**STAT515 Mid-Project: Incorrect Data Visualization and Redesign**

**Group #2**

**Mykola Signayevskyy**

**Zhang Huairuo**

**Zeeshan Khalid**

**George Mason University**

**Professor Dassanayake**

**March 9, 2023**

**Introduction**

A vital technique for conveying complex information in an understandable and visible way is data visualization. Data visualizations, however, are not all made equal. Inappropriate or poorly constructed visualizations can be deceptive, ambiguous, or even outright misleading, resulting in wrong inferences and decisions. Graphs are one of the most frequent causes of poor visualization.

Although popular and effective, graphs can be misleading and confusing if they are not used properly. An improper interpretation of the data can be produced by a badly constructed graph, which can result in erroneous conclusions. Such misunderstandings can have major repercussions, particularly in industries like public policy, finance, or medicine.

The good examples of poor created graphs that can lead to giant public disinformation are old FOX TV channel graphs. (FOX News, 2013).

Graphical user interface

Description automatically generated

The graph represents the tax rate of two different dates. “This is a terrible Fox News graphs, including one that tried to show the top marginal tax rate exploding under former President Barack Obama, even though the actual increase was from 35 percent to 39.6 percent.” (Reed, 2021). The main problem with the graph is that the Y scale was not taken from zero but from a value of 34% and ends at 42%. Such a scale distorts the visualization and shows a small difference of a few percent as a huge gain.

Figure 1

A picture containing text, scoreboard

Description automatically generatedIn 2019, FOX News channel showed the graph (FOX News, 2019) that represents the percentage of Americans that identifying as Christians. “The graph only showed a modest decline from 77 percent to 65 percent, even though Carlson's graph made the decline appear far steeper than it was.” (Reed, 2021). The examples above are only a drop in an ocean of incorrect charts that manipulate data.

The purpose of this project is to examine several incorrectly created graphs that manipulate data through improper visualization and propose alternatives to visualize the incorrect graphs.

Figure 2

**Redesign and Methodology**

***Average Female Height***

The first graph (Hooper, 2021) we will look at shows the average growth of the female population of the 5 countries. The graph is remarkable in that it can be misleading to young and incompetent people who may indeed be influenced by an incorrect portrayal of growth and have a false impression of people of different races.

Chart, diagram, histogram

Description automatically generated

In this graph, the Y-axis data is so distorted that the graph shows that females from India are several times shorter than females from Latvia. Despite the nice colors and good design, the Y-axis was shown from 5'0 to 5'7 feet, which led to the wrong average height displaying. We redesigned the original graph into 3 different options.

Figure 3

The first redesign shows how the author of the original graph could avoid the mistakes and create appropriate graph.

Chart, bar chart, histogram

Description automatically generatedThe graph was created with the R “ggplot2” package. The chart above shows the real difference in average height of females from the countries. The dataset was created with the information represented on the original graph to show how simple and nice the graph could be.

Figure 4

The second redesign shows the possibility of the R package “leaflet”. The library allows us to visualize the map of the world to show much more than just numbers.

Map

Description automatically generated

Figure 5

The created graph is interactive and shows the location of the countries from the original graph. The markers show the names of each country while popups show the average female height in the countries in centimeters to make graph more understandable for the rest of the world. The graph has a widow for improvement; show the average height of most countries, for example. Indeed, we did not want to overwhelm the graph with information.

However, the map above gives a clear understanding that the original graph took a country from each continent to show the average female height except North America. We want to show that R packages are accessible way to show the verity of designs and opportunities to show more insights than it could seems firstly.

Chart, box and whisker chart

Description automatically generated

Figure 6

The interactive boxplot was created to show the average female height not just in a few countries, but in each region of the world. The chart represents the average height, maximum and minimum average height of countries in each region. Moreover, the dots on the graph show the density of countries in each height category.

***Average Number of Actual Weekly Hours***

The disadvantage of Figure 6 is that the starting point of the X-axis is not zero. The first impression given by this image is that German workers are more motivated and work longer hours than workers in other EU countries.

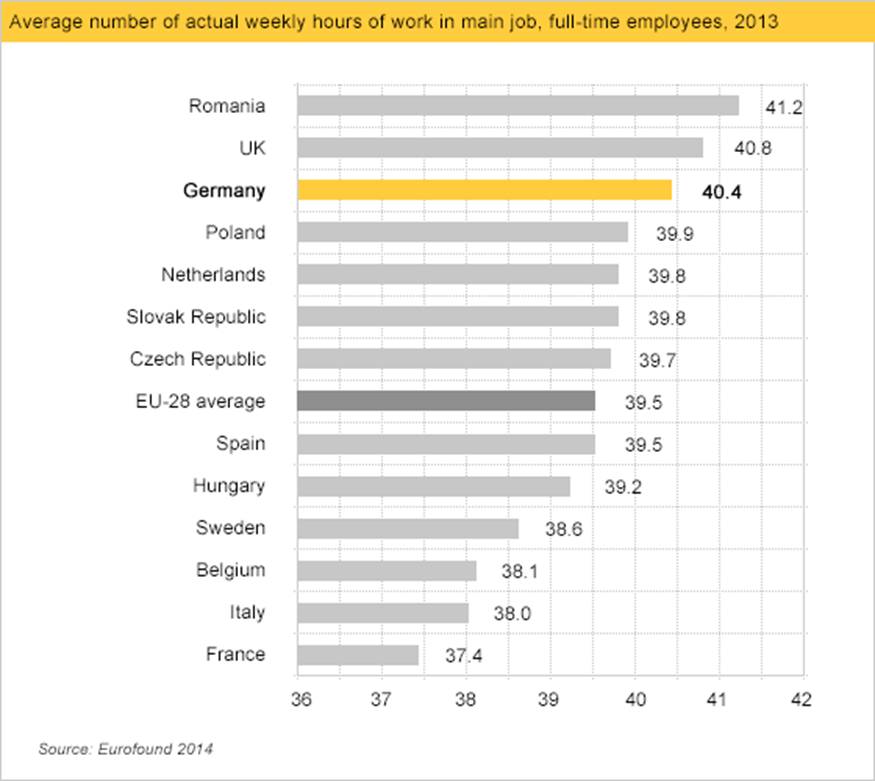


Figure 7

(Bergstrom & West, 2020)

Bar chart

Description automatically generated

Figure 8 is plotted by ggplot2 in R. Figure 7 completes the missing x-axis origin of Figure 6 and colors Germany with red. The countries with the longest average working hours are shown in blue, and the countries with the shortest working hours are represented in gray. The average working time of all EU members is shown in yellow and marked out with a straight line in the plot it.

Figure 9

Figure 8

Chart, line chart

Description automatically generated

Figure 9

The idea of creating a scatter plot (Figure 9) is like that of a bar chart, but the displayed results are not very intuitive. The maximum value is marked with blue dots, the minimum value with green dots, Germany with red dots, and the EU average with purple dots and a purple line to mark the average.

Figure 9

Map

Description automatically generated

Figure 10

Considering that picture 6 relates to some EU countries, we import the world map from map\_data, and intercept the information of 13 EU countries from it, and combine with the database established before, to map the average working time in EU. Finally, use gradient colors to mark different countries on the map. Use color differences to show the division between countries, and get Figure 10.

***Groceries Spending Graph***

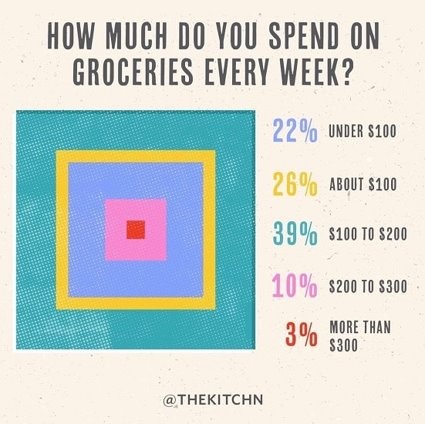


Figure 11

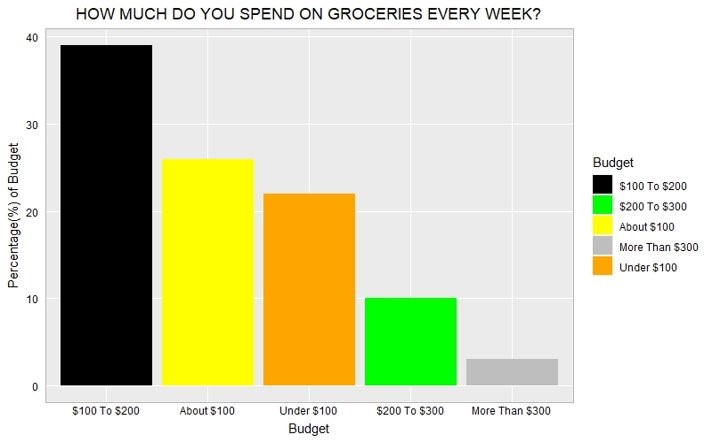


Figure 12

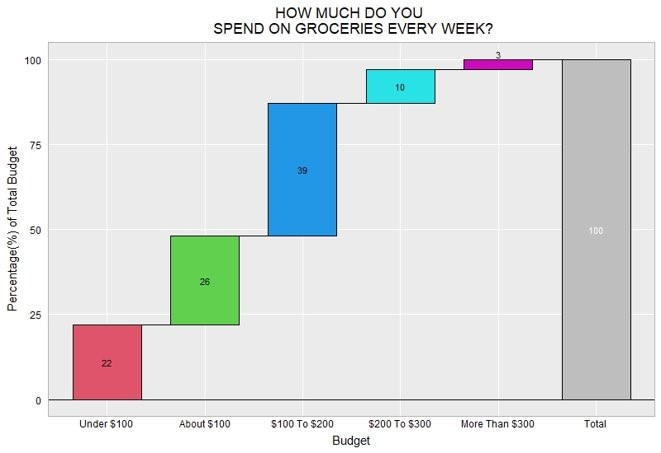


Figure 13

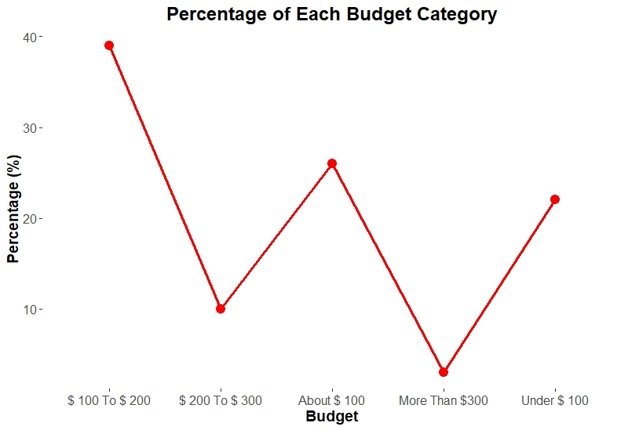


Figure 14

**Relevance**

This project is concerned with transforming bad visual images into optically appealing visual images. It is important for viewers who are easily misguided. A graphically appealing image conveys information more effectively than a poorly visualized image. It allows viewers to understand what message has been presented to them easier.

Reduces misinterpretation with data. A misleading data visualization leads viewers to interpret wrongly the data which is presented. Viewers may reach inaccurate findings or decisions made based on erroneous facts or their mind may have been influenced by the graphic designer.

Regain the viewer's trust. When viewers realize that the data visualization is biased, they erode their trust in the sources of information. It may damage the reputation of the individual or organization responsible for the manipulation of misleading visualization. An image that displays a data visualization that is completely informative creates trust in the image designer on the part of the viewer.

**Conclusion**

***Average Female Hight***

Starting the Y-axis of the image from 0 reveals that the average height difference of women between the five countries is not as large as the original image shows.

In addition, after analyzing the heights of women living in six continents, it was found that the average height of European women was the highest among the six continents with a mean of 166 cm. the average height of African women was the lowest at 158 cm. the lowest value of the overall sample was 151 cm, located in North America.

***Average Number of Actual Weekly Hours***

After redesigning the graph of average working hours in the EU countries, it was found that the average working hours in the 13 countries in the comparison were not very different. Like the average height graph, the EU average weekly working time graph is missing the origin of the x-axis. The redesign allows the average working hours for each country to be displayed in full on the bar chart, with Germany's average working hours being only 0.9 hours more than the average for the entire EU.

***Groceries Spending Graph***

The redesigning bar chart allows us to visualize the difference between each consumer price point. The redesigned image shows the magnitude of each variable more clearly than the original graph. The scatter plot shows the exact value of the overall percentage of what each variable account for.

**References**

Bergstrom, C, T., & West, J. (2020, July,12). *Misleading axes on graphs*. Calling Bullshit. https://www.callingbullshit.org/tools/tools\_misleading\_axes.html

Reed, B. (2021, January 26). *Journalism professor posts hilarious history of Fox News's worst and most misleading graphs*. Raw Story. Retrieved from https://www.rawstory.com/fox-news-misleading-graphs/

Hooper, L. (2021, September 27). *Bad Infographics: The Worst Infographics of 2020 (+ Lessons for 2021)* Retrieved https://venngage.com/blog/bad-infographics/