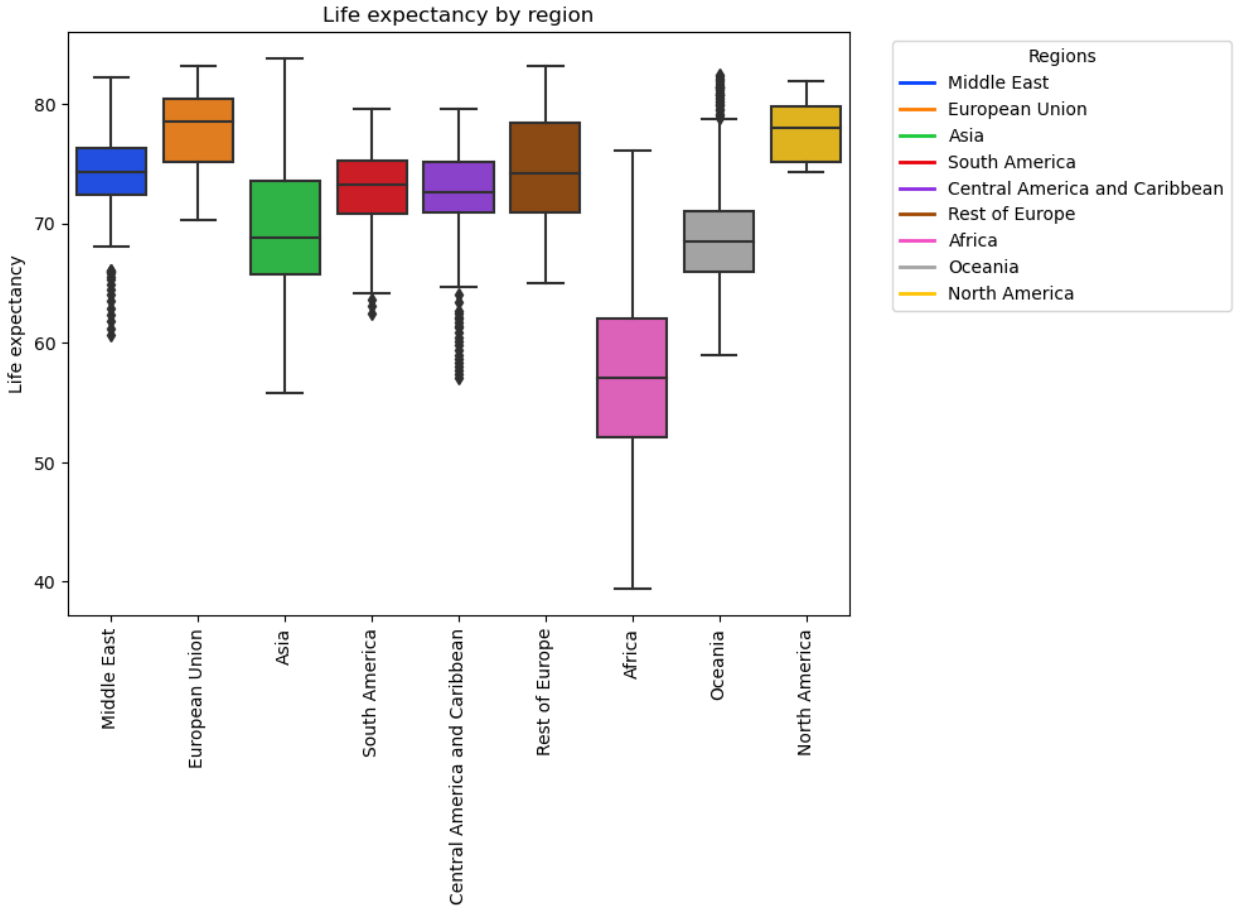
INFSCI 2415: Information Visualization Midterm Report

Vishruth Pradeep Reddy

vpr8@pitt.edu

Forecasting Life Expectancy

Link to dataset: https://www.kaggle.com/datasets/lashagoch/life-expectancy-who-updated/data



The above `box plots (or box-and-whisker plots)` are a graphical representation of my dataset's summary statistics. The above boxplot shows the life expectancy of the people in various continents.

- `Middle East`, `South America`, `Central America`, and `Oceania` have `outliers`.
- `Africa` has the `widest and the lowest range` of life expectancy whereas, `North America` has the `narrowest and the highest range` of life expectancy compared to the other regions.
- The legend on the side displays the color used to represent each region in the boxplots.

About the boxplots:

1. Box: The central part of the plot is a rectangular box. The box represents the interquartile range (IQR), which includes the middle 50% of the data. The bottom and top of the box mark the first quartile (Q1) and third quartile (Q3), respectively. The width of the box shows the spread of the data within this middle 50%.

- **Q1 (First Quartile)**: The 25th percentile of the data, i.e., the value below which 25% of the data falls.
- **Q3 (Third Quartile)**: The 75th percentile of the data, i.e., the value below which 75% of the data falls.
- **IQR (Interquartile Range)**: The range between Q1 and Q3, representing the middle 50% of the data.

2. Median: A vertical line or marker inside the box represents the median (Q2) of the data, which is the middle value when the data is sorted.

- **Q2 (Median)**: The middle value of the data set.

3. Whiskers: Lines extending from the top and bottom of the box, indicate the range of the data.

- **Whisker Range**: The range between the whiskers, which typically covers the entire data range if `whis` is set to `[0, 100]`.

4. Outliers: Any data points outside the whiskers are considered outliers and are usually plotted as individual points. Outliers are values that are significantly different from the rest of the data and can provide insights into data anomalies.