Final Documentation

Design Process:

      Our group decided to use the countries dataset from the list of recommended data sets in assignment two. This dataset contained 263 data objects and 45 attributes. The data objects were the countries and the attributes were all quantitative data. The attributes covered economic, health, industry, and societal factors for each country. Because there were so many attributes, may countries were missing values. To clean up the dataset, we decided to narrow down the attribute selection to something more manageable. We ended up selecting three economic attributes and three biological attributes to identify relationships in the data. The attributes selected were Birth Rate, Death Rate, Life Expectancy, GDP Per Capita, Unemployment Rate, and Public Debt. Further, we removed any country will null values for any of these attributes, leaving us with 102 countries to use. From here, we started to explore different ways to answer the question: Is there a relationship between biological factors and economic factors within the dataset.Our first idea for answering this question was to create an interactive world map paired with a scatter plot that would be used to show the relationships in the data. The map would have used colors and gradients to represent the different attributes on the map, while the scatter plot would show the relationship between the attributes being shown on the map. After doing some research on how to make all of those ideas happen, we decided that we needed to scale back our idea for the sake of time. We decided to create an interactive scatter plot that would allow us to identify the relationships in the attributes we had chosen.

Design Rationale:

      Each dot on the the scatter plot represents a different country, during the early iterations of the scatter plot, each dot was assigned a different color. With there being a total of 102 countries being displayed on the scatter plot, using different colors as a way to identify each country would decrease the effectiveness of the overall chart, so we made the design decision to only use the color orange for the dots, to simplify how each country is identified. To allow a user to view each country a tooltip was added to each dot, allowing the viewer to hover their mouse over a dot to display, the name of the country, the two attributes being compared on the chart, and the values the dot represents. Along with the tooltip appearing, the dot that the user hovers over will grow in size so the user can easily identify the exact dot they are looking at. Viewers of the visualization can make comparisons between seven attributes. Country is the main attribute within the visualization, so users do not decide on what countries are seen on the visualization. Viewers are able to manipulate which biological and economic attributes they want to compare by manipulating the 12 buttons encoded on the interface. There are six attributes on the x axis and six attributes y axis, each axis has the same six attributes. We decided to encode the same six attributes on both the y and x axis to give the user more freedom of comparison. We wanted to viewer to able to make as many comparisons as they wanted with the selected six attributes. Formatting the buttons this way allows the user to make comparisons of

biological v. economic attributes, economic v. economic attributes, and biological v. biological attributes. Giving this freedom, viewers of the visualization will be able to get a better understanding of the dataset.

Insights Gained:
Our final scatter plot shows unique relationships between the economic and biological attributes of a country. In our comparisons, we placed biological factors on the x axis and economic factors on the y axis. The results below show the relationships we were able to visually identify in the scatter plot.

Positive Correlation:
Birth Rate v. Unemployment Rate
Life Expectancy v. GDP Per Capita

Negative Correlation:
Birth Rate v. GDP Per Capita
Life Expectancy v. Unemployment Rate
Death Rate v. GDP Per Capita
Death Rate v. Unemployment Rate
Life Expectancy v. Public Debt

No Relationship:
Death Rate v. Public Debt
Birth Rate v. Public Debt

As you can see, the majority of the economic and biological attributes had a negative correlation. Looking at the biological attributes, we noticed that when birth rate increased, the GDP Per Capita decreased, and the unemployment increased. These relationships show us that people who live in countries with higher birth rates usually have less money and are more likely to be unemployed. When Death Rate increased, we noticed that GDP Per Capita and the Unemployment Rate decreased. This tells us that people who live in a country with a high death rate are more likely to have less money and also more likely to be employed. Lastly, we observed that an increase in life expectancy showed an increase in GDP Per Capita, and a decrease in Unemployment and Public Debt. This means that people who live in countries with a higher life expectancy generally have more money, are more likely to be employed, and hold less public debt. This analysis of the scatter plot shows us that there is a relationship between biological and economic attributes in the data set.