**HW 1: Lessons Learned from Misleading Graphs** <https://www.statisticshowto.com/misleading-graphs/>

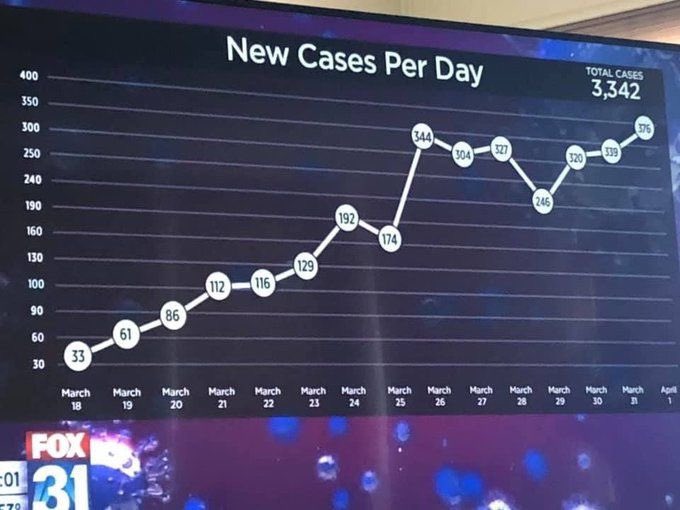
Write a reflective essay about the most common mistakes made in data visualizations. You can use the article linked in this assignment as a guide. However, also contribute an example you find from any source. It could be your own past attempt at a data visualization that went poorly or an example you find from another source. Be certain to provide links and citations when relevant. Expected length: ½ page to 1 page of writing and extra space for any graphics.

Ari Kassin-Fuentes - My essay

The type of data visualization mistakes discussed in the article are often seen in the explanatory phase of a project, i.e., when a person is trying to present an idea and a graph is used to support it. These errors, whether on purpose or not, can mislead the audience into an idea that is not supported by the data, moreover, even if the data is displayed at some extend and reflects a specific point of view, the graph can actually produce the opposite point of view if dimensions, labels and other properties are misused. This is a very serious phenomena because audiences don’t often focus in the whole story while listening to it. Often, a person listens to the first sentences to know what the story is about and then it sees a visualization to try to come up with a conclusion. If the graph misrepresents the data, the visualization part is more likely to make an impact on a person even if the storyteller talks about data. This is a quick and wrongful technique to persuade audiences to a story even if there is no data to back it up.

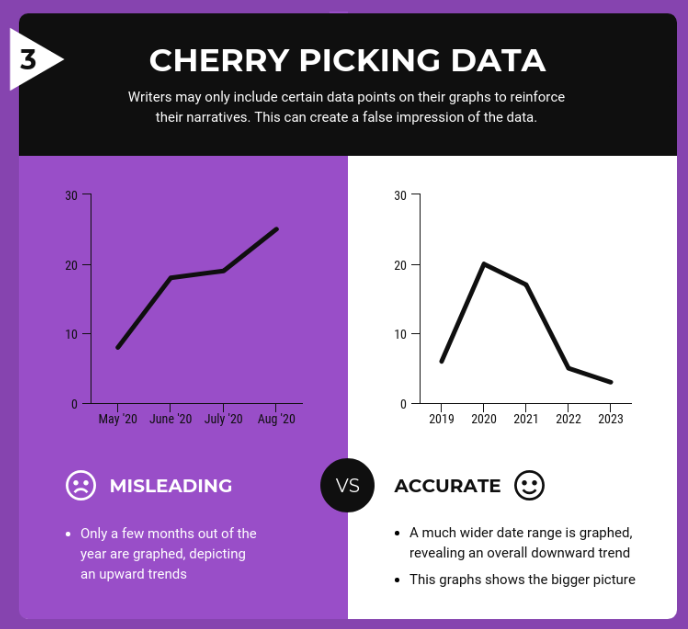
According to [1] The most common errors are caused by 1) misuse of visual proximity, 2) data manipulation and 3) incorrect/missing labels. Misuse of visual proximity.

An example of misuse of visual proximity [2] is shown below:



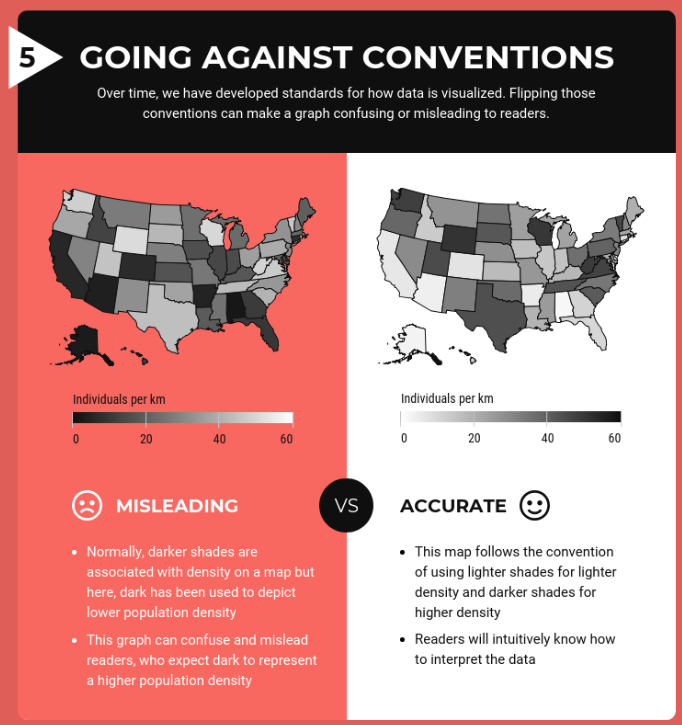
Note the y-axis. It doesn’t start at zero and has uneven intervals. The graph is manipulated to have a flattened looked but the reality shows that the line is moving upwards.

An example of data manipulation [3] can be shown in the next graph:



Hiding portions of a dataset or not fully explaining how it was collected or analyzed qualifies as data manipulation.

Below is an example of incorrect labels [4]. In a heatmap, darker shades mean high density or a high value and lighter shades low density and low numbers however, the maps has this convention backwards and can lead the audience to make a wrong conclusion.



In conclusion, data visualization needs to be treated with as much rigor as any other phase in the research methodology as it is not only useful to present a story but it has the potential to mislead an audience whether on purpose or not.

Reference

[1] Doan, S. (2021). Misrepresenting COVID-19: Lying With Charts During the Second Golden Age of Data Design. Journal of Business and Technical Communication, 35(1), 73–79. <https://doi.org/10.1177/1050651920958392>

[2] <https://www.datasciencecentral.com/the-worst-covid-19-misleading-graphs/>

[3] <https://venngage.com/blog/misleading-graphs/>

[4] <https://venngage.com/blog/misleading-graphs/>