A Qualitative Evaluation and Taxonomy of Bar Chart Annotations

Md Dilshadur Rahman*
University of Utah

Ghulam Jilani Quadri†
University of North Carolina at Chapel Hill

Paul Rosen‡
University of Utah

ABSTRACT

When sharing visualizations, annotations provide valuable insights into the data by focusing attention on important visual elements. As a result, annotations have become an essential part of visualizations, primarily when externalizing data or engaging in collaborative analysis. Therefore, it is crucial to understand how people annotate visualizations. This study investigates how visualization students annotate bar charts when asked to answer high-level questions about the data in the charts. The resulting annotations were coded and summarized into a taxonomy with several interesting findings. For example, we notice that several annotation types were broadly applied, while others were just used in special cases. In addition, we found ensembles of annotations were required to sufficiently annotate some tasks. Our findings provide an early framework for contextualizing the usage of annotations, further providing guidance for best practice uses.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods

1 INTRODUCTION

In 2020, a Georgia Department of Public Health (GDPH) visualization became a viral sensation for its misleading nature [2]. While technically accurate and well-labeled, the grouped bar chart was seen as deceptive for its unorthodox ordering of bars, ordered highest-to-lowest instead of chronologically. One approach to resolving such ambiguity is clear labeling of the graphic [4], which leads to the question of what constitutes clear labeling of the graphic? In other words, how might someone annotate the graphic to make it clearer?

To better understand the process and options available for annotating a graphic, we set about to study the forms of annotations used by undergraduate and graduate visualization students when prompted with high-level questions about the data. We coded and summarized the resulting annotated visualizations (see Fig. 1), which showed that 5 primary annotation categories were used: enclosure, connectors, text, marks, and color, on the 5 low-level tasks: retrieve value, filter, compute derived value, find extremum, and sort.

2 METHODOLOGY & STUDY

As our goal was to understand how people annotate bar charts when prompted with questions about the data, we evaluated annotation patterns on grouped bar charts based on the GDPH visualization.

Participants The participants are the students from the "Interactive Data Visualization" course taught by the authors at the University of South Florida in Spring 2022. The course was a cross-listed elective for senior-level undergraduate and Master's- and Ph.D.-level graduate students with a total of 39 students (21 undergraduate and 18 graduate). The assignment was given approximately halfway through the semester.

*e-mail: dilshadur@sci.utah.edu
†e-mail: jiquad@cs.unc.edu
‡e-mail: prosen@sci.utah.edu

Study Procedure Each students was given an assignment that had 3 bar charts, each with 4 questions/high-level tasks (12 questions in total), which was to be completed individually. For each question in the assignment, subjects were asked to: 1) identify the low-level tasks needed to answer the questions and 2) annotate the visualization to make answering that question easier. We began with a brief in-class discussion (10-15 minutes) about the GDPH visualization. We then verbally reviewed the written assignment instructions and answered any questions. Students had 7 days to complete the assignment, which was done out of class. We did not require them to use any specific tool to complete the work.

Datasets & Visualization We wanted each student to have different data but similar trends within the data. Therefore, the datasets we used were generated using a random number for each county from the GDPH visualization from April 16 to May 9. We ensured that each random number fell within a specific range, e.g., for Fulton on May 4, the range was set to a minimum of 33 and a maximum of 56. We used Vega-Lite [3] to create the grouped bar charts from the generated data (e.g., see Fig. 2). As with the GDPH visualization, we had 5 bars, 1 per county, of different colors for each day from April 16 to May 9. We used 2 variants of the x-axis, a non-chronological version (like the viral GDPH visualization) and a chronological version (like the corrected GDPH visualization). In total, we generated 150 bar charts, half chronological and half non-chronological. We built 40 unique assignments of 3 bar charts each. Half of them had 2 bar charts with chronological dates, 1 with non-chronological dates, whereas the other half had 2 bar charts with non-chronological dates and 1 with chronological dates.

High-Level Tasks Students were asked to annotate the visualization based on 4 types of question about the data:

- finding a specific value (e.g., how many COVID cases are there in Hall County on May 5?);
- filtering some values from others (e.g., which counties have fewer than 40 cases of COVID on May 1?);
- aggregating (e.g., how many total COVID cases does Dekalb County have from May 1 to May 4?); and
- sorting (e.g., sort the counties in descending order based on the number of COVID cases on May 8).

2.1 Subject Tasks

Low-Level Task Identification We developed the high-level tasks so that the students could answer them by performing one or more low-level analysis tasks. We used 5 of Amar et al.'s [1]

Task	Enclosure				Connector		Text			Mark		Color				Other
	Ellipse	Bracket	HalfBox	Rectangle	Arrow	Line	Description	Values	Legend	Identifier	Legend	Highlights	Category	Questions	Values	
Retrieve value	15	4	0	14	18	52	19	11	4	0	1	0	33	14	2	0
Filter	20	18	10	31	34	47	35	4	11	19	10	7	30	55	2	0
Compute Derived Value	1	6	0	8	0	9	11	8	4	0	1	0	15	3	1	5
Find Extremum	11	0	0	5	11	15	21	7	3	0	3	0	14	9	0	0
Sort	0	0	0	1	6	9	14	1	3	0	1	0	6	3	0	0
Total	47	28	10	59	69	132	100	31	25	19	16	7	98	84	5	5

Figure 1: Our taxonomy of annotations with the frequency of different tasks. Colors indicate frequency: 1-9 10-19 20+

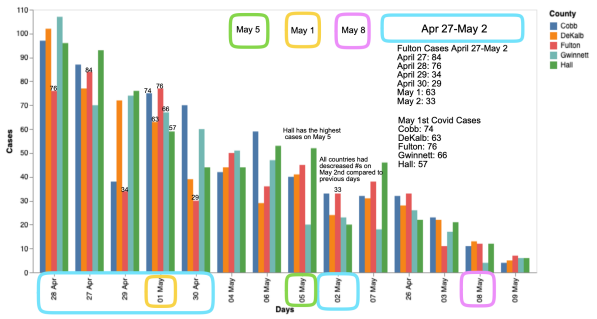

Figure 2: An example of annotated bar chart submission.

frequently used low-level visual analysis tasks, including: *retrieve value (RV)*, *filter*, *compute derived value (CDV)*, *find extremum (FE)*, and *sort*. Students were asked to enumerate which of these low-level tasks were used to answer each of the high-level questions.

Annotating the Visualization We instructed the students to annotate the charts while looking for answers to the questions associated with the charts. Their goal was to annotate the charts in a way that made the questions as easy as possible to determine without necessarily writing the answer on the visualization.

3 Evaluation

A total of 38 students completed the study. Once we received all the worksheets, we anonymized them, assigned each a random number, and performed quality checks. We excluded 18 submissions from the analysis (13 had no annotations, and 5 had minimal annotations).

To evaluate and summarize the submitted assignments, 2 co-authors went through all the annotated charts and hand-coded them in several iterations. In the first iteration, they separately identified a set of annotation types. Next, all authors discussed their findings and agreed on an initial taxonomy of annotation types. In the second and subsequent iterations, the 2 coding co-authors independently revisited each submitted chart, re-categorizing the annotations, and the group revised the taxonomy. This process continued for several iterations until a common taxonomy was converged upon.

4 Taxonomy

Ultimately, we identified 15 annotation types which were grouped into 5 top-level annotation categories: *Enclosure*, *Connector*, *Text*, *Mark*, and *Color*. The resulting taxonomy is shown in Fig. 1.

Enclosure Enclosure is an annotation that uses a partially or fully closed boundary, including the *ellipse*, *bracket*, *half-box*, and *rectangle* categories. Enclosure was generally used for RV and filter tasks, and CDV and FE tasks to a lesser extent. Ellipse and rectangle annotations were used most frequently.

Connector A connector is an annotation that uses a line (e.g., solid, dotted, or directional line) to answer questions about data in the bar charts, which includes *arrow* (directional) and *line* (undirected) categories. Students used connector annotations for RV, filter, and, to a lesser extent, FE tasks. Some of them also used connectors for sorting. Notably, most of the time, connectors did not appear alone but were combined with other annotations for a given task.

Text Text is an annotation that uses words or sentences to answer questions about the data, with second-level categories of *descriptions*, *values*, and *legends*. A *description* annotation is defined in our taxonomy as text that describes a process, information, computation, or derivation that supports a particular task annotation. *Value* is another annotation that is a specific text-based annotation used to highlight exact data of the bar where a number is used to represent the bar length. *Legend* is the final text-based annotation, which was used for any labels on legends, such as using a specific color on the enclosure for a particular date or annotating the trend in a number of cases in a particular county. Legends also comprise the text used marks to identify particular questions or annotations. From Fig. 1, description-based text annotations were broadly applied for all 5 tasks. Text enumerating values was used frequently for RV tasks. Finally, legend text was frequently used for filter tasks. Overall, text annotations seemed to be used for elaborating on what other annotations were highlighting and as a last resort when no other annotations were suitable.

Mark Mark is an annotation that uses symbols or shapes to answer questions about data, which is divided into the second-level hierarchy that includes *identifiers* and *legends*. *Identifiers* are when a mark was used to denote a particular category (e.g., a county or group). Whereas the *legend* mark occurred specifically when marks were used in a legend. From Fig. 1, mark-based annotations were primarily used for filter tasks.

Color Color annotations are any property of color (generally but not exclusively hue) to answer questions about data, which is divided into second-level hierarchy categories, including *highlights*, *category*, *questions*, and *values*. The *highlight*-based annotations use different colors to highlight an area, bars, or group of bars that helps annotate a given task. Color for *values* annotations are an annotation type that aligns with text for value annotations (similar to a color map). The *category*-based color annotations highlight different categories (e.g., counties, dates/months, etc.) and are typically used in collaboration with mark- and enclosure-based annotations. The *question*-based color annotates the different questions (e.g., Q1, Q2, etc.) used in coordination with mark-based annotations. From Fig. 1, color-based annotations were used heavily for all tasks, except sort, while highlights were only used for the filter task.

5 Discussion

Our results provide an early framework for contextualizing annotations and low-level tasks they support, which provide guidance for best practice usage. Within this context, we found that most annotation types could be used for several tasks, specifically, RV and filter. For the other low-level tasks, specifically, CDV, FE, and sort, we observed more targeted use of a subset of annotation types.

There are several important lines of future work. First, we would like to understand which annotation types are most effective. Also, we would like to study how people annotate other chart types, e.g., line charts, scatter plots, pie charts, etc., and validate those results. Finally, often several annotations were used as an ensemble, e.g., mark-color, enclosure-connector-text, connector-text. These combinations occur because individual annotations cannot stand on their own, e.g., the text might need a connector for spacing reasons. Our taxonomy does not directly capture these, but classifying them remains an interesting direction for future work.

Acknowledgments

This work was partially supported by the National Science Foundation (IIS-1845204 and DUE-2216227).

References

[1] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization (InfoVis)*, pp. 111–117, 2005.

[2] S. Collins. Georgia's covid-19 cases aren't declining as quickly as initial data suggested they were. Vox, 2020.

[3] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer. Vega-lite: A grammar of interactive graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):341–350, 2016.

[4] E. R. Tufte. The visual display of quantitative information. *The Journal for Healthcare Quality (JHQ)*, 7(3):15, 1985.