# HW5: Persuasion through Visualization

I have chosen the data file titled "gender.csv" from the *Gender Equality Indicators 1960–2017* dataset.
The dataset is a comprehensive collection of data with 15,551 entries and 53 columns. It includes various indicators to study the extent of gender equality worldwide, with details for each country and year.

Good Visualization
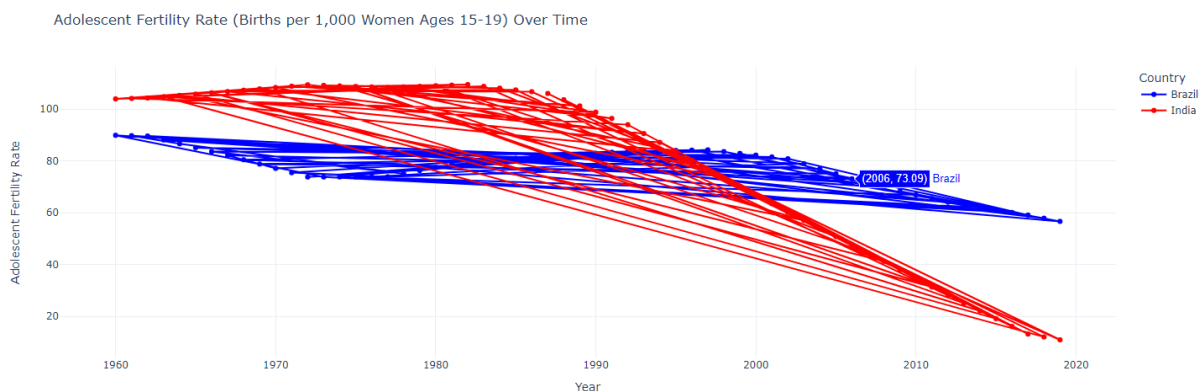
The good visualization aims to provide a clear, accurate comparison of adolescent fertility rates between Brazil and India over several decades. I chose a line graph for its effectiveness in depicting trends over time. Each country is represented by a distinct color (blue for Brazil, red for India) to maintain clarity and aid in distinction between the two datasets. Data points were marked on the graph to highlight specific yearly values, enhancing readability and providing exact figures for more in-depth analysis.The visualization above represents the Adolescent Fertility Rate (births per 1,000 women ages 15-19) over time for Brazil and India.

Rationale:
The visualization was designed to be intuitive and informative, catering to a general audience interested in understanding trends in gender and social issues across countries. By clearly labeling the axes, including a descriptive title, and providing a legend, the graph adheres to best practices in data visualization. This approach helps avoid misinterpretation and allows for straightforward comparison between the two countries' fertility trends, highlighting significant changes or patterns over time.

Data transformations:
1. First, the original dataset was filtered for the "Adolescent fertility rate" indicator.
2. Then, brazil_data was created by selecting rows where the 'Country Name' is 'Brazil'.
3. Similarly, india_data was created by selecting rows where the 'Country Name' is 'India'.

Accordingly, the yearly data values are plotted for both the countries. A different color is chosen for each so that they are easily distinguishable. This graph is designed to be clear and easy to interpret:

Trend Lines: Each line represents one of the countries, with Brazil and India clearly distinguished by different colors and markers.
Data points: Each data point either represents Brazil's adolescent fertility rate or India's.
Axes and Labels: The x-axis represents the year, and the y-axis represents the adolescent fertility rate. Both axes are clearly labeled.
Title and Legend: The title at the top clearly states what the graph shows, and the legend distinguishes between the two countries.

Evil Visualization

Survival to Age 65: Female (% of cohort) across All Countries (2010-2020)
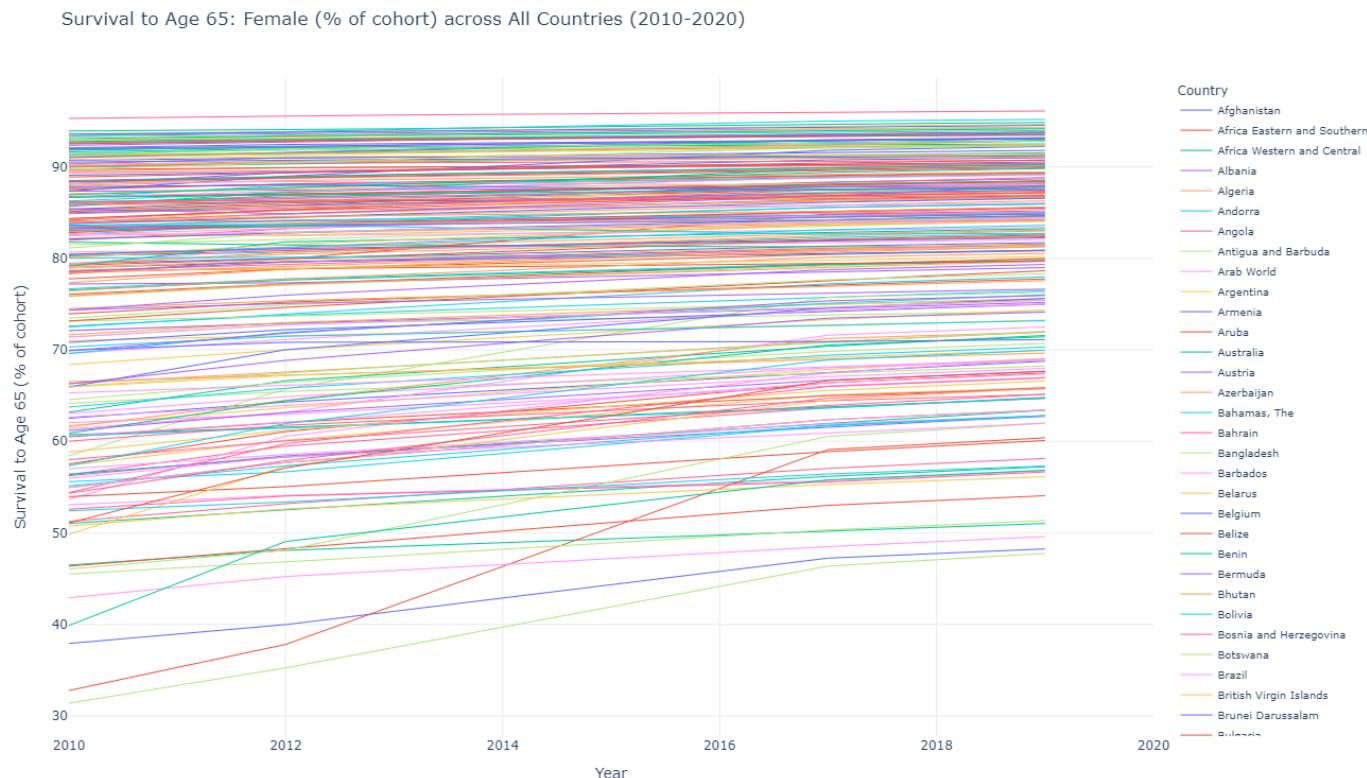
For the "evil" visualization, I intentionally created a cluttered and confusing graph, plotting survival to age 65 for females in all available countries over the last decade. Each country's data was plotted as a separate line without aggregation, leading to a dense, overlapping design. The choice of color was arbitrary, with no strategic differentiation, contributing to the visualization's complexity and lack of clarity.

Rationale:
This visualization serves as a negative example, demonstrating how not to represent complex datasets. The design choices were made to overwhelm the viewer deliberately, making it nearly impossible to extract meaningful insights or distinguish between individual countries' data. The misleading title suggests a comprehensive analysis, whereas the actual graph offers no clear information, illustrating the importance of thoughtful data presentation and the ethical implications of visualization choices.

Data transformations:

1. <u>Data Filtering</u>: The dataset was filtered to include only the data from the last decade. This means selecting rows where the 'Year' is greater than or equal to 2010, focusing on the most recent 10 years available in the dataset.
2. <u>Pivoting Data:</u> The filtered data was then pivoted to change its structure. Specifically, the dataset was restructured so that each country's data becomes a separate column. This pivot operation changes the dataset from a long format, where each row contains a single year-country pair, to a wide format, where each row represents a year and each column represents a country.

Accordingly, for each country a line was added to the plot representing the survival rate over time. This was done using a loop that iterates over all the countries and plots a line for each. The visualization deliberately employs techniques such that the different data points are cluttered thus preventing a meaningful understanding of the plot. Not only are all the countries plotted, neither are the y-axis values not spaced well enough.

<u>Data points:</u> Each country's data series was plotted as a separate line on the same graph. With a large number of countries, this results in significant overlap, making individual lines hard to discern and interpret.

<u>Labels and Axes:</u> While the axes themselves might not be misleading, the sheer volume of data renders them useless as reference points due to the clutter.

<u>Legends:</u> The legend includes every country, making it overwhelmingly large and impractical to use for identifying specific lines within the graph.

<u>Title:</u> The title of the graph might suggest a clear and comprehensive analysis ("Survival to Age 65: Female (% of cohort) across All Countries"), which is misleading because the actual visualization does not provide clarity or a comprehensive understanding due to its complexity and clutter.