discrepancies* (seeking outliers [RM90,ZF98]), and *relation-seeking* (seeking one of the aforementioned relations for individual objects; see [Cas91,HS12]). While we initially coded for *discrepancy*, this code was removed from our analysis due to poor agreement between coders.

High-level characteristics code specific judgments of high-level data attributes afforded by summary visualizations. While Schultz et al.'s taxonomy does not explicitly define these characteristics, we used the following definitions that were agreed upon by the coders after iteration:

Trends	Estimate high-level changes across a dependent dimension,
Outliers	Identify items that do not match the modal distribution,
Clusters	Identify groups of similar items,
Frequency	Determine how often items appear,
Distribution	Characterize the extent and frequency of items, and
Correlation	Identify patterns between data dimensions.

These analysis tasks provide a representative proxy for understanding the informational utility of a summary visualization. Each of these three categories (summarized in Table 1) is measured as a combination of binary codes (task supported/unsupported).

Data: We coded for data type using Shneiderman's data type taxonomy [Shn96], with one-dimensional and temporal data collectively coded as *sequence data*, encompassing one-dimensional data on a common axis (e.g., temporal, genomic, or ranked data). While we considered data size as a potential code due to the utility of summarization for complex datasets, most systems did not provide bounds on the number of datapoints supported, and tended to design their methods for use with more than one dataset. For these reasons, coding for data size would have required significant extrapolation on the part of the coders, limiting the validity of the resulting data. We therefore did not consider data size in our analysis, but it is an important consideration for future work.

Other: We recognize that a codebook constructed *a priori* may not account for all elements of designs and tasks of summarization. To capture traits of summary visualizations not captured by this initial set of codes, we allowed coders to note additional observations about each summary for further exploration.

4. Survey Results

We used these codes to quantify factors leading to different design decisions for summary visualizations. In this section, we use our research questions (§3, **Q1–4**) to organize our findings from the coding process and generate 16 themes (**T1–T16**) characterizing common patterns in summary visualization design. These themes highlight core practices and opportunities in the design of summary visualizations.

We randomly selected 180 papers from the visualization research literature, 104 (58%) of which contained summary visualizations. Of these papers, 64 (36% of the original corpus of 180) provided sufficient detail within the paper to concretely apply codes describing our four factors of interests. For simplicity, we refer to these 64 examples as *fully-coded summaries*. The remaining 40 papers containing summaries primarily describe scientific visualization systems focused on rendering and provide little to no description of the target purpose or analytic tasks supported. To avoid over-extrapolation, we only coded those systems for data summarization methods. The full analysis results are available online at `http://graphics.cs.wisc.edu/Vis/vis_summaries/`.

4.1. Q1: Methods of Data Summarization

Our first question (**Q1**) asks whether our taxonomy of four categories of data summarization methods sufficiently covers the range of methods used in summary design. We found that all 104 summaries used at least one of these methods (Figure 2), and none used techniques that could not be expressed as a combination of the four categories. **Most summaries used more than one data summarization method (T1)** (63 summaries of 104, 61%). Overall, we found a strong correlation between summarization methods and tasks: each summarization method tended to favor a particular set of tasks, and designs often combined methods in order to leverage the strengths of individual techniques and increase the breadth of tasks supported by the summary.

Aggregation — Aggregation summarizes data by combining related values into representative statistical or graphical structures. **Most surveyed visualizations (74%) use aggregation to reduce data (T2)**, with 27% exclusively using aggregation. **Visualizations frequently used aggregation to support tasks characterizing the**

Axis — contributing factors	**Theme** — trends identified from survey data	
Data Summarization	T1	Most summaries use more than one data summarization method.
Data Summarization	T2	Most summaries use aggregation to summarize data.
Data Summarization × Task	T3	Summaries using aggregation support tasks characterizing the entire dataset.
Data Summarization	T4	Aggregation is common across all data types.
Data Summarization	T5	Filtering can be used across all data types.
Data Summarization × Task	T6	Projection and filtering emphasize similar data characteristics.
Data Summarization	T7	Summaries using subsampling are most common for scientific visualization.
Data Summarization × Task	T8	Subsampling supports tasks that are statistically robust to random sampling.
Purpose	T9	Summaries serve as a starting point for analysis.
Purpose × Task	T10	Exploratory summaries encode a broad set of data characteristics.
Purpose	T11	Confirmatory summaries support exploration.
Purpose × Task	T12	Presentation summaries emphasize a small set of specific characteristics.
Purpose × Data Summarization	T13	Designs for communicating specific, known information use aggregation.
Purpose × Task	T14	Summaries using subsampling emphasize exploration.
Task	T15	Summaries act as roadmaps to guide detailed exploration by interaction.
Task	T16	Summaries emphasize patterns that characterize all data and dimensions.

Table 2: *Our analysis revealed sixteen common design themes in examples of summary visualization. Taken collectively as observations, these themes highlight the challenges in the design of summaries. We use these challenges to reason about the trade-offs in existing designs and to identify underexplored areas of the design space to inform new summary designs.*

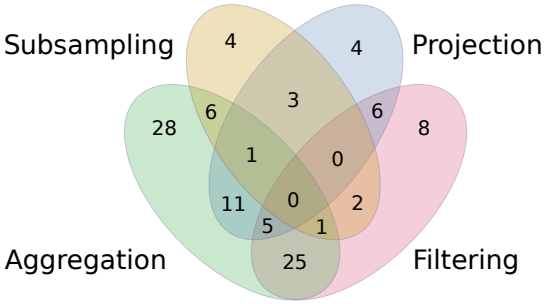

Figure 2: *The distribution of summary designs using each data summarization method across 104 coded visual summaries.*

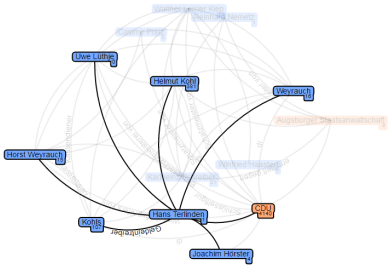

Figure 3: *Most network summaries, such as Networks of Names [KLB14], combine aggregation and filtering to summarize data. The system aggregates different relations across pairs of entities and filters these patterns according to their frequencies to encode the relationships that best characterize the dataset.*

entire dataset (T3): aggregation commonly enabled inference into both data distributions (42 of 54, 78%) and clusters (43 of 54, 80%).

Visualizations often used both aggregation and filtering to take advantage of trade-offs in both methods: while aggregation emphasizes characteristics describing multiple datapoints, filtering can help tailor these characteristics towards relevant collections. For example, Network of Names [KLB14] first aggregates recurring relations in social networks and then filters out uncommon relations to emphasize dominant patterns in large actor networks (Figure 3). We found that summaries without aggregation targeted individual value judgments, such as identifying outliers (supported by 70% of non-aggregate visualizations).

We found that **aggregation was commonly used across all data types (T4)**. The dominance of aggregation agnostic to data type indicates that it is a "default" used in many visualization systems. However, summaries using aggregation exchange flexibility for specificity: analysts can quickly and accurately complete tasks related to the chosen aggregates, but do so at the expense of other tasks. For example, replacing a set of values with their average removes analysts' abilities to reason about data variance or outliers. Critically examining this trade-off may offer new opportunities for visualizations (see Section 5).

Filtering — Filtering subsets data based on predefined properties, allowing analysts to remove potentially irrelevant data. Summaries can also use filtering to reintroduce important data values lost through aggregation, such as outliers in a scatterplot aggregated by density [MG13] (Figure 4). 47 visualizations (44%) used filtering; however, filtering was seldom used in isolation (only 17% of all filtering summaries, supporting **T1**). Filtering in visualizations allowed analysts to identify clusters (23 of 47 filtering visualizations, 82%), characterize distributions (22 of 47, 79%), and evaluate correlation (17 of 47, 61%).

Like aggregation, **filtering supported summary designs for all data types (T5)**. However, visualizations leveraging filtering provided analysts with little information about how filtering might bias

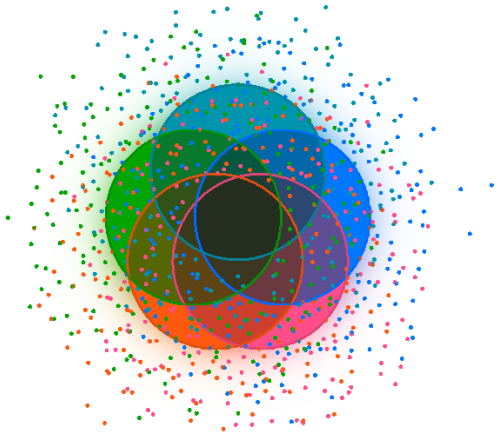

Figure 4: *Splatterplots [MG13] represent two-dimensional points by combining a kernel density estimation with filtering and subsampling of representative outlier points. Combining aggregation and filtering takes advantage of the trade-offs between these methods to support a broader variety of tasks.*

potential interpretations. This choice again exchanges flexibility for specificity: filtering allows analysts to closely analyze specific sets of values at the expense of the rest of the data.

Projection — Projection allows analysts to explore data in a simplified subspace. 30 examples (28%) used projection to summarize data. Similar to filtering, projection was seldom used in isolation (**T1**), and was commonly paired with either aggregation or filtering (24 summaries, 80%). Most examples used projection to summarize large collections of documents (7 of 30, 23%), 3D data (9 of 30, 30%) and multi-dimensional datasets (10 of 30, 33%). This bias in data type highlights projection's common use for high-dimensional data (**Q4**): projection methods can synthesize patterns across dimensions to support further summarization in low-dimensional spaces. For example, text visualizations can use topic modeling to project document vectors into a lower dimensional space and then aggregate documents according to topics (e.g., Cui, et al. [CLWW14]).

Projection summaries support characteristics similar to filtering (T6): locating clusters (17 of 19 summaries, 89%), characterizing distributions (16, 84%), and estimating correlation (14, 74%). However, projection frequently also enables outlier analysis (15, 79%). Visualizations can combine filtering and projection to help highlight critical patterns in complex data. For example, Progressive Insights [SPG14] projects data patterns onto statistical axes and filters the strongest patterns along each axis.

Projection was seldom used for presentation (2 of 20 presentation visualizations, 10%), but instead supported exploratory visualizations like Progressive Insights. We hypothesize that the mathematical complexity of many projection methods makes it difficult to clearly communicate meaningful narratives about the data. However, our corpus included few examples of presentation-oriented visualizations and a deeper exploration of the role of projection in presentation is important future work.

Subsampling — Subsampling reduces datasets by stochastically removing values from the dataset. Similar to projection, subsampling is commonly used as a composite operation to reduce data to manage the complexity of the resulting visualization: while only 16 of the 104 surveyed summary visualizations used subsampling, subsampling was commonly paired with other summarization methods (aggregation: 8 visualizations, 47% of subsampled examples; filtering: 3, 18%; and projection: 4, 24%).

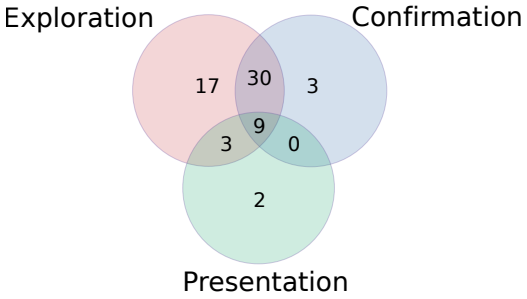

Figure 5: *The distribution of summaries designed for each purpose over 64 fully-coded summaries.*

Subsampling was predominantly used for spatial visualization (T7) (11 of 17 examples, 65% of subsampling use), where it reduced the visual complexity of aggregated structural data. In this context, subsampling is primarily used to assist rendering (minimizing noise), so only six subsampling visualizations were fully-coded. These visualizations primarily support trend analysis (5 of 6, 83%) and characterizing distributions (5, 83%), suggesting that **subsampling can support summarization where analysis tasks are statistically robust to random sampling (T8)**. This correlation implies subsampling may be a powerful tool for summaries for novel exploratory visualizations, especially when the target tasks or properties of interest are unknown *a priori*.

4.2. Q2: Purpose

Q2 addresses how the purpose of a summary visualization affects its design. Purpose codifies whether summaries are designed for exploration, confirmation, or presentation (Figure 5). Most fully-coded summaries supported exploration (59 of 64, 92%), allowing viewers to analyze large collections of data without any *a priori* goals. 66% (42) were designed for directed analysis (confirmation), while only 22% (14) were explicitly designed to communicate known results (presentation). The dominance of exploration indicates that **summaries frequently serve as a starting point for detailed analysis (T9)**. 95% (56 of 59) of exploratory summaries allowed analysts to actively navigate the dataset.

Additionally, **exploratory summaries support a broader set of data characterization tasks (T10)**, such as identifying trends, outliers, clusters, frequency, distribution, and correlation. 70% (41) of exploratory summaries enabled viewers to explore more than half of the coded characterization tasks, compared to 43% (6 of 14) for presentation summaries. 12% of exploratory summaries (7) supported all six. As an example, Chen, et al. [CYW*16] uses a set of summarization methods to visualize different patterns across geo-tagged social media data (Figure 6). Analysts can use the system to explore aggregated movement trends and interact with the

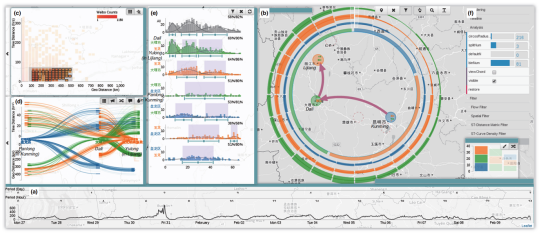

Figure 6: *A visual summary in the system built by Chen, et al. [CYW*16] uses both aggregation and filtering in order to support a wide range of high-level analysis tasks.*

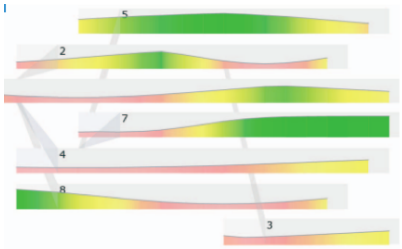

Figure 7: *World Lines [WFR*10] aggregates spatial data across different simulation runs to allow viewers to directly search for the simulation with the best outcome.*

summary to analyze data distributions, frequency, and geospatial clusters.

Confirmatory summaries (those used to validate prior hypotheses) were often also exploratory: 61% of summaries (39 of 64) supported both exploration and confirmation while none were designed for confirmation or presentation alone. Like exploratory designs, confirmatory designs support a broader array of data characterization tasks than presentation-oriented summarization: 68% supported more than half of our coded tasks. These correlations suggest that **summaries designed for confirmation also support exploration (T11)**: confirmatory tools generally allow analysts to not only confirm specific hypotheses about data, but also to further refine and develop additional hypotheses.

In contrast, **presentation summaries often emphasize a small set of data characteristics (T12)**. 57% (8 of 14) of presentation summaries communicated three or fewer coded characterization tasks, and only one design communicated all six (Domino [GGL*14], which also supports exploration). All coded presentation summaries used aggregation to summarize data. Of these, 50% (7) used aggregation alone and 35% (5) used aggregation plus filtering. This pattern suggests that **designs communicating specific, known information heavily rely on aggregation (T13)**. It is important to note that our choice to survey only examples from the visualization literature biases our analysis towards exploratory visualizations. However, we anticipate the heavy use of aggregation we observed in presentation summaries will extend to examples in other outlets, such as news organizations, as it aligns with modern guidelines for effective presentation. More specifically, aggregation can summarize data into a small number of precise features to emphasize known findings, encouraging effective presentation [Kos16]. We discuss trade-offs of this focus in Section 5.

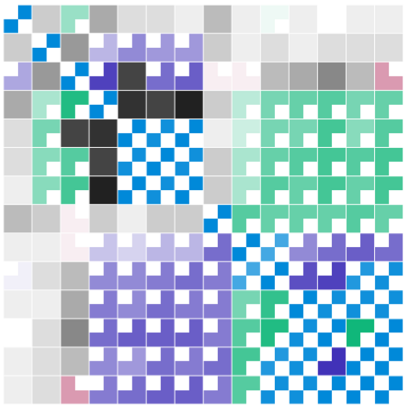

Figure 8: *Summaries act as roadmaps for exploration, starting at a high-level of abstraction and letting viewers drill down into data. Glyph SPLOMs [YWS*14] summarizes clustering patterns in component SPLOM scatterplots to help identify scatterplots to further explore.*

Only five coded summaries were not explicitly designed for exploration. All five were confirmatory visualizations using aggregation, and none used subsampling. This bias indicates a potential trade-off between purpose and subsampling. **Subsampling methods favor exploration (T14)** as directed search may be inhibited by stochastically reducing data. Aggregation alternatively helps guide analysts by presenting precise summarized values for well-defined tasks. For example, World Lines [WFR*10] uses aggregation to summarize parallel simulations of temporal events enabling comparison across known metrics for disaster planning (Figure 7).

4.3. Q3: Tasks

Q3 asks how the design of a summary visualization affects the tasks that it supports. While the previous subsections touch on interactions of methods and purpose with supported tasks, here we identify several trends to help inform how summarization affects possible avenues of analysis (Figure 9). From the 64 fully-coded visualizations, we found themes describing how designs allow viewers to navigate the dataset, how summarizing different data types prioritize different analyses, and data characteristics summaries universally preserve.

Means of Navigation — Through our survey, we found that most summaries present information at high levels of granularity and allow analysts to drill down into data to uncover specific details within the data. This indicates that summary visualizations, like other forms of overview, generally provide a starting point for analysis, allowing analysts to browse for both unknown (58 of 64, 91%) and expected patterns (48 of 64, 75%). The use of summaries as a starting point for navigation implies that **effective summaries can act as roadmaps to guide user interactions with the data (T15)**. As an example, glyph SPLOMs [YWS*14] summarize distributions within specific SPLOMs so that viewers can identify scatterplots to explore in detail (Figure 8). This raises an important

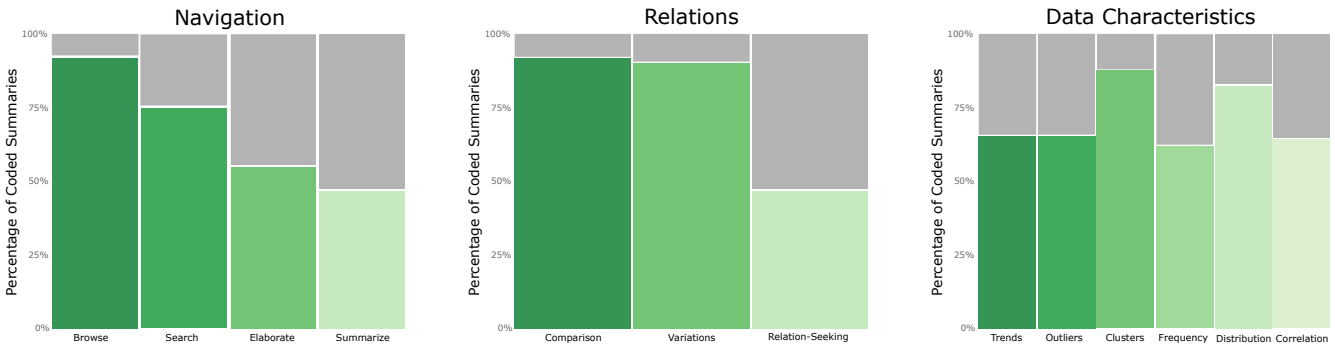

Figure 9: *The distribution of summary designs supporting different kinds of analysis tasks across 64 fully-coded summaries.*

challenge for visualization designers: what properties of the data might make for an effective starting point for analysis?

Existing summary visualizations often choose to first emphasize distributions (48 of 64, 75%) and clusters (51, 80%) within data. Analysts can then navigate these structures to identify specific properties and values of interest. Drilling down into data generally takes three forms: changing the data granularity (*elaborating* in our codebook; 35 of 64, 55%), changing the visual representation and/or summarization method (28 visualizations, 44%), or adding supplemental information to the existing display. Designers may choose from these strategies based on the parameters of the data and analysis tasks; however, summaries must actively balance the need to support different tasks and granularities with potential challenges introduced from inconsistent visual representations [QH17].

Means of Relation — Most summary visualizations enable viewers to identify similarities (89%) and differences (88%) between collections of datapoints. However, significantly fewer support relation-seeking between individual items (45%), with most of these being network visualizations, which prioritized important relationships over specific structures within the data. We found no notable relationships between relation seeking tasks and purpose or summarization methods. We anticipate this is because summary methods tend to support analyses of large collections of data. These analyses naturally privilege tasks emphasizing higher-level attributes of the data (e.g., data distributions and relationships between data classes) rather than individual data values, which quickly grows intractable as the datasets grow larger.

Data Characteristics — While the prior sections discuss interactions between data characteristics and other aspects of summary design, we found that summary visualizations generally emphasized data clusters (80%) and distributions (75%), two characteristics that describe entire sets of data points. Trends (59%), outliers (59%), frequency (56%), and correlation (58%) were roughly equally supported across all visualizations. The bias towards clusters and distributions suggests that **summarization often emphasizes descriptive aggregate patterns across all data values (T16)**, rather than patterns in individual values or relationships between specific dimensions. 11% of coded summary visualizations support all characterization tasks (7 of 64).

4.4. Q4: Data Type

Q4 asks how data type affects summary design. We found that the underlying data systematically influenced summary visualization design. For example, nine of the ten coded visualizations for one-dimensional data used aggregation and support cluster analysis. For 2D data, summary visualizations frequently support discovering trends (7 of 8, 88%) and frequency patterns (6, 75%). However, 3D data summaries emphasize data distributions (5 of 7, 71%) rather than trend or frequency (3 and 1 of 7, 43% and 14%, respectively).

Neither multidimensional nor network data used subsampling frequently (5 summaries of 24, 21%; 0 of 10, 0%). This methodological bias is likely a result of common analytic practices for these kinds of data. For example, stochastically removing information in networks could potentially remove critical structures in the data, such as relations between different levels of hierarchy. We observed that nearly all summarizations of network data utilized aggregation (through collapsing important collections of nodes or edges) and filtering (through selecting meaningful or common connections) to emphasize relation-seeking between salient entities. For example, Networks of Names [KLB14] highlights relationships between large collections of entities by first aggregating all entity relations and then filtering on these aggregate frequencies to visualize the most common relations in the dataset. While these patterns suggest that designers employ common strategies based on the target data type, we hypothesize that our framework will allow designers to consider novel summary approaches that transfer design elements across data types and domains by systematically considering trade-offs offered by different summarization approaches.

5. Discussion

We use QCA to answer four research questions pertaining to how designers consider a visualization's purpose, summarization methods, data type, and target analysis tasks in constructing summary visualizations. Our analysis resulted in 16 design themes (Table 2) commonly used for summary visualizations in visual analytics. We also confirm that our taxonomy of four methods of summarization is sufficient to describe the set of summarization techniques used in summary visualizations (**Q1**). We can use the themes identified through this survey to synthesize common design practices and identify opportunities for innovation in summary design.

Q2: Purpose of Summary Visualizations — Designers frequently use summary visualizations as a starting point for analysis (**T7**), serving as a roadmap to guide analysts' interactions with data (**T15**). The ways summaries support these goals tie tightly with their purpose: a summary can allow analysts to explore a broad set of data features (**T10**) or emphasize particular characteristics of interest (**T12**). We find that exploratory visualizations generally use the former strategy, while presentation visualizations use the latter.

While these patterns follow conventional visualization strategies, we found few examples that invert this paradigm. For example, supporting larger numbers of tasks can overwhelm the viewer. In some exploratory situations, the viewer may not know what questions to ask, or even how to "read" a visual paradigm. In these situations, designers might choose to target anticipated analyses with the goal of focusing exploration. Future research should explore strategies for achieving this guidance without inadvertently biasing analyst workflows. As an example, compositing multiple summarization methods can help to focus exploration on relevant subsets of data, but maintain the flexibility of focusing on disparate sets. Alternatively, presentation summaries that support a broader set of tasks may allow viewers to construct a deeper understanding of the narrative argument offered by that visualization. Presentation summaries that carefully balance clarity of target attributes with supplemental features could potentially lead to increased trust in the data [CRMH12].

Q3: Tasks Supported by Summary Designs — Our exploration highlighted heavy use of task *specificity* in existing designs. Summary visualizations often use aggregation to emphasize specific data characteristics, such as clusters (**T13**). Alternatively, less common subsampling strategies tend to provide greater flexibility in analysis, retaining global characteristics of the data (**T14**).

While existing designs tend to emphasize specificity, we argue that the trade-off of task specificity and flexibility offers an interesting consideration for designers. Specificity can bias analysts toward only considering a particular set of characteristics. Aggregation is an example of this—aggregation systematically combines multiple values into a single representation, which can influence data interpretation (noted by Saraiya *et al.* [SNLD06]). The specificity bias indicates the need to look toward more holistic views of creating summaries. By noting the full set of tasks, designers can construct a summary that provides sufficient specificity to support these tasks, while also allowing analysts to pivot between multiple views or representations to support serendipitous discovery [THC12]. As an example, summary visualizations can use visual encodings that allow analysts to visually estimate features of data distributions from individual datapoints rather than encoding these features directly. Such visual aggregation may enable a more holistic view on data [SHGF16], though we found no summaries explicitly leveraging this strategy.

Q4: Data Type Drives Summary Design — Summaries can represent a wide array of data types. Aggregation and filtering are especially flexible for constructing summaries regardless of data type (**T4**, **T5**). The broad use of these methods suggests that visualizations could adapt many summarization approaches across data domains. However, properties of the data and analysis context may preclude the use of some methods. For example, we found that projection methods were commonly used to summarize text and spatial data (1D and 3D data, respectively; **T7**). However, we did not see these methods used in sequence data despite conceptual similarities between these data types (e.g., texts are sequences of words). Considering how summarization methods and other summary techniques might transfer across data types may offer new approaches for summary visualization in different domains.

Aggregation methods in particular may offer novel summary approaches. For example, continuous 2D data can be meaningfully summarized using kernel-density estimation (KDE); however, a kernel does not easily map to hierarchical data. Alternatively, Elmqvist & Fekete [EF10] demonstrate how hierarchical aggregation can be applied to non-hierarchical data. Our survey identified several characteristics not conventionally supported by summaries for specific data types, such as frequency in three-dimensional data or subsampling in network or multi-dimensional data. The lack of common design strategies in these areas highlight potential scenarios that require innovative design approaches for intuitive summary visualization.

5.1. Limitations

This work provides a systematic analysis of common practices for summarization in visual analytics. Our exploratory survey characterizes the use of four design factors in summary visualization. However, our data-driven approach is limited by the need to use sampling to manage the scale of our corpus. Although we anticipate that the collected systems and themes characterize summarization more broadly, we cannot make absolute claims about the generalizability of our results.

Our analysis focuses on high-level design factors for summary visualization such as summarization methods rather than specific encoding choices. This high-level focus provides a general overview of current practices in summarization, but prevents us from generating more prescriptive design recommendations. For example, our analysis revealed that the use of specific encodings in summary visualizations is a secondary step to determining the primary summarization method. This means that we could not perform a conclusive analysis of encoding choices without first understanding summarization. Our results provide a necessary scaffold for future work in understanding how our four design factors may inform effective encoding design.

We sampled from the visualization literature in order to codify trends that represent the state-of-the-art in summary visualization. This choice also meant that summaries designed for presentation, common in journalism and professional dashboards, are underrepresented in our survey [Kos16]. Expanding the corpus studied here to both include examples from practitioner communities and to use stratified sampling to select underrepresented design features could help to develop a broader set of design patterns for both exploratory and presentation summary visualizations.

6. Conclusion

As datasets grow in size and complexity, effectively leveraging summarization becomes increasingly critical for visual analytics

systems. In this paper, we characterize common approaches to summary visualization design using four factors. Through a structured survey, we identified the importance of summary visualization and 16 design themes relating data summarization methods, visualization purpose, analysis tasks, and data types. The survey results highlight trade-offs in the use of different summarization methods and biases in their applications in existing designs. Our results can inform effective summary design and introduce a more principled understanding of summary visualization.

Acknowledgments

We acknowledge the fruitful discussions with Robert E. Roth and Michael Correll, as well as the feedback of the anonymous reviewers. This work was supported by NSF award IIS-1162037.

References

[AA06] ANDRIENKO N., ANDRIENKO G.: *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer, Heidelberg, 2006. 3

[AES05] AMAR R., EAGAN J., STASKO J.: Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* (2005), IEEE, pp. 111–117. doi:10.1109/INFVIS.2005.1532136. 3, 5

[AMM*07] AIGNER W., MIKSCH S., MÜLLER W., SCHUMANN H., TOMINSKI C.: Visualizing time-oriented data—a systematic view. *Computers & Graphics 31*, 3 (2007), 401–409. doi:10.1016/j.cag.2007.01.030. 3

[Ber10] BERTIN J.: *Seminology of Graphics: Diagrams, Networks, Maps*. ESRI Press, Redlands, CA, 2010. 3, 5

[BKS*12] BRUBAKER J. R., KIVRAN-SWAINE F., ET AL.: Grief-stricken in a crowd: The language of bereavement and distress in social media. In *Sixth International AAAI Conference on Weblogs and Social Media* (Dublin, Ireland, 2012), The AAAI Press, pp. 42–49. URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/viewFile/4622/4965. 4

[BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2376–2385. doi:10.1109/TVCG.2013.124. 3

[BS06] BERTINI E., SANTUCCI G.: Give chance a chance: Modeling density to enhance scatter plot quality through random data sampling. *Information Visualization 5*, 2 (2006), 95–110. doi:10.1057/palgrave.ivs.9500122. 3, 4

[Cas91] CASNER S. M.: Task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics 10*, 2 (1991), 111–151. doi:10.1145/108360.108361. 5

[CGM*17] CENEDA D., GSCHWANDTNER T., MAY T., MIKSCH S., SCHULZ H.-J., STREIT M., TOMINSKI C.: Characterizing Guidance in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 111–120. doi:10.1109/TVCG.2016.2598468. 3

[CLWW14] CUI W., LIU S., WU Z., WEI H.: How hierarchical topics evolve in large text corpora. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2281–2290. doi:10.1109/TVCG.2014.2346433. 7

[CM97] CARD S., MACKINLAY J.: The structure of the information visualization design space. In *Proceedings of the Information Visualization Conference* (1997), pp. 92–99. doi:10.1109/INFVIS.1997.636792. 2, 4

[CRMH12] CHUANG J., RAMAGE D., MANNING C., HEER J.: Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), ACM, pp. 443–452. doi:10.1145/2207676.2207738. 10

[CWRY06] CUI Q., WARD M. O., RUNDENSTEINER E. A., YANG J.: Measuring data abstraction quality in multiresolution visualizations. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 709–716. doi:10.1109/TVCG.2006.161. 2

[CYW*16] CHEN S., YUAN X., WANG Z., GUO C., LIANG J., WANG Z., ZHANG X. L., ZHANG J.: Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 270–279. doi:10.1109/TVCG.2015.2467619. 7, 8

[ED07] ELLIS G., DIX A.: A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1216–1223. doi:10.1109/TVCG.2007.70535. 2, 4

[EF10] ELMQVIST N., FEKETE J. D.: Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics 16*, 3 (2010), 439–454. doi:10.1109/TVCG.2009.84. 3, 4, 10

[FP02] FEKETE J.-D., PLAISANT C.: Interactive information visualization of a million items. In *IEEE Symposium on Information Visualization* (2002), vol. 2002, pp. 117–124. doi:10.1109/INFVIS.2002.1173156. 2

[GAW*11] GLEICHER M., ALBERS D., WALKER R., JUSUFI I., HANSEN C. D., ROBERTS J. C.: Visual comparison for information visualization. *Information Visualization 10*, 4 (2011), 289–309. doi:10.1177/1473871611416549. 5

[GGL*14] GRATZL S., GEHLENBORG N., LEX A., PFISTER H., STREIT M.: Domino: Extracting, comparing, and manipulating subsets across multiple tabular datasets. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2023–2032. doi:10.1109/TVCG.2014.2346260. 8

[HS12] HEER J., SHNEIDERMAN B.: Interactive dynamics for visual analysis. *Queue 10* (2012), 30. doi:10.1145/2133416.2146416. 5

[Jol02] JOLLIFFE I. T.: *Principal Component Analysis*, second ed. Springer-Verlag, 2002. doi:10.1002/9781118445112.stat06472. 4

[JYSJ07] JI SOO YI, YOUN AH KANG, STASKO J., JACKO J.: Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1224–1231. doi:10.1109/TVCG.2007.70515. 3, 5

[Kei02] KEIM D. A.: Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics 8*, 1 (2002), 1–8. doi:10.1109/2945.981847. 3

[KH13] KEHRER J., HAUSER H.: Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics 19*, 3 (2013), 495–513. doi:10.1109/TVCG.2012.110. 3

[KK96] KEIM D. A., KRIEGEL H.: Visualization techniques for mining large databases: A comparison. *IEEE Transactions on Knowledge and Data Engineering 8*, 6 (1996), 923–938. doi:10.1109/69.553159. 3

[KLB14] KOCHTCHI A., LANDESBERGER T. V., BIEMANN C.: Networks of Names: Visual exploration and semi-automatic tagging of social networks from newspaper articles. *Computer Graphics Forum 33*, 3 (2014), 211–220. doi:10.1111/cgf.12377. 6, 9

[Kos16] KOSARA R.: Presentation-oriented visualization techniques. *IEEE Computer Graphics and Applications 36*, 1 (2016), 80–85. doi:10.1109/MCG.2016.2. 3, 8, 10

[LA93] LEUNG Y. K., APPERLEY M. D.: E3: Towards the metrication of graphical presentation techniques for large data sets. In *Proceedings of EWHCI '93* (1993), Bass L. J., Gornostaev J., Unger C., (Eds.), Springer, pp. 125–140. doi:10.1007/3-540-57433-6_44. 3

[MG13] MAYORGA A., GLEICHER M.: Splatterplots: Overcoming overdraw in scatter plots. *IEEE Transactions on Visualization and Computer Graphics 19*, 9 (2013), 1526–1538. doi:10.1109/TVCG.2013.65. 6, 7

[QH17] QU Z., HULLMAN J.: Keeping multiple views consistent: Constraints, validations, and exceptions in visualization authoring. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (2017), 468–477. doi:10.1109/TVCG.2017.2744198. 9

[RAW*16] RIND A., AIGNER W., WAGNER M., MIKSCH S., LAMMARSCH T.: Task cube: A three-dimensional conceptual space of user tasks in visualization design and evaluation. *Information Visualization 15*, 4 (2016), 288–300. doi:10.1177/1473871615621602. 3

[RLF98] RIFFE D., LACY S., FICO F.: *Analyzing Media Messages: Using Quantitative Content Analysis in Research*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA, 1998. 2, 3

[RM90] ROTH S. F., MATTIS J.: Data characterization for intelligent graphics presentation. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (1990), 193–200. doi:10.1145/97243.97273. 5

[SBM92] SPRINGMEYER R. R., BLATTNER M. M., MAX N. L.: A characterization of the scientific data analysis process. In *Proceedings of the IEEE Conference on Visualization* (1992), pp. 235–242. 5

[SHGF16] SZAFIR D. A., HAROZ S., GLEICHER M., FRANCONERI S.: Four types of ensemble coding in data visualizations. *Journal of Vision 16*, 5 (2016). doi:10.1167/16.5.11. 10

[Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings IEEE Symposium on Visual Languages* (1996), IEEE Comput. Soc. Press, pp. 336–343. doi:10.1109/VL.1996.545307. 2, 3, 5

[SNHS13] SCHULZ H. J., NOCKE T., HEITZLER M., SCHUMANN H.: A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2366–2375. doi:10.1109/TVCG.2013.120. 3, 4, 5

[SNLD06] SARAIYA P., NORTH C., LAM V., DUCA K. A.: An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics 12*, 6 (2006), 1511–1522. doi:10.1109/TVCG.2006.85. 10

[SPG14] STOLPER C. D., PERER A., GOTZ D.: Progressive visual analytics: User-driven visual exploration of in-progress analytics. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1653–1662. doi:10.1109/TVCG.2014.2346574. 7

[THC12] THUDT A., HINRICHS U., CARPENDALE S.: The Bohemian Bookshelf: Supporting serendipitous book discoveries through information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), ACM, pp. 1461–1470. doi:10.1145/2207676.2208607. 10

[WFR*10] WASER J., FUCHS R., RIBICIC H., SCHINDLER B., BLOSCHL G., GROLLER E.: World lines. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1458–1467. doi:10.1109/TVCG.2010.223. 8

[YWS*14] YATES A., WEBB A., SHARPNACK M., CHAMBERLIN H., HUANG K., MACHIRAJU R.: Visualizing multidimensional data with Glyph SPLOMs. *Computer Graphics Forum 33*, 3 (2014), 301–310. doi:10.1111/cgf.12386. 8

[ZF98] ZHOU M., FEINER S.: Visual task characterization for automated visual discourse synthesis. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (1998), pp. 392–399. doi:10.1145/274644.274698. 3, 5