Basic Info

Title: **Of the People** - a visualization of the demographics of congress

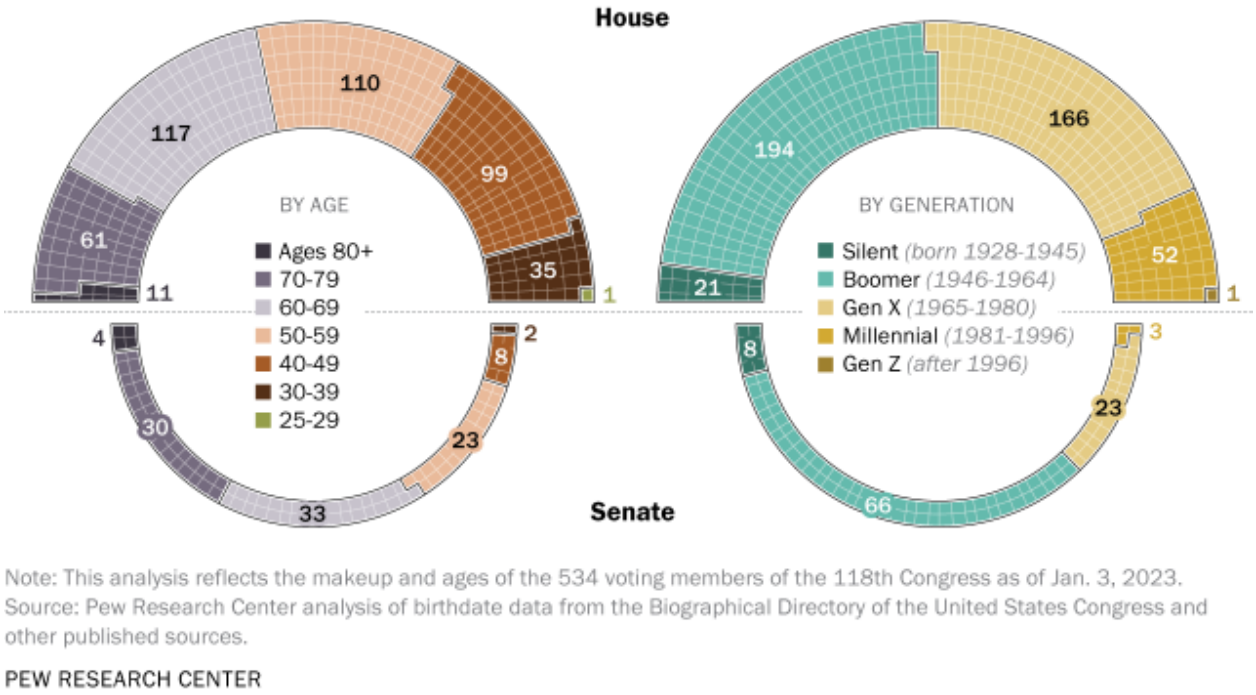
Seth Doubek - u1285625@utah.edu, u1285625

Peter Bruns - u1350429@utah.edu, u1350429

GitHub Repository - https://github.com/dataviscourse2023/final-project-congressdemographics

Background and Motivation

One of the main points of inspiration for this project was this article by NPR. Summarized, it points out how the 118th congress is one of the oldest in recent history and how that will affect what bills and laws are passed before them. We even get some decent visualizations out of this article shown below.



Many articles have been published recently about the age of members of Congress, and we were wondering if we could expand this further to include more demographic details. For example, the Pew Research Center who made the above visualization actually put out another related article that looked further into a variety of demographic data that includes visualizations as well. It is important to know who is representing us, to know what one of the driving forces looks like so we can get a better understanding of the inner workings of our democracy. Looking at how Congressional demographics have changed over time will also provide information about

how broader societal movements have impacted the representation that citizens receive from their Congressional members.

Project Objectives

The primary question that we are trying to answer with these visualizations is "Who is representing us in Congress?" We hope to provide the user with visualizations that clearly relay demographic information about the members of Congress. This information will allow users to gain a better understanding of how well the makeup of Congress reflects the country as a whole, and how Congress may respond to certain issues. Viewing historical data will also show whether Congressional representation has changed to better represent as many people as possible. During the project we would like to learn and accomplish the following:

1. Learn fundamentals of programming in D3
2. Create stimulating, engaging, and educational visualizations
3. Learn how to best incorporate interactive elements to visualizations
4. Learn to create a large variety of visualizations, this data seems like a good fit for this point
5. Come out with a project that can display technical chops that we've learned during this semester

Data

Data exists for this project, but it seems like there might be a few barriers to access the data. There appear to be two different routes we can take. One is to use resources like this one which provides us with links that we can scrap demographic data from. This dataset is maintained through a partnership by FiveThirtyEight and ProPublica among other organizations. The other option is to find a data source that has everything we need. The Pew Research Center makes some of their data available to download (found here), but you need an account to access this data and I'm unsure if we could get access. Here are some of the possible data sources that we have identified and their feasibility/usability.

https://github.com/unitedstates/congress-legislators

Pros:

- Information on almost all of the members of congress
- Long date range going back to the first Congress
- More customizable
- This dataset is maintained by members of several notable political sites including FiveThirtyEight and ProPublica

Cons:

- Data scraping might be intense
- Some early members are missing birth date information

https://www.pewresearch.org/politics/datasets/

Pros:

- Might have super easy and ready to use data

Cons:

- Might just not have data we need
- Need to gain access

https://www.brookings.edu/articles/vital-statistics-on-congress/

Pros:

- Has some easily accessible data that we can use

Cons:

- Only has 117th congress (2021-2022) not those currently serving
- Some of the data is extremely sparse, example below

Table 1-21 Gender Identity and Sexual Orientation of Representatives, 117th Congress, 2021

	117th (2021)		
	D	**R**	**Total**
Sexual Orientation			
Bisexual			
Gay	7		7
Lesbian	2		2
Gender Identity *(see note below)*			

https://crsreports.congress.gov/product/pdf/R/R46705

Pros:

- Has a lot of what we're looking for in terms of attributes

Cons:

- Is a PDF, we would have to manually enter data in
- Mostly summary statistics with no information on individuals

CQ Roll Call

Pros:

- Pew Research Center pulled a lot of data from this source

Cons:

- We might not be able to access the data and may need to consult the library
- Because we cannot see it, we do not know for certain that the data contains what we need

Data Processing

Data processing will very much depend on how we're able to gather our data. Hopefully we can find a source where cleanup is rather easy. Such a dataset probably exists somewhere on the internet, and we will continue to search for one throughout the project. We have data that can be used to calculate age for all members of Congress throughout its history, but we would like to find datasets that would allow us to connect individuals to other demographic information instead of just using the summary statistics that we have found. Combining datasets would likely be done in Python using Pandas dataframes.

Another possible source of data cleanup might come from an abundance of missing values. For example, the first datasource listed above is missing some birthday information for early members of Congress. Because this is all historical data, we will probably be able to find the missing information or an appropriate substitute somewhere on the internet. Filling in those missing values will likely need to be done by hand.

Most derived attributes will probably come from sub-sectioning our data. For instance, one derived attribute could be the comparison of age distribution of Democrats vs Republicans. Such attributes seem like a good fit for our data, and this might be fairly common in our project.

Visualization Design

The anticipated design for the full page and ideas for individual sections are shown in the sketches section below. The full page will consist of one main visualization section and smaller sections to display more details. The user will be able to select what the main section shows from a variety of summary statistics. For example, this may show a trendline of average age or racial makeup overtime, histograms of age for a certain Congressional session, or bar charts for other demographic information for a single session. We chose to present this type of summary data in the main section to give a clear overview of Congressional demographic information. If a user wants to view changes over time, trend lines seem appropriate to clearly visualize changes. Histograms will allow for displaying quantitative information for a single session in a way that clearly shows distributions. Bar charts will be used for categorical or ordinal data to display counts.

The user can then choose what details to show below this main chart. For example, they could display a grid showing information about all members of a certain session individually, choose to compare trends in the House and Senate, or compare trends between states. Each visualization will allow the user to mouse over or click to get more information. For example, clicking on a trendline in the main plot would autofill the more detailed sections with information from that time. These sections will use similar encodings to the main section for much of the data but will also include grids or fields to show individual members. This seems to be the standard for representing all members of Congress on their own, and will allow users to select individuals for further information. We will use colors to separate the members by whatever characteristic (e.g., party) the user selects.

We want all of the visualizations present to have some level of interactivity. A lot of the similar visualizations lack this quality. This means we can show more than just summary statistics and allow the user the ability to control and explore the dataset at their own pace.

Must-Have Features

1. Interactivity, what would separate our project from the existing graphs about this data is if we can make it dynamic. All the other graphs, while well produced, are static in nature. I think using interactivity we could provide more insights
2. Trend graphs for all demographic data over time and the ability to plot different demographic features against each other
3. The ability to view data for the House and Senate separately or together along with the ability to select which states to include in the visualization

Optional Features

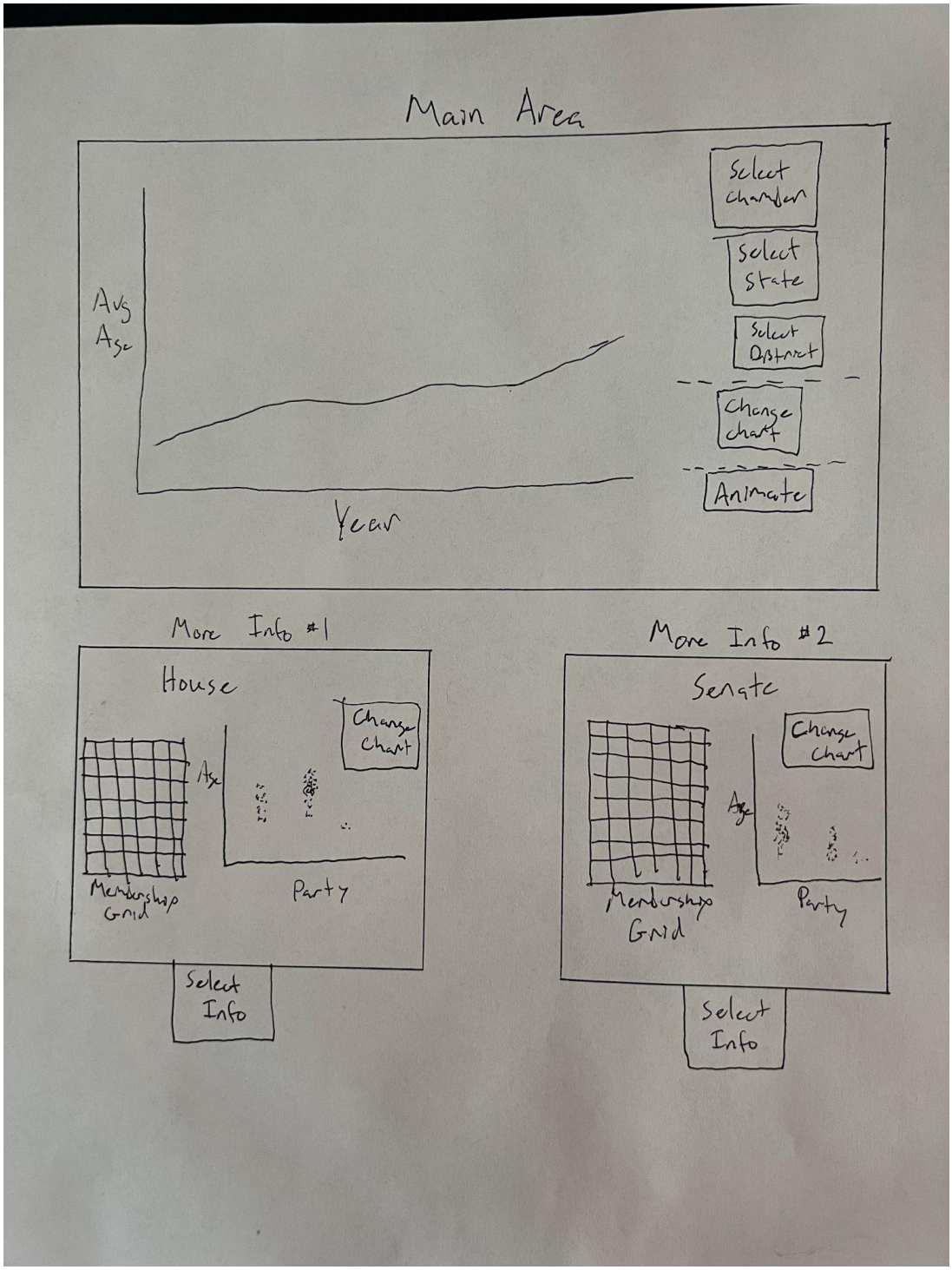
1. Animations that let the user see how demographic data has changed over time
2. Interactive maps to allow users to pick states to focus on for more specific visualizations or to plot demographic differences geographically

Project Schedule

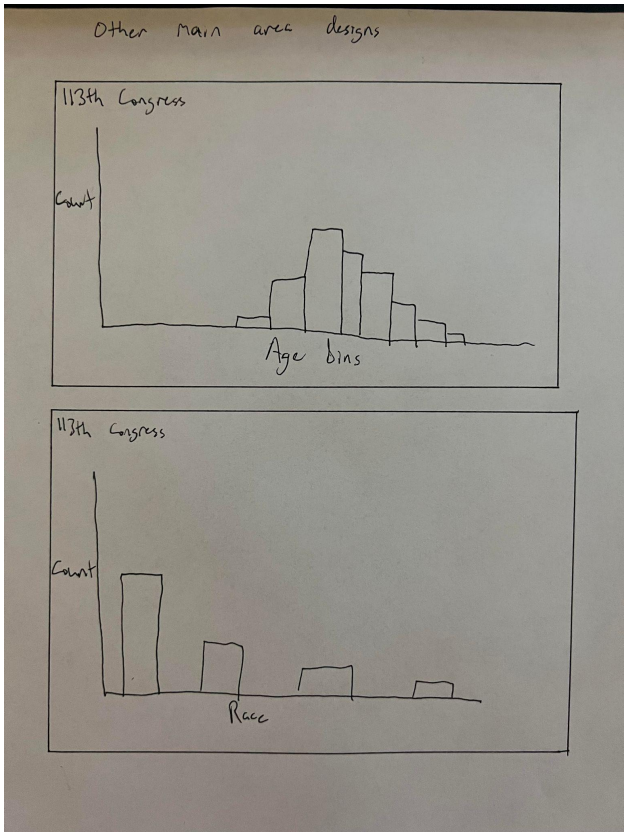
- ☐ **Sep 15, 2023 Project Proposal**
- ☐ **Sep 22, 2023 Project Review with Staff**
- ☐ **Sep 29, 2023 Have data sources identified and start data cleaning**
- ☐ **Oct 6, 2023 Finished collecting and cleaning data**
- ☐ **Oct 13, 2023 Fall Break**
- ☐ **Oct 20, 2023 Have basic examples done for each visualization**
- ☐ **Oct 27, 2023 Interactivity added to main visualization plots**
- ☐ **Nov 3, 2023 Milestone**
- ☐ **Nov 10, 2023 Added interactivity to examples of the more detailed visualizations**
- ☐ **Nov 14, 2023 Added interactivity to all visualizations**
- ☐ **Nov 21, 2023 Added at least one stretch goal feature**
- ☐ **Nov 28, 2023 Finished adding stretch goal features**
- ☐ **Dec 1, 2023 Final Project Submission**

Visualization Sketches

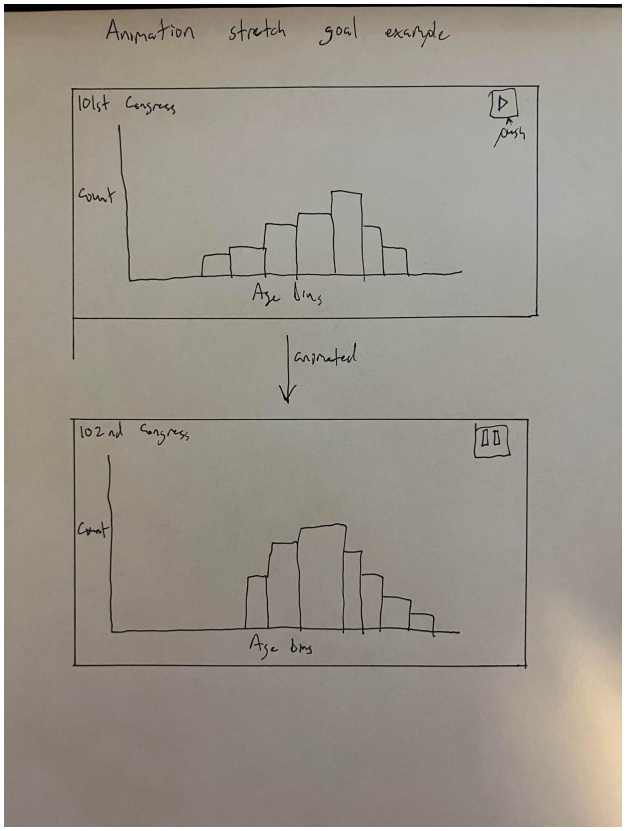
We plan to have the overall page laid out like this



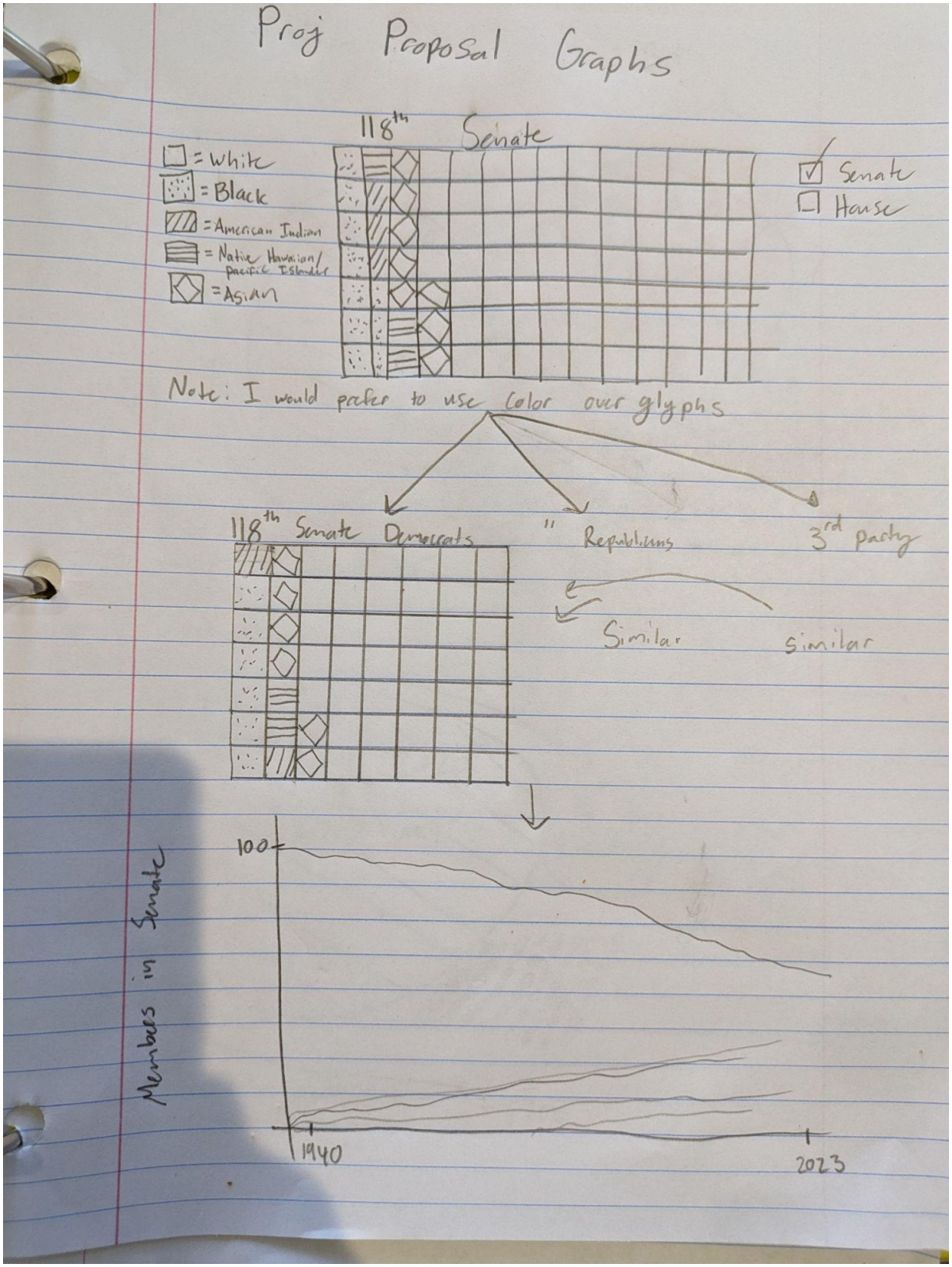
The main plot will be interactive and allow users to select data to display in the smaller windows below.



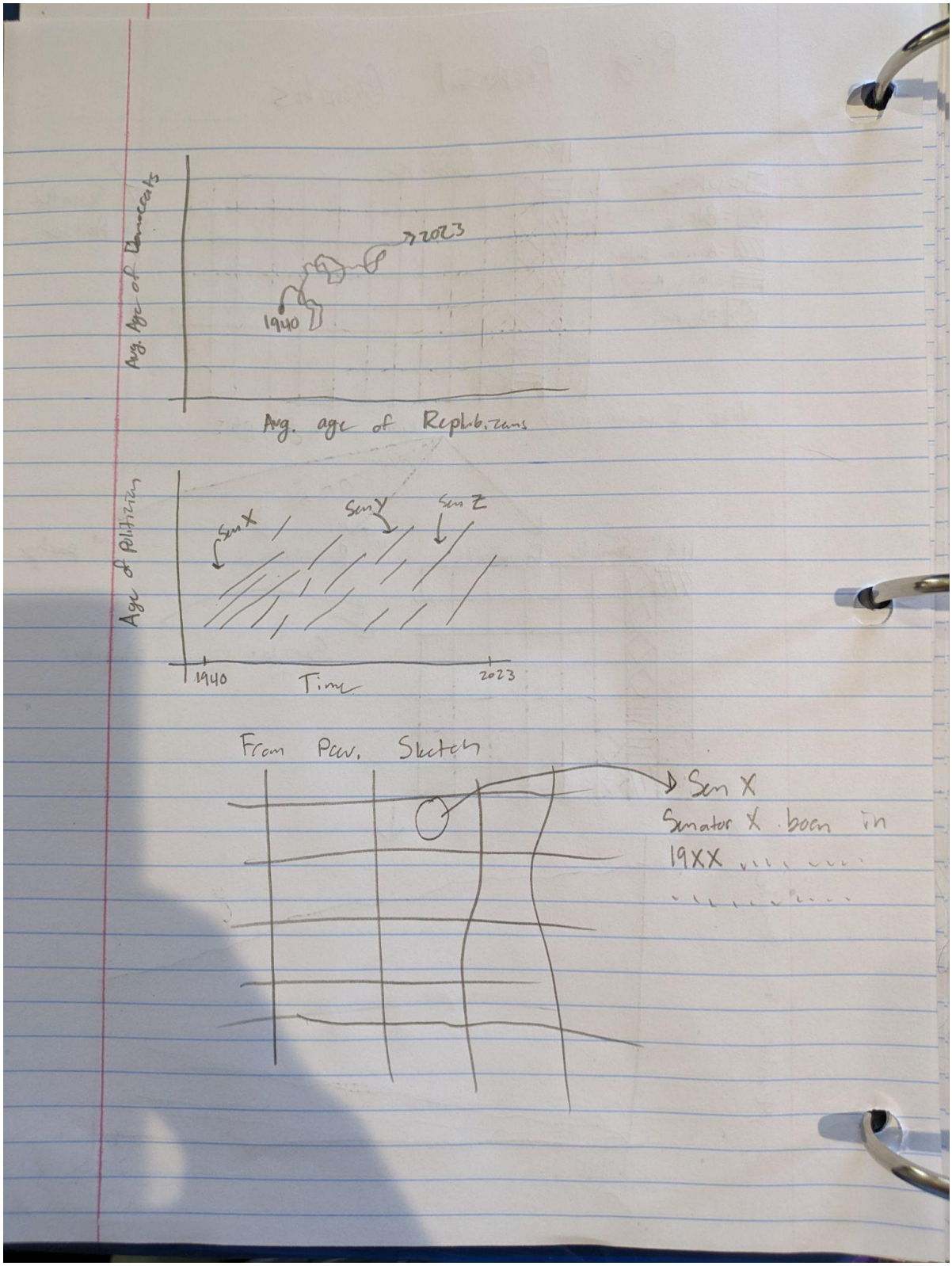
Other ideas for main visualization plots



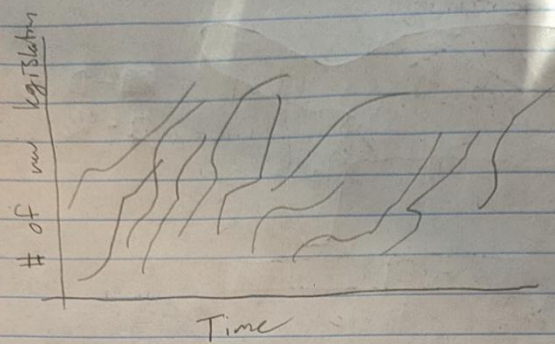
One potential animation transition

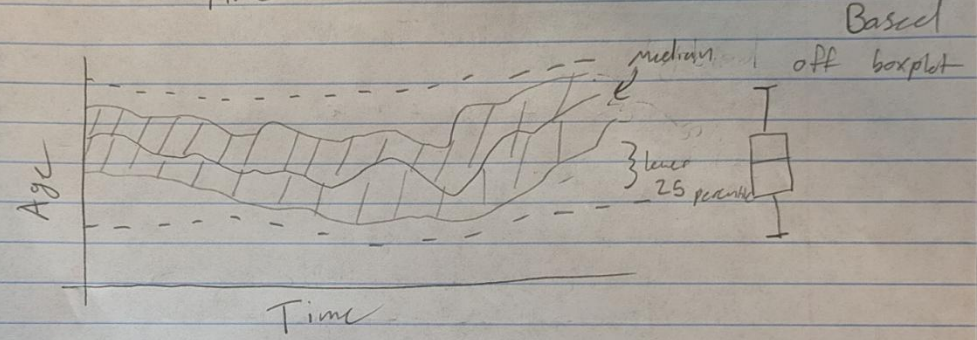


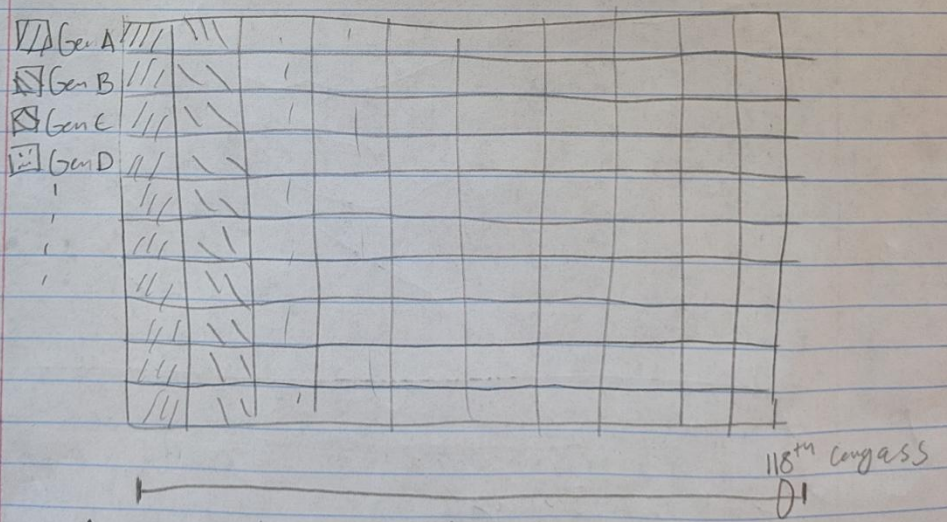
Such a strategy as above could be easily iterated for gender and even age.



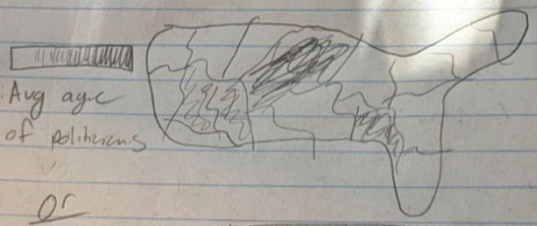
Above are some more plots we can use in our visualization's detailed sections to tell more of a story







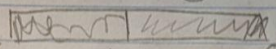
As you slide, see older generations being replaced by newer ones



▨ - Democrat
□ - Republican

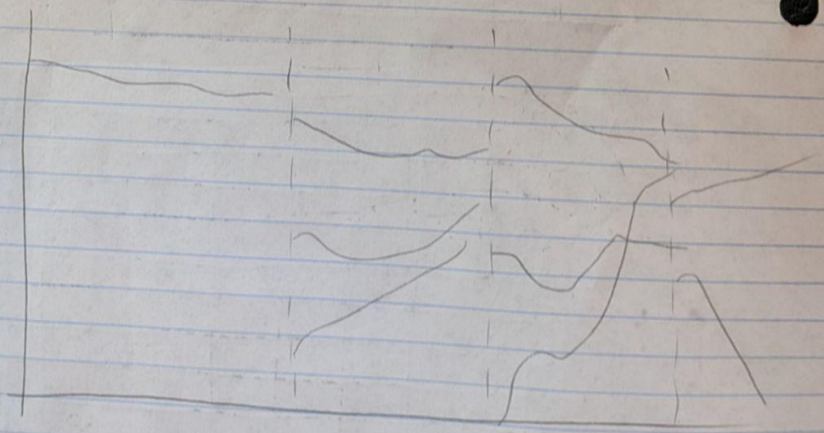
Or

Similar but



Click on state, gives more info

- State view
- Regional view



Scrolling Journey
Start mostly white males ——→ end at current
Demographics change dynamically, just one long line chart