Politivis: Process Book

Amanda Boss, Hillary Do, Takehiro Matsuzawa

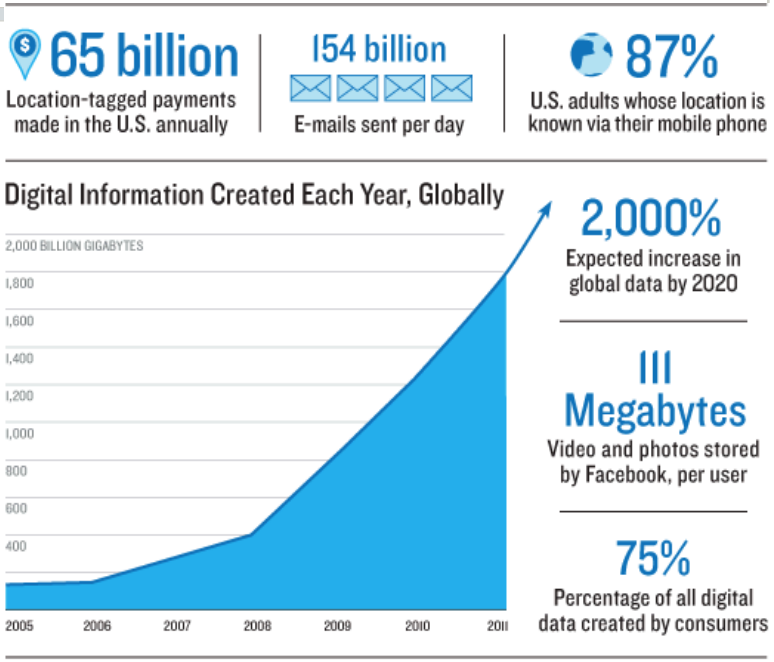
Overview and Motivation

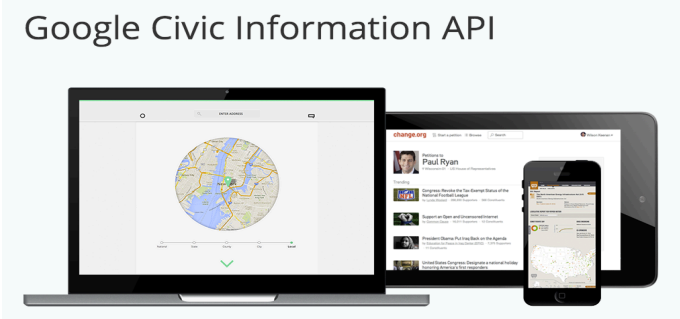
The motivation behind this project is linked to a desire to understand the implications of redistricting, the process of redrawing United States electoral boundaries. The process occurs every 10 years after the U.S census is completed. More often than not, especially in the southern regions of the U.S, redistricting has become more and more susceptible to unfair partisan gerrymandering. When redistricting is being decided, states require that their own redistricting committees comply with the federal Voting Rights Act. However, the impact of redistricting over the years is not entirely transparent to many Americans.

A solid link between redistricting and the movement of socioeconomic transformations within the newly created districts has not been fully established, despite readily available civic data. We aim to combine available data into a presentable form that will allow users to view redistricting in the U.S on a larger scale (and over time).

Motivation: New Data

The internet continues to rapidly expand and with that expansion, a plethora of new information becomes available each day. One such example is the Google Civic Information API which provides comprehensive data about representatives for each district. However, there also exists other lesser known sources such as the US Census which has many datafiles that are often difficult to comb through, but nevertheless provide invaluable information that acts as a strong complement to sources like the Google Civic Information API. Our project combines these two data sources to create a powerful and telling visualization about an often unknown topic that impacts Americans yearly.





Goals

Redistricting is something that is mentioned frequently in the media, but does not garner a lot of visibility by the typical user. We want to ensure that we answer questions that we had as well as answer the questions of those with less familiarity:

- **What is redistricting?**
- **Does redistricting target specific areas the most?**
- **What are abovementioned areas' demographics?**
- **Is there a link between redistricting and education?**
- **What about redistricting and the distribution of age?**
- **What about redistricting and race?**
- **What about redistricting and its relation to unemployment?**
- **How does migration fit into all of this? Are certain districts more prone to unpredictable migration due to changing of districts?**
- **What about political affiliation?**

Benefits of Goals

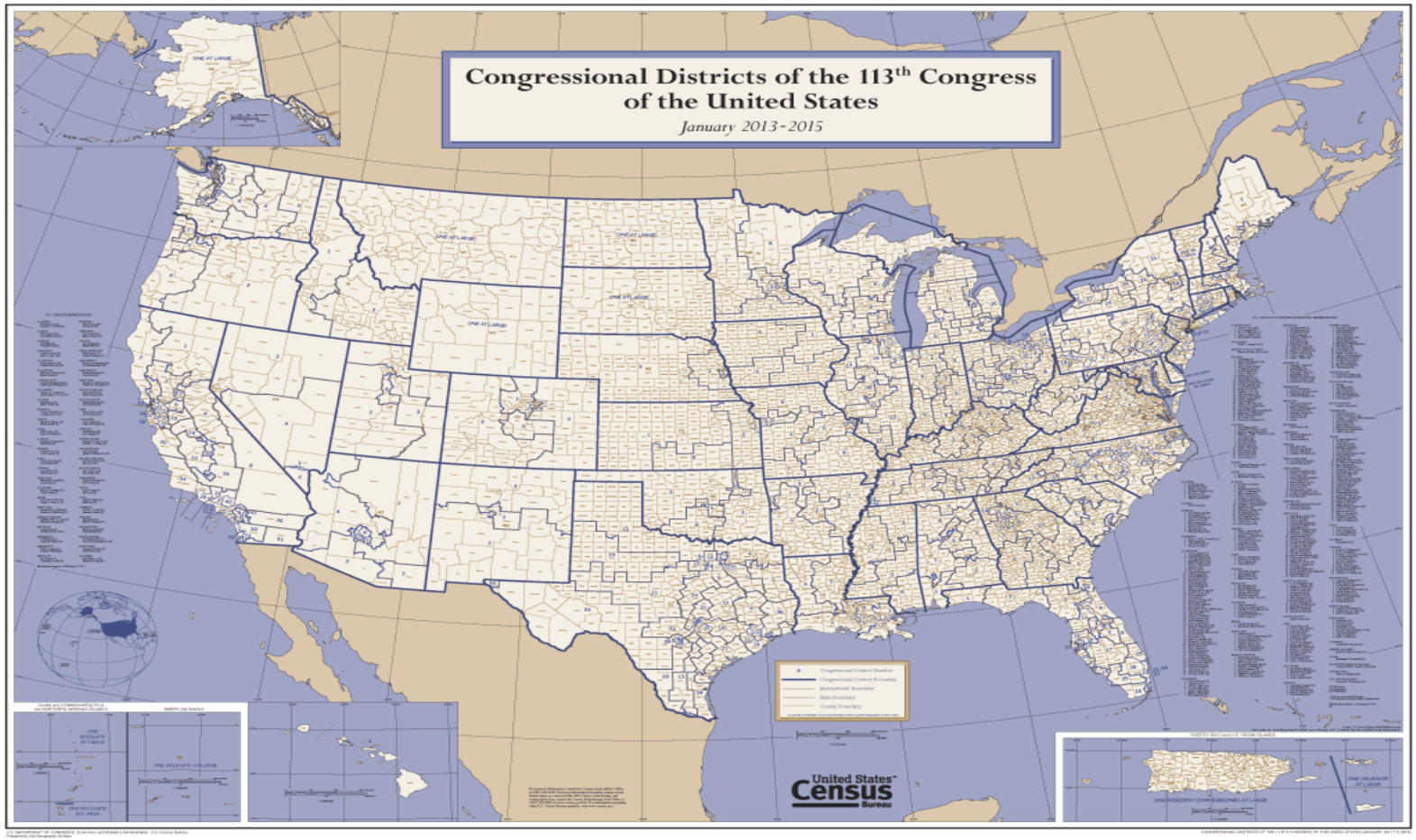
The benefits of answering these questions through a visualization include, but are not limited to:

- **Greater understanding of a topic that has a tremendous impact on voting, especially in states that do not particularly lean towards a certain party**
- **Allows any user who is interested in the visualization of data to understand a comprehensive and often unknown topic (learn what redistricting is in a comprehensive, but intuitive manner)**
- **Notice any of the abovementioned patterns and see if redistricting is unintentionally restricting one's ability to vote.**
- **Gain a broader comprehension of domestic migration in the U.S over a decade (between the 2000 and 2010 census)**

Related Work

We were inspired by two websites, one being the US census website (**www.census.gov**) which has myriad of redistricting and demographic data and the other being Mike Bostock's redistricting visualization (**https://f.cloud.github.com/assets/230541/522393/db95f2d2-c029-11e2-88e8-713345e1693d.png**)

Related Work – Census Data



Related Work – Census Data

- This expansive map inspired the motivation for the idea of visualizing redistricting in the United States.
- However, there was no information provided as to how to create such a map (thus the introduction of Mike Bostock's visualization, which will be mentioned further in the data section). Furthermore, the map was static and we wanted to observe different patterns (see Background and Motivation).
- In addition, we wanted to use d3 for interactivity, thus furthering the motivation for using Mike Bostock's interactive topojson format .

Questions

As mentioned previously, we had quite a few questions that we wanted to answer as part of our goals for the project (see Overview and Motivation).

As our visualization progressed, we realized that we wanted to deviate away from socioeconomic indicators such as education and unemployment. This was for two reasons:

1.) We felt that this data was too variable to present in a cohesive manner. Further research indicated that a change in unemployment would be attributed more to the recession. Education itself was hard to visualize.

2.) There is no data about educational attainment for the 108th congressional district. In addition, there is very limited unemployment data for each congressional district and it is limited to the latest 113th congressional district. Our goal was to show the major congressional redistricting between the 2000 and the 2010 census.

Besides those socioeconomic factors, the questions that we wanted to answer remained the same. New questions that we considered in the course of our analysis focused on the age aspect. We were interested in the age demographic for each district from the beginning, but wanted to see the divide by sex, which we displayed in our final version.

Data

- This leads us to the next question – how did we obtain the data?
- The U.S census has a plethora of shapefiles (files that eventually can be converted into a map) that contain the id of the congressional district and its coordinates for boundary drawing.
- However, we had to consider a few factors before integrating the shapefiles into our visualization.

Data Wrangling – Congressional Districts

1.) Obtain shapefiles from US Census TIGERFiles database for 108th and 111th congressional districts

2.) Convert the shapefile to GeoJSON format.

3.) Merge the fields in the newly created GeoJSON format to obtain the GEOIDs for each congressional district. (GEOIDs are unique numerical identifiers for each district that can be linked to demographic data from other sources)

4.) Convert the GeoJSON file to TopoJSON (a format that allows for a map to be rendered)

Data Wrangling – Congressional Districts

```
// convert Shapefile  (obtained from U.S Census)  for 2010 year to GeoJSON format

Ogr2ogr –f GeoJSON districts.json tl_rd13_us_cd111.shp

//convert GeoJSON to TopoJSON and add GEOID field

Topojson –s 7e-9 –id-property=+GEOID10 –o  us-congress-113.json –districts.json

//We then use congress-111.json as our source for creating the map for the 2010 congressional district
```

Data Wrangling – Demographic Data

- Unfortunately the US Census TIGERFiles do not contain any information about demographic data for each district. Those files only contain the geographic boundaries. However, as mentioned previously, each geographic boundary has a GEOID that is universal.
- GEOID is a four digit number that represents the following:
- STATE (2 digits) + Congressional District Number (2 digits)
- This is universal identifier for congressional districts and is used in many different APIs and data files.

Data Wrangling – Demographic Data: 108th Congressional District

- There is a very limited amount of **compiled** demographic data for the 108th congressional district.
- The American Community Survey conducted simultaneously with the 2000 Census contains the most information about the 108th congressional districts in one single file.
- **However, this survey only contains 205/500 congressional districts, despite being the most comprehensive survey.**
- Despite this obstacle, data wrangling was very simple, with minor formatting needed to match the GEOIDs from the large csv file.

Data Wrangling – Demographic Data: 111th Congressional District

- The 111th Congressional District has much more comprehensive data with respect to the 2010 census.
- Demographic data was compiled from a json file that contained information about age for each congressional district. The topojson file created originally was updated so that its properties contained race data from the 2010 census.

Data Wrangling – Representatives and Parties

- The maps for the congressional districts contain information about the representative for each district (when hovered), while the color of each district represents its party affiliation
- This data was gathered from the GovTrack API which provides programmatic access to information about the U.S congress.
- Like the other data sources, data was matched via GEOID
- Standard data from an API

Data- Summary

This data conversion method allowed us to successfully show maps of two of the most pivotal congressional districts in the 21st century (since redistricting is aligned with the census, only the 108th and 111th congressional districts were significant).

Summary of our data:

1.) Congressional district boundaries – Converted into TopoJSON form from US Census TIGERFiles (shapefiles)

2.) Demographic Data – Compiled from US Census (111th CD) and American Community Survey (108th CD)

3.) Party affiliation and Representatives for each district: GovTrack API

Design Evolution

- Generally, the different visualizations that we considered did not deviate much from the original sketches that we had submitted, with the exception of the removal of education and employment (see Questions)
- However, there were certain aspects of the visualization that we considered and then changed.

Design Evolution – Change Over Time

- Since our visualization tracks a change over time (in our case the 108th CD and the 111th CD which correspond to the 2000 and 2010 Census), we initially decided upon a slider.
- However, as we implemented our visualization we realized that this was not the best idea. Namely, redrawing the congressional boundaries wouldn't allow for a seamless transition that a slider would require. Secondly, we are only showing two main redistricting districts of the 21st century, a slider would make sense for more than two congressional districts. So instead we replaced the slider with easy to view buttons.

Design Evolution - Congressional Districts

- At first, we considered having a blank map with just the congressional boundaries (as viewed with Mike Bostock's visualization of the 113th congressional district -- See Related Work).
- However, we recognized that this would not be visually appealing to the user and that we could use saturation to represent party affiliation. This would present that information in a manner that was not distracting or overwhelming to a viewer. Plus, like many visualizations regarding political affiliations, color to represent the parties is a familiar and common concept.
- As mentioned in lecture, color is a good way to label qualitative data such as party affiliation. It helps break the monotone color of the map.

Design Evolution- Zoom

- One of the primary features that we depicted in our sketches was the zoom feature. This feature involved zooming in on a congressional district and viewing the district up close as well as visualizing the graphs.
- We implemented the zoom in and zoom out feature, but included the graphs right below the zoomed in congressional district (outside the svg element). This was because we wanted to allow enough space for the viewer as well as accommodate congressional districts that are landlocked. Our original sketch involved having the graphs on the svg element itself.

Design Evolution - Hover

- We originally did not include this feature in our sketches, but as our visualization progressed, we realized it would be helpful to obtain representative information (name and picture) when hovering over each congressional district. This would complement the party affiliation for each congressional district as well as establish a link between race of representatives and concentration of redistricting.
- We wanted to use a hover feature for our design, since we already implemented a click and zoom feature. Furthermore, representative data is more of a quick view rather than an in depth view (i.e looking at the demographic data an individual congressional district), and thus a hover feature seemed more suitable.

Design Evolution – Pie Chart: Race

- As mentioned in both the Overview and Motivation as well as the Goals section, one of the primary factors we were concerned with was race. We felt that a pie chart would best visualize this data for the following reasons:

  1.) We wanted to show proportional data (i.e slice of pie for each race) - Originally our sketch displayed race as a bar chart, but we wanted to hone in on proportion

  2.) Pie charts are acceptable for around 6 categories or less (We had 4 categories) 3.) A more visually appealing way to represent nominal data that would normally be viewed as a table without an idea of proportion

Design Evolution – Bar Chart: Age

- Both our sketch and implementation featured displaying age as a bar chart. For implementation, we implemented sex by age to gain a broader understanding of the gender dynamic. We utilized a bar chart for the following reasons:

1.) Numerous categories - We had over 20 age ranges presented in the data, making a pie chart impossible.

2.) Nominal Data – A scatter plot nor a line graph would best show the amount of people per age group

3.) Proportional – Height is proportional to the amount of people in each age group.

We displayed the two genders side by side and changed color for females to differentiate the two.

Implementation

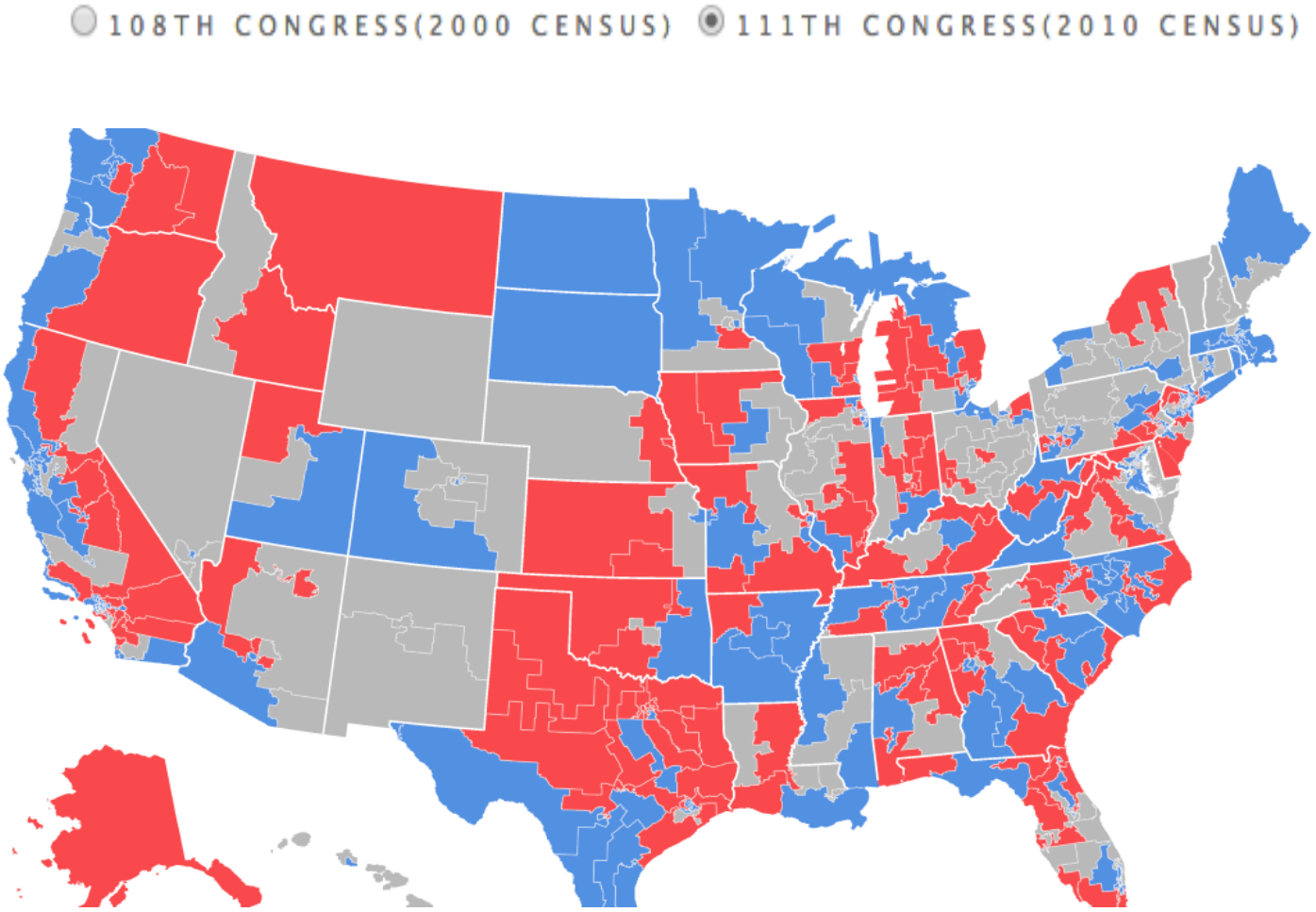
(Fig. 1) The first view that the user sees is the overall map reflecting the political affiliations for the 111th district. If the user wants to see the oldest available congressional data for the 21st century, they can click on the 108th district.

The colors represent the following political affiliations:

- Red : Republican
- Blue: Democrat
- Gray: No Party Affiliation Available

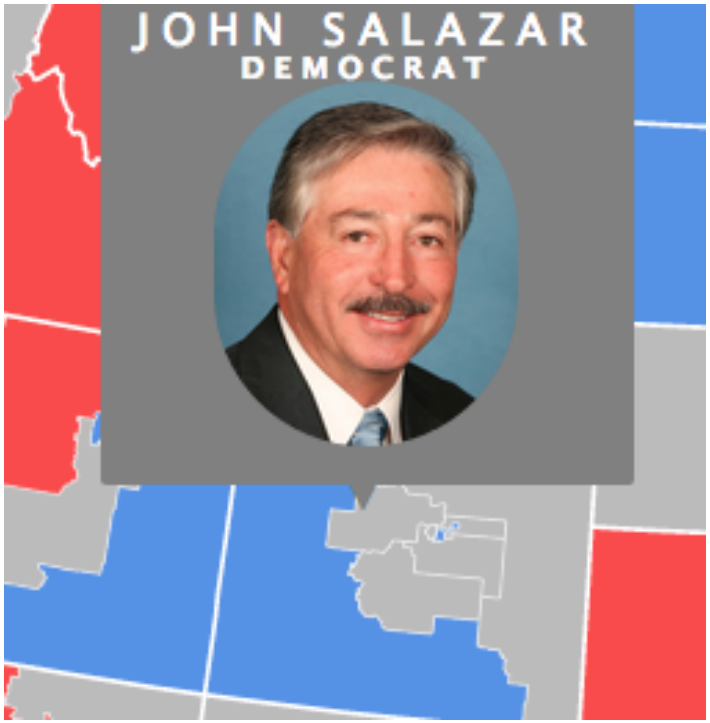
As explained in the Design Evolution, we used standard party affiliation colors. The map is a bit off center to accommodate for the hover feature which is discussed on the following page.

Figure 1



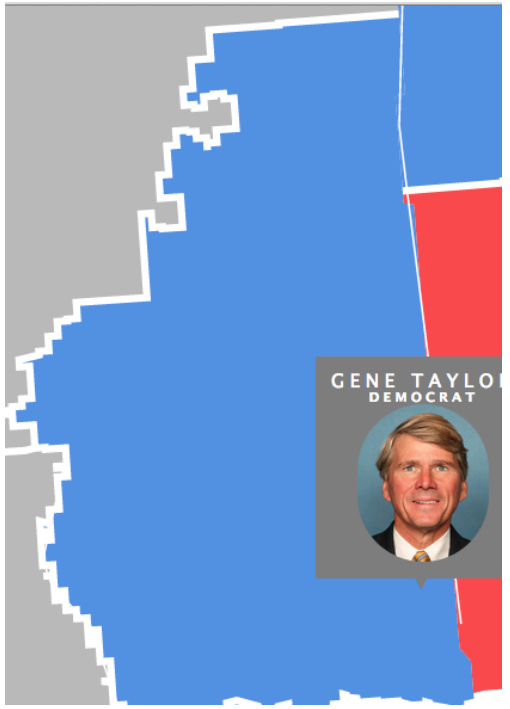
Implementation

The user can hover over the individual districts from either the 111th or 108th to see basic information about their representative.



Implementation

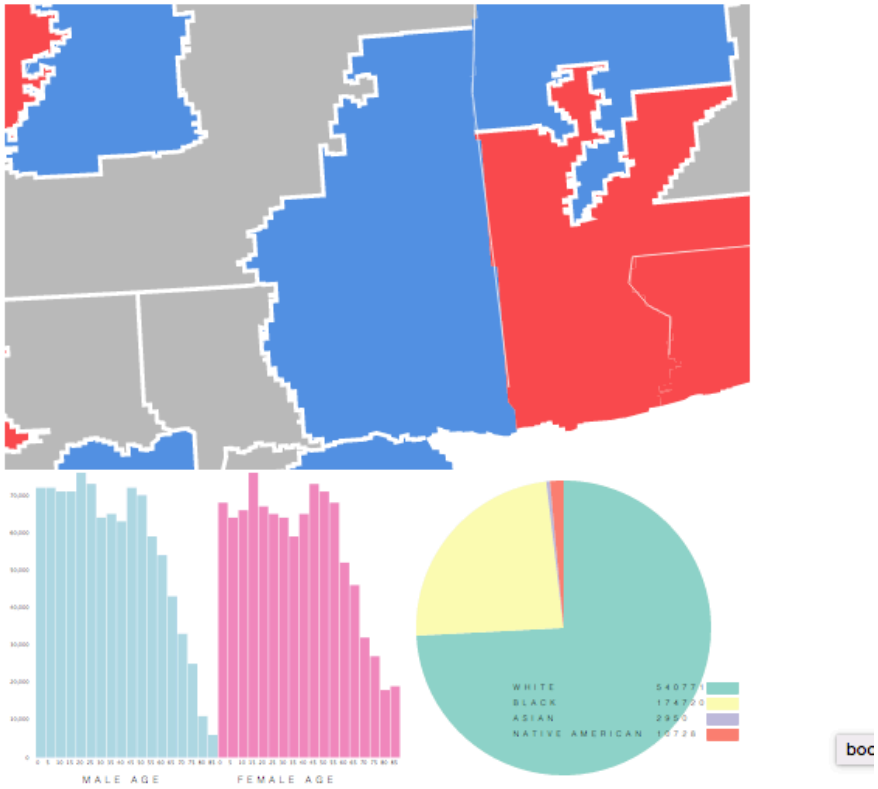
In addition to being able to view the congressional districts over those two times, the user can also click to zoom in on a congressional district.



Implementation

When the user is zooming in on the congressional district, they can scroll down to see demographic data.

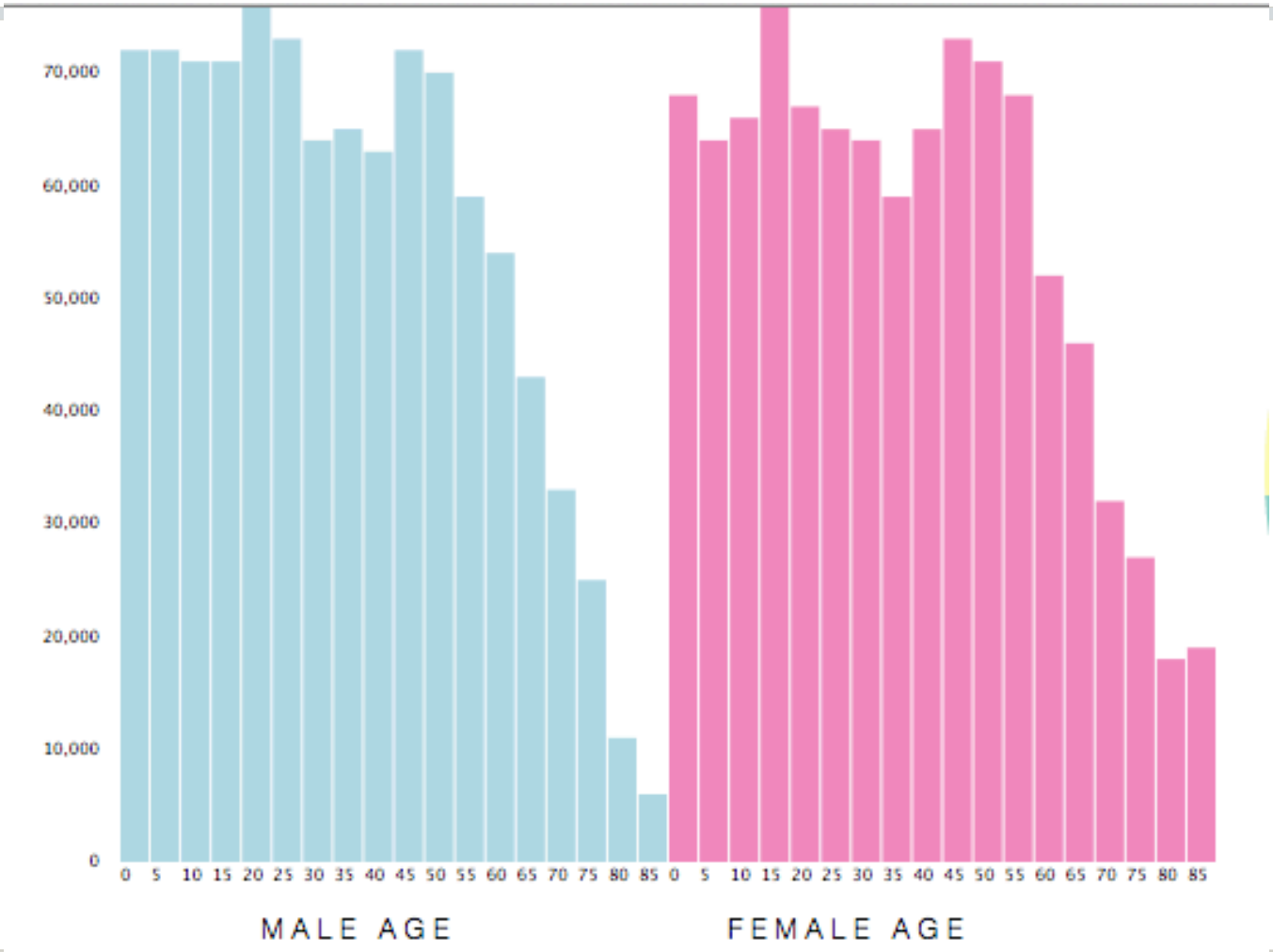
Overall view:



Implementation – Demographic Data

One of the views that the user can see is the composition of age by sex from each congressional district (compiled from individual csv and json files from the US Census). The pink colors represent female, while the blue colors represent male. On the x-axis are the 20+ age ranges represented by the US Census. They are 0-5, 5-10, etc. (See Figure 2)

Implementation – Figure 2

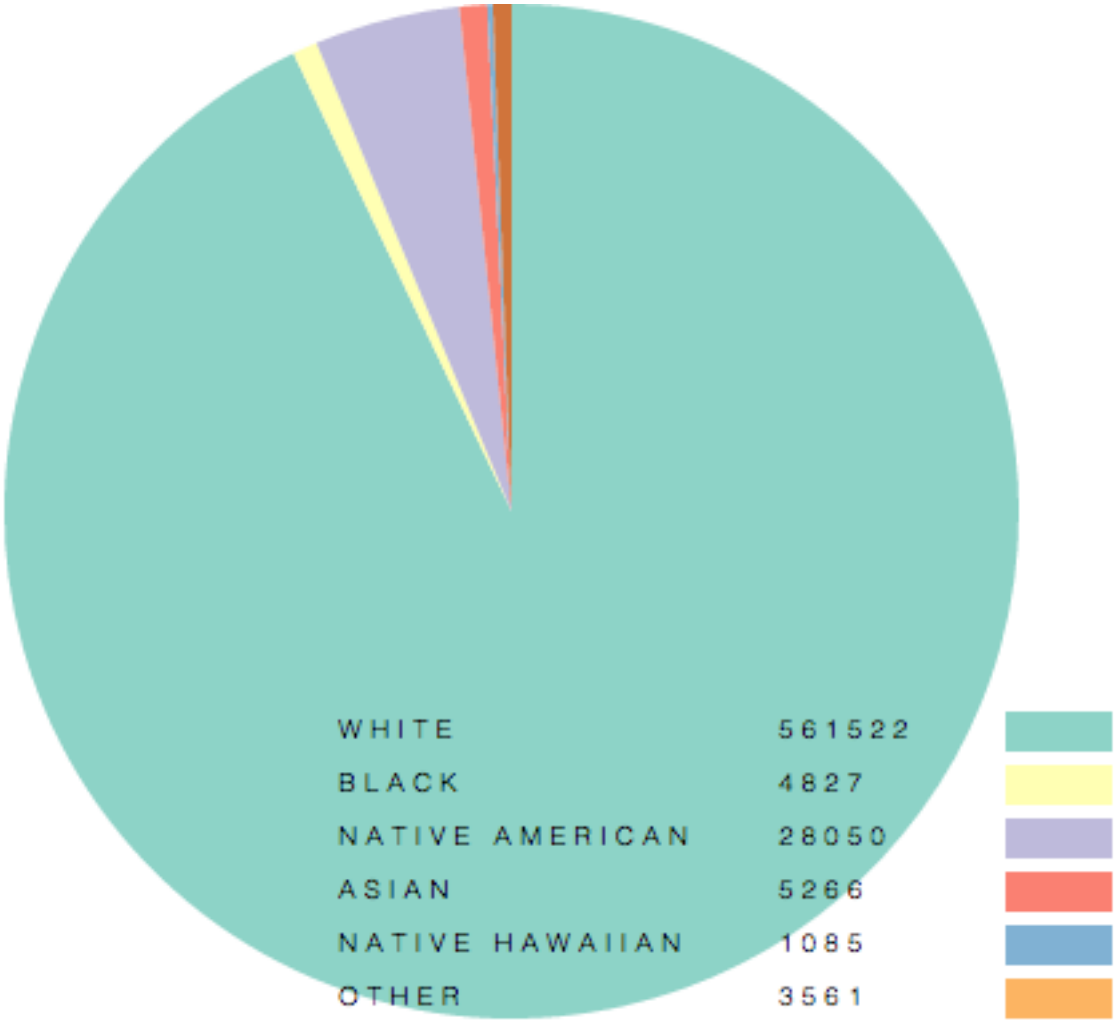


Implementation – Demographic Data

One of the other views that the user can view for each congressional district is race, which is represented by a pie chart (see Design Evolution). Each slice/color represents a race as detailed in the legend (See Figure 3).

Next to the race is the actual amount of people who identify as that race.

Implementation – Figure 3



Implementation – Zooming In and Out

Users can zoom out of the map by clicking again in that same region or clicking outside of it. From there, they can zoom in on more congressional districts to find out more information.

Note: Due to the nature of the visualization, if you zoom in on a district and click on an adjacent one, that district may need to be zoomed out and then in again to properly view the demographic data.

Evaluation - Observed Trends

With the views that we presented, we were able to observe the following general trends regarding redistricting from the 2000 Census to the 2010 Census:

- States that are primarily affiliated with the Republican party have gone through a significant amount of redistricting.
- Redistricting is much more common in the South than the North.
- States like Texas that go under more redistricting have a greater age by sex disparity. As redistricting occurred, there is a noticeable shift in age, more people in their 20s than before. That could be due to migration shifts that occurred as a result of redistricting
- States like West Virginia, which are more susceptible to gerrymandering and have higher levels of poverty, have greater racial diversity.
- If racial minorities increased in a region, the state tended to shift left.

Evaluation

Redistricting is a tricky thing to visualize and it cannot be compared to any "aggregate" data. Congressional district numbers are always changing, so it was often hard to see that change over time, especially since it was with respect to only the 2000 and 2010 Census.

However, we were able to see race and age data for each congressional district in 2010 (close to 500 congressional districts total) as well as representative and party affiliation. The project was able to replicate that data for 2000, which has never been done before since there is such scarce data. However, as mentioned previously, many of the congressional districts do not have that much data. The American Factfinder site has a similar visualization to our 111th CD, but nothing for 108th CD, this visualizations is creating something entirely new.

Evaluation

Overall, we felt that the following aspects of our implementation worked well :

1. Allowing for three entirely different data sources to work together to deliver information about hundreds of congressional districts.
2. The overall view of the map allowed political affiliation, representative info, and congressional boundaries to be presented in a digestible manner.
3. Answered our questions, about how race was affected (see previous page),

Evaluation

As mentioned previously, there were some aspects of our project that we felt needed improvement:

1. We were working with a very limited dataset, so we don't have that much data for congressional districts from the 2000 Census ~200/500. (Demographic data is only available per census)

2. There is a lack of concrete socioeconomic data available. We felt that redistricting is an important topic, as mentioned in Goals and Motivations, but were surprised towards the end that our data really only focused on race and age for the 2010 census. There wasn't a lot of data that mapped it to the congressional district.

3. If we had more time, another nice option would be to see the maps side by side, but once again it is hard to have comparison baseline for redistricting.

States that have information for CD108

- Maine
- Montana
- South Dakota
- Wyoming
- Arizona
- California