CALIFORNIA STATE UNIVERSITY, FRESNO
DEPARTMENT OF COMPUTER SCIENCE

February 16, 2020

Class:	Big Data Analytic (CSCI 191T)		Semester:	Spring 2020
Points		Document author:	Madhuri Pyreddy	
		Author's email:	madhuripyreddy @mail.fresnostate.edu	
		Assignment number:	1	

1 Statement of Objectives

The objective of this assignment is about visualizing and analyzing data from over 4 different data sets. When we visualize data, we look for unique trends, which helps you to identify the relationship between the independent variable (x axis) and the dependent variable (y axis). Other factors involved also help contribute to the relationship between since they can determine whether it will have positive growth or negative growth. Based on these factors, we can make conclusions based off the knowledge we gather from the data sets itself. It helps us to better understand what we are viewing from the graphs.

1.1 Data Set 1:- Literacy Rate, adult total (percent of people ages 15 and above)

The first data set has 20 rows of countries and 38 columns of the percentage of people age 15 and above who are literate from the years 1975-2011. In this data set, I am visualizing and analyzing the data to see which country has the highest literacy rate and the lowest literacy rate in the year 2011. Also, I want to see the overall change in literacy rate from 1975-2011.

1.2 Data Set 2:- Income per person (GDP/capital, PPP inflation based)

The second data set has 20 rows of different countries and 242 columns of how much GDP there is in each of the different countries from the years 1800-2017. In this data set, I am visualizing and analyzing the data to see which country has the highest income per person(GDP) and the lowest income per person (GDP),in the year 2017. Also, I want to see the overall change in literacy rate from 1800-2017.

1.3 Data Set 3:- Exports (percent of GDP)

The third data set has 20 rows of different countries and 59 columns of how many exports they exported out of their countries from the years 1960-2017. In this data set, I am visualizing and analyzing the data to see which country has the highest exports (percent of GDP) and the lowest exports (percent of GDP),in the year 2017. Also, I want to see the overall change in literacy rate from 1960-2017.

1.4 Data Set 4:- CO2 emissions per person

The fourth data set has 20 rows of different countries and 216 columns of carbon dioxide emissions per person in each of the countries from the years 1800-2018. In this data set, I am visualizing and analyzing the data to see which country has the highest carbon dioxide emissions per person and the lowest carbon dioxide emissions per person,in the year 2018. Also, I want to see the overall change in literacy rate from 1800-2018.

2 Experimental Procedure

For this report, I navigated through https://www.gapminder.org and opened the tools to look for the data. Then, I searched for the data sets that I was doing my report on and In this report, I looked at the data sets from the state of California website to get a better understanding of what to expect from them. Then, I searched the different data sets, using gapminder for the visualization. In gapminder, we we are going over 4 different graphs from 4 different data sets. Each of the four sets have an x axis and y axis, along with different variables that influence how the graph moves. For instance, a linear, bubble, trend, graph can have positive growth because as the numbers for independent variables continues to grow, so does the numbers for dependent variables.

2.1 Procedure

In this procedure, I need to find the information for visualizing and downloading data.

A) Literacy Rate, adult total (percent of people ages 15 and above

1) Go to https://www.gapminder.org/
2) Go to Data->Search
3) Search for the data and click "Literacy Rate, adult total (percent of people ages 15 and up)
4) Click the link to download the data (.csv) on Excel.
5) Click visualize to display data

B) Income per Person(GDP/capital, PPP inflation based)

1) Go to https://www.gapminder.org/
2) Go to Data->Search
3) Search for the data and click "GDP/capital, PPP inflation based")
4) Click the link to download the data (.csv) on Excel.
5) Click visualize to display data

C) Exports (percent of GDP)

1) Go to https://www.gapminder.org/
2) Go to Data->Search
3) Search for the data and click "Exports (percent of GDP)")
4) Click the link to download the data (.csv) on Excel.
5) Click visualize to display data

D) CO2 emissions per person

1) Go to https://www.gapminder.org/
2) Go to Data->Search
3) Search for the data and click "CO2 emissions per person")
4) Click the link to download the data (.csv) on Excel.
5) Click visualize to display data

3 Analysis

Literacy Rate The map graph displays countries that have a percentage of how many adults 15 and above can read from 1975-2011. It includes different countries like Afghanistan, Albania, Bangladesh, Angola, Ecuador, Guatemala, and Trinidad and Tobago. In 1975, Afghanistan had a literacy rate of 11 percent. As the years increased by 2011, Afghanistan increased its literacy rate to 99.7 percent. In Albania, literacy rate was shown at 9 percent in 1975. In 2011, Albania's literacy rate jumped to 95.4 percent. In 2003, Angola's literacy rate was at 68 percent. In 2011, Angola's literacy rate rose to 70.4 percent. From 1999-2011, Bangladesh increased from 39 percent to 51.4 percent. From 1999-2011, Trinidad and Tobago's

literacy rate increased from 50 to 87 percent. From 1983-2011, Ecuador's literacy rate increased from 84.1 to 91.6 percent. From 1996-2011, Guatemala's literacy rate increased from 65.4 to 75.9 percent. Afghanistan has the highest literacy rate and Angola has the lowest literacy rate.

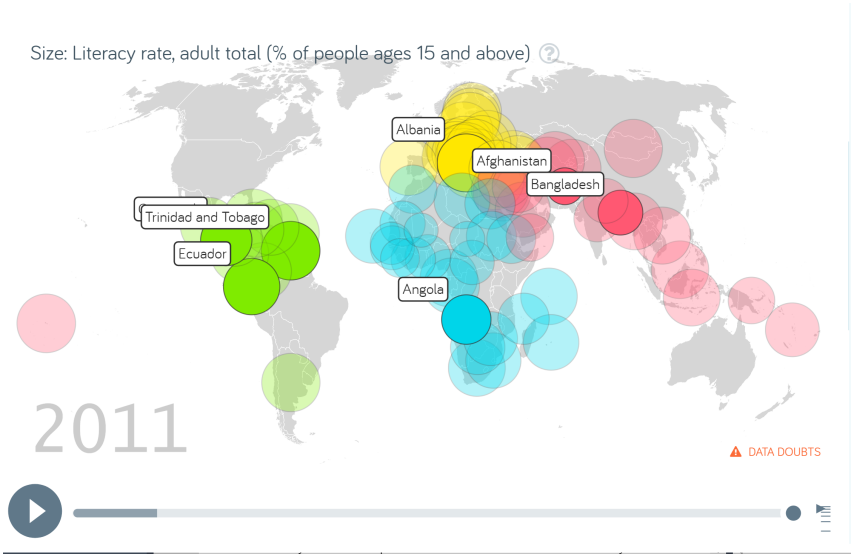

Figure 1:

Income per Person(GDP) The trend graph displays countries that have GDP. From 1800-2019, the country's GDP increased significantly. United States's GDP increased from 2,000 to 64,000 dollars. France's GDP and Japan's GDP increased from 1,000 and 2,000, respectively, to 32,000. Zimbabwe's GDP increased from less than 1,000 to 2,000. United States has the highest GDP and Zimbabwe has the lowest GDP.

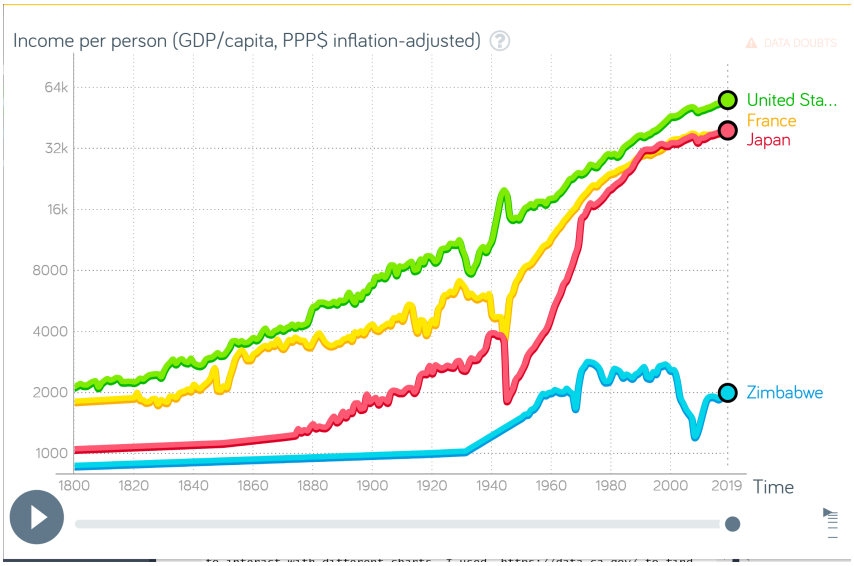

Figure 2:

Exports(percent of GDP) The rankings graph ranks countries from highest exports to lowest exports. The country, Luxembourg has the highest exports (with GDP) with 230 percent of the exports. Estonia has the lowest exports (with GDP) with 7.74 percent. The data proves that Luxenborg exports more goods and Estonia exports less goods from their respective countries.

CO2 emissions per person The bubble graph has x axis, which is carbon emissions per day, and the y axis, which is babies per woman. The bubble graph shows that as the number of carbon emissions per day increase, the total fertility of babies per woman decreases for each country. Asian Countries like India, Papua New Guinea, and Malaysia hav decreased population of babies from 5 to 4 as carbon emissions

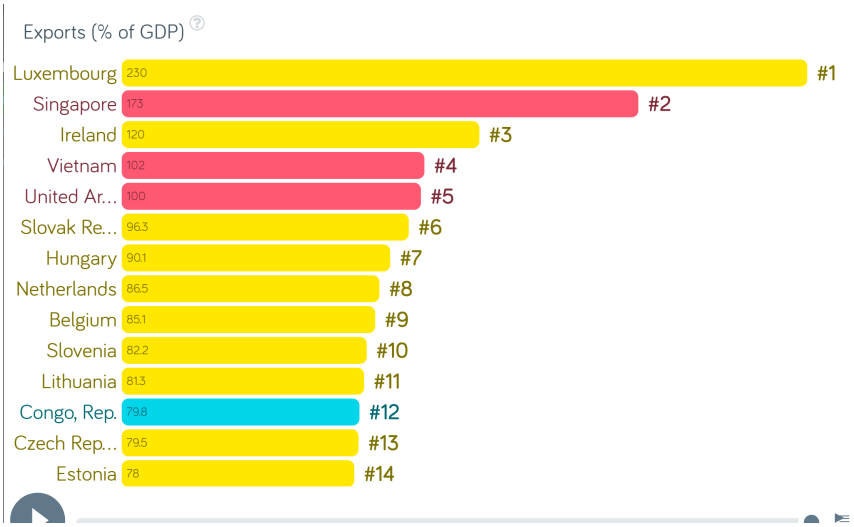

Figure 3:

increase from 1 to 10. African countries like the Congo and Ethiopia have decreased population of babies from 7.34 to 2.51 babies as carbon emissions increase from 0.1 to 10. North American countries like the United States have decreased population babies from 3.03 to 1.75 babies as carbon emissions increase from 0.1 to 10. The United States has the lowest population of babies per woman.

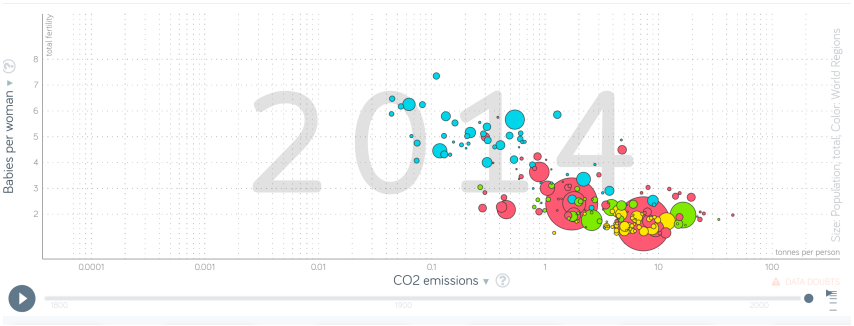

Figure 4:

4 Encountered Problems

The problem with looking at the data sets is that the there were data doubts. In other words, the data doubts pretty much stated that they weren't sure whether they were correct or not. I overcame that by clicking on the tutorial to see how to use the data sets to graph them on gapminder. By reading those tutorials, I had a better understanding of how to read and anlyze the trends from the graphs.

5 Conclusions

In this assignment, I learned how to visualizing data from different data sets. In order to visualize data, I used gapminder for visualizing different graphs for different data sets. Once I finished using gaminder for visualizing graphs, I decided to analyze the trends for each of the data sets. There are positive and negative trends in each of the graphs, allowing me to draw conclusions based on the data shown. In the literacy map graph, more countries in the Middle East read more. In the trend graph, the GDP for each country increases over time. In the ranking graph, the data proves that Luxenborg exports more goods and Estonia exports less goods from their respective countries. In the bubble graph, as the number of carbon emissions per day increase, the total fertility of babies per woman decreases for each country.

6 References

For references, I used www.gapminder.org to find tutorials for graphing different graphs for different data sets. I used Gapminder tools offine to interact with different charts. I used https://data.ca.gov/ to find different data sets.