

How Demographics Affect Voter Turnout in Massachusetts

CSCI 1710 Visualization
Final Project

Anitej Thamma
Max Bahar
Stefan Chu
Ted McCulloch

Proposal

Title

How do local demographics affect voter turnout in Massachusetts?

Abstract

In any election, it is paramount that all citizens are able to cast their vote and participate in shaping their government. Therefore, resources should be dedicated to maximize voter turnout and increase democratic engagement. To this end, it is important to understand the dynamics behind voter turnout; specifically, how voter turnout is influenced by demographic factors such as race, ethnicity, gender, and income. Unraveling how the elements of one's identity can influence the decision to vote or not vote can help increase the effectiveness of voter turnout programs.

Team Agreement

We, Max Bahar, Stefan Chu, Ted McCulloch, and Anitej Thamma, are putting forward the following agreement about the CS 171 Final Project.

Meetings and Communication

Each member of the team agrees to make time to meet either virtually or in person when needed.

- Meeting Frequency: Meeting frequency will depend on the point in the semester as well as the upcoming deadlines. We anticipate that it should average at least once a week. Group members are expected to conduct most of the project work outside of these meetings.
- Meeting Format: For major decisions (such as design), we agree to meet in person. Otherwise, we can meet virtually via Google Hangouts.
- Communication Channels: Primary communication will be conducted through iMessage. Code collaboration will be conducted through GitHub.

Decision Making

Group agrees to conduct decision making collectively and democratically.

Accountability

Each member of the group agrees to uphold their agreed upon portion of the project as well as active contribution to the success of our project.

Responsibilities

Since Max, Stefan, and Ted are co-conducting the project with AC 209a, their visualizations are expected to be more involved with advanced applications of the data. This will be considered when agreeing about the allocation of different visual designs.

Shared Responsibilities

1. Initial EDA/familiarity with dataset
2. Contribute to the design of visualizations
3. Complete the implementation of their allocated visualizations
4. Meet all submission deadlines for project

Typed Signatures Below

Theodore McCulloch

Maximilian Bahar

Stefan Chu

Anitej Thamma

Project Plan

Basic Info

Title: How do local demographics affect voter turnout in Massachusetts?

Group Members:

- Anitej Thamma (anitej_thamma@g.harvard.edu)
- Max Bahar (maxbahar@g.harvard.edu)
- Stefan Chu (stefanchu@g.harvard.edu)
- Ted McCulloch (tmcculloch@g.harvard.edu)

Background and Motivation

Research conducted by [Pew Research Center](#) has shown that the demographics of voters and nonvoters are vastly different at the national level. On average, voters were more likely than nonvoters to be older, identify as White, and have a college degree. We believe that everyone should have the ability to cast their vote and participate in shaping their government. Furthermore, low voter turnout may exclude citizens with certain political views, skewing the results of the elections. Therefore, it is important to understand the dynamics between voter turnout and demographics, particularly at the local level, so efforts to increase voter turnout can be tailored to each area.

Related Work

There are lots of great data visualization examples surrounding US elections:

- Many focus on the winning party in a specific area, such as this map from [the Washington Post](#).
- [The Census Bureau](#) has an entire library of plots dedicated to understanding voter turnout over time and by demographics at the national level.

We hope that our project can contribute to existing work by providing an analysis and visualization of voter turnout and demographics at a smaller scale.

Data

We will be utilizing data from [RedistrictingDataHub.org](#), specifically:

- CSV file for 2022 Massachusetts Voter File Elections Turnout Statistics at the 2020 Census Blocks level.
- Shapefile for Massachusetts block PL 94-171 2020 (Census Blocks).

Access to the data requires signing up for a free account.

Data Cleanup

We expect to have to do a good amount of data cleanup. While election data and demographic data have already been attached to 2020 Census Block IDs, these IDs will need to be matched with those from the Shapefile. Variables in the dataset may also need to be encoded in the right format to work with our choice of model. Furthermore, since this is a joint project with AC209A (Introduction to Data Science), additional data cleaning may be needed to incorporate supplementary demographic data for a more thorough analysis.

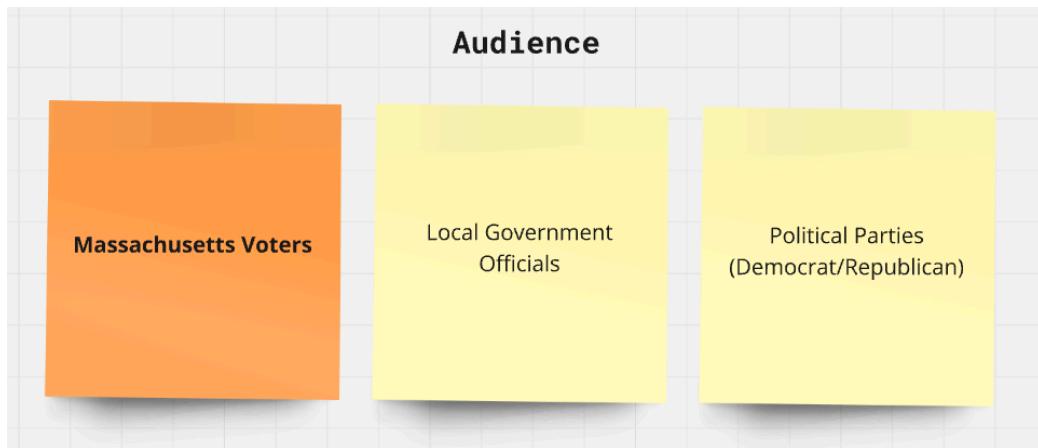
We hope to use the racial/ethnicity demographic information and voter turnout rates from the Census Survey that are both present in the Voter Turnout CSV file.

Data processing will be conducted via OpenRefine and Python scripts.

Project Map

Team

1. Who is your audience? Come up with at least three options and pick one target audience.



Three potential audiences for our visualization project are:

- Massachusetts Voters
- Local Government Officials
- Political Parties

For our final project, we will focus on Massachusetts voters, creating a data story that informs people about voter turnout for the state they live in.

2. Describe your target audience in more detail. What do they know? What are their interests? What visualization literacy do they have? At what level of detail will you present information to them?

Voters in Massachusetts will know the area that they live in, but may not be so familiar with other parts of the state. Similarly, they may be very interested in their immediate vicinity or neighborhood but less so in parts of the state that are very far away. In terms of visual literacy, they will likely be able to comprehend visualizations that are on the simpler side. We hope to present information at the level of Census Blocks, which are smaller than 5-digit ZIP codes. However, we will also aggregate this information to Census Block Groups, Tracts, or Counties for simpler visualizations at the state level.

3. *What questions about your data will be interesting for your audience? Come up with a list of interesting questions that your audience may have about your data. The more, the better, but your team should come up with at least ten questions.*

Questions		
How does demographics affect voter turnout in Massachusetts?	What is voter turnout in my neighborhood like?	How does voter turnout vary in Western, Central, and Eastern Massachusetts?
How does voter turnout vary in different counties?	How does the size of Metropolitan Statistical Area (MSA) (or size of the city) affect voter turnout?	How does party affiliation affect voter turnout in Massachusetts?
Is voter turnout by demographic group representative of the actual population?	How does voter turnout vary for different elections?	Did young voter's behavior vary in the 2020 election compared to other elections?
	How is voter turnout correlated with distance to the nearest polling location?	

Some interesting questions our audience might have about our data:

1. How does demographics affect voter turnout in Massachusetts?
 2. What is voter turnout in my neighborhood like?
 3. How does voter turnout vary in Western, Central, and Eastern Massachusetts?
 4. How does voter turnout vary in different counties?
 5. How does the size of Metropolitan Statistical Area (MSA) (or size of the city) affect voter turnout?
 6. How does party affiliation affect voter turnout in Massachusetts?
 7. Is voter turnout by demographic group representative of the actual population?
 8. How does voter turnout vary for different elections?
 9. Did young voter's behavior vary in the 2020 election compared to other elections?
 10. How is voter turnout correlated with distance to the nearest polling location?
-
4. *What data do you have? Download the data you picked from the website linked in the PDF that describes the data (available on Canvas, week 2). Look at it in Excel or Google spreadsheet and briefly describe each attribute and its data type (categorical, ordinal, or quantitative) in your process book. It's OK if you are unsure about the data type for some attributes - you can simply describe them (e.g., geographic location).*

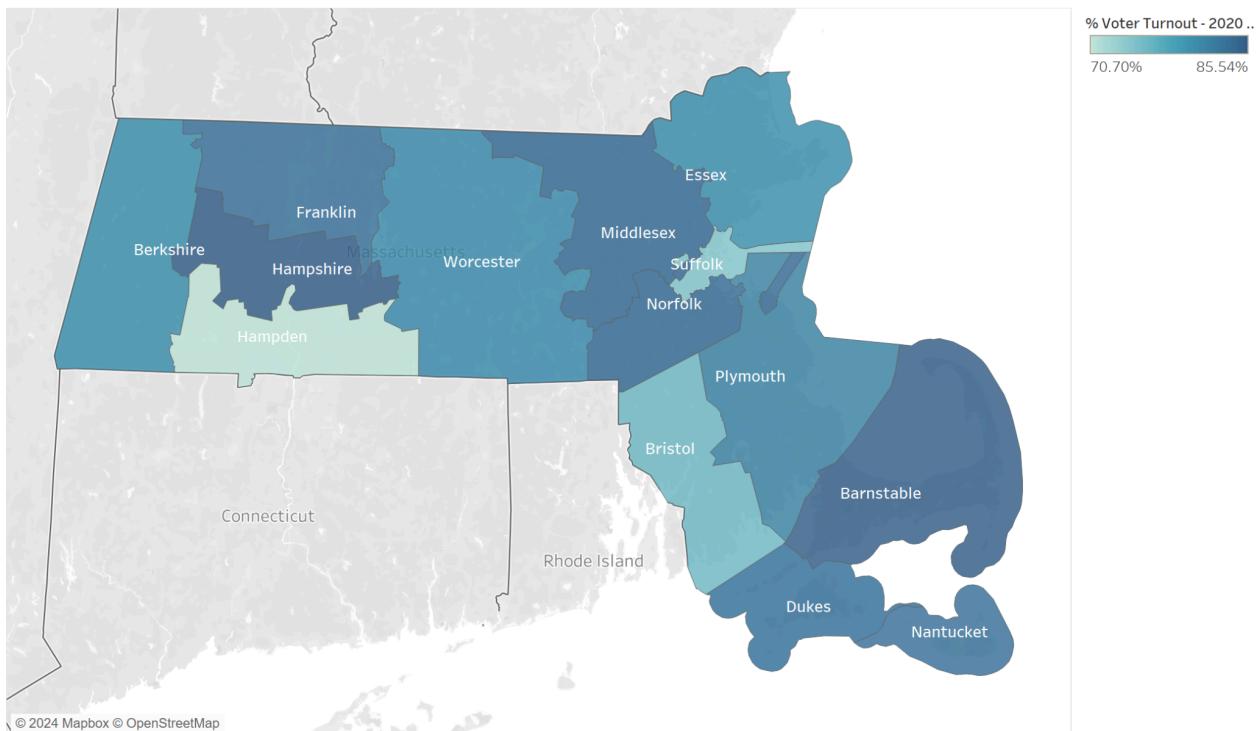
We've conducted some exploratory data analysis in AC209A, most of our data are quantitative variables as they are counts of people at the Census Block level. A descriptive table of our variables can be found below:

range	name	type	description
0	geoid20	15 char string	15-character GEOFID corresponding to 2020 Census Blocks, based on L2 geo-referencing of individual voter addresses
1	total_reg	int	Count of total registered voters in the County, as geo-referenced by RDH from L2 voter file dated above
2-11	age_*_*	int	count of people in age range in that census block
12-14	voters_gender_*	int	gender count
15-24	party_*	int	count of voters registered in party
25-30	eth1_*	int	count of members in broad ethnic category
31-56	eth2_*	int	count of those in narrow ethnic category
57-62	languages_description_*	int	count fluent in language
63	commercialdata_estimatedhhincomeamount_avg	num	Average of modeled data for estimated household income reported by L2 for individuals in the following ranges: \$1,000-\$14,999/ \$15,000-\$24,999/ \$25,000-\$49,999/ \$50,000-\$74,999/ \$75,000-\$99,999/ \$100,000-\$124,999/ \$125,000-\$149,999/ \$150,000-\$174,999/ \$175,000-\$199,999/ \$200,000-\$249,999/ \$250,000+
SKIP	g20221108_*	num	not covering this election
SKIP	p20220906_*	num	not covering this election
SKIP	s20211214_*	num	not covering this election
SKIP	s20211130_*	num	not covering this election
SKIP	s20211102_*	num	not covering this election
SKIP	s20210330_*	num	not covering this election
SKIP	s20210302_*	num	not covering this election
*	g20201103_*	num	2020 general election
SKIP	p20200901_*	num	not covering this election
SKIP	s20200602_*	num	not covering this election
SKIP	pp20200303_*	num	not covering this election - 2020 presidential primary

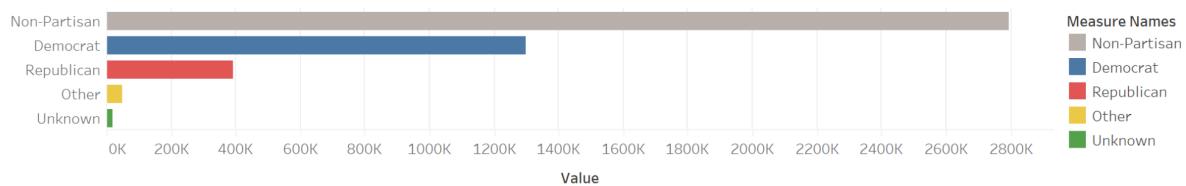
Individual

Max Bahar

Voter Turnout for the 2020 Presidential Election by County

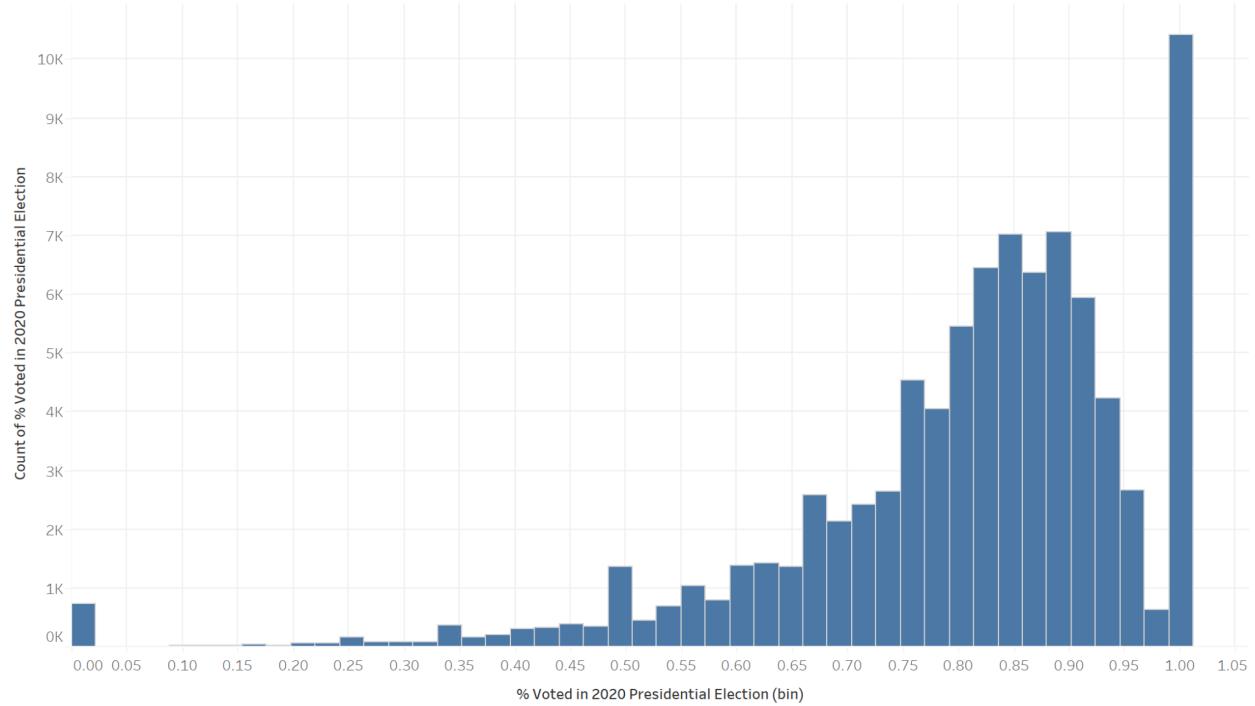


Distribution of Voters by Party Affiliation



Non-Partisan, Democrat, Republican, Other and Unknown. Color shows details about Non-Partisan, Democrat, Republican, Other and Unknown.

Distribution of Voter Turnout in Census Blocks in the 2020 Presidential Election



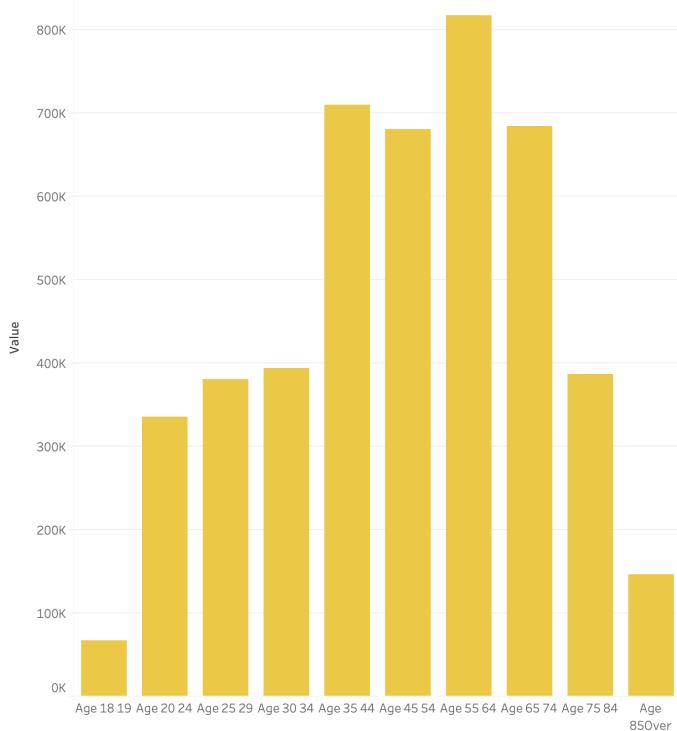
The trend of count of % Voted in 2020 Presidential Election for % Voted in 2020 Presidential Election (bin).

The first visualization shows voter turnout for the 2020 presidential election by county, answering the question "How does voter turnout vary by county?" The second visualization shows the party affiliation of voters in Massachusetts, answering the question "What is the distribution of party affiliation in Massachusetts voters?" The final visualization shows the distribution of voter turnout in the 2020 presidential election by Census Blocks, answering the question "What is the typical voter turnout for Massachusetts Census Blocks?" I think that the map is the best among all the visualizations, as it provides valuable context for the viewer, allowing them to see how voter turnout is where they live. Some of our original questions like "How does party affiliation affect voter turnout in Massachusetts?" require some additional work to establish causality, and is outside the scope of how we've been utilizing Tableau so far.

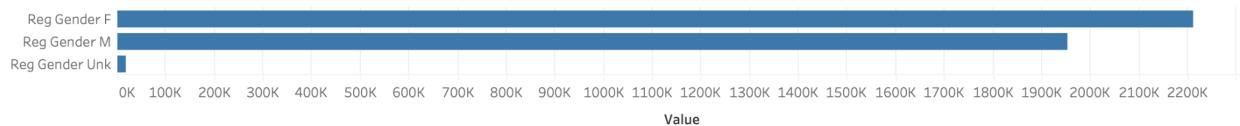
Anitej Thamma

Theodore McCulloch

Distribution of Voters by Age



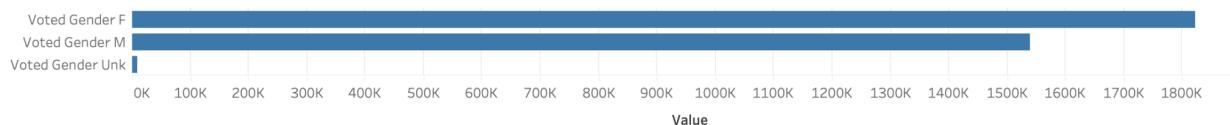
Registration on or before 2020 General Election



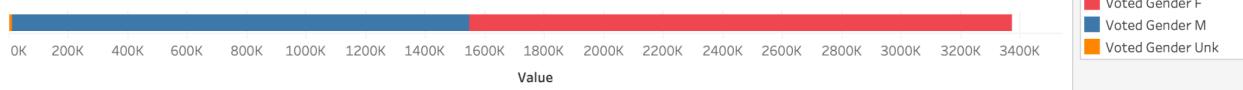
Registration on or before 2020 General Election



Votes by Gender 2020 General Election



Votes by Gender 2020 General Election

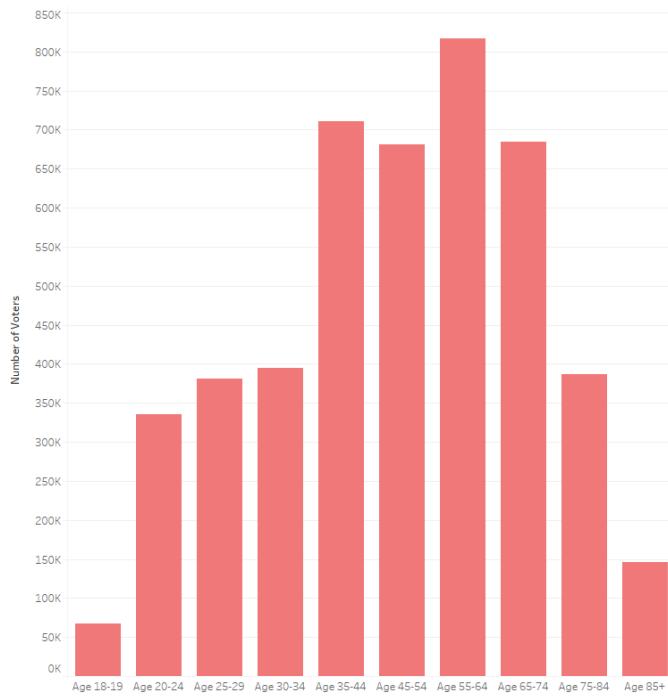


My first question was what does the voting population look like by age? This is addressed by the histogram. My next question: what was both the voter turnout and registered population separated by gender? I have four graphs for this; two of the graphs are different displays of the same information. Many of our questions will involve some form of regression or model fit,

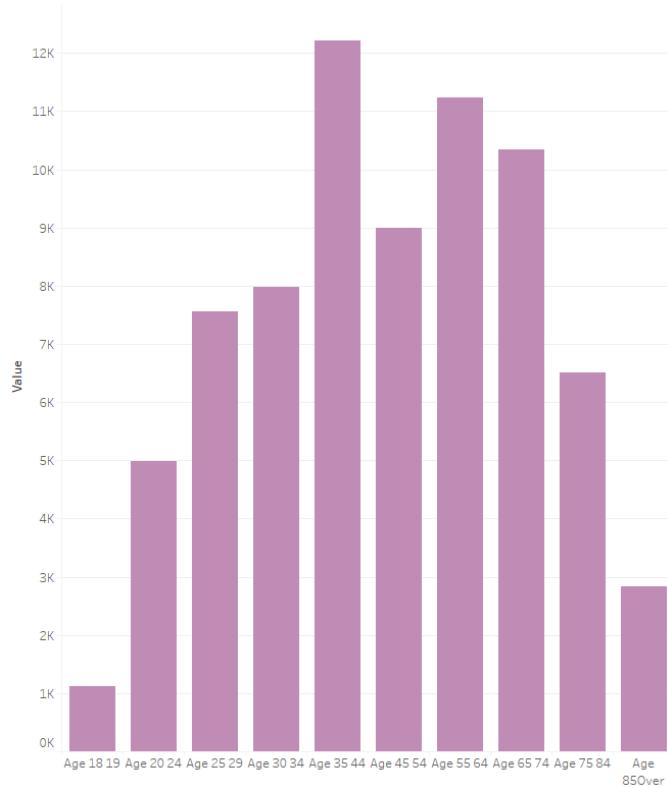
which is outside of how I've used Tableau up until now. That's why I focused on more narrower questions that involved count comparisons of different categories. You can glean a lot by looking at the shape of data, but determining relationships or causality requires more statistical footwork.

Stefan Chu

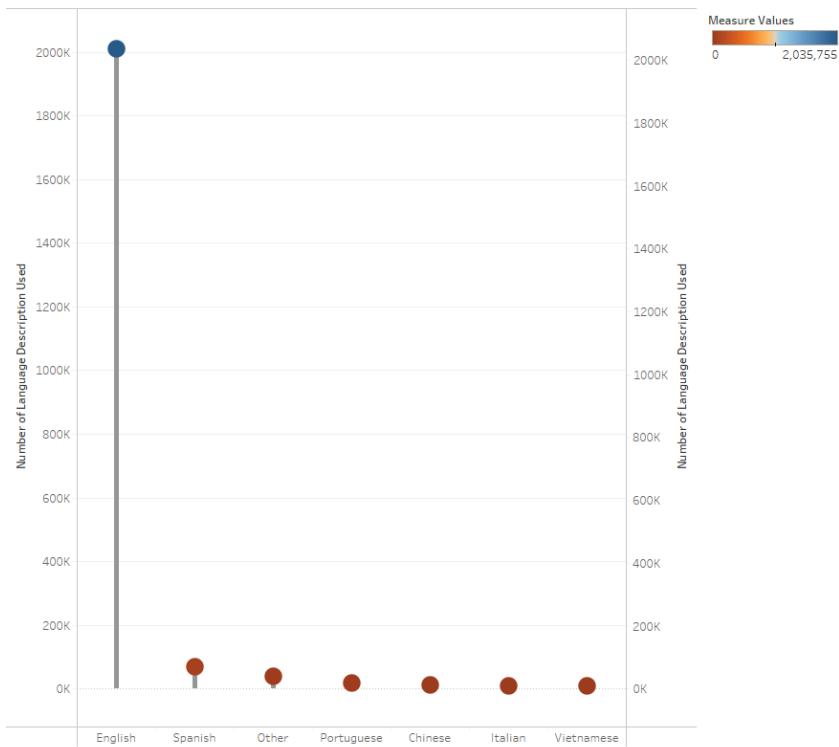
Voter Histogram by Age Brackets



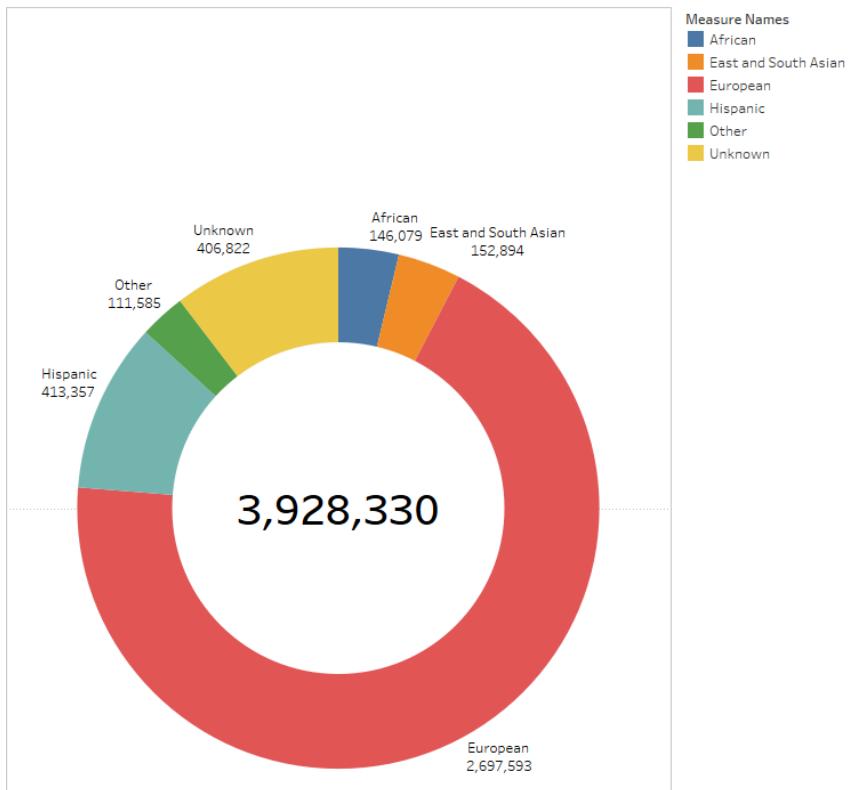
Dropped Voter Histogram by Age Brackets



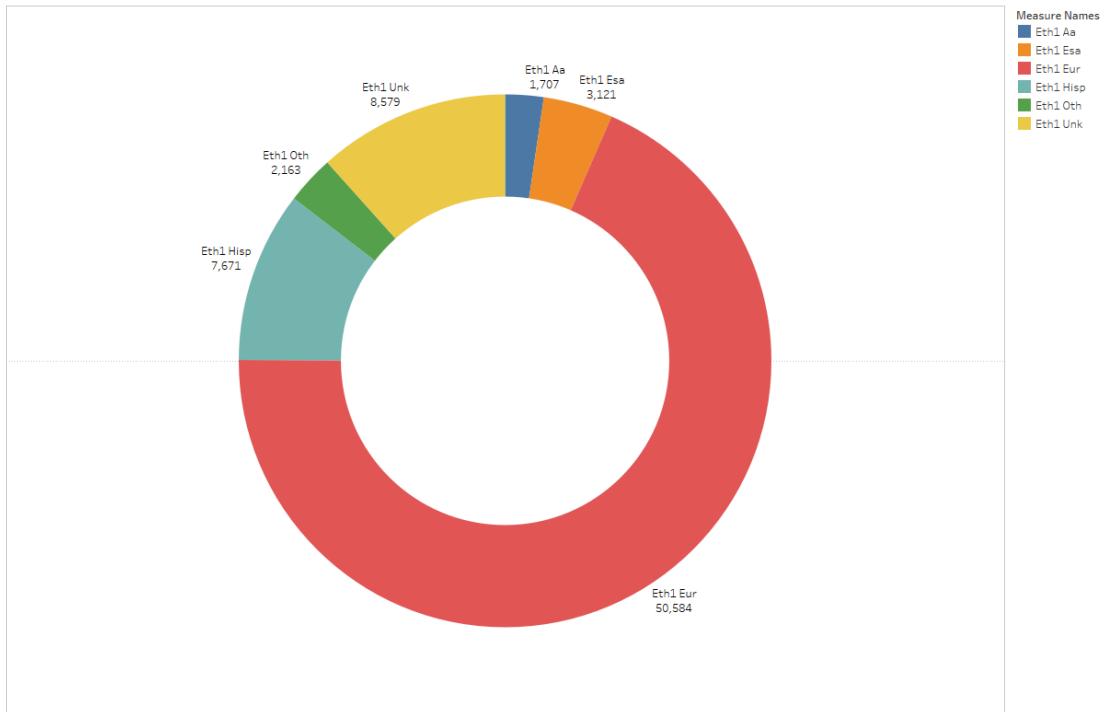
Language Description Lolipop Chart



Voter Count by Ethnicity Donut Chart



Dropper Voter Count by Ethnicity Donut Chart



Total: 73825

I looked at the distribution of voters by age. It is interesting that the Census blocks do not bracket the ages by the same span of years. It could be worth looking into. We are interested as a group in seeing whether many young voters are coming out to vote. Personally, I was interested in not only how demographic features such as ethnicity affect voter turnout, but also the availability of language resources. Many bilingual voters vote in English, but there are many citizens who prefer to vote in a language other than English, with Spanish being the 2nd most prominent language in the United States. The lollipop chart shows the vast difference that may not reflect the actual proportion of the population. Finally, the donut chart at the end shows the proportion of Massachusetts voters by ethnicity. One of our groups questions is whether this is reflective of the true proportions of each ethnicity within each county.

Data, Sketches, Decide & Storyboard

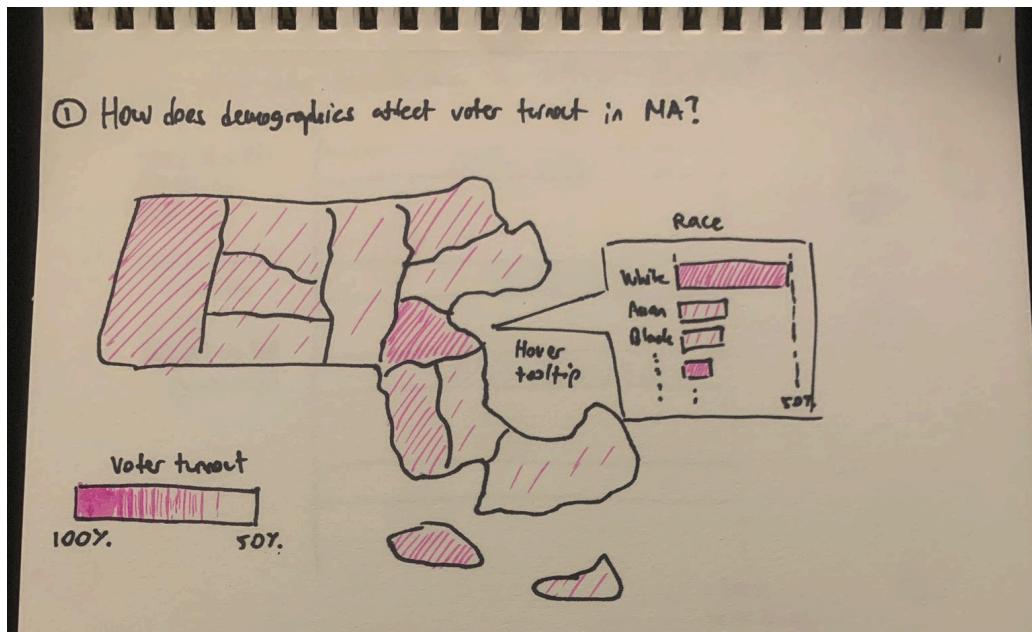
Sketches

Questions

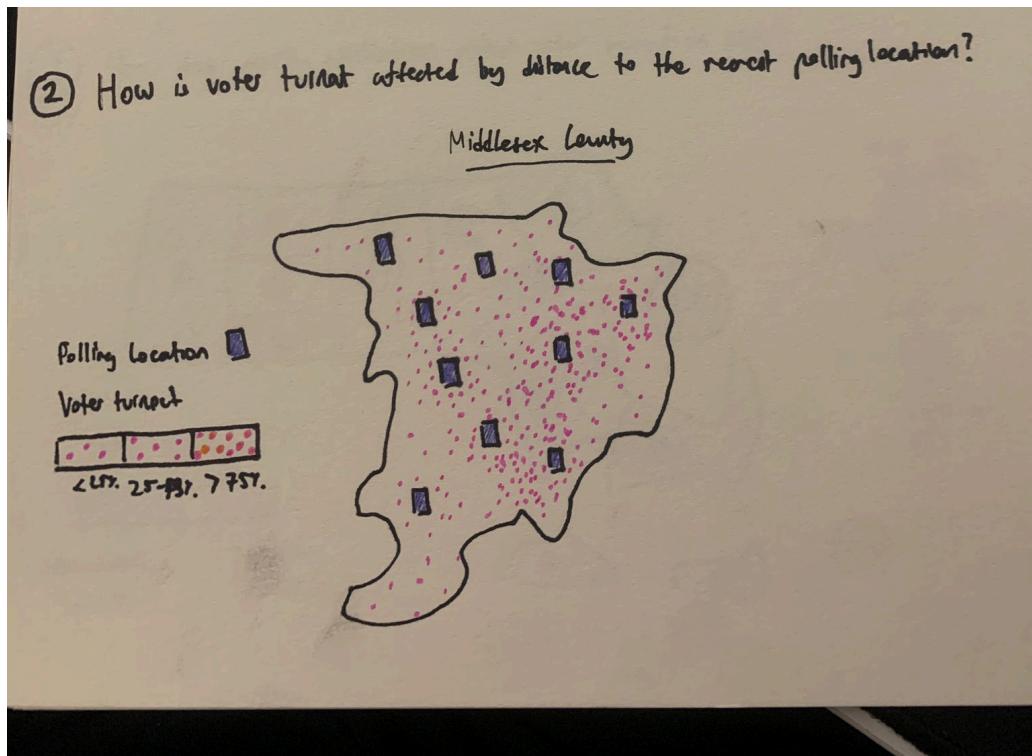
1. How does demographics affect voter turnout in Massachusetts?
2. What is voter turnout in my neighborhood like?
3. How does voter turnout vary in Western, Central, and Eastern Massachusetts?
4. How does voter turnout vary in different counties?
5. How does the size of Metropolitan Statistical Area (MSA) (or size of the city) affect voter turnout?
6. How does party affiliation affect voter turnout in Massachusetts?
7. Is voter turnout by demographic group representative of the actual population?
8. How does voter turnout vary for different elections?
9. Did young voter's behavior vary in the 2020 election compared to other elections?
10. How is voter turnout correlated with distance to the nearest polling location?

Max

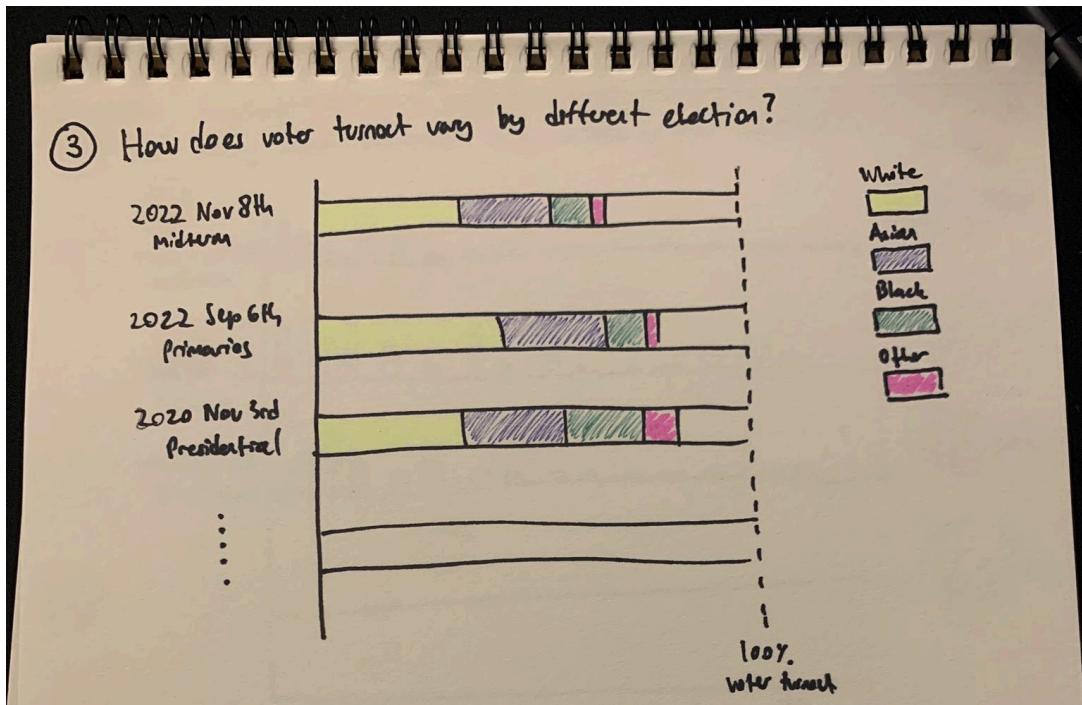
Sketch 1: Questions 1, 3, 4



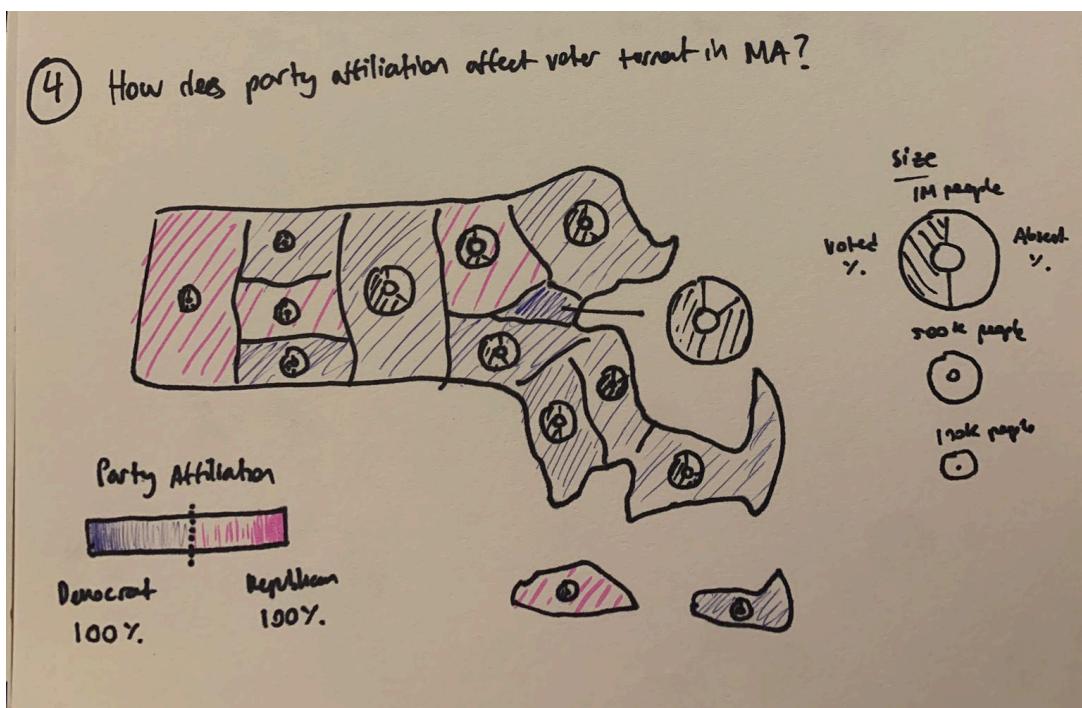
Sketch 2: Questions 2, 10



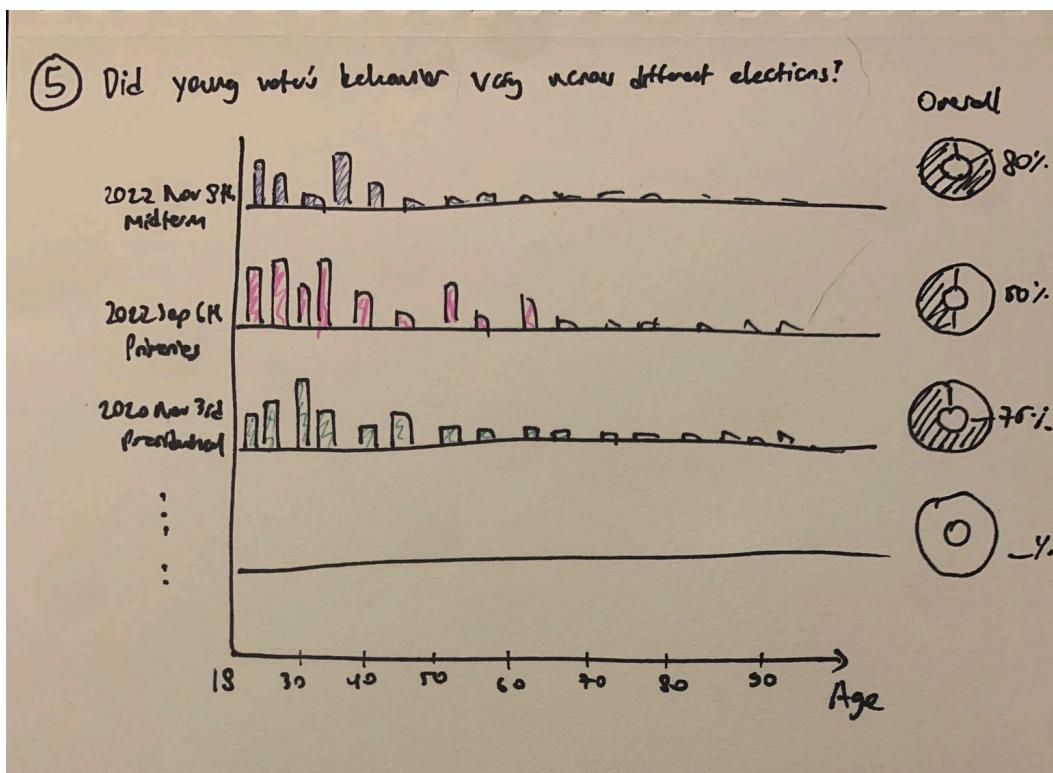
Sketch 3: Questions 1, 7



Sketch 4: Questions 3,4,6

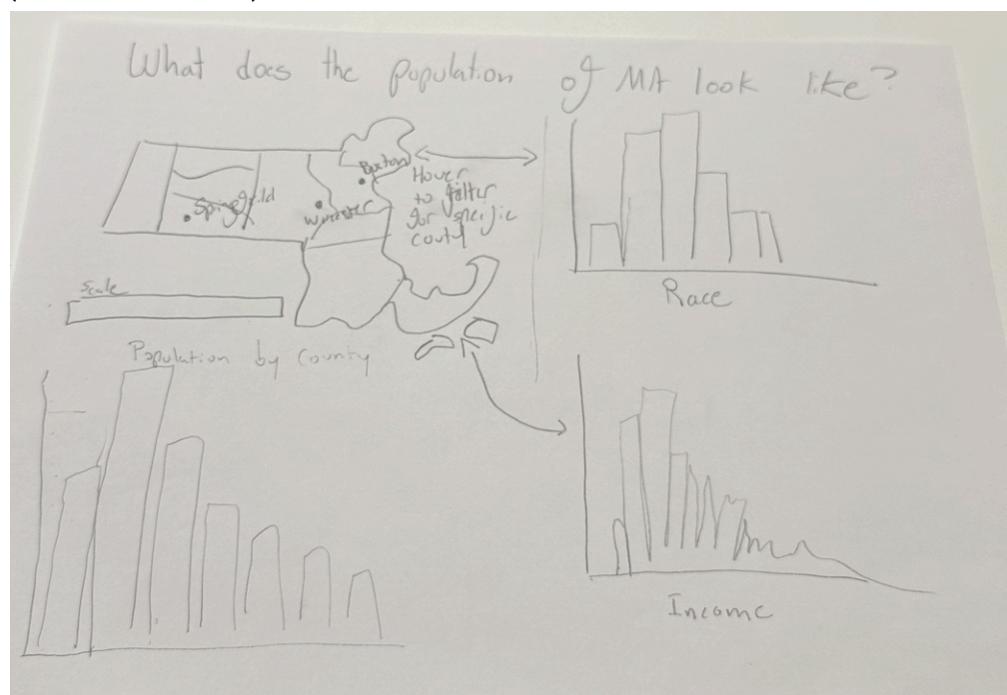


Sketch 5: Questions 1, 8, 9

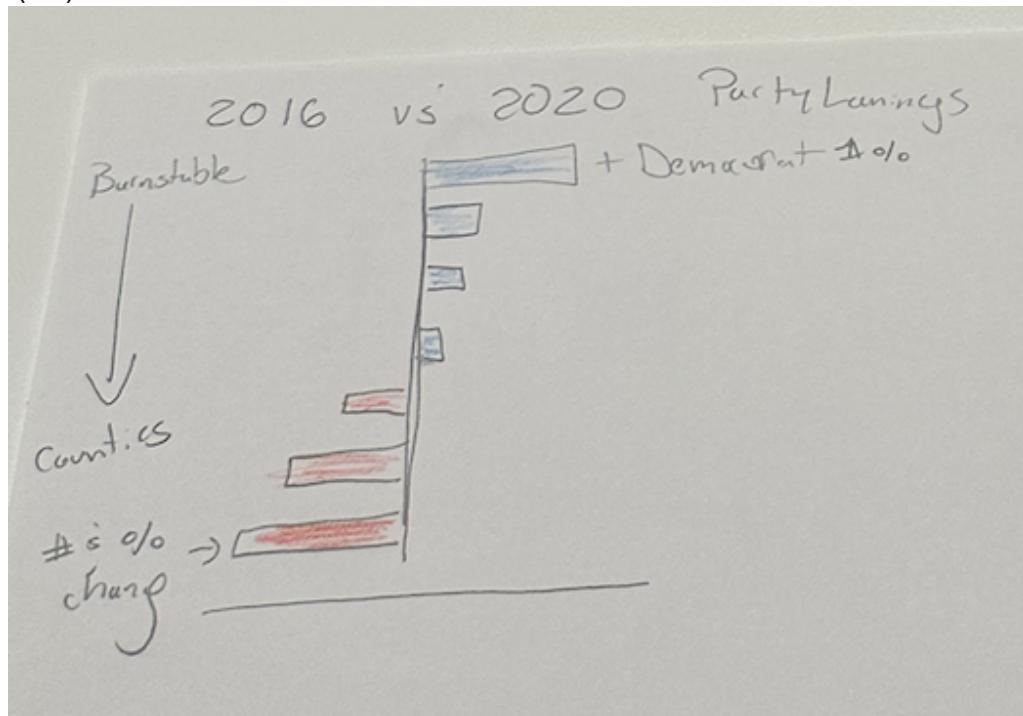


Ted (Sketches 6-10)

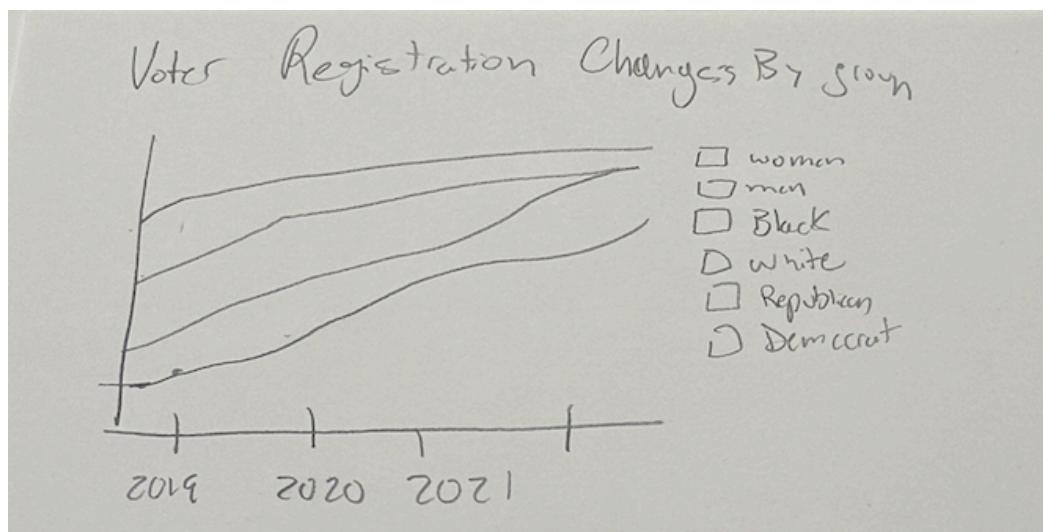
Sketch 6 (Q1, Q2, Q3, Q4, Q5)



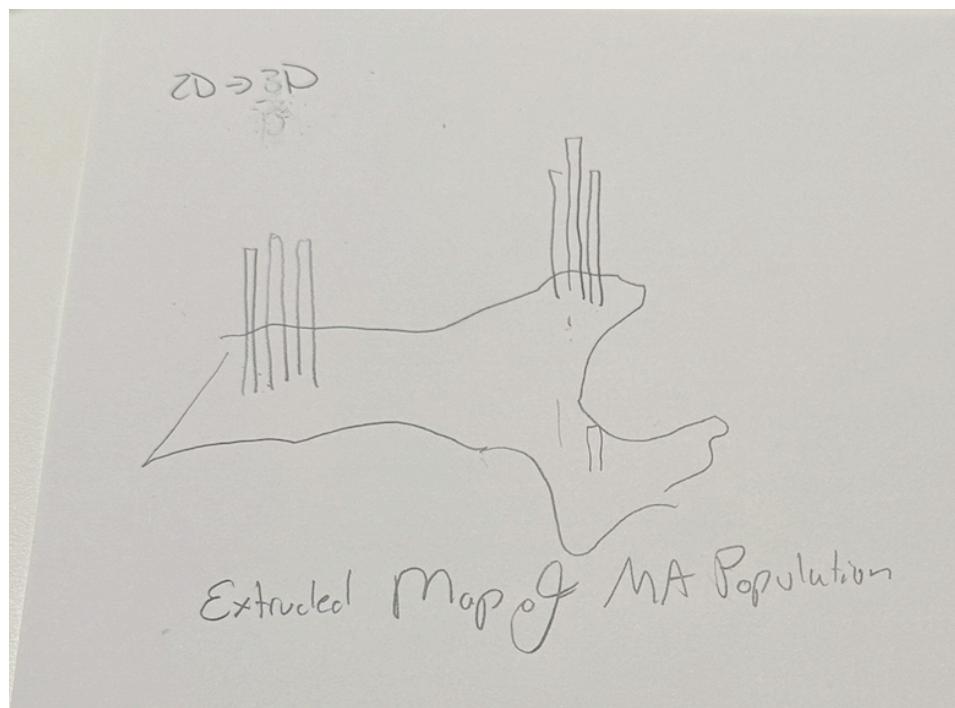
Sketch 7 (Q8)



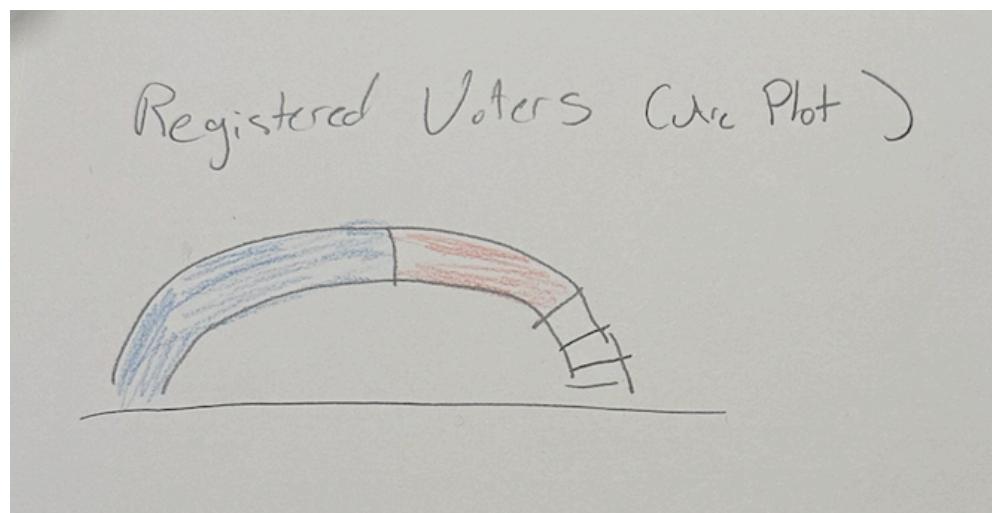
Sketch 8 (Q1, Q2, Q8, Q9)



Sketch 9 (Q1, Q2, Q5)

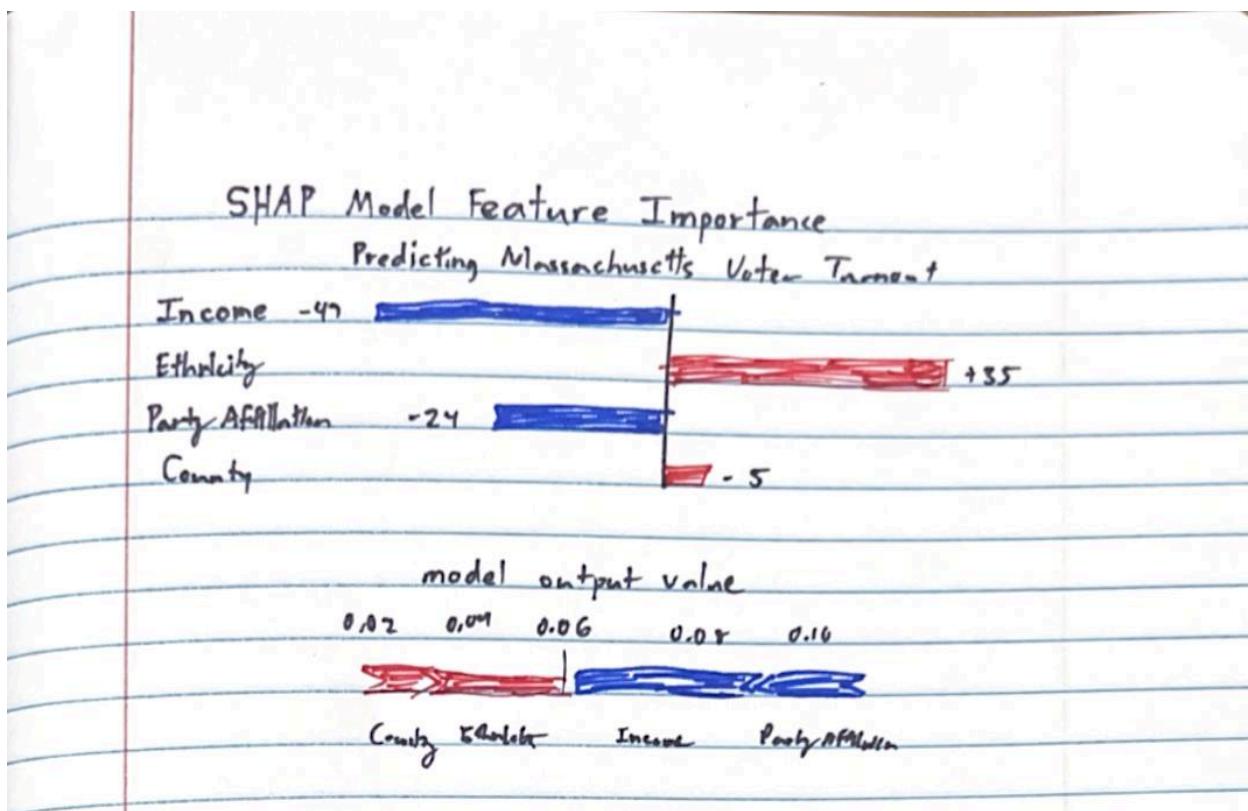


Sketch 10 (Q1, Q6, Q7)



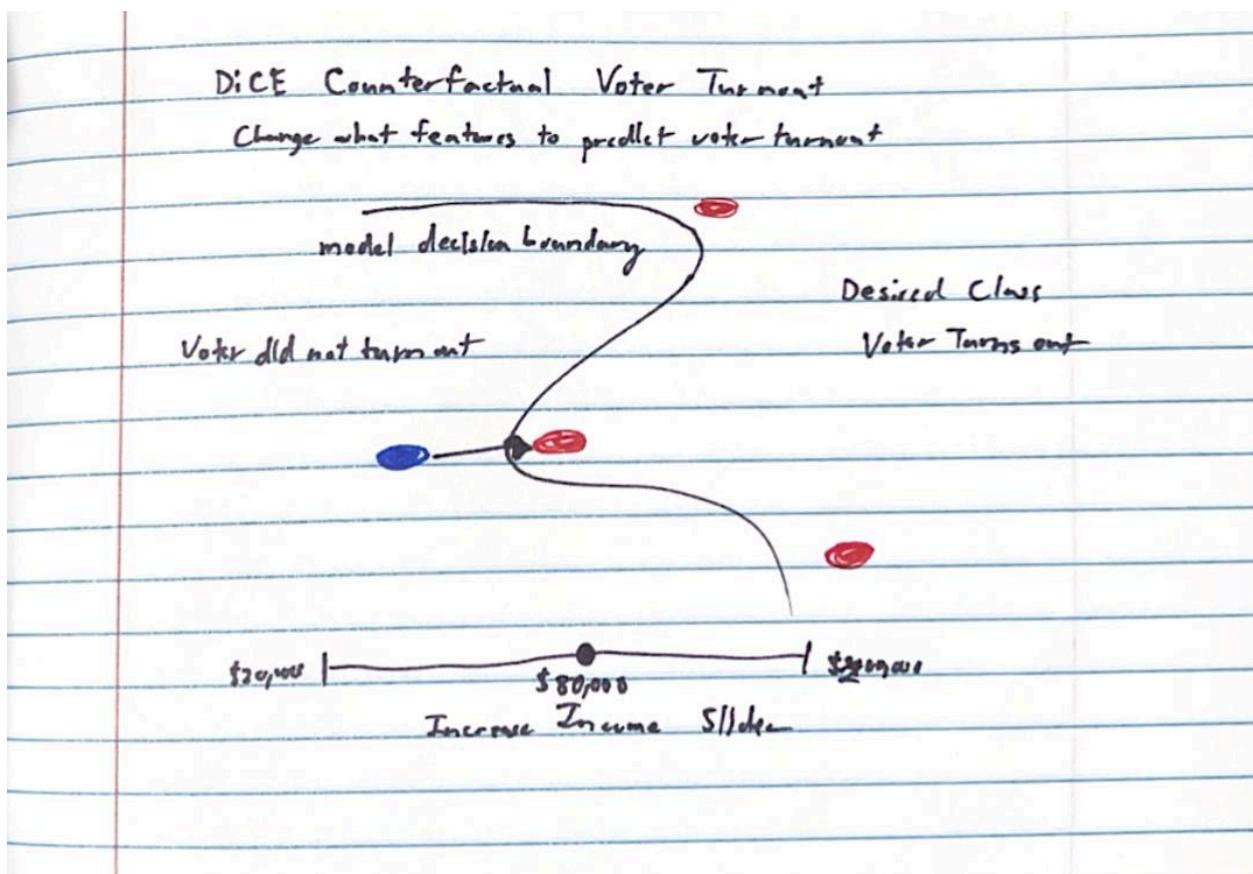
Stefan

11. How does demographics affect voter turnout in Massachusetts?

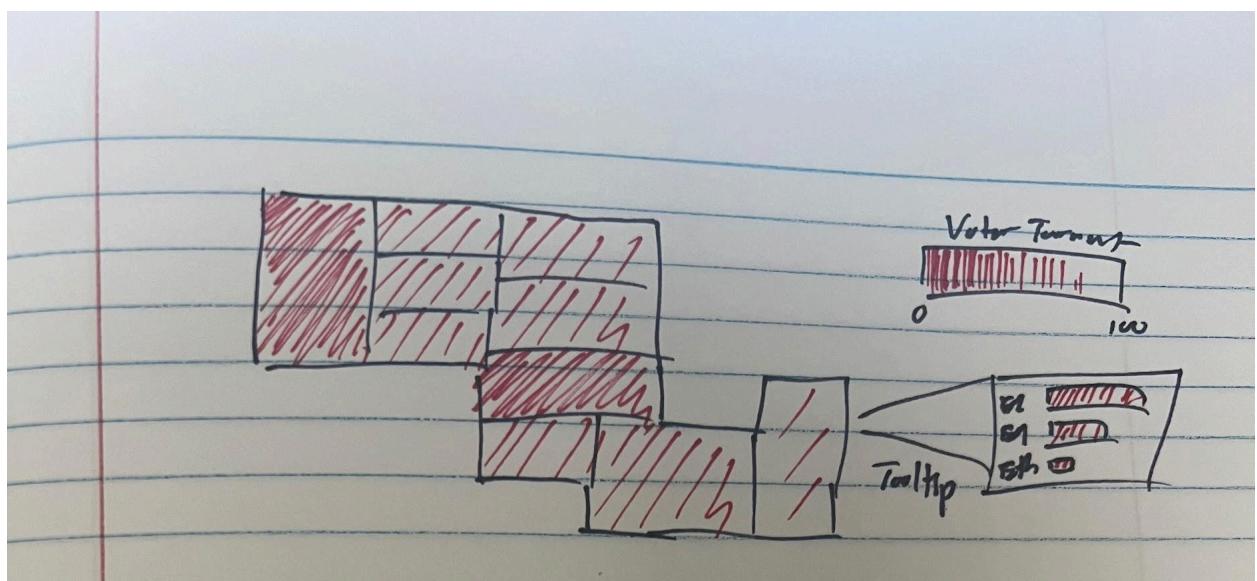


12. How does demographics affect voter turnout in Massachusetts?

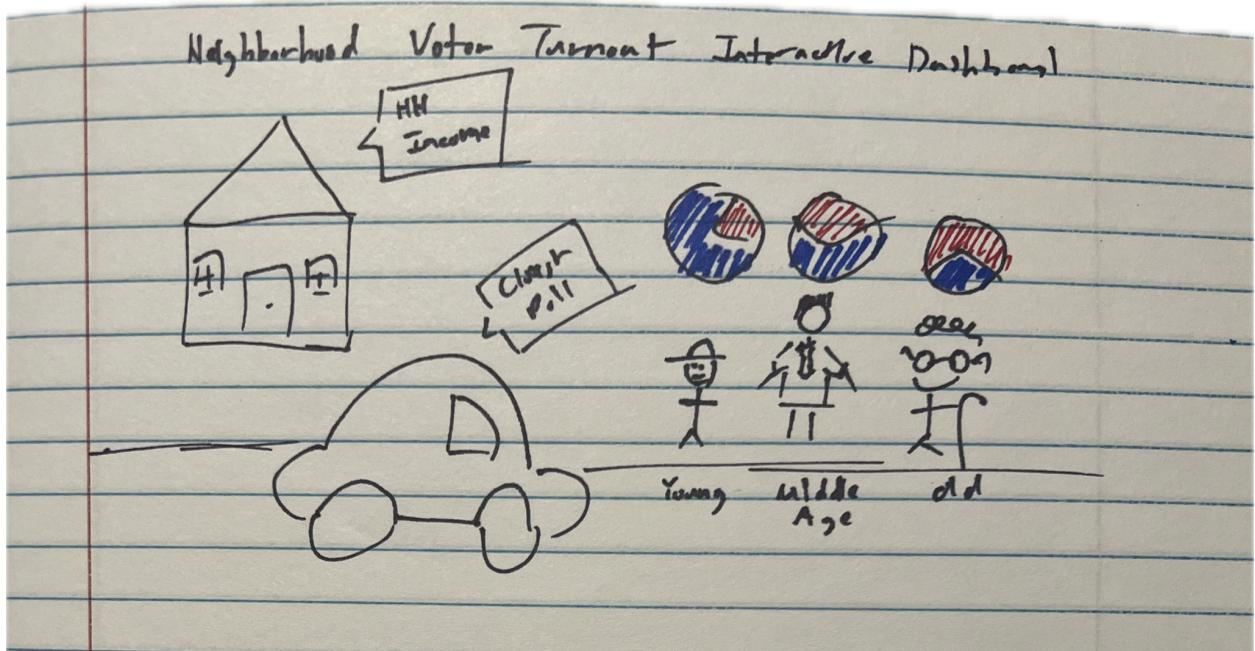
Note: At census block level, what needs to change to increase voter turnout?



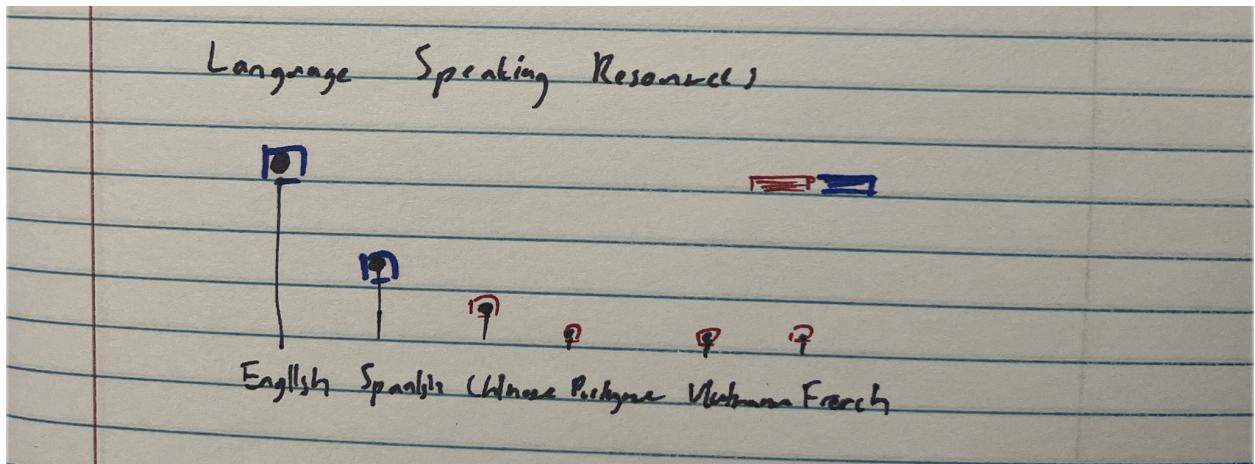
13. How does demographics affect Voter Turnout in Massachusetts?



14. What is voter turnout in my neighborhood like?

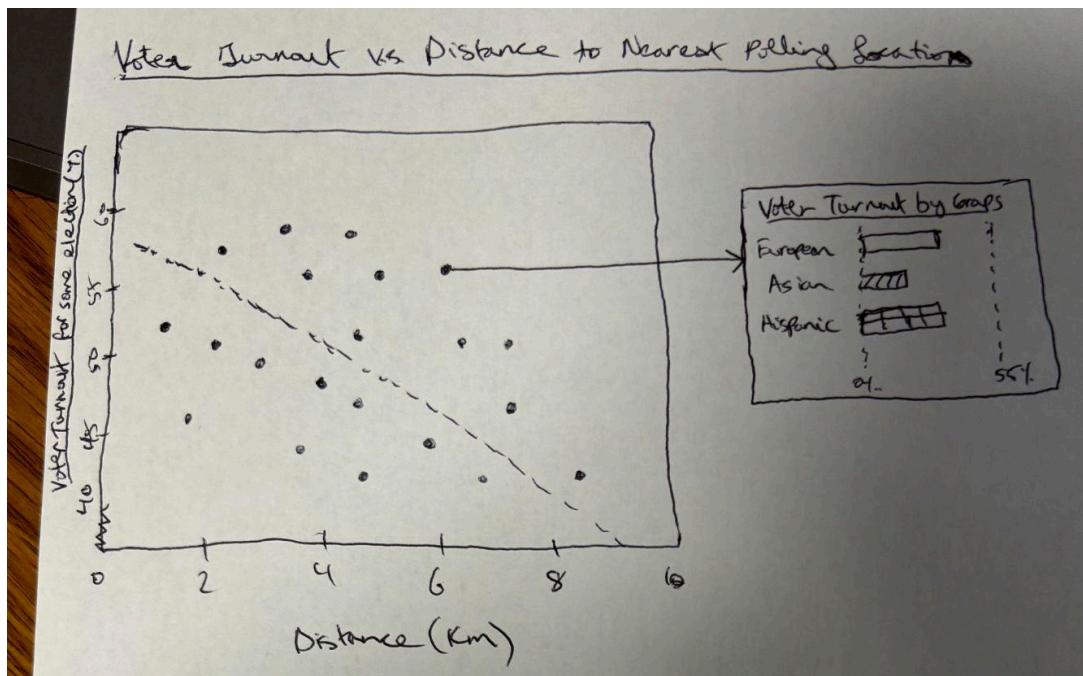


15. Language Resources in Massachusetts



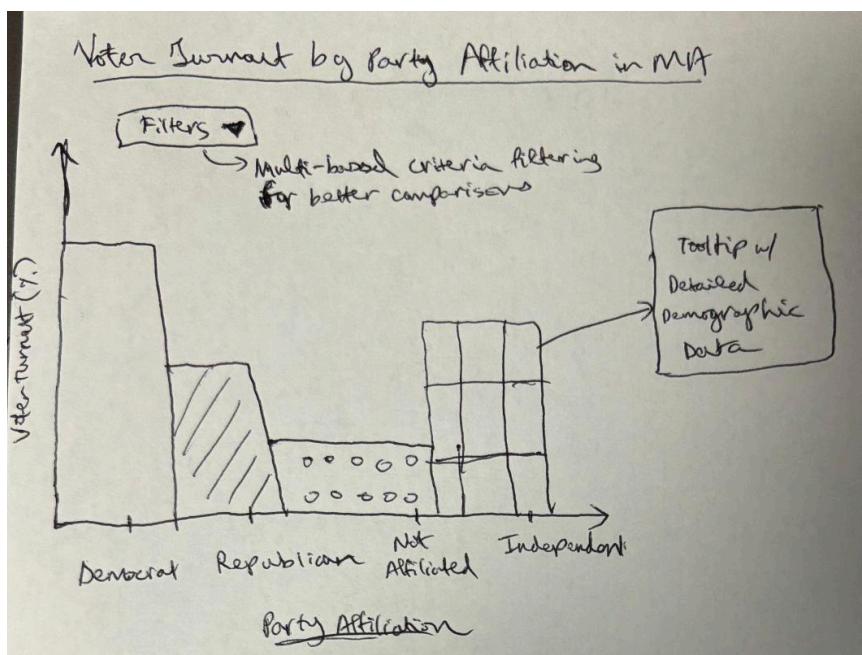
Anitej:

Sketch 16 (Q10)



- Hovering over data points shows more detailed data on ethnic groups voter turnout (%)
- We can add potential filtering based on election year, zip code, etc.

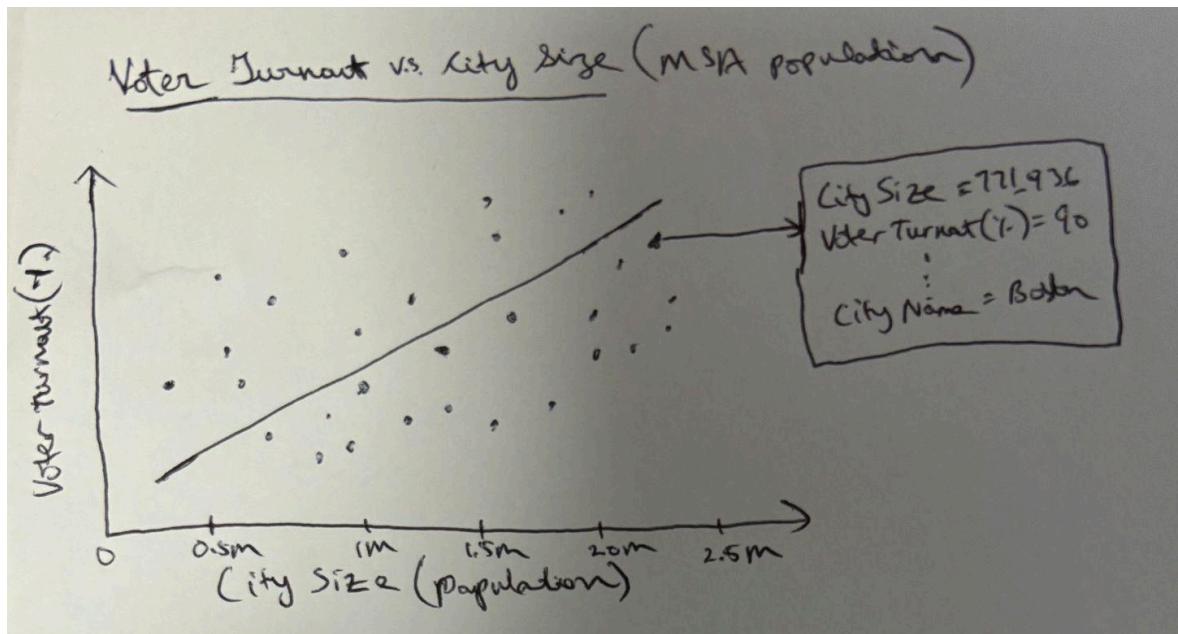
Sketch 17 (Q6)



- Tooltip for detailed demographic data

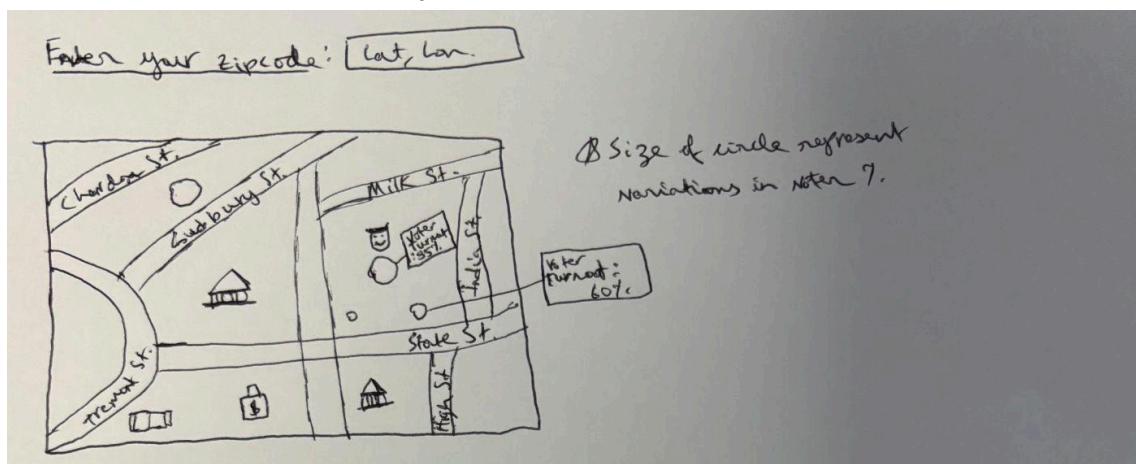
- Multi-based filtering for a more specific comparison

Sketch 18 (Q5, Q4)



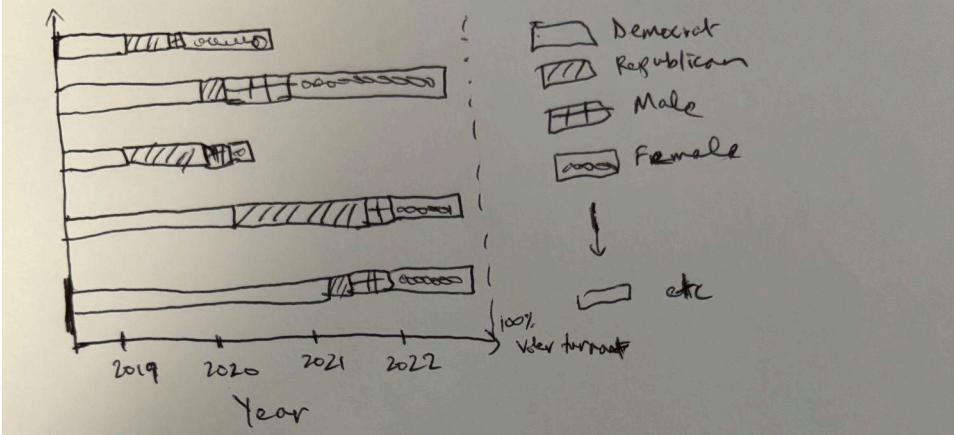
Sketch 19 (Q2)

Title: What is voter turnout in my neighborhood like?



Sketch 20 (Q1, Q2, Q9):

Voter Turnout Changes by Identity Groups



Decide

Affinity Diagramming

Sketch ID	Question ID	Author	Votes
11	1	SC	4
12	1	SC	
7	8	TM	3
2	10	MB	
16	10	AT	
9	1, 2, 5	TM	2
19	1, 2, 5	AT	1
1, 4, 6, 13	1, 3, 4, 6	MB, MB, TM, SC	4
17	1, 3, 4, 6	AT	
18	1, 5	AT	
3	1, 7	MB, SC	
10, 14	1, 7	TM, SC	1
15	1, 7		1
5	1, 8, 9	MB	2
8	1, 8, 9	TM	2

We are choosing to implement the following visualizations:

- Feature importance plot will enable viewers to better interpret our modeling results, seeing the correlation between demographic attributes and voter turnout
- Map visualization will provide much needed context to our viewers, allowing them to place themselves in the data (for example, someone living in Boston can see the demographics and voter turnout there).
- The diverging bar plot will allow users to easily see changes in value across different races, and we can add filtering for user interaction.
- Histogram of age or demographic characteristics allows viewers to see how demographics associated with voter turnout changes over different races and time.

Storyboarding

Insights

Max

- Voter turnout in the 2020 general presidential election seems to be positively correlated with mean household income at the Block Group level, Block Groups with higher income have higher voter turnout rates on average.
- Voter turnout in the 2020 general presidential election seems to be negatively correlated with the proportion of Hispanic voters at the Block Group level, Block Groups with higher percentage of registered Hispanic voters have lower voter turnout rates on average.

Ted

- The data organization by block has led to many empty blocks. For regression, we will likely work in block groups; for visualization, we may default to county.
- All counties went electorally blue in the 2016, 2020, and 2024 elections, so a map with simple party allocation will likely not be visually or informationally interesting. We need to work in degrees of change (or counts) between two elections.

Stefan

- A lot of blocks were empty and had 0 registered voters. In addition to blocks that had missing or invalid data, a total of 76,085 records were dropped, comprising 1.6% of the total population of registered voters in the state of Massachusetts. While this is only a small proportion of the dataset, it is important to gain an understanding of why the data was inaccurate.
- The census block groups that were dropped for having problems such as invalid geographic coordinates and null values showed a similar age and ethnicity distribution to the whole, suggesting that data was missing at random for reasons such as a person incorrectly inputting their address.

Anitej

- There is a significant variation in voter turnout across different ethnic groups. For instance, census blocks with higher European-descent populations show a relatively consistent turnout rate across elections, while blocks with higher Hispanic or African American populations display more fluctuation.
- Gender-specific voter turnout shows that female voter participation is consistently higher in general elections compared to males, especially in 2022.

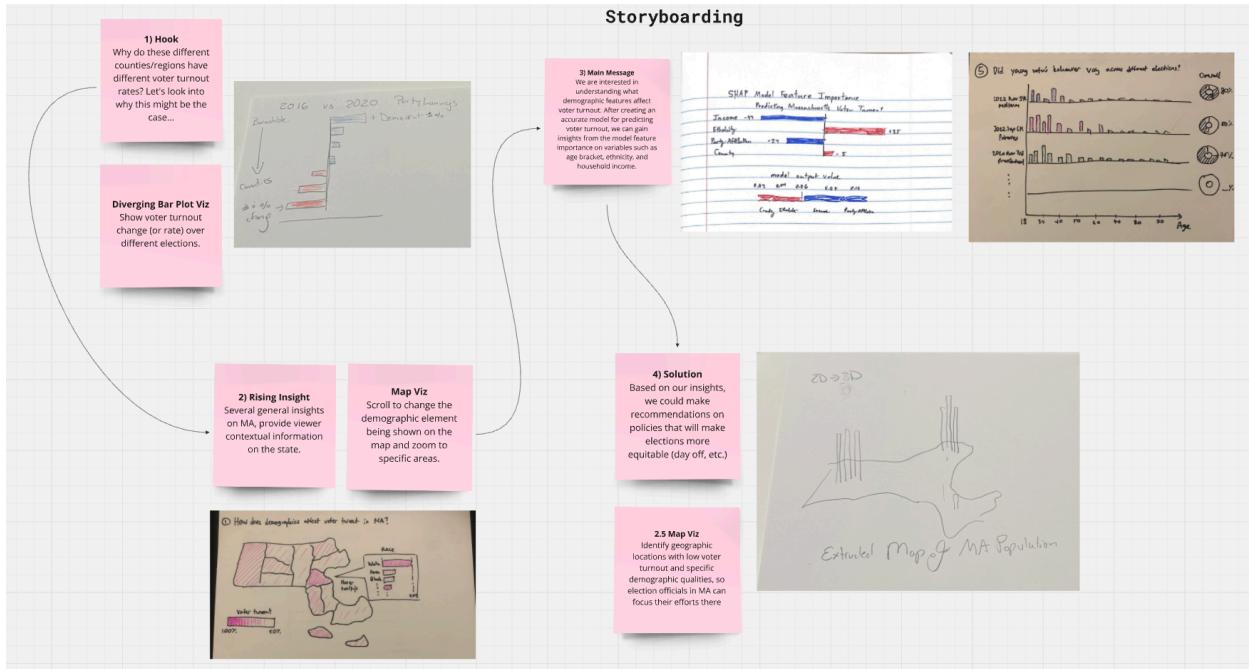
Main Insight

Different demographic groups have different rates of voter turnout, it is important to understand why to promote a fair and just democracy.

We chose this topic because:

- Multiple team members identified discrepancies in voter turnout associated with the different demographic groups.
- As a Massachusetts voter, it is important to understand that the state is composed of many different kinds of people, and how these differences in one's identity can result in inequitable outcomes.

Storyboard



Prototype v1

[Instructions](#)

Name of students that worked on prototype v1 submission

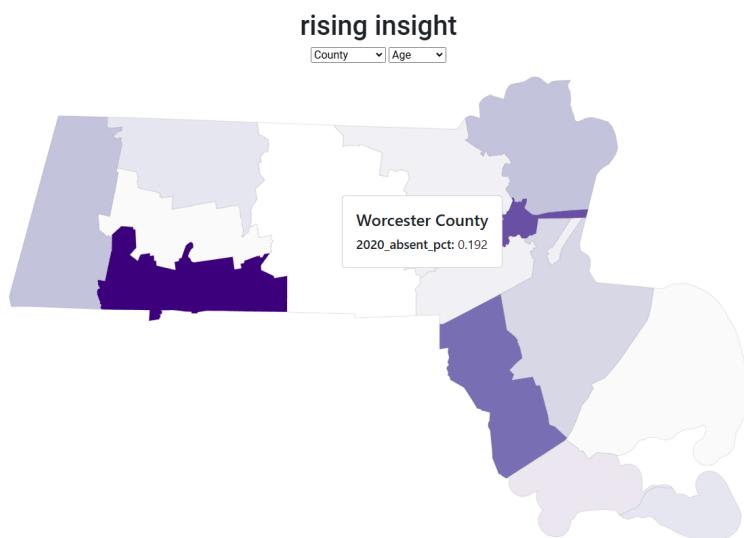
Anitej Thamma
Max Bahar
Stefan Chu
Ted McCulloch

Data scraping and cleaning complete (using the real data sets)

data/ folder contains geojson files used for visualization.

At least two D3 visualizations already partly implemented

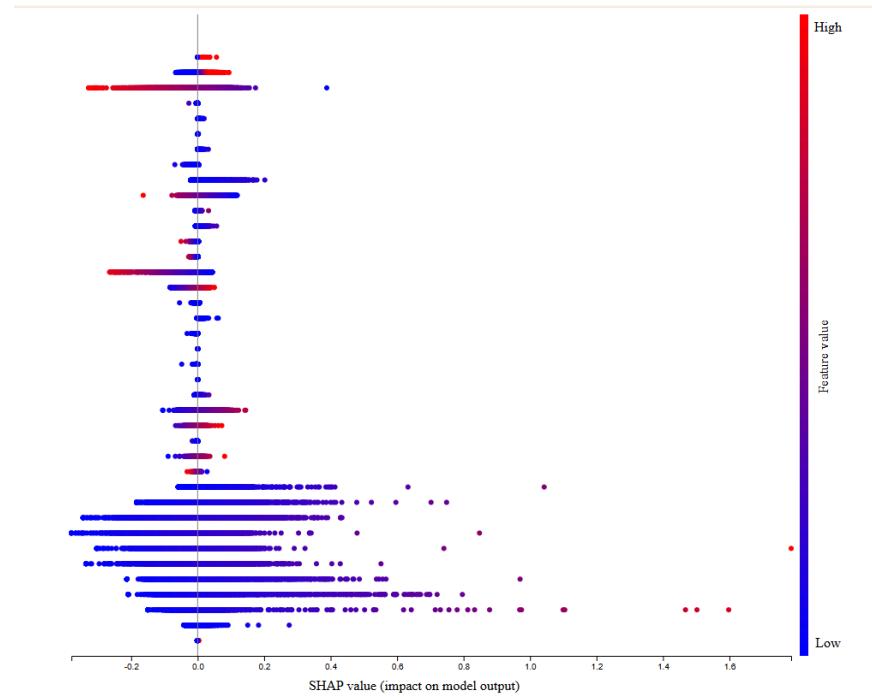
- Map visualization



- Kernel density plot

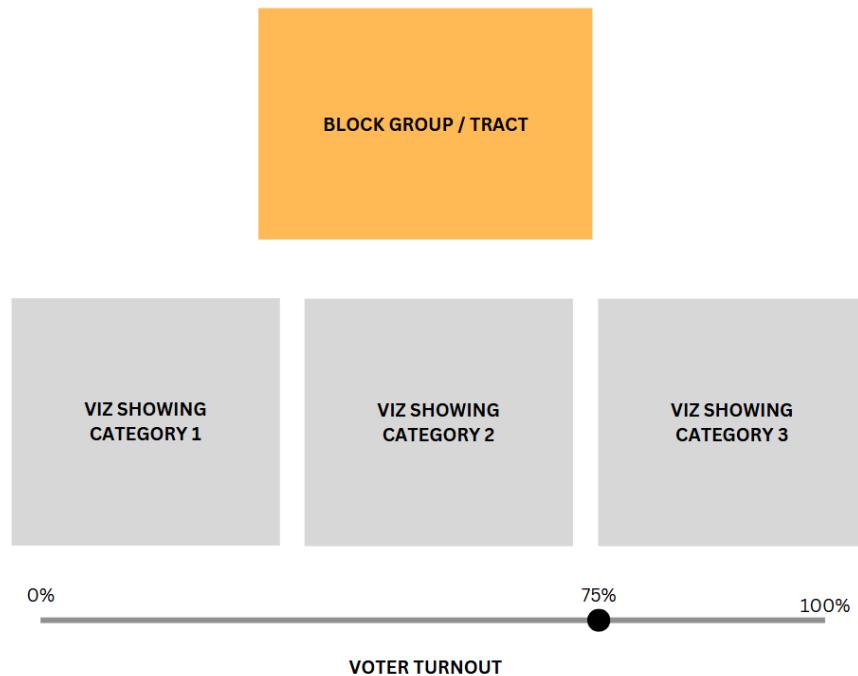


- Feature importance



Detailed drafts for 2-3 more visualizations

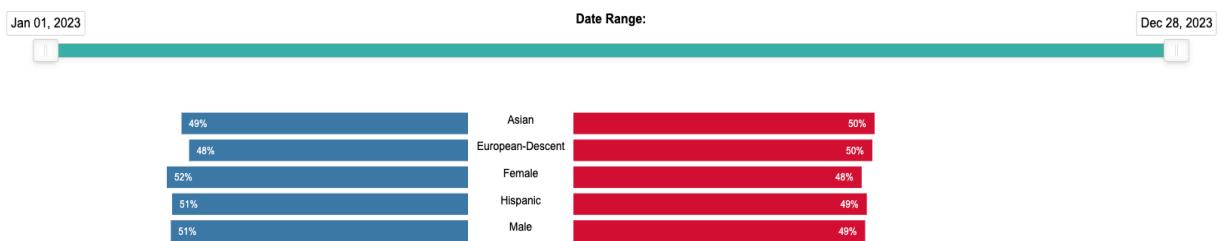
- Slider visualization



- Diverging bar plot

Demographics in Regions with Differing Party Affiliation

What are the distribution of votes?



Rough webpage design and structure

Scrolling behavior and storytelling structure is implemented.

Storytelling is clear

Hook:

- Visualization showing demographic categories in a Census Block Group, asks users to guess:
"Given this demographic composition, what would you predict voter turnout to be?"

Rising Insight:

- *Map visualization* showing that voter turnout/absence is different in different regions.
- *Map visualization slideshow* showing different demographic
- *Diverging bar plots* showing demographics by party affiliation, separate Block Groups into those with more than % of rep/dem voter registrations.
- *Kernel density plots* showing distribution of demographic proportions and correlation with voter turnout.

Main Message:

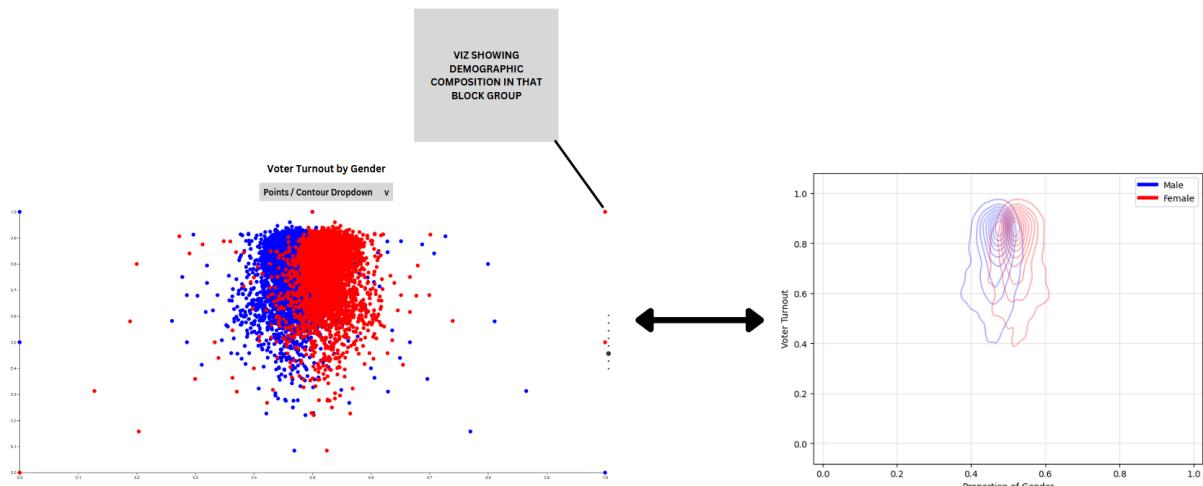
- *Feature importance plot* showing the different predictors in our model and how relevant each are.

Solution:

- How did your prediction compare to our model's prediction?

The first design of an innovative view

- *Kernel density plot*, toggle between points and contour lines.



Interactions (e.g., filtering, brushing, etc.) have to be designed

- Map visualization, interactive drop downs.
- Hook page, interactive demographic selection and voter turnout slider.

	INTERACTIVE	<table border="1"><tr><td style="width: 50%;">CATEGORIES Party Affiliation Democrat, Republican, Independent, etc. Gender Male, Female, Unknown</td><td style="width: 50%; text-align: right;">Select up to THREE <input type="button" value="..."/></td></tr><tr><td>Average Income High, Middle, Low Spoken Language English, Spanish, Italian, etc.</td><td style="text-align: right;">CATEGORIES CONTINUED <input type="button" value="..."/></td></tr><tr><td>Age Younger, Older, etc. Ethnicity White, Black, Asian, etc.</td><td style="text-align: right;"><input type="button" value="..."/></td></tr></table>	CATEGORIES Party Affiliation Democrat, Republican, Independent, etc. Gender Male, Female, Unknown	Select up to THREE <input type="button" value="..."/>	Average Income High, Middle, Low Spoken Language English, Spanish, Italian, etc.	CATEGORIES CONTINUED <input type="button" value="..."/>	Age Younger, Older, etc. Ethnicity White, Black, Asian, etc.	<input type="button" value="..."/>	
CATEGORIES Party Affiliation Democrat, Republican, Independent, etc. Gender Male, Female, Unknown	Select up to THREE <input type="button" value="..."/>								
Average Income High, Middle, Low Spoken Language English, Spanish, Italian, etc.	CATEGORIES CONTINUED <input type="button" value="..."/>								
Age Younger, Older, etc. Ethnicity White, Black, Asian, etc.	<input type="button" value="..."/>								