



Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

CS7DS4 / CSU44065 Data Visualization 2023-24

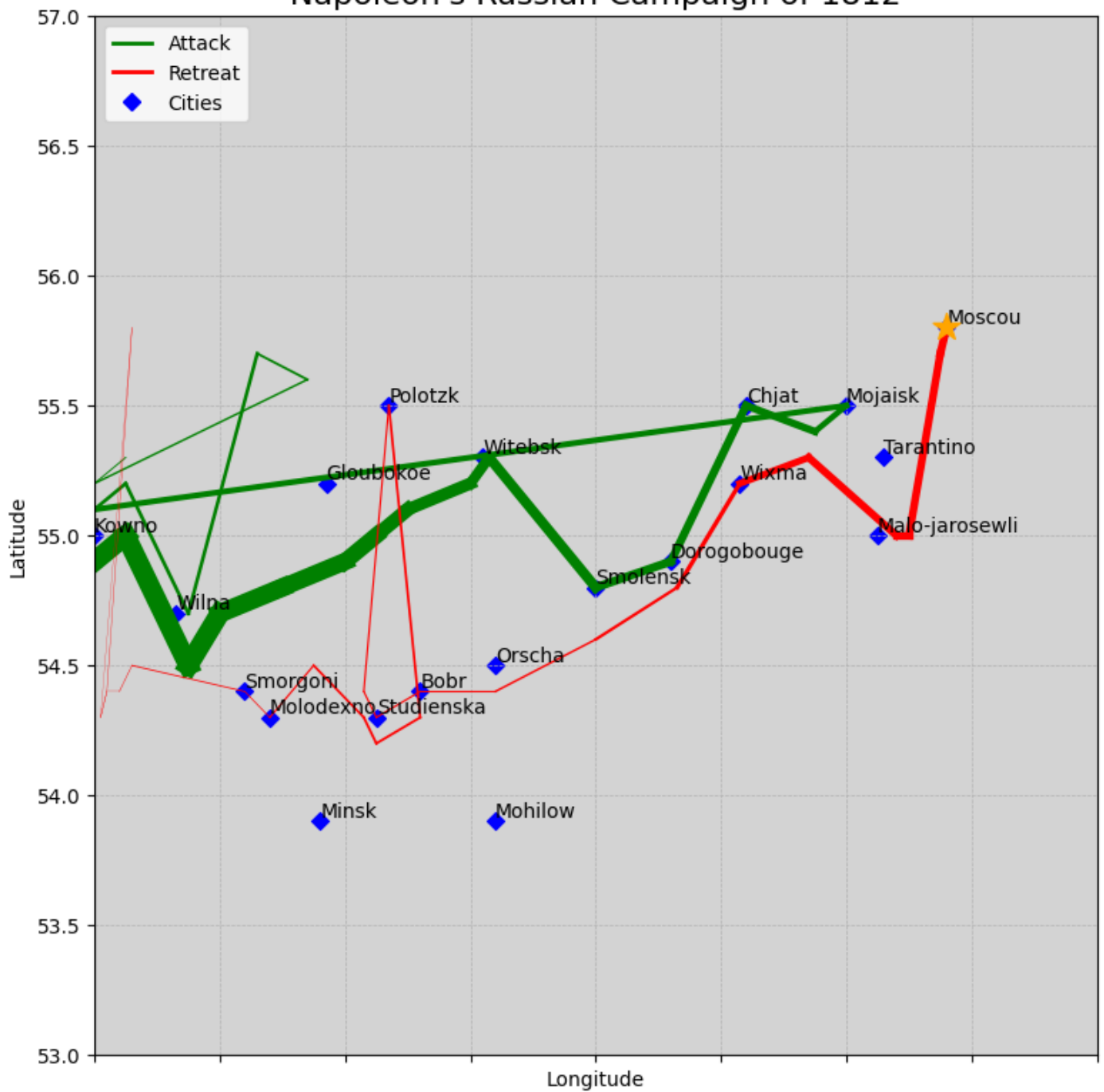
Assignment 1.1 & 1.2

Siddharth Lala

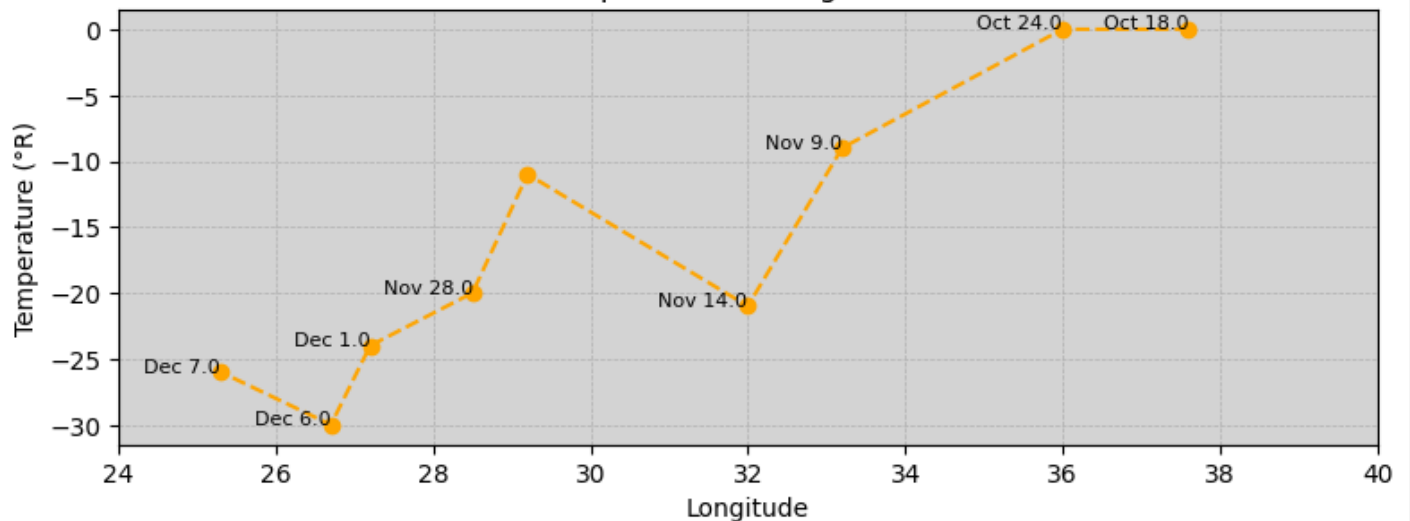
20315464

DECLARATION: I understand that this is an **individual** assessment and that collaboration is not permitted. I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at <http://www.tcd.ie/calendar>. I understand that by returning this declaration with my work, I am agreeing with the above statement.

Napoleon's Russian Campaign of 1812



Temperature During Retreat



To construct Charles Minard's famous visual display of Napoleon's Russian title charge, I used python with matplotlib libraries. Additionally I used pandas to read and scale the data and to load the various excel files. To make the Minard-data.xlsx file more convenient to use, I divided it into 3 different excel workbooks.

The first subplot showcases the path of Napoleon's army during the attacking and retreating phases. The line width encodes the relative size of the army at each stage, showing its decline as the campaign progresses. The Russian cities are named on the chart and are also plotted as blue diamonds. Since Moskow is of the highest significance as compared to the other cities on this chart, I highlighted the city with an orange star.

The second subplot of the same chart depicts the extreme weather conditions faced by the soldiers during their retreating phase. The line plot contains temperature points with corresponding dates to show the correlation between the retreating army (100,000 to 6,000) and the extremely harsh weather conditions.

Since the Minard data.xlsx file only had longitude coordinates (LONT) for the days the temperature was recorded on, I used a common x axis for both plots. Both these sub plots consistently use a common x-axis that is longitude which is an interesting addition to the visual appeal of the chart as it makes it scalable.

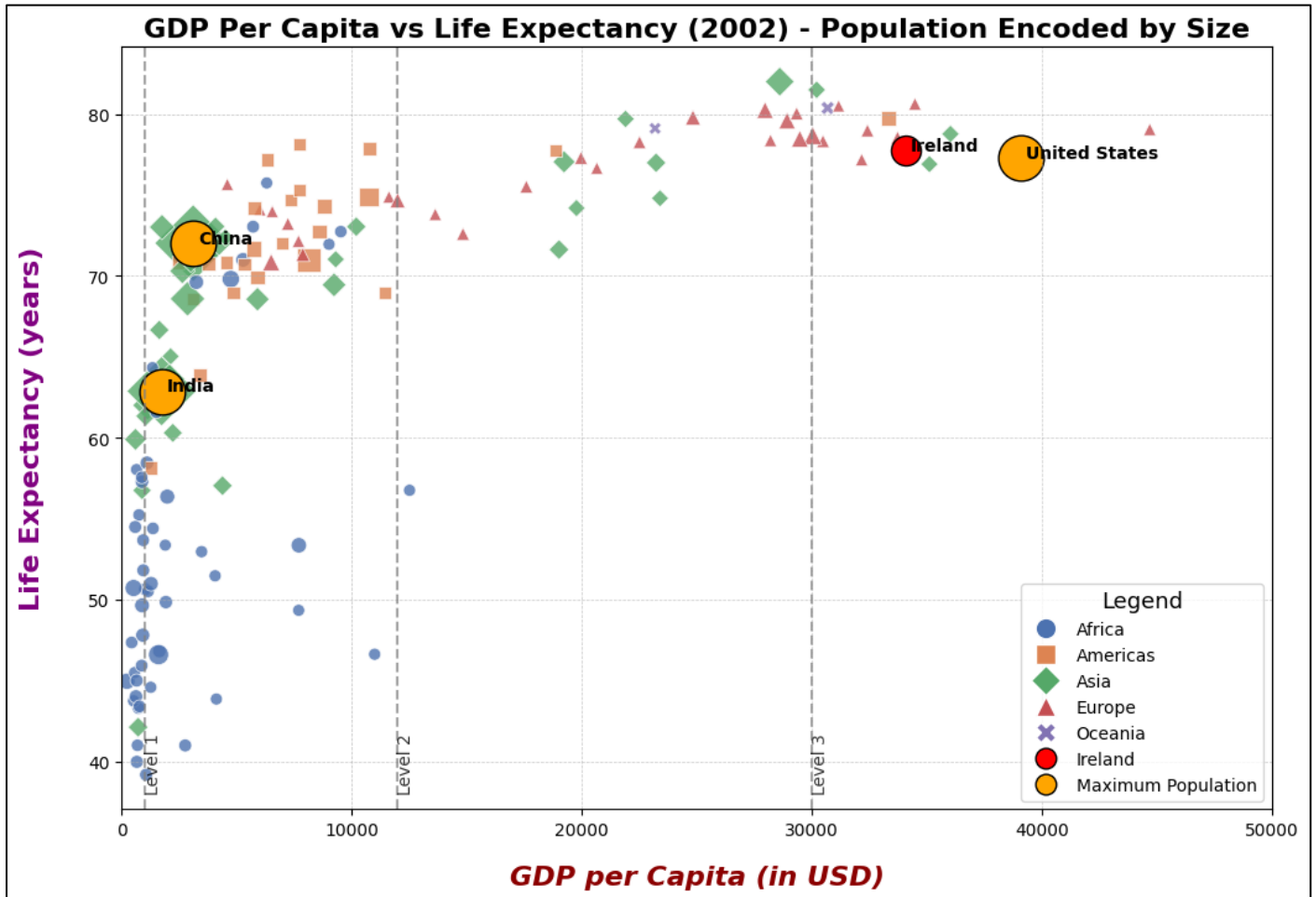
I set the background color of the visualisation to light gray to improve the contrast of the chart. I also included grid lines to help the reader to understand the temperature and army data better. By encoding the army size according to the line width, it exhibits the tragic scale of losses.

In my opinion, this visualisation provides a clear and irrefutable depiction of Napoleon's campaign. It captures his failures as well as the harsh conditions that largely contributed to his defeat. Further improvements could be made to improve readability in packed areas.

Assignment 1.2

Part A: Visualize the correlation between wealth and health in the year 2002

Chart #1 – Population encoded by size



I used google colab as the software to generate all these charts and I also used matplotlib and seaborn as the main tools. I used marker size to encode the population which helps to communicate visually and clearly about which countries are more populous.

The addition of colour intensity for GDP provides added information to give a more comprehensive idea of the data. Chart #1 offers a clear understanding of the countries having the highest populations. Chart #2 displays a more balanced comparison across all population sizes. Chart #3 may be tougher for viewers to interpret due to its opacity method.

Chart #2- Population encoded by color intensity

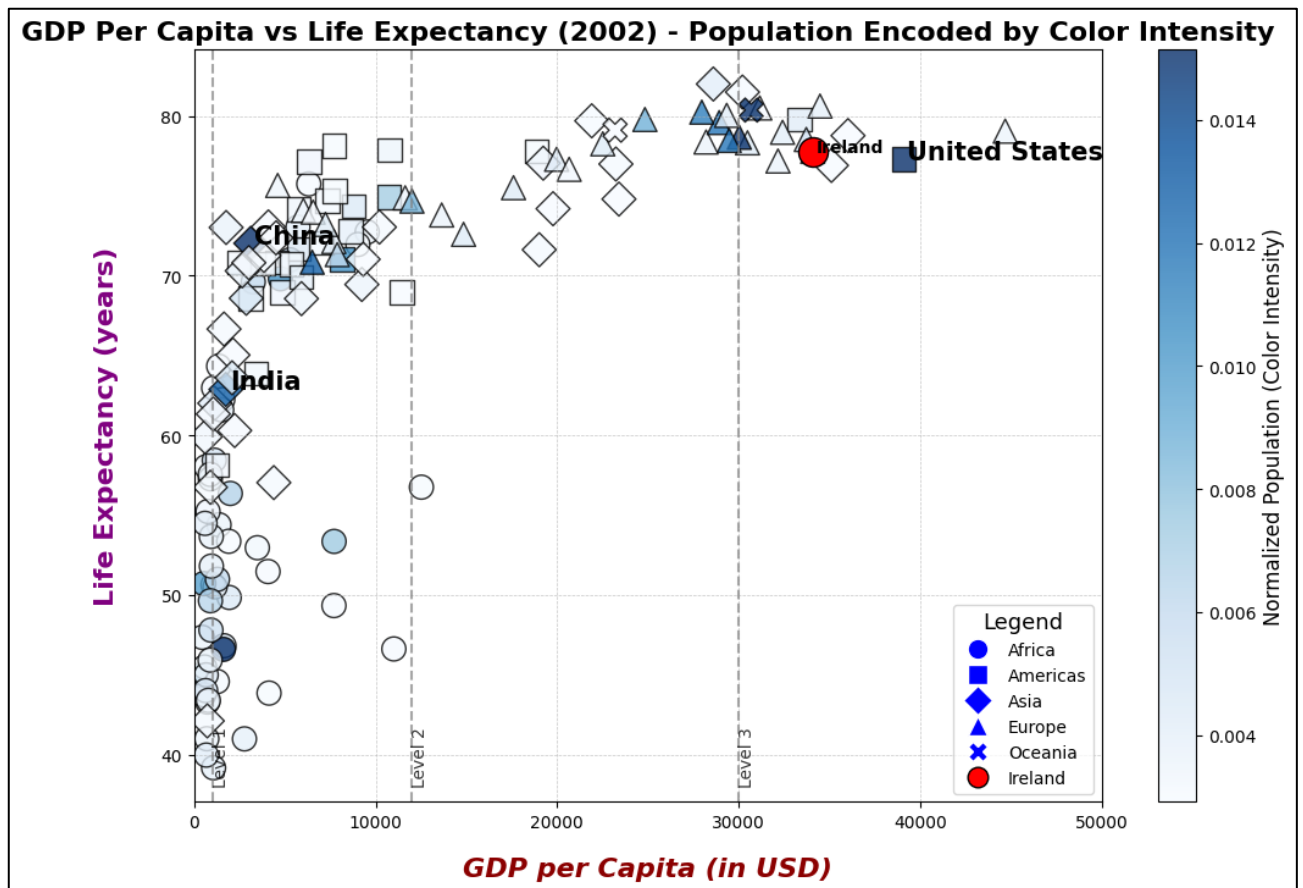
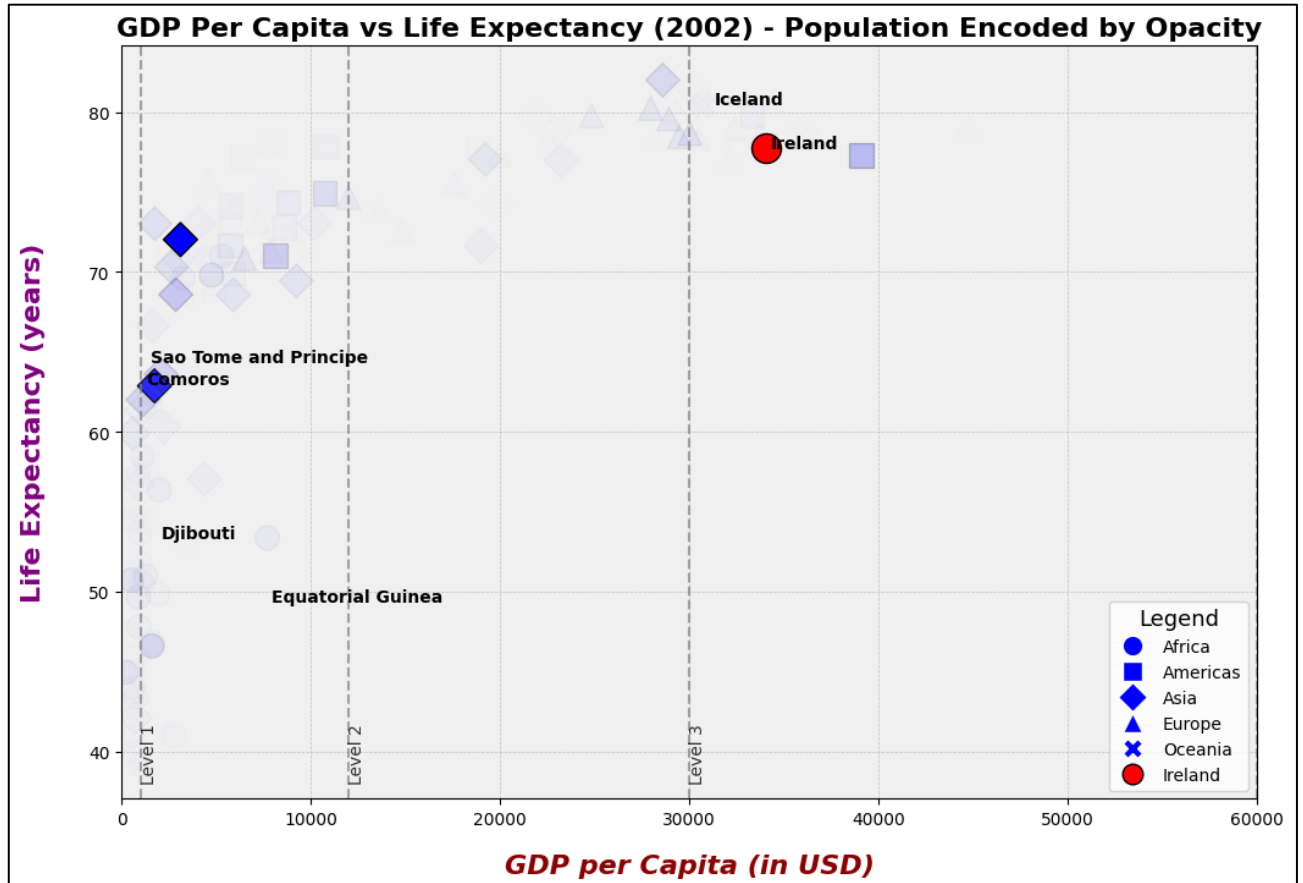
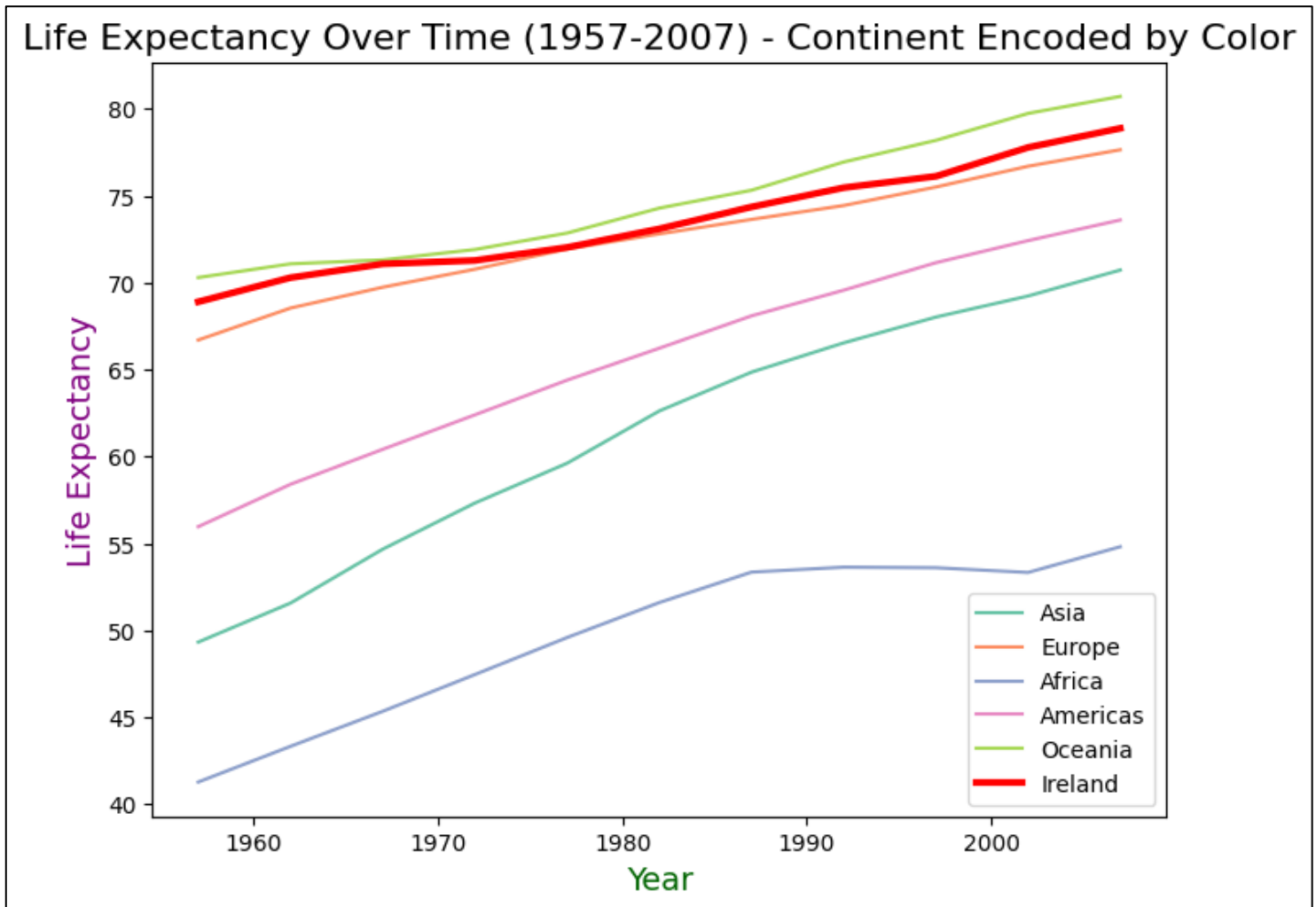


Chart #3- Population encoded by Opacity



Part B- Visualize the evolution of life expectancy over the years (1957-2007)

Chart #4 – Continent coded by color



Encoding by line type could become tougher to visualise if the line types are not sufficiently varied. Hence encoding by color type is the most visually intuitive. Chart #4 is effective to compare regional trends. Chart #5 uses line type coding. Chart #6 uses shape encoding which offers additional cues.

Chart #5 – Continent coded by line type

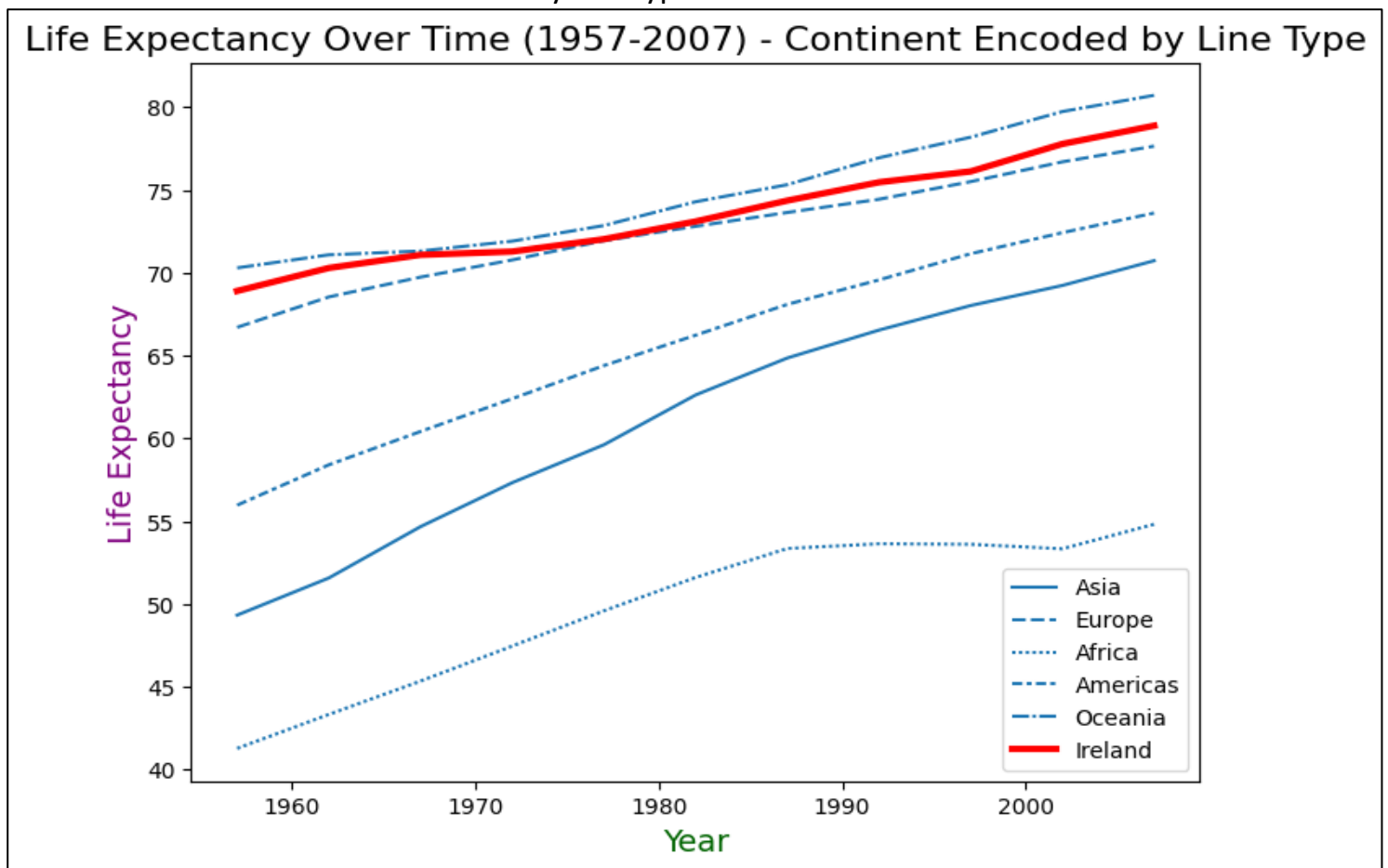
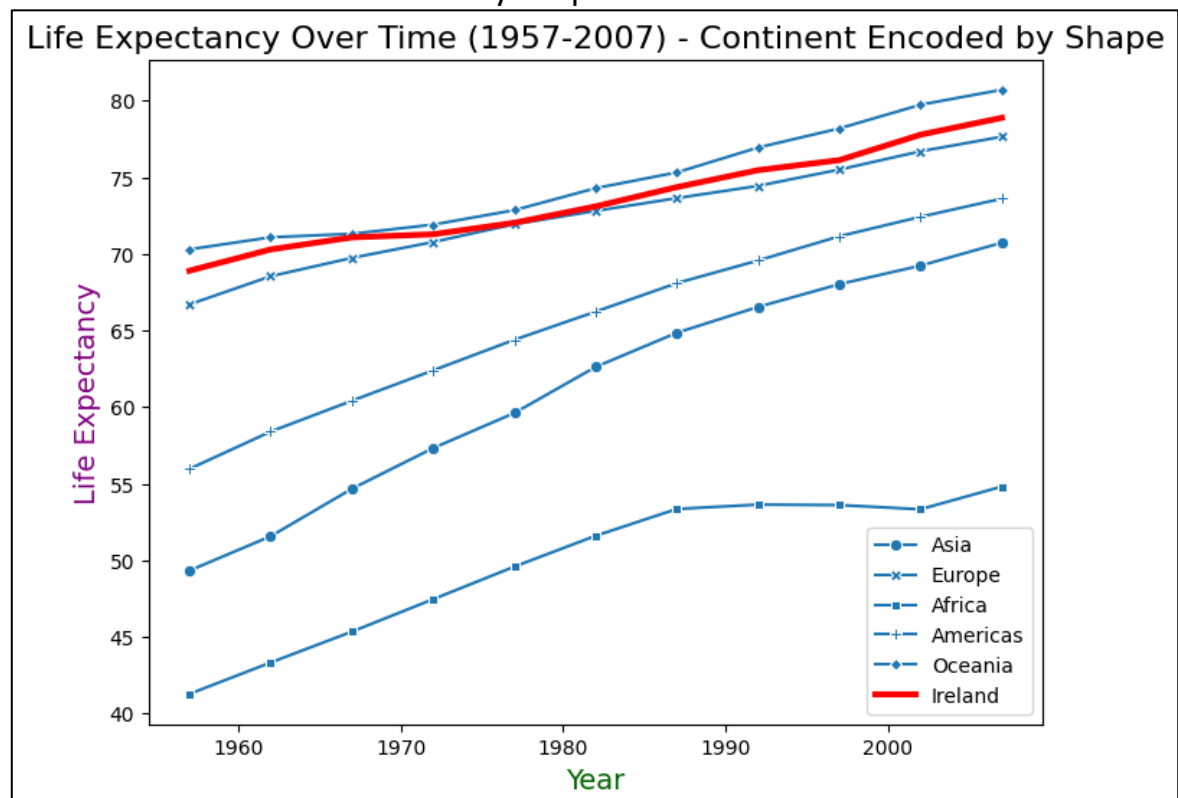
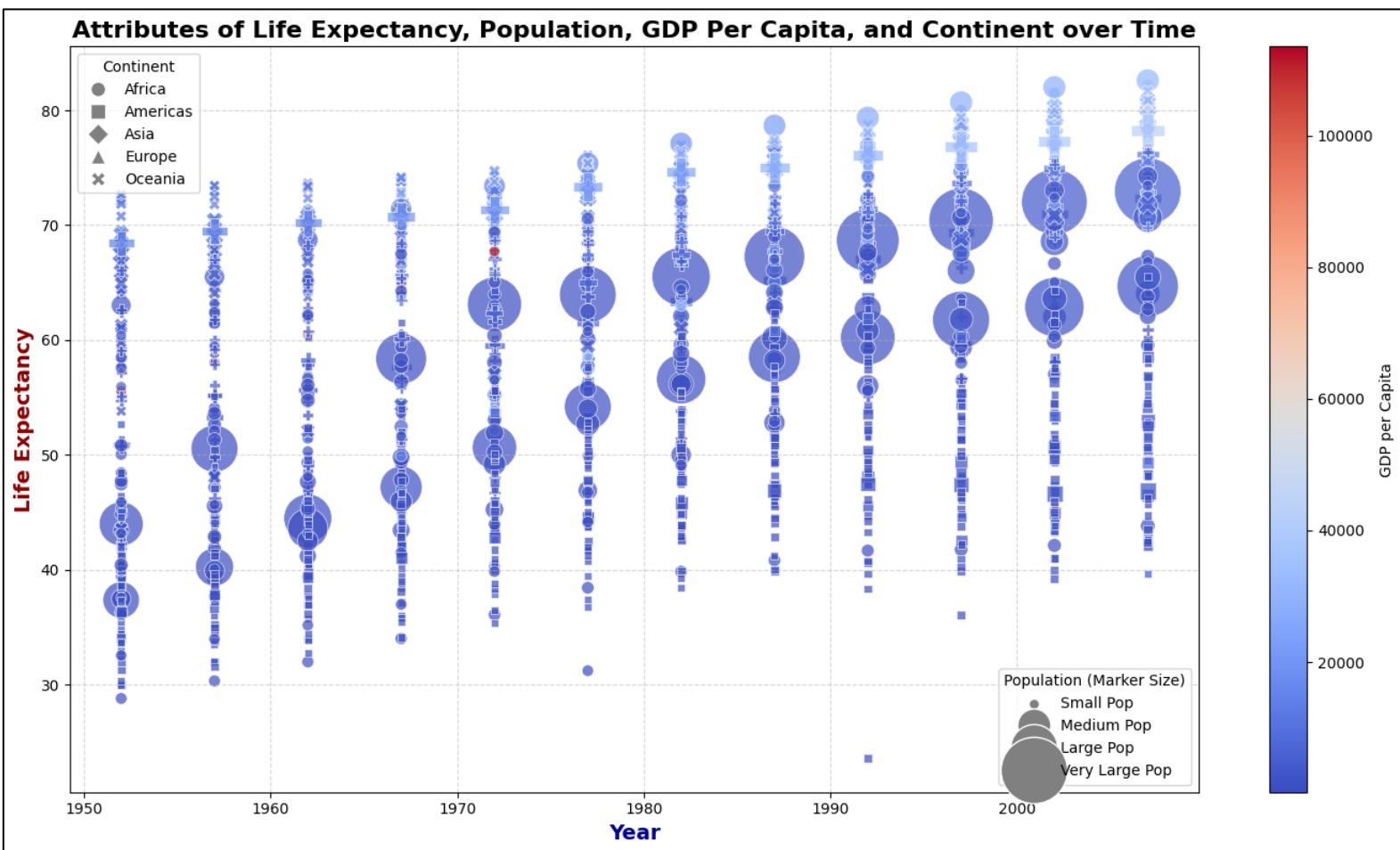


Chart #6 – Continent encoded by shape



Part C – Exploratory visualisation of the full data set
Chart #7



Chart#7 uses multiple encoding channels like marker shape, marker size and color gradient. Each continent is represented by a different shape. The population is encoded by marker size where the larger markers represent countries having higher populations. The GDP per capita is encoded using a color gradient.

The main drawback of this chart is that since there is so much data packed into a single chart, the viewer may have difficulties in interpreting the chart.