Efficiency of Stream graphs: A Comparison to the Popular Stacked Bar Graph

Holly Echlin & Samantha Roberts

PSYC 6135 Data Visualization
Professor: Dr. Michael Friendly

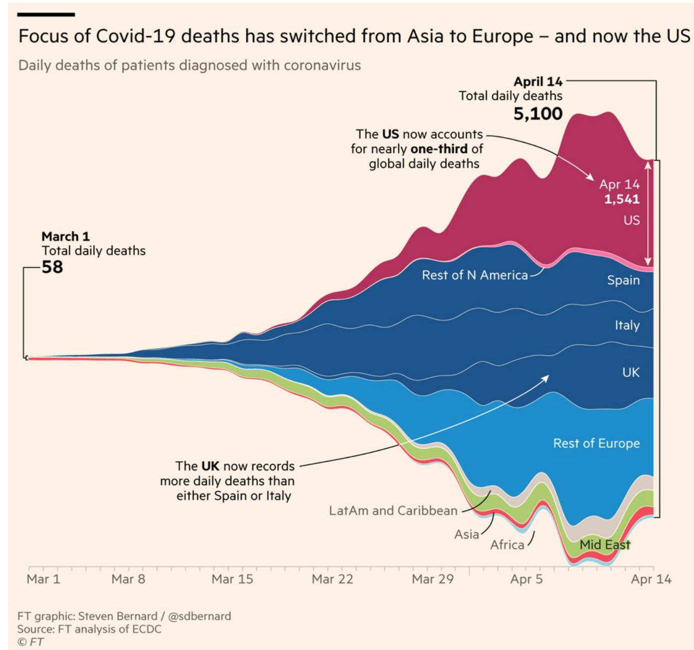
Introduction

There are several different data visualization techniques for depicting information to the public for scientific and other reasons, each with their own benefits and drawbacks. A popular method has been the traditional bar graph and further, the stacked bar graph. The stacked bar graph has become widely used for showing data in scientific communities, as well as on global news platforms in a neat and organized fashion. Despite the popularity, there is growing recognition regarding the readability and transparency of data findings in situations where other graphs may be more appropriate. Emerging in acceptance, the streamgraph provides an alternate form of data visualization with some literature supporting the idea that streamgraphs improve the efficiency and readability of data findings among certain situations of which stacked bar graphs are less equipped. In attempt to understand this further, the current study aims to examine the efficiency of the streamgraph and stacked bar graph, as well as offer an understanding into factors that may contribute to response types of readers.

A streamgraph is a variation of the stacked area chart and the stacked bar chart that alternatively plots all data values (Bartolomeo & Hu, 2016). Streamgraphs were first popularized in 2008, when a New York Times journalist published an article displaying box office revenues of Blockbuster movies (Bloch, Byron, Carter & Cox, 2008). The distinct streamgraph in this article had dozens of small streams representing the seasonal trends and popularity of movies over time. This streamgraph gained traction in the media for its visually appealing look and novelty. Consequently, streamgraphs have begun to flood the internet and news outlets, with particular emphasis on the current coronavirus pandemic. A notable alteration to the streamgraph compared to other graph types, is that rather than plotting all values along a conventional x and y axis, the streamgraph forms asymmetrically around the x axis to create these flowing, organic shapes, known as *streams* (Byron & Wattenberg, 2008). These streams display change over time for set of data points making them visually appealing (Byron & Wattenberg, 2008; Thudt, 2016). Streamgraphs sort individual steams of data over time in a way that smooths the distortions by reducing the data points 'wiggle-factor' (Hochheiser & Shneiderman, 2004); this gives the graphs their asymmetric outer shape. Some researchers have suggested that the reduction in the distortions is what improves that readability over other graph types such as the ThemeRiver graph or the stacked back graph (Hochheiser & Shneiderman, 2004; Byron & Wattenberg, 2008). To date, there is a dearth of research that examines elements of the streamgraph and its readability to the public despite the growing use.

There are several benefits to using streamgraphs for presenting data in the scientific community and the media. The use of streamgraphs allows the authors the ability to plot multiple groups of data overtime, rather than using marginal data points. To the reader, this allows you to see all the data, distribution pattern, magnitude at specific time points, as well as a comparison of the change over time for each of the varying groups. Additionally, as previously stated, one of the major benefits of streamgraphs is that it minimizes the distortions and wiggle in data points that are seen in other graph types and charts (Byron & Wattenberg, 2008). In a comparative lens, this increases the visual appeal and increases the use of streamgraphs in the data visualization community (Thudt et al., 2016). Another benefit of the streamgraph in comparison to the stacked bar graph is that there is an abundance of additional information provided when all data is provided. Stacked bar graphs provide a snapshot of descriptive data at a specific time point (e.g., week 1 and week 2). However, what you don't see is what is happening between that week 1 and week 2 on stacked bar graphs. There can be all kinds of variations to the data between those two time points that may be important for the reader to know. The streamgraph provides this type of information to the reader.

A famous journalist, Steven Bernard, for the Financial Times provided an excellent example of a streamgraph at use (Bernard, 2021). As shown in the picture to the right, you can see that the focus of the streamgraph was to highlight the number of COVID-19 daily deaths by country from March 1st, 2021 to April 14th, 2021. Looking at the streamgraph, you are able to see the inflection or spikes of daily deaths by country over time. This graph provides the reader with the opportunity to see the proportions of individuals passing away from COVID-19 daily within each country over time, as well as draw comparisons between countries. Specifically, you can see how at one point Italy had the most significant number of daily deaths; however, within those six weeks the United States quickly surpassed the other countries. Despite the benefits of streamgraphs, there are several limitations that are important to consider, alike all other data visualization techniques.



Although many people perceive streamgraphs to be easy to read, intuitive, and visually appealing, there are some downsides and limitations to consider. As with any stacked bar chart, they can easily get cluttered if you try and add too many categories at once. In typical practice it is recommended that stacked bar graphs do not surpass more than 3 categories or else the readability of the data becomes exponentially more difficult. Although less so, streamgraphs share this limitation. Streamgraphs are able to capture more categories without reaching the saturation point of readability as quickly; however, the more categories you include (i.e., the more streams there are), the more challenging it is to read and understand these graphs. A solution to this limitation is the addition of interactivity to the graph such as adding numerous labels, legends, and marginal data points that would be captured on other graphs. Streamgraphs can only represent positive values and do not support negative values or a combination of negative and positive values together (Heer & Agrawala, 2006). The reason for this is because there is no clear baseline to differentiate positive and negative data values, and therefore making it difficult to depict data that may hold negative and positive values with equal importance to the story the author is trying to convey (Bartolomeo & Hu, 2016). Another limitation of the streamgraph is that it is challenging to read the scale at a glance and understand what the difference means between lines. This is due to the fact that it lacks a y-axis. Readers are able to view the graph and draw comparisons between categories over time, yet it is exceptionally difficult to put a value to the difference between groups. Furthermore, when the difference between groups or categories is not noticeably different, it can be difficult to differentiate which is larger due to the lack of y-axis. Lastly, streamgraphs are based on the perception of curved slopes and slope ratios (Hochheiser, Shneiderman, 2004; Javed, McDonnel, Elmqvist, 2010;

Munzner, 2014) making them susceptible to perceptual biases such as the *sine illusion*. The sine illusion is a perceptual compromise between the vertical extent and the greater overall dimensions of the wave figure that is being examined (Day & Stecher, 1991), which holds similar principles to the Muller-Lyer illusion. The sine illusion occurs when the human perception leads to systematic bias in the assessment of the optical stimulus being presented. More clearly, when we see waves or *streams* on the streamgraph, our visual system is biased to accidently conclude that the distance between lines at the upper and lower portions that are bending/curved are larger than during the vertical ascending or descending (straighter) sides – when in fact they are the same distance. This is problematic for streamgraphs due to the natural '*eb and flow'* of the streams on the graph. It is possible that the readers may conclude a significant difference when in fact there is not due to the natural perceptual bias, as well as the natural increase in error on area judgment tasks (Cleveland & McGill, 1994; Hollands & Spence, 1998).

The paucity of research that examines streamgraph has focused on the visual appeal and readability of the data in comparison to other graph types, with results being equivocal. In studies that have examined the readability of streamgraphs researchers have concluded that the way streamgraphs smooth distortions of each stream improve the overall readability over similar ThemeRiver graphs but not when compared to stacked bar graphs (Thudt, 2016). Other researchers examined the time-series distortions of graphs and found that the streamgraphs performed better than two other graph types, including the traditional stacked bar graph, and justified that the difference is aligned with the fundamental graphical perception studies (Byron & Wattenberg, 2008; VanderPlaus, 2015; Cleveland & McGill, 1994). Contradicting these findings, another study examined several examples of causal streamgraphs that were published on the web and concluded that streamgraphs are difficult to read due to their uncommon shapes and without interactions to mitigate this problem, the readability significantly declines (Kirk, 2010). Another group also concluded that despite the fact that streamgraphs have been found to be more aesthetically appealing, the distortion of streamgraphs based on the perception of curved slopes and slope ratios have been found to not improve the readability over stacked bar graphs (Bartolomeo & Hu, 2016). To date, it remains unclear whether streamgraphs improve the readability of data and whether they are a reliable data visualization technique. Furthermore, there are no studies that provide information on the efficiency of streamgraphs and the amount of time needed to read a streamgraph. Given that Hollands and Spence proposed an incremental estimation model to account for speed of processing arguing that less overlap in images was related to faster responses (J. G. Hollands & Spence, 2001), it is important to test these theories within the context of streamgraphs. Therefore, the current study aims to provide more clarity surrounding the readability and efficiency of streamgraph with the following objectives.

Aims

1. To understand whether a stream graph or stacked bar graph is more efficient (i.e., based on accuracy and the time participants took to view each graph and respond to prompts) at relaying COVID-19 information to a sample of university students.

2. To evaluate whether other factors contribute to timing and accuracy of responding, including both data- and graph-related variables.

Methods

Participants

The experiment will include healthy young adults (60 female and 60 male) from the York University Research Participant Pool, with an age range of 17 to 25 years. Participants will be eligible to participate in the study if they meet the following inclusion criteria: 1) normal or corrected to normal vision, 2) ability to physically move arms and legs without delay, and 3) no neurological conditions or associated medical conditions impacting their cognition. With their consent, participants' demographic data will be obtained from their participant pool accounts. This will include their age, gender, years of post-secondary education, and number of statistics courses taken. All participants who indicate that they wish to volunteer for the study must provide written informed consent.

Procedure

Participants will sit in a height- and distance-adjustable chair in a room with low lighting so that they can comfortably view the targeted screen areas where graphs will be presented. Participants will begin each session with a training trial which will include two trials consisting of the presentation of a graph. The participants will be presented with the same data in both of these training trials, displayed in two different graph types. Participants will press the spacebar to indicate they are finished viewing each graph and then a separate screen with both graphs will appear along with a question (i.e., "Do these graphs show the same or different data?") and a true/false answer type. Participants will be provided with corrective feedback if needed and will be given one-minute debrief about the main differences between the two graphs.

After the training, participants will be presented with a question on the screen about the upcoming graph which they will indicate they have read by pressing the spacebar. The instructions on this page will also indicate that participants should answer as quickly and accurately as possible. After participants press the spacebar, the graph will be displayed on the screen with a series of five multiple choice questions directly below the graph and the question from the previous screen appearing above the graph. Each graph will be displayed for a maximum of 30 seconds and will contain real data published on government websites about global numbers of COVID-19-positive individuals. Participants will be asked to select an answer from the multiple choice (i.e., five possible answers with one response indicating "I don't know") about the graph after each individual graph is shown. The questions will pertain to information derived from viewing the graph. Some sample questions include, "which country had the most COVID-19-positive tests on April 14th?", "which country had the third highest number of COVID-19-positive adults in this graph?", or "which country showed the greatest decrease in CVOID-19 cases between June and August?"

Two graph types will be used for comparison (i.e., a stream graph and a stacked bar graph). An example of a stream graph containing COVID-19 data is displayed in figure 1 on the left and the same data is displayed in a stacked bar graph on the right.

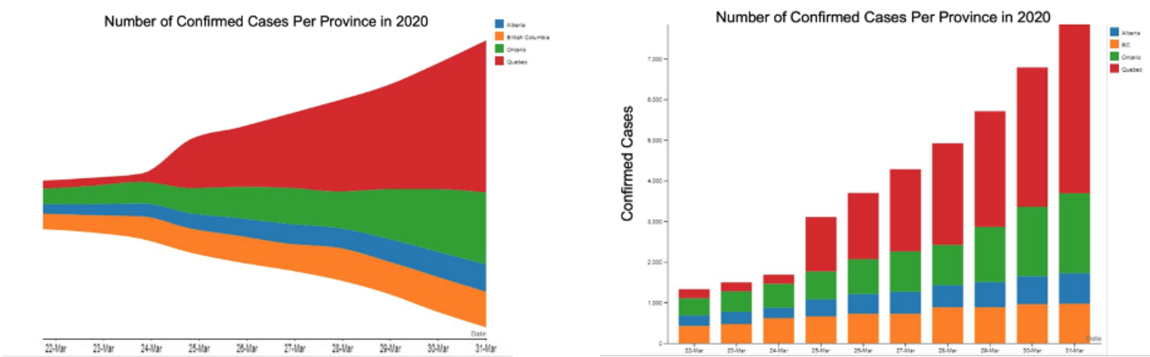


Figure 1. Sample containing COVID-19 cases over a week across four provinces depicted in a stream graph on the left and stacked bar graph on the right.

There will be 60 trials in total (i.e., 30 stream graphs and 30 stacked bar graphs) and each stream graph will have a corresponding stacked bar graph showing the same data, which will be presented in sequential order. The order of trials will be counterbalanced such that participants will either be in group A or group B so that one group will be presented with one graph type first and the other group will be presented with the opposite graph types first, showing the same data. A sample of the order of trial presentation is depicted in table 1.

Table 1

Order of Counterbalancing for Both Participant Groups

Group A	Stacked bar (1st question)	Stream graph (1st question)	Stacked bar (2nd question)	Stream graph (2nd question)	Stacked bar (3rd question)
Group B	Stream graph (1st question)	Stacked bar (1st question)	Stream graph (2nd question)	Stacked bar (2nd question)	Stream graph (3rd question)

Note. The order continues on for 60 trials including and beyond the order shown here.

Several different question types will be examined to understand whether there are instances wherein one type of graph is more efficient than the other. These include data-related factors and graph-related factors. The data-related factors that will be explored will include the slope, magnitude, the types of data (e.g., continuous or discrete) and the relationships between groups. The graph-related factors will include perceptual factors specific to the graph type as well as the number of groups depicted. Ten trials will be used to examine each of these specific question types (e.g., ten trials specific to the slope will be given, with five shown as stacked bar graphs and five stream graphs), totaling 60 trials.

Data Analysis

Results will be analyzed based on participants' average time to respond to each question and accuracy (i.e., correct or incorrect response). Across the questions, mean response times will be calculated for each participant, as well as intra-individual variability (IIV) using the individual standard deviations (ISD; Wojtowicz et al., 2014) to measure stability of performance across questions. Regression analyses will be conducted in order to predict the outcome variables of interest. It is anticipated that in the majority of cases, participants may answer the questions accurately, and thus reaction time (i.e., analyzed using linear regression models) will be the primary variable of interest and accuracy may be used to determine validity. However, for more challenging questions, logistic regression analyses with be used to predict correct versus incorrect responses while accounting for the type of graph, age, sex, and familiarity with statistics (i.e., this will be quantified according to the number of statistics courses taken). Some of the independent variables of interest include the graph type, question type, as well as students' familiarity with statistics. Age and sex will also be controlled for in all analyses. Significant findings will be followed up with appropriate post hoc analyses. All data analyses will be conducted with Statistical Package for the Social Sciences (SPSS) version 24 (IBM corp., 2016). For all analyses, two-tailed alpha values of less than .05 will be considered statistically significant.

Implication of Results

Given the rise in popularity and dearth of research surrounding streamgraphs, the results from the proposed study will allow for a better understanding of the efficiency and accuracy of reading streamgraphs in the community. Specifically, it will allow us to understand how accurate readers are drawing information from the streamgraph in comparison to the stacked bar graph providing clarity on the current equivocal literature. Furthermore, the results from the study will allow us to understand how quickly readers are able to look at the graph and pull the important information in comparison to the more familiar stacked bar graph. These results will directly impact the use of streamgraphs within literature and the media. It is particularly important to understand these aspects of streamgraphs given the importance of accurate and fast information – especially when they are being used frequently in the media for important topics such as the COVID-19 pandemic across the globe.

References

Bartolomeo, Marco Di; Hu, Yifan (2016). "There is More to Streamgraphs than Movies: Better Aesthetics via Ordering and Lassoing". *Computer Graphics Forum.* ***35*** (3): 341–350. doi:10.1111/cgf.12910. ISSN 1467-8659. S2CID 17724977.

Bernard, S. (2021). Daily deaths of patients diagnosed with coronavirus. https://www.ft.com/content/a2901ce8-5eb7-4633-b89c-cbdf5b386938

Bloch, M., Byron, L., Carter, S., & Cox, A. (2008). The Ebb and Flow of Movies - Box Office Receipts 1986 — 2008. *The New York Times*. Retrieved from http://archive.nytimes.com/www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531–554. https://doi.org/10.1080/01621459.1984.10478080

Day, R.H., Stecher, E.J. (1991). Sine of an illusion. *Perception* 20:49–55

Heer, J. and Agrawala, M. (2006) Multi-Scale Banking to 45 Degrees. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 12(5). Sep/Oct 2006.

Hochheiser, H., Shneiderman, B., (2004) Dynamic Query Tools for Time Series Data Sets, Timebox Widgets for Interactive Exploration, *Information Visualization 3, 1*

Hollands, J. G., & Spence, I. (2001). The Discrimination of Graphical Elements. *Applied Cognitive Psychology*, *15*(4), 413–431. https://doi.org/10.1002/acp.714

Hollands, J., & Spence, I. (1998). Judging proportion with graphs: The summation model. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *12*(2), 173–190. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1099-0720(199804)12:2%3C173::AID-ACP499%3E3.0.CO;2-K

Javed, W., McDonnel, B., Elmqvist, N. (2010). Graphical perception of multiple time series. *IEEE Transactions of Visualization and Computer Graphics, 16*(6), 927-934.

Kirk, A. (2010). Making sense of streamgraphs. *visualizingdata.com*. Retrieved from http://www.visualisingdata.com/index.php/2010/08/making–sense–of–streamgraphs/

Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press. Retrieved from https://books.google.ca/books?hl=en&lr=&id=NfkYCwAAQBAJ&oi=fnd&pg=PP1&dq=munzner+2015+Visualization+Analysis+%26+Design&ots=ZA1jRD6kj5&sig=G3-

Aw3UW4IanP6mlzVt3aEEAG_A#v=onepage&q=munzner 2015 Visualization Analysis %26 Design&f=false

Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, *82*(398), 454–465. https://doi.org/10.1080/01621459.1987.10478448

Thudt, A., Walny, J., Perin, C., Rajabiyazdi, F., Macdonald, L., Vardeleon, D., … Vardeleon, R. (2016). Assessing the Readability of Stacked Graphs. https://doi.org/10.20380/GI2016.21ï

Wojtowicz, M. A., Ishigami, Y., Mazerolle, E. L., & Fisk, J. D. (2014). Stability of intraindividual variability as a marker of neurologic dysfunction in relapsing remitting multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, *36*(5), 455–463.

VanderPlas, Susan Ruth, "Perception in statistical graphics" (2015). Graduate Theses and Dissertations. 14438. https://lib.dr.iastate.edu/etd/14438