Data Visualization and Charts

Armand Junior DONGMO NOTUE

College of Engineering and Technology, Grand Canyon University

DSC-510-O500: Advanced Probability and Statistic

Edward Ofori

03/13/2024

Life expectancy and socioeconomic:

Life expectancy at birth indicates the number of years a newborn would live if the mortality patterns prevailing at the time of birth remained the same throughout their life. This is a key indicator for assessing the health of the population. In this exercise, we aim to better understand the factors affecting life expectancy. For this, we will first study the distribution of life expectancy around the world and compare life expectancy in different regions of the world; we will then make a comparative study of the factors affecting life expectancy in North America and, finally, we will study the evolution of life expectancy in the United States over time.

Data collection:

The data set used for this analysis includes various variables such as(Country Name, Country Code, Region, IncomeGroup, Year, Life Expectancy World Bank, Prevelance of, Undernourishment, CO_2, Health Expenditure %, Education Expenditure %, Unemployment, Corruption, Sanitation, Injuries, Communicable, NonCommunicable). It has been downloaded from the Kaggle website.
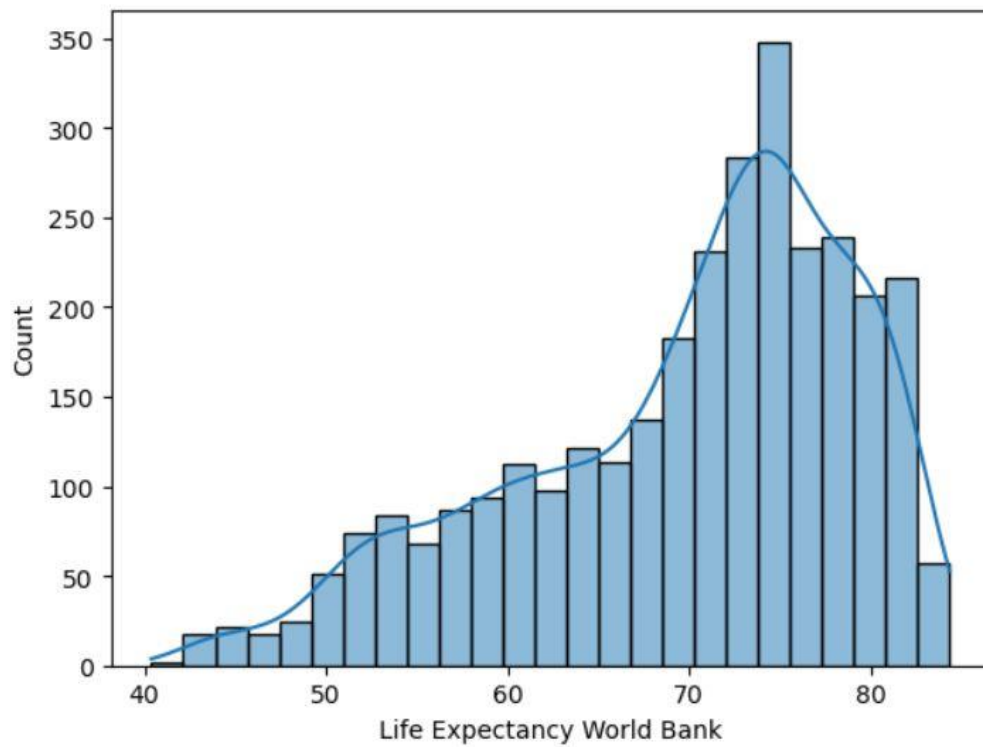
Data Visualization:

We employed Python libraries like Matplotlib and Seaborn to create insightful visualizations for each variable in the dataset.
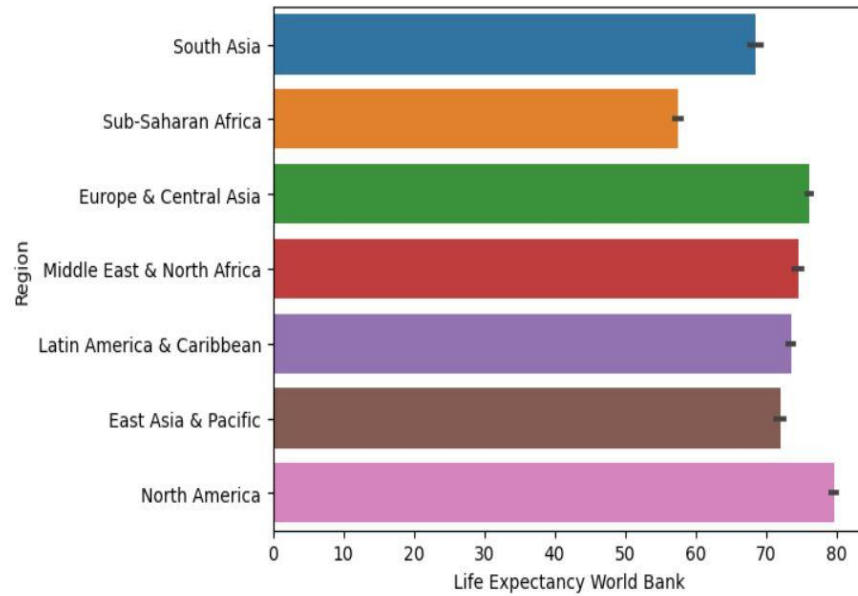
Let's take a look at these visualizations:

- Life Expectancy World Bank (world)

- Histogram: it allowed us to see the distribution of life expectancy around the world.
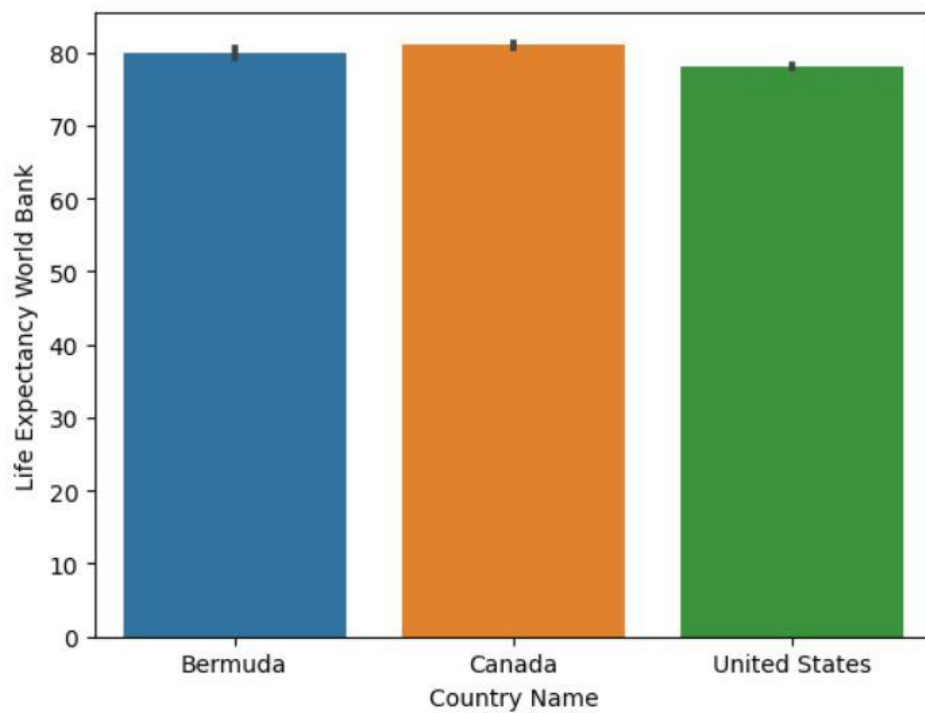

- Bar plot: it enabled us to compare life expectancy in different parts of the world.
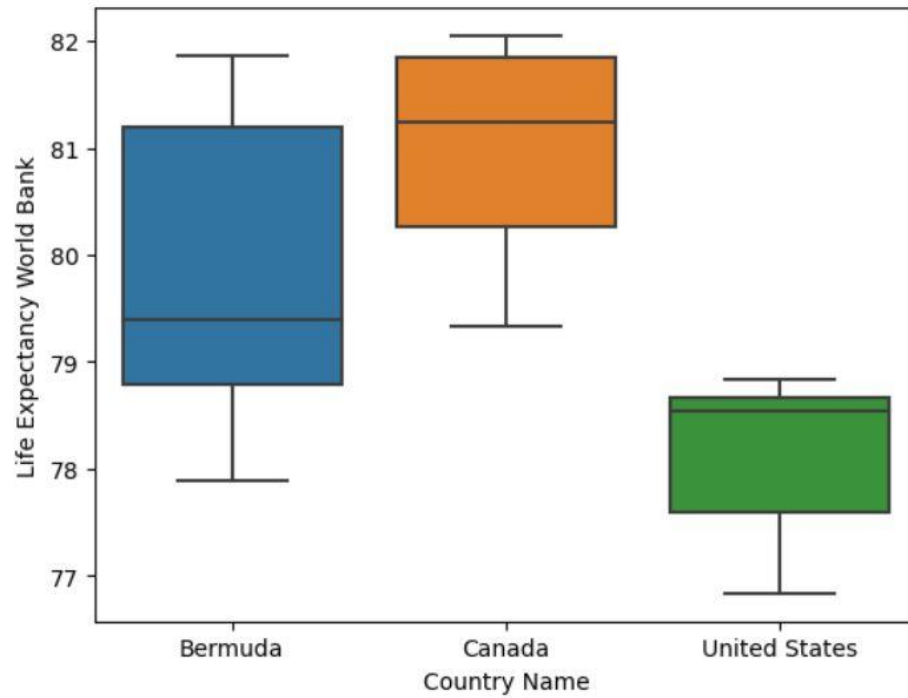
- Life Expectancy World Bank (North America)

  - Barplot: The diagram shows that Canada has the highest life expectancy, followed by Bermuda, and that the USA has the lowest.
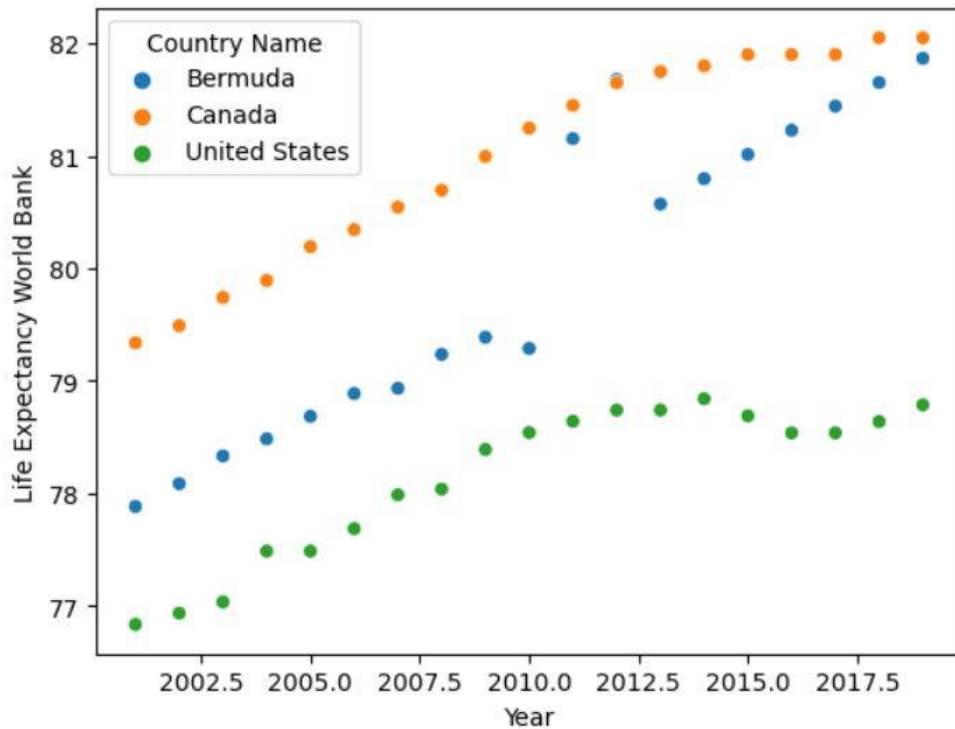
- Box Plot: it enabled us to compare Canada's maximum, minimum and median life expectancy, which is higher than in other countries.

- Line Plot: it has enabled us to visualize life expectancy in the three countries over the years.
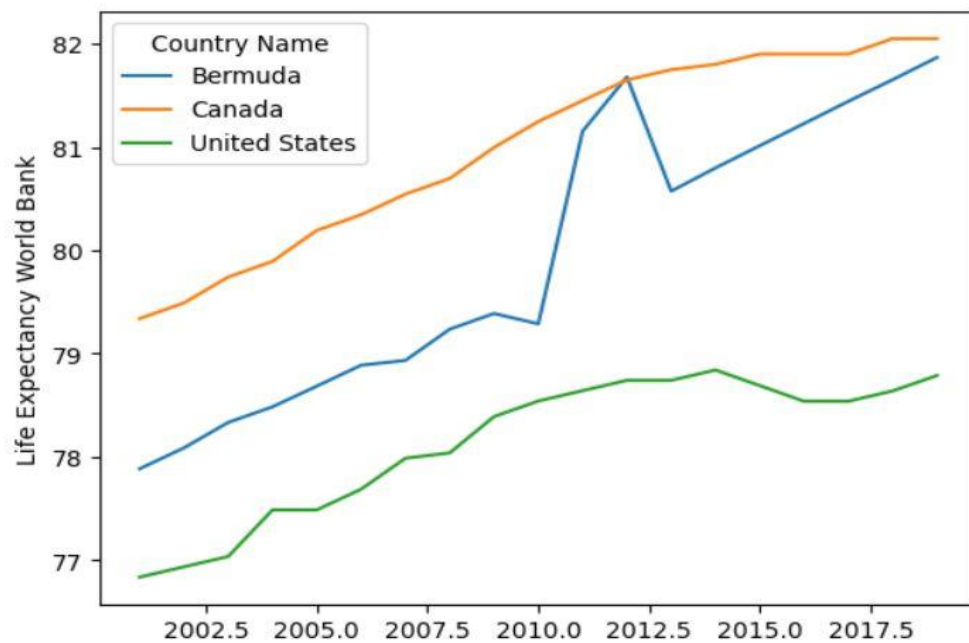
- Scatter Plot: it has enabled us to visualize life expectancy in the three countries over the years.

- CO2 emissions:
    - Bar plot: it allowed us to compare co2 emissions between canada and the united states
    - Scatter Plot: it enabled us to visualize co2 emissions in the three countries over the years.

- Health Expenditure :

    - Bar plot: it allowed us to compare Health Expenditure between canada and the united states
    - Scatter Plot: it enabled us to visualize Health Expenditure in the three countries over the years.

- Unemployment :
    - Bar plot: it allowed us to compare unemployment between canada and the united states
    - Line Plot: it enabled us to visualize Unemployment in the three countries over the years.

- Sanitation :

  - Bar plot: it allowed us to compare sanitation between Canada and the United States
  - Line Plot: it enabled us to visualize sanitation in the three countries over the years.
- Injuries :

  - Bar plot: it allowed us to compare Injuries between Canada and the United States
  - Line Plot: it enabled us to visualize Injuries in the three countries over the years.
- communicable :

  - Bar plot: it allowed us to compare communicable diseases between Canada and the United States
  - Line Plot: it enabled us to visualize communicable diseases in the three countries over the years.

Evaluation of visualizations:

Each visualization was evaluated based on our results. For example:
- Histograms compare data effectively, but this comparison is simply done in a general way; If we want to compare other elements of the data, we can use another chart.
- The boxplot usually allows us to quickly assess the distribution, variability, and presence of outliers in a data set, but in this case it allowed us to effectively compare the data by comparing the Median, Max, and Min values.
- Line plots have been optimized to better represent trends over time, ensuring they are easy to interpret.
- Scatterplots have been refined to better represent trends over time, reducing clutter and improving readability, providing a clearer overview of relationships between variables, but the line chart is more effective in this case.

In conclusion, this visual analysis of factors affecting life expectancy provides valuable

information for decision making. After looking at the graph, we can conclude that the elements

that most reduce life expectancy are increased CO2 emissions and increased communicable

diseases.

GitHub: https://github.com/ARMAND-cod-eng/Data-Visualization-and-Charts/blob/main/Data Visualization and Charts.ipynb

Video: https://www.loom.com/share/93b0e80e74f24f748dc8c26d43001b3c

Video2: https://www.loom.com/share/ddb2f6d468b544ada455cc20ee488181

data set: https://www.kaggle.com/datasets/mjshri23/life-expectancy-and-socio-economic-world-bank/data

References

Datacamp. (2023). *11 Data Visualization Techniques for Every Use-Case with Examples.* *https://www.datacamp.com/blog/data-visualization-techniques*

Jack, D & Ilya, I. (*2024). Hands-On Data Visualization.* https://handsondataviz.org/

Practical Data Science. (2021). *How to visualize correlations using pandas and Seaborn*. https://practicaldatascience.co.uk/data-science/how-to-visualise-correlations-using-pandas-and-seaborn

Will, K.(2018). *Visualizing Data with Pairs Plots in Python.* Towardsdatascience. https://towardsdatascience.com/visualizing-data-with-pair-plots-in-python-f228cf529166