

US Census Visualization Process Book

CS171-Visualization

Dingchao Zhang, Simon Malian

dingchaoz.github.io

Index

- 1. Overview and Motivation
- 2. Related Work
- 3. Questions
- 4. Data
- 5. Exploratory Data Analysis
- 6. Design Evolution
- 7. Implementation
- 8. Evaluation
- 9. Libraries used
- 10. Thank you

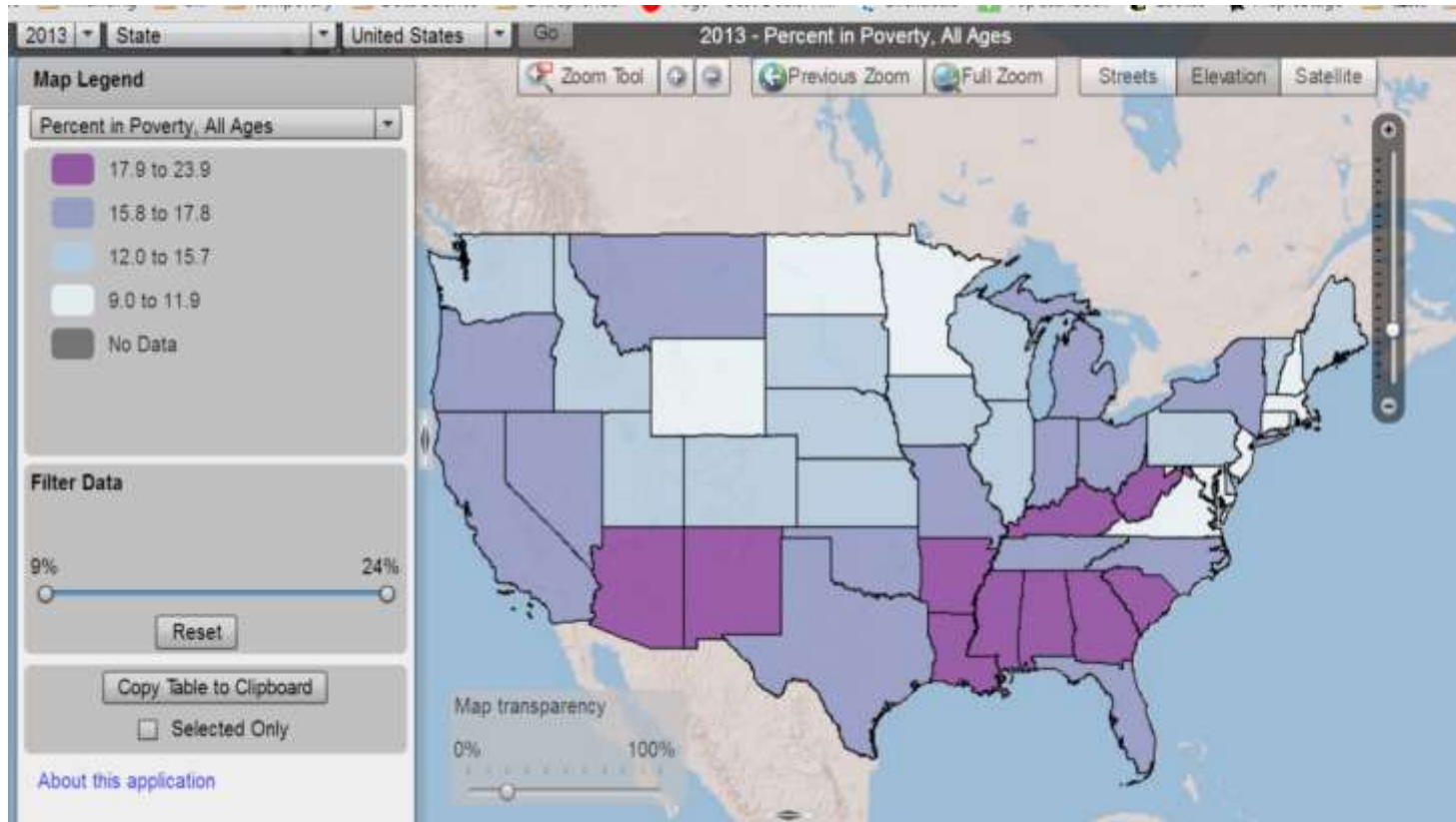
1. Overview and Motivation

- 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:



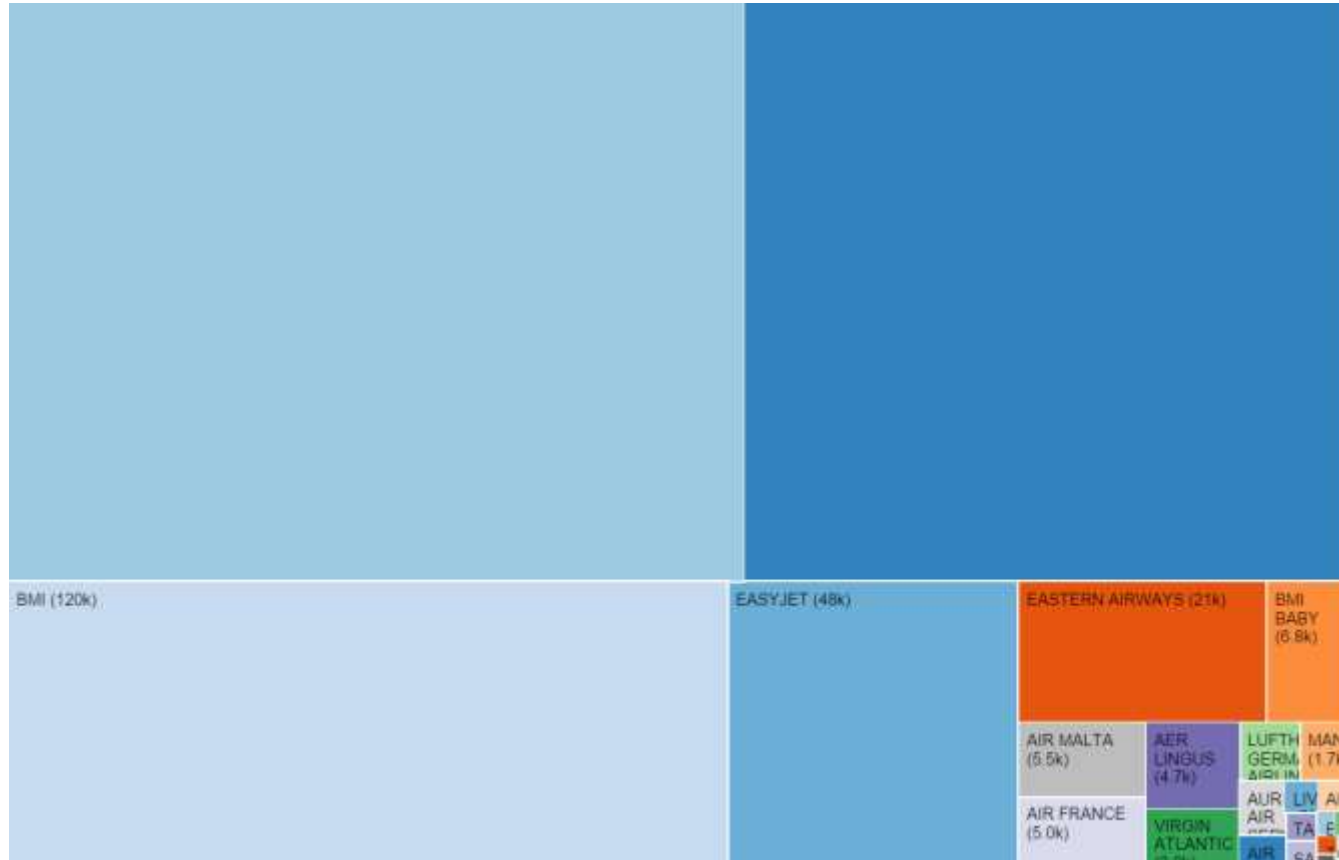
What we can do better than US Census:

- 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 structure, 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 and ethnicity group distribution.
- As the project goes, we would like also to consider new questions such as besides ethnicity distribution on location wise, would there be other patterns of age, gender, year dimension related to ethnicity group demographics change.

4. Data

- Source

We pulled available CSVs with all 2010 Census data from US Census Bureau Web; the CSV is released separately for each state

<http://www.census.gov/popest/data/counties/asrh/2013/CC-EST2013-ALLDATA.html>.

A screen shot of data table:

SUMLEV	STATE	COUNTY	STNAME	CTYNAME	YEAR	AGEGRP	TOT_POP	TOT_MAL	TOT_FEM	WA_MALE	WA_FEM	BA_MALE	BA_FEM	IA_MALE	IA_FEM	AA_MALE	AA_FEM	H_MALE	H_FEMALE
50	1	1	Alabama	Autauga County	1	0	54571	26569	28002	21295	22002	4559	5130	119	139	200	284	694	616
50	1	1	Alabama	Autauga County	1	1	3579	1866	1713	1411	1316	362	317	5	3	13	15	88	62
50	1	1	Alabama	Autauga County	1	2	3991	2001	1990	1521	1526	399	374	14	8	17	21	68	74
50	1	1	Alabama	Autauga County	1	3	4290	2171	2119	1658	1620	431	406	15	12	23	18	66	64
50	1	1	Alabama	Autauga County	1	4	4290	2213	2077	1628	1585	502	424	12	7	25	14	60	51
50	1	1	Alabama	Autauga County	1	5	3080	1539	1541	1201	1184	293	312	6	9	8	7	65	42
50	1	1	Alabama	Autauga County	1	6	3157	1543	1614	1234	1223	277	350	1	3	7	23	66	44
50	1	1	Alabama	Autauga County	1	7	3330	1594	1736	1289	1298	276	378	1	10	12	25	61	62
50	1	1	Alabama	Autauga County	1	8	4157	2004	2153	1596	1645	338	432	11	9	30	30	56	45
50	1	1	Alabama	Autauga County	1	9	4086	1974	2112	1625	1658	303	384	9	14	16	36	51	39
50	1	1	Alabama	Autauga County	1	10	4332	2174	2158	1810	1744	315	353	14	11	19	27	34	47
50	1	1	Alabama	Autauga County	1	11	3873	1866	2007	1508	1573	323	385	11	12	10	17	24	21
50	1	1	Alabama	Autauga County	1	12	3083	1524	1559	1265	1232	237	302	7	9	6	9	20	13
50	1	1	Alabama	Autauga County	1	13	2777	1279	1498	1086	1237	171	220	6	13	7	13	9	15
50	1	1	Alabama	Autauga County	1	14	2277	1014	1263	873	1056	128	177	3	8	4	9	12	17
50	1	1	Alabama	Autauga County	1	15	1736	807	929	688	796	109	114	4	5	2	8	6	9
50	1	1	Alabama	Autauga County	1	16	1251	546	705	498	597	46	92	0	6	1	7	3	7
50	1	1	Alabama	Autauga County	1	17	731	295	436	268	368	26	63	0	0	0	5	4	2
50	1	1	Alabama	Autauga County	1	18	551	159	392	136	344	23	47	0	0	0	0	1	2
50	1	1	Alabama	Autauga County	2	0	54571	26569	28002	21295	22002	4559	5130	119	139	200	284	694	616
50	1	1	Alabama	Autauga County	2	1	3579	1866	1713	1411	1316	362	317	5	3	13	15	88	62

4. Meta data

VARIABLE	DESCRIPTION
SUMLEV	Geographic Summary Level
STATE	State FIPS Code
COUNTY	County FIPS Code
STNAME	State Name
CTYNAME	County Name
YEAR	Year
AGEGRP	Age group (See code below)
TOT_POP	Total population
TOT_MALE	Total male population
TOT_FEMALE	Total female population
WA_MALE	White alone male population
WA_FEMALE	White alone female population
BA_MALE	Black or African American alone male population
BA_FEMALE	Black or African American alone female population
IA_MALE	American Indian and Alaska Native alone male population
IA_FEMALE	American Indian and Alaska Native alone female population
H_MALE	Hispanic male population
H_FEMALE	Hispanic female population

The key for AGEGRP is as follows:

- 0 = Total
- 1 = Age 0 to 4 years
- 2 = Age 5 to 9 years
- 3 = Age 10 to 14 years
- 4 = Age 15 to 19 years

The key for the YEAR variable is as follows:

- 1 = 4/1/2010 Census population
- 2 = 4/1/2010 population estimates base
- 3 = 7/1/2010 population estimate
- 4 = 7/1/2011 population estimate
- 5 = 7/1/2012 population estimate
- 6 = 7/1/2013 population estimate

-
- 5 = Age 20 to 24 years
 - 6 = Age 25 to 29 years
 - 7 = Age 30 to 34 years
 - 8 = Age 35 to 39 years
 - 9 = Age 40 to 44 years
 - 10 = Age 45 to 49 years
 - 11 = Age 50 to 54 years
 - 12 = Age 55 to 59 years
 - 13 = Age 60 to 64 years
 - 14 = Age 65 to 69 years
 - 15 = Age 70 to 74 years
 - 16 = Age 75 to 79 years
 - 17 = Age 80 to 84 years
 - 18 = Age 85 years or older

4. Data cont

- Data wrangling

Besides cleaning unnecessary columns, understanding 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.

Initial 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	173600	86800

Change structure

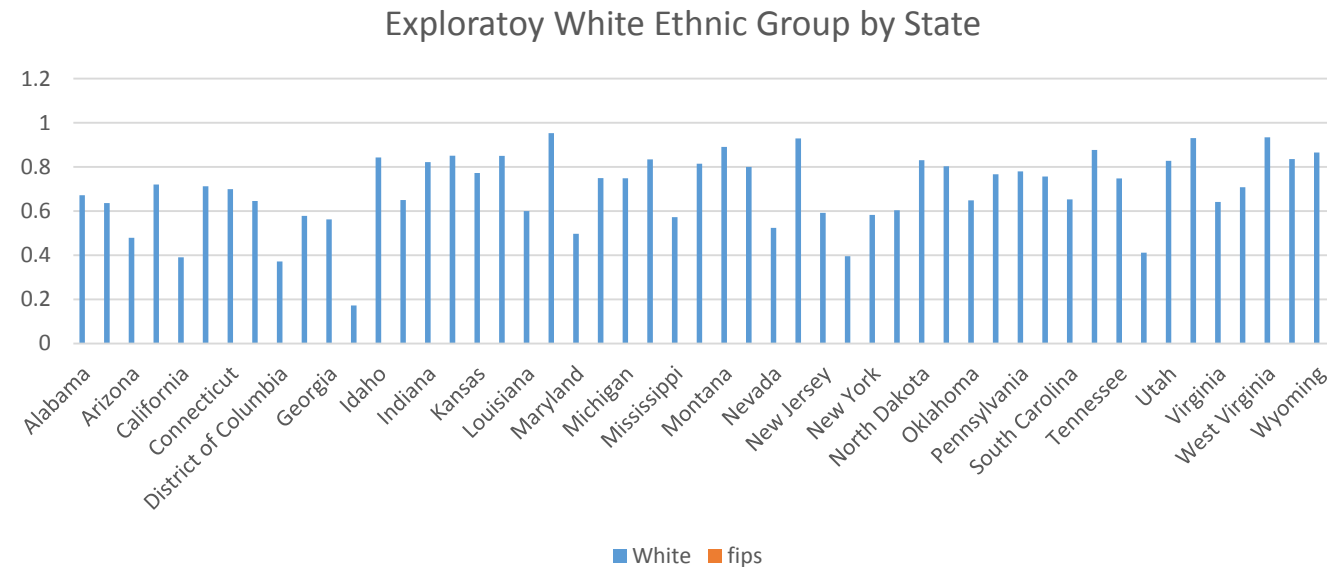
Transformed Hierarchical data

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

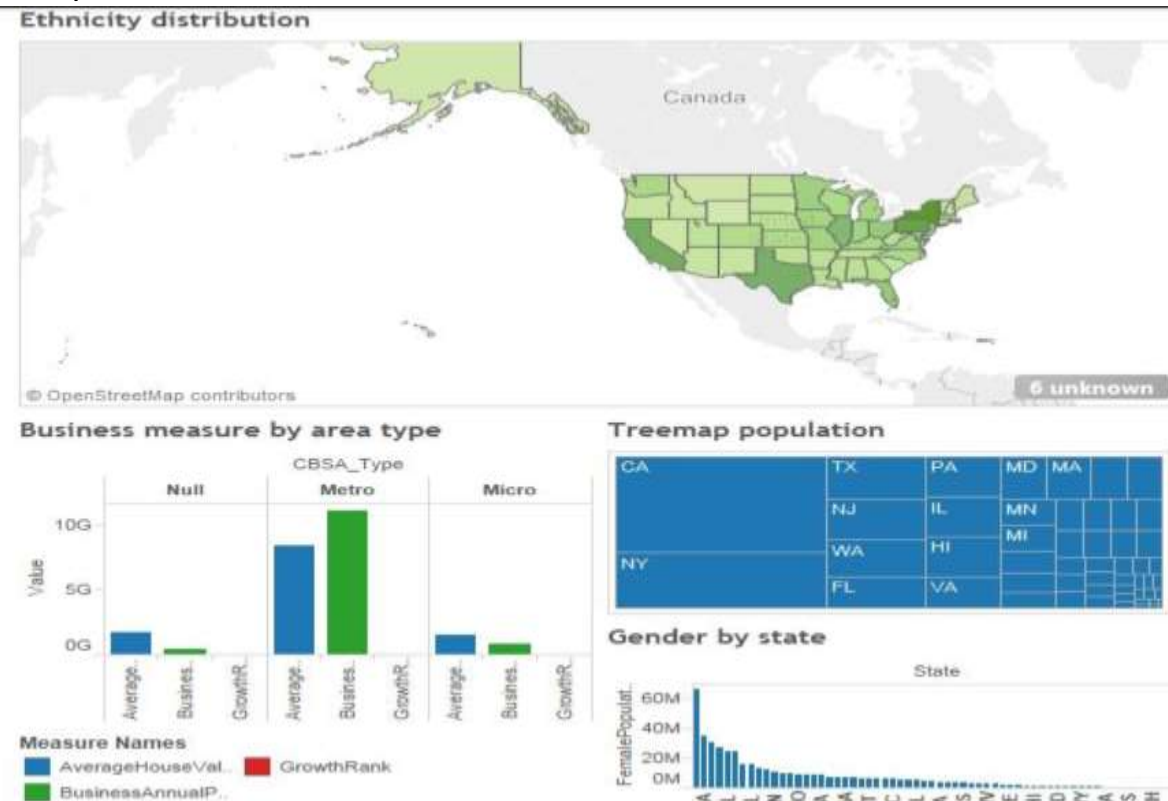
- We first used rudimentary, simple visualizations to take a better look at our data, this included looking at the overall data over different metrics such as gender, race, income; to understand its max, mean, min, median values.

For example, to look at the white group by state composition percentage data, we generate a bar chart in Excel



6. Design Evolution

- 6.1 Proposed Sketch
- 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.

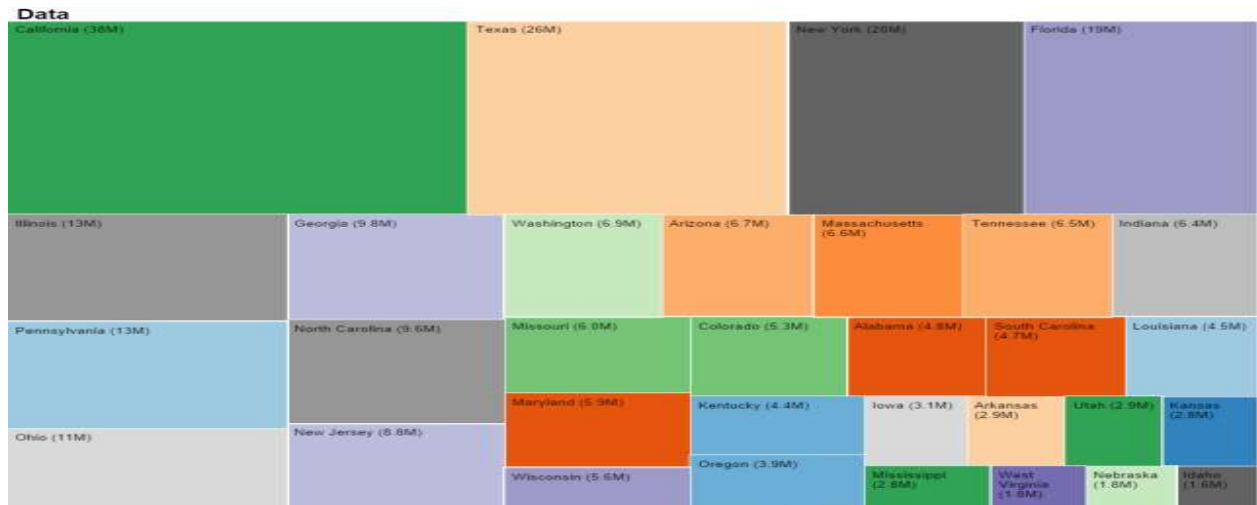
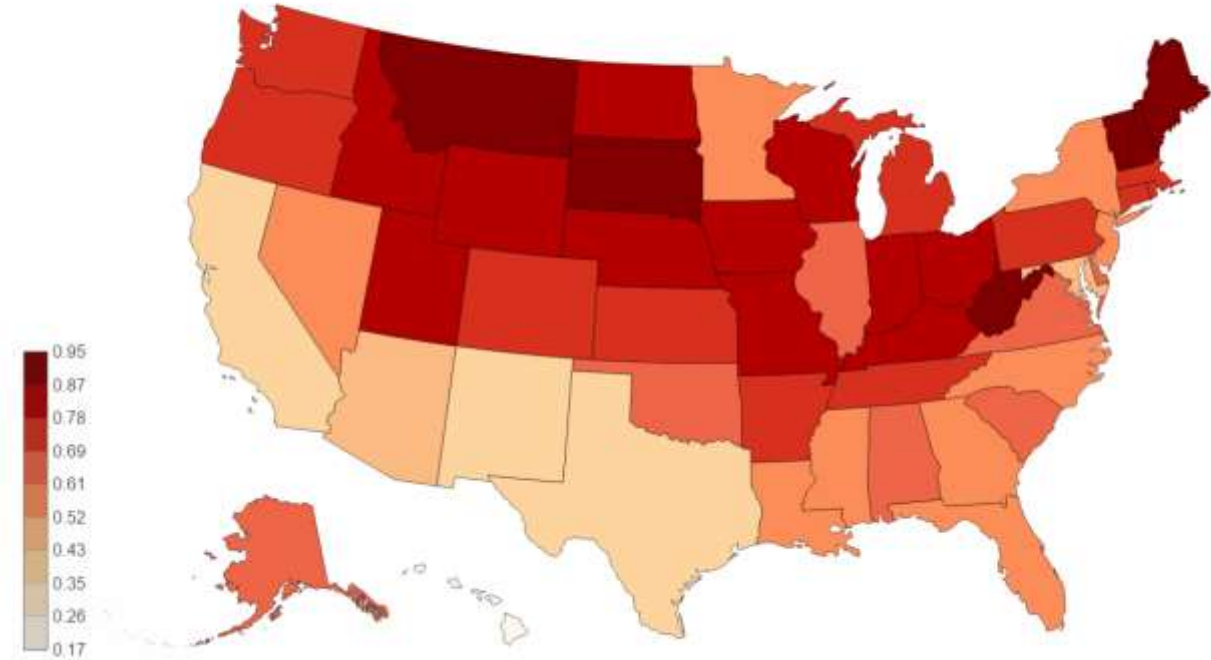
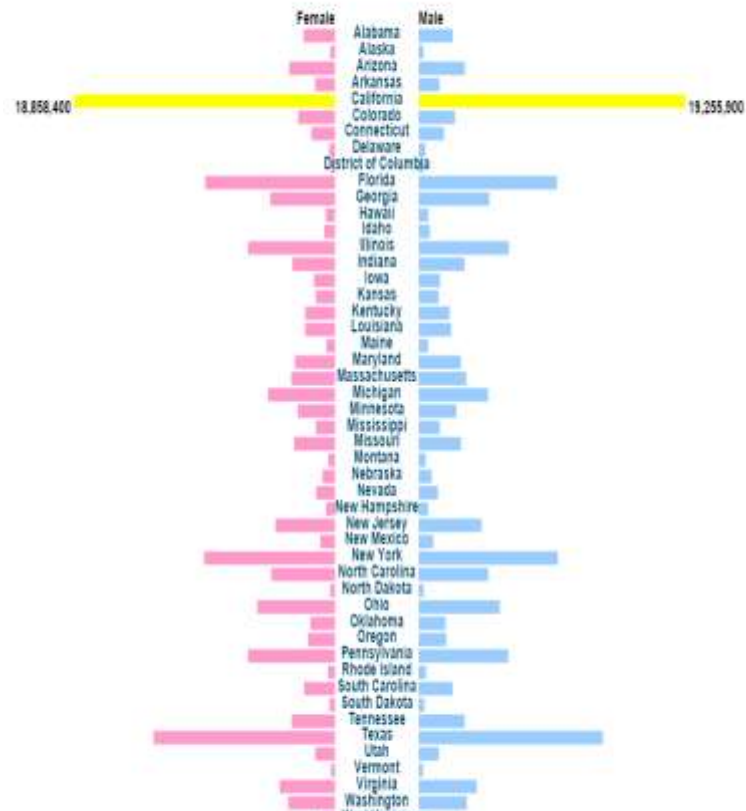


6. Phase 1 Design – Milestone 04/10/2015

White population density 2013

- 6.2 Milestone 1 Design Overview
- Choropleth map, Multiple layer Tree Map, Pyramid Bar charts are created

Pyramid Chart Gender by State



6. Phase 2 Design – Milestone 04/16/2015

US Census Visualization

Year Selected: 2008

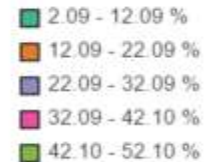
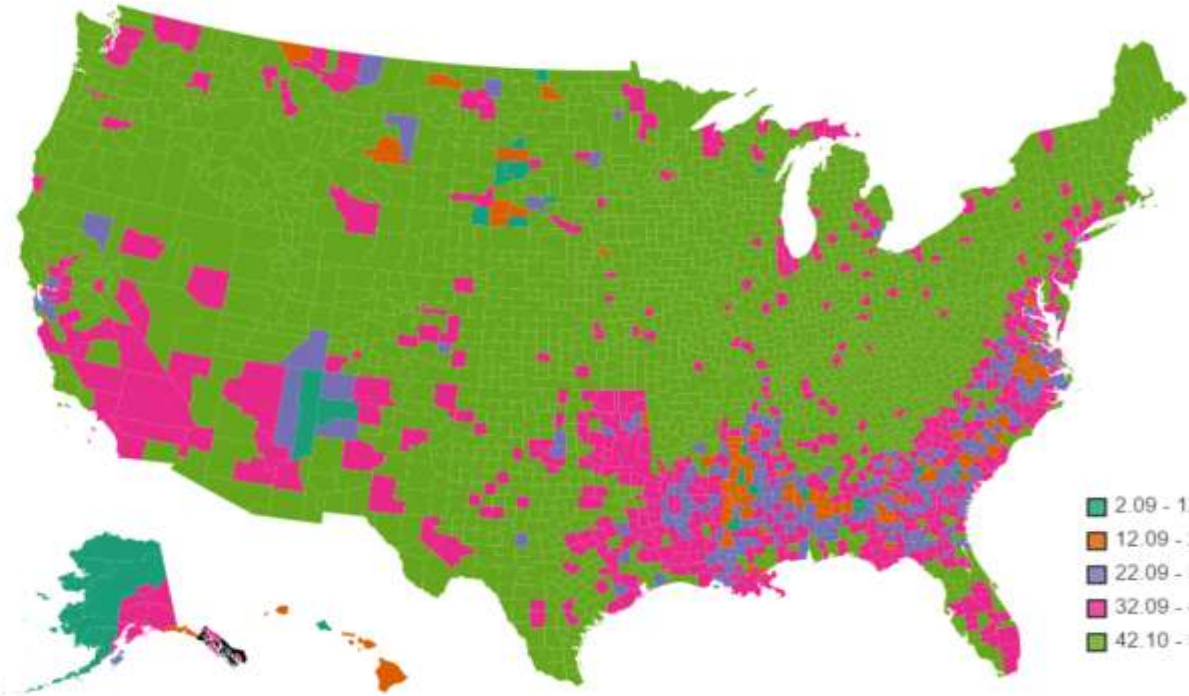


Ethnic group by:

☒ White ☒ Black ☒ Hispanic ☐ Asian ☒ Native

Filter by Age group: All ages ▼

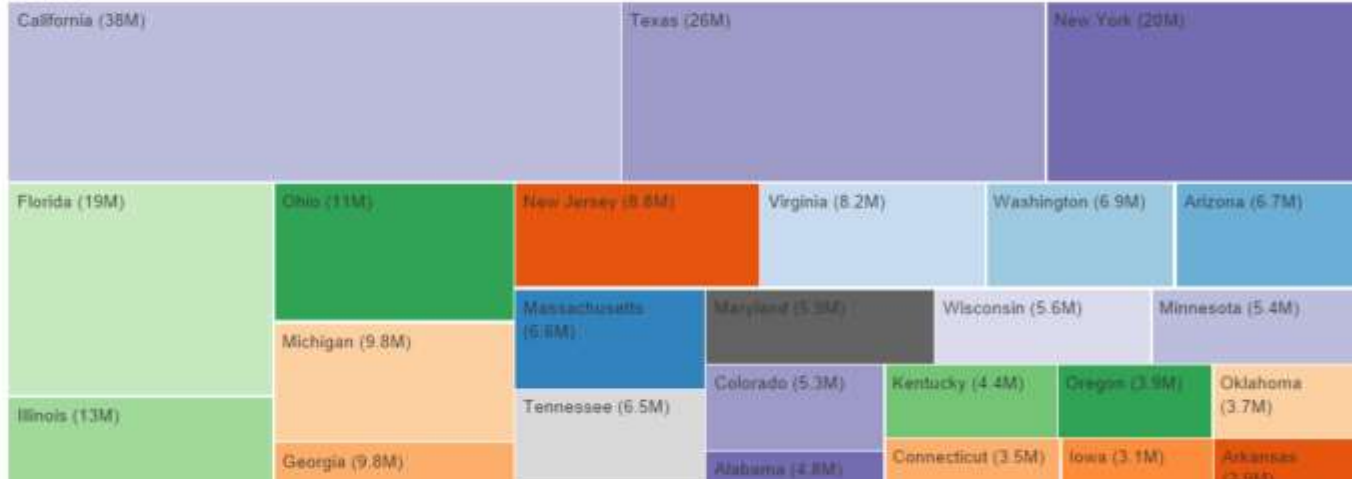
Filter by gender: ☐ Male ☒ Female



Comparing to Design 1, Change 1:
Added an interactive menu
including slider, checkboxes

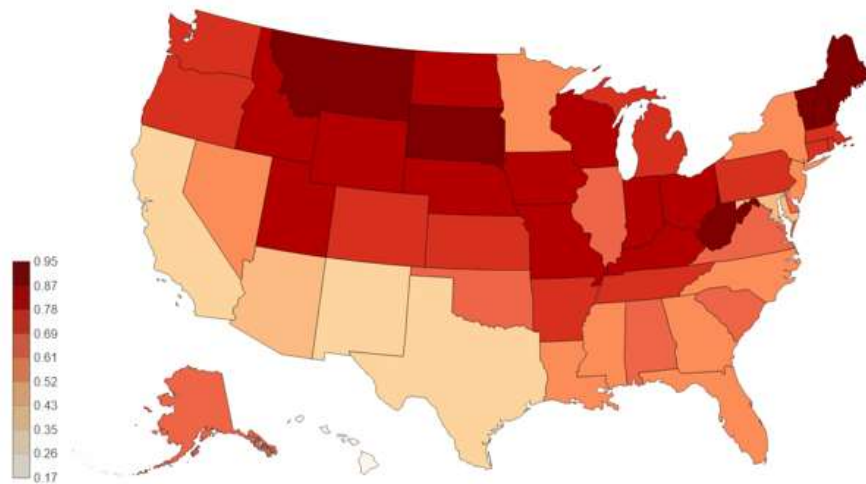


Data



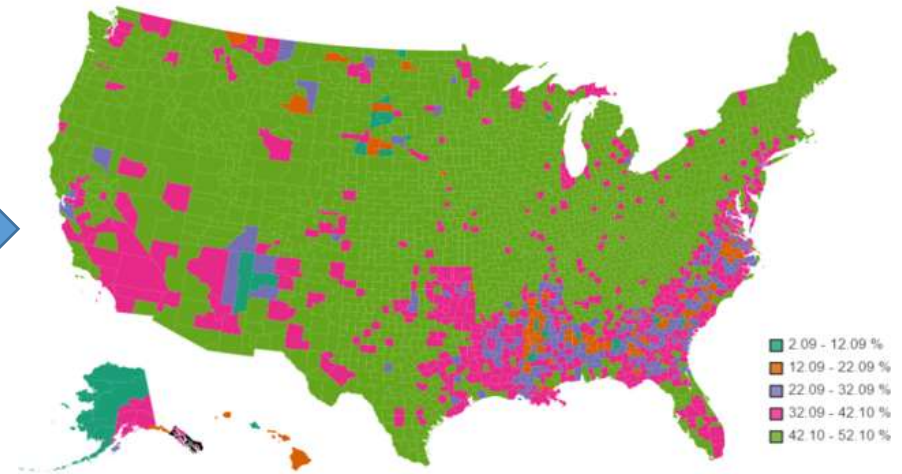
Comparing Design 1 vs Design 2' s choropleth map

Design 1 Choropleth Map



Comparing to Design 1:
Improved data granularity from
state level to county level

Design 2 Choropleth Map



6. Phase 3 Design – Milestone 04/20/2015

US Census Visualization

Year Selected: 2008



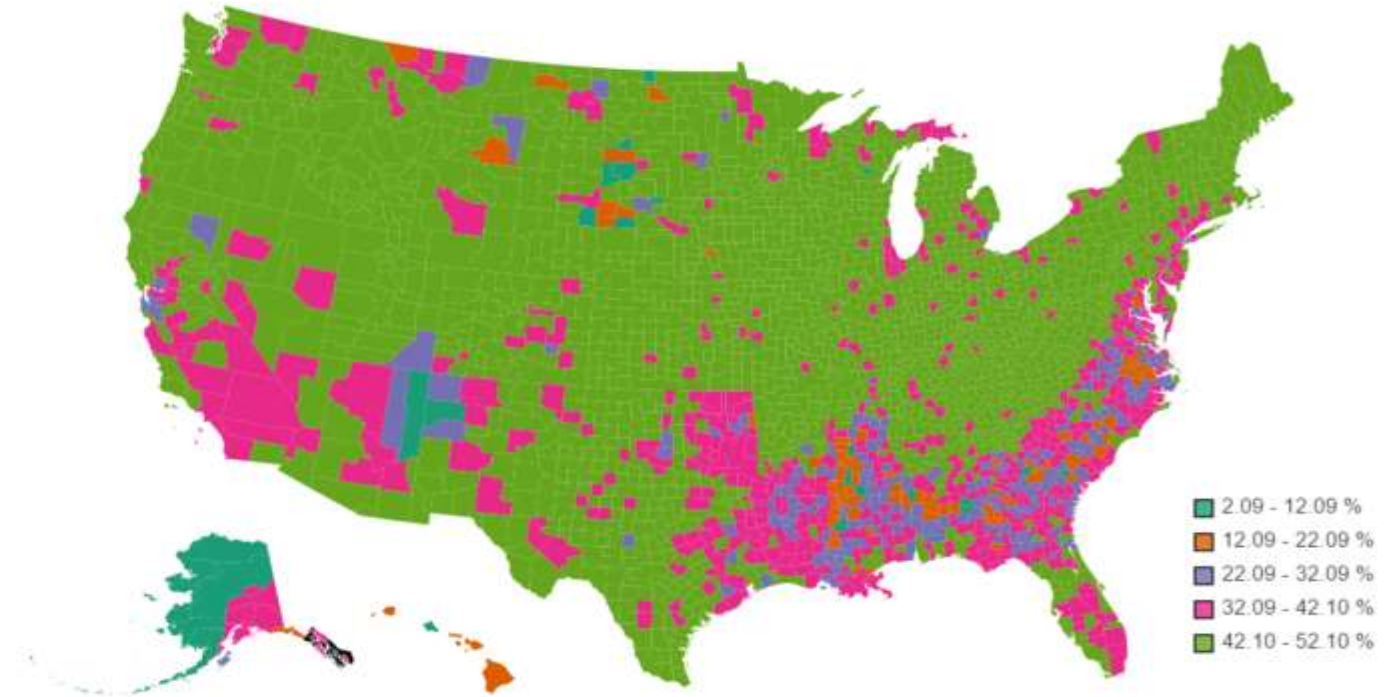
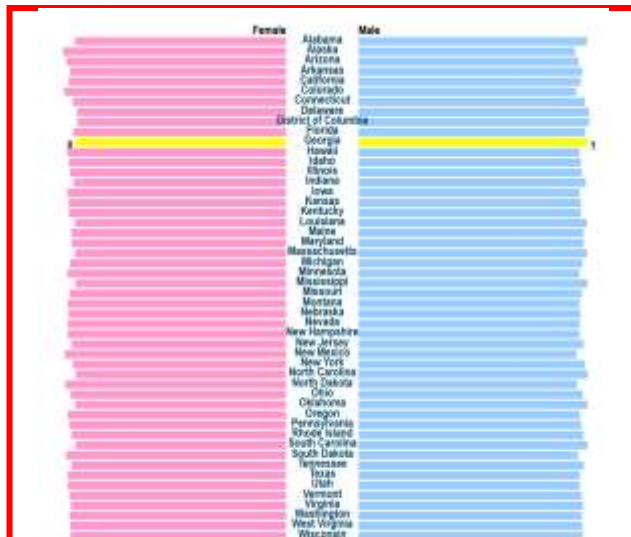
Ethnic group by:

☒ White ☒ Black ☒ Hispanic ☐ Asian ☒ Native

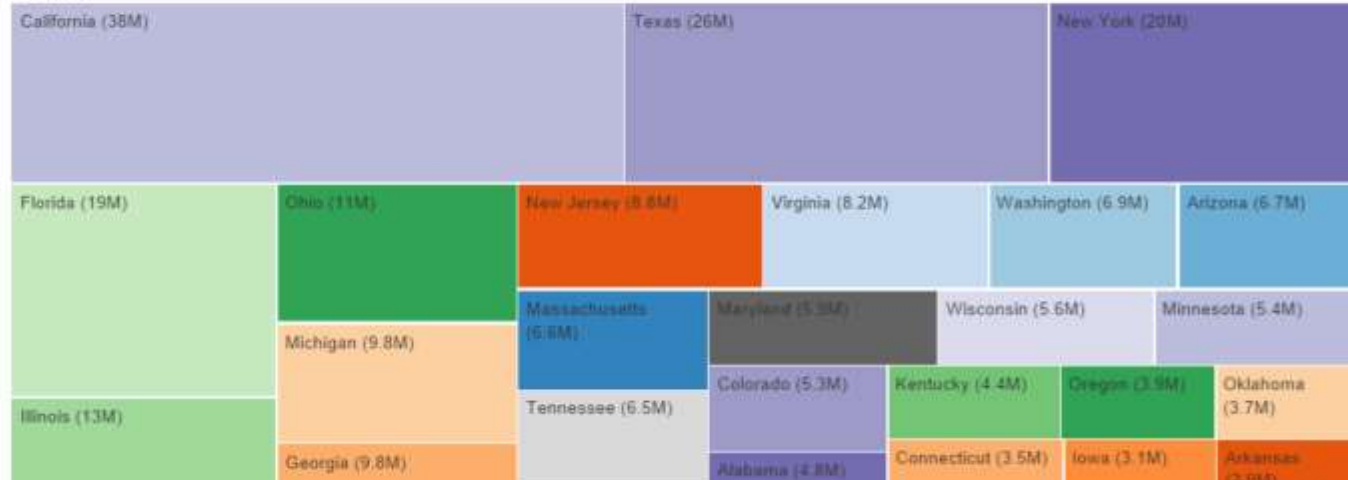
Filter by Age group: All ages ▼

Filter by gender: ☐ Male ☒ Female

Comparing to Design 2, Change:
Added an interactive menu
including slider, checkboxes



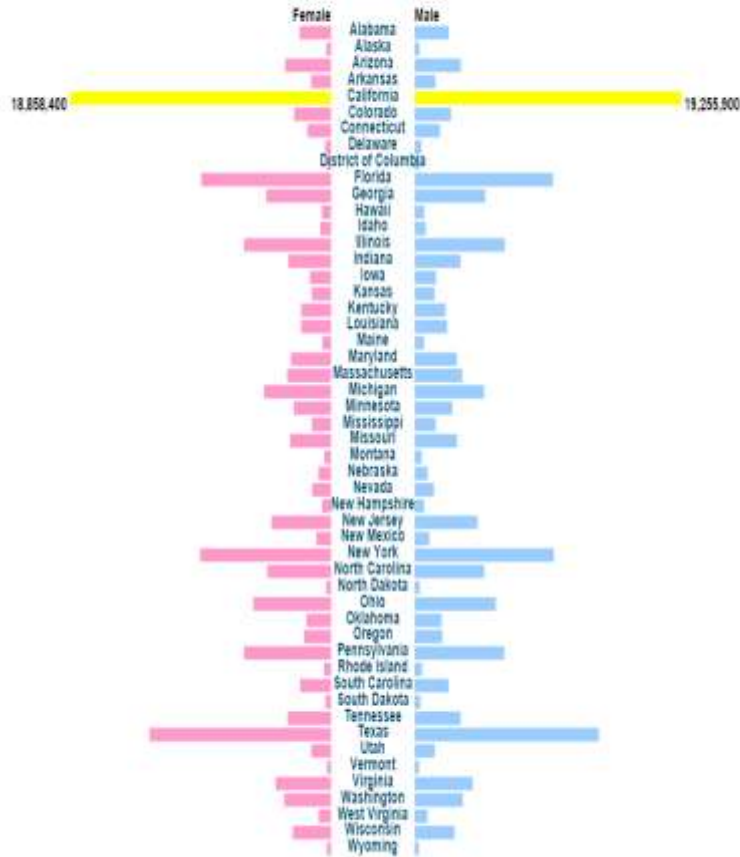
Data



Comparing Design 3 vs Previous Designs' Pyramid bar chart

Previous Design Pyramid Bar

Pyramid Chart Gender by State

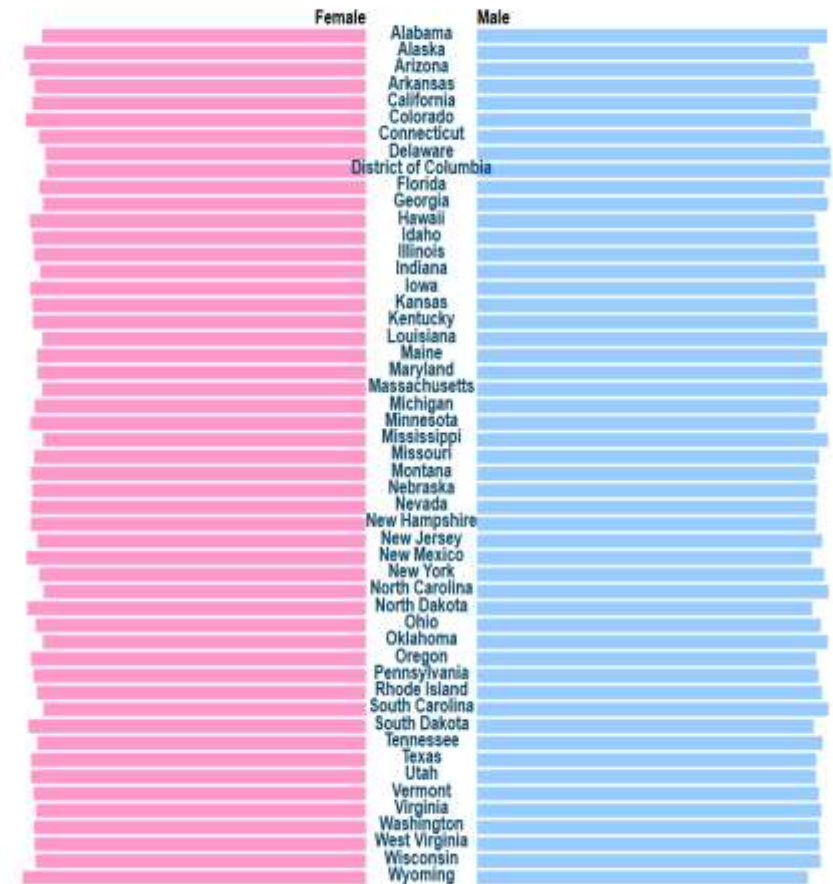


Comparing to previous designs : Gender comparison goes from absolute number to percentage. This turns out not to be a better design, as the difference of gender ratio varies little from state to state.



Design 3 Pyramid Bar

Pyramid Chart Gender by State



6. Phase 4 Design – Milestone 04/26/2015

US Census Visualization

Year Selected: 2008

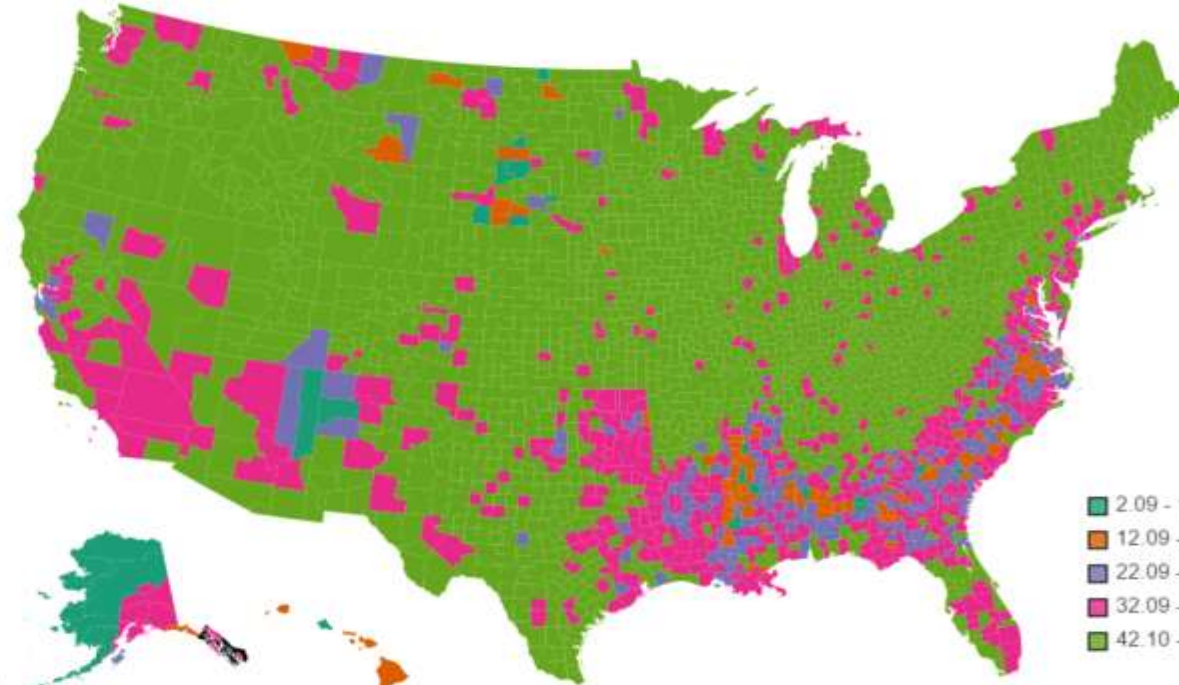


Ethnic group by:

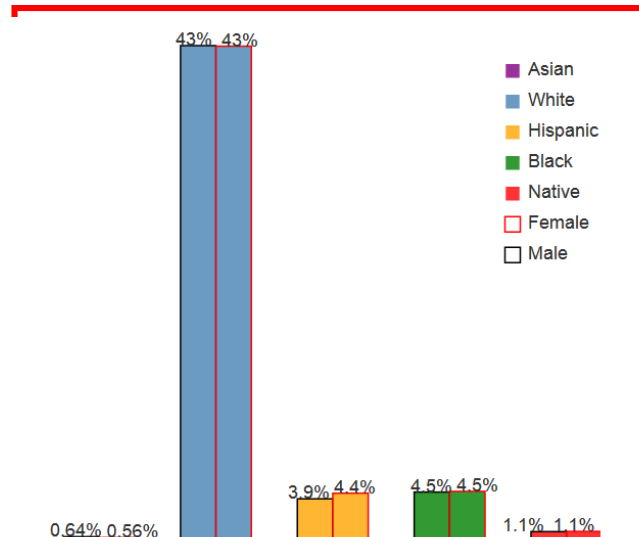
☒ White ☒ Black ☒ Hispanic ☐ Asian ☒ Native

Filter by Age group: All ages

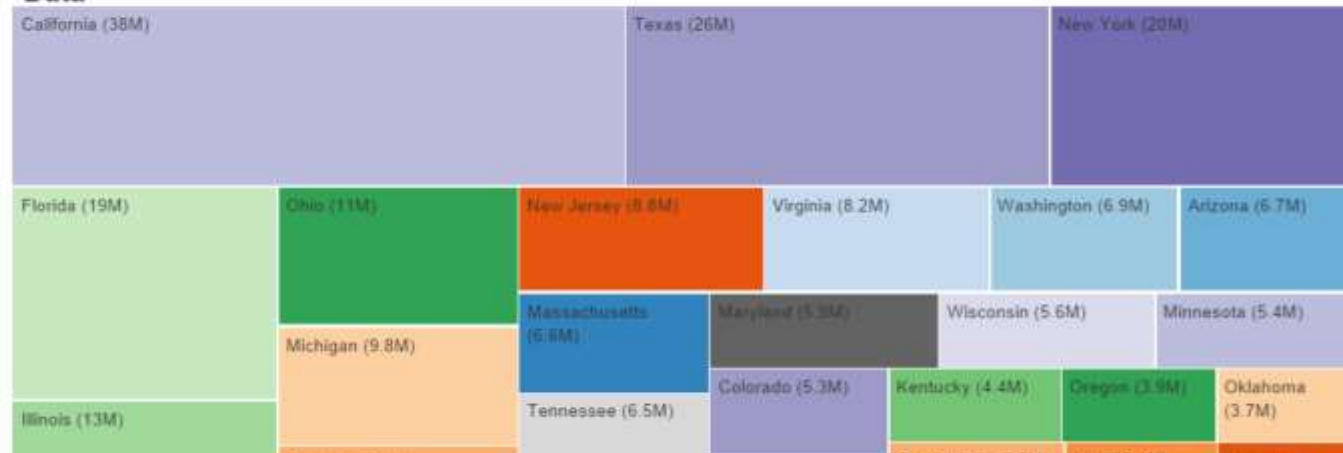
Filter by gender: ☐ Male ☒ Female



Comparing to Previous Design ,
Change:



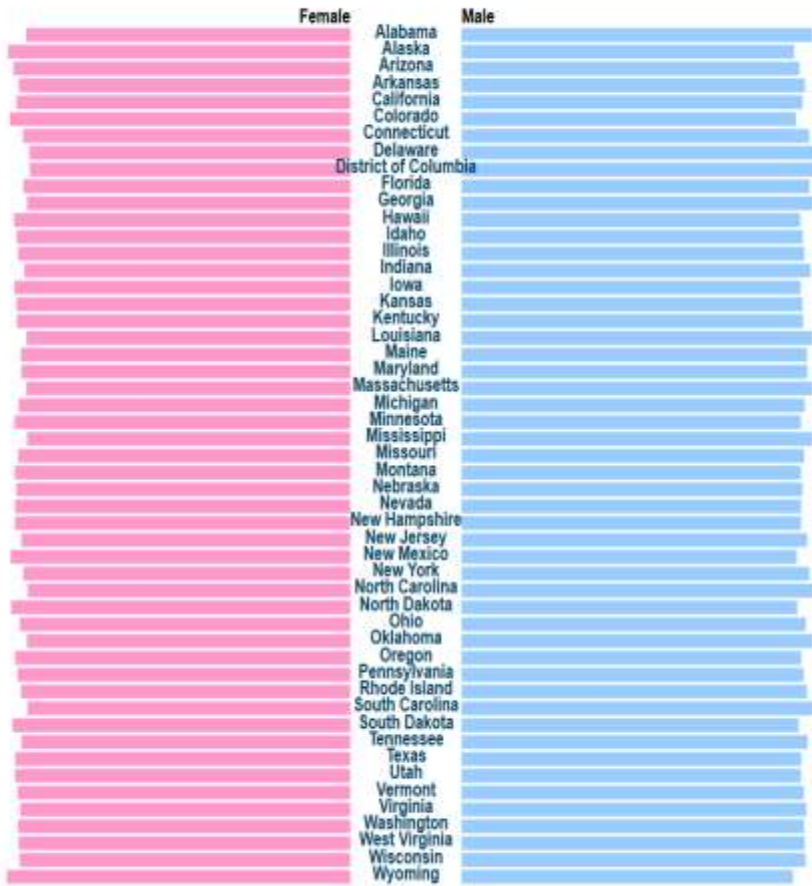
Data



Comparing Design 4 vs Previous Design 's Pyramid bar chart

Previous Design Pyramid Bar

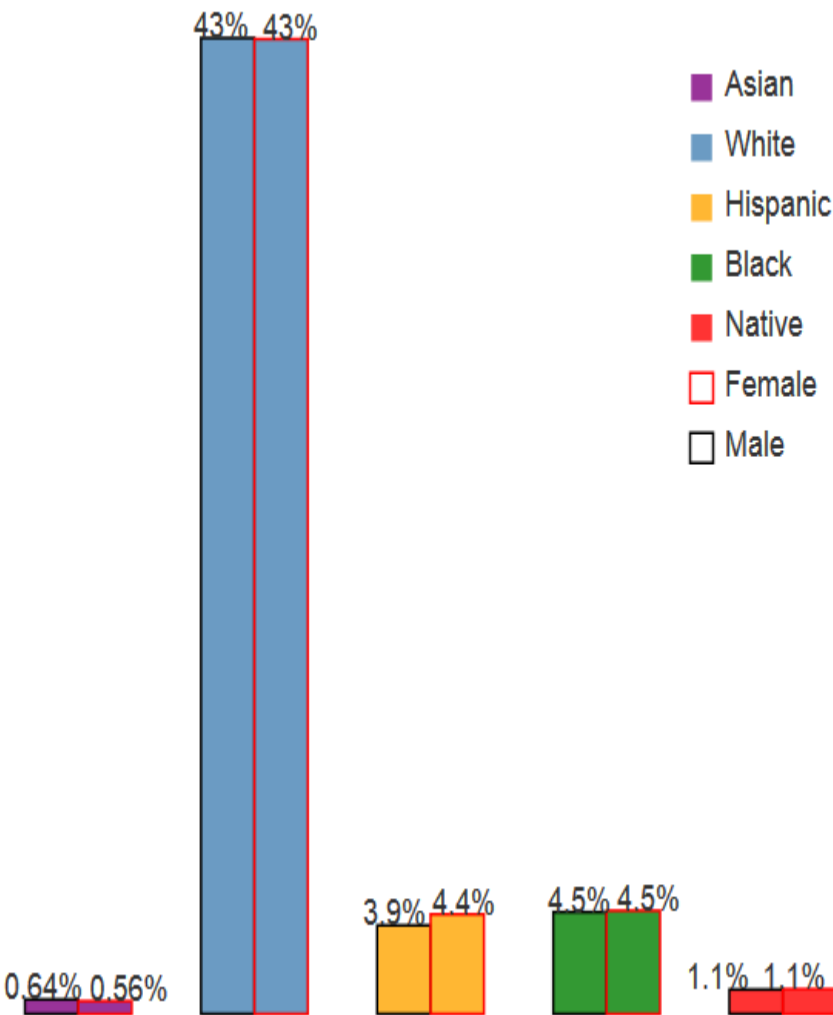
Pyramid Chart Gender by State



Comparing to previous designs:
Since state level gender ratio differs very small, we instead compare ethnic group's gender difference



Design 4 Pyramid Bar



Final Design – 05/05/2015

Year Selected: 2008

Ethnic group by:

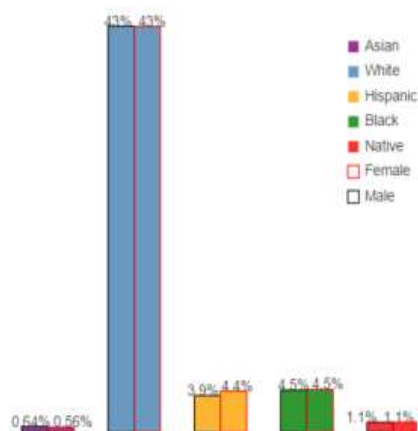
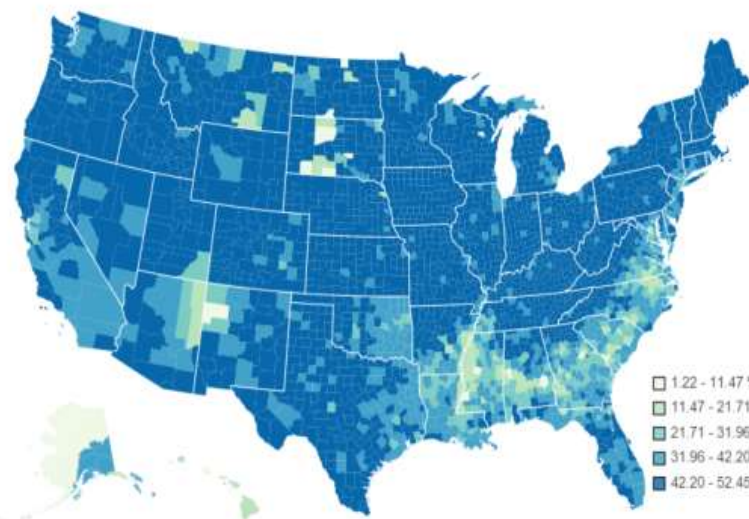
☒ White ☐ Black ☐ Hispanic ☐ Asian ☐ Native

Filter by Age group: All ages

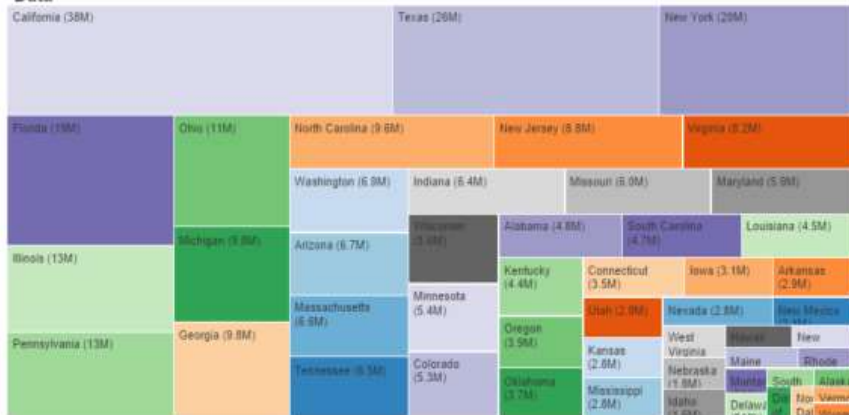
Filter by gender: ☐ Male ☒ Female

Mouse behave: ☐ Tooltip ☒ Brush

To explore United States population demographics changes from 2008 to 2013, use above interactive menu or brush window in the choropleth map. This visualization is created using data released from [United States Census Bureau](#).

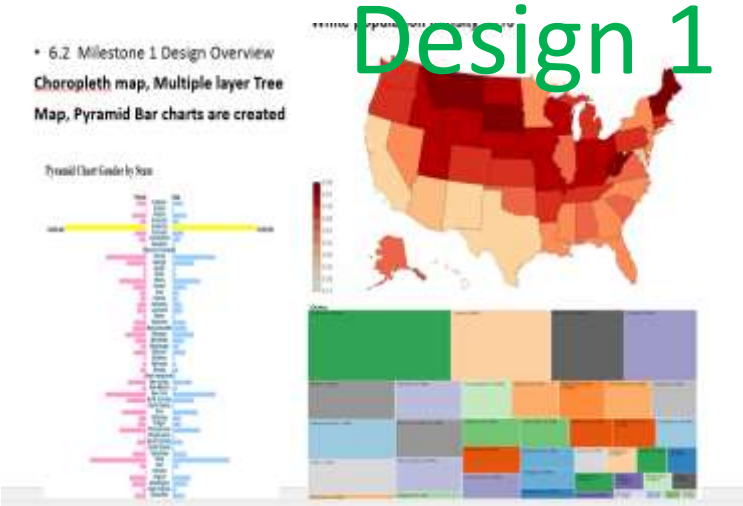


Data



Design Evolutions Recap

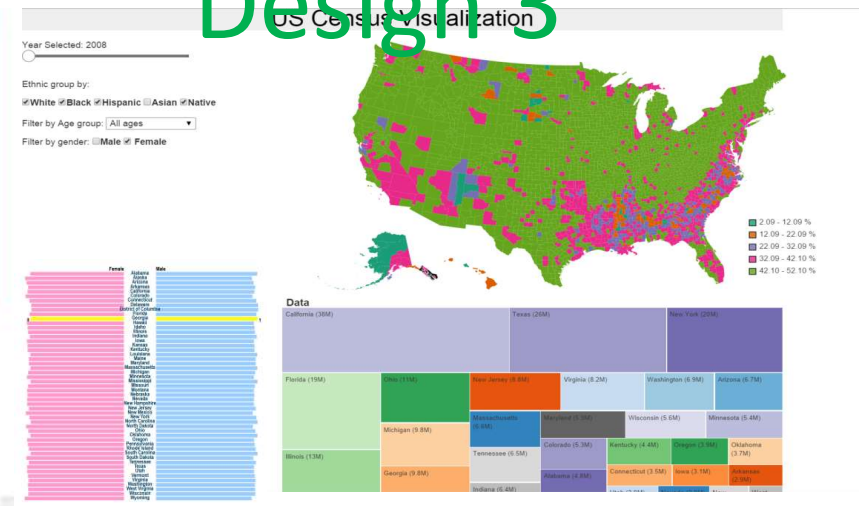
Design 1



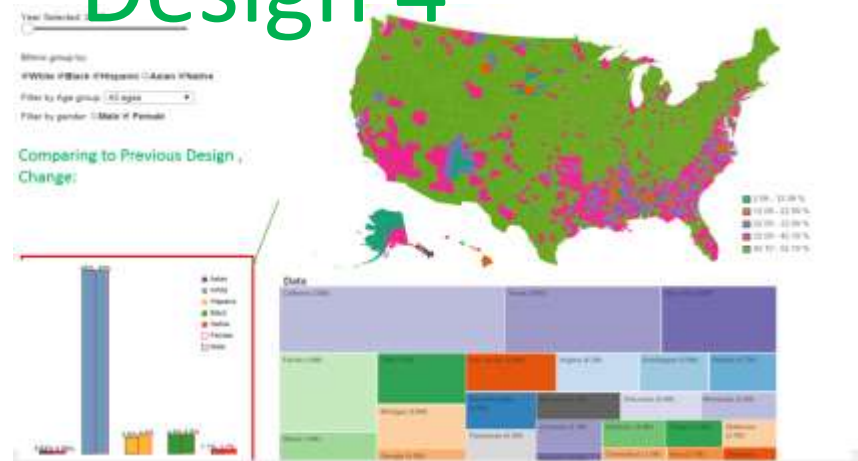
Design 2



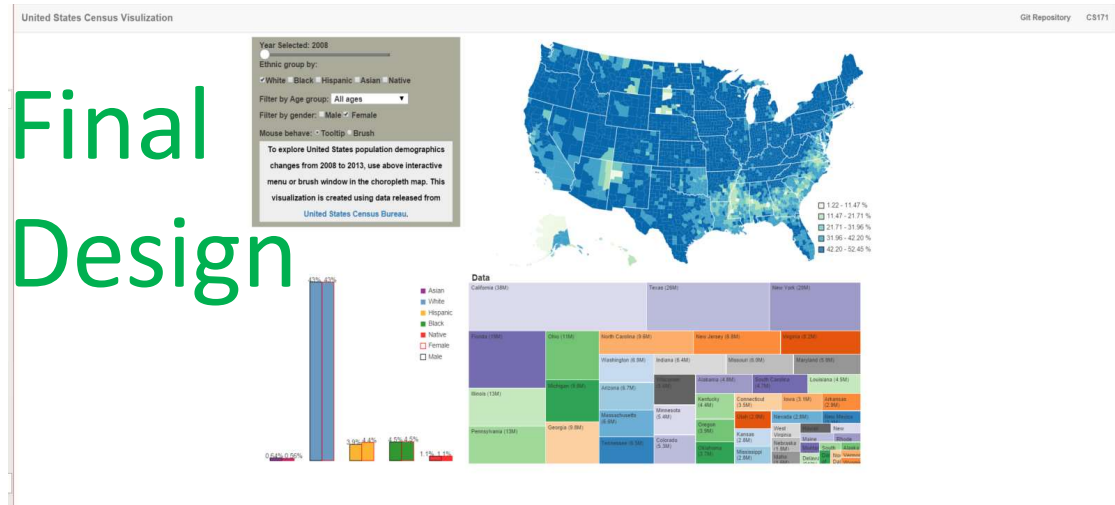
Design 3



Design 4



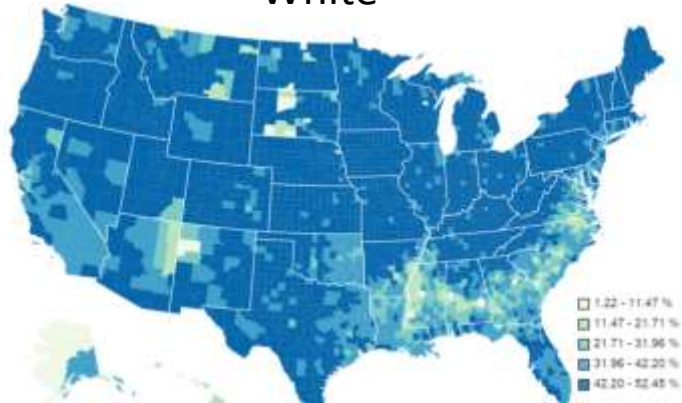
Final Design



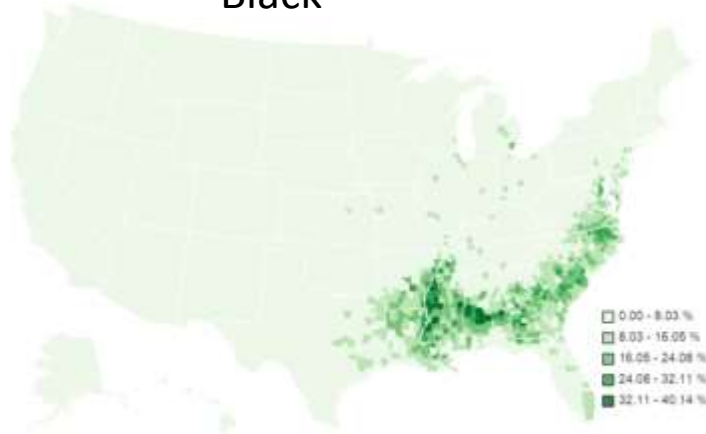
7. Implementation Choropleth

- 5 ethnic group's choropleth map's color schema are in 5 different colors.

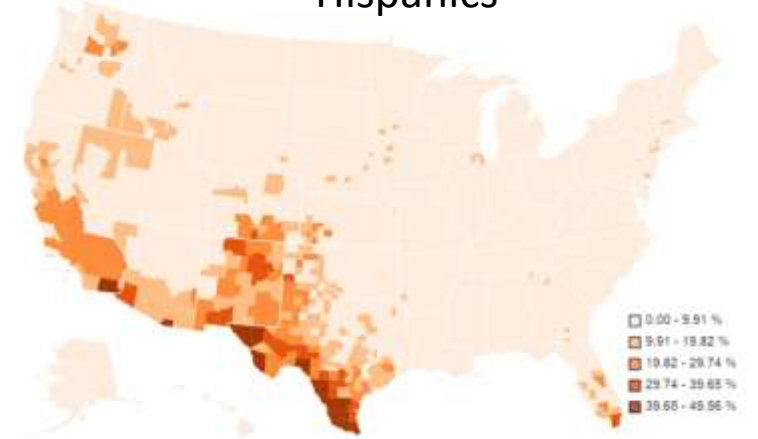
White



Black



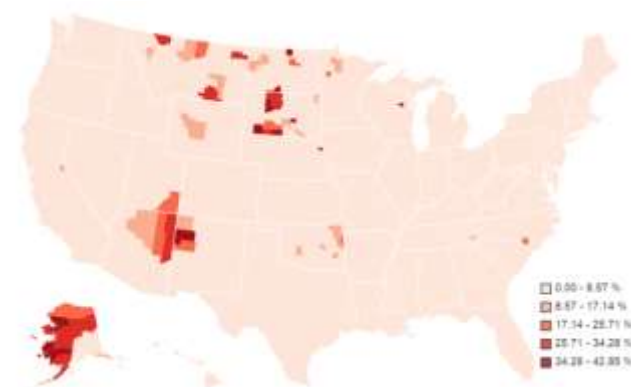
Hispanics



Asian



Natives



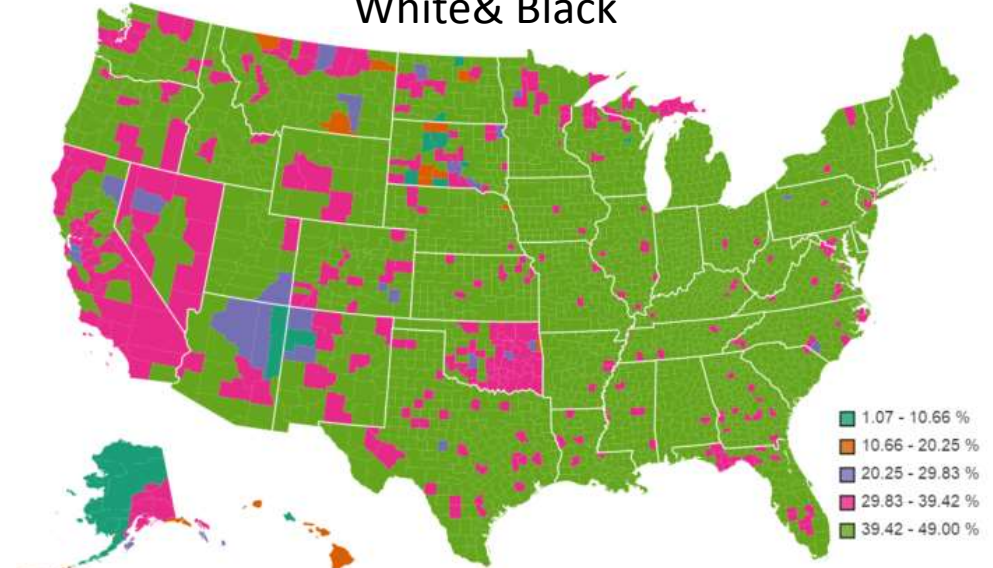
7. Implementation Choropleth cont

- User can also select multiple ethnic groups display at the same time

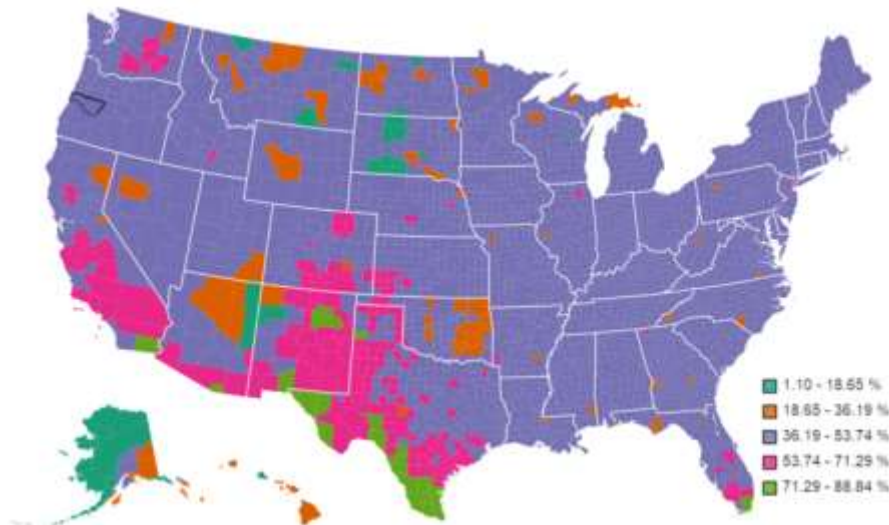
Asian& Natives



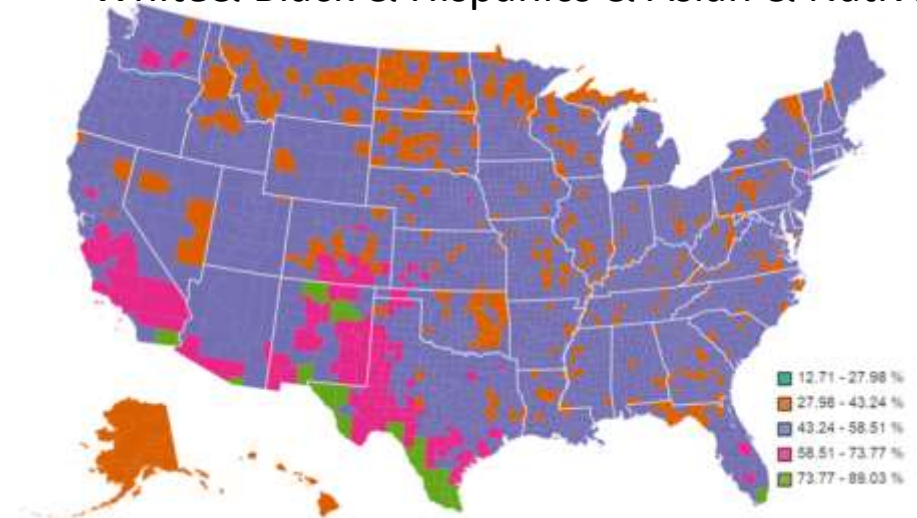
White& Black



White& Black & Hispanics



White& Black & Hispanics & Asian & Native



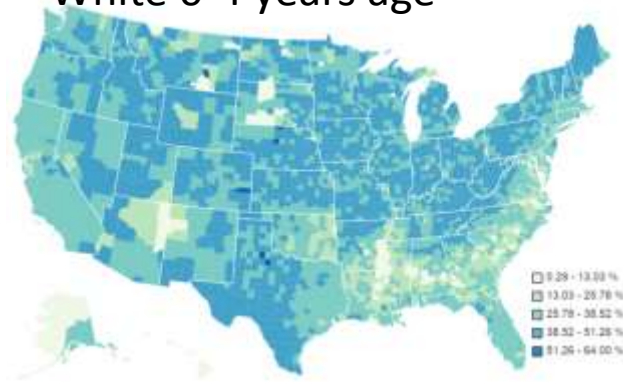
7. Implementation Choropleth cont

- User can also select certain age group and choose year

Hispanics 0-4 years age



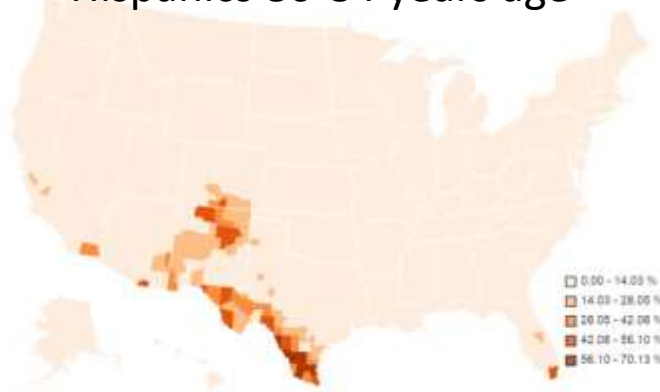
White 0-4 years age



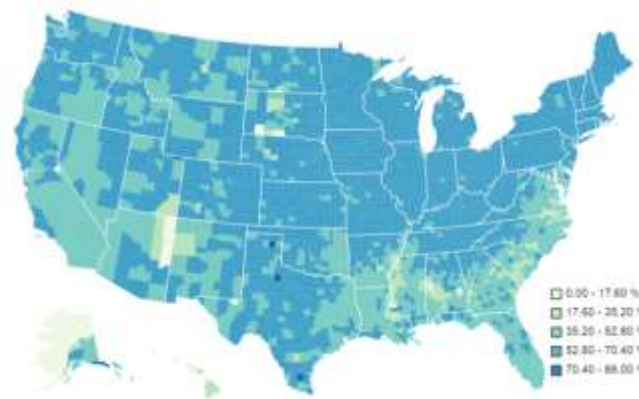
Asian 0-4 years age



Hispanics 80-84 years age



White 80-84 years age

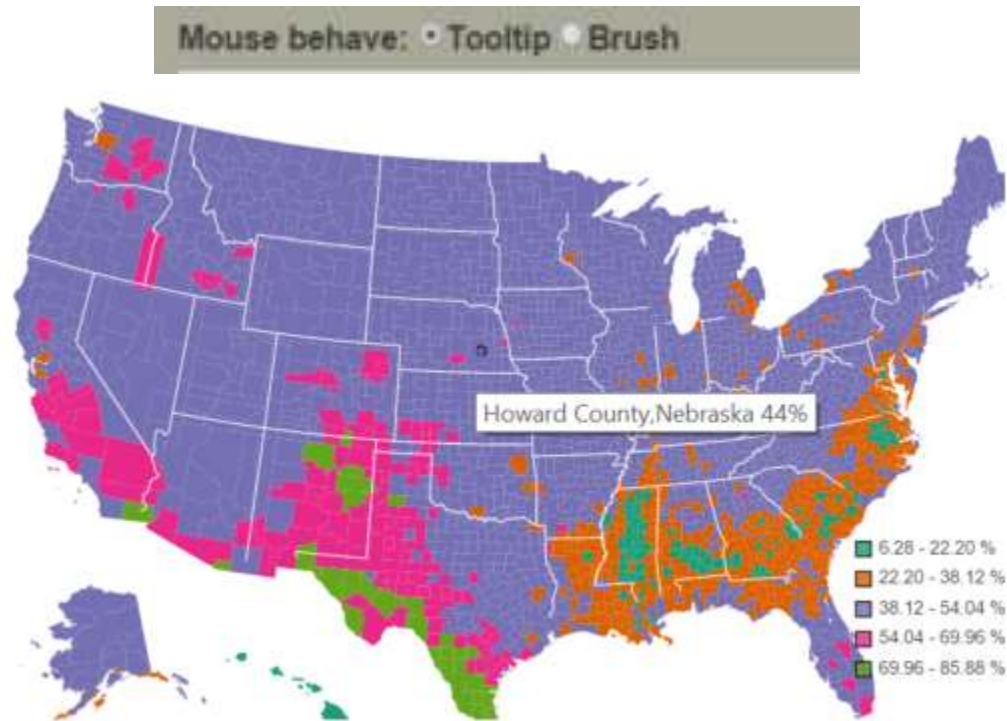


Asian 80-84 years age

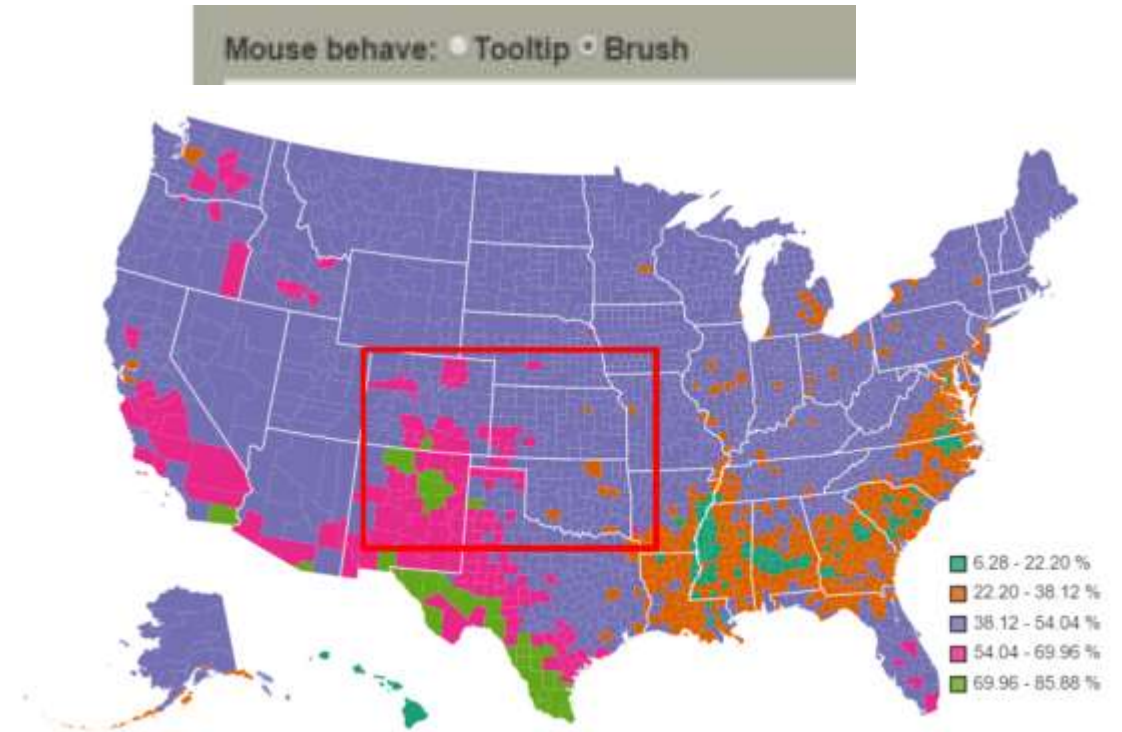


7. Implementation Mouse Behavior

User can see each county's detail by hovering mouse on it

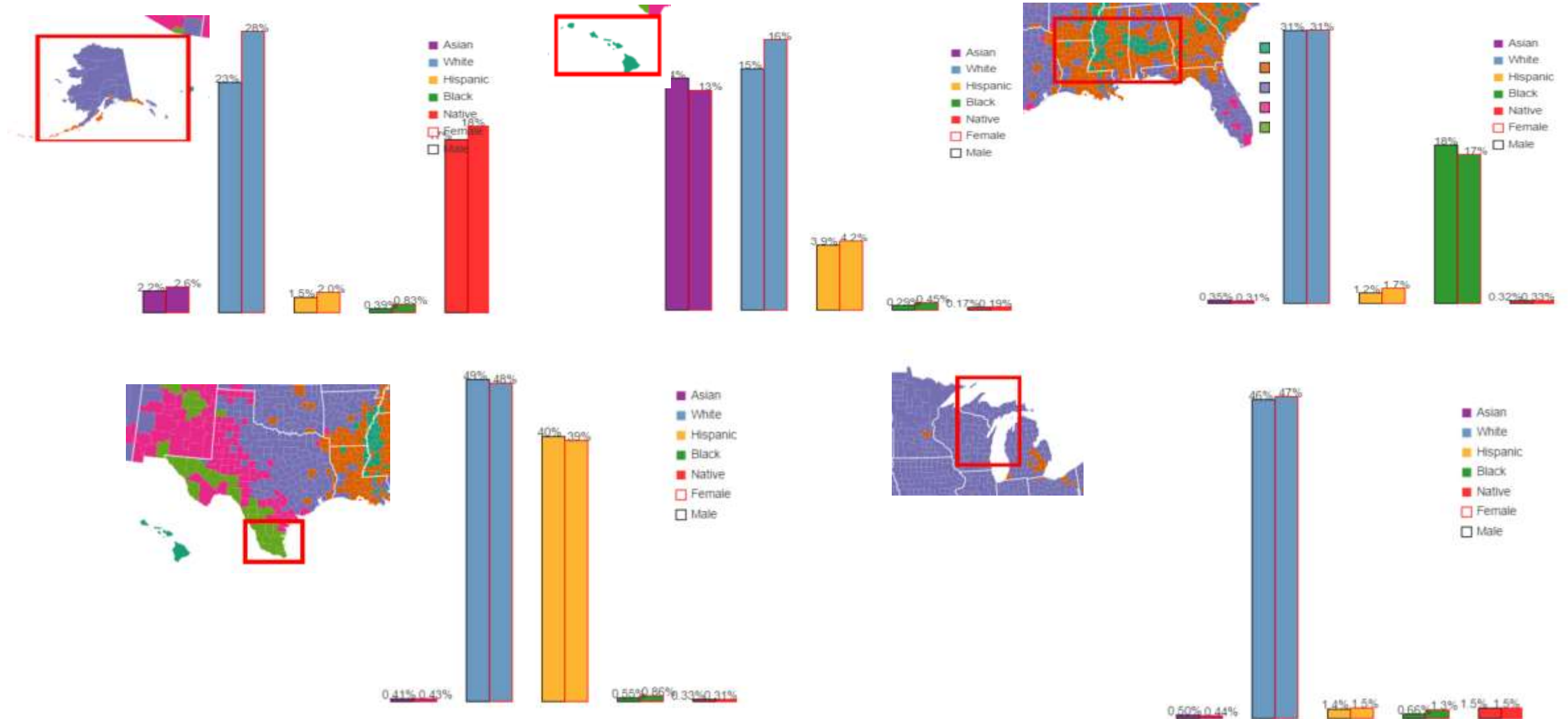


User can select to see multiple counties data by changing mouse behavior to brush in menu and create an brush



7. Implementation Bar Chart

- Bar Chart will be updated in real time corresponding to brush window



7. Implementation- Treemap



- Treemap is a very good method to display hierarchical data. The color and size dimensions are correlated with the tree structure, 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.

7. Implementation- Treemap cont

- If we click any state, the Treemap will go 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.

User can click head line to go back up one level and continue explore other states' data

Data.Illinois




```
▼ 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
```

8. Evaluation

There were several key patterns and trends as part of exploring census data over different ethnic groups, ages , gender and years.

- Minority group are speeded over biasedly heavily in certain areas, for example , Asian group is below 1% in most areas except in Bay Areas of California, Ney York, other few major cities and Hawaii.
- Overall female and male population are equal considering all ages throughout the years, however female population increases the ratio as age group goes higher, in beyond 85 years group, female is 10-20% more than male.
- 0-4 years new born babies overall speaking , there are more girls than boys for all ethnic groups, except Asians.
- Hispanics 0-4 years group population increases significantly through the years and at more areas, while White ethnic group is aging more as fewer areas with higher percentage of 0-4 years population, more areas with higher percentage of 80 years older.
- The design and implementation works very well we think, and it can be further improved to have more features such as comparing averages with selected areas, sum up number of counties at each quantize level, linking treemap with choropleth map , etc if time allows.

9. Libraries used


 **dingchaoz / uscensusvisualization** PRIVATE
forked from gajooge/uscensusvisualization

[Unwatch](#) 3 [★ Star](#) 0 [Fork](#)






branch: master ▾ **uscensusvisualization / libs / +** [≡](#) [🔍](#)

This branch is 40 commits ahead, 1 commit behind gajooge:master [Pull Request](#) [Compare](#)

created a by county choropleth.html with auto legend, added colorbrew... [...](#)

 **dingchaoz** authored 11 days ago latest commit 43803b54b6 [🔗](#)

..

 bootstrap	created coordinate view main.html,added treemap to it, however the 2n...	13 days ago
 d3	added a js libraries conflict issue document	13 days ago
 jquery	created coordinate view main.html,added treemap to it, however the 2n...	13 days ago
 colorbrewer.js	created a by county choropleth.html with auto legend, added colorbrew...	11 days ago
 legend.js	created a by county choropleth.html with auto legend, added colorbrew...	11 days ago

10 . Thank you

- We thank our Teaching Fellow Jordan Dominguez for feedback and guidance throughout the project.