

Vraj Mistry

CSCI191T: Big Data

109951458

Mistryvraj98@mail.fresnostate.edu

## Assignment 1

### Statement of Objectives

a) Dataset1 :- “Income per person / Time” from 1800-2019

The data set was opened in MS Excel and it has data of 193 countries around the global and it contains all the income per person from 1800- 2019. The objective for this dataset is to analyze the data and measure the value of everything produce by that country per year divided by the number of population. In other terms, it is also called GDP per capita.

b) Dataset2 :- “People living with HIV/ Time” from 1980-2010

The data set was opened in MS Excel and it has data of 149 countries around the global and it contains number of people living with HIV virus over the time period from 1980 – 2010. There are many countries, such as Brazil, India who were not registered until 1990 so they will appear after 1990 year. Likewise, the objective for this dataset is to analyze the data and see which countries has highest amount of people living with HIV over their time period.

c) Dataset3:- “population aged 5-9, both sexes (1%)” from 1950-2019

The data set was opened in MS Excel and it has data of 184 countries around the world and it contains number of population aged 5-9, of both sexes of that country. So, the objective for this dataset is to analyze and visualize to see the growth of both sex in number around the country. Also, we will see how countries like China and India rapidly went up in numbers.

d) Dataset4 :- “Nuclear electricity production, per person / Time” from 1960-2011

The data set was opened in MS Excel and it has data of 150 countries which contains the amount of electricity produced by nuclear power plants per person in that given years. The energy was calculated in the tons. The objective for this dataset is to visualize and analyze the data to see the trend in energy production for each year and which countries are improving or how much production is they are stagnating.

### Experimental Procedure

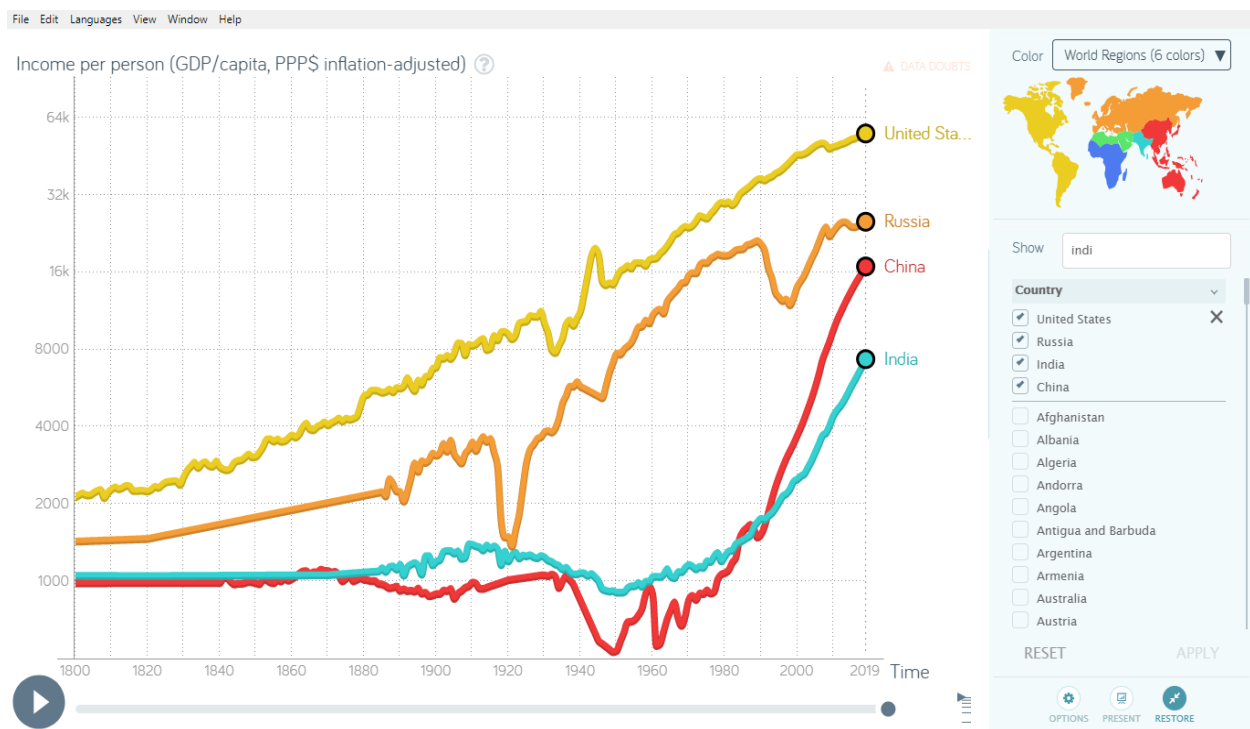
a) Navigating to [www.gapminder.org](http://www.gapminder.org), Open the tab tools , then facts, then data.

- in the tab data find the dataset you are interested in and download the csv file from it
- open the file in MS excel and analyze the data from it.

- After analyzing it, open the file in the gapminder app and choose the graph style you want.

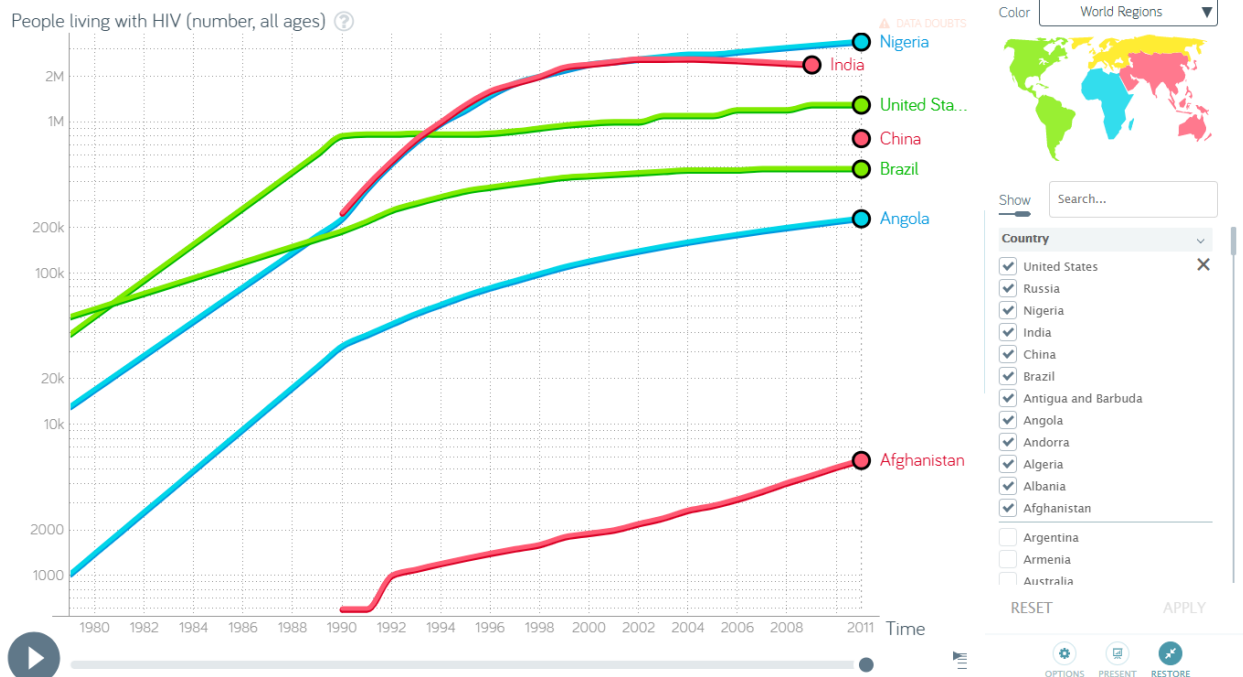
## Analysis

a) Since, my dataset1 had 194 countries, I chose 4 main countries to focus on the graph. So, in this following graph, I observed that United States since 1800 until 2019, have gradually increase their income per person over time. On the other hand, China was constant on the income until 1930 and then it was suddenly drop until year 1940. Later, China gradually went back up in income and by 2019 it almost close to Russia's income. Furthermore, India was constant until 1950 and then after, it gradually increases until 2019. The graph clearly states that almost all the countries have seen increase in the given time period. Likewise, United States has highest income per person among the countries listed.

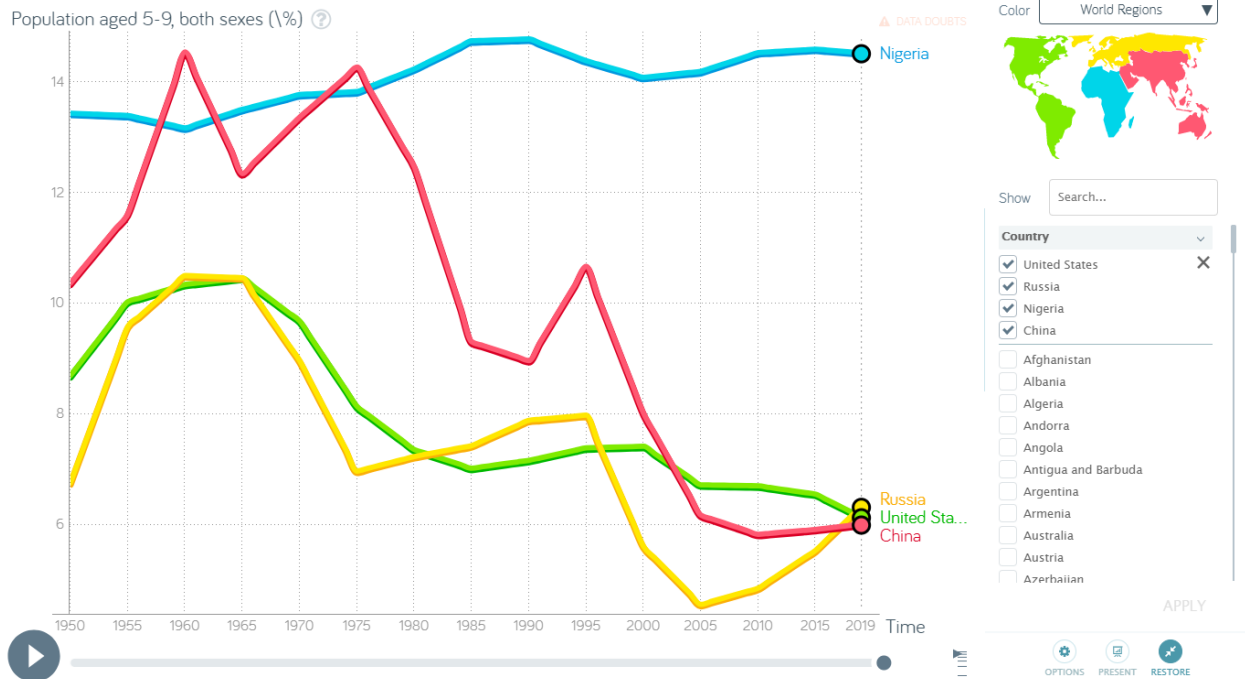


b) The following graph shows the distribution of the people living with HIV over the time period of country. In this graph, we can see that Europe is one continental where most of the people with HIV are living with highest income. Nigeria was highest in 2011 amongst the world to have most of the people

living with HIV. Furthermore, as per the trends, Nigeria had 3.4M people by 2011. Also, there were some countries like India, which came in the report after 1990 as per UNAIDS.



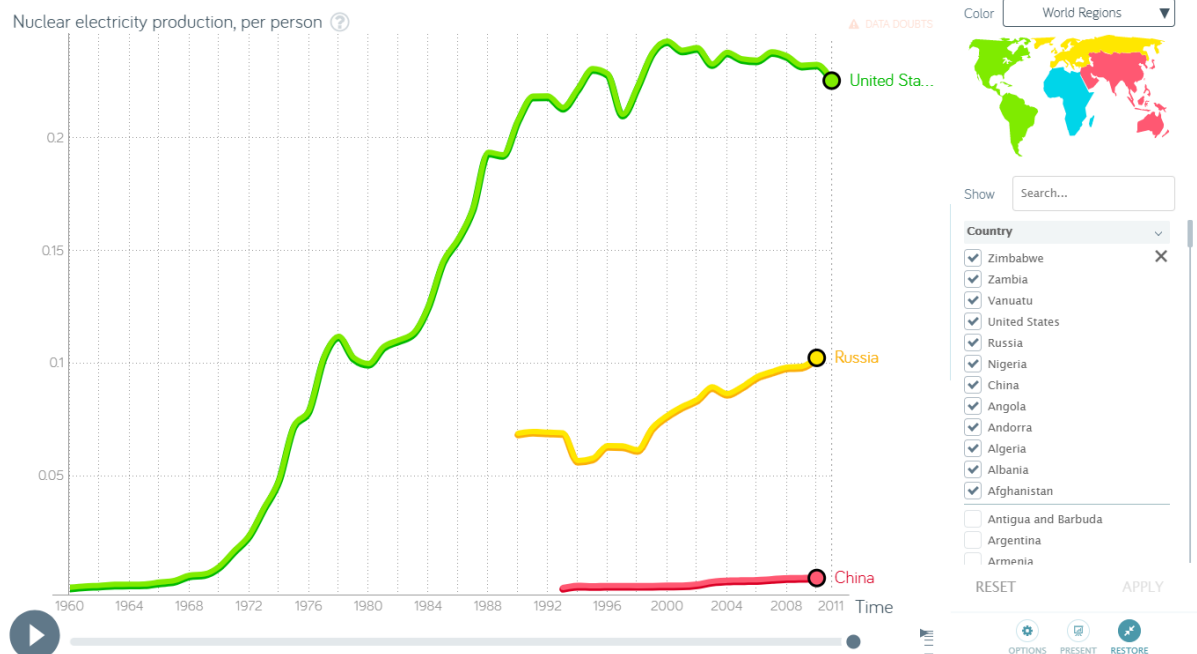
c) The following graph is about the population age from 5-9 of both sexes around the globe. According to this graph Nigeria was highest in the population amongst the selected countries and it was increasing little by little and after the year 1990, the population when low until 2000 and slowly going up until 2019. If you look at the line of China, it was going having rapid increase and then rapid decrease and by 2019 they are almost near United States and Russia in controlling world population. Likewise, Nigeria was highest in the population for both sexes over the given timeframe.



d) the following graph is showing the comparison between countries of nuclear electricity production. This time France is taking the lead in having more nuclear electricity production in tons per person. It is 0.693 tons per person on average income of 32k currency. On the other hand, United States are gradually growing their productions since 1975 until 2011 and they are second country to have highest production of nuclear electricity.



The following graph is showing the main countries focusing their growth on nuclear electricity production.



Thus, China is also focusing on the nuclear electricity production since 1992 and Russia is also growth their production since 1996 until today.

## Encountered problems

- When we download the files from the website <https://data.ca.gov/>, the files were not able to open in the gabminder and we need to rearrange the data and clean it.
- Moreover, even after the arrangement where done, it was not loading the data the way it should be, So we used the data from gabminder.
- No issues were faced while performing visualization of data.

## Conclusion

Performed analysis on the dataset "Income per person / Time" which were taken from the website [www.gapminder.org](http://www.gapminder.org) from the data section. By creating the graph, we conclude that all the countries are constantly increasing their income over time from 1800 to 2019. Likewise, United States was the only country which was growing exponentially in the income rate among other countries.

Performed analysis on the dataset "People living with HIV/ Time" which was taken from the website [www.gapminder.org](http://www.gapminder.org). By the creating the bubble chart on this data, I conclude that China was the first country to have highest people living with HIV over their income and also Ireland was the one of the countries where least people with HIV were living. India and other countries like Brazil, they came in the

consideration after 1990 by UNAIDS. Likewise, this data includes people who are affected by AIDS and healthy people after the HIV infection.

Performed analysis on the dataset "population aged 5-9, both sexes (1%)" which was taken from the website [www.gabminder.org](http://www.gabminder.org). By the creating the trends on the graph, I conclude that Nigeria was one of the top countries to have highest population of both sexes of age 5 to 9 from 1950 to 2019. On the other hand, countries, like United States, China and Russia, they were having lesser population from 1960. Likewise, we can also conclude that these countries are trying to control the birth by having polices and spreading awareness.

Performed analysis on the dataset "Nuclear electricity production, per person / Time" which was also taken from the website [www.gabminder.org](http://www.gabminder.org). After making bubble graphs and trends I conclude that North America and South America are the continentals that are producing highest amount of Nuclear electricity production in the world. Moreover, Russia and China have also started working on the Nuclear electricity production since 1990 and will try to maintain their trend up with united states.

## References

(n.d.). Retrieved from <https://data.ca.gov/>

(n.d.). Retrieved from <https://www.gapminder.org/>