v-plots: Designing Hybrid Charts
for the Comparative Analysis of Data Distributions

Michael Blumenschein[iD], Luka J. Debbeler[iD], Nadine C. Lages[iD], Britta Renner[iD], Daniel A. Keim[iD], and Mennatallah El-Assady[iD]

University of Konstanz, Germany

(a) *v-plot*

(b) *Selection of common charts to visualize and compare data distributions.*
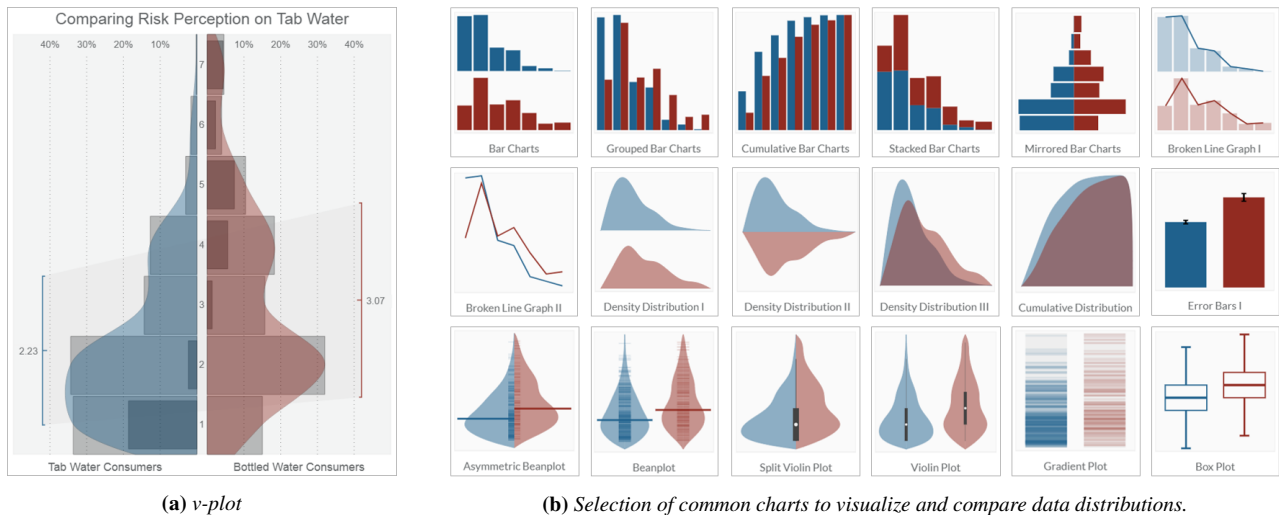
Figure 1: *The* v-plot designer *enables the creation of custom hybrid charts* (v-plots) *for the comparative analysis of data distributions. Given a set of analysis tasks, v-plots can be tailored to highlight particular distribution properties (on a* local, global, *and* aggregated level) *using a guiding wizard. All the charts above represent the same data, showing the perceived risk of two groups for tap water consumption [DGB*18].*

Abstract

Comparing data distributions is a core focus in descriptive statistics, and part of most data analysis processes across disciplines. In particular, comparing distributions entails numerous tasks, ranging from identifying global distribution properties, comparing aggregated statistics (e.g., mean values), to the local inspection of single cases. While various specialized visualizations have been proposed (e.g., box plots, histograms, or violin plots), they are not usually designed to support more than a few tasks, unless they are combined. In this paper, we present the v-plot designer; a technique for authoring custom hybrid charts, combining mirrored bar charts, difference encodings, and violin-style plots. v-plots are customizable and enable the simultaneous comparison of data distributions on global, local, and aggregation levels. Our system design is grounded in an expert survey that compares and evaluates 20 common visualization techniques to derive guidelines for the task-driven selection of appropriate visualizations. This knowledge externalization step allowed us to develop a guiding wizard that can tailor v-plots to individual tasks and particular distribution properties. Finally, we confirm the usefulness of our system design and the user-guiding process by measuring the fitness for purpose and applicability in a second study with four domain and statistic experts.

CCS Concepts

• Human-centered computing → Visualization design and evaluation methods;

1. Introduction

Analyzing and exploring empirical data and its distribution is a core task in descriptive statistics across various research disciplines, and often serves as a foundation for in-depth analyses. Comparing data distributions is a multi-faceted process comprising a variety of different analysis tasks ranging from *global* aspects, such as identifying the type, shape, and skewness of a distribution, to *local* aspects,

such as comparing value frequencies, or identifying differences on an instance level. Alongside global and local tasks, users also want to analyze *aggregated* statistical properties such as comparisons of mean and interquartile ranges. However, even with the abundance of statistical techniques, the visual inspection of distributions is essential to a successful analysis [Tuk77, FMF12] and can help to generate hypotheses, select appropriate statistical methods, and support the understanding and communication of analysis results.

Global, aggregated, and local analysis tasks focus on different properties of data distributions, which has led to the development of a broad range of *specialized* charts within the statistics and InfoVis communities. Some of the most prominent include box plots [Tuk77], violin plots [HN98], and bar charts. Each has its own strengths and weaknesses: for example, while box plots are useful for comparing medians and quartiles, they do not show whether a distribution is bi-modal, or in which value ranges two distributions differ most. Violin plots support these questions, but they are less useful for local tasks where users are interested in the frequencies of individual values (e.g., in discrete distributions). However, most analytical processes combine a number of different tasks. Analysts wishing to derive effective designs for a given dataset and a particular combination of tasks are faced with one of two choices: either using multiple charts to communicate their analysis on the different levels; or designing an expressive hybrid chart tailored to their analysis needs. Although hybrid designs have shown to be more effective for targeted analysis [BKC*13], most people do not have access to them and must use multiple charts instead. In addition, analysts have to rely on experience and knowledge about descriptive statistics when choosing appropriate charts, making this design space troublesome for non-experts to navigate.

To support the comparative analysis of data distributions for a wider audience, we need to make the design of hybrid charts more accessible and make task-dependent customization in the visualization of descriptive statistics less reliant to the user's expertise. Our chart authoring system, the *v-plot designer*, is developed to generate custom, widely applicable hybrid charts (*v-plots*) that concurrently support combinations of global, local, and aggregated tasks. v-plots consist of five (**"v"**) layers, which inspired its name. We strove to ground our system design in an expert survey that aimed to externalize the implicit knowledge and experience of domain experts into *guidelines* used to inform a guiding wizard component. In this paper, we reflect on the **multistage research process** of our system design that tackled the following research question: *How can we make hybrid charts for the visual comparison of data distributions (1)* **simultaneously support local, aggregation-based, and global analysis tasks**; *and (2)* **accessible to analysts***?*

We provide a **grounding for our visualization design** by reviewing and categorizing existing tasks and charts for the comparative analysis of data distributions (Section 3). This lays the foundation for our expert survey (Section 4), which *induces guidelines* and an *automatic chart recommendation* for the selection of existing visualization techniques for specific task combinations.

In addition to this design study, we also provide a rationale on our **system design and guidance**. This is based on the review of the related work (Section 2), and describes the reasoning behind v-plots as custom hybrid charts that are designed to support a variety of concurrent tasks (Section 5). Based on the externalized guidelines, a *guiding wizard* in the *v-plot designer* can automatically tailor the v-plot through different visualization layers and transparency levels to highlight particular distribution properties. The guidelines are interchangeable and can be extended by new findings or the requirements of specific communities in the future. Single v-plots can also be combined to a *v-plot matrix* that is sorted according to the similarity of their visual structures and support the comparison of all pairwise distributions.

To summarize, the contribution of this paper is two-fold. (1) We contribute **guidelines** based on a representative **expert survey** on the applicability and usefulness of the 20 most commonly-used statistical charts to 20 analysis tasks for descriptive statistics. For transparency and reproducibility, we make the survey and its results available at osf.io/jk8rp. (2) To make the acquired knowledge accessible while supporting the creation of custom hybrid charts, we contribute the *v-plot designer* (v-plot.dbvis.de). This **chart authoring approach** relies on a *guiding wizard* to enable users to adjust v-plots, which combine mirrored bar charts, direct difference encodings, a distribution shape, labels, and axes with statistic values.

2. Related Work

In addition to the landscape of comparative distribution analysis (described in Section 3), our work lies within the context of other chart authoring tools, as well as guiding and chart comparisons.

2.1. Chart Authoring Tools

Designing visualizations to communicate patterns is one of the key tasks in data science and descriptive statistics. However, implementing new visualizations for each analysis is not feasible for analysts. Therefore, in contrast to the vast amount of visualization coding libraries (e.g., Lyra [SH14] or Vega [SRHH16]), chart authoring systems have been developed to streamline design process and aid non-visualization-experts. Most of these systems are commercial applications [BSS*19]. Most prominently, Tableau [Tab20] has positioned itself as an easy-to-use, toolkit-based chart authoring system for the masses. More recently, Charticulator [RLB19] has been presented as a chart authoring tool for variable, user-defined layouts, enabling more flexibility with chart designs. RAW-Graphs [MEC*17] is an example of a authoring system providing users with an open-access API. Lastly, approaches such as Voyager [WMA*16, WQM*17] combine the flexibility of declarative user-defined visualization designs with the ease of template-based authoring systems. While such tools target the broad spectrum of data-driven visualization designs, none are specifically focused on authoring visualizations for the comparative analysis of data distributions for descriptive statistics. Additionally, these approaches do not encode knowledge on choosing appropriate designs to provide useful constraints on the outputs of the systems, which has proven to be effective for non-visualization-experts [MWN*19].

2.2. Guiding and Chart Comparisons

Earlier research has addressed the weaknesses of particular visualizations for distribution analysis. For example, Silverman [Sil86], Tapia & Thompson [TT78], and Scott [Sco92] discuss the problem of histograms for continuous data. In the InfoVis community, many empirical user studies exist which compare charts or visual elements for distribution analysis. To name a few, Correll et al. [CG14] evaluate error bars, box plot variations, gradient plots, and violin charts for judging mean and standard error. Ondov et al. [OJEF19] evaluate the comparison of frequencies in distributions using different layouts such as mirrored and separate bar charts. Correll et al. [CLKS19] investigate the identification of outlier-distributions using bar charts, density plots, and density distributions. Gschwandtner et al. [GBFM16] evaluate six different

visual encodings, including gradient and violin plots, to visualize temporal uncertainty. Finally, Skau et al. [SHK15] evaluate the impact of embellishments in bar charts. While some of the links between analysis tasks and visualizations have been addressed, we still lack a set of guidelines for selecting appropriate charts for comparative distribution analyses, particularly if we are faced with a combination of tasks which should be combined in a single chart.

3. Landscape of Comparative Distribution Analysis

The goals of comparative distribution analysis are manifold. Exploratory data analysis [Tuk77] aims to identify interesting patterns across multiple dimensions, while other applications need sanity checks [CLKS19] to identify missing values, outliers, or skewed distributions that influence the analysis. Comparing distributions is another important aspect of descriptive statistics [FMF12] which can support data cleaning, choosing appropriate statistical models, and understanding potential reasons for significance.

Examples such as Anscombe's Quartet [Ans73] and the recent paper by Matejka & Fitzmaurice [MF17] illustrate that summary statistics are not enough to analyze distributions: visualizations are also needed to discover patterns, support analyses, and communicate results [FMF12]. Nevertheless, established charts often only support a limited number of tasks. Since a comprehensive statistical analysis typically comprises a combination of different aspects, it is often necessary to create and explore several different charts.

To guide analysts towards choosing an appropriate chart for a given application, we first provide a summary of the design space by categorizing existing analysis tasks (Section 3.1) and distribution charts (Section 3.2). Then, in Section 4, we conduct an expert survey to link analysis tasks with charts and derive guidelines.

3.1. Analysis Tasks

Many statistical books (e.g., [FMF12]) enumerate a subset of analysis tasks. However, we are unaware of an overview of all tasks for comparative distribution analysis. Often, tasks are not formally introduced, and different wordings are used for the description. We therefore establish a common vocabulary and introduce a classification of tasks, which is shown in Table 1. We group all tasks by their scope and type, classify them into four complexity levels, and discuss the involved distributions as elaborated below. In Table 1, we also provide examples from psychology applications for all tasks.

Scope of analysis tasks. We group the analysis tasks into three different scopes, which comprise the general analysis focus:
Local tasks concentrate on particular instances, for example reading the frequency of one value (L1) or comparing them across two or more distributions (L4). Local tasks are particularly interesting in discrete distributions (e.g., questionnaire results) in which the characteristic of specific values is of high interest.
Global tasks take the majority of data values into account and analyze the entire distribution. Typical questions are the identification of the distribution type (G1), or the comparison of skewness and kurtosis in different distributions (G4).
Aggregation tasks can be seen as the link between global and local tasks. They focus on aggregated statistical measures of a distribution, such as the identification of mean (A1) or median (A2), or the comparison of quartiles (A8) and standard errors (A10).

Analysis type and involved distributions. We classify the tasks into two analysis types and investigate their involved distributions. **Comparison tasks** ⬌ analyze the *relationships between distributions*. Some tasks compare the frequencies of one distribution (● vs ●), e.g., when identifying the most and least frequent values (L2). Others compare them across different distributions (● vs ●). **Identification tasks** ⌕ focus on reading, measuring, and estimating the individual *properties of one distribution* (●). Examples include extracting the frequency of one value (L1), the identification of the distribution type (G1), and its skewness (G2).

Different analysis types are often interlinked and build on top of each other. For example, to identify the value(s) with the largest and smallest distribution difference (L5), one must first identify the frequency of each value (L1), then compare them across distributions (L4), and finally find the highest difference.

Complexity. We categorized the analysis tasks into four different complexity levels. We define complexity as the number of atomic, consecutive identification and comparison tasks needed to reach an analysis goal. It is not defined as the difficulty of extracting the relevant information from an (optimal) visual representation.
Complexity 1 ●○○○ comprises single identification tasks such as reading the frequency of a value (L1) within one distribution (●).
Complexity 2 ●●○○ are tasks which compare frequencies *within* a distribution (● vs ●). Examples are the identification of the least frequent value (L2), or the distribution type (G1).
Complexity 3 ●●●○ summarizes tasks which compare and relate frequencies *across* different distributions (● vs ●), e.g., the comparison of the distribution shapes (G3). Identifying aggregated statistics (A1–A5) is also considered as complexity 3.
Complexity 4 ●●●● quantifies the similarity and differences between two distributions (● vs ●) or aggregated statistical properties, e.g., identifying the value ranges with the largest and smallest distribution differences (G5), or comparing standard errors (A10).

3.2. Visualization Techniques

We summarize charts, particularly those designed for comparative distribution analysis. We first discuss charts for two distributions, and then for three or more distributions. We group charts into *histogram*, *shape*, and *statistical property based* approaches, as well as *hybrid methods*. We structure them perpendicularly by the taxonomy of Gleicher et al. [GAW*11]: **Juxtaposition** ⓙ designs use a separate chart for each distribution. **Superposition** ⓢ designs represent multiple distributions within the same chart, supporting their comparison. **Explicit Encoding** ⓔ computes the similarity between distributions and directly encodes their difference.

Histogram-based charts represent the frequency as the height of bins. They are particularly useful for discrete distributions, as every value can be depicted by one bar. Binning is usually applied for continuous data. Figure 2 (top) depicts different arrangements of the bins. Separate (a) *bar charts* ⓙ place one chart per distribution next to each other, while (b) *grouped bar charts* ⓢ place bins of different distributions with the same value in one group. Similarly, (c) *stacked bar charts* ⓢ arrange these groups on top of each other, while (d) *mirrored bar charts* ⓢ place bins of the same value next to each other, mirrored by a horizontal axis. The charts can also be rotated (e) for a horizontal layout. The center of each bin can be connected with a (f) *broken line graph* ⓙ. Sometimes the bins are hid-

Table 1: *Classification of analysis tasks for comparative distribution analysis. Tasks are grouped by their scope (*local*, aggregation, and* global*) and ordered by complexity (●○○○ – ●●●●). We differentiate between identification (🔍) and comparison tasks (⬌), and highlight the involved distributions: lookup in one distribution (●), and comparison within (● vs ●) and across distributions (● vs ●).*

		analysis task for descriptive statistics	example	type	complexity	distribution
local	L1	Identify the *frequency* of one value.	[WYRG12]	🔍	●○○○	●
	L2	Identify the *most and least frequent value(s) of* one distribution.	[WYRG12]	🔍	●●○○	● vs ●
	L3	Compare *frequencies within* one distribution.	[WYRG12]	⬌	●●○○	● vs ●
	L4	Compare *frequencies across* multiple distributions.	[PDCM*16]	⬌	●●●○	● vs ●
	L5	Identify the *value(s)* with the *largest and smallest difference*.	[PDCM*16]	🔍	●●●●	● vs ●
aggregation	A1	Identify the *mean* of one distribution.	[vdL15]	🔍	●●●○	●
	A2	Identify the *median* of one distribution.	[SPA11]	🔍	●●●○	●
	A3	Identify the *quartiles* of one distribution.	[MW07]	🔍	●●●○	●
	A4	Identify the *standard deviation* of one distribution.	[vdL15, SPA11]	🔍	●●●○	●
	A5	Identify the *standard error* of one distribution.	[vdL15]	🔍	●●●○	●
	A6	Compare the *means* of multiple distributions.	[vdL15, SPA11]	⬌	●●●●	● vs ●
	A7	Compare the *medians* of multiple distributions.	[GGN*11]	⬌	●●●●	● vs ●
	A8	Compare the *quartiles* of multiple distributions.	[VPBD98]	⬌	●●●●	● vs ●
	A9	Compare the *standard deviations* of multiple distributions.	[dFDPH09]	⬌	●●●●	● vs ●
	A10	Compare the *standard errors* of multiple distributions.	[vdL15, DRC13]	⬌	●●●●	● vs ●
global	G1	Describe and identify the *shape* and *type* of one distribution.	[vdL15]	🔍	●●○○	●
	G2	Describe and identify the *skewness* and *kurtosis* of one distribution.	[vdL15]	🔍	●●○○	●
	G3	Compare the *similarity*, *shape*, and *type* of multiple distributions.	[VSBK11]	⬌	●●●○	● vs ●
	G4	Compare the *skewness* and *kurtosis* of multiple distributions.	[REVN15]	⬌	●●●○	● vs ●
	G5	Identify the *value ranges* with the *largest and smallest difference*.	[BMBP05]	🔍	●●●●	● vs ●

den. In other variations, the (g) lines can be arranged as a superposition 🅢, similar to a line-chart, which allows a better comparison. Finally, (h) *cumulative bar charts* accumulate the frequencies of values and can be arranged as separate 🅢 or grouped bar charts 🅙.

Statistic-property based charts directly depict summary statistics. Popular examples are *box plots* [Tuk77] 🅙 and *error bars* 🅙 (Figure 1 (b), bottom right). Box plots indicate the median and interquartile range (IQR) as a box, representing 75% of the data. Additionally, the whiskers indicate $1.5\times$ IQR and outliers are marked with a dot. Variations have been proposed for the design, such as *notched* and *variable width box plots* [MTL78], along with numerous approaches for coping with skewed distributions [HV08]. Error bars typically represent the mean as a histogram and add the spread of standard error as an interval on top. Further designs comprise the mean and standard deviation, confidence interval, or any other uncertainty measure. A popular variation of the error bars shows only the uncertainty interval without the histogram underneath.

Shape-based charts estimate the distribution using a probability density function. Most implementations use a kernel density estimation (KDE), which depends on the kernel type and bandwidth parameter. Both need to be carefully selected to provide reliable representations [Sil86]. As shown in Figure 2, simple *density distributions* compute a KDE and place the distributions next to each other 🅙 (i), on top of each other 🅢 (j), or in a mirrored fashion 🅢 (k). Violin-type charts are inspired by the idea of opening up the 'black box' of a box plot, illustrating the shape of the distribution. Early work comprised the *histplot* 🅢 and *vaseplot* 🅢 [Ben88],
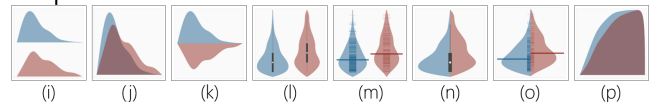
Histogram-based charts
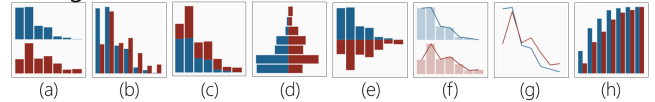
Shape-based charts

Figure 2: *Overview of charts for comparative distribution analysis.*

and the development of the *violin plot* (l) 🅢 [HN98] and *beanplot* (m) [Kam08] 🅢 was based on those ideas. Both encode mean or median and the distribution of the data in the form of a box plot or gradient plot. It can also be extended to an (n) *split violin plot* 🅢 or (o) *asymmetric beanplot* 🅢, similar to the mirrored bar chart. The shape can also be used to encode the cumulative distribution (p) 🅢.

Hybrid charts are used to visualize different aspects of distributions in a single plot. Combinations of a histogram and a kernel density estimation, histogram and box plot, or boxplot with jitter are often used. Potter et al. [PKRJ10] introduce a so-called Summary plot which combines a mirrored histogram with the majority of summary statistics, such as mean, skewness, and quartiles. These hybrid visualizations have proven to be effective in supporting multiple different analysis tasks at the same time. However, as discussed in Section 2, they are typically difficult to generate, particularly if the set of analysis tasks changes between applications.

Visualizations to compare multiple distributions are limited. Most approaches use Juxtaposition 🔵 or Superpositon 🔵 and 'just' concatenate single charts together, e.g., multiple box plots, error bars, stacked or grouped bar charts, line charts, and violin plots. We are not aware of direct encodings 🔵 of multiple distributions.

Related visualization techniques. Many other charts have been created which are often used to represent some aspects of distributions and their properties. A comprehensive summary can be found in the book by Wilkinson [Wil05]. Examples comprise *stem-leaf diagrams*, which follow the idea of histograms but use consecutive numbers to encode bins. This helps to understand the spread of aggregated values. Gradient plots use a transparent line (or point) for each data record. These plots explicitly use overplotting to help seeing all data records, as well as the overall shape. Other, partly related charts are stacked and jittered line plots [CCKT83], linear and nonlinear dotplots [SR96, RW18], and flow charts.

4. Connecting Analysis Tasks to Visual Representations

We now interlink the analysis tasks for descriptive statistics with the design space for visualizations that support distribution comparisons. Our aim is to find out which visualizations and visual elements support users for any given set of analysis task(s) to provide guidelines for application experts. To address these questions, our survey is based on the following methodology and survey design.

Methodology. Since we intend to investigate the usefulness of the most common visualizations for descriptive statistics for different analysis tasks, we selected a set of 20 representative charts covering histogram, shape, and statistical-property based visualizations. Part of our selection is shown in Figure 1b, which is extended by a rotated version of the mirrored bar chart and a commonly-used error bar representation without the median bar. As tasks, we use all local (L1 – L5) and all global tasks (G1 – G5). However, we selected only two representative examples from the aggregated group (A1 + A2). The reason is twofold: Firstly, the pilot study showed that with 20 tasks × 20 charts it would take participants almost two hours to complete the survey. And secondly, feedback from the pilot participants suggested that one would need the direct encoding of the statistical properties to successfully solve an aggregated task. This is true for all aggregated tasks and was also reflected in the results of the pilot study. We also limited the survey to the comparative analysis of two data distributions and included both InfoVis and statistic experts as participants to make sure we covered both statistical and information visualization expertise.

Survey design. For convenience, we designed the survey on paper. It comprises 16 pages and is divided into background material and five parts. Every analysis task was described, and a concrete example given. The participants first indicated the usefulness of every chart for a each analytical task using a Likert scale. Then, they selected up to three techniques which they felt were most useful for the respective task. At the end of the survey, we asked all the participants (a) whether any important and commonly used visualizations were missing; (b) which three plots they would be most likely to use in a paper when addressing most of the tasks, and why; and (c) which plots are most common in the literature, regardless of their usefulness. The survey and the results are available at osf.io/jk8rp.

Participants. 20 participants (6 female) participated in the survey. 17 were PhD students and 3 PostDocs. 13 participants reported their primary background in computer science (CS), 4 in psychology, and 3 in both CS and psychology. Their ages ranged from $25 - 52$ with a mean of 29.8 and a standard deviation (std) of 6.3. Their average experience (and std) was reported as 4.1 years (3.7) in statistics, and 4.0 years (3.0) in information visualization.

Survey procedure. We conducted several pilot runs and iteratively improved the descriptions and tasks in the survey. Since every participant was able to complete the survey at their own convenience we do not know whether the survey was completed in one go, or whether they had any additional help. However, as we are primarily interested in the participants' assessments rather than evaluating their knowledge about the different visualization techniques and supported analysis tasks, we do not see this as a limitation.

Data collection and cleaning. We manually extracted the answers and removed two participants from the results. Since the first had more than 50% missing values in the assessment of the different charts, we did not consider them as 'knowledgeable' in the topic. The second had a lot of outliers in the assessments of local analysis tasks, and most of the answers did not match the average of the other participants. For example, he/she said that bar charts are not useful for identifying the value with the highest frequency - which is obviously possible. As these outliers occurred across all local analysis tasks, we believed that the participant misunderstood local and global analysis tasks, and so he/she was removed. On average, all remaining participants have 3.3% missing values ($std = 4.9\%$) in the assessment of the different visualization techniques.

4.1. Usefulness of Visualizations Across all Analysis Tasks

We report the results of the last tasks to provide general findings.

Most common charts. Based on the survey, bar charts (17 participants), box plots (15), error bars (4), broken line graphs (4), and grouped bar charts (3) are generally most common in the literature.

Visualizations most useful across all tasks. No single visualization technique was considered most useful for all analysis tasks. Many participants reported that the density distribution III is "*good to compare shapes*", "*great for global comparisons*", and enables the "*focus on difference analysis (difference is darker)*". Also, separate bar charts have been reported as useful because they are seen as "*simple*", "*a faithful representation of discrete data*", and "*easy to understand at first glance*". Box plots are considered useful as "*everybody knows them*", and "*you can see the median and quartiles*". Mirrored bar charts are "*easy to understand, have a clear baseline [compared to stacked bar charts], and good to compare distributions*". Finally, the asymmetric bean plot comprises "*comprehensive information in a single graph*", "*looks pretty and captures a lot of information*", and makes it "*easy to compare distributions, means, and non-aggregated values*".

4.2. Usefulness of Visualizations for Different Analysis Tasks

In the following, we analyze the results of the survey on both a task level and across local, global, and aggregated analysis scopes.
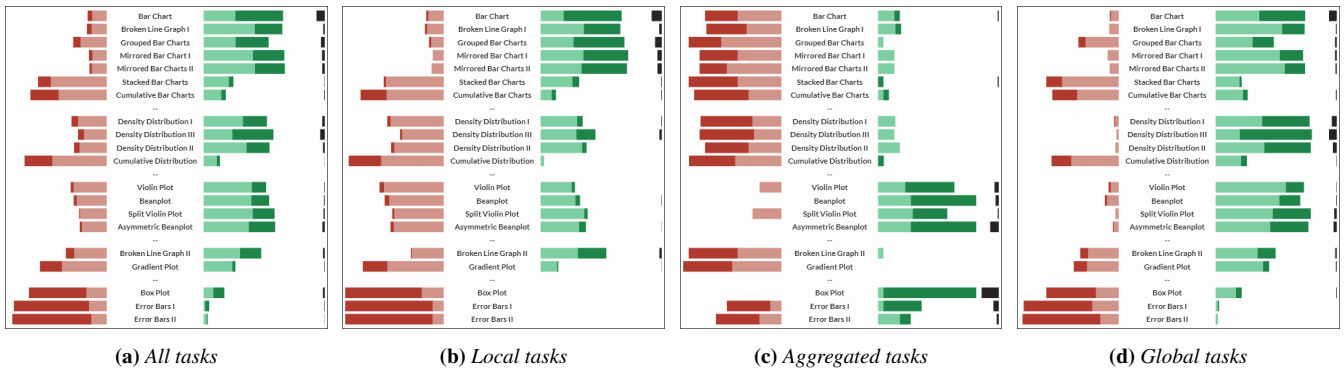
(a) *All tasks* **(b)** *Local tasks* **(c)** *Aggregated tasks* **(d)** *Global tasks*

Figure 3: *Overview of expert survey results. Each chart represents a summary of a task scope according to Table 1. Every participant indicated the usefulness of each visualization for a particular task with (++) very useful, (+) useful, (-) not useful, and (- -) not possible. The vertical black bar charts on the right side represent the three visualizations, marked as most useful by the participants. We can see a clear difference for the usefulness of local, aggregated, and global analysis tasks. Generally, histogram representations are considered more useful for local tasks, while shape representations, such as a density distribution or violin plots are more useful for global tasks. Box plots and error bars, although often seen for comparing distributions, are only considered useful for aggregated tasks such as the identification and comparison of mean values and quartiles. The results of all individual tasks can be found in the supplementary material.*

Visual representation of (un)usefulness. We realized that the participants had different encoding strategies. While some marked almost all techniques that are useful with (++), others applied mainly (+), and (++) only for the most useful technique per analysis task. Therefore, we visualized the results of the survey (see Figure 3) for its exploration. While we cannot rely on the exact proportion between (+) and (++), and (-) and (- -), Figure 3 can provide us with a clear tendency towards (un)usefulness. In the following, we discuss the most important findings (**F**) of the survey.

F1: Charts differ across local, aggregated, and global tasks. Considering the combined results in Figure 3, we can see that there are significant differences in the usefulness of the charts for local, aggregated, and global analysis tasks. In Figure 3b we can see that histogram-based charts are considered most useful for local analysis tasks. While they are also considered useful for global tasks (Figure 3d), the participants seemed to prefer shape-based visualizations such as density distributions and violin-type charts. For aggregated tasks, there was a clear preference for charts that directly encode statistical properties, such as box plots and error bars.

Another interesting observation is that six visualizations were not seen as being useful for most of the tasks. These were stacked and cumulative bar charts, cumulative distributions, box plots, and variations of error bars, as shown in Figure 3a. Surprisingly, while box plots and error bars are often used in the literature, they seem to be of limited usage for considering task variations.

F2: Local analysis tasks. On average, most histogram-based charts are considered useful for local analysis tasks, as shown by the green color in Figure 3b. As expected, stack and cumulative bar charts were an exception, being generally considered as not useful across all tasks, not only those with a local scope. We can see some differences of the other histogram charts within the scope of local tasks: For tasks considering a single distribution (L1 – L3), there is not much difference between single bar charts (with or without a broken line graph) and mirrored bar charts. They all support the identification of a value, as well as their comparison. Grouped bar

charts are considered marginally less useful for these tasks. There was some disagreement on whether shape-based charts are useful for identifying the most (in)frequent values (L2).

Grouped bar charts, as well as density distribution III and broken line graph II, are considered most useful for comparative tasks across distributions, as they directly encode the similarity and difference between the distributions, making comparisons easier.

F3: Global analysis tasks. While many participants considered histograms to be useful for global tasks, the shape-based approaches outperformed them. In particular, when investigating the individual tasks, we can see that they are more useful, especially when comparing distribution shapes (G3), skewness (G4), and identifying differences (G5). Density distribution III, split violin, and asymmetric beanplots seem to be particularly useful for comparisons across distributions. Besides density distribution III, grouped bar charts are also considered very useful for identifying the value ranges with the largest and smallest differences (G5). Single and mirrored bar charts seem to be as useful as density distributions for the identification of the distribution type (G1), skewness, and kurtosis (G2), while grouped bar charts hinder the comparison of frequencies within a distribution and are therefore less useful.

Across the global tasks, we do not see much difference between the simple density distributions (I, II, and III), and the more complex violin and beanplot variants. Participants do not agree on the usefulness of the broken line graph II and gradient plots.

F4: Aggregated analysis tasks. Figure 3c shows that only charts which directly encode statistical measures are considered useful (i.e., all violin-typed charts, box plots, and error bars). The box plot is favored, followed by the asymmetric bean plot. While many participants rated error bars as one of their favorite charts for aggregated tasks, quite a few noted that error bars are not useful for identifying mean or median values. This is surprising, as the height of the error bars can directly encode this value.

F5: Impact of rotating charts. There seems to be no impact on the rotation of charts. Both versions of the mirrored bar charts are almost identical across all tasks. The same is true for the split violin plot and the density distribution II, which are the same except for the rotation (and an additional encoding of the mean value).

4.3. Summary

None of the presented charts support all tasks at the same time. Charts encoding statistical measures are either simplified to these values (e.g., error bars) or part of violin-typed charts, missing a histogram for local tasks. Vice versa, histograms support tasks on single values, but often miss comparisons on a global level or of statistical measures. Furthermore, charts that include a direct encoding of differences (e.g., grouped bar chart or density distribution III) support the comparison of multiple distributions, but they are more complex regarding the analysis of a single distribution as the frequencies of each distribution cannot be followed easily.

Based on our findings, we envision an ideal plot for comparative analysis of data distributions combining (1) a histogram representation for local analysis tasks, (2) a shape-representation for the global tasks, (3) a direct encoding of the differences between two or more distributions, and (4) a representation of statistical measures. Perpendicularly, such a chart should unify aspects of 🅢 and 🅔 in a single chart. While superposition 🅢 layouts help to identify the properties of single distributions, explicit encoding 🅔 helps to compare distributions by highlighting their differences.

5. v-plots: Hybrid Distribution Chart Design

Based on the findings of our survey, we develop the *v-plot designer*; a chart authoring approach which facilitates the design of customizable hybrid charts, so-called *v-plots*. In particular, these charts combine the advantages of several established representations to support local, global, and aggregated tasks in a unified representation. One v-plot compares two distributions, while a combination of different distributions can be arranged in a *v-plot matrix* which can automatically be sorted by similarity using a matrix reordering algorithm. In the following, we discuss the design rationale of v-plots and introduce the guiding wizard of the v-plot designer, which automatically tailors the style of a v-plot to a given set of analysis tasks to highlight particular distribution properties.

5.1. Design Rationale

The fundamental design rationale of v-plots is a layered representation, similar to the idea of the Summary plot [PKRJ10]. Each layer supports different analysis tasks. The total number of layers is five (**"v"**), which gives this hybrid plot its name. The order of layers and their style can both be customized to focus on specific analysis scenarios. All layers are based on well-established visualizations, which makes them easy to use and interpret while still supporting a combination of complex analysis tasks. By default the v-plots are configured to enable all five layers to be visible but not highlighted. Users can adjust the layers and their highlighting by selecting certain tasks, either manually or through the guiding wizard, as described in Section 5.2.

(i) Mirrored bar chart 🅢. The first layer is a mirrored bar chart which supports local tasks on single distributions, such as the identification of frequencies (L1) or their comparison (L3). In discrete distributions, every bin corresponds to one particular value. For continuous distributions, an adjustable equal-width binning is applied. Small values are at the bottom; high values are at the top. The height of each bin corresponds to its relative frequency in the distribution, which also allows the comparison of distributions of different sizes.

(ii) Density distribution 🅢. The global properties of distributions, such as type (G1) and skewness (G2), can be analyzed by a shape-based density distribution. This layer supports two implementations: (1) the center of each bin is used as a control point for a Catmull–Rom spline [CR74]; (2) a kernel-density estimation (KDE) with selectable parameters for the bandwidth and kernel type can be selected. The first option is the default, as it has three advantages compared to the KDE: (1) it is parameter-free, (2) it shows all peaks and valleys properly, and (3) it is directly linked to the underlying mirrored bar chart, linking global and local tasks together.
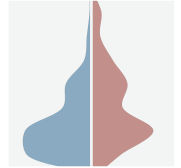
(iii) Direct difference encoding 🅔. We chose a vertical layout as the v-plots' default for supporting the comparisons between both distributions. The symmetrical arrangement allows one to easily see if the two distributions align or not. We also encode the differences of the distributions directly into the inner part of the mirrored bar chart as a *difference shape* or *difference histogram*. This encoding represents the absolute difference between the two relative frequencies and, for example, allows the highest difference between bins (L5), or the value ranges with low differences (G5), to be identified. While the difference histogram supports a direct comparison between bins, the difference shape supports the analysis of more general patterns and is often used in the v-plot matrix as shown below.
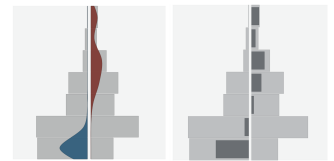
(iv) Statistic measures 🅢 + 🅔. As a fourth layer, we support the encoding and comparison of statistical measures. For each distribution on the left and right side, the analyst can represent a value of central tendency (i.e., mean or median) and the spread of data (i.e., standard deviation, interquartile range, or standard error). The properties can be connected and highlighted with color for a comparative analysis.
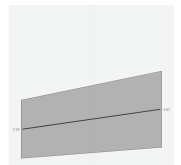
(v) Labels. The final layer comprises various labels, such as the chart title, name of the distributions, a grid, and labels of the bins with the respective frequencies for a detail analysis. The size and position of the labels can be interactively adjusted.

5.2. Guiding Wizard for Task-Dependent v-plot Customization

Based on our survey results, we can guide analysts towards an optimized v-plot for specific analysis task combinations. After uploading their data, the user selects all tasks (c.f. Table 1) that are relevant

for the current analysis question. Using different radio-buttons, the user can specify whether a task is *irrelevant*, *relevant*, or if the visualization should particularly *highlight* the corresponding property.

Automatic v-plot optimization. The guiding wizard is implemented as a rule-based system. While the complete set of rules is documented in the supplementary material, we want to highlight the general concepts in the following: If all local tasks are marked as *irrelevant*, then the mirrored histogram layer is removed (the same holds for density distribution and statistical properties regarding global and aggregated tasks). Tasks marked as *relevant* will typically add elements to the v-plot. For example, if the user wants to identify the frequency of one value (L1), a grid and labels with the bin height are added. If the user wants to identify (A1) and compare the mean values (A6), the statistical layer is added and the mean values are connected, as shown in Figure 4. Changing a task from *relevant* to *highlight* usually results in a darker color and a higher level of the visualization layer. For example, if the user wants to highlight the differences (L5, G5), then the opacity of the difference histogram or difference shape is increased. We also change the difference histogram to the difference shape if only G5 is selected.

Automatic recommendation for basic charts. The v-plot designer also provides an automatic chart recommendation which proposes basic charts which fit best for the given task combination. This allows for a comparison with the optimized v-plot. Based on the selected tasks, the system automatically provides a ranking of all visualizations (Figure 4 bottom) based on a score for each chart. This score is computed using a weighted linear combination of the Likert scale results in the survey, i.e. a chart considered *very useful* and *not possible* are weighted higher than charts considered *(not) useful*: $score_{vis_i} = w_1 \cdot ratio_{(++)} + w_2 \cdot ratio_{(+)} - w_3 \cdot ratio_{(-)} - w_4 \cdot ratio_{(- -)}$. Here, $ratio_{(++)}$ corresponds to the ratio of participants rating the visualization as very useful $(++)$. $w_1 \ldots w_4$ are the weights and by default are set to $w_1 = w_4 = 1.5$ and $w_2 = w_3 = 1.0$.

The ranked visualizations all show the distributions of the uploaded data so the user can compare them easily. The system also provides two perpendicular views of the ranked charts. First, we customize a usefulness chart based on the selected tasks and show the usefulness of every visualization to this task combination (Figure 4 top right). Second, we create a similar representation illustrating the usefulness of every selected task to each chart (Figure 5 bottom). Since not all task combinations can be covered with the existing charts, the user can then select one or multiple charts that cover the underlying analysis question, based on the automatic recommendation and the supportive charts.

While the findings in our survey are the grounding for the guiding wizard and chart recommendation, we want to highlight that this basis is interchangeable. New findings based on other quantitative user studies or the recommendation and guidelines of specific communities can be exchanged by replacing CSV files in our publicly available source code of the v-plot tool (see Section 5.4).

5.3. v-plot Matrix

v-plots are particularly designed to compare *two distributions*. For the comparative analysis of *many distribution* pairs, we extend the v-plot designer to generate a *v-plot matrix*, which arranges all pairwise distributions in a matrix (see Figure 6). This layout allows an-
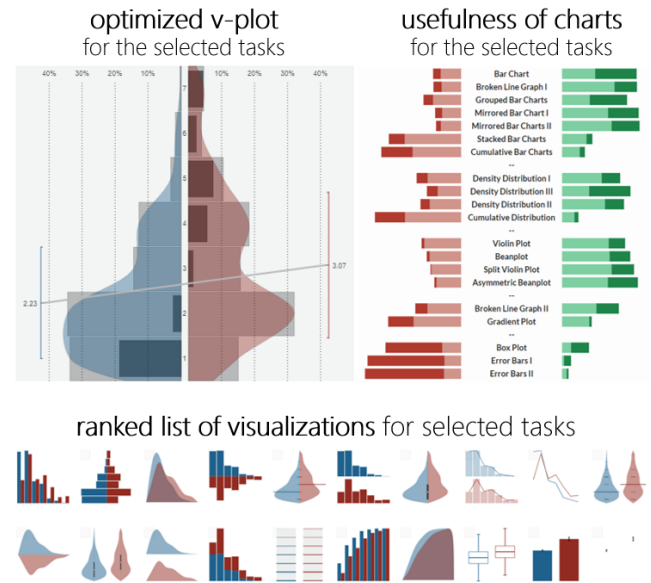
Figure 4: *Based on a set of analysis tasks, the v-plot designer automatically adjusts the v-plot to highlight distribution properties. We show 20 alternative visualizations using the same data, ranked by the perceived usefulness of the task combinations. For each chart type, we also select the top-ranked visualization (see Figure 5).*

alysts to compare one distribution against all others (one row or column), but also helps to find similarities and differences across all distribution pairs. To improve the perception of similarities, we allow users to apply a matrix reordering algorithm to sort rows and columns such that similar distributions or similar difference patterns appear close together. However, rows and columns can also be arranged by semantics, as we show in Figure 6.

Matrices are symmetric, i.e., the upper and lower triangles depict the same distribution pairs. Hence, we support different v-plot styles for both triangles. Each style can be designed independently, either manually or with the help of the guiding wizard. In the example of Figure 6, we are interested in the general shape of the distributions (G1) and the frequency of each value (L1) in the upper triangle, and we want to know in which value ranges the distributions differ most (G5) in the lower triangle. Our guiding wizard therefore proposes starting with a *histogram + shape representation*, and a *difference encoding*. This layout has also generally proven to be useful for providing an overview and initiating a detailed analysis.

5.4. Usability and Implementation Details

We implemented the v-plot designer using d3 and angular.js. The tool is available at v-plot.dbvis.de, along with the source code, which will allow researchers to adjust the guiding wizard and the chart recommendation engine. Users can upload CSV files and directly compare the corresponding v-plot with 20 alternative charts, all using the same data for a useful comparison. All properties can be changed in an interactive menu, which immediately updates them in the v-plot. As shown in the supplementary video, users can,

for example, change the size and aspect ratio, add a tile, modify the grid granularity, and adjust the color and transparency of the histogram, distribution shapes, and the statistical measures. Together with reordering the layers, users can thereby tailor the v-plots to highlight specific distribution properties. The resulting v-plot, as well as the customized style, can be downloaded.

6. Evaluation

To show the effectiveness and usability of the v-plot designer and guiding wizard, we conducted a qualitative expert user study and show a use case with survey results from a psychology application. All designs and interactions in this use case are inspired by the participants of our expert user study.

6.1. Use Case: Risk Perception in Drinking Water

Consider health psychologists investigating why many people buy water in (plastic) bottles, despite the numerous advantages of tap water. The psychologists conducted an experiment [DGB*18] in which tap and bottled water consumers reported their perceived risk with respect to *self-vulnerability* and *worry* when consuming water from the tap, glass, or plastic bottles. To investigate differences and similarities between the two consumer groups on a general level, we need to compare the distributions of 2 consumer groups (tap water and bottled water consumers) × 2 attributes (self-vulnerability and worry) × 3 drinking water categories (water from the tap, a glass bottle, or a plastic bottle) = 12 × 12 = 144 combinations. To do so, we create a v-plot matrix (see Figure 6) and manually sort rows and columns by consumer groups and water types. By simultaneously analyzing the general shape and differences of the groups we make three interesting observations, as highlighted in Figure 6.

❶ When analyzing *tap water consumers* and comparing *worry* of consuming bottled water in *glass vs. plastic*, we can see that tap water consumers are more worried about plastic packaging than glass. We can also see the same distribution difference in *worry* of consuming *water from plastic bottles* (more worried) vs. *water from the tap*. In summary, it seems that tap water consumers are generally more worried about drinking water from plastic bottles, compared to glass bottles or water from the tap.

❷ When comparing *worry* for tap and bottled water consumers, we see that both consumer groups have very similar data distributions. This is also reflected in the difference encoding (lower-left triangle). This means that both tap and bottled water consumers have a similar risk perception about water from plastic bottles. Both distributions are also skewed towards smaller values, indicating a general tendency towards a low risk perception for both groups.

❸ When comparing both consumer groups w.r.t. *self-vulnerability* when *drinking water from the tap*, we can see that the distribution of tap water consumers (blue distribution) is visually more skewed towards smaller values while the distribution of bottled water consumers is skewed towards higher risk perception values. This may indicate that bottled water consumers see water from the tap as a higher risk than consumers that generally drink tap water.

We investigate the last observation in more detail, focusing on the identification and comparison of mean values (A1 + A6), the frequencies across the distributions (L1 + L4), comparisons of distribution shapes (G3), and, in particular, highlighting distribution
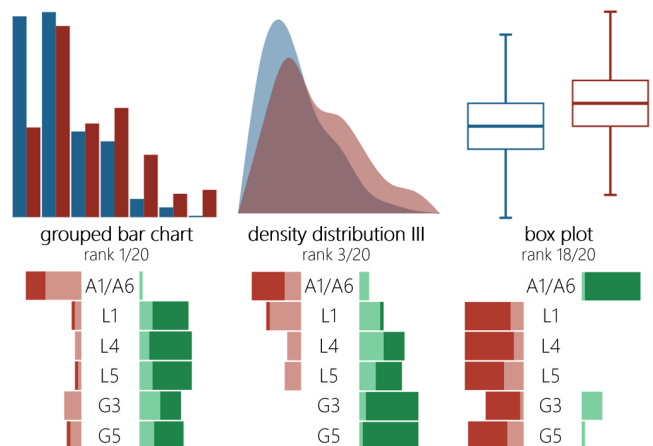

Figure 5: *Based on a set of analysis tasks, we select the top-ranked visualizations for each chart type and show their fitness to all tasks. For a comparison, we select the same tasks as in Figure 4.*

differences (L5 + G5). These tasks can be seen as an often-used combination for a comprehensive comparison of two data distributions. We start the guiding wizard, select the tasks above, and receive a v-plot tailored towards the underlying tasks. Figure 4 shows the v-plot, along with a ranked list of basic visualizations. For each chart type, we also select the top-ranked visualization (see Figure 5). Below each chart, we show its usefulness for each of the selected tasks. We can clearly see that these basic charts only support a subset of the tasks. For example, the density distribution support the global tasks to identify and compare the shapes of the distribution, while box plots support the identification and comparison of median values. The optimized v-plot from the same data, also shown in Figure 1a, supports all analysis question in a single chart.

6.2. Qualitative Expert User Study

We conducted a pair analytics study [KF14] with four domain experts, DE1 – DE4, to evaluate and get feedback on our v-plot designer. All participants were PhD students in the field of psychology. Two were female, the age range was 25 to 33 years, and the reported experience in statistics varied from 4 to 10 years. We explained our aim of evaluating a chart authoring visualization to compare data distributions and asked all participants to bring a dataset which they are currently exploring. Each session took one hour and was structured as follows: The DE first explained the dataset and the visualizations that s/he commonly uses. Then, the visualization expert (VE) introduced v-plots with the different layers and parameter settings, and the DE analyzed this own dataset. Thereafter, the VE introduced the guiding wizard and the DE started tailoring the analysis toward specific task combinations. In a second step, we provided a new dataset [KGSR17], not known to the DE before, and let the DE create hypotheses and explore the data freely, allowing us to observe the participant's action and approach. As a last step, we introduced the v-plot matrix and let the participants explore the pair-wise relationships. We ended the study with questions about the general assessment of the usability of the v-plot designer with the guiding wizard. Occasionally, the

VE asked for feedback during the study and guided the **DE** towards new tasks. The **DE** was encouraged to think aloud during the study so that we were able to capture their thought process.

Findings. Due to privacy constraints of the participants' own datasets, we will only summarize general findings. All participants reported that they normally use histograms as a first approach to get an overview of the data. **DE2** said that his/her biggest challenge is "*to keep an overview over the data and not get lost in the jungle of variables*". **DE2** also mentioned that s/he is satisfied with the visualizations s/he uses, but sometimes they are difficult to generate. **DE4** uses Tableau [Tab20] on a regular basis to create more advanced visualizations to compare global vs. local aspects.

After introducing the v-plots, **DE2** mentioned that having the distributions and aggregated statistics combined in one graph is useful. **DE2+3** liked that changes in the menu are translated directly to the chart, making step-by-step adjustments easily possible. **DE2** said "*the adjustment of the v-plot is very easy. Particularly compared to R and SPSS, where you need 100 clicks to do 10 changes. Here you need 10 clicks to do 10 changes.*" **DE3+4** raised some skepticism as the v-plots may show more than they actually need, which can make the analysis more complex. But they liked that "*there is lots of information that you usually only get by combining information from two different graphs and one table*". All participants particularly endorsed the difference encoding. Spotting differences was particularly easy with this feature, and the participants agreed, that this is a core task in their common analyses that is not supported well in the charts they typically use. All participants were able to find new insights into their data. **DE3** and **DE4** felt that even if the mean values did not differ between groups, there were interesting differences between the distributions that they spotted with the help of the v-plots and did not know beforehand.

We then introduced the guiding wizard, along with our classification of analysis tasks. By exploring different tasks and task combinations, the participants automatically optimized the v-plots and compared them to the alternative visualizations. We regularly asked for the visualization they liked most. **DE1** said, "*if I wanted to display one attribute (i.e. standard error), I would choose the box. For displaying a single task, I do not think the v-plots are necessarily better compared to other visualizations. But if I want to display multiple tasks, I would choose the v-plots. They are very good at displaying the combination of tasks.*" **DE4** agreed and added "*[for multiple tasks] I would definitely choose the v-plot, because the other charts only display single attributes. The more complex the attributes get [...] the more I would tend towards the v-plot.*" **DE2** preferred the bar charts for the global tasks, because they entailed more information in the view. Adding an aggregated task to a global task, s/he would manually add the mean level to the graph.

The wizard was generally liked by the participants. **DE3+4** stated that the wizard was particularly useful when getting started, because it suggests a quick and good starting point based on what is important for the current analysis. The alternative visualizations were positively emphasized by all participants. **DE4** said that "*the wizard shows me which visualization I can normally use and how attributes are displayed by the v-plot*". **DE3** said that the recommendations of alternatives are reasonable.

After introducing the v-plot matrix, **DE3** and **DE4** said that they initially found the v-plot matrix overwhelming. But all partic-
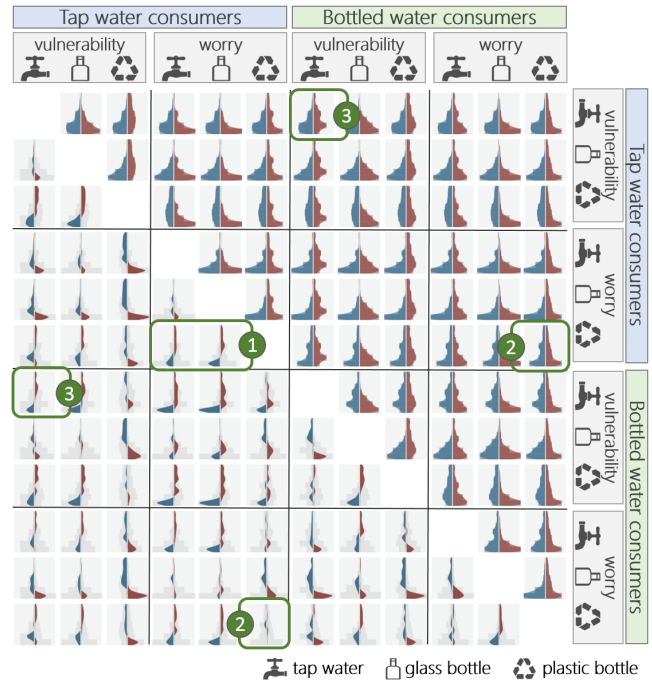

Figure 6: *A v-plot matrix comparing the risk perception of tap and bottled water consumers w.r.t. their* self-vulnerability *and* worry *when consuming water from a tap, a glass, or a plastic bottle. The upper triangle of the matrix shows a density distribution for global analysis, the lower triangle a difference encoding for comparison.*

ipants agreed that if one worked through the matrix and explored the different patterns, it's a good way to get an overview over the dataset and extract interesting attribute combinations. **DE2** particularly liked the matrix: "*I do not have to create v-plots one by one, I get the combination of all plots right away. I like it!*"

We asked the participants for general feedback at the end of the study. All mentioned that the v-plot designer was very intuitive, but one needs a few minutes to understand all the v-plots' layers. All participants agreed that they would use the tool for more than just getting an overview over new datasets; they would also try to incorporate the v-plots in a paper, poster or presentation as an eye-catcher and a dense source of information, if enough time or space for explanations was given. **DE1** said "*I would include a v-plot in my paper if displaying the combination of several attributes was important to me*".

7. Discussion

This section summarizes the main findings, lessons learned, and limitations by reflecting on our multistage research process, in particular summarizing the results of the qualitative expert study.

7.1. Summary and Lessons Learned

Our evaluation shows the advantages and usefulness of the v-plots and v-plot matrix when *simultaneously analyzing different analysis tasks*. Based on feedback from the participants and our own observations, we can summarize the following lessons learned.

Understanding the visual elements. Even though v-plots are designed in an easy way by combining existing charts into a unified representation, one still needs some training to understand and interpret all visual elements. **DE1** said that while s/he would also like to use the v-plots in papers and conference talks they are not very well-known, meaning extra time would be needed for explanations.

Usability and direct feedback. All participants in our study liked that changes in the menu are directly reflected in the v-plots. This helps to understand the impact of specific parameters and to adjust and tailor the v-plot design incrementally.

Difference encoding. All participants repeatedly highlighted that the difference encoding of the v-plots (i.e., the difference histogram for local analysis and the difference shape for global analysis tasks) is one of the most important visual elements. Compared to other representations, this facilitates concentrating on the differences (only) and so helps significantly when comparing distributions.

Single vs. combination of analysis tasks. Some participants reported that v-plots might be too complex for single analysis tasks. For example, they would prefer a box plot if the main goal is to compare the median and quartiles of different distributions. However, our user case and the feedback of the participants also show that existing visualizations often fall short when a combination of local, global, and aggregated tasks is required. In this case, the layered concept of the v-plots supports a comprehensive analysis.

Guided analysis and automatic chart recommendation. Comparing the v-plots with alternative charts was well received by the participants. In particular, our system automatically proposes the most useful charts, after which the participants could make an educated decision on whether v-plots are the appropriate technique for a specific task. Furthermore, they liked that the v-plots can be tailored to the analysis by highlighting specific distribution properties.

7.2. Limitations and Future Work

Our main focus is to design a guided authoring approach for hybrid charts, supporting comparative distribution analysis. To make such visualizations accessible and useful for non-experts, a central part of the v-plot designer is the guidance component that is grounded in a design study of analysis tasks and visualization techniques. In this section, we discuss four limitations of the current state of our approach and highlight potential for future research.

Coverage of visualization techniques. As presented in Section 3.2, there is a wide range of visualizations available for the analysis of data distributions. In this paper, we deliberately focus on analyzing elements of some of the most-commonly-used, basic charts. However, we plan to extend this work to include more charts and visual elements. Most notably, we plan to include *other visual representations* such as dotplots as a potential additional layer to the customizable hybrid plots. We also intend to further investigate approaches for communicating *uncertainty* in the v-plot design.

Guidelines limited to scope of expert survey. To establish a foundation for user guidance, we relied on surveying the usage and analysis patterns of 20 practitioners. In addition to the theoretical foundation provided by the related work, the design guidelines were grounded in our survey. Further research and replication studies are needed to avoid potential sample biases based on the number of survey participants and their background. We provide all data from our survey at osf.io/jk8rp for transparency and have implemented the recommendations of our guiding wizard to be modular, i.e., subject to *adaptation and renewal through the availability of new findings*.

Lack of quantitative evidence. Going beyond the qualitative analysis and evaluation of our approach, there is a research opportunity to examine the *cognitive effects* of combining chart elements. Our evaluation suggests that the correct interpretation of the v-plot layers might be explained by the familiarity of the chart elements, as well as learning effects through usage. However, more studies are needed to determine the usefulness of individual components and their combinations. In particular, we plan a quantitative user study to evaluate the performance of the v-plots and to identify when it is beneficial to switch from a simple representation to the v-plots. This can further improve our guiding wizard.

Complexity of the v-plot matrix. We also plan to further improve our v-plot matrix. Some participants mentioned that presenting so many charts at the beginning of the analysis might be overwhelming. We therefore plan to add interaction concepts such as linking and brushing, highlighting, and attribute filtering directly to the matrix. We further want to automatically highlight interesting v-plots, for example with pattern matching and similarity search, as well as an automatically applied statistical analysis which only extracts significantly different distribution pairs.

8. Conclusion

*How can we make hybrid charts for the visual comparison of data distributions which (1) **simultaneously support local, aggregation-based, and global analysis tasks**; and are (2) **accessible to analysts**?* Our paper addresses this research question by first classifying existing tasks for comparative distribution analysis and exploring the design space of appropriate visualizations. Based on a representative expert survey with 20 participants, we develop an automatic chart recommendation which proposes appropriate charts for a given combination of analysis tasks. As a second main contribution, we develop the v-plot designer as a chart authoring tool for hybrid v-plots, allowing data distributions to be compared simultaneously on global, local, and aggregated levels. Furthermore, we introduce a guiding wizard which tailors the style of the v-plots towards given analysis tasks. Our evaluation shows that this wizard helps to design effective v-plots through highlighting specific distribution properties. Once a combination of analysis tasks is relevant, v-plots outperform other techniques.

Acknowledgments

References

[Ans73] ANSCOMBE F. J.: Graphs in statistical analysis. *The American Statistician 27*, 1 (1973), 17–21. doi:10.2307/2682899. 3

[Ben88] BENJAMINI Y.: Opening the box of a boxplot. *The American Statistician 42*, 4 (1988), 257–262. doi:10.2307/2685133. 4

[BKC*13] BORGO R., KEHRER J., CHUNG D. H. S., MAGUIRE E., LARAMEE R. S., HAUSER H., WARD M., CHEN M.: Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In *Eurographics - State of the Art Reports* (2013), Eurographics Association, pp. 39–63. doi:10.2312/conf/EG2013/stars/039-063. 2

[BMBP05] BIRSAN M.-V., MOLNAR P., BURLANDO P., PFAUNDLER M.: Streamflow trends in Switzerland. *Journal of Hydrology 314*, 1 (2005), 312 – 329. doi:10.1016/j.jhydrol.2005.06.008. 4

[BSS*19] BEHRISCH M., STREEB D., STOFFEL F., SEEBACHER D., MATEJEK B., WEBER S. H., MITTELSTÄDT S., PFISTER H., KEIM D. A.: Commercial Visual Analytics Systems – Advances in the Big Data Analytics Field. *IEEE Trans. Vis. Comput. Graph. 25*, 10 (2019), 3011–3031. doi:10.1109/TVCG.2018.2859973. 2

[CCKT83] CHAMBERS J. M., CLEVELAND W. S., KLEINER B., TUKEY P. A.: *Graphical Methods for Data Analysis*. Wadsworth, 1983. doi:10.1201/9781351072304. 5

[CG14] CORRELL M., GLEICHER M.: Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error. *IEEE Trans. Vis. Comput. Graph. 20*, 12 (2014), 2142–2151. doi:10.1109/TVCG.2014.2346298. 2

[CLKS19] CORRELL M., LI M., KINDLMANN G. L., SCHEIDEGGER C.: Looks Good To Me: Visualizations As Sanity Checks. *IEEE Trans. Vis. Comput. Graph. 25*, 1 (2019), 830–839. doi:10.1109/TVCG.2018.2864907. 2, 3

[CR74] CATMULL E., ROM R.: A class of local interpolating splines. In *Computer aided geometric design*. Academic Press, 1974, pp. 317–326. doi:10.1016/B978-0-12-079050-0.50020-5. 7

[dFDPH09] DE FRANÇA DORIA M., PIDGEON N., HUNTER P. R.: Perceptions of drinking water quality and risk and its effect on behaviour: A cross-national study. *Science of The Total Environment 407*, 21 (2009), 5455 – 5464. doi:10.1016/j.scitotenv.2009.06.031. 4

[DGB*18] DEBBELER L. J., GAMP M., BLUMENSCHEIN M., KEIM D., RENNER B.: Polarized but illusory beliefs about tap and bottled water: A product- and consumer-oriented survey and blind tasting experiment. *Science of The Total Environment 643* (2018), 1400 – 1410. doi:10.1016/j.scitotenv.2018.06.190. 1, 9

[DRC13] DREWNOWSKI A., REHM C. D., CONSTANT F.: Water and beverage consumption among children age 4–13y in the United States: analyses of 2005–2010 NHANES data. *Nutrition Journal 12*, 1 (2013), 85. doi:10.1186/1475-2891-12-85. 4

[FMF12] FIELD A., MILES J., FIELD Z.: *Discovering statistics using R*. Sage publications, 2012. 1, 3

[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. 3

[GBFM16] GSCHWANDTNER T., BÖGL M., FEDERICO P., MIKSCH S.: Visual Encodings of Temporal Uncertainty: A Comparative User Study. *IEEE Trans. Vis. Comput. Graph. 22*, 1 (2016), 539–548. doi:10.1109/TVCG.2015.2467752. 2

[GGN*11] GORELICK M. H., GOULD L., NIMMER M., WAGNER D., HEATH M., BASHIR H., BROUSSEAU D. C.: Perceptions About Water and Increased Use of Bottled Water in Minority Children. *Archives of Pediatrics & Adolescent Medicine 165*, 10 (2011), 928–932. doi:10.1001/archpediatrics.2011.83. 4

[HN98] HINTZE J. L., NELSON R. D.: Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician 52*, 2 (1998), 181–184. doi:10.1080/00031305.1998.10480559. 2, 4

[HV08] HUBERT M., VANDERVIEREN E.: An adjusted boxplot for skewed distributions. *Computational Statistics & Data Analysis 52*, 12 (2008), 5186 – 5201. doi:10.1016/j.csda.2007.11.008. 4

[Kam08] KAMPSTRA P.: Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. *Journal of Statistical Software, Code Snippets 28*, 1 (2008), 1–9. doi:10.18637/jss.v028.c01. 4

[KF14] KAASTRA L. T., FISHER B. D.: Field experiment methodology for pair analytics. In *Proceedings of the Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization* (2014), pp. 152–159. doi:10.1145/2669557.2669572. 9

[KGSR17] KÖNIG L. M., GIESE H., STOK F. M., RENNER B.: The social image of food: Associations between popularity and eating behavior. *Appetite 114* (2017), 248 – 258. doi:10.1016/j.appet.2017.03.039. 9

[MEC*17] MAURI M., ELLI T., CAVIGLIA G., UBOLDI G., AZZI M.: RAWGraphs: A Visualisation Platform to Create Open Outputs. In *Proceedings of the Conference on Italian SIGCHI* (2017), ACM, pp. 1–5. doi:10.1145/3125571.3125585. 2

[MF17] MATEJKA J., FITZMAURICE G. W.: Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. In *Proceedings of the Conference on Human Factors in Computing Systems* (2017), pp. 1290–1294. doi:10.1145/3025453.3025912. 3

[MTL78] MCGILL R., TUKEY J. W., LARSEN W. A.: Variations of box plots. *The American Statistician 32*, 1 (1978), 12–16. doi:10.2307/2683468. 4

[MW07] MCCABE G. J., WOLOCK D. M.: Warming may create substantial water supply shortages in the Colorado River basin. *Geophysical Research Letters 34*, 22 (2007). doi:10.1029/2007GL031764. 4

[MWN*19] MORITZ D., WANG C., NELSON G. L., LIN H., SMITH A. M., HOWE B., HEER J.: Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco. *IEEE Trans. Vis. Comput. Graph. 25*, 1 (2019), 438–448. doi:10.1109/TVCG.2018.2865240. 2

[OJEF19] ONDOV B. D., JARDINE N., ELMQVIST N., FRANCONERI S.: Face to Face: Evaluating Visual Comparison. *IEEE Trans. Vis. Comput. Graph. 25*, 1 (2019), 861–871. doi:10.1109/TVCG.2018.2864884. 2

[PDCM*16] PINKET A.-S., DE CRAEMER M., MAES L., DE BOURDEAUDHUIJ I., CARDON G., ANDROUTSOS O., KOLETZKO B., MORENO L., SOCHA P., IOTOVA V., ET AL.: Water intake and beverage consumption of pre-schoolers from six European countries and associations with socio-economic status: the ToyBox-study. *Public Health Nutrition 19*, 13 (2016), 2315–2325. doi:10.1017/S1368980015003559. 4

[PKRJ10] POTTER K., KNISS J., RIESENFELD R. F., JOHNSON C. R.: Visualizing Summary Statistics and Uncertainty. *Comput. Graph. Forum 29*, 3 (2010), 823–832. doi:10.1111/j.1467-8659.2009.01677.x. 4, 7

[REVN15] RICE J. S., EMANUEL R. E., VOSE J. M., NELSON S. A. C.: Continental U.S. streamflow trends from 1940 to 2009 and their relationships with watershed spatial characteristics. *Water Resources Research 51*, 8 (2015), 6262–6275. doi:10.1002/2014WR016367. 4

[RLB19] REN D., LEE B., BREHMER M.: Charticulator: Interactive Construction of Bespoke Chart Layouts. *IEEE Trans. Vis. Comput. Graph. 25*, 1 (2019), 789–799. doi:10.1109/TVCG.2018.2865158. 2

[RW18] RODRIGUEZ N., WEISKOPF D.: Nonlinear Dot Plots. *IEEE Trans. Vis. Comput. Graph. 24*, 1 (2018), 616–625. doi:10.1109/TVCG.2017.2744018. 5

[Sco92] SCOTT D. W.: *Multivariate density estimation: theory, practice, and visualization*. Wiley, 1992. doi:10.1002/9780470316849. 2

[SH14] SATYANARAYAN A., HEER J.: Lyra: An Interactive Visualization Design Environment. *Comput. Graph. Forum 33*, 3 (2014), 351–360. doi:10.1111/cgf.12391. 2

[SHK15] SKAU D., HARRISON L., KOSARA R.: An Evaluation of the Impact of Visual Embellishments in Bar Charts. *Comput. Graph. Forum 34*, 3 (2015), 221–230. doi:10.1111/cgf.12634. 3

[Sil86] SILVERMAN B. W.: *Density estimation for statistics and data analysis*, vol. 26. CRC press, 1986. doi:10.1201/9781315140919. 2, 4

[SPA11] SAYLOR A., PROKOPY L. S., AMBERG S.: What's Wrong with the Tap? Examining Perceptions of Tap Water and Bottled Water at Purdue University. *Environmental Management 48*, 3 (2011), 588–601. doi:10.1007/s00267-011-9692-6. 4

[SR96] SASIENI P. D., ROYSTON P.: Dotplots. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 45*, 2 (1996), 219–234. doi:10.2307/2986156. 5

[SRHH16] SATYANARAYAN A., RUSSELL R., HOFFSWELL J., HEER J.: Reactive Vega: A Streaming Dataflow Architecture for Declarative Interactive Visualization. *IEEE Trans. Vis. Comput. Graph. 22*, 1 (2016), 659–668. doi:10.1109/TVCG.2015.2467091. 2

[Tab20] Tableau, 2020. https://www.tableau.com/, accessed on 2020-02-13. 2, 10

[TT78] TAPIA R. A., THOMPSON J. R.: *Nonparametric probability density estimation*. Johns Hopkins University Press, 1978. 2

[Tuk77] TUKEY J. W.: *Exploratory Data Analysis*. Pearson, 1977. 1, 2, 3, 4

[vdL15] VAN DER LINDEN S.: Exploring Beliefs About Bottled Water and Intentions to Reduce Consumption: The Dual-Effect of Social Norm Activation and Persuasive Information. *Environment and Behavior 47*, 5 (2015), 526–550. doi:10.1177/0013916513515239. 4

[VPBD98] VEGA M., PARDO R., BARRADO E., DEBÁN L.: Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research 32*, 12 (1998), 3581 – 3592. doi:10.1016/S0043-1354(98)00138-9. 4

[VSBK11] VILLARINI G., SMITH J. A., BAECK M. L., KRAJEWSKI W. F.: Examining Flood Frequency Distributions in the Midwest U.S. *JAWRA Journal of the American Water Resources Association 47*, 3 (2011), 447–463. doi:10.1111/j.1752-1688.2011.00540.x. 4

[Wil05] WILKINSON L.: *The Grammar of Graphics (Statistics and Computing)*. Springer-Verlag, Berlin, Heidelberg, 2005. doi:10.1007/0-387-28695-0. 5

[WMA*16] WONGSUPHASAWAT K., MORITZ D., ANAND A., MACKINLAY J. D., HOWE B., HEER J.: Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations. *IEEE Trans. Vis. Comput. Graph. 22*, 1 (2016), 649–658. doi:10.1109/TVCG.2015.2467191. 2

[WQM*17] WONGSUPHASAWAT K., QU Z., MORITZ D., CHANG R., OUK F., ANAND A., MACKINLAY J. D., HOWE B., HEER J.: Voyager 2: Augmenting Visual Analysis with Partial View Specifications. In *Proceedings of the Conference on Human Factors in Computing Systems* (2017), ACM, pp. 2648–2659. doi:10.1145/3025453.3025768. 2

[WYRG12] WRIGHT J. A., YANG H., RIVETT U., GUNDRY S. W.: Public perception of drinking water safety in South Africa 2002–2009: a repeated cross-sectional study. *BMC Public Health 12*, 1 (2012), 556. doi:10.1186/1471-2458-12-556. 4