Final Project: Conor Jones, Hunter Setzler, Tejasvani Venkatkumar, Tristen Odham

I. INTRODUCTION

The Fifteenth Amendment to the United States Constitution prohibits denying the right to vote based on race or color. Ever since its passage, there have been efforts to undermine the vote of minorities. In the Jim Crow era, these efforts came in the form of intimidation, poll taxes, and literacy tests designed to suppress the black vote. In 1965, after years of activism and protest, Congress passed the Voting Rights Act, which allowed the federal government to enforce the Fifteenth Amendment.

Despite the strides toward equality that these legal and political actions have produced, Americans of color still turn out at lower rates than white Americans. This disparity is attributed by some to policies of voter suppression. Some say that these policies, which include strict Voter ID laws, restrictions on early voting and voting by mail, and fewer polling places, restrict the minority vote. During the 2020 election, those people often pointed toward long lines to vote in majority-minority voting districts as evidence of that kind of racist voter suppression in an attempt to drive minorities away from polls. We hypothesize that we will find that non-white Americans have to wait longer than white Americans to vote. We believe this because we recognize that Americans of color face systemic obstacles in many areas. Additionally, minorities are more likely to live in low-income urban environments that have fewer polling places per capita.

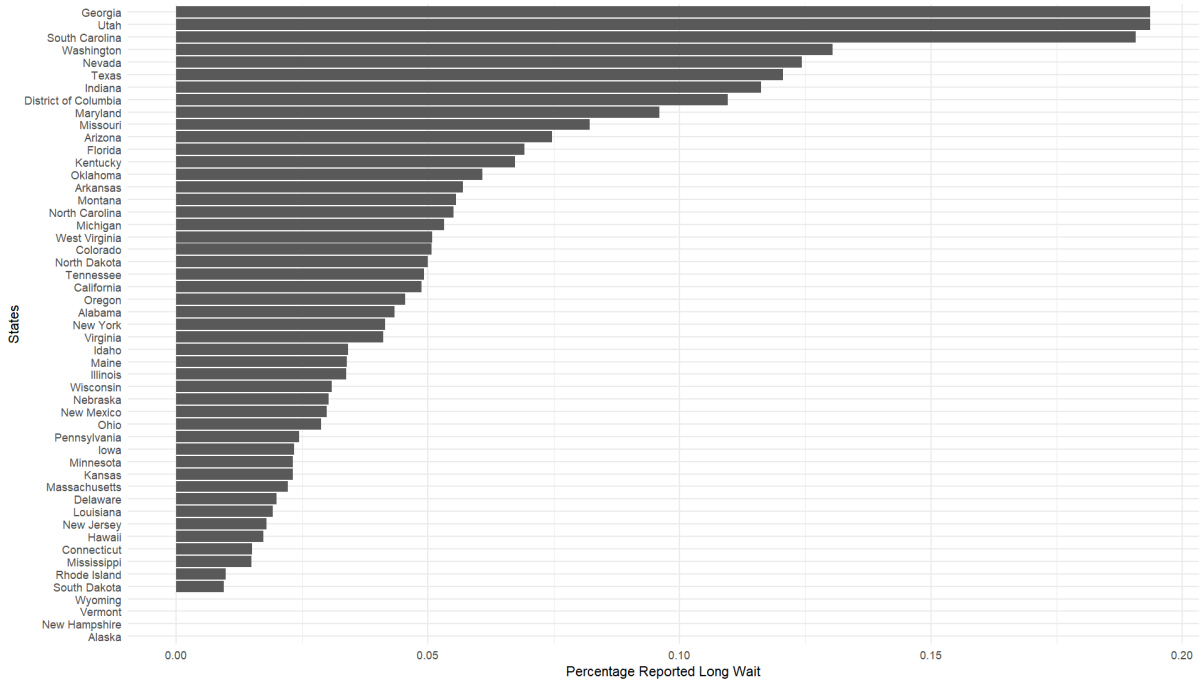
II. HOW LONG DO PEOPLE WAIT?

	No wait time	Fewer than 10 minutes	10-30 minutes	31-60 minutes	Greater than 60 minutes
Number of respondents	13470	10354	5002	1503	385
Percentage of total respondents	43.90%	33.70%	16.30%	4.90%	1.30%

First, we removed the respondents who didn't know how long they waited to vote. The lack of clear responses to the question of our dependent variable made their responses practically worthless for this exercise. Additionally, these respondents only represented 0.13% of our data, meaning we did not significantly alter our sample by removing them. We then decided that 3, 4, and 5 were all going to be classified as long waiting times. We determined that the mean in the data set was 1.8 which shows that the average person waited for less than ten minutes. Although what makes a long wait time is subjective, we considered how much time categories 3,4, and 5 would take out of a working person's day. Despite the broad 10-30 minute range for Category 3, we designated it as a prolonged wait time. This choice was

influenced by the potential for individuals waiting 30 minutes at a polling place to face disruptions in their work schedule, likely exceeding an hour when considering travel time and other factors. Both the mean and median wait times were fewer than 10 minutes. Most of the people in the data set did not have to wait a long time, showing that generally, election infrastructure was strong enough that long wait times were an anomaly. However, there were enough of them to provide a considerable sample size to consider if they were disproportionately experienced by people of color.

III. LONG WAITS BY STATE

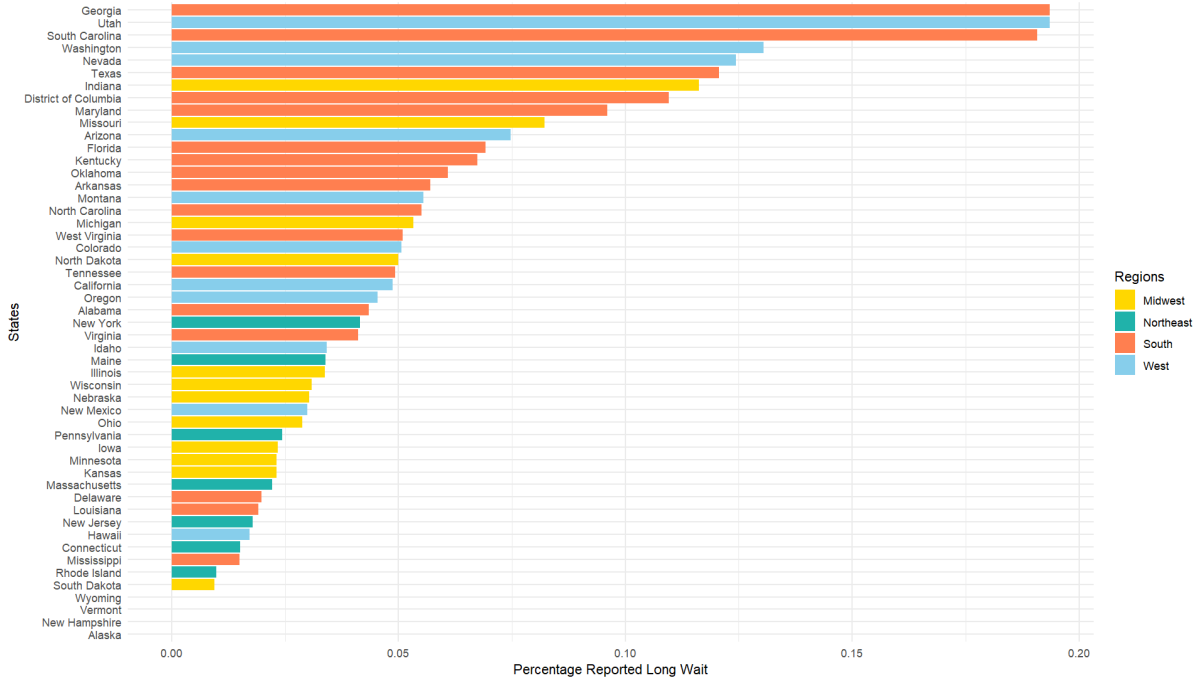


The depicted bar graph illustrates that the proportion of individuals experiencing prolonged wait times ranges from 0% to 18%. Notably, there is a significant disparity in reported percentages after the initial three states, with the prevalence of long wait times declining from 17-18% to below 15%. There does not seem to be an explicit pattern on which states have long or short wait times. We compared the data gathered in this code to state populations and racial populations by state. There is not a strong correlation. In an ideal scenario, where voting access and polling place distribution aligns with state populations, one would anticipate shorter wait times in more populous states and longer wait times in less populous ones. Despite this expectation, a strict pattern is not evident. According to the US Census, the least populated states include Wyoming, Vermont, New Hampshire, Alaska, North and South Dakota, Delaware, Rhode Island, Maine, and Montana. Interestingly, a concentration of these least populated states is observed among the bottom five states in the graph, which reported minimal to no long wait times.

When examining racial data, the top five states with the highest White populations—Maine, West Virginia, Vermont, New Hampshire, and Montana—include some that report minimal to no wait times. Despite these observations, establishing a clear pattern between population

size, racial composition, and the incidence of long wait times remains challenging based on the presented graph.

IV. WAITING TIMES BY STATE AND REGION



After incorporating color codes and regional distinctions into the bar graph, a trend emerges, highlighting the dominance of the South in terms of reported longer wait times. The visual representation facilitated by colors reveals that 17 Southern states reported extended wait times, surpassing 12 in the Midwest, 7 in the Northeast, and 11 in the West. This color-coded presentation enhances the clarity of these regional variations. Among the top five states with the lengthiest waiting times, three belong to the West, while two are from the South, aligning with the broader data indicating that Southern and Western states more frequently reported prolonged wait times. In contrast to our non-color-coded graph, this version significantly simplifies the identification of patterns. While acknowledging the absence of a strict pattern, the graph displays a higher frequency of Southern states reporting extended wait times.

V. WAITING TIMES BY PRIOR VOTE

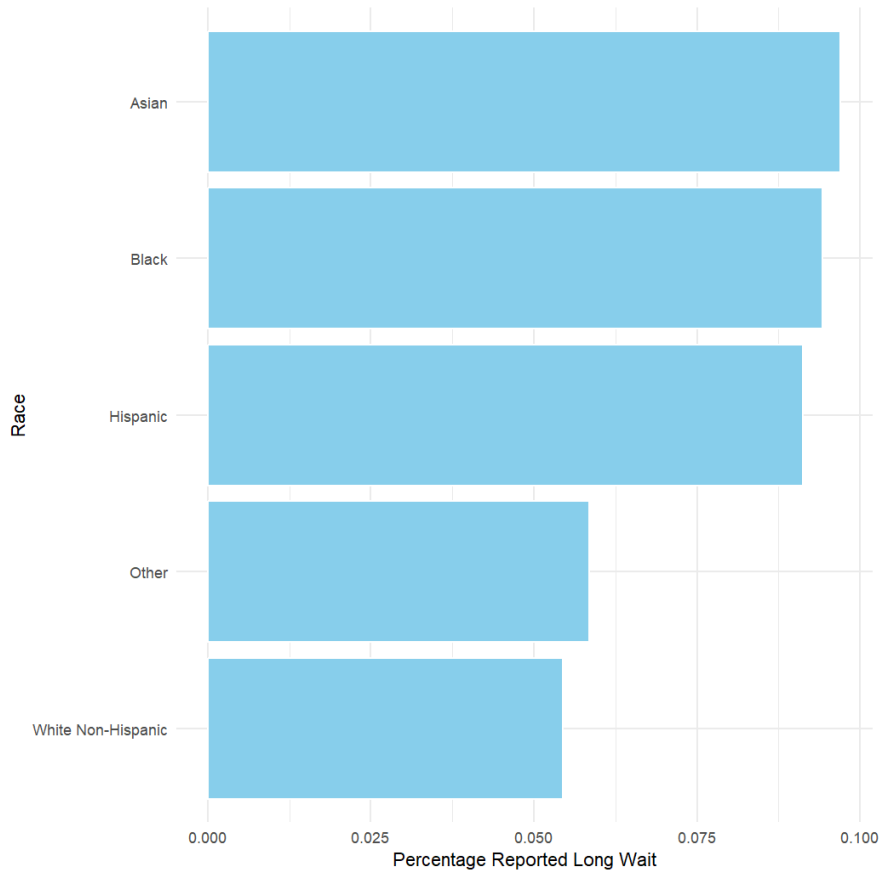
First, we removed voters who voted for a third-party candidate, no candidate at all, or couldn't remember the candidate they voted for. Although there was high third-party turnout in 2016, it is hard to translate those votes directly into ideology, especially since the data doesn't say if the voters voted for a more liberal third-party candidate, such as Jill Stein, or a more conservative one, such as Gary Johnson. We determined that voting for Donald Trump was indicative of being a conservative and voting for Hillary Clinton was indicative of being a liberal. Although these indicators are not perfect predictors of ideology we determined they

were strong enough to use. The data shows that liberals were more likely to wait in longer lines than conservatives.

	Long wait time %
Liberal	3.5
Conservative	2.4
	5.9

The data shows that liberals were more likely to wait in longer lines than conservatives. This discrepancy suggests that there may be factors influencing voting accessibility that vary between political affiliations. Possible factors could be higher voter turnout in liberal-leaning areas, inadequate polling resources, or systemic issues affecting certain demographics. Policy should aim to make this process more efficient and more equal among the two groups.

VI. WAITING TIMES BY RACE



The generated bar chart illustrates the percentage of individuals reporting extended wait times during voting, categorized by race. This visual representation allows us to examine the potential correlation between race and voting wait times. Notably, White Non-Hispanic

respondents reported a lower proportion of long wait times compared to individuals of Asian, Black, Hispanic, and Other races. The graph effectively supported our hypothesis, indicating that minority groups generally experience lengthier voting wait times. While combining Native American, Mixed, Middle Eastern, and Other categories may have reduced the specificity of representation for these groups, it had minimal impact on our analysis, as these groups had less data compared to White, Black, Hispanic, and Asian categories. Reintroducing these combined groups would not have significantly altered the overall observation of lower reported wait times among White non-Hispanics compared to other racial groups.

VII. WAITING TIMES BY INCOME

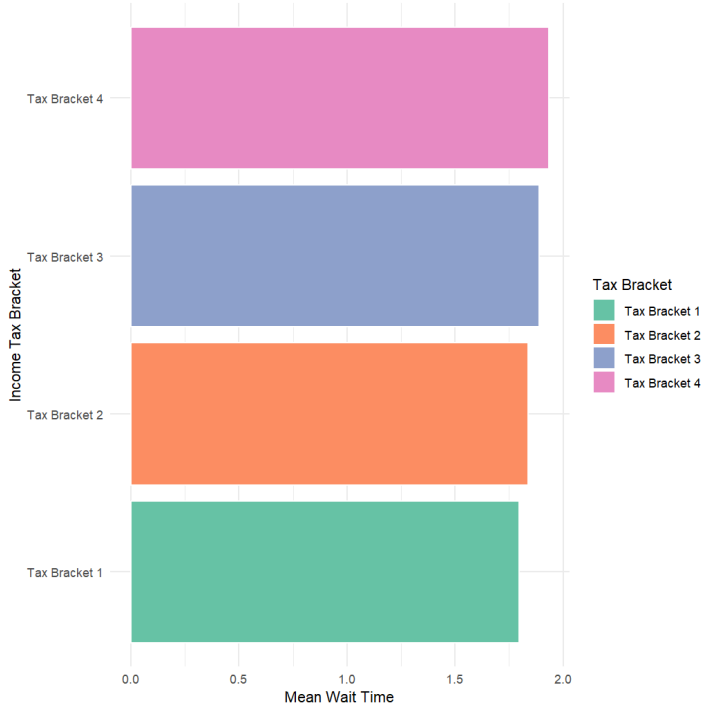
For income to be a confounder when looking at the effect of race on wait time, there must be an association between the confounding variable (income) and the independent variable (race). In this context, it implies that there is a link between income levels and the racial groups being considered. Secondly, the confounding variable (income) should also be associated with the dependent variable (voting wait times), independently of the main variable (race). This means that income levels should independently influence reported wait times, regardless of racial identity. If income is a confounder and is not properly controlled for, it could introduce bias in the initial analysis by attributing observed differences in voting wait times solely to race, when income may be a contributing factor. This bias may either exaggerate or underestimate the impact of race on wait times. Before conducting an analysis, the possibility of income being a confounder is likely. Since our hypothesis focuses on comparing White Non-Hispanics to all racial groups, there will be inherent differences between the racial groups we combine to be minorities. It is very possible that income could be one of these differences.

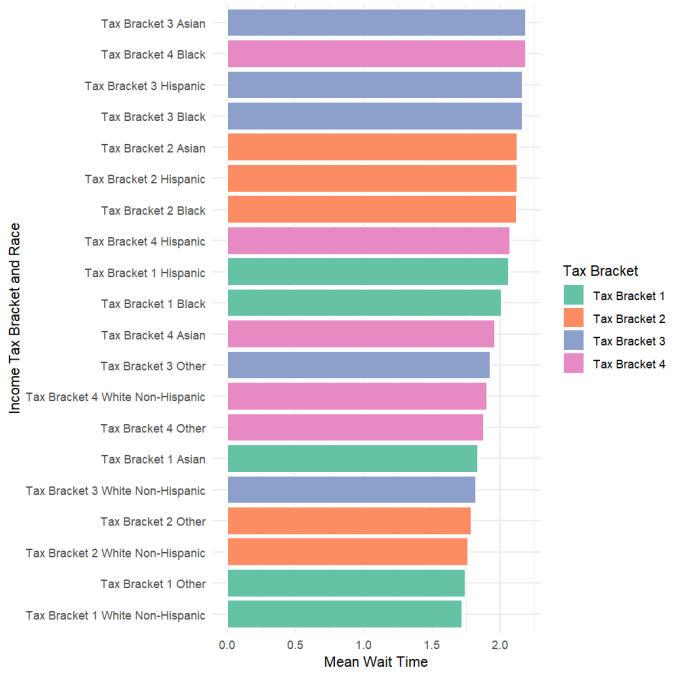
We looked at outside sources for our coding decision regarding faminc. Our categorization of tax brackets aligns with the tax rate schedule provided by the IRS for married filing jointly status individuals, since that was the most representative of a family household income. We also looked at the distribution of household incomes in the United States in 2022 for some more insight on our tax brackets. We excluded individuals who chose not to disclose their income, ensuring a comprehensive and representative analysis. The data presented in our table reflects the distribution of individuals across various tax brackets within our dataset. Upon examining the table, it was revealed that a significant majority of respondents during the voting period fell within Tax Bracket 3. This observation prompted us to delve deeper into the demographics, leading us to explore the racial composition within each tax bracket. This additional analysis seeks to uncover potential correlations or patterns between income levels and racial representation.

Tax Bracket 1	Tax Bracket 2	Tax Bracket 3	Tax Bracket 4
9.004414	29.650532	53.15462	8.190435

Tax Bracket	Race	Percentage	Tax Bracket	Race	Percentage
Tax Bracket 3	White Non-Hispanic	42.1243036	Tax Bracket 1	Black	1.50495623
Tax Bracket 2	White Non-Hispanic	22.75522755	Tax Bracket 3	Asian	1.15042327
Tax Bracket 4	White Non-Hispanic	6.66377252	Tax Bracket 2	Other	1.11786412
Tax Bracket 1	White Non-Hispanic	6.33818103	Tax Bracket 1	Hispanic	0.61138847
Tax Bracket 3	Black	4.33398452	Tax Bracket 2	Asian	0.47753419
Tax Bracket 3	Hispanic	3.68280153	Tax Bracket 1	Other	0.45944577
Tax Bracket 2	Black	3.38976919	Tax Bracket 4	Black	0.45582809
Tax Bracket 2	Hispanic	1.91013675	Tax Bracket 4	Hispanic	0.43773967
Tax Bracket 3	Other	1.86310687	Tax Bracket 4	Other	0.34006222
Tax Bracket 1	Black	1.50495623	Tax Bracket 4	Asian	0.29303234

VIII. EXAMINING INCOME



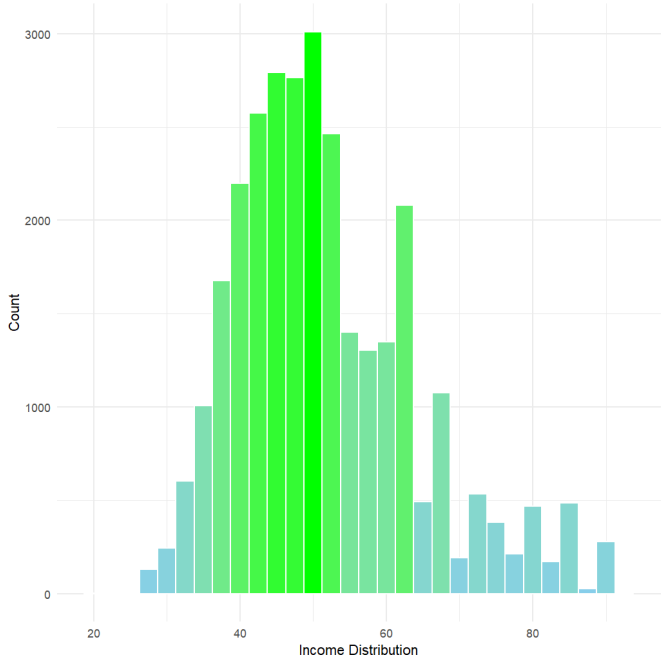


We opted to generate figures to assess whether income could potentially act as a confounding variable in our analysis for two primary reasons. Firstly, income is a factor that can influence political attitudes and participation. By examining income distribution across tax brackets, we aimed to discern whether income variations could confound the relationship between voting behavior and demographic characteristics. Secondly, as income often correlates with disparities in access to resources and opportunities, understanding its potential confounding effect is important to find the interplay between income, demographics, and voting patterns. We chose to make graphs that showed the mean wait time by tax bracket and one that showed the mean wait time by race and tax bracket. Upon analyzing the figures derived from our dataset, certain trends have emerged. While a substantial number of respondents fall within Tax Bracket 3, we observed a diverse distribution of racial backgrounds within this bracket and across others. This suggests that while income plays a role in shaping political engagement, its impact does not disproportionately favor or disadvantage a particular racial group. Consequently, our concern regarding income acting as a confounder has diminished since the figures indicate a more complex relationship between income, demographics, and voting behavior. While the relationship is more complex, having a lower income may make a long wait more impactful. As the adage says, time is money, and so lower-class Americans may have less to spare.

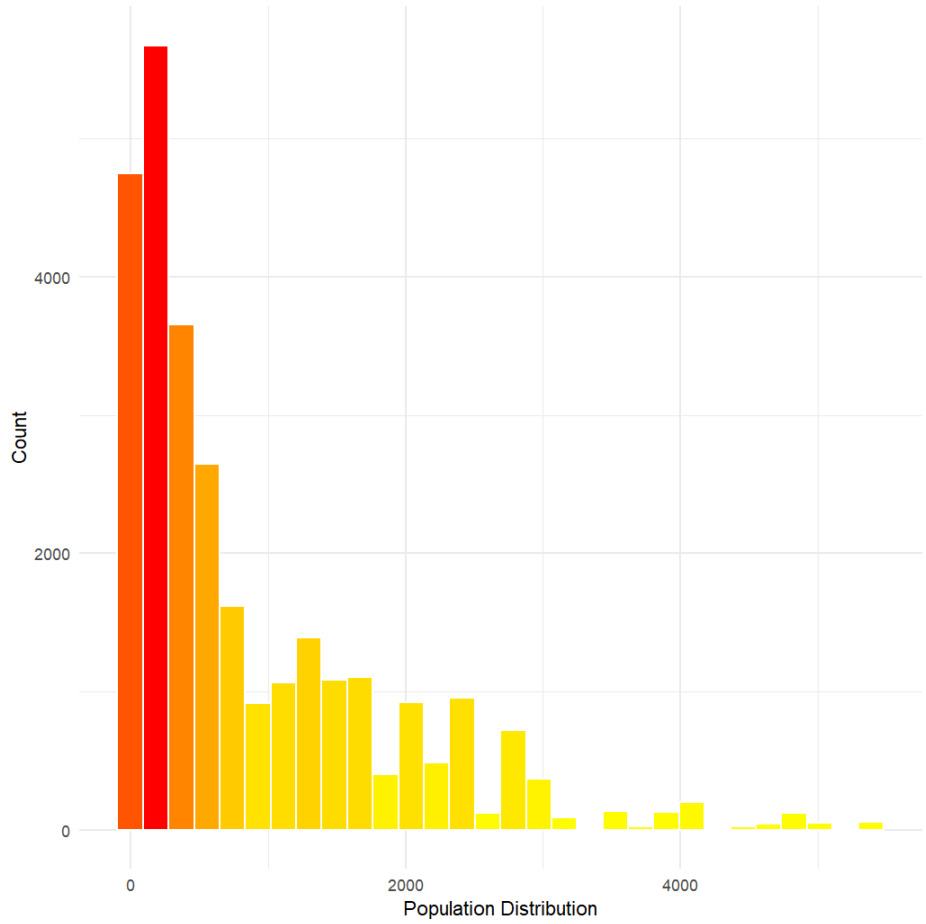
IX. USING REGRESSION TO ACCOUNT FOR MULTIPLE CONFOUNDERS – PART 1

County socioeconomic status and population density can be considered potential confounding factors in the relationship between race and voting wait times due to their influence on the social and environmental contexts in which individuals cast their votes. County socioeconomic status, reflecting the economic prosperity of the area, may impact the

availability and allocation of resources dedicated to the voting process. Higher socioeconomic status could potentially lead to better-funded and more efficiently managed polling stations, resulting in shorter wait times. Lower socioeconomic status areas might face resource constraints, contributing to longer wait times. Secondly, county population density may also play a role in the race/wait time relationship. High-density areas might experience greater demand for voting resources, potentially leading to longer wait times due to higher voter turnout. Low-density areas may face challenges in managing voting polls. Both county socioeconomic status and population density could explain variations in voting wait times that are independent of racial identity, potentially confounding the observed relationship between race and wait times if not appropriately controlled for.



Upon examining the distribution of the county socioeconomic status variable (income_county) through a histogram, it became apparent that there are extreme values that may be considered outliers. To address this, we chose to put a "ceiling" on the variable. Specifically, we decided to recode observations with very high values by assigning them the income value associated with the 95th percentile, getting rid of the top 5% percent of respondents. This decision was made to mitigate the influence of outliers on our analysis. We chose option two over option one because removing observations identified as outliers (option 1) can reduce the sample size. Keeping the observations but capping extreme values at a reasonable percentile ensures that the sample size remains relatively intact, which moderates its impact on the overall distribution while still considering their contribution to the dataset.



For the density variable, we filtered out any data that was above the 95th percentile because we were concerned that that data could skew our findings. In the context of the density variable, we used the 95th percentile to focus on the majority of cases while mitigating the influence of counties with exceptionally high population densities. We made this decision to get a representative and stable estimate of population density and its distribution across counties.

X. USING REGRESSION TO ACCOUNT FOR MULTIPLE CONFOUNDERS – PART 2

Residuals:

Min	1Q	Median	3Q	Max
-1.1347	-0.8061	0.1391	0.1939	4.1939

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.806107	0.006094	296.352	< 2e-16 ***
black	0.328549	0.018684	17.585	< 2e-16 ***
Hispanic	0.319545	0.022462	14.226	< 2e-16 ***
asian	0.315205	0.038818	8.120	4.83e-16 ***
other	0.054799	0.027839	1.968	0.049 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9468 on 30749 degrees of freedom
(6 observations deleted due to missingness)
Multiple R-squared: 0.01652, Adjusted R-squared: 0.01639
F-statistic: 129.1 on 4 and 30749 DF, p-value: < 2.2e-16

Residuals:

Min	1Q	Median	3Q	Max
-1.5295	-0.7860	-0.0179	0.2817	4.3126

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.854e+00	2.481e-02	74.732	< 2e-16	***
black	3.026e-01	2.115e-02	14.310	< 2e-16	***
hispanic	2.949e-01	2.442e-02	12.075	< 2e-16	***
asian	3.008e-01	4.504e-02	6.678	2.48e-11	***
other	3.579e-02	3.077e-02	1.163	0.245	
faminc_reg	1.304e-02	1.855e-03	7.028	2.14e-12	***
income_county	-4.114e-03	4.800e-04	-8.570	< 2e-16	***
density954	9.520e-05	6.948e-06	13.702	< 2e-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9405 on 25858 degrees of freedom
(4894 observations deleted due to missingness)
Multiple R-squared: 0.02584, Adjusted R-squared: 0.02557
F-statistic: 97.98 on 7 and 25858 DF, p-value: < 2.2e-16

In both Model 1 and Model 2, the intercept term represents the predicted wait time for whites. Since there is no explicit variable for whites, the intercept captures the baseline wait time for this group. The coefficients for the race categories (Black, Hispanic, Asian, Other) in both models represent differences in wait times compared to whites. Since the coefficients are positive, it suggests longer wait times for the respective racial group compared to whites. This affirms that racial minorities experience longer wait times compared to whites, confirming our initial hypothesis. In Model 2, three control variables are introduced: personal income, county socioeconomic status, and population density. These control variables aim to capture additional factors influencing wait times beyond race. The positive coefficient for personal income suggests that higher income is associated with increased wait times. This did not coincide with our hypothesis that lower-income groups face longer wait times. The negative coefficient for county socioeconomic status indicates that living in a wealthier county is associated with shorter wait times. The positive coefficient for population density implies that higher population density is associated with longer wait times. Comparing the coefficients for race variables between Model 1 and Model 2 reveals changes in the values. The decrease in the coefficients for blacks, Hispanics, Asians, and others from Model 1 to

Model 2 suggests that the control variables are influencing the relationships. This indicates that personal income, county socioeconomic status, and population density contribute to disparities in wait times. Most of the data from this table affirms our initial hypotheses, the only thing that is off is the table's indication that a higher income indicates a longer wait time.

XI. CONCLUSION

In conclusion, as a result of our analysis, we can reject our null hypothesis that there is no race-based impact on the wait time to vote and accept our alternative hypothesis that non-white Americans have to wait longer than white Americans to vote. This finding shows that despite the progress toward racial equality we have made as a nation, more must be done to become a state that embraces the value of democracy and moves past its racist history. This analysis is limited in scope by only focusing on a single year and a single survey. A more complete analysis would be based on more data that would allow the researcher to map the change over time in these trends. Additionally, the analysis could go further in taking note of how race impacts various aspects of life. There may have been Americans who were not included in this survey because they did not go vote because they simply had no time in their day or lacked the identification required to vote in their state.

To help remedy the racial disparities that we found through our research, we suggest a number of policy changes. One easy policy change that would help would be making election day a federal holiday. Even if this wouldn't necessarily fix long wait times, it would make them less disruptive in the lives of federal employees and encourage employers to give the holiday off. A more direct solution to the issue of long wait times would be to increase the number of polling places, especially in areas that have suffered long wait times in the past. Additionally, increasing the availability of early and mail-in voting will make turnout easier, especially for working-class Americans.

Our insights deepened with this project, especially regarding the relationship between the income variable and the race variable. We assumed that income would be a causal mechanism explaining how race impacted voting wait times. The regression we ran showed, however, that the relationship is more complex. Additionally, we learned about how geography impacts voting wait times. The concentration of states with long wait times in the South shows that much of the work to be done is right in our backyard and that we can make a big difference on the issue within our local communities.

Works Cited

"Population Distribution by Race/Ethnicity." *KFF*, 27 Oct. 2023, www.kff.org/other/state-indicator/distribution-by-raceethnicity/?dataView=0&currentTimeframe=0&selectedDistributions=white&sortModel=%7B%22colId%22%3A%22White%22%2C%22sort%22%3A%22desc%22%7D.

Published by Statista Research Department, and Nov 3. "Distribution of Household Income U.S. 2022." *Statista*, 3 Nov. 2023, www.statista.com/statistics/203183/percentage-distribution-of-household-income-in-the-us/.

Washington, Kemberley. "2023-2024 Tax Brackets and Federal Income Tax Rates." *Forbes*, Forbes Magazine, 10 Nov. 2023, www.forbes.com/advisor/taxes/taxes-federal-income-tax-bracket/.

Where Do I Fall in the American Economic Class System?, money.usnews.com/money/personal-finance/family-finance/articles/where-do-i-fall-in-the-american-economic-class-system. Accessed 11 Dec. 2023.