Minh Nguyen – 01597092 CIS 602-02 Data Visualization

**Project Report**

**1- Dataset**

Natality Information (Births)

The dataset “Natality” represents the birth counts of residents and non-residents along with the infant death rate happening within the United States regions. The number of given births are calculated from state, census regions, and mother’s risk factors such as tobacco use. All the data were collected from 2007 to 2015.

Source of dataset: Centers for Disease Control and Prevention (CDC)

United States Department of Health and Human Services (US DHHS)

National Center and Health Statistics (NCHS)

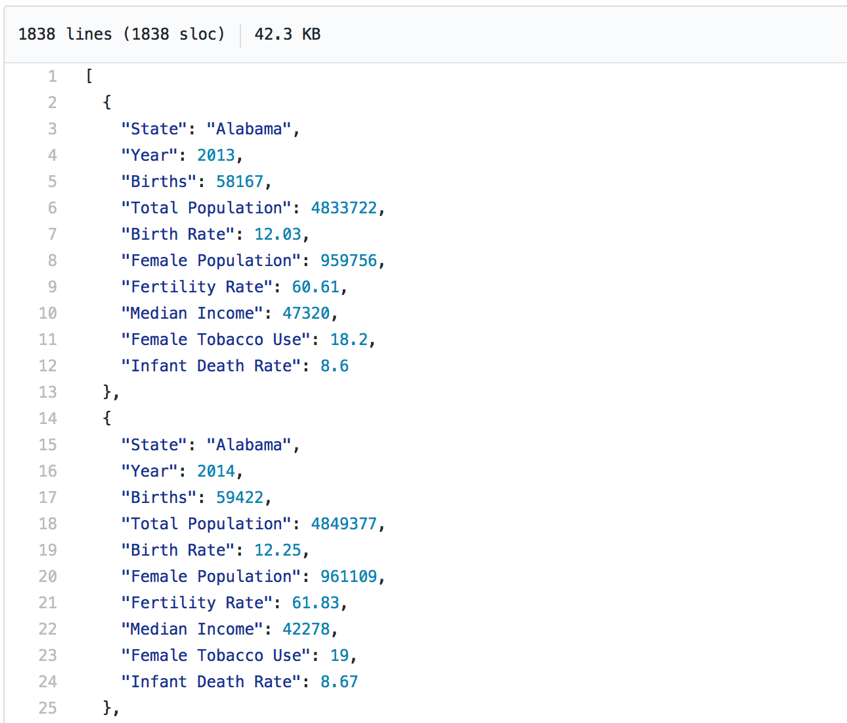
Natality public-use data on CDC WONDER Online Database.

The dataset in the website CDC WONDER provides table, and data extracts according to the chosen Group results. When requesting the data, I can limit and index the data by any and all of the data variables. Follow is the description of types and semantic of datasets:

***natality\_1.json***

|  |  |
| --- | --- |
| **Header** | **Definition** |
| State | The states of the United States |
| Year | The year of data (2013-2015) |
| Births | The total number of births are given |
| Total Population | The total population for each US State |
| Birth Rate | The number of births per 1000 population |
| Female Population | The total number of female in the US |
| Fertility Rate | The number of born children per 1000 women over her lifetime |
| Median Income | The median household income by State |
| Female Tobacco Use | The percentage of female using tobacco while given birth |
| Infant Death Rate | The percentage of infant death per 1000 live births |

Here is a look at the data:



Natality Data

The raw data is from 130MB to 200MB in zip files.

When unzip, the data is around 5GB. Here is the link to obtain the raw data:

<https://www.cdc.gov/nchs/data_access/VitalStatsOnline.htm#Births%EF%BB%BF%EF%BB%BF%EF%BB%BF%EF%BB%BF%EF%BB%BF%EF%BB%BF%EF%BB%BF>

Median Income Data

The raw data (included in zip file) can be obtained from:

<https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-households.html>

Tobacco Use

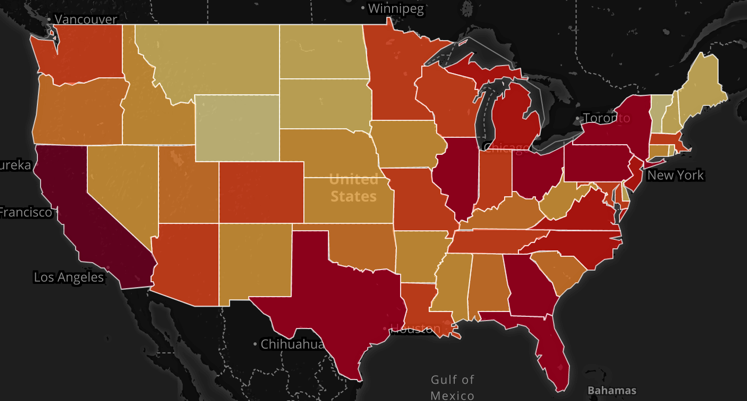
The raw data (included in zip file) can be obtained from:

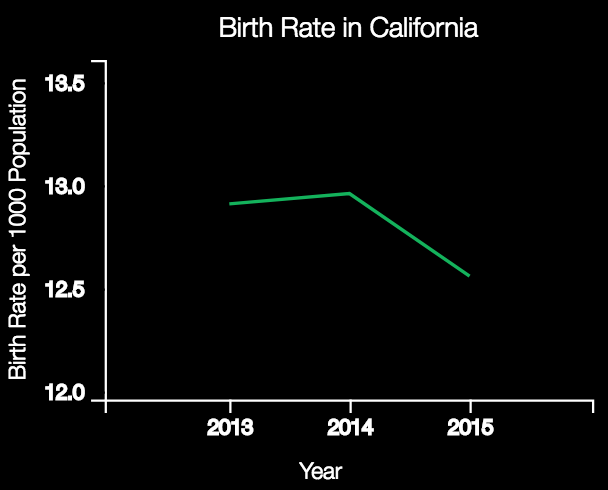
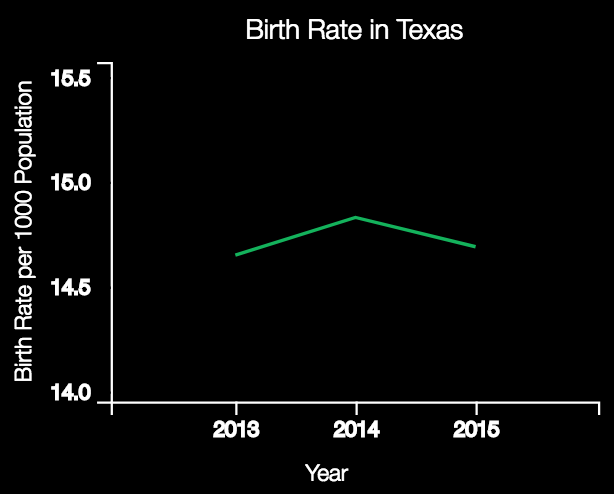
<https://www.healthdata.gov/dataset/behavioral-risk-factor-data-tobacco-use-2011-present>

This dataset is behavioral risk factor data represent tobacco use from 2011 to present. All information about leading causes of death are collected from CDC, State Tobacco Activities Tracking and Evaluation (STATE) system. BRFSS Survey Data. I extracted and filtered the data and get the current smoking status of female only from 2013 to 2015.

**2- Question**

1. What state in the US see the highest density of birth over the past 3 years? What is the trend? On the choropleth map, California accounts for the highest number of birth with nearly five hundred thousand, follow by Texas with approximately four hundred thousand. On the line chart, in overall, we can see that the birth rate in California is lower than in Texas with nearly 2 births per 1000 population. California shows decreasing trend but Texas shows the stable balance.



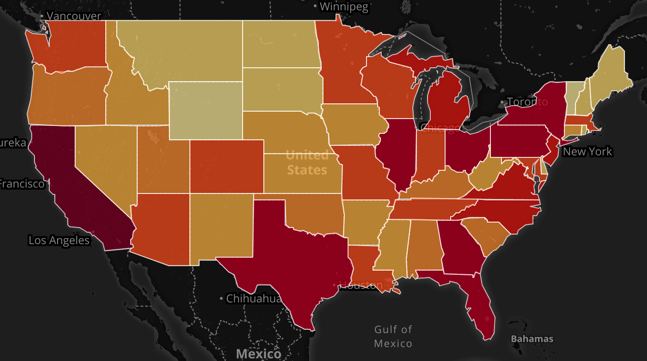
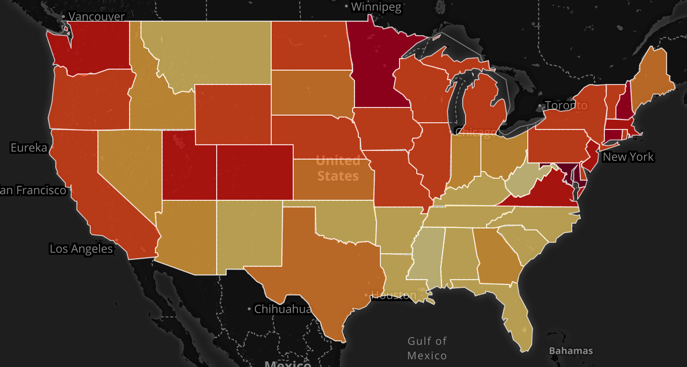
 

1. What is the difference between each census regions based on the number of births? Do the census regions affect the number of given birth?

Most of the Eastern states have the highest density of birth.

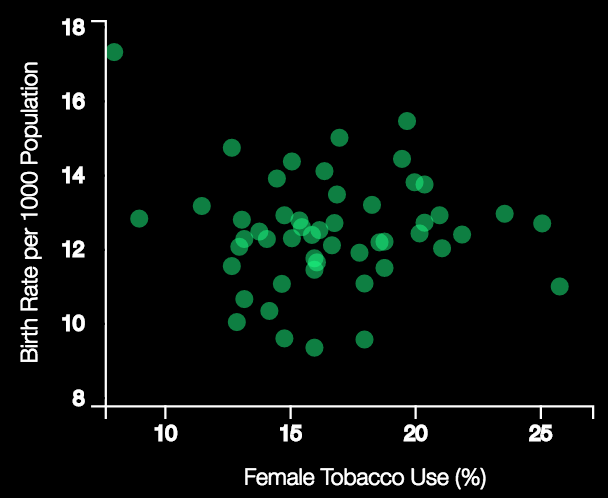
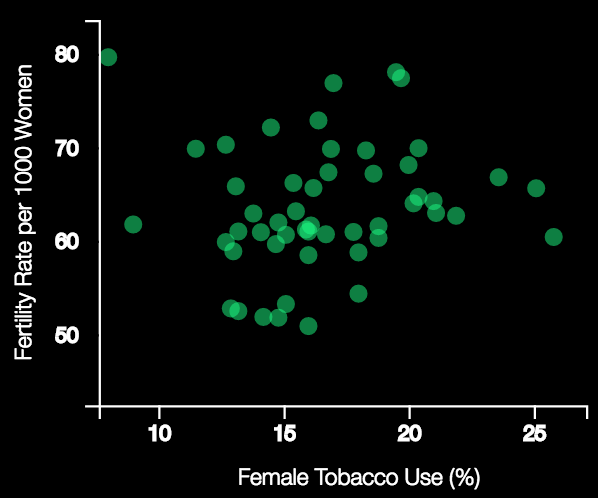
1. What is the difference between median income and population growth data?

According to the choropleth map with two variables *Population Growth* and *Median Income*, we can see that states with higher income will mostly have lower number of births. In the figure below is the Population Growth (left), and the Median Income (right).

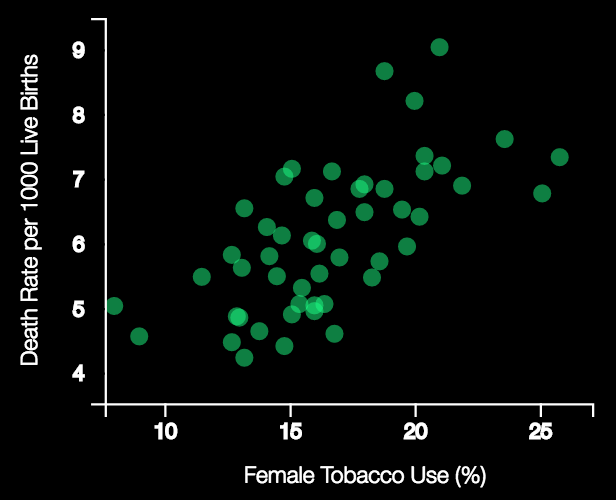
 

1. Does the tobacco usage affect the birth rate, the fertility rate, and the infant death rate? What is it trend?

Yes. Smoking during pregnant can harm the baby. It slows the growth of the baby before birth and health problem after birth. So that the population growth and tobacco use comparison does not show any trend in scatterplot, but it forms a cluster in the middle. From this point, the graph shows that the average between birth rate and tobacco use is around 12 per 1000 and 15% respectively. And the same as fertility rate, 60 per 1000 for fertility rate compare to 15% tobacco use.

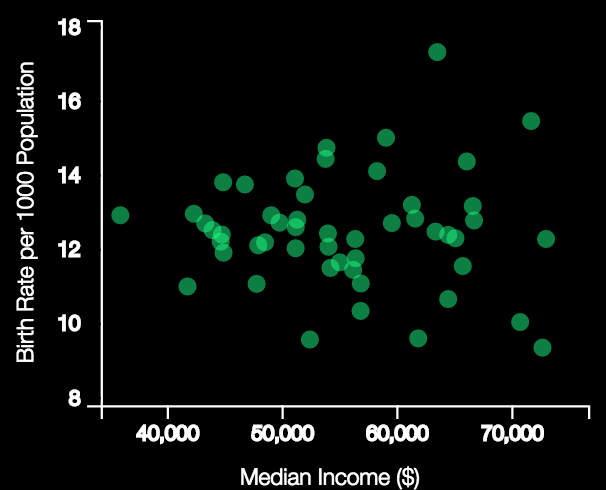
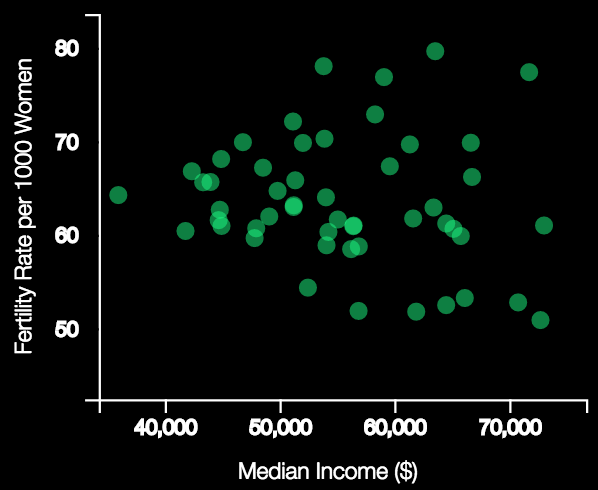
 

However, smoking does effect the rate of infant death, the higher the tobacco usage is, the higher the death rate per 1000 live birth is. Thus, the scatter spot shows the increase trend.

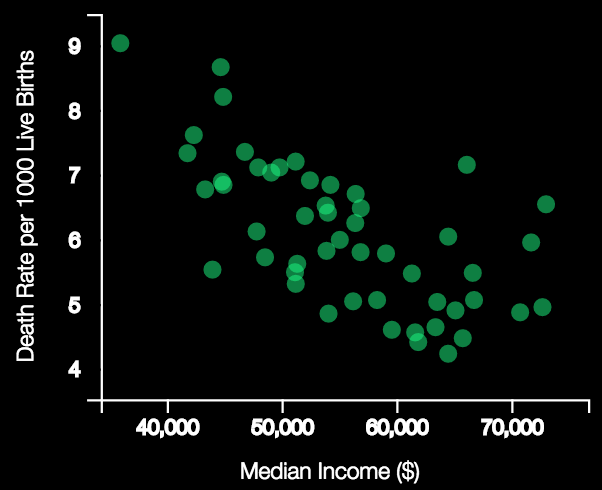


1. Does the median income affect the birth rate, the fertility rate, and the infant death rate? What is it trend?

Yes. According to the graph, the dots form a cluster at the low income with the birth rate value of around 12 per 1000 population. As the income goes higher, the distance of dots become further. So the same as fertility rate.

However, we can see clearly that the higher the income is, the lower the infant death rate is.

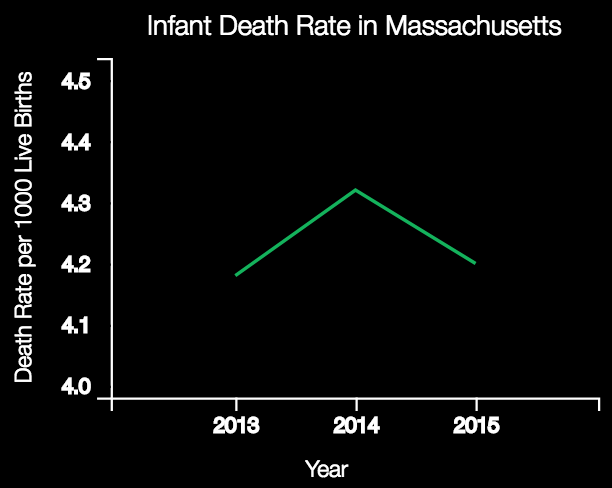
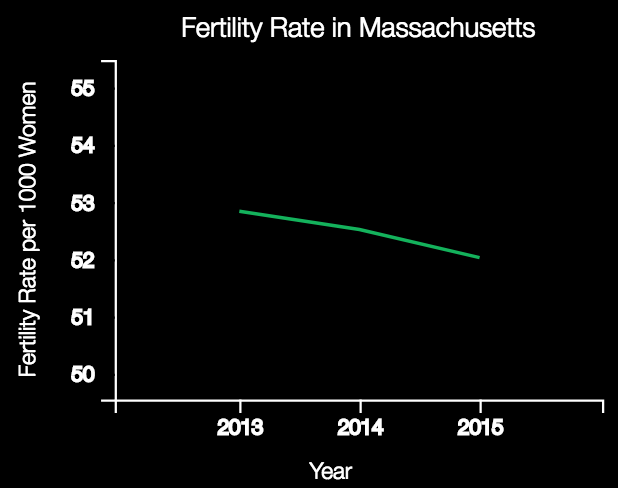


1. How many baby was born in each state from 2013 to 2015? What is the average number of births in all states in the US?

By hover the mouse on each state, we can see the number of births for each over 3-year period. The average number of births is around two hundred thousand children in the US.

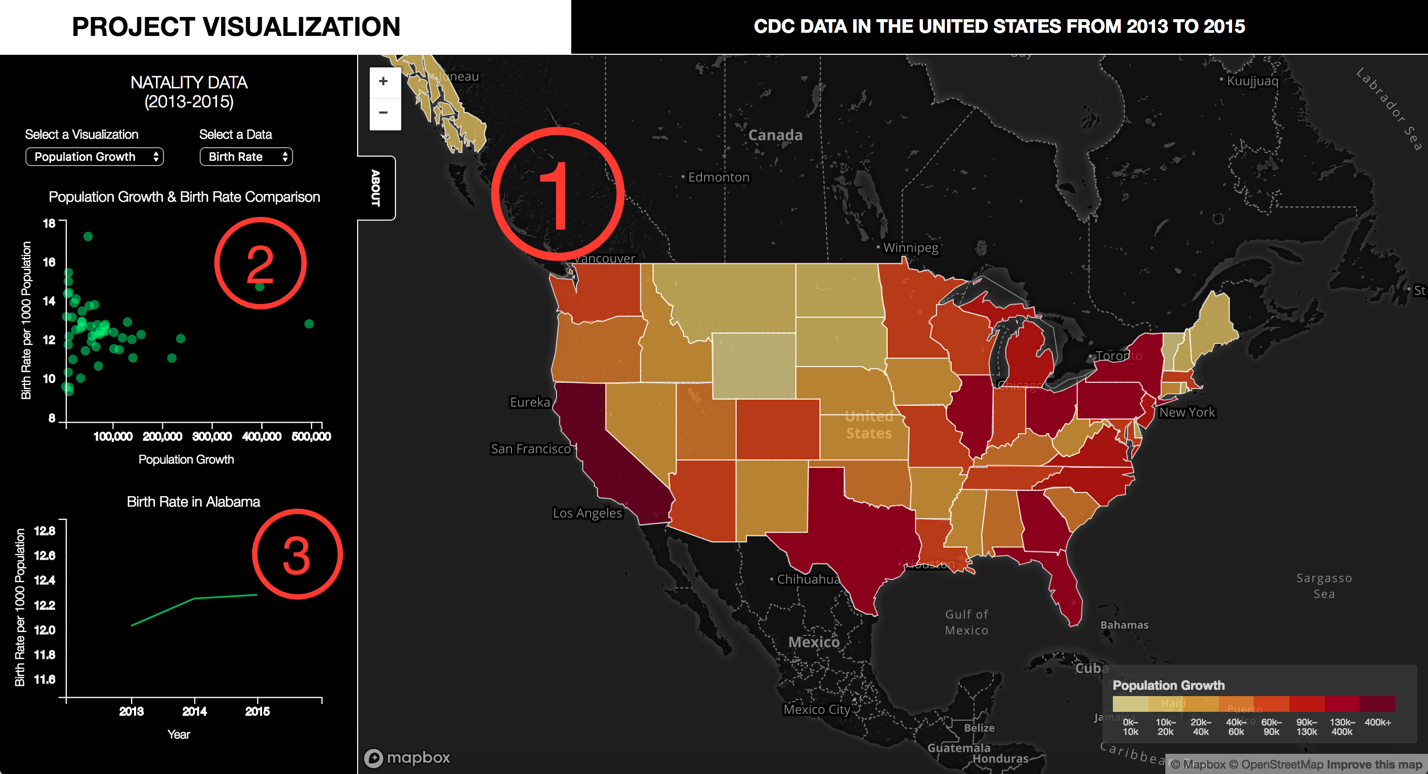
1. What is the trend of infant death rate and fertility rate for each state in the US from 2013 to 2015?

The line chart will show the trend for each state. Below is the infant death rate and fertility rate in Massachusetts.

**3- Visualization**

My visualization is an interactive web application designed to help users visualize spatial and statistical relationships between births, fertility rate, infant death rate, median income, and tobacco use.



There are three sections in the visualization.

1. Choropleth Map

Using the dropdown menus to select a visualization to view the different variables on the map. From the map, we can compare between each state based on the color. Hover to each state to see the specific statistic values. There is a legend to explain the meaning of each section. Using the color saturation to separate each state.

2. Scatterplot Graph

The dropdown menus from the selected visualization will change the variables of x axis of the graph. Scatterplot is used to find the correlation between the two variables. User can then see clearly the trend that the map want to show.

3. Line Chart

With the x axis represents the 3 year periods, and y axis represents the information need to be showed according to the selector such as birth rate, fertility rate, and death rate. I used line to demonstrate the change throughout years, along with transition to make it more interactive.

This project used Multiple-view technique with juxtaposition.