Design Principles for Data Visualization

Jennifer Smith

February 10, 2022

For this research paper, I chose to research how design principles apply in data visualization, as I am working towards my Master of Science in Business Analytics and have a special interest in data visualization. I took a deep dive into visualization design best practices to learn the dos and don'ts of the field. This is a broad topic, and much is beyond the scope of this assignment, but I learned many valuable lessons and summarize some high-level guidelines below.

Chart Types: Choose Carefully

The first important aspect of designing a visualization for a dataset is selecting the right chart for the message you aim to convey. Countless types of charts and graphs exist for visualizing data, and there are many resources available online to help determine the best type of visualization for the dataset at hand. A great place to start is the Financial Times Visual Vocabulary chart, which displays thumbnails of many types of charts organized in columns for the various dataset attributes you want to communicate, such as deviation, ranking, change over time, magnitude, etc.

It is important to understand that information can be either very clear or very difficult to comprehend depending on how it is presented visually. For example, consider the example shown in Figure 1 from Evergreen Data (*The Most Important Question)*.

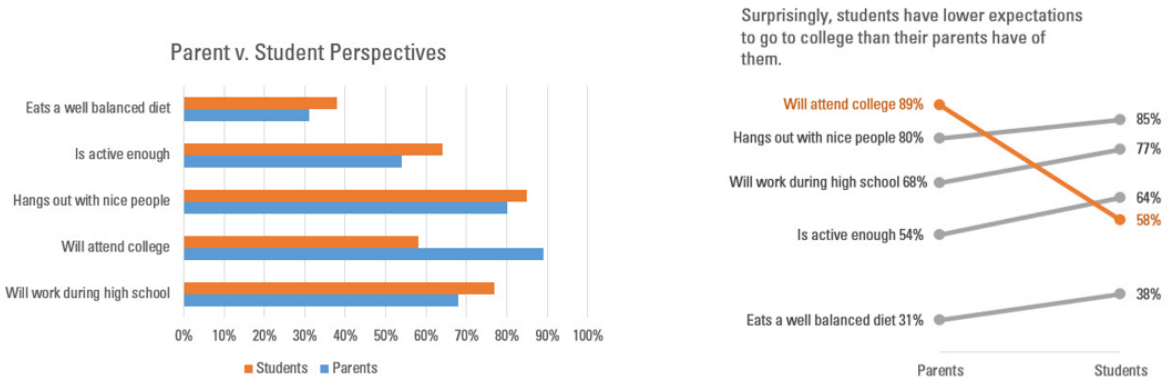

Figure 1

The charts in Figure 1 show the same data in two different visualizations. In the bar chart on the left, it's difficult to tell if there is an important takeaway in the data. If we studied it long enough, we could find it, but visualization should make data easier to understand, and this example does not. In the slope graph image on the right, however, it's easy to see that the category highlighted in orange is following a trend that is opposite to the rest of the categories. The message is clear in this visualization because it contains a great deal of contrast, both in the directions of the lines and in the single line of orange against the gray. Our eyes immediately pick up on the message, so we don't need to spend time analyzing it.

A visualization's design must help our eyes to quickly see the message in the data. If the image is too busy, our brains can't distinguish what is important. Figure 2 below shows another example from Evergreen Data that demonstrates how showing data with less visual clutter makes it easier to understand (*Declutter Dataviz*).

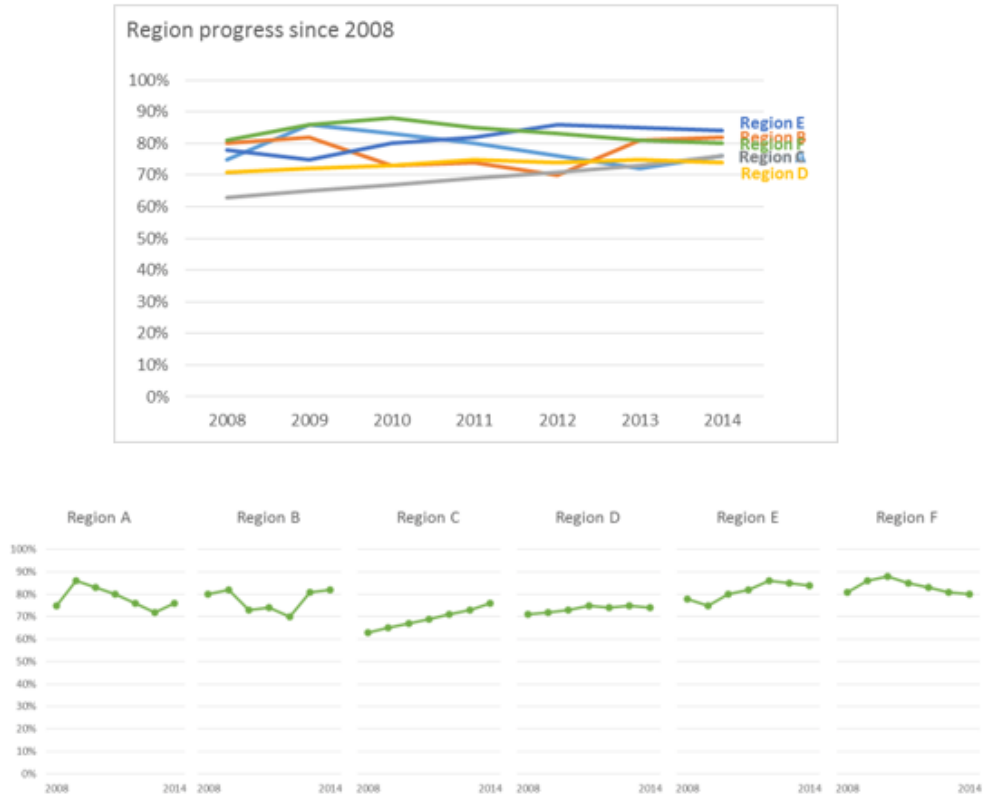

Figure 2

In the top line graph in Figure 2, there are so many overlapping lines so close together, our eyes can't see trends for any of them. But when the same data is broken apart into several small charts that are close together and share the same scale, as in the bottom image, it is very easy for us to see the trend of each region. This technique makes use of several design principles to communicate the data. Proximity lets us know that all of the data is related, alignment helps us understand that the x-axis on the left applies to all of the charts, and repetition of the same chart design allows us to quickly comprehend the trend of each region.

One piece of advice I saw many times in my research was to be wary of pie charts because the human eye is not good at comparing the size of two shapes that aren't aligned. For this reason, bar charts are nearly always superior to pie charts. Figure 3 below illustrates this nicely with a side-by-side example.

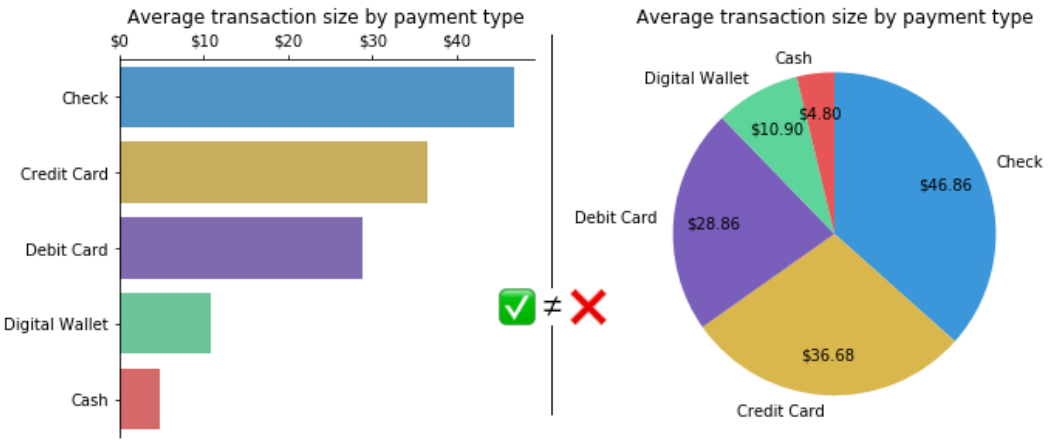

Figure 3

In the pie chart on the right, it's difficult to tell the size difference between the tan credit card and purple debit card pieces. In the bar chart on the left, however, our eye quickly understands the size difference because they share alignment on the left (David, 2020). For comparative visualizations, it's nearly always better to use a bar chart than a pie chart.

Color: Be Intentional and Cautious

Once we have our data in visual form, perhaps the next logical step is to consider how we should add color. Some organizations may have detailed brand guidelines that require the use of brand colors within data visualizations (Nightingale has a great post about this) but if no brand guidelines exist, then similar to selecting the type of chart, we must consider our specific dataset and the main point of our visualization when selecting our color palette. Theresa-Marie Rhyne explains in an episode of The Present Beyond Measure Show (Pica, 2021) that we would choose different palettes for sequential, diverging, or categorical data. If we choose the wrong one, we could inadvertently show a story or connection in the data that isn't there. On the other hand, color can be a very effective tool to thoughtfully highlight key pieces of information that we want to convey to our audience. In Figure 4 below, I created a set of simple charts to demonstrate some of these points.

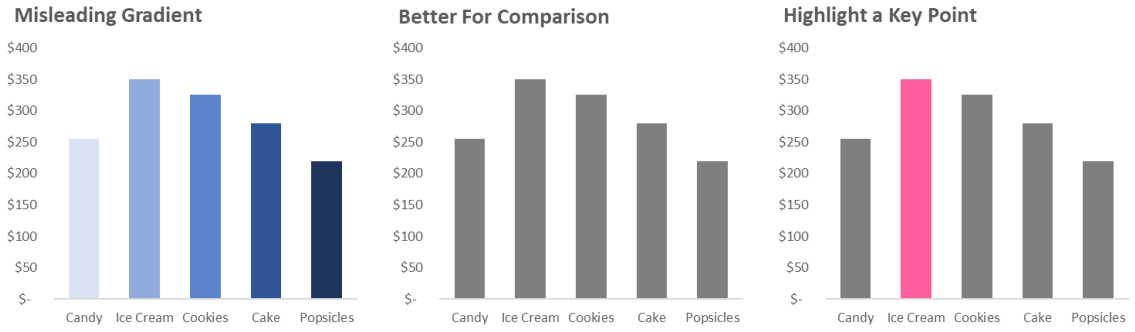

Figure 4

In the image above, the first chart illustrates how a color scheme could be misleading if used incorrectly. In this example, a gradient scheme suggests that the data is somehow intensifying as it moves from left to right when in reality each bar is simply a separate category. The middle chart shows a better option, illustrating how a uniform color across the categories communicates more clearly that the information being shown is simply the same measure of different categories. The right chart is better still, showing how color can be used to highlight a key point you want to make clear, in this case, the fact that ice cream sales are the highest among dessert categories.

One thing to be mindful of when choosing a color scheme for data visualization is how the image will appear to someone with color vision deficiency (CVD). If the wrong colors are used, some visualizations are meaningless to someone with CVD. Likely the most common example of this in business settings is the use of red-green-yellow to signify bad-good-risk. An example of this can be seen below in Figure 5, a snapshot from Evergreen Data's post *Friends Don't Let Friends Use Stoplight Color Schemes.*

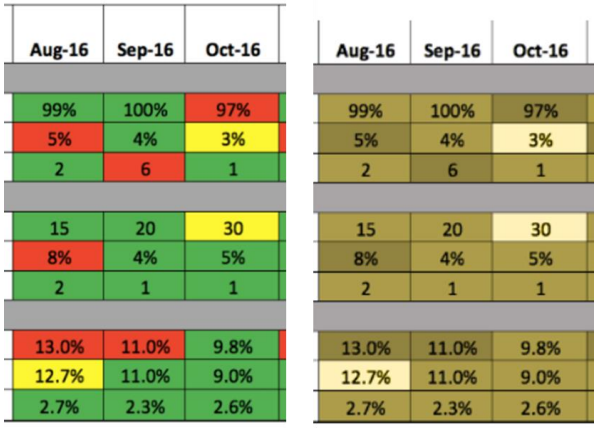

Figure 5

The red-green-yellow image on the left would be clearly understood by many businesspeople, but we must consider how it might look to someone with CVD, which is simulated in the right image. Any message that the red-green-yellow scheme was intended to communicate is completely lost. To make our visualization accessible, we must use a palette that is still distinguishable to someone with CVD. There are several tools available online that can help with this, such as http://www.vischeck.com/.

We should also consider how a data visualization will look if it is printed in black and white. Even in today's digital age, we can't assume that our visualization will only be viewed digitally in full color. It is best to use a color scheme with varying values so the intended message is still clear if viewed in black and white, and it should be noted that any text included in the visualization should also be legible if printed in black and white. This is easy to check by simply looking at the print preview in grayscale.

Typography: Enhance Clarity and Comprehension

Evergreen Data has a very useful visualization checklist that includes a section on text (*Checklist*). Evergreen lists practical advice on using text such as positioning the title in the upper left corner, because that's how western cultures are accustomed to reading. She advises that a clear hierarchy should be established with font size so that the title stands out the most, followed by the subtitle, then labels, then axis labels. She also recommends that all text be horizontal and that labels be placed near the data rather than in a separate legend. These guidelines all serve to make a visualization clean and easy to understand.

As for fonts to use in data visualization, best practices suggest selecting a font that has a large x-height and a shorter line length (i.e., condensed), but that still possesses a stable, open counter (the open space in letters such as "a" and "p") (France, 2020). These attributes help a font be readable at small sizes, which is very important for data visualization, as most fonts will be displayed quite small. Also for this reason, and because most visualizations are viewed digitally, sans serif fonts are also most commonly recommended.

Numbers are another important consideration when selecting a font for use in data visualization. Typeface numbers are either tabular (each digit takes up the same amount of space) or proportional (space per digit varies to look best when used with text). Tabular numbers are easier to read in visualizations and are designed to work with data (France, 2020), though we must be aware that there are differences even among tabular fonts. Some tabular fonts center the digit in the columns, whereas others offset them. The offset tabular fonts create more even spacing and are therefore easier for our brains to understand. See Figure 6 below (France, 2020).

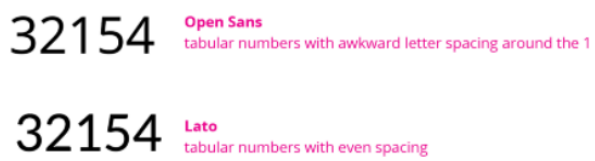

Figure 6

A final recommendation for numbers is to avoid fonts that feature old-style numbers, where some of the digits descend below the baseline (see Figure 7 below). These fonts are harder to read in data visualization and should not be used.

Figure 7

Other Design Elements: Keep it Clean

Finally, after the larger considerations of chart type, color scheme, and typography, there are additional design guidelines to consider for data visualizations that enhance the clarity of a message. Perhaps the most succinct way I found these aspects explained was simply to "avoid chart junk" (*Data: Four Principles),* meaning unnecessary graphics, labels, colors, backgrounds, or other design flourishes that don't support the key points of the visualization.

Evergreen Data provides some practical advice for accomplishing this, including never using 3-D or other visual elements that might distort or obscure data, deleting borders around charts, and avoiding gridlines or tick marks if possible (or keeping them faint if you must include them) (*Checklist*).

Summary

I learned a great deal in the process of researching design principles for data visualization and amassed a large catalog of resources in the process. I can now see how prevalent poor choices in data visualization are, with common violations including not putting enough thought into color schemes (especially the frequent use of red and green), using a legend instead of directly labeling data, and including unnecessary borders and gridlines on charts. Simply adhering to a few simple rules when selecting chart type, color, typography, and design elements can help data visualizations communicate key points as clearly as possible.

Resources

Cesal, A. (2019, July 19). What are data visualization style guidelines? *Nightingale.* **https://medium.com/nightingale/style-guidelines-92ebe166addc**

Data: Four principles for better data design and infographics. (n.d.). Opus. Retrieved February 10, 2022, from **https://opusdesign.us/wordcount/data-design/**

David, M. (2020, October 15). 5 data visualization best practices: The secrets behind easily digestible visualizations. *Chartio.* **https://chartio.com/learn/business-intelligence/5-data-visualization-best-practices/**

Evergreen, S. (n.d.). *Data Visualization Checklist.* Evergreen Data. **https://stephanieevergreen.com/data-visualization-checklist/**

Evergreen, S. (n.d.). *Declutter dataviz with small multiples.* Evergreen Data. **http://stephanieevergreen.com/declutter-dataviz-with-small-multiples/**

Evergreen, S. (n.d.). *Friends don't let friends use stoplight color schemes.* Evergreen Data. **https://stephanieevergreen.com/friends-dont-let-friends-use-stoplight-colorschemes/**

Evergreen, S. (n.d.). *The most important question in data visualization.* Evergreen Data. **https://stephanieevergreen.com/most-important-question/**

Financial Times. (n.d.). *Visual vocabulary: Designing with data.* Retrieved February 9, 2022 from **https://github.com/Financial-Times/chart-doctor/tree/main/visual-vocabulary#readme**

France, T. (2020, June 8). Choosing fonts for your data visualization. *Nightingale.* **Choosing Fonts for Your Data Visualization | by Tiffany France | Nightingale | Medium**

Pica, L. (Host). (2021, September 16). Data visualization color best practices with Theresa-Marie Rhyne (No. 70) [Audio podcast episode]. In *The present beyond measure show.* The Present Beyond Measure Show: Data Storytelling, Presentation & Visualization.