SUMMARY

I have chosen this dataset for analysis because it provides a comprehensive view of factors influencing **life expectancy** across different countries and years, covering the period from 2000 to 2015. It combines health indicators (such as mortality rates, disease prevalence, immunization coverage, and nutrition), economic measures (like GDP and health expenditure), and social factors (including schooling and income composition of resources). With 2,944 records from 195 countries, the dataset offers a rich mix of both developed and developing nations, allowing for meaningful comparisons and trend analysis. The diversity and depth of variables make it well-suited for exploring relationships between health, economic, and social development, and understanding how these elements contribute to differences in life expectancy worldwide.

|	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	...	Polio	Total expenditure	Diphtheria	HIV/AIDS	GDP	Population	thinness 1-19 years
0	Afghanistan	2015	Developing	65.0	263.0	62	0.01	71.279624	65.0	1154	...	6.0	8.16	65.0	0.1	584.259210	33736494.0	17.2
1	c###bd	2014	Developing	59.9	271.0	64	0.01	73.523582	62.0	492	...	58.0	8.18	62.0	0.1	612.696514	327582.0	17.5
2	Afghanistan	2013	Developing	59.9	268.0	66	0.01	73.219243	64.0	430	...	62.0	8.13	64.0	0.1	631.744976	31731688.0	17.7
3	Afghanistan	2012	Developing	59.5	272.0	69	0.01	78.184215	67.0	2787	...	67.0	8.52	67.0	0.1	669.959000	3696958.0	17.9
4	Afghanistan	2011	Developing	59.2	275.0	71	0.01	7.097109	68.0	3013	...	68.0	7.87	68.0	0.1	63.537231	2978599.0	18.2

5 rows × 22 columns

Visuals

Histplot
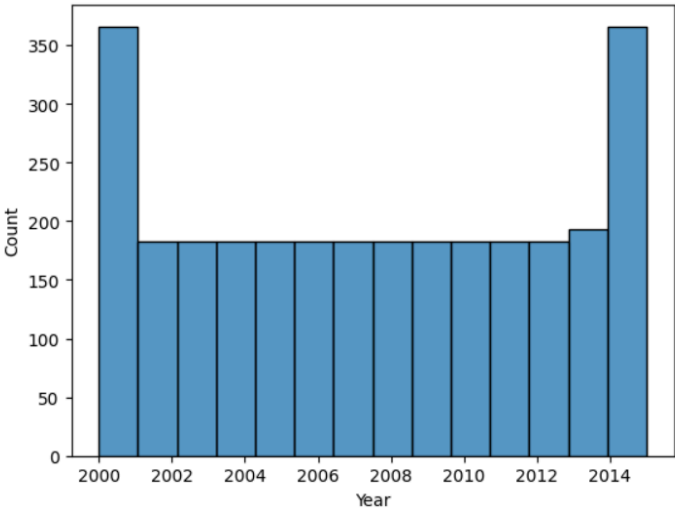
Boxplot
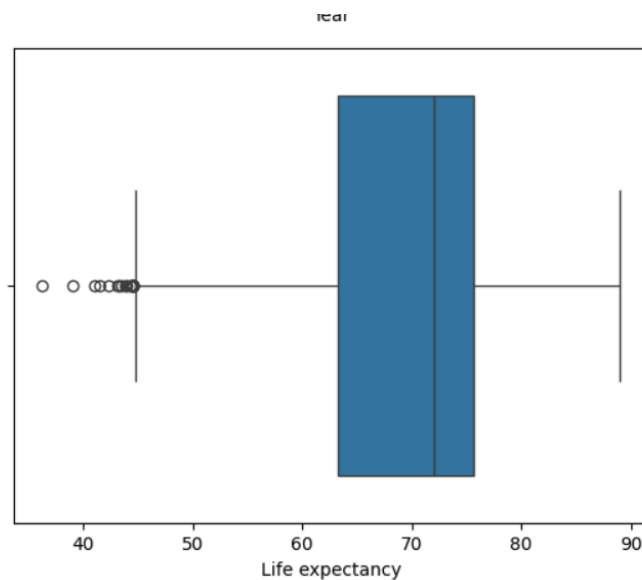
Heatmap
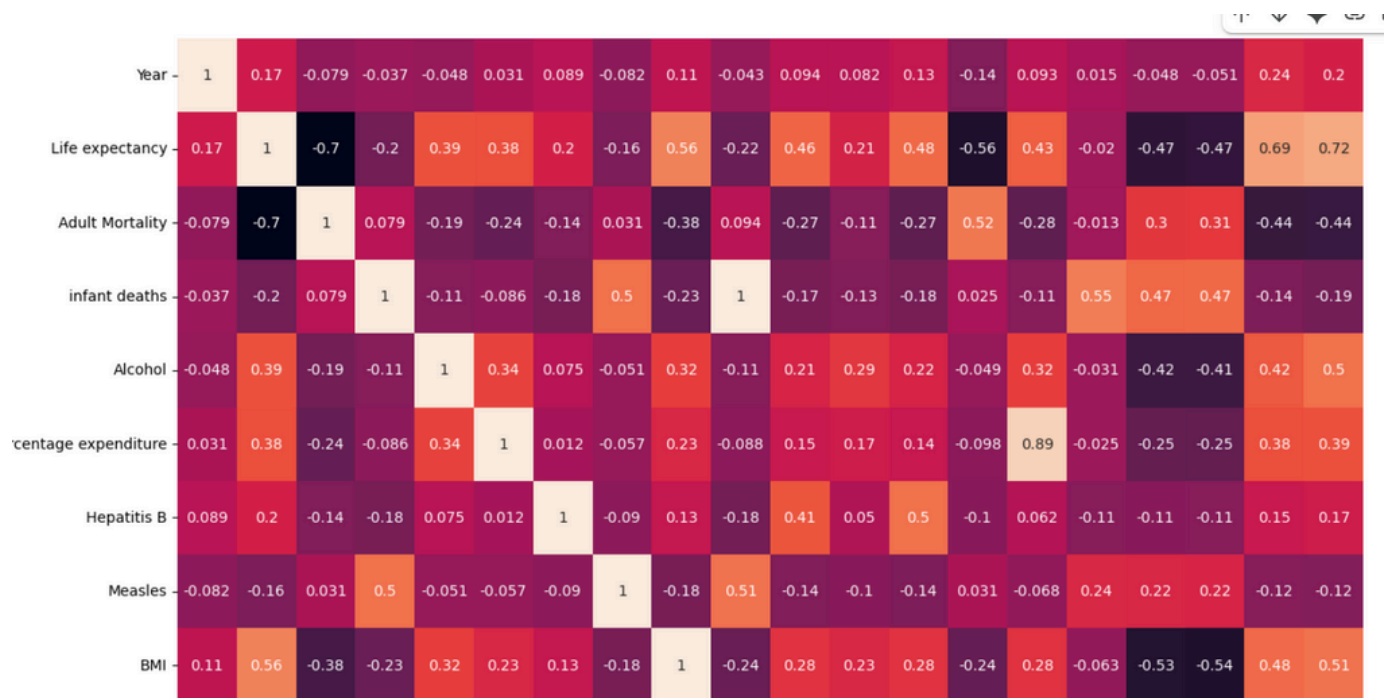
Conclusion

In analyzing this dataset, I selected histograms, heatmaps, and scatter plots because each provides a distinct and valuable perspective on the data. Histograms are ideal for understanding the distribution of key variables such as life expectancy, GDP, and mortality rates, revealing patterns like skewness or concentration of values. Heatmaps are powerful for visualizing correlations between multiple variables at once, allowing us to quickly identify strong positive or negative relationships, such as the link between education levels and life expectancy or between health expenditure and mortality. Scatter plots, on the other hand, are essential for examining pairwise relationships in greater detail, highlighting trends, clusters, or outliers in variables like GDP versus life expectancy or schooling versus mortality rates. Together, these visualization techniques complement each other, offering

both an overall understanding and targeted insights into the health, economic, and demographic dynamics captured in the dataset.