Final Project

AUTHOR
Kaijie(KJ) WU, Carey Cui, XiaoTian Tang

PUBLISHED
December 7, 2024

respective sections

Kaijie(KJ) Wu: education chart plotting, weakness, policy implication, future direction, and conclusion.

Carey(Zheng) Cui: age chart plotting, data collection and cleaning, slides.

Xiaotian Tang: shiny app, geo map plotting, data collection.

Github Username:

Kaijie(KJ) Wu: jameswkj

Carey(Zheng) Cui: CUI19999

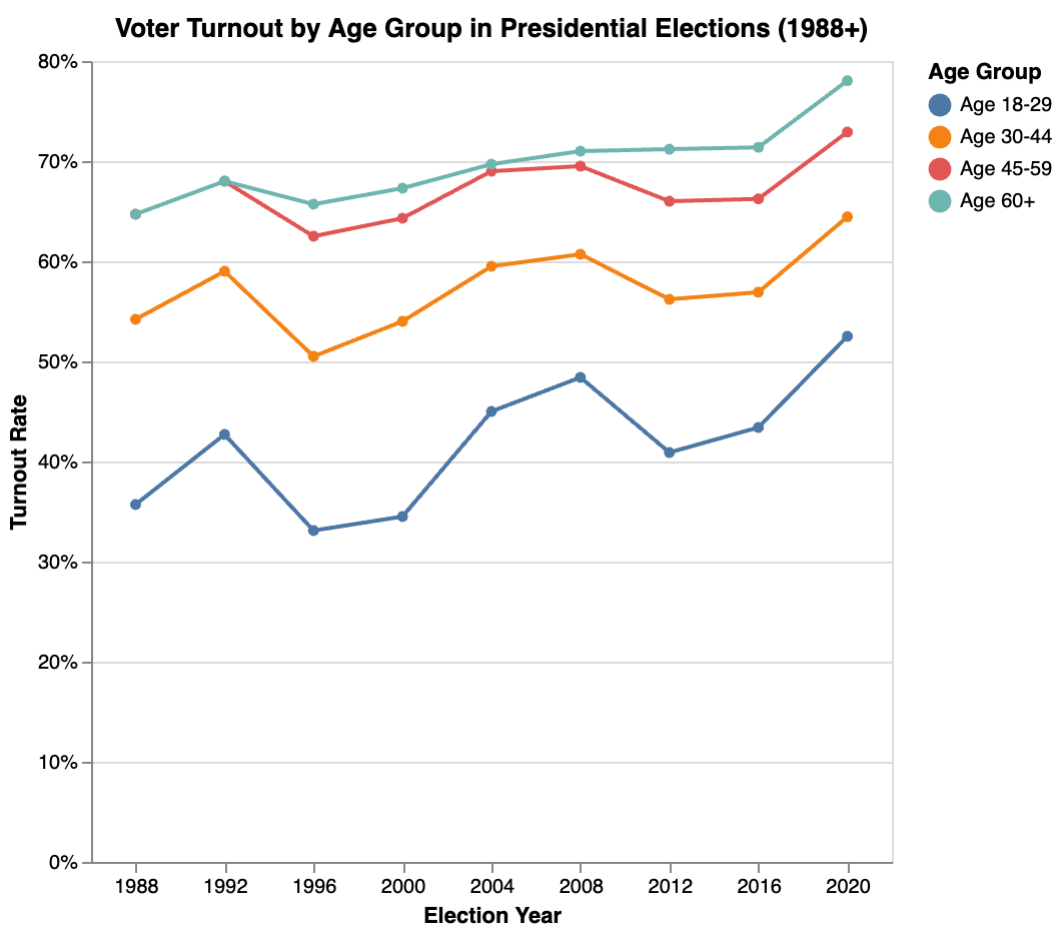
Xiaotian Tang: xiaotiantangishere

Factors that affect turnout rate

This year, 2024, is an election year. We have witnessed a lot of big news: the withdrawal of Joe Biden and Donald Trump's return to the White House. Some states indicate that the turnout rate is lower than the election yeat in 2020. Alabama even declares the lowest turnout rate in over 30 years. These events make us interested in turnout rate. Our research questions are What factors affect turnout rate and how these fatocrs affect turnout rates.

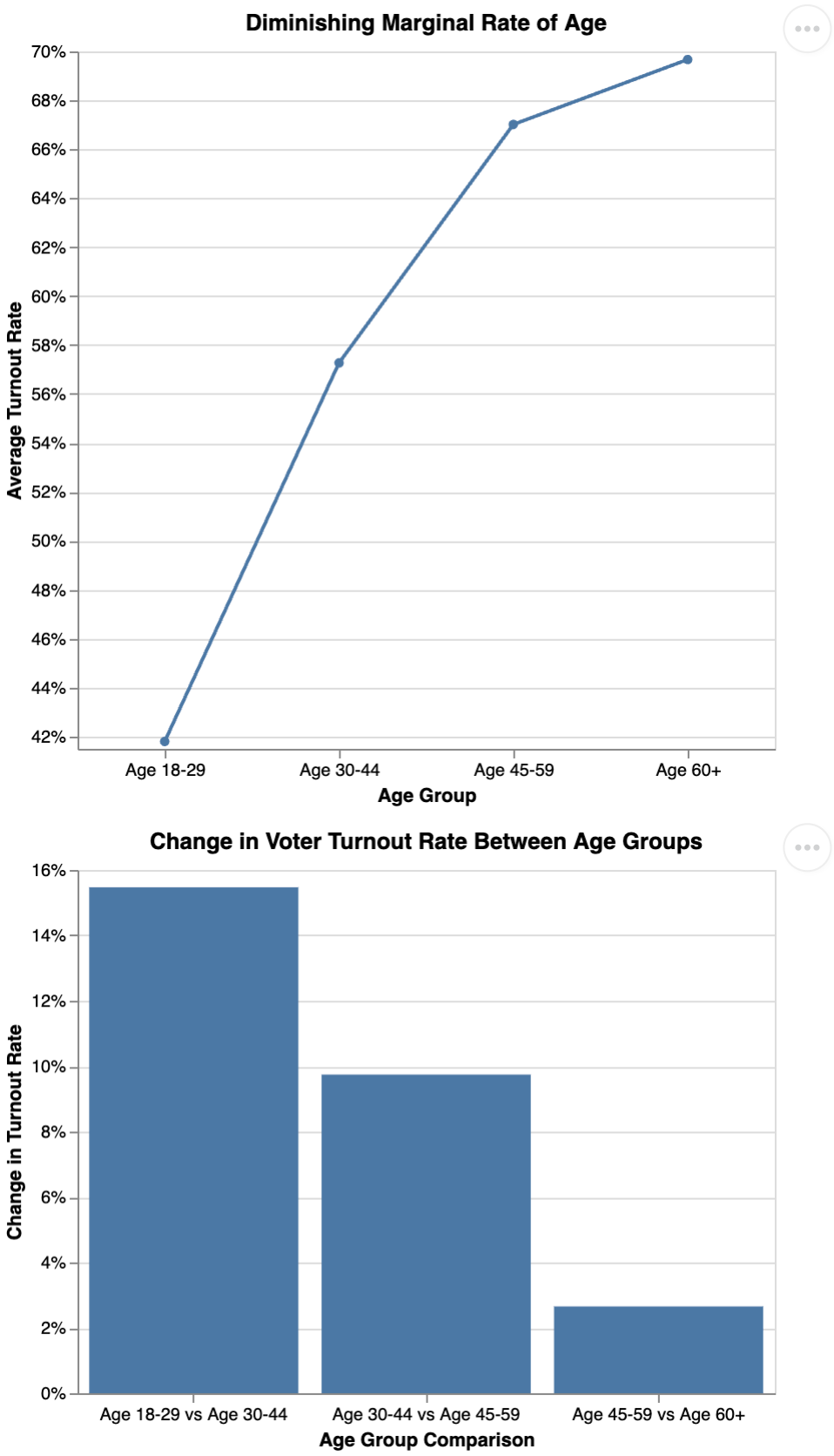
The first factor that we are interested in is *State*. We plan to draw a map reflecting turnout rate in the state level. We use the **State and National Turnout Rates for Presidential Elections** data from the University of Florida Election Lab where it provides statewise presidential elections turnout rate data from 2012 to 2024. The geodata we use is Census shapefile, specifically the zipfile "gz_2010_us_040_00_500k.zip". In the data cleaning process, we basically compared the 'State' columns in the four turnout rate dataset with the geodata using *set* function, to make sure the data be merged correctly. The app_fanciest.py is the shiny app that we created. Running this file with code *$shiny run –reload shiny/app_fanciest.py* in the terminal will display a shiny app with a drop-down menu selecting years, a check box to decide whether comparing the rate with nation level, and a toggle decide whether to showcase the swing states only. The results demonstrate that the Northern part of the US seems to have higher turnout rates than the Southern part, and the swing states, in most cases, have higher turnout rates than the national average.

The second factor we are interested in is age. We want to analyze how age affects turnout rates. We use the Reweighted Current Population Survey Demographic Turnout Statistics data to solve this problem. The data are reweighted by the author using a procedure proposed by Hur and Achen(2013). To start, we observe that the data is distributed over the year, with each year having a different turnout rate for various age groups. Based on these observations, we use Altair to plot the first diagram. We set the election year as the x-axis and the turnout rate as the y-axis, using different colors for each age group. Here is what the first diagram looks like:



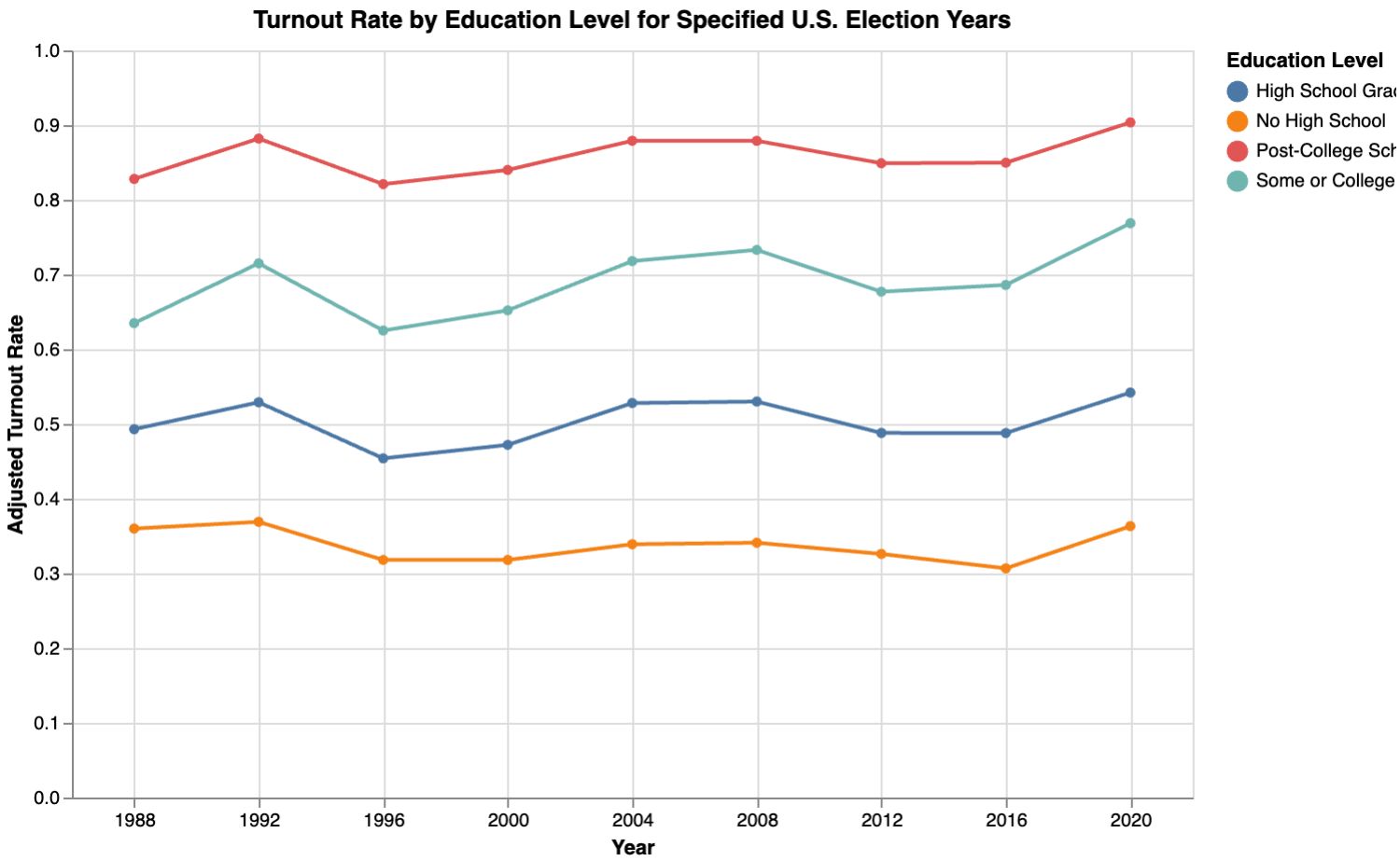
Based on the first diagram, we can see that the age group 18-29 has a relatively lower turnout rate than other age groups over the election years. Additionally, the gap in turnout rates decreases as people age. However, the diagram only provides a general picture of how the turnout rate changes over the election years for each age

group. Because the gaps in turnout rates are not an intuitive way to show how age affects political participation, we decided to average the turnout rates over the election years and calculate the change in turnout rate between each age group. We can then use these results to plot the diminishing marginal rate of age (the change in turnout rate between each age group). We use Altair to create the diagram, setting age as the x-axis and the change in turnout rate as the y-axis. Here are the following plots:



Based on the diagrams above, we can intuitively see the change in turnout rate between different age groups. As people getting older, the turnout rate tends to increase, and the differences in turnout rates diminish. Thus, we can confirm that age does have an impact on turnout rates.

The third factor we are interested in is education. We again use the Reweighted Current Population Survey Demographic Turnout Statistics data. We are looking for the connection between the education level and the turnout rates. As a result, we have four education levels: post-college graduate, some or college graduate, high school graduate, and no high school. We filter the data and have a new data set of turnout rates by different eduation level. Finally, we draw the line chart with four different colors to present the change in turnout rates by education level from 1988 to 2020.



From the above chart, we can see that the citizens with higher education level have higher turnout rates. The citizens with post-college school degrees have the highest turnout rate, but the citizens without high school education have the lowest turnout rates. Additionally, the gaps of turnout rates between each education level are similar between 1988 and 2020 (around 15%).

To increase the turnout rates, there are three potential policy implications. The first one is to optimize the election policies between states. This one would help increase the turnout rates in the non-swing states. The second is to target the young voters. The turnout rates in young people are much lower than the old people. Motivating young people to vote can dramatically increase turnout rates. The third one is to focus on the low-education groups. The people with low education level have low turnout rates. Sensitizing them to the importance of voting might be one way to increase turnout rates.

However, there are some limitations in the project. The first one is the omitted variable bias. There would be other variables correlated with turnout rates and the above three factors, so biased outcome. Besides, Due to the data limitation, age, education cannot be classified as states level.Therefore, interaction effects between states and demographic variables could not be explored.

In order to study this topic in greater depth, the future direction should concentrate on exploring additional variables, using advanced modeling techniques, and expanding outcome measures about political participation. Additional variables would help to address the problem of biased outcome. The advanced modeling techniques, such as nonlinear regression, would show the correlation between these fators and turnout rates. Finally, turnout rates cannot be 100 percent to represent political participation, so the consideration of other measurements is also significant.

In conclusion, we believe that states, age, and education level are three factors affect the turnout rates. Although there are some limitations in the project, our project deserves a further research and more concentration.