Abhilash Kumar Dikshit   
R Practice: ALY 6010

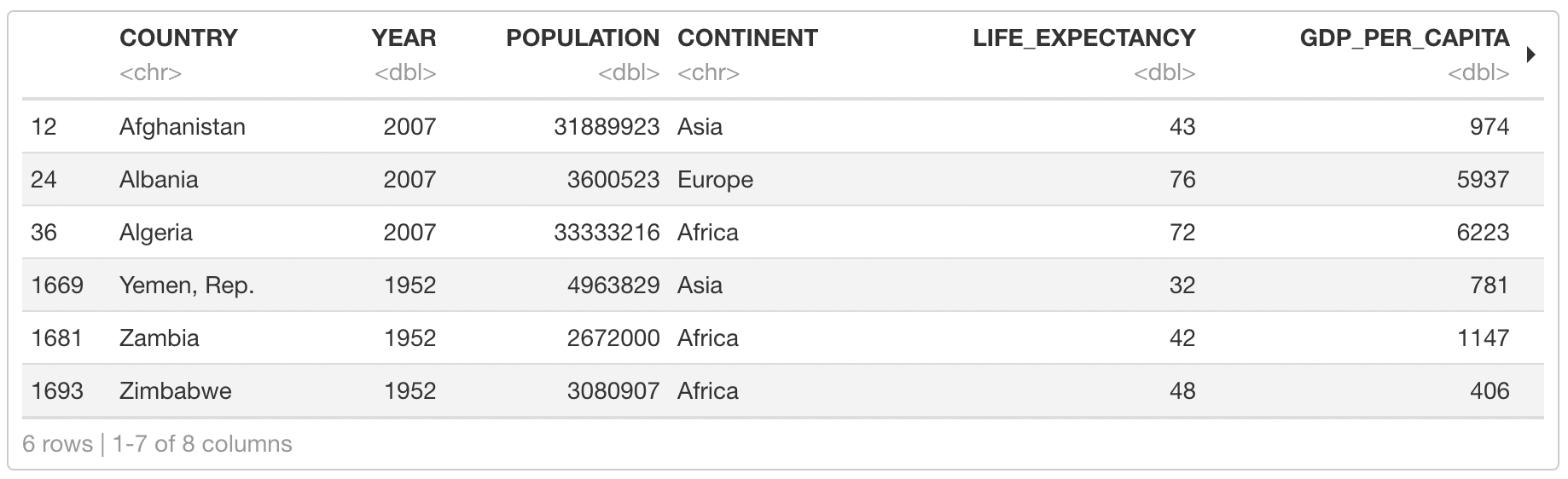
Module 1

Week 1

Initially, I began by downloading the raw dataset *“gapminderDataFiveYear.csv”* from [GitHub](https://raw.githubusercontent.com/plotly/datasets/master/gapminderDataFiveYear.csv) and using the libraries below for further data analysis report.

|  |
| --- |
| Graphical user interface, application  Description automatically generated  *Fig1: Top and Bottom 3 Of Raw Dataset* *(1704 observations and 6 attributes)* |

Fig 2: In order to clean up the data, it was sorted by descending *year*; column headers were converted to capital letters, and the regular expression was removed using gsub; column class type was changed from character to integer and the final stage was to add the life expectancy percentage column.



*Fig2: Top and Bottom 3 Of Cleaned Dataset*

|  |  |
| --- | --- |
| Text  Description automatically generated  *Fig3: Data set’s variable and respective data types* | Fig 3: As we can see from the data frame, we have 1704 observations and 7 attributes available for our data analysis report and the class types of the following attributes have 2 characters, 3 integers and 2 numeric data types. |

Abhilash Kumar Dikshit

R Practice: ALY 6010

|  |  |
| --- | --- |
| Table  Description automatically generated with low confidence  *Fig4: Summary of data frame* | Fig 4: The mean population value across every country from the year 1952 till 2007 is 30 million and the median is 70 million. The minimum Life expectancy is 23 and maximum is 83, and the average life expectancy is at 59 while the median is 60. |

|  |  |
| --- | --- |
| A picture containing text  Description automatically generated  *Fig5: Frequency table for Continent*    *Fig6: Frequency table with Proportions*    *Fig7: Frequency table with Percentages*    *Fig8: Frequency table with Percentages (readability)*  *Text  Description automatically generated* *Fig9: Cumulative Frequency Table* | Fig 5: Looking at the frequency table data for the continent, African continent has the highest frequency of 624 as compared to the other continent.  Fig 6: Then we converted the frequency table to a proportion table. The output shows the relative proportions of each value in our example vector.  Fig 7: For this, we multiplied the proportion table that we have initialized with 100. The output shows the percentages of each character element in our example vector.  Fig 8: The previous R code has rounded the percentages to two digits, and it has added a percentage sign after each value for better readability.  Fig 9: To find the Cumulative frequency, we applied the “cumsum” function. The output shows a cumulative frequency table of our input data. |
| Abhilash Kumar Dikshit  R Practice: ALY 6010 |  |

|  |  |
| --- | --- |
| Fig 10: For the cross-tabulation report, Oceania continent for the year 2007 has been taken as reference.  As shown in the table, we only had 2 observations for the respective continent i.e., for Australia and New Zealand for the year 2007. | Table  Description automatically generated *Fig10: Cross tabulation for Oceania (2007)* |

|  |  |
| --- | --- |
| Fig 11: For the life expectancy report of each continent, first we identified the average life expectancy using mean functionand the output was 58.9689 which is denoted as Red Dash Line in boxplot.  Based on the data visualization from year 1952-2007, all the continents median values are above the average life expectancy except African continents due to scarcity of food and water that plays a major factor for the illness in that region. Moreover, both maximum and minimum life expectancy can be seen in Asia. | Chart, box and whisker chart  Description automatically generated  *Fig11:* *Life Expectancy of Each Continent* |

|  |  |
| --- | --- |
| Chart, box and whisker chart  Description automatically generated  *Fig12: GDP per capita of Each Continent* | Fig 12: For the GDP per capita report for each continent, first we identified the average GDP per capita using mean functionand the output was 7214.837 which is denoted as Red Dash Line in boxplot. The median value of Europe and Oceania are living above the average GDP per capita. Countries in Africa had the lowest GDP per capita due to climate change as well as lack of hospitality and money bank. |

Abhilash Kumar Dikshit

R Practice: ALY 6010

|  |  |
| --- | --- |
| Fig 13: Here the density plot of Life expectancy for each continent has been shown because histograms will block each other. Due to the transparency, we can clearly identify Oceania and Asia were having higher expectancy as compared to other continents. We will be discussing more later using other plots by comparing the populations. | *Fig13:* *Density Plot of Life Expectancy for Continent* |

|  |  |
| --- | --- |
| Chart, line chart  Description automatically generated  *Fig14:* *Highest Population Based on Latest Year* | Fig 14: To get the highest population for the top 10 countries for the latest year, which was 2017, the data was sorted with index value 2 and 3 for Population and Year column together and then then column 1 to 3 was considered which included country, year, and population for our data.  Based on the given visualization, China had the highest population with more than 1.3 billion population in the year 2007 followed by India whereas Mexico was at the least. |

|  |  |
| --- | --- |
| Fig 15: The respective histogram shows a total of 276 countries which had the life expectancy of more than 70 and there are only 7 countries with life expectancy of more than 80 while there were only 6 countries with expectancy below 30 years of age. | Chart, histogram  Description automatically generated  *Fig15: Frequency of Life Expectancy* |

Abhilash Kumar Dikshit

R Practice: ALY 6010

|  |  |
| --- | --- |
| Fig 16: For this bar plot, we considered the year 2007 and chose only European countries to show their population rate.  Based on the visualization, Germany was having 82 million population with life expectancy of 79 years, followed by Turkey with 71 million and 71 years of life expectancy. Countries like Slovenia and Albania had the population below 4 million with 76 and 77 years of life expectancy. | *Fig16: Population of Countries in Europe (2007)* |

|  |  |
| --- | --- |
| Chart, scatter chart  Description automatically generated  *Fig17: Average life expectancy and GDP for Asian countries in each year* | Fig 17: To find the average life expectancy and GDP (not GDP per capita) for Asian countries for each year, the mean value was calculated as below which shows Europe had the highest GDP per capita among all. |

|  |  |
| --- | --- |
| Fig 18: Pie char was considered to show the broder picture of populations in Asian countries and based on the visualization report, China occupied 34.6% of the total population and India had 29.1 %. Bahrain was having the least i.e., 0.05% population in the Asian continent. | *Fig18: Population of Countries in Asia (Year 2007)* |

Abhilash Kumar Dikshit

R Practice: ALY 6010

|  |  |
| --- | --- |
| Fig 19: Scatter plot was considered to show the GDP per capita income for individual countries. When we hover over the dots, we can clearly identify Kuwait was having the highest GDP per capita as shown here. | Chart  Description automatically generated  *Fig19: GDP per Capita for Individual Countries* |

|  |  |
| --- | --- |
| Chart, bar chart  Description automatically generated  *Fig20: Life Expectancy of Population* | Fig 20: Here the bar plot is used to showcase the Life expectancy for each continent based on their population. As the data shows, Asia is clearly winning in terms of population and if we look closely at the stack of life expectancy, the expectancy rate is also high. For more accurate analysis refer Fig 11 with box plot representation. |

**References:**

1. *GitHub: Datasets Used in Plotly Examples and Documentation* ([November 5, 2022](http://127.0.0.1:64200/rmd_output/4/#ref-R-Kaggle));
2. *Descriptive Statistics and Graphics- Easy Guides* ([October 28, 2022](http://127.0.0.1:64200/rmd_output/4/#ref-R-Material)); Kervizic ([Feb 9, 2020](http://127.0.0.1:64200/rmd_output/4/#ref-R-Career));
3. *Data Visualization Tips for More Effective and Engaging Design*. n.d. Tableu. <https://www.tableau.com/learn/articles/data-visualization-tips>.
4. *Descriptive Statistics and Graphics- Easy Guides*. October 28, 2022. <http://www.sthda.com/english/wiki/descriptive-statistics-and-graphics>.
5. Kabacoff, Robert I. May 2015. *R in Action, Second Edition*. <https://www.manning.com/books/r-in-action-second-edition>.
6. Kervizic, Julien. Feb 9, 2020. *Data Science Career Path & Progression*. <https://medium.com/analytics-and-data/data-science-career-path-progression-ab5140cfbc84>.
7. *Mapping in r: Latitude and Longitude Coordinates*. October 29, 2019. <https://map-rfun.library.duke.edu/01_georeference.html>.