# Journal Paper Summary (25 points total)

| | |
|---|---|
| Paper Title | Infovis and Statistical Graphics: Different Goals, Different Looks |
| Author Name(s) | Andrew Gelman, Antony Unwin |
| Student Name | Luna McBride |
| Student ID | 107607144 |

1. What do you think the paper is about in layman's terms? What are the main arguments presented and any recommendations/suggestions about the tension and need for collaboration between practitioners in the fields of statistical graphics and information visualization? Do you agree with the recommendations/suggestions? Why or why not? [5 points]

| Focus of this question |
|---|
| • This question encourages you to evaluate the arguments, evidence, assumptions, and conclusions about key issues (i.e. think critically about the paper)<br>• This question encourages you to develop your own knowledge, comprehension and conceptual understanding and to connect, synthesize, and/or transform your ideas into a new form (i.e. be a creative thinker and contribute your ideas and thoughts)<br>• This question challenges you to consider diverse perspectives (gender, political, ethnic, racial, etc) during class or in assignments |

Write your answer below:

This paper discusses the prevalence of flashy yet difficult to read visualizations in the public sector, pointing to some pretty nasty ones in the process. It also discusses how clear cut yet boring to look at the normal statistical models are. The authors propose the idea that both diverging paths, yet are also inherently necessary. They state that the flashy visualizations draw attention to the concept while the statistical graphs provide the actual substance. I can agree with this notion to an extent, but it is quite naïve given the current state of culture.

To be fair to the authors, this was written in 2012. This would put the article in the lull between the flashy blogs and experiments of the 2000's and the current trend of making everything sanitized just so advertisers can bombard our senses. The internet has come to

a point where the flashy is becoming a negative, since it is being linked with advertisers and soulless companies targeting the iPad kids of generation alpha. Even statistical models in the space are trading information for being flashy, which is worse when misinformation has run rampant in the modern era. A boring looking graph with fewer issues to induce sensory overload is a breath of fresh air given the current climate, and may have some of the same alluring power that the ultra-flashy graphs had back then.

2. The authors present a discussion of the key differences between the two approaches of information visualization and statistical graphics. Think about these differences, suggest and discuss four real-world applications of one key difference between information visualization and statistical graphics. [5 points]

| Focus of this question |
| --- |
| • This question encourages you to connect your learning to "real world" issues or life experiences and consider diverse perspectives for the application of concepts in the paper to the real-world<br>• This question encourages you to develop your own knowledge, comprehension and conceptual understanding and to connect, synthesize, and/or transform your ideas into a new form (i.e. be a creative thinker and contribute your ideas and thoughts) |

Write your answer below:

- Consider the only major infographic shown in the 50-chart challenge (number 46). It is not fair to use the same visualization techniques as the others, since making flags, berries, and meats all the same color in different hues would not work. It says numbers, yes, but the sources or dates the numbers were recorded are not heavily scrutinized. Also, many people would not question why lamb specifically was chosen instead of overall meats to represent the amount imported. This works with the argument of aesthetic versus data display, as many people would take the infographic at face value while looking for errors in the average bar chart.
- Propaganda has been taking the style over function motto and running with it for years. This can take the form of removing context like with certain historical maps being used, modern day memes imparting modern western values on non-western people or historical figures, or even posters that mean nothing and are only meant to boost morale, like with Rosie the Riveter. The propaganda game is all about utilizing dubious data storytelling methods and emotional responses to move public opinion in one direction. It takes the idea of form over function in visualizations and weaponizes it for specific agendas. Statistical graphs may try to do this too, but they are much more likely to be called out on it due some of the terrible methods to skew such visualizations being pretty clear to see and call out (starting a bar chart not at 0, making some bars abnormally high for their value, etc.)
- Smogon, a major competitive Pokémon battling format, releases a graphic on which Pokémon have moved up and down tiers each month. These are interesting to look at in a void as someone not in competitive Pokémon, but they do not tell an outsider much about the format or strategies that changed in order to cause these changes. They do not show how usage has grown or shrank for certain key threats to cause

the domain shift. While they look good, at the end of the day, this makes them have no actual substance to them. March 2024 version: (https://twitter.com/SmogonU/status/1763595048657252353)

- The Pikmin series is actually very good at utilizing visualizations at the end of each day of gameplay. They usually have both charm to fit the world of the game and function well to show how you did on a given day. Some graphs, however, like the Pikmin 2 Pikmin count change graph, provide a pretty look for minimal function. The simple counts at the end of the day numbers from other titles do an excellent job using numbers to show where you stand.

3. Choose two “Discovery goals” and two “Communication goals” (See pages 8 – 9) of the paper and discuss at least three (3) implications for the visual display of quantitative information not mentioned in the paper. [15 points]

| Focus of this question |
| --- |
| • This question encourages you to connect your learning to “real world” issues or life experiences and consider diverse perspectives for the application of concepts in the paper to the real-world<br>• This question encourages you to reflect on what you are learning<br>• This question encourages you to contribute your ideas and thoughts<br>• This question challenges you to consider diverse perspectives (gender, political, ethnic, racial, etc) during class or in assignments<br>• This question challenges you to develop and present your own knowledge, comprehension, and conceptual understanding |

Write your answer below:

Discovery:

- Getting an Overview
    - The paper discusses this sort of exploration as having only yourself as an audience, but does not discuss how this information needs to be pulled into presentations at some point. The simple play done in previous courses in computer science is to simply copy them into the presentation, but that does not communicate the expected outcome correctly for all audiences. Doing the overview stage with the required presentation scrutiny would not work either, as that is just a whole lot of work for many graphs that the stakeholder will not see. As such, an intermediate step should exist like how a pipeline exists in data engineering.
    - It is almost impossible to gain an overview from categorical variables with many entries. At a certain point, labels in bar charts and pie charts devolve into a black, messy block. This means that the column should not be visualized in the first place, but instead used as a filtering tool, bucketized further if given a key, or removed altogether.
    - Means, medians, and similar metrics can be a good starting point, but they should not be the be all end all. Means especially can be shifted by outliers, so these values need to represent a starting point instead of leading the charge.
- Exploration
    - The authors had such high hopes for interactive data. If it cannot be made intuitive enough, the user is bound to have some frustration given the freedom to explore the data themselves in an interactive setting.

There are very few occasions where users should be trusted with that, if the speedrunning scene in gaming has shown us anything. They may even break your product *for fun*.
    - If anything, the point of exploration should be to lead into explanation and storytelling. It is one of the key points of using Jupyter Notebooks in the modern era. Considering these ideas separate undersells exploration as a whole.
    - Simple statistical methods need to be used in these early stages of discovery rather than the flashy versions that can commonly be used in this step. If the selected graph is known for being harder to read (like with pie charts and angles,) the explanation is likely to miss out on some interesting aspect that would not be found otherwise.

Communication:

- Communication to Self and Others
    - It may be difficult to accept that you do not process things like others do. My case stems more from medical disorders and sensory processing disorders, but even age or research domain could change how you would look at something compared to anyone else in the room.
    - You may not have direct access to stakeholders or your audience before any presentation. In such a case, certain safety measures should be considered to reduce the risk of losing the audience due to common ailments. These can include single colors/blue to account for red/green color blindness, keeping a gridline in a bar chart to help those who process sensory inputs differently, or removing pie charts since all people are not the best at judging angles. It really depends on the data and your model. Though, if you can ask, ask.
    - “They’ll know what I mean” you say, but no they will not. They have not seen the data you have, and they do not have the background you have. If you have even the slightest doubt, do not lean on the assumption of knowledge.
- Attracting Attention and Stimulating Interest
    - The very existence of a chart amongst text draws attention toward it, as it is something different compared to the text around it. The assumption that it needs to have many red clickbait arrows and circles around it is simply overbearing. Of course the clickbait circles and arrows are an exaggeration in this context to represent flashiness, but the comparison should show the issue being caused.
    - A modern news article risks a flashy graph blending in with the background noise of the constant ads on the sides. If anything, a more

subdued graph could net more attention since it is more clearly not part of the flashy ad bombardment around it.

- The use of noise reduction strategies and pre-attentive attributes in the graph should be relied on to make it less intimidating rather than using flashiness to draw eyes in. This article may have been made before the modern internet advertising landscape got as bad as it has, but even at that time it should have been clear that less is more.