Basic Info:

- **Project Name:** Demographic Analysis
- **Names:**
    - Christopher Mertin – u1010077 – cmertin@cs.utah.edu
    - SeyedMajid RasouliPichahi – u1013493 – maj.rasouli@gmail.com
    - Ashkan Bashardoust – u1011913 – u1011913@utah.edu
- **Repo:** https://github.com/cmertin/DemographicAnalysis

Background and Motivation:

Migrating to new places is always hard. All three of us had the same problem that we did not have enough information before moving to a new place for graduate studies. Creating this kind of visualization would help other people to decide which areas suit their criteria to move.

Project Objectives:

Since this data provides statistical analysis on certain categories around the US, it will provide relevant information on the types of people in various areas of the United States. You can therefore select certain attributes or areas such that you can tailor your business to your current demographic or see to where you want to expand.

1. **Public usage:** This project could be used for finding a place to migrate, according to a specific lifestyle. For instance, a young person may have higher tendency to go to a place with younger range of people to have more fun activities and nightlife. Or to find a place having people with higher university degrees. One would prefer to live in a city which has people with higher income rates.

2. **Journalists** can use this project to conduct researches and use them as background for articles.

3. **Commercial companies** can use this visualization as a basis for market research and to figure out the supply and demand ratio for their businesses.

Data:

The data will be provided by census.gov, http://factfinder.census.gov

Data Processing:

The U.S. census provides the data in easy .csv format with the info we need from 2010-2015 with the appropriate categories. The only "data cleanup" that needs to be done is to turn the columns of the data selections/attributes into percentages so that we can perform statistical analysis on those columns to give the user appropriate numbers.

For example, if a county has a population of 1 million, 450,000 of which are male, and 4,000 of the original 1 million are native American, we will change the data such that it reads 1 million for the population, 0.45 for the males, and .004 for the number of native Americans, and so forth for each of the data points. Therefore, if the user asks for the number of native American males around the US, we can say for that given county that it's approximately 1 million * .45 * .004 or approximately 1,800 Native American Males in that county on average. This can be done for all the attributes so the statistics can be calculated on the fly as the user selects the data.

Visualization Design:

Different options for the design:

1. We could use states for changing the line chart, but we thought it would give more accurate information to the user, if we use counties instead.
2. We thought about using donut chart, but since there are 50 states, it would be hard to read the chart.
3. We were considering using a bar chart to show the main information for each state, but decided to change that to stacked bar chart to have information even for subcategories.

The user should be able to select multiple attributes, such as age, gender, income, race, etc and be able to tell which states and counties would best fit their selection. There will be sliding scales to select ranges, from which the data will be parsed and the results will be populated on the screen.

The idea is to have a color scale on the map such that it will act as a "heat map" that will show the data all around the US. Then, you can also click on a certain state and it will "zoom in" to the county level, for which it will have more precise colors as well for each individual county.

Underneath this map, there will also be a line chart which will show the selected county over the years that are given by the US Census, and also have a check-box that will allow the user to select that will populate the line chart with the "top 5" in the given selection range to plot against the county that the user selected.

Underneath these, there is a stacked bar-chart, for which the user can choose a category to see the information for that category for all states on that chart. Then, he can choose a state to see the information for all counties of that state.

Must-Have Features:

Some features that are needed for this project to be successful is a good color scale such that the data can be fairly represented and easily understood. On top of this, we need to make sure that it is customizable enough that the user can get the results that he wants and that it will represent the country appropriately. While some of these attributes won't be completely independent (for example income and race), this should provide a good enough approximation to the data.

Optional Features:

We could potentially let the user decide the color scale they would like to use and also the type of chart they would like to see for the info on the given county (donut chart, stacked bar chart, pie chart, etc). This is not required to complete the project, but some user customization on this aspect would make it better for the user in some aspects.

Project Schedule:

1. Downloading all the datasets until October 27th

2. Cleaning the datasets until October 30th

3. Writing the code for the outline of the website including empty charts until November 5th

4. Analyzing the data and the drawing charts until November 29th