US Census Visualization Process Book Milestone

CS171-Visualization

Simon Malian , Dingchao Zhang

Index

1. Overview and Motivation

- Our project is to visualize a variety of United States demographic and socio-economic data including gender, race, age, income , venture capital, etc. gathered from US Census Bureau and other accredited agencies.

- The benefits of understanding the current population characteristics will influence user's understanding of each location.

- The work can be used by others – e.g. public policy students – it will help facilitate developing realistic projections regarding further trends and challenges (e.g. aging population) that may be utilized, neutralized or reduced through proper planning

2. Related Work

- We found and drew inspirations of Choropleth map from US Census Bureau provided data maps.

 For example, we would follow a similar design of this poverty map:
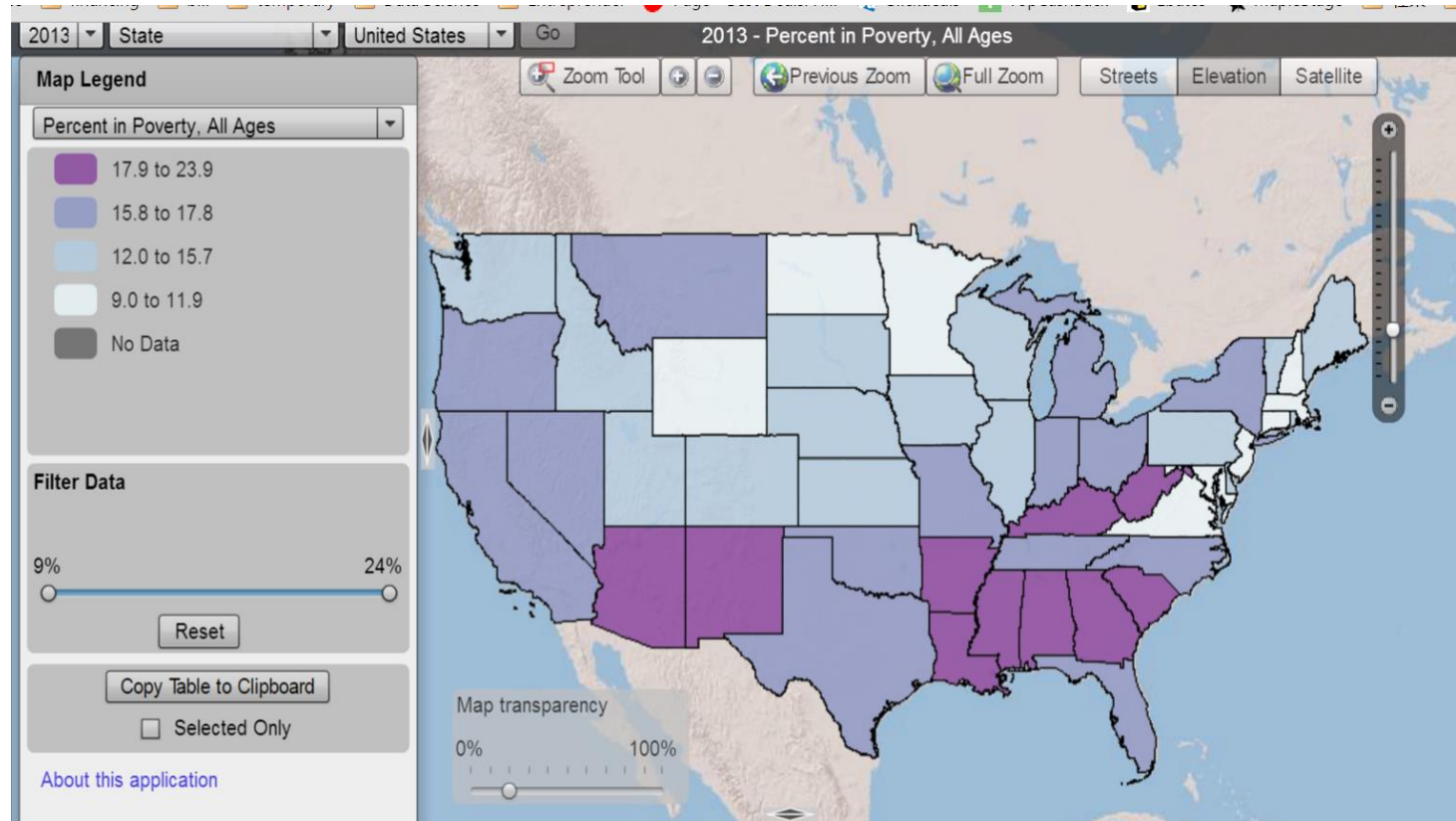
What we can do better than US Census:

- The US Census map seems all built upon flash player technology which is slow and takes long to load data, so we believe our D3 powered technology has advantage of responsiveness to it.
- There is no coordinated view map of different aspects of data provided by US Census, and our project will fill that gap.

2. Related Work - conti

- We found and drew inspirations of Tree map from various D3 blogs.

  For example, we would follow a similar design of this UK flight data map:
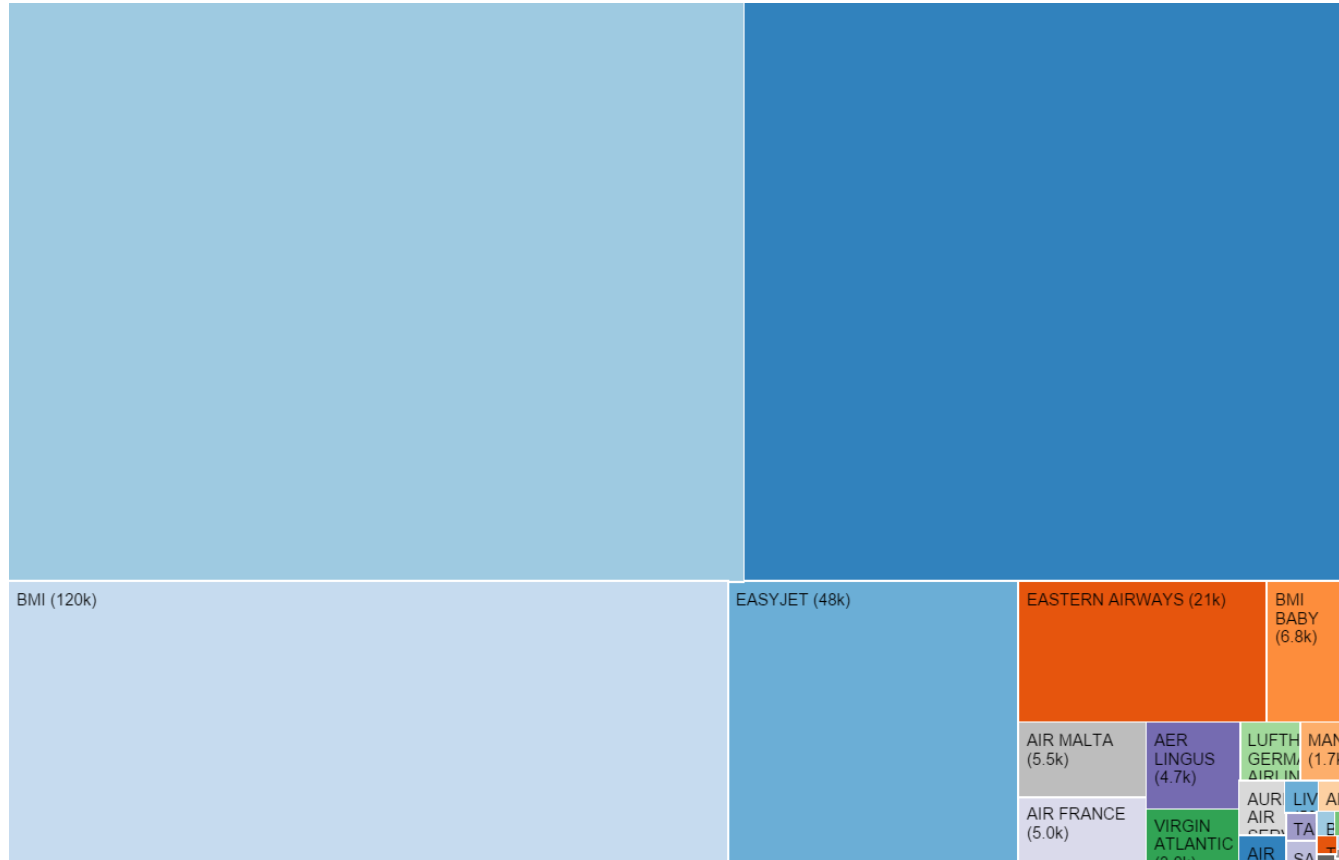
The advantage of using Treemap:
- A very good method to display hierarchical data.
- The color and size dimensions are correlated with the tree strcture, one can easily see patterns that would be difficult to spot in other ways.

3. Questions

- Questions we attempt to answer with this visualization are whether there are any correlations between location, socio-economic distribution, ethnic origin, household income, and other . For example one of the initial questions is what are some interesting stories that the visualization say about deviations in states' ethnicity compositions and their venture capital activities

- As the project goes, we would like also to consider new questions such as besides ethnicity distribution, would education level, age or gender other factors correlated with state's economic performance.

4. Data

- Source

 We initialed pulled available CSVs with all 2010 Census data from US Census Bureau Web; the CSV is released separately for each state, and each CSV is hundreds megabytes, and therefore unsuitable for a d3 project and also un-pushable to Github. Thorough cleaning and de-sizing work would be required to process the data.

 We found there are a few foundations, academic and business agencies summarized and published demographics and economy data on their websites, and The Henry J. Kaiser Family Foundation is one of them. We were able to download CSVs including population by age, gender, race/ ethnicity, income, for each state.

 We were able to found Science and Engineering Indicators 2014 State Data Tool from National Science Foundation and downloaded economy data such as venture capital deals for each state.

4. Data cont

- Data wrangling

Besides cleaning unnecessary columns, meta data from downloaded CSV files, we need to change the structure completely for the data that will be used in Treemap visualization.

Because Treemap data requires to have a multiple layer hierarchical structure, which has "grandparents", "parents", "children", etc. While the downloaded CSV files' rows are aggregated mostly, so to transform to a Treemap data, we need to *de-flat* the data and create multiple layers using various Excel operations.

Original flat data

Location	White	Black	Hispanic
United States	1.95E+08	37782700	54253200
Alabama	3192900	1252100	227200
Alaska	445400	22000	59200
Arizona	3189300	289000	2501300
Arkansas	2119200	458100	239300
California	14894200	2078100	14832900
Colorado	3772600	187700	1085900
Connecticut	2478900	383000	401200
Delaware	586100	172600	96800

Change structure →

Transformed Hierarchical data

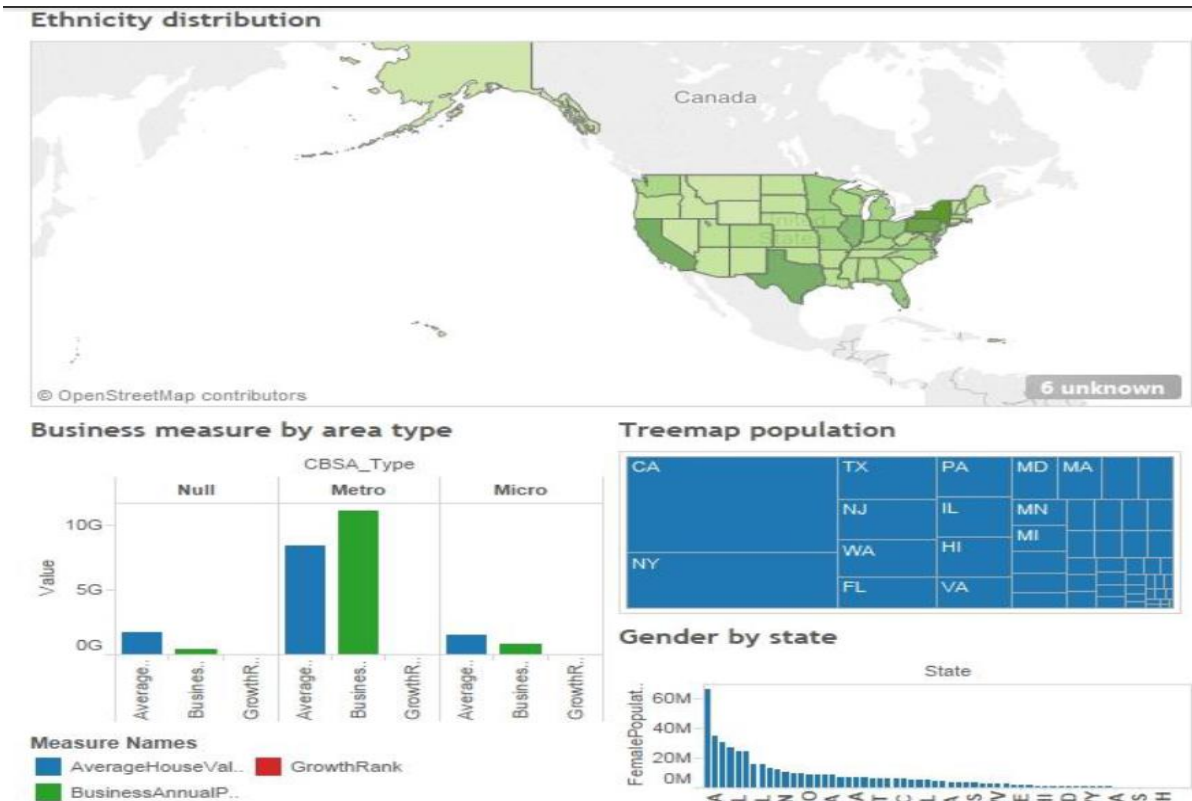| State	Race	Population
Alabama	White	3192900
Alabama	Black	1252100
Alabama	Hispanic	227200
Alabama	Other	83100
Alaska	White	445400
Alaska	Black	22000
Alaska	Hispanic	59200
Alaska	Other	173200
Arizona	White	3189300
Arizona	Black	289000
Arizona	Hispanic	2501300
Arizona	Other	675300
Arkansas	White	2119200
Arkansas	Black	458100
Arkansas	Hispanic	239300
Arkansas	Other	124500

5. Exploratory Data Analysis

- We first used rudimentary, simple visualizations to take a better look at our data, this included looking at the overall data over different metrics such as gender, race, income; to understand its max, mean, min, median values.

  For example, to look at the white group by state composition percentage data, we generates a bar chart in Excel
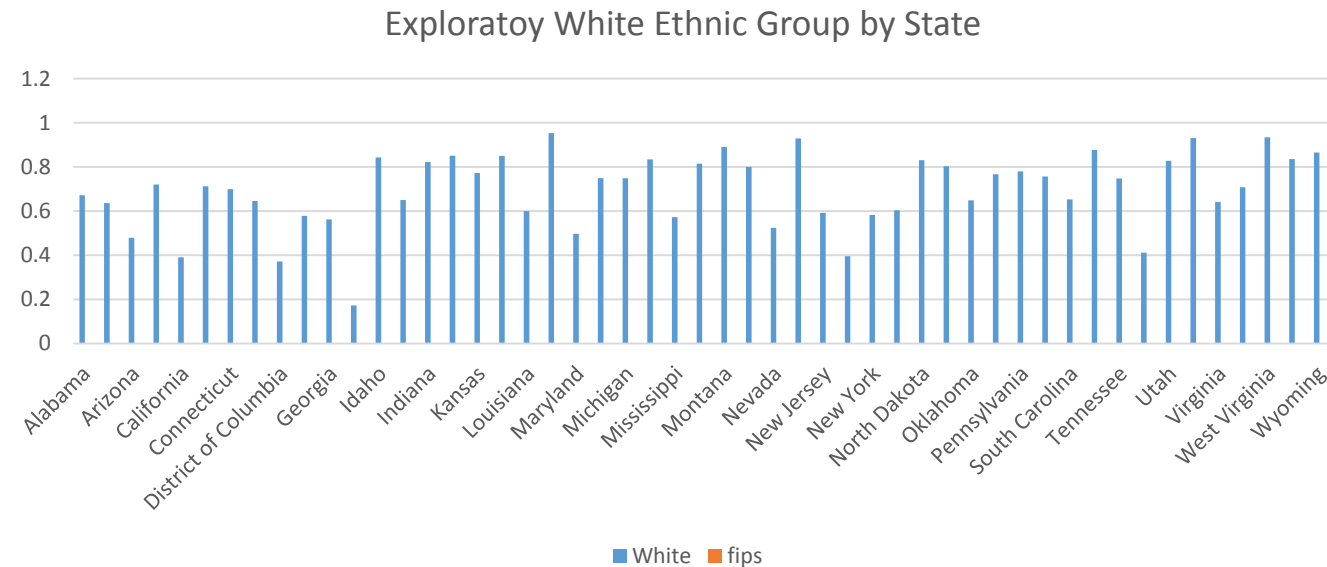
6. Design and Implementation

- The main component of visualization will be in the form of a choropleth map , showing distributions of different data attributes, and we also consider introducing control elements to allow interactive exploration of the data (e.g. show the data for zipcode to analyze whether it is above/below certain threshold) as well as other linked coordinated views such as Treemap, Pyramid bar chart, area chart, etc.
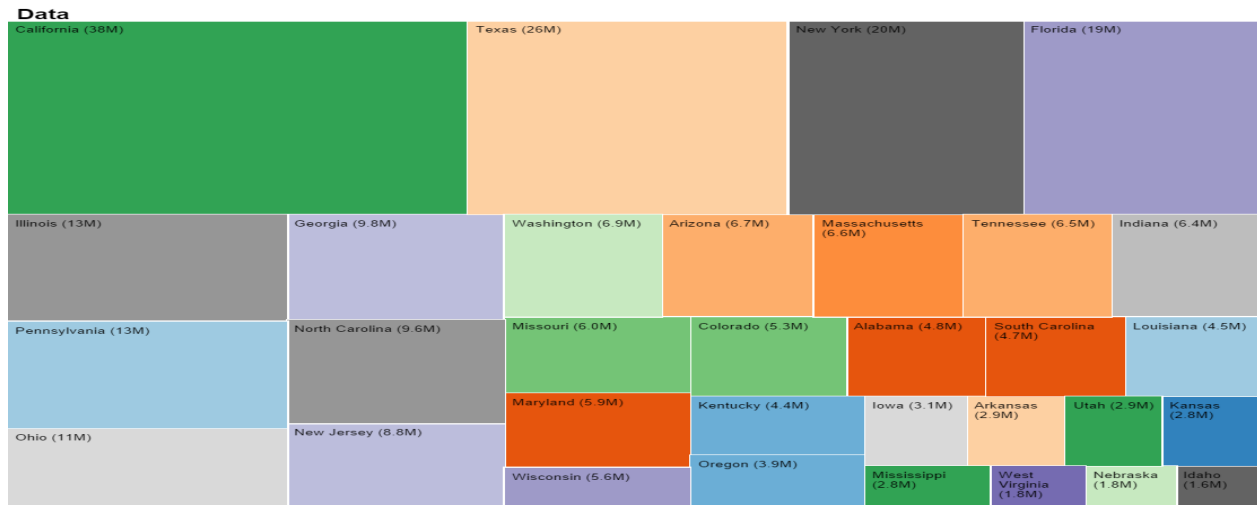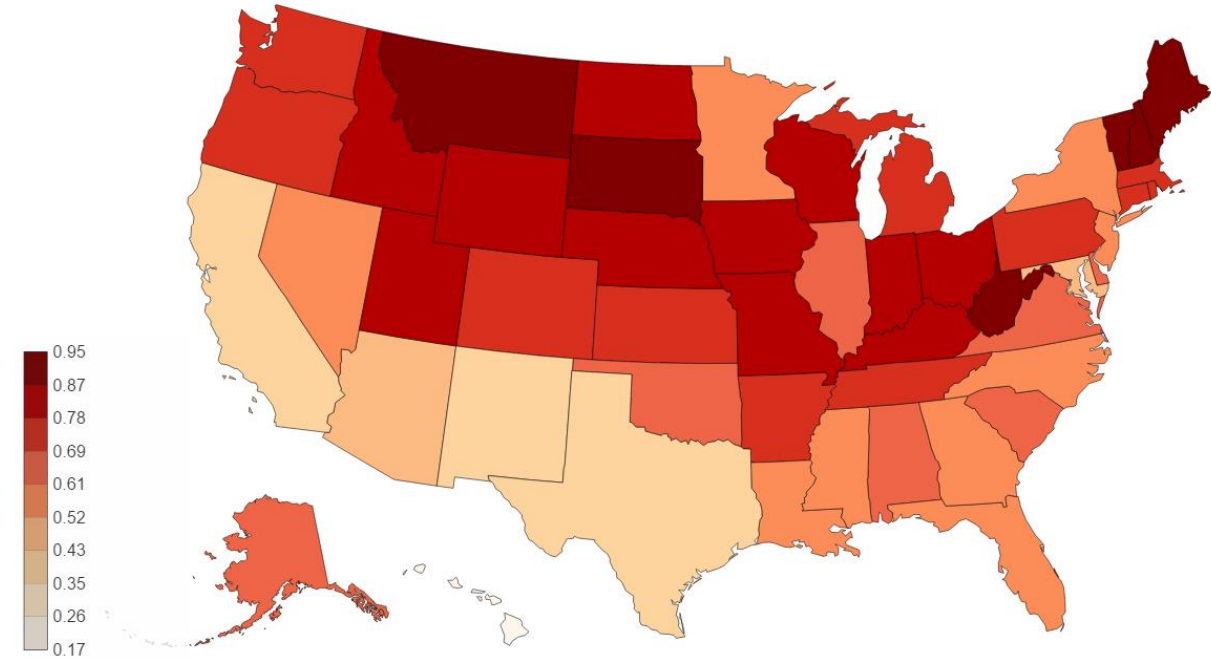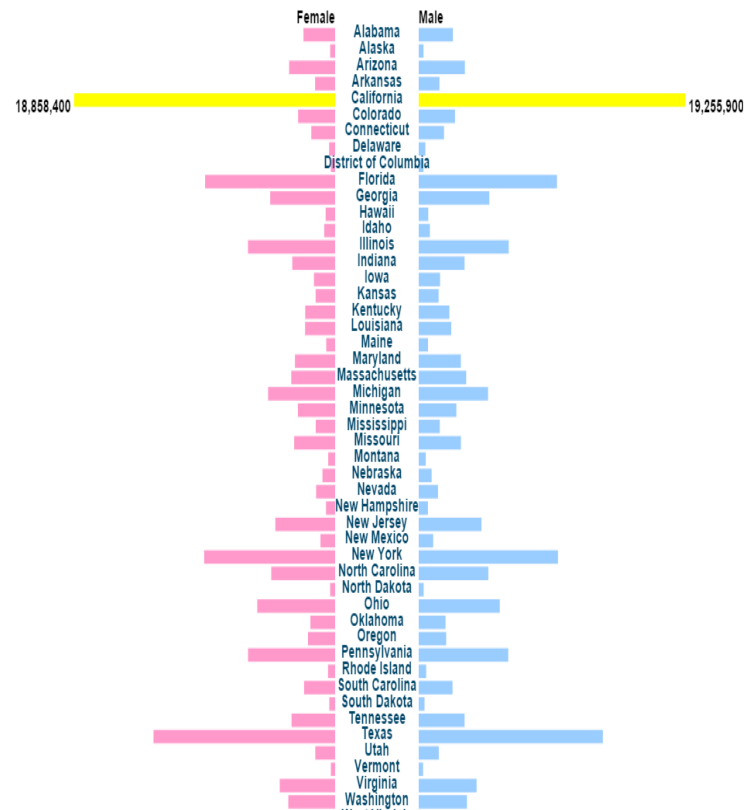
- 6.1 Proposed Sketch

6. Design and Implementation cont

- 6.2 Milestone 1 Design Overview

Choropleth map, Multiple layer Tree Map, Pyramid Bar charts are created
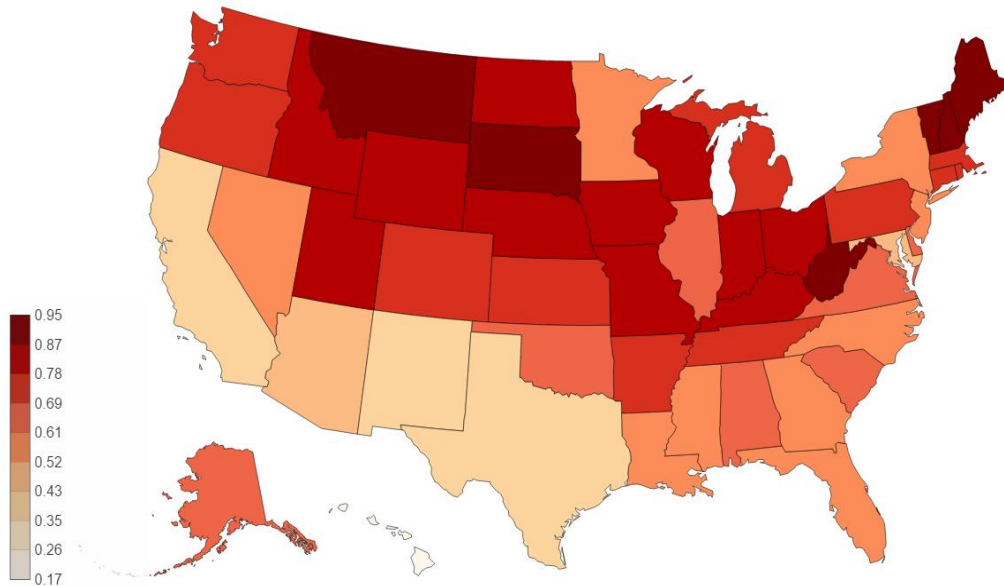

White population density 2013


Pyramid Chart Gender by State

6. Design and Implementation cont

- 6.2.1   Milestone 1 Design Choropleth map overview

- We created two Choropleth, the left one shows the white group population composition, range from 0 to 1, the higher, the more white in the state's population, and the corresponding color ranges from gray to red, the more red, the more white in terms of composition in that state.

- The right one shows the venture capital total spend per state , and the corresponding color ranges from gray to red, the more red, the more venture capital spending in that state, the CA and NY state's data is missing in the map, and we plan to fix that in the next step.

White population density 2013


Venture Capital total spend 2012
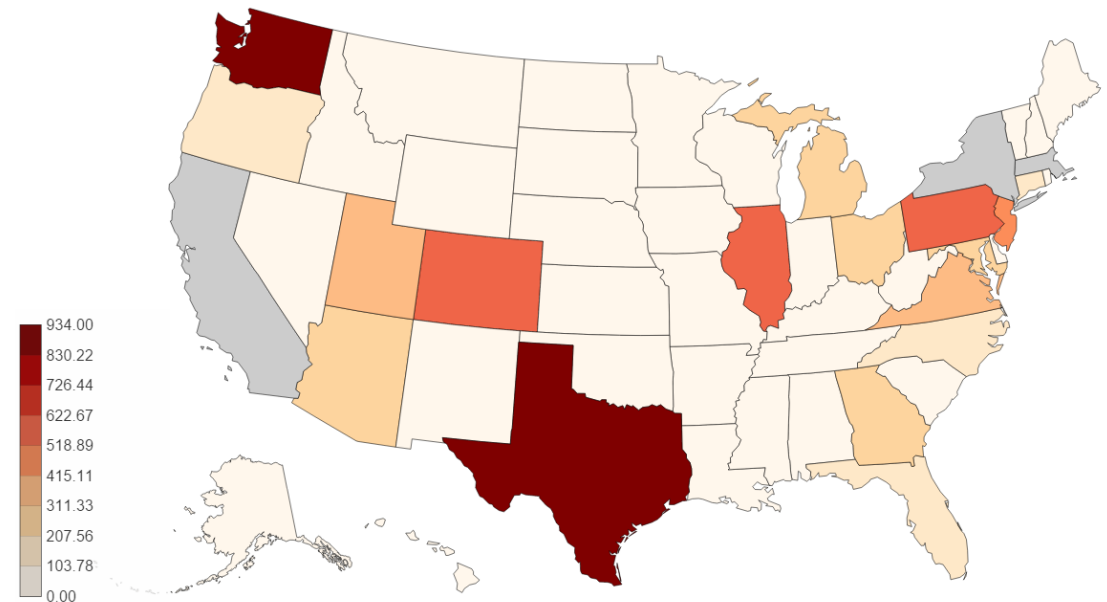
6. Design and Implementation cont

6.2.2   Milestone 1 Design Choropleth map overview cont

- **Zoomable and Hover over feature**:   If we click on the map, the view will be zoomed in and show the details of the map, also, the state's details will be presented in a tooltip window as the following example shows Arizona's data.

White population density 2013
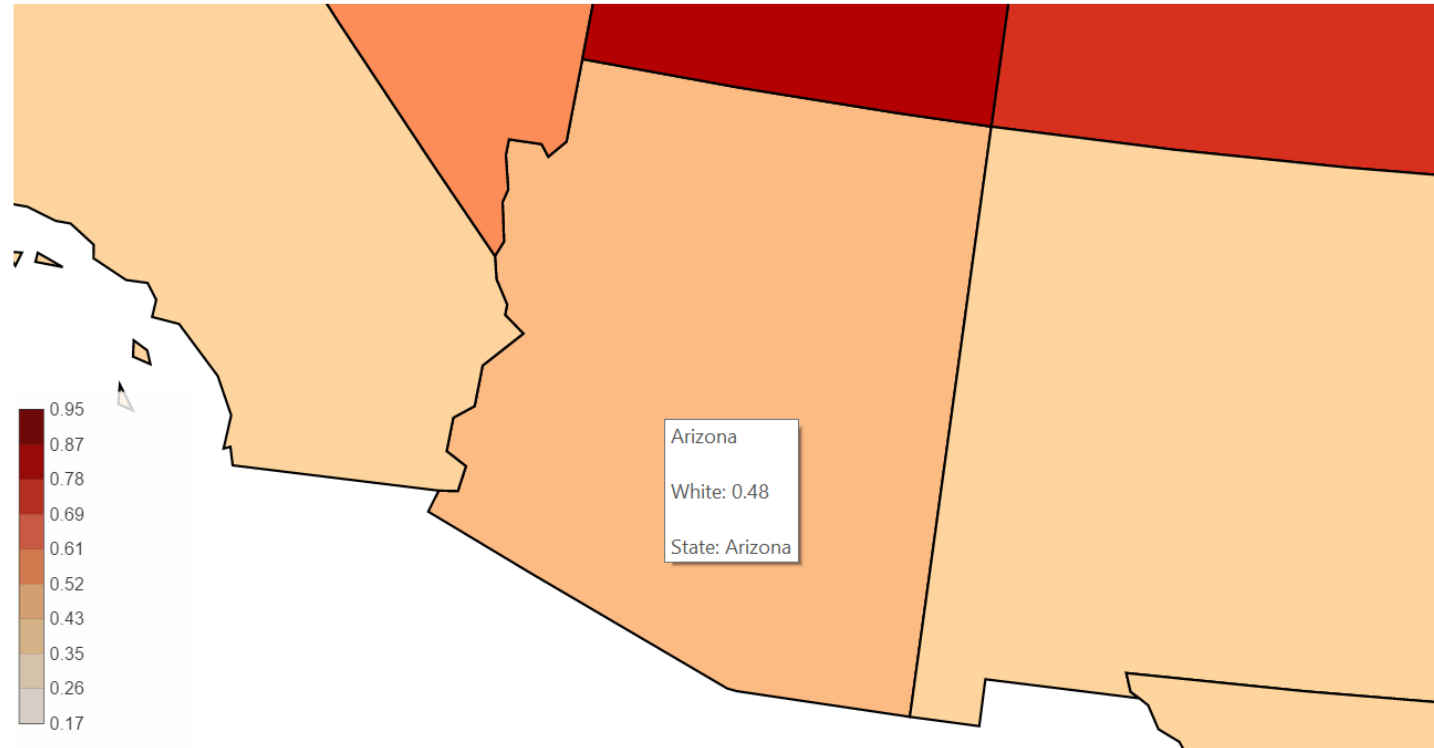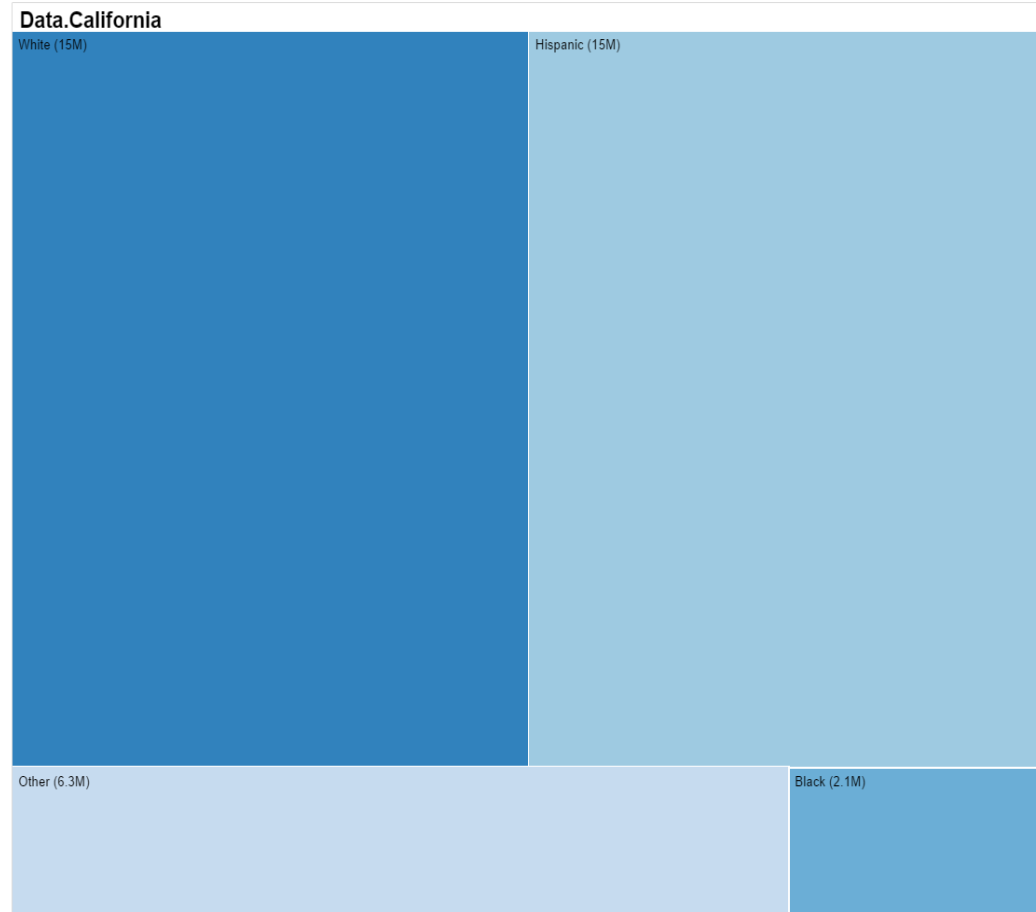
6. Design and Implementation cont
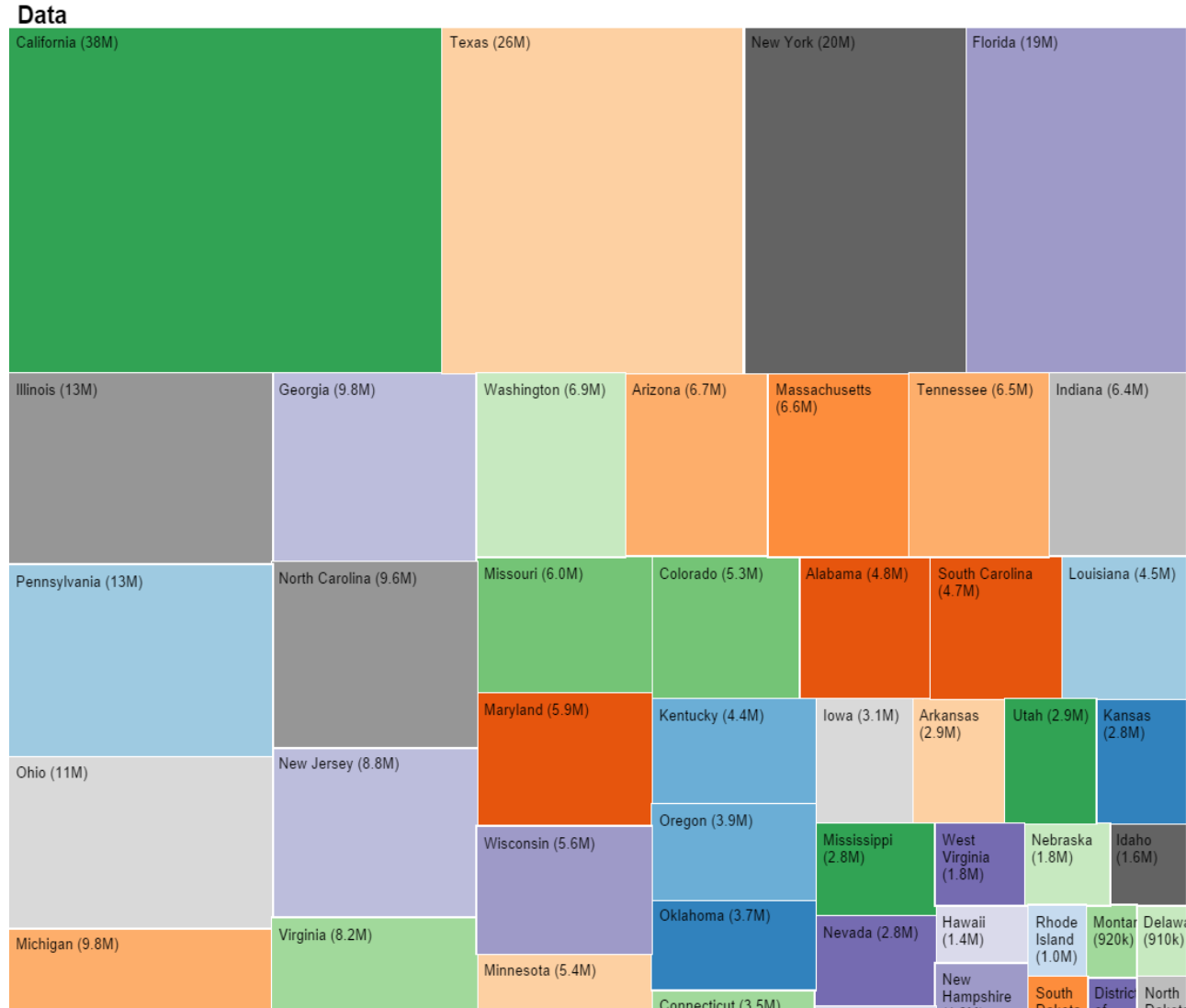
- 6.2.3 Milestone 1 Design Tree map overview

- Treemap is a very good method to display hierarchical data. The color and size dimensions are correlated with the tree strcture, one can easily see patterns that would be difficult to spot in other ways.

- Here we use 20 colors to differentiate each state, and the size corresponds to the population at parent level, and at child level, each ethnicity group's population is presented in different sizes of square as well.

6. Design and Implementation cont

6.2.3   Milestone 1 Design Tree map overview cont

- If we click any state, the Treemap will goes one layer down to show its race population in squares, in the following example we use clicked California's square and it presents CA's race group population.
- The right side DOM shows the hierarchical structure of the data used in Treemap.


```
Data.California
White (15M)              Hispanic (15M)




                                        
                                        
                                        
                                        


Other (6.3M)            Black (2.1M)
```

```
▼root: Object
  ▼children: Array[51]
    ▼0: Object
      ▼children: Array[4]
        ▼0: Object
          ▼children: Array[1]
            ▼0: Object
                Race: "White"
                State: "Alabama"
                value: 3192900
              ►    proto  : Object
              length: 1
            ►    proto  : Array[0]
            name: "White"
          ►    proto  : Object
        ▼1: Object
          ▼children: Array[1]
            ▼0: Object
                Race: "Black"
                State: "Alabama"
                value: 1252100
              ►    proto  : Object
              length: 1
            ►    proto  : Array[0]
            name: "Black"
          ►    proto  : Object
```

6. Design and Implementation cont

- 6.2.4 Milestone 1 Design Pyramid Bar Chart Overview
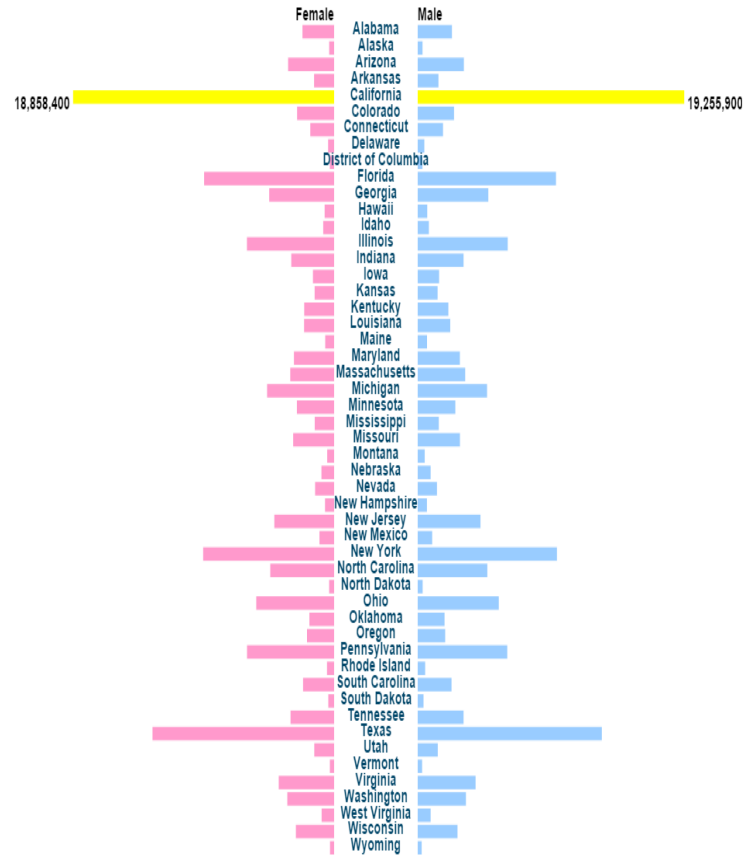
- Pyramid bar chart is a very good method to show two contrasting group's data in different unit, it allows user to compare at two levels: horizontal and vertical, therefore we use this model to compare Female vs Male population at 50 states.

- We also introduced hover feature , when mouse over, the bar will be highlighted and show the exact population numbers at two gender groups.

7. Design and Implementation To Do

- Create a coordinated view framework to link all different visualization in one graph using event handler, brushing and other d3 methods.

- Create interactive tools such as filter, slider, selection box to allow users to view datasets of their interests

- Add more datasets such as income, age, education, etc to enrich the visualization

- Create ways to help user find particular patterns and correlations among different attributes of data.

- Fix the venture captial Choropleth map CA,NY data showing NaN error

- Add interactive features to allow users to select

- Add a year slider to venture capital to show yearly change

- Add a area or line chart to show annual vc spend trend the whole country and then by state when selected by user

- Add a linked vis between age and gender distribution, e.g. brush a certain agerange, will update the gender distribution of that age range

- Other design and implementations as project evolves.