Data Set Description

The title of the Data set is part of the American FactFinder II, a large set of datasets and tables surveyed from the US population. This data (found on this site, here) comes directly from the Bureau of the Census in the US Department of Commerce. It describes many different aspects of American populations, including but not limited to: Births, Deaths, Immigration, Migration, Total Populations, etc. With this diverse spread of data, there are many facets of populations that can be analyzed.

To fully understand the data set, it is helpful to read the documentation. This documentation covers each broad category represented in the data, some of which were mentioned before, as well as how they are represented in the data. For example, the 'Population Estimates' portion of the data contains estimates of populations by town, city, county, or state. It also describes additional information including historical and intercensal estimates, evaluation estimates, demographic analysis, and what the queried data and values mean. Using this additional information, one can ask many different questions of the data, only needing to know how to search the documentation to find the information they are interested in.

Analysis Questions, one question per team member.

One question that can be asked of this data set is: "Does the amount of migration to a region in the US cause an increased rate of deaths for that region?" Since the data set provides information on migration and deaths, this is a feasible question to answer. We could look at each region or the US as a whole, and see how much a change of environment influences the death rate of a population.

Another question that can be asked of this data set is: "Is there locations where more female deaths occured than male in the state of Washington" The data set provides demographic characteristics of age, sex, and race. Additionally the data sets contains information about population change such as death, births, and migration. These characteristics make it possible to answer the question. This question is of interest because we can see the correlation of population growth and gender.

One more question that can be asked of this data set is: What percentage of states received an increase of births from 2015 to 2017? Since the data set contains years we will be able to see the increase of births. Having births and years make it possible to see increases or decreases.

Yet another question that can be asked of this data set is: How does intranational migration affect international migration? This data set provides numbers for international and national migration

each year, so putting the two side by side may give a clue as to whether people moving between states in the US affects how many people immigrate to the US.

Presentation Description

This presentation will be a Shiny app; an interactive visualization viewable on the web. The data is originally well formatted, so not much wrangling will need to be done. Most of the wrangling will be filtering out irrelevant data to the question of analysis. Depending on the question, we will be looking at mins, maxes, and averages. We will also be making interactive graphs and displays of the analysis. For example, a graph plotting deaths against migration, with the option to see different regions, a specific region, or the US as a whole.

The data will have quantitative analysis, for example we are going to have to see the max deaths, mean of deaths, average of sexes in each county, numbers of migrations, and so forth. For the locations where more females deaths occured we are going to have to filter out for Washington state, and compare with comparison operators to see where female death columns were higher than males.

The hardest part of this will most likely be the interactivity component, as well as the analysis of the data for each question. This is because plotly is very poorly documented, and each question requires critical thinking to figure out which statistical transformation is most appropriate to answer the question. Some of the questions will require quantitative analysis, for example finding the maximum deaths in a county.