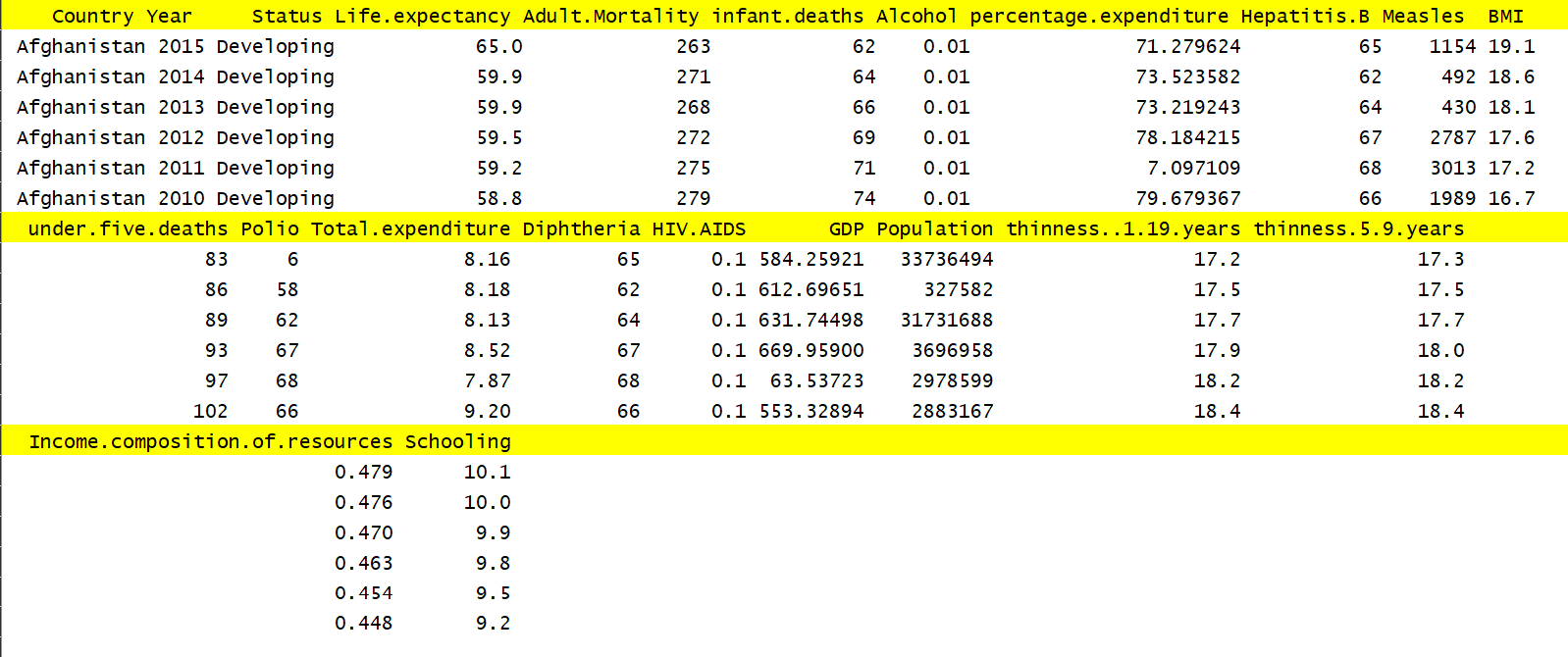
# **Exploratory data analysis**

Exploratory Data Analysis (EDA) is a critical first stage in every data analysis project, including the life expectancy case study. Its goal is to obtain a deeper knowledge of the dataset, uncover patterns, linkages, and potential concerns, and lead us in developing hypotheses for future quantitative research. In the life expectancy case study, we will use EDA to better understand the dataset and highlight crucial insights.

**Data Collection and Loading**

Begin by obtaining a life expectancy dataset from a trustworthy source or database. The dataset should include factors relevant to life expectancy, such as nation, year, GDP, healthcare spending, and so on. Load the dataset into R environment using appropriate libraries and review the first few rows to confirm proper loading. ( Figure 1)

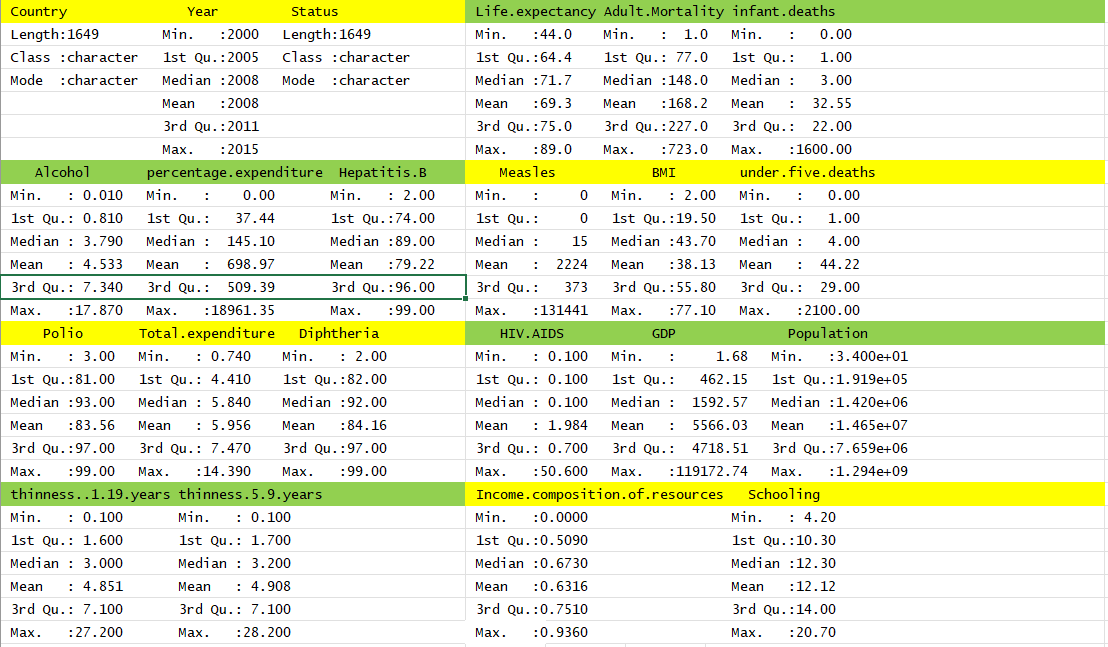


Figure

**Data Summary**

We have generated a numerical summary of the dataset to get an initial overview. This includes mean, median, standard deviation, minimum, and maximum values for relevant numerical variables (e.g., life expectancy, GDP, schooling, etc.).

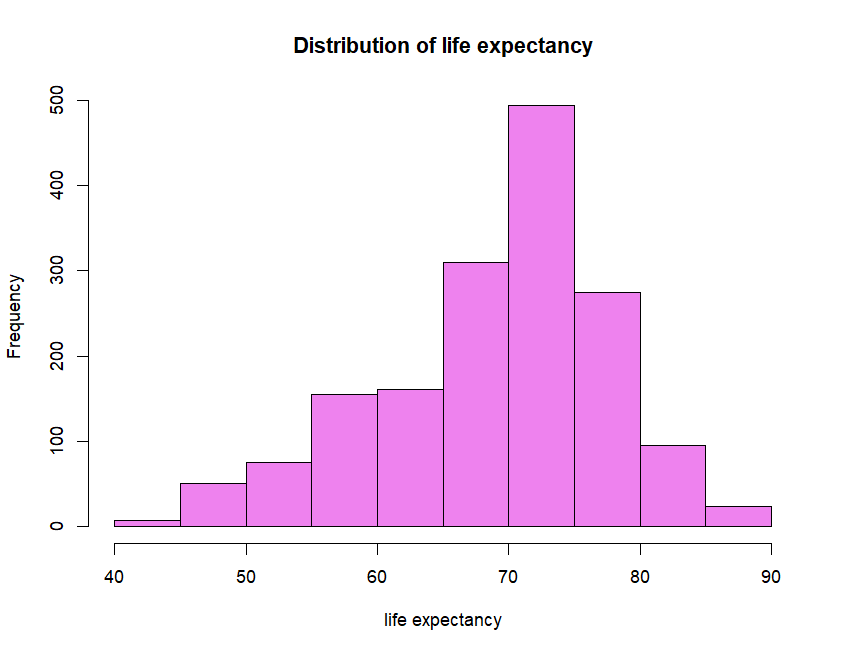
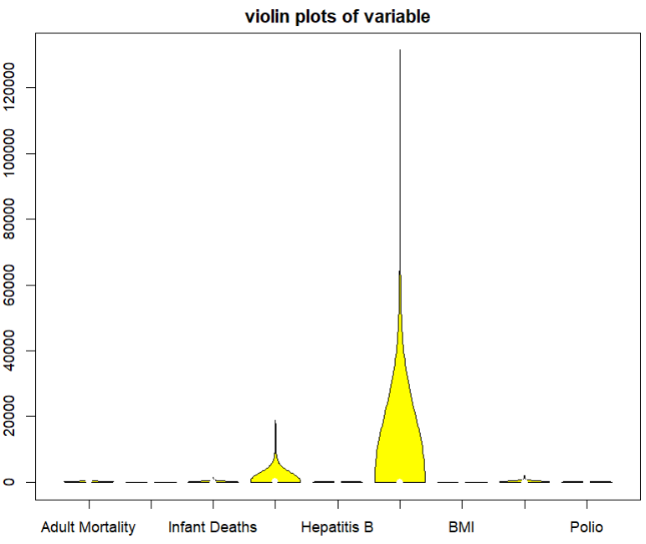
(Figure 2)



Figure

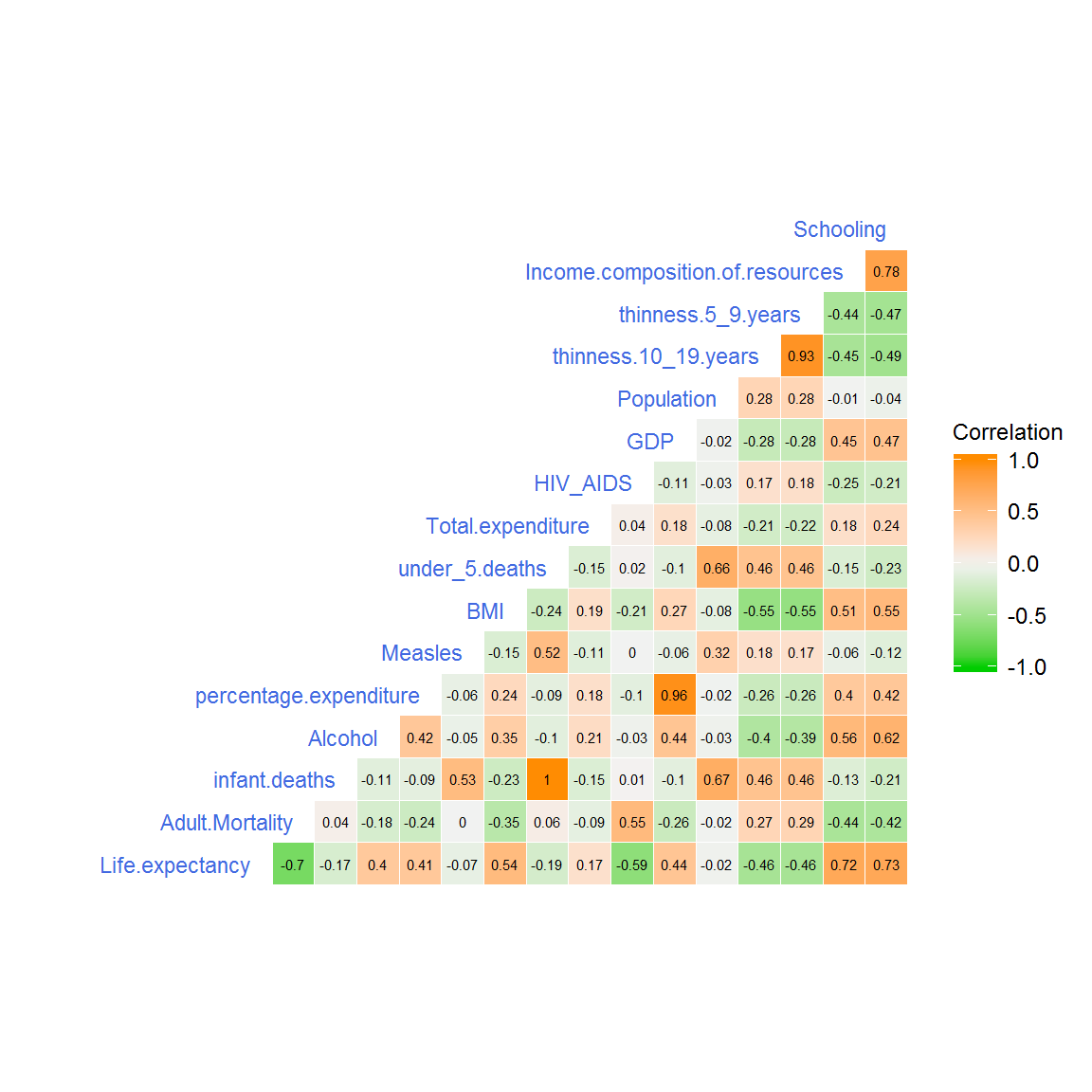
**Univariate Analysis**

To display the distributions of crucial numerical data (such as life expectancy), we plotted histograms, violin plots. This will aid in determining whether they follow any particular pattern (normal, skewed, etc.).To display the distributions of crucial numerical data (such as life expectancy), we plotted histograms. This will aid in determining whether they follow any particular pattern (normal, skewed, etc.).



**Data Visualization**

To uncover patterns of linkage between various numerical variables, we generated visualizations such as a correlation matrix and a heat map (figure 3). Using interactive visualization tools to go deeper into data and unearth insights.



Figure

The life expectancy dataset exploratory data analysis offered useful insights that will guide our future quantitative research. We discovered intriguing trends and linkages, such as positive correlations between life expectancy and GDP and healthcare expenditures, while highlighting differences among nations and regions. The histograms for life expectancy indicated a broadly normal distribution. The analysis laid the groundwork for additional examination, allowing us to make educated judgments and obtain a better knowledge of the facts in this case study.