DATA WAREHOUSING & BUSINESS INTELLIGENCE

FINAL PROJECT REPORT

(Team Members: Mayank Vijaywargia, Jyotika Roychowdhury, Harshit Gupta)

Introduction

In this project we have performed visual analytics of the Central Intelligence Agency Factbook dataset. It is a dataset published by the Central Intelligence Agency (CIA). It is a publicly available dataset that provides information about country profiles. The World Factbook provides data about 267 international entities. It gives details about the people, history, geography, economy, government, military, transportation, communications and transnational issues across the world. It is also partially updated on a weekly basis.

We created five dashboards with each dashboard providing insights for a area of the factbook, namely "People and Society", 'Economy'," Energy", "Communications" and "Transportation" and also illustrated their characteristics. Those insights included tracking different KPIs which are actionable and may vary over time.

Project Objectives

The main objectives of our project were:

- Exploring the Central Intelligence Agency (CIA) factbook dataset and deciding on important metrics for analysis.
- Analyzing the data set by filtering and extracting relevant data and also identifying KPI'S across different categories for tracking and deriving meaningful insights from the processed data.
- Visualizing the data using Tableau and building dashboards to reflect the same effectively. Also, deriving correlation between the parameters to perform deductive analysis.

Challenges Encountered and their Resolution

We searched online and ended up finding a GitHub repository containing 267 international entities each representing a different country. But the issue with the repository was that we had 76 csv files in total since individual variables were formalized into different datasets. Therefore, we had to integrate all the columns together to create a unified dataset and to visualize it using Tableau.

To get a resultant consolidated dataset, we used Excel's power query editor to integrate the data sources and then used outer join to combine the columns which resulted in a lot of redundant null values. We then cleaned the data by performing an inner join which produced our final dataset containing 186 rows and 45 columns. This was the final database that we imported into Tableau.

Key Performance Indicators Selected

We have created our dashboards based on five attributes within the key categories:
- **Economy**: GDP, GDP per capita income
- **Energy:** Crude Oil, Nuclear Energy, Electricity
- **People & Society**: Birth Rate, Population Growth, Life Expectancy, Infant Mortality
- **Communication**: Internet Users, Landline Users and Mobile Users
- **Transportation**: Airports and Roadways

Result of The Project:

Electricity Metrics Dashboard
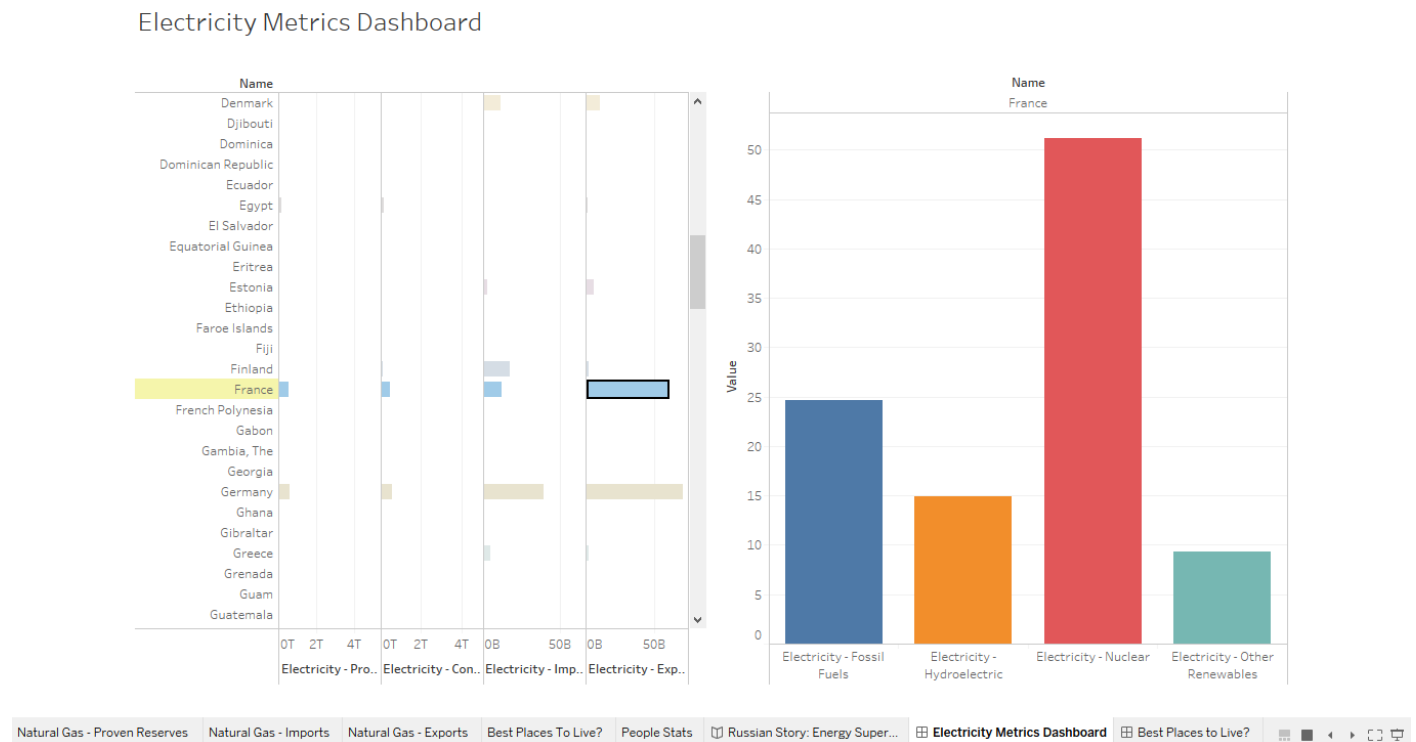
The above figure visualizes the electricity metrics included in the dataset. We are categorizing the electricity metrics into two semantically related domains. They are:

- Electricity metrics pertaining to economy
- Electricity metrics pertaining to electricity source.

In the figure, the left chart shows the electricity metrics related to economy such as:

- Electricity Production
- Electricity Consumption
- Electricity Imports

- Electricity Exports

And the right chart shows the metrics related to the electricity source namely:

- Electricity - Fossil Fuels
- Electricity - Hydroelectric
- Electricity - Nuclear
- Electricity - Other Renewables

The above figure shows the example of **France** which has one of the highest **nuclear** electricity productions and is also a leader in the exports of nuclear electricity power. The visualization of other countries such as USA, China, India show high electricity imports as well as high consumption of electricity and primary reliance on **fossil fuels** for electricity.

Russian Story: Energy Superpower

In this story we emphasize the energy self-sufficiency of Russia and its implications. Energy plays an important role in the geopolitics of different nations. The energy sufficiency and high exports of Russia help its to shape global politics. We will show this across three energy sources namely crude oil, refined petroleum products and natural gas. We first begin with crude oil.
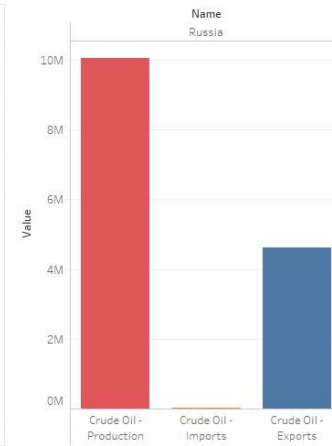
Russian Story: Energy Superpower
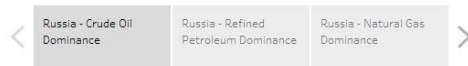
World Map - Crude Oil Exports

Crude Oil

As we can see from the above map, Russia has a very high crude oil production and exports while at the same time low imports of crude oil. Saudi Arabia leads in this category which can be seen from the map. We next move onto Refined Petroleum products category.

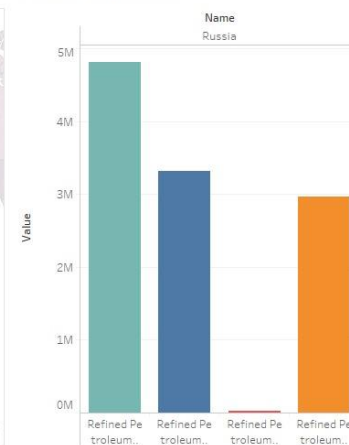Russian Story: Energy Superpower
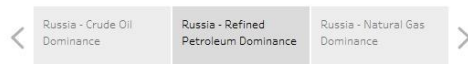
World Map - Refined Petroleum Exports

Refined Petroleum

As can be seen from the above chart, Russia also has high refined petroleum production and exports. We can see the consumption is high as well but that does not affect the exports of petroleum products. The ratio of imports is similar as that of crude oil. Lastly, we move onto natural gas category.
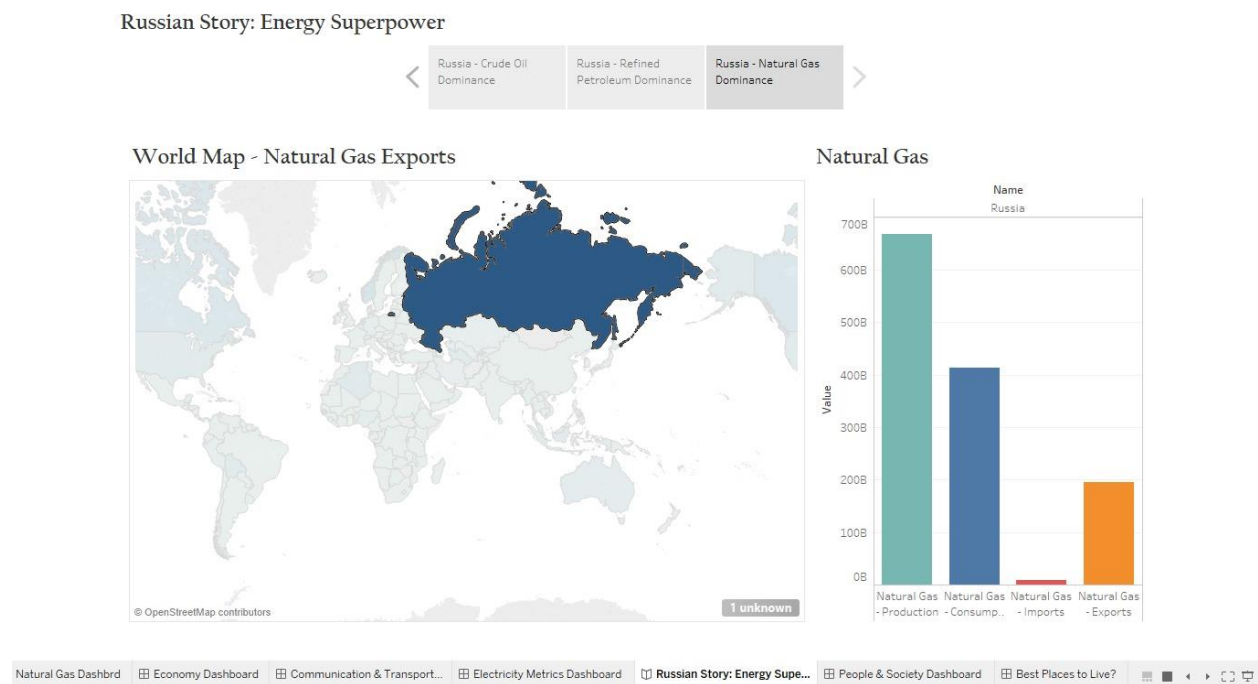
As we can see, the same story repeats with the Natural Gas. High production, high consumption, high exports with low imports. We can see from the above charts that the energy needs of Russia are largely met internally. Also, high exports among all the three categories makes Russia play an important role in these markets as well as global trade, the reason for it being call "Energy Superpower".

Economy Dashboard

Since economy of any country is an important metric that tells us about its wealth or prosperity dynamics. So, we took economy as one of the key performance indicators for our Visualization of CGI data on Economy Dashboard.

We represented three graphs:

- **Comparison of GDP for Top 25 Countries**
- **Comparison of GDP Per Capita Income for Top 25 Countries**
- **Comparison of Difference Between GDP and GDP Per Capita Income for Top 50 Countries**

By using All these Graphs on our Dashboard, we got some interesting details about

- **the Variation in GDP for Different Countries,**
- **Its Impact on the Economy and Prosperity of that Country,**
- **Difference among different countries with respect to economic growth and average income for an Individual.**

GDP is defined as an inflation-adjusted measure that reflects the value of all goods and services produced by an economy in a given year, expressed in base-year prices. And at the same time **GDP per capita** is a measure of a country's economic output that accounts for its number of people. It divides the country's **gross domestic product** by its total **population.**

China with highest population has the highest GDP of 176300000,00,000 coming among top 25 countries while Bermuda having a highest GDP Per Capita of 86,000 coming among Top 25 countries in this metrics.

So, what we perceived from our analysis is that GDP came out as a measure of an economic health while GDP per capita considered as the reflection of such economic health corresponding to an individual citizen as a perspective. GDP measured the nation's wealth while GDP per capita roughly determined the standard of living in a country. GDP normally increases as the population increases while GDP per capita may decrease when population increases.

Transportation and Communication Dashboard

A Strong, reliable communication and transportation systems play a major role in the economic development of countries and regions. We observe that in areas where both communication networks and transportation systems are efficiently integrated, the fact what's more interesting and strengthening about them is that those regions

- Have the ability to swiftly communicate
- Provide mobility access to people, and transport goods thereby providing a strong base for the growth of trade and hence is a crucial factor in developing a stronger economy

By contrast, if a country or region does not have strong communication and transportation networks, trade and the economy are less robust.

Communication & Transportation Dashboard
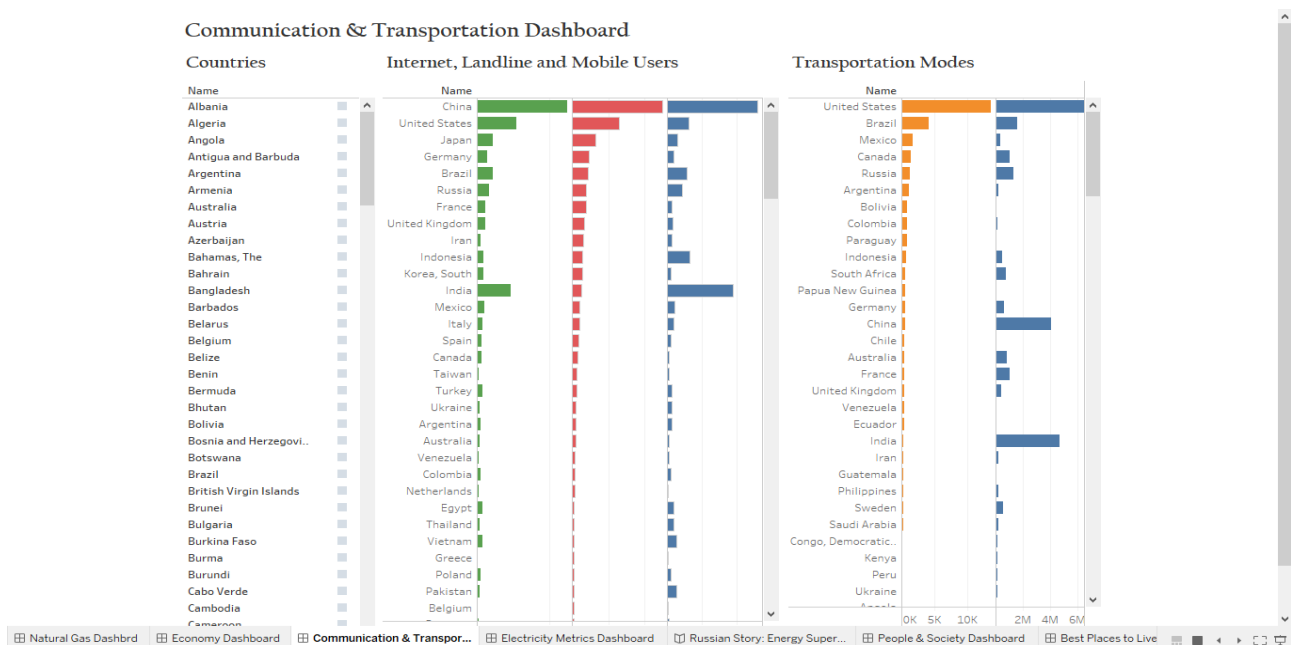
For Our Data We analyzed 3 important modes of communication being Internet, Landline Telephones and Mobile Users. While in Transportation we compared the Number of Airports and Roadways for different countries.

Countries

Name

Slovenia
Solomon Islands
South Africa
Spain
Sudan
Suriname
Swaziland
Sweden
Switzerland
Sri Lanka
Taiwan
Tajikistan
Tanzania
Thailand
Togo
Tonga
Trinidad and Tobago
Tunisia
Turkey
Turkmenistan
Uganda
Ukraine
United Arab Emirates
United Kingdom
United States
Uruguay
Uzbekistan
Vanuatu
Venezuela
Vietnam
West Bank
Yemen
Zambia

Internet, Landline and Mobile Users

Name

United States

100M 200M	
Internet Users	

Name:	United States
Internet Users:	2766,00,000
Telephones - Fixed Lines:	1294,00,000
Telephones - Mobile Cellular:	3174,00,000

Transportation Modes

Name

United States

0K 5K 10K	0M 2M 4M 6M
Airports	Roadways

Natural Gas Dashbrd | Economy Dashboard | **Communication & Transpor...** | Electricity Metrics Dashboard | Russian Story: Energy Super... | People & Society Dashboard | Best Places to Live

---

Countries

Name

Slovenia
Solomon Islands
South Africa
Spain
Sudan
Suriname
Swaziland
Sweden
Switzerland
Sri Lanka
Taiwan
Tajikistan
Tanzania
Thailand
Togo
Tonga
Trinidad and Tobago
Tunisia
Turkey
Turkmenistan
Uganda
Ukraine
United Arab Emirates
United Kingdom
United States
Uruguay
Uzbekistan
Vanuatu
Venezuela
Vietnam
West Bank
Yemen
Zambia

Internet, Landline and Mobile Users

Name

United States

100M 200M	0M 50M 100M	0M 200M
Internet Users	Telephones - Fixe...	Telephones - Mob...

Transportation Modes

Name

United States

0K 5K	
Airpor	

Name:	United States
Airports:	13,513
Roadways:	65,86,610

Natural Gas Dashbrd | Economy Dashboard | **Communication & Transpor...** | Electricity Metrics Dashboard | Russian Story: Energy Super... | People & Society Dashboard | Best Places to Live

The United States has total number of Internet Users as 2766,00,000, Telephone Users (Fixed Lines) as 1294,00,000 and Cellular Phones as 3174,00,000 which gives us an impression of it being the best country in terms of well-integrated communication network which makes communication swift and efficient across different channels.
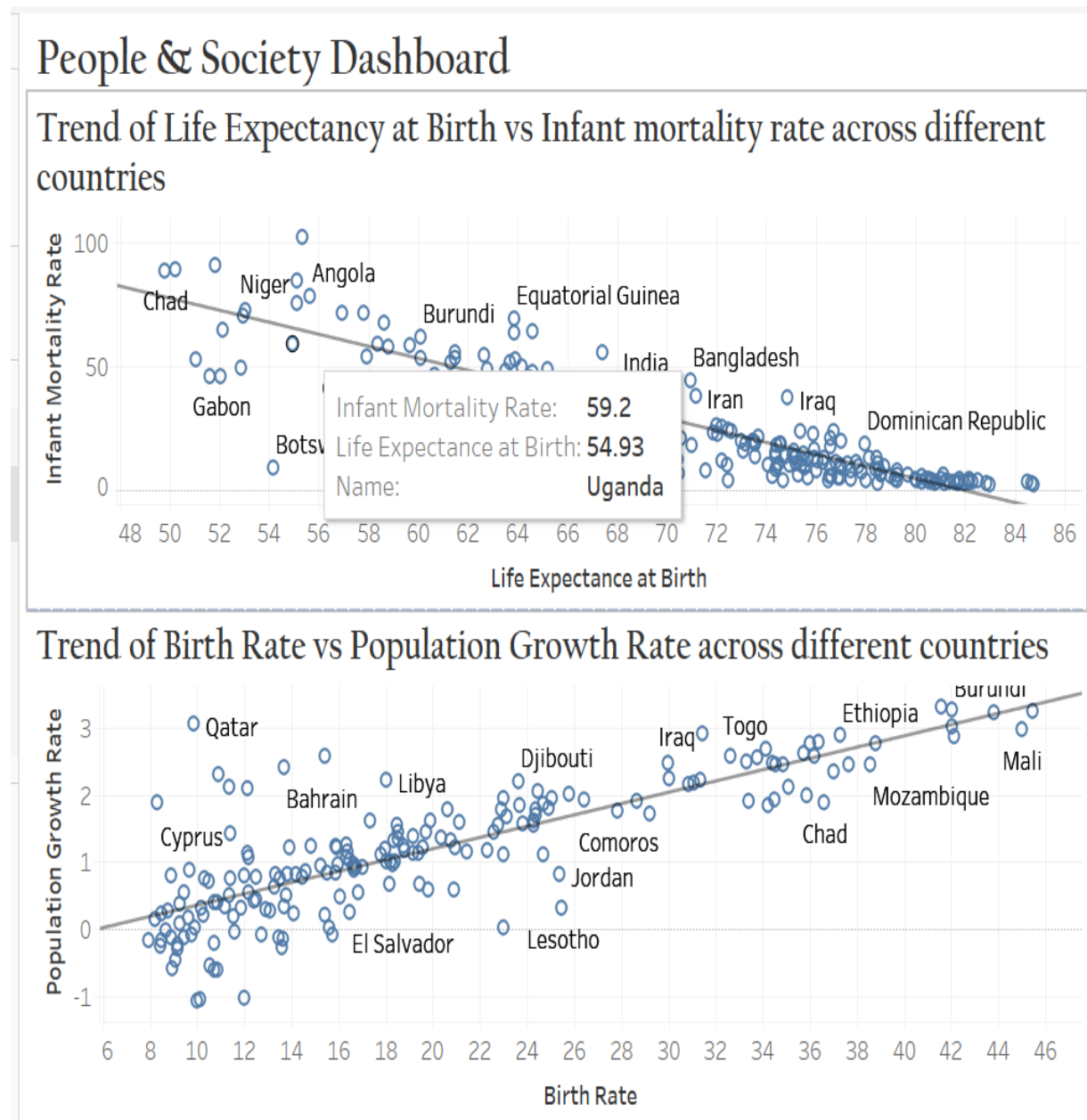
The Transportation modes being Roadways and Airports in the United States are quite significant in number. There is a total of 65,86,610 Roadways while there is a total of 13,513 Airports making it the most mobile in accessibility and having a strong base for trade.

Following we are analyzing further important aspect being **People and society**.

Now People and Society is one key aspect for understanding the demographic well by comparing and analyzing the people and society for different geographical locations mainly different countries based on Life Expectancy and Infant Mortality rate for that area. Life expectancy will let us know about the mean expected value of one's future lifetime. Where Half people will live longer, half will live shorter while Mortality of Infant will let us know about number of deaths in children under one year of age, expressed per 1 000 live births and comparing both will show us a descriptive analysis of trend in different locations.

People & Society Dashboard

Trend of Life Expectancy at Birth vs Infant mortality rate across different countries

Trend of Birth Rate vs Population Growth Rate across different countries

In the people and society dashboard, we analyzed the important metrics related to people across different countries. One area of interest has been the relationship of birth rates to infant mortality, which has led to several studies especially in developing nations.

We arrived on four parameters, namely, Life Expectancy at Birth, Infant mortality rate, Birth rate and Population Growth rate which are defined as follows:

- **Life Expectancy at Birth** is defined as measure of the average time a person is expected to live, based on the year of its birth, its current age and other demographic factors including gender. [1]
- **Infant mortality rate** is defined as the death of young children under one year of age per 1000 live births. [2]
- **Birth rate** is defined as the total number of live births per 1,000 in a population in a year or period. [3]
- **Population Growth rate** is defined as the rate a versus which the number of individuals in a population increases in a given time period, expressed as a fraction of the initial population. [4]

In order to find trends, we used scatterplots where we compared "Life Expectancy at Birth" versus "Infant mortality rate". Then we drew a trend line across it. We found that these two parameters were inversely proportional to each other. At the higher end, there were countries like Niger (Infant Mortality rate - 84.6, Life expectancy at birth - 55.13) and towards the lower end we had countries like Japan (Infant Mortality rate - 2.1, Life expectancy at birth - 84.74). We further analyzed the reasons for this based on the population growth rate.

Similarly, we drew a scatterplot for "Birth rate" versus "Population Growth Rate". Drawing a trend line across the plot, we found that these two parameters were directly proportional. At the lower end there were countries like Japan (Birth rate - 7.93, Population Growth Rate - 0.16) and towards the higher end we had countries like Niger (Birth rate - 45.45, Population Growth Rate - 3.250). We found that the developing countries like Niger with high population growth rate had high birth rates, infant mortality rates and low life expectancy rates whereas developed countries like Japan had low birth rates, infant mortality rates and high life expectancy rates. This could be due to the lack of medical resources and prenatal care in the developing countries.

Concluding Dashboard - Best Places to live

Thus, we found out which are the best places to live in.
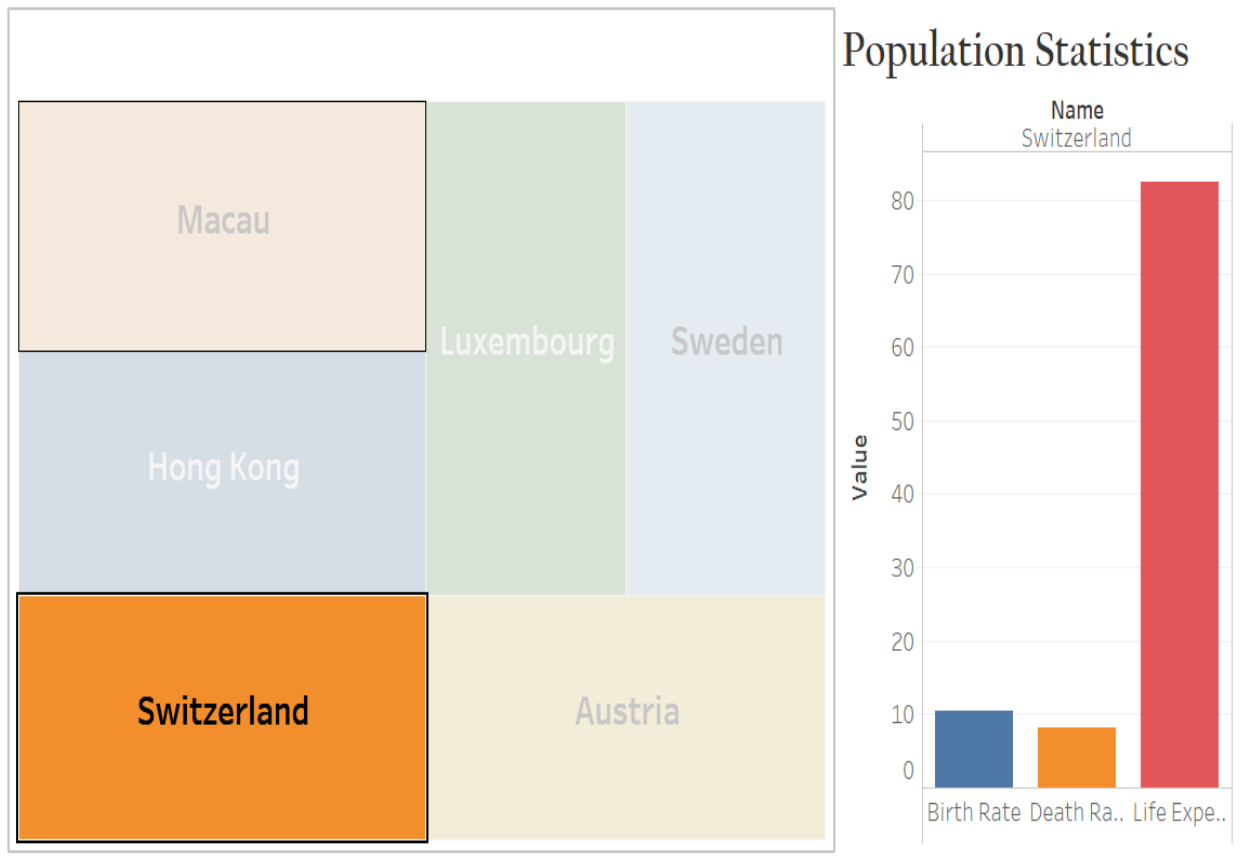
We wanted to find out which countries were the best countries to live in. To determine the standard of living, we used four parameters, namely, GDP, life expectancy, death rate and birth rate. Here, we filtered out these countries based on the top twenty countries with a union of the highest GDP and life expectancy combined with the bottom twenty countries with the intersection of lowest birth rate and death rate. We used a tree map for visualization since the number of countries after filtering were six, namely, Macau, Hong Kong, Switzerland, Austria, Luxembourg and Sweden.

References:

[1] Data Source: https://github.com/thewiremonkey/factbook.csv

[2] S. Shryok, J. S. Siegel et al. *The Methods and Materials of Demography*. Washington, DC, US Bureau of the Census, 1973.

[3] "Under-five Mortality". *UNICEF*. 2017-03-07.
Retrieved from: https://data.unicef.org/topic/child-survival/under-five-mortality/

[4] "World Birth rate - Demographics". *Indexmundi.com*. 17 October 2011.
Retrieved from: https://www.indexmundi.com/world/birth_rate.html

[5] Association of Publication Health Epidemiologists in Ontario. 2008-05-22
Retrieved from: https://www.apheo.ca/
https://web.archive.org/web/20080522092307/http://www.apheo.ca/index.php?pid=61