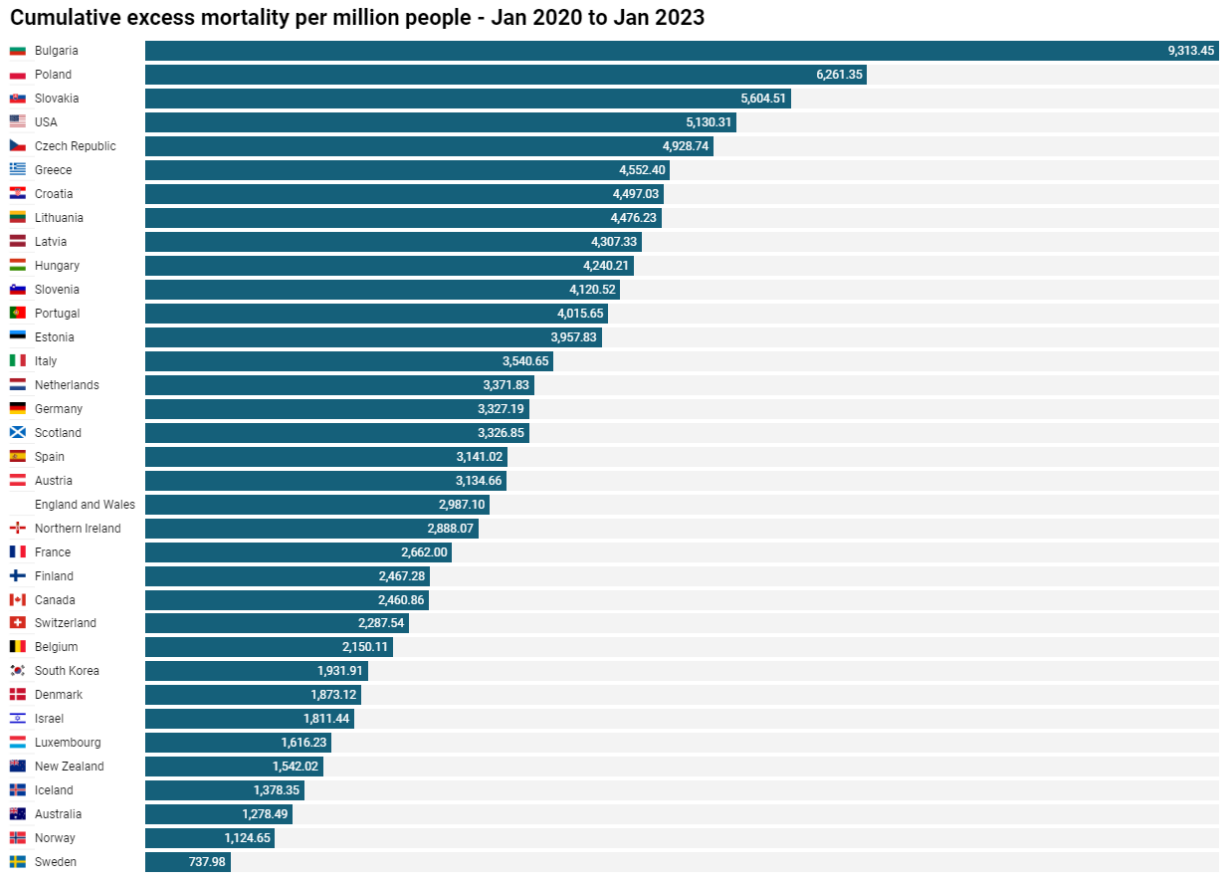
Data Visualization Critiques

Name : Rahul Kumar Nalubandhu
Email : nalubanr@oregonstate.edu

Critique 1:
Referenced :
https://www.reddit.com/r/dataisbeautiful/comments/10q6hrt/oc_total_excess_mortality_per_million_people/



The bar graph in the visualization shows the total excess mortality per million people in various countries. The y-axis displays the countries, and the x-axis displays the excess mortality per million people.

1. Is the data representation accurate? If not, what is the lie factor?

   The data representation in this visualization appears to be accurate, as there is no indication of any manipulation or distortion of the values. This means that the heights of the bars are proportional to the true values of the data, and there is no cherry-picking or manipulation of the data to emphasize certain values or patterns. It is crucial for visualizations to

accurately represent data, as this helps the reader to understand the data and make informed conclusions. Inaccurate representation of data can lead to misunderstandings and incorrect interpretations, which can have negative consequences.

2. Do the number of dimensions of the visualization match the number of dimensions of the data? If not, is it a problem?

   The number of dimensions in the visualization matches the number of dimensions in the data, with two dimensions: one for the countries and one for the excess mortality per million people. Having the number of dimensions match the number of dimensions in the data ensures that all relevant information is included in the visualization and makes it easier for the reader to understand.

3. Is there wasted ink?

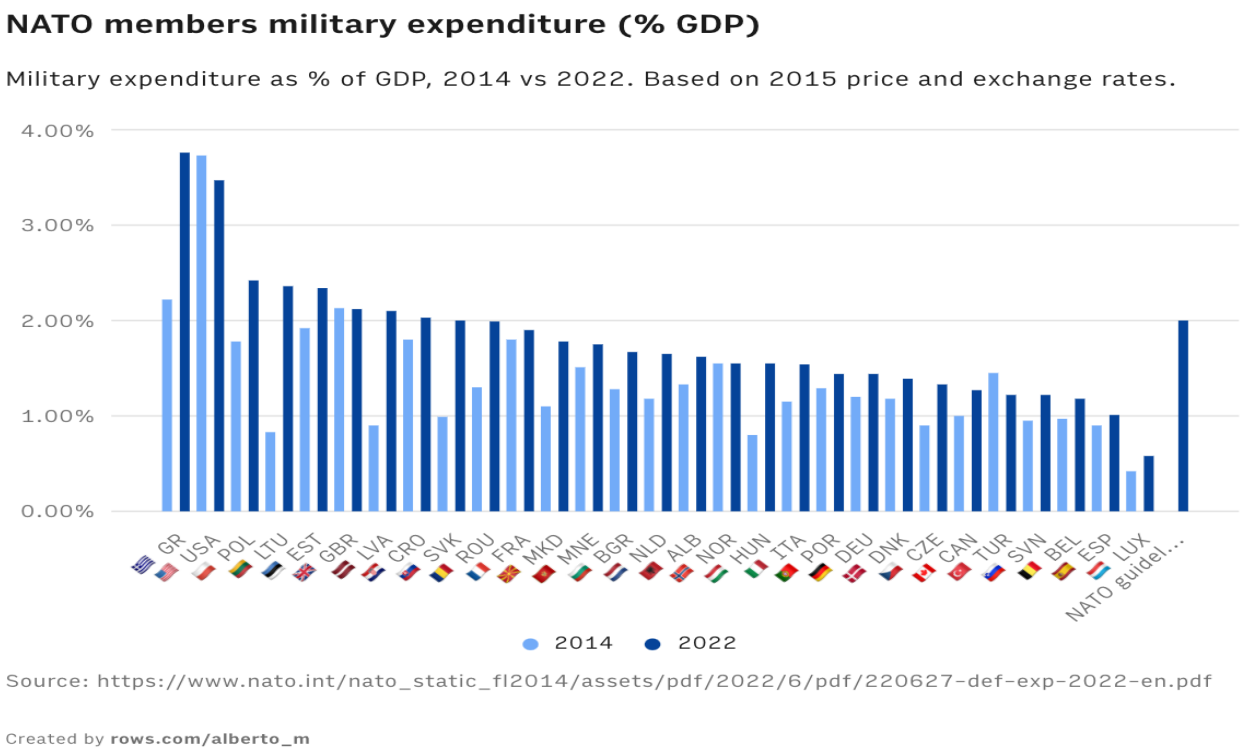
   The background of the visualization is clean and minimalistic, which helps to reduce visual clutter and emphasize the data. The gridlines, axis labels, and bar labels are clear and easily readable, which ensures that the information is easily accessible to the reader. There is no wasted ink in this visualization.

4. Any other of Tufte's list of points which seem relevant:

   The bar labels provide the exact values, which makes it easy to see the precise values for each country and helps to increase the accuracy of the visualization. The use of a bar graph is an effective way to compare the values between countries. However, the order of the countries on the y-axis is not sorted, which could make it more difficult for the reader to compare the values and draw conclusions also the labels on both axes are not mentioned which makes it difficult to decide for one with less knowledge on visualizations. Sorting the countries on the y-axis could improve the visualization and make it easier for the reader to compare the values.

Critique 2:

Referenced from :
https://www.reddit.com/r/dataisbeautiful/comments/10qpagj/oc_nato_members_military_expenditure_gdp_2014_vs/



The visualization is a bar graph comparing the military expenditure as a percentage of GDP for NATO member countries in 2014 and 2022. The y-axis displays the military expenditure as a percentage of GDP and the x-axis displays the NATO member countries. The bars are colored to show the military expenditure of each country in 2014 (light blue) and 2022 (dark blue).

1. Is the data representation accurate? If not, what is the lie factor?

   The data representation appears to be accurate based on the information provided. There is no indication of a lie factor.

2. Do the number of dimensions of the visualization match the number of dimensions of the data? If not, is it a problem?

   The number of dimensions in the visualization matches the number of dimensions in the

data. Therefore there are two dimensions: one for the NATO member countries and one for the military expenditure as a percentage of GDP.

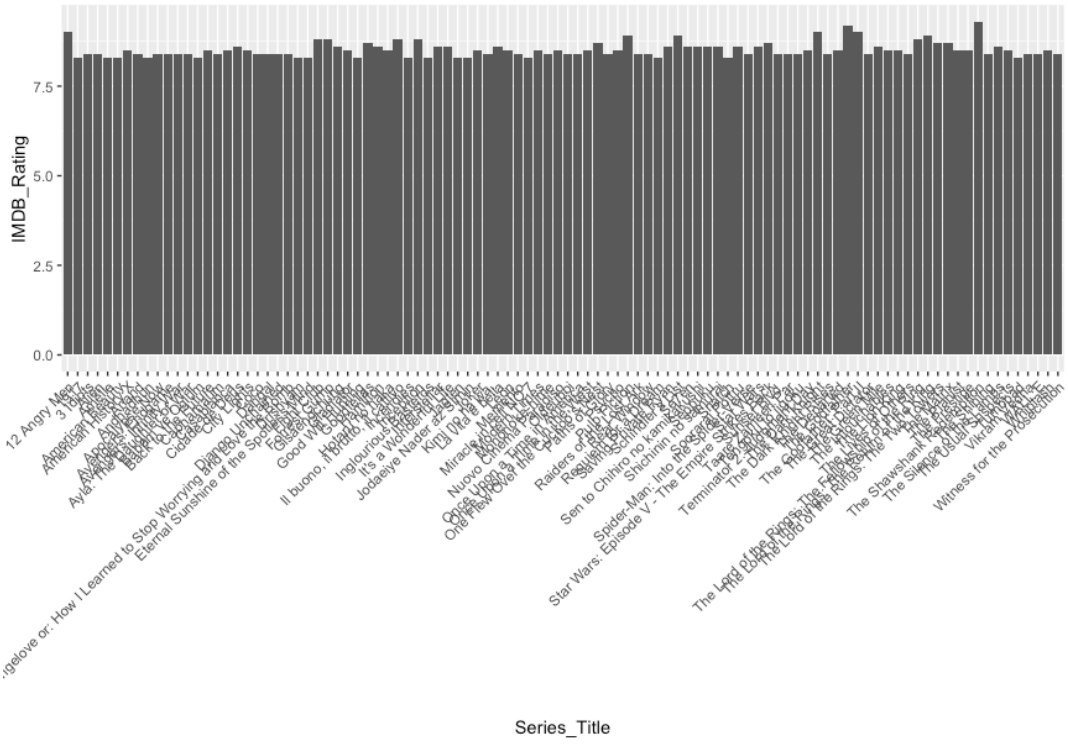
3. Is there wasted ink?

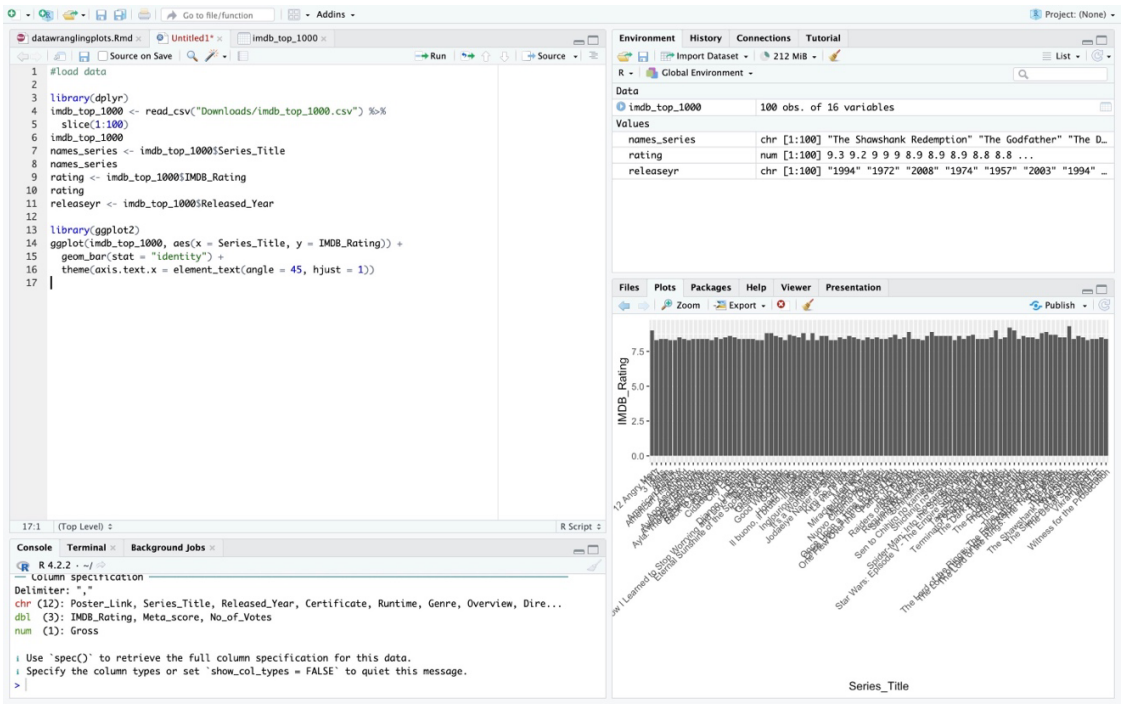
   The visualization is clean and minimalistic, with no wasted ink. The labels on the x-axis and y-axis are clearly visible and provide sufficient context for understanding the data. Additionally, the use of color to distinguish between the military expenditure in 2014 and 2022 makes it easy to compare the two sets of data.

4. Any other of Tufte's list of points which seem relevant:

   The labels on the x-axis and y-axis provide sufficient context for understanding the data and the visualization makes efficient use of space by showing a large amount of information in a compact format. However, the visualization could benefit from additional labeling and context. For example, it would be helpful to have the exact percentage of military expenditure labeled on the y-axis, rather than relying on the reader to infer this information from the height of the bars. Additionally, including a title or description of the data could provide additional context for the reader. Overall, the bar graph provides a clear visual representation of the data, making it easy for the reader to understand the relationships between the different variables. However, adding additional labeling and context would make the visualization even more effective.

Critique 3:





The above visualization is from IMDB Movies Dataset, and I used R to do analysis and I extracted only first 100 lines from the original data set and generated graph for 1st 100 movies with reviews.

1. Is the data representation accurate? If not, what is the lie factor?

   The graph appears to be a bar plot of IMDB ratings of 100 TV series, with the x-axis representing the series titles and the y-axis representing the IMDB ratings. The x-axis labels are angled at 45 degrees to prevent overlap, and the just argument is set to 1 to align the labels to the right.

   However, this graph may not effectively communicate the data as it is difficult to read the series titles due to the angle and possible overlap of the labels. A better option could be to use a horizontal bar plot or to limit the number of characters in the series titles to reduce overlap. Additionally, the graph could benefit from labeling the y-axis and adding a title to better contextualize the information being presented.

2. Do the number of dimensions of the visualization match the number of dimensions of the data? If not, is it a problem?

   The number of dimensions of the visualization matches the number of dimensions of the data, which is two. The x-axis represents the series titles, and the y-axis represents the IMDB ratings.

3. Is there wasted ink?

   There is wasted ink as the angled labels on the x-axis could be overlapping, making it difficult to read. May be a horizontal bar plot or limiting the number of characters in the series titles could help reduce the wasted ink.

4. Any other of Tufte's list of points which seem relevant :

   The graph could benefit from reducing the amount of ink used on the x-axis labels to improve clarity and focus on the main data representation.

   The angled labels on the x-axis and the possible overlap of labels could be considered chart junk as they distract from the main data representation and make it difficult to read.

   The graph could benefit from adding a title to contextualize the information being presented and labeling the y-axis to provide more information about the data. The rating was just above 7.5 but not exactly how much it is. It could be helpful if we do the scaling for the rating count on y-axis.

   Dataset reference: https://www.kaggle.com/datasets/harshitshankhdhar/imdb-dataset-of-top-1000-movies-and-tv-shows?resource=download