Analyzing CIA Factbook Data

Malik Williams

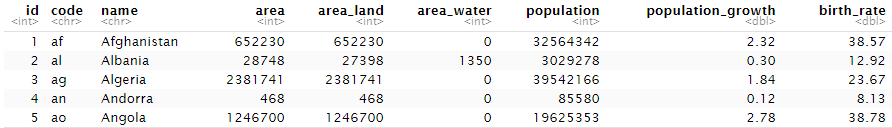
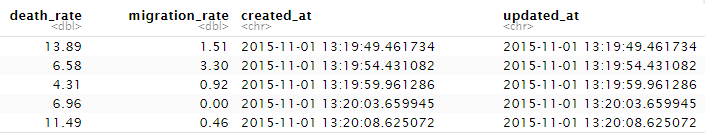
July 27, 2019

For this project we will be exploring CIA World Factbook data. It contains a compilation of statistics over the countries in 2015. We will be looking at data from the facts table. Below is a table detailing each variable in the facts table.

## About the Data

The CIA Factbook contains a compilation of statistics over the countries. We will be looking at data from the facts table. Below is a table detailing each variable in the facts table.

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| name | The name of the country |
| area | The total land and sea area of the country |
| area\_land | The country’s land area in square kilometers |
| area\_water | The country’s water area in square kilometers |
| population | The country’s population |
| population\_growth | The country’s population growth as a percentage |
| birth\_rate | The country’s birth rate, or the number of births a year per 1,000 people |
| death\_rate | The country’s death rate, or the number of death a year per 1,000 people |
| migration\_rate | The country’s migration rate, the ratio between immigrants and emigrants throughout the year |

Now let’s take a look at the first five rows of the facts table.

Below each variable is its variable type and the first 5 values of that observation.

Finding

Now let’s take a look at the population data by finding the extrema (minimum and maximum) values of the population and population\_growth columns.

|  |  |  |
| --- | --- | --- |
| **Extrema of Population & Population Growth** | Minimum | Maximum |
| Population | 0 | 7,256,490,011 |
| Population Growth | 0 | 4.02 |

That doesn’t seem right. It says the minimum population of a country is 0, and the maximum population of a country is greater than 7 billion. We know that the entire world’s population is roughly 7.2 billion so this can’t be right.

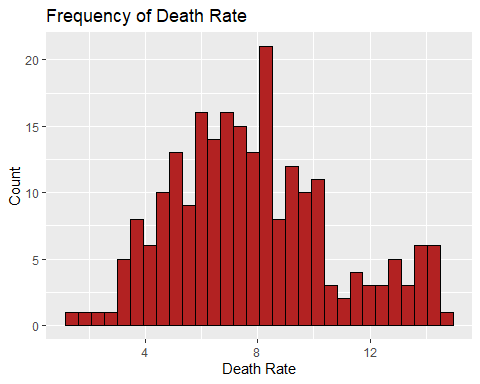
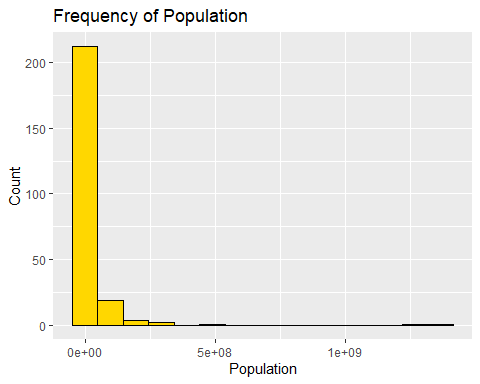
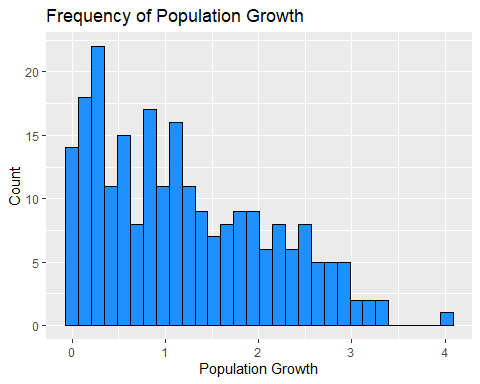
Below, we will find which countries gave those strange values.

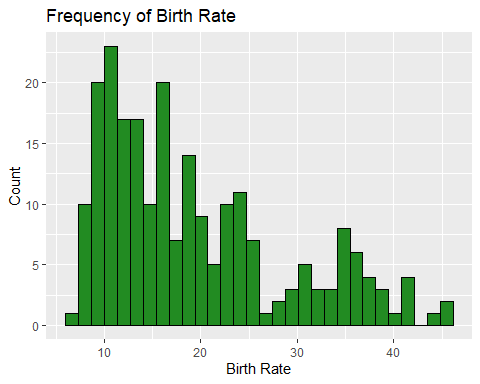
|  |  |  |
| --- | --- | --- |
| Name | Population | Population Growth |
| Antarctica | 0 | NA |
| World | 7256490011 | 1.08 |

The country with a population of 0 is Anarctica, and the observation with 7.2 billion is the World. This explains why there are such extrema maximum and minimum values. No humans permanently live on Antarctica, so its population of 0 makes sense.

Although these two observations are correct, they are also outliers that will skew our results. Therefore, we will remove the outliers before creating visualizations.

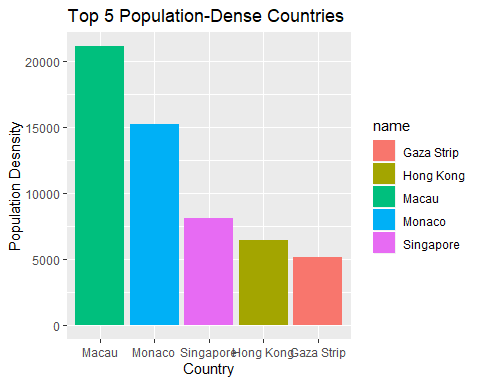
## Creating Visualizations

Now that we have gotten rid of the outliers, let’s generate histograms for the birth\_rate, death\_rate, population, and population\_growth. This will give us insights on how populations change.

The birth rate, population, and population growth histograms are all right skewed. This means that their median values are greater than the means. The death rate is close to being normally distributed.

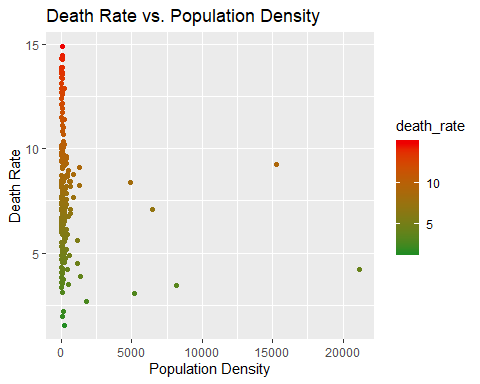
## Finding the Countries With the Highest Population Density

Now that we have gotten an idea of the distributions of the population-related variables, we want to find the most population dense countries. The population density is a measurement of population per unit area or unit volume. In other words, it is the population divided by the area.

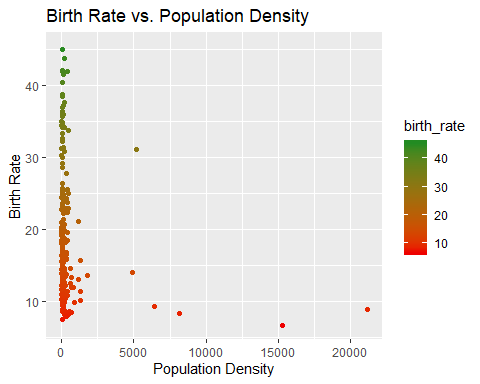


From this data, we see that the most population-dense countries are Macau, Monaco, Singapore, Hong Kong, and Gaza Strip. If we compare our findings with what is listed on [Wikipedia](https://en.wikipedia.org/wiki/Population_density), we see that the top 4 countries match, but on Wikipedia, the 5th country is Gibraltar instead of the Gaza Strip.

## The Relation Between Population Density and Birth Rate/Death Rate

Now that we have explored population density a bit, let’s see if it has any discernable correlations with birth rate or death rate.

Correlation Coefficient: -0.1540137 Correlation Coefficient: -0.1149941

From these scatterplots and correlation coefficients, there appears to be a slight negative correlation between Birth Rate and Population Density. This means that as population density increases, birth rates start to decrease. This makes sense because in more population dense, people are less likely to want have children. There is an even smaller negative correlation between death rate and population density.

## Conclusion

Looking at the CIA Factbook Data seems to indicate that having a large population density correlates with having a smaller birth rate. Death rate seems to be unaffected by population density. This is probably because medicine has advanced to accommodate large populations.