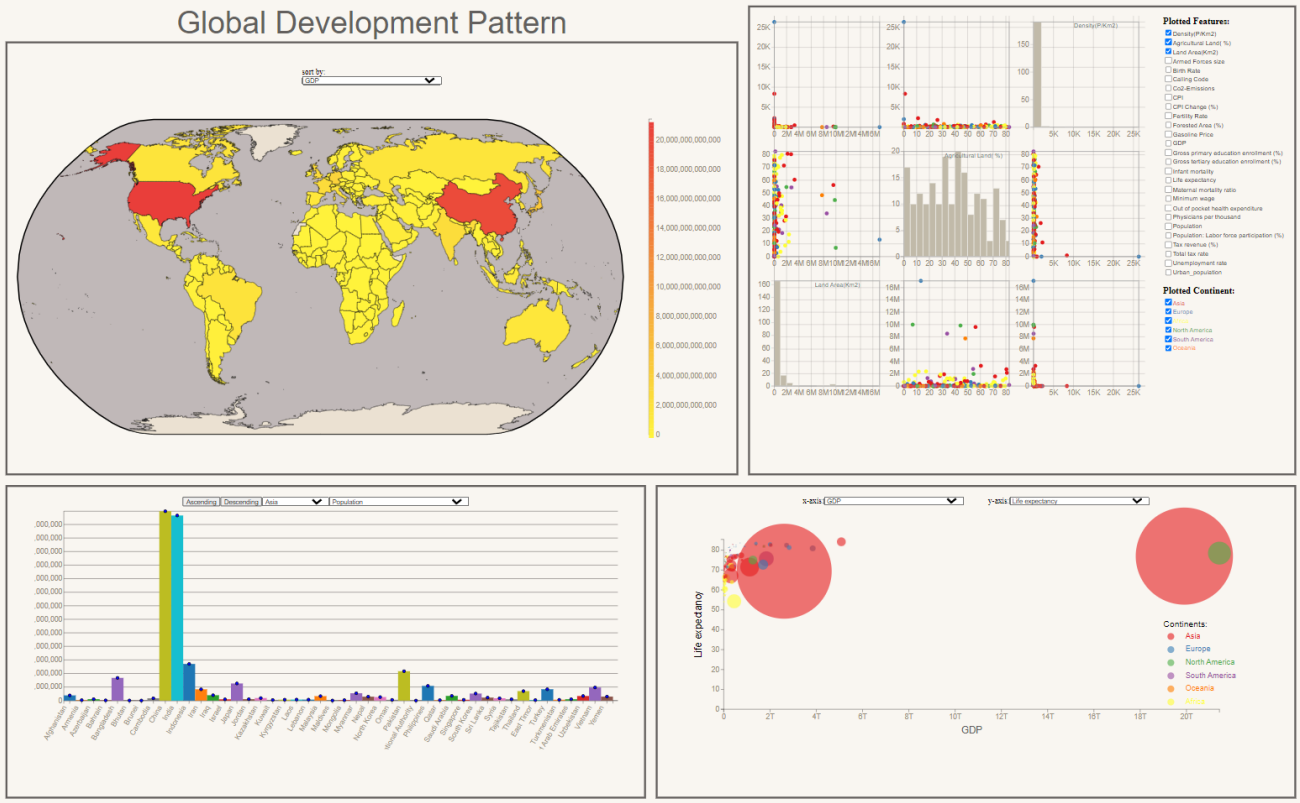
Data Visualization Final Project Report

李嘉玲 *(109550186)*, 梁巍濤 *(109550187)*, 林家輝 *(109550201)*

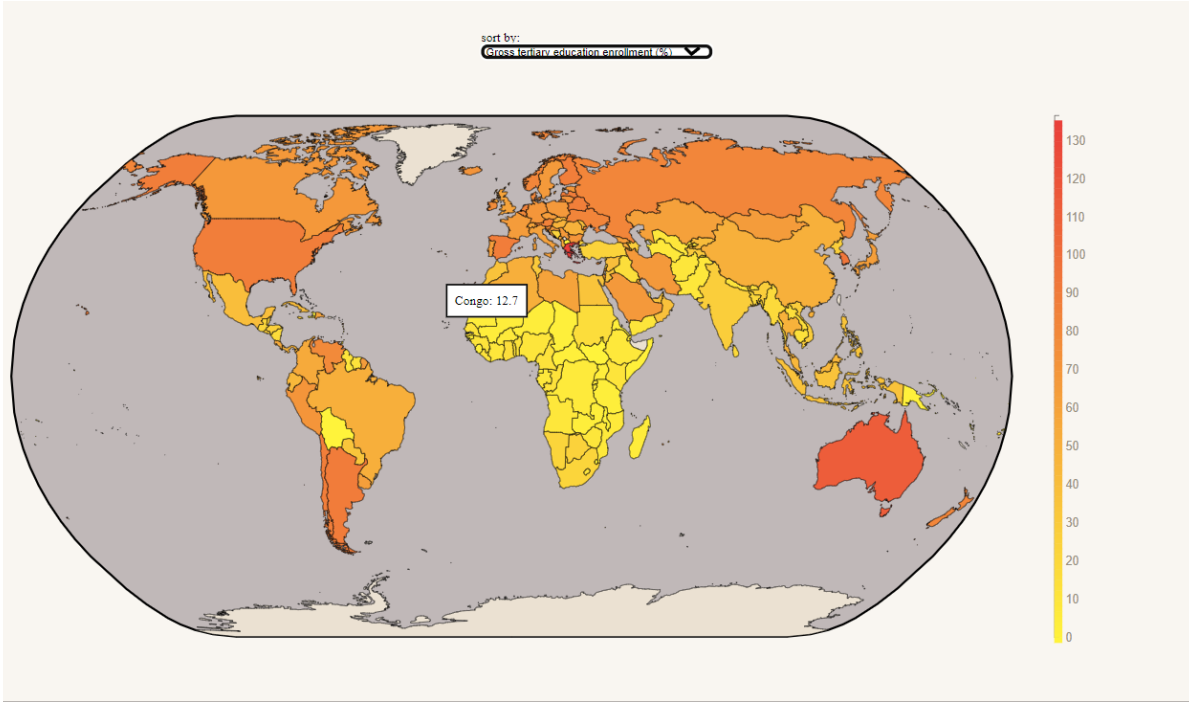
For this final project, we used a dataset that consists of detailed information for each country. The dataset consists of 35 columns providing vital information and giving insights ranging all the way from fertility-mortality rates to the GDP of 193 countries in this world. Our final project consists of 4 different visualizations, which are choropleth map, scatter plot matrix, bubble chart, and bar chart. Each of the visualizations are interactive and are designed to showcase important statistics of the countries in this world. For example, our bar chart can clearly showcase the difference of each feature between countries in either descending or ascending order. Whereas using a scatter plot matrix can show us a visualization of correlation between the life expectancy and physicians per thousand in a country.



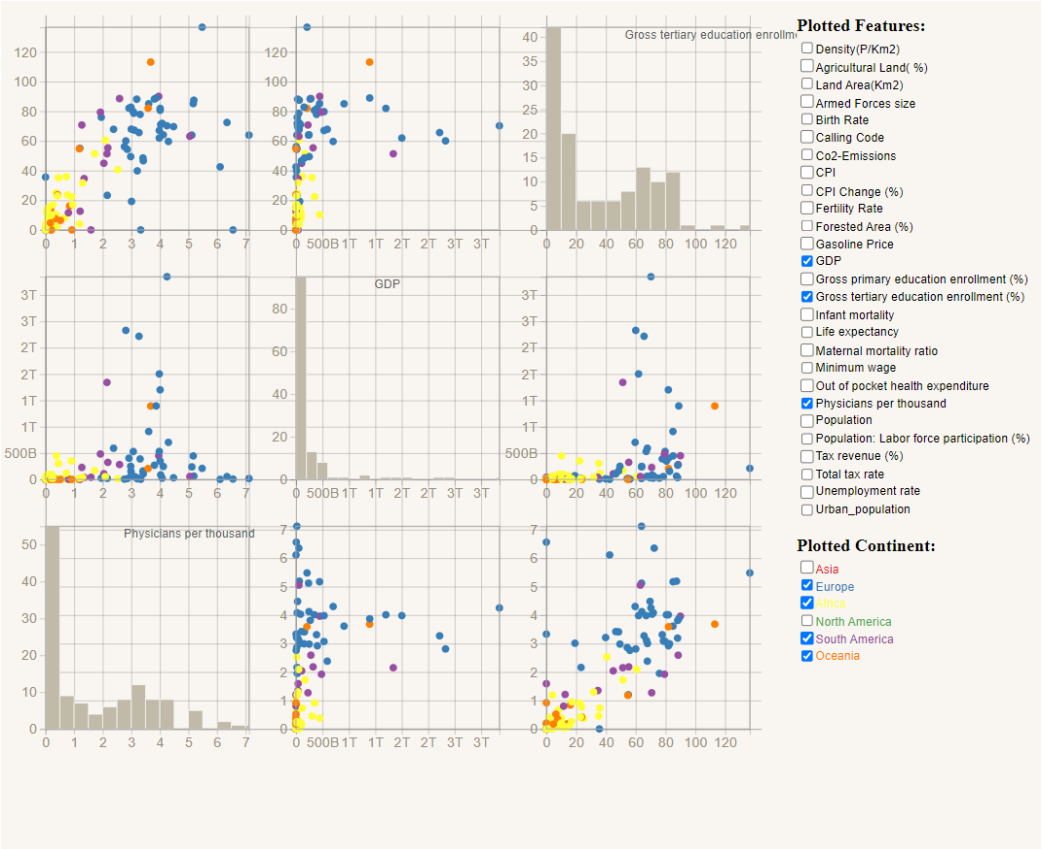
Through this visualization system, we aim to provide educational information that can demonstrate the development patterns in each country from different aspects. For example, researchers studying international relations, economics and sociology can

use this system to compare different metrics between countries. Companies considering expanding globally can assess their potential markets by understanding the various socio-economic factors in different countries.

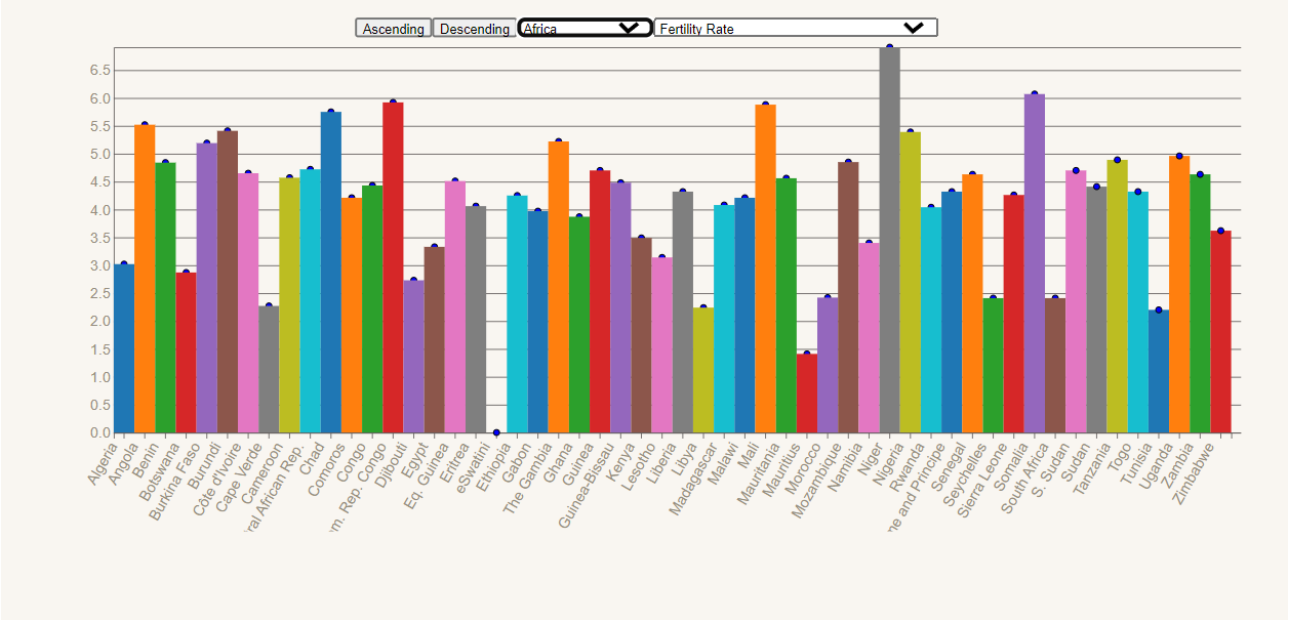
Before proceeding with the visualizations, data preprocessing is one of the most crucial steps to produce clean and neat data. In this stage, we added a new column for country regions [‘Continent’] with the help of outside resources. The newly added column helps in visualizing and classifying the countries into regions. Other preprocessing include modifications of country names, numerical value hashing, and the update of latitude and longitude.



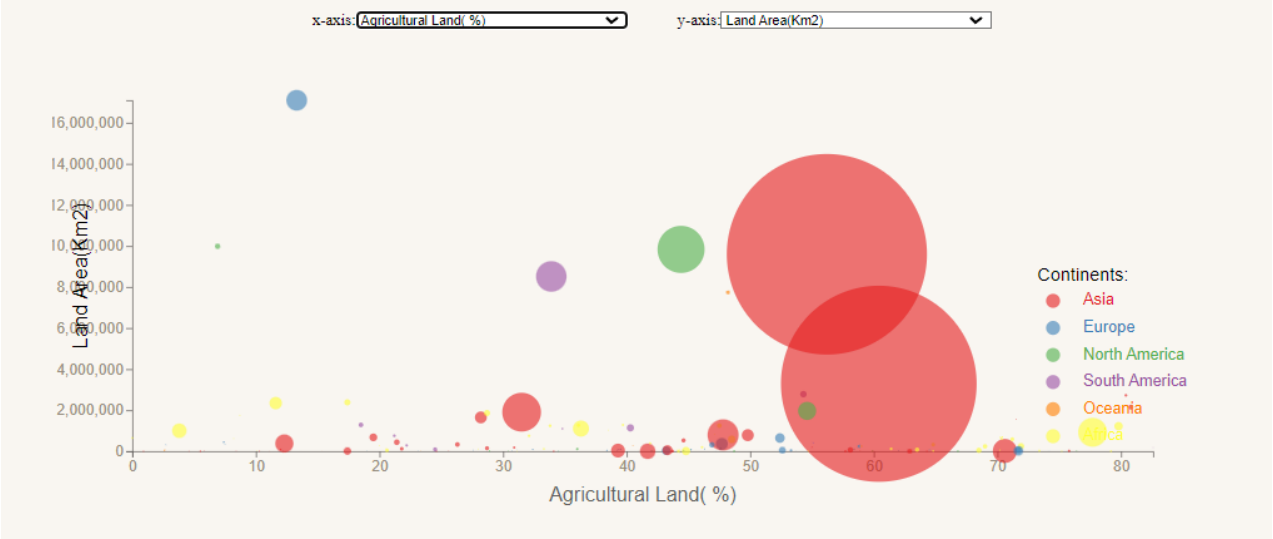
The first visualization is the choropleth map, which shows the world map and color gradients from red to yellow that indicate the difference in each metric of the countries. Users can interact with the system and choose the metric by drop down list. This choropleth map can be easily understood through the help of tooltips when users hover through a certain country, which then shows the country name and its respected metric chosen in the drop down list.



Next is the brushable scatter plot matrix, which shows the correlation between different metrics through their coordinates in each scatter plot. The checkboxes located in the right side of the visualization gives users the freedom to choose which metrics they would like to include in the matrices and what regions to include. The brushable feature also enables them to identify the brushed coordinates in other plots, which shows the relationships between multiple indicators simultaneously.



Following is the bar chart, which consists of four interactive features: two buttons to sort the data displayed into ascending or descending order, and two dropdown menus to choose what region and values of the countries shall be presented. Users are able to experiment and find out the ranks of each value for each region; for example, one can pick the 'GDP' value & region 'Asia' in 'Descending' order to see China as the furthermost left bar, followed by Japan, India, South Korea, and so on.



Lastly, there is a bubble chart with adjustable x and y axes. These axes can be set through the drop down list to visualize the relationship between two metrics. The bubble color differs between each region, positioned by their values in the axes, and size that shows the country's population size. Therefore, it results in a large number of smaller bubbles compared to several gigantic bubbles. For the ease of reading data, tooltips are included to make out the country's name, region, and both axes values. This visualization mainly focuses on the population size and how several countries have a tendency to form a cluster due to similar values and patterns.

Our visualization system with several visualizations combined in one page gives users the ease of reading and analyzing the data. Through the visualizations that we provided, relationships between attributes ranging from geographical, demographic to socio-economic can be effortlessly studied.