Project 2023

- Part 1 Visualisation [8 marks]

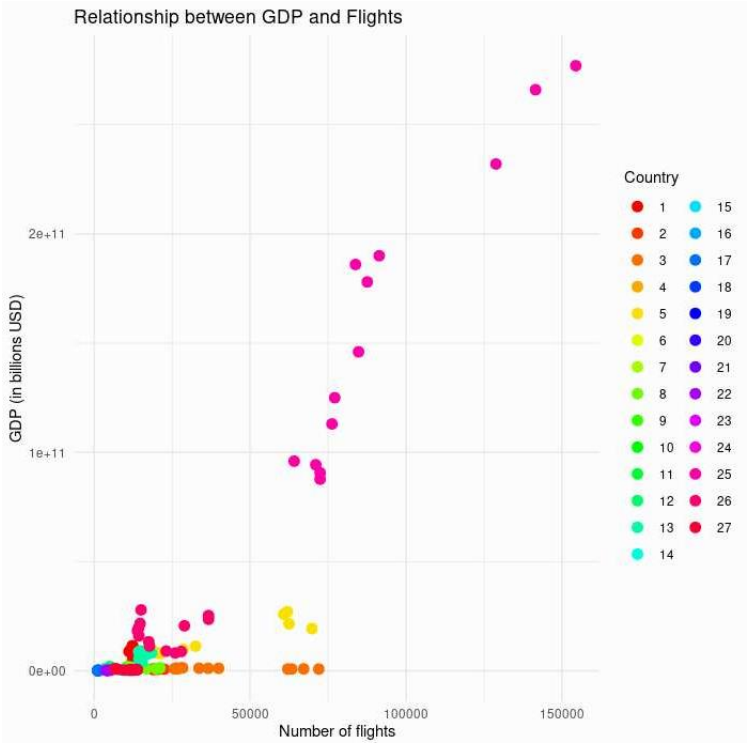
Story: The story of the importance of the tourism industry for the economic growth of island nations. Does tourism affect the island nation's economy?

Pre-processing:

- I first decided to replace empty data values with NA (an alternative I found you can use as. E.g., numeric and it changes it to NA by coercion)
- Then renamed the flights - WB column to flights (make the attribute name nice)
- as factor to necessary columns -> convert all character columns
- check for hidden spaces in strings.
- Also experimented with "data <- data[complete.cases(data),]" and "data[is.na(data)] <- 0
- Later learned and experimented with "dplyr" (filtering out)

Questions:

- How has the tourism industry impacted the economic growth of island nations over time?



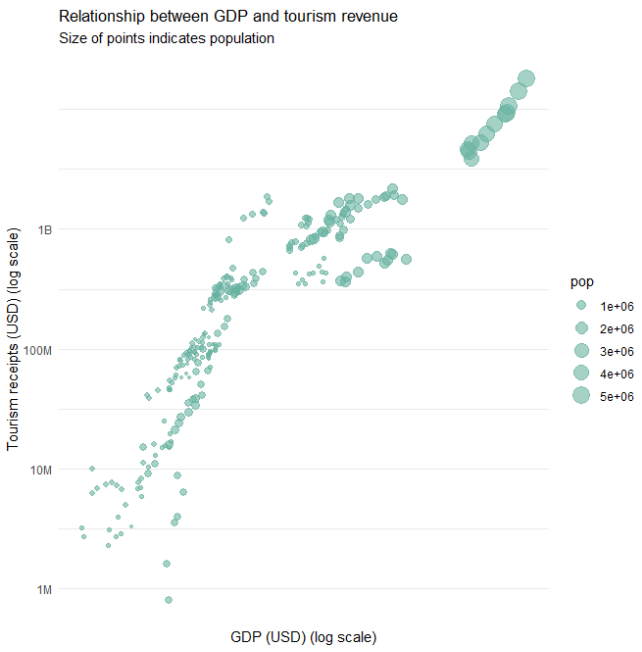
One way to explore the impact of tourism on economic growth could be to create a scatterplot of GDP against maybe tourism popularity for each island nation. This would allow us to see if there is a correlation between how popular a tourist destination is and its economic growth over time. Colour coding to represent different regions of the world. In this case it is a scatterplot of GDP vs. flights, with the colour of each point determined by the country. This allows us to differentiate between countries in the plot and see if there are any trends or patterns in the relationship between GDP and flights.

The scatterplot of GDP against tourism popularity, represented by the number of flights, is a good way to explore the impact of tourism on economic growth. The use of colour coding to differentiate between countries and regions of the world is also a useful visualization technique.

However, it is important to note that the number of flights alone may not be the best indicator of tourism popularity. Other factors, such as the number of international tourist arrivals or hotel occupancy rates, could also provide valuable insights into the relationship between tourism and economic growth. Additionally, it would be helpful to explore the data over a longer period of time to identify any trends or patterns in the relationship between GDP and tourism. Furthermore, while the use of colour coding can be useful in differentiating between countries and regions, it may also

be helpful to explore other visualization techniques that can reveal more detailed information. For example, a heat map or choropleth map could provide a visual representation of the distribution of GDP and tourism across different countries and regions.

- How does the economic situation of island nations affect their tourism industry, and are there any trends in terms of GDP and tourism revenue?



Although I managed to plot this to a large scale due to time constraints, I could not quite work out the exact limits to set so I used arbitrary ones, thus the x axis. In terms of GDP might be obscured or down-played due to your visualisation design. I also decided to experiment and use “dplyr” to filter out missing values, that I renamed to NA. This has had a clear impact on the graph as there are not many outliers.

The x-axis represents GDP, and the y-axis represents total tourism receipts. Each point on the graph represents a different country, and the size of the point represents the population or land area of the country. This visualisation helps to show that there is a positive relationship between GDP and tourism revenue, as well as some outliers or exceptions to this trend. The data is presented in a way that is easy to read and understand, with clear and concise labels that help to convey the key insights.

The use of a scatterplot to explore the relationship between GDP and tourism receipts is a good visualization technique, and the inclusion of country population as a size variable adds an extra layer of information. However, as noted, the arbitrary limits on the x-axis may obscure or downplay certain aspects of the data. It would be helpful to carefully choose appropriate limits for the x-axis to better represent the range of GDP values in the data.

Additionally, the use of "dplyr" to filter out missing values may have a clear impact on the graph by reducing the number of outliers, but it is important to consider the potential impact of this data cleaning step on the overall analysis. Careful consideration must be given to the reasons why certain data may be missing and whether excluding these values could potentially bias the results.

Overall, the visualisation effectively conveys the positive relationship between GDP and tourism receipts, but as with any visualisation, it is important to carefully consider design choices and potential biases that may impact the interpretation of the data.