The Impact of Boston University Student Backgrounds on Voting Habits

Marcos Sasson, Enoch Ngan, Gui Marques, Annabella Calle, Austin Gerrard

This study examines the factors influencing voting behavior among Boston University (BU) undergraduate students, specifically investigating how personal backgrounds and academic focus might impact students' likelihood to vote. Voting participation among young adults has consistently lagged behind older age groups, even though young adults significantly shape future policies[1]. Prior research by scholars such as Niemi and Hanmer suggests that traditional predictors of voting, such as age, income, or education, do not fully account for the unique challenges college students face, including living away from home and balancing academic and social obligations[2]. Our analysis is based on survey data from BU undergraduates and gathers data on students from varying swing states, socioeconomic backgrounds, parental voting habits, and majors vote differently. Using demographic and qualitative data, this report aims to clarify voting patterns among BU students. More specifically, our hypotheses consider whether a student from a swing state, a student with parents in blue-collar jobs, a student with an age over 20, and a student with a major in social science have significant voting habits.

Introduction

College students are often viewed as less engaged in voting, but there's a lot to learn about what actually influences students to participate in elections. This study focuses on understanding what might drive or discourage BU undergraduates from voting by collecting data on their backgrounds. Past research suggests that factors like personal beliefs, social influences, and even specific demographic traits can affect voting decisions. By analyzing these areas among BU students, key trends are uncovered that show what makes some students more likely to vote than others.

Demographic details were the primary point of interest in our survey, providing information into BU students' motivations and barriers for voting. An analysis of these responses provides insights on ways to increase college student turnout for future elections. Therefore, the following main questions were used to identify the influences that sway a BU students' choice to vote.

Independent Variables: Demographic Questions	**Dependent Variable:** Qualitative Question
1) **Age**: How old are you? 2) **Gender**: What is your gender? 3) **College**: Which college/school is your major in? 4) **Swing State**: Which U.S. state are you	1) **Vote**: Are you going to vote in the 2024 presidential election?

[1]File, Thom. 2021. "Voting in America: A Look at the 2016 Presidential Election." *Census.Gov*. October 8. https://www.census.gov/newsroom/blogs/random-samplings/2017/05/voting_in_america.html.

[2] Niemi, Richard G., and Michael J. Hanmer. 2010. *Voter Turnout among College Students: New Data and a Rethinking of Traditional Theories | Request PDF*. June. https://www.researchgate.net/publication/227377195_Voter_Turnout_Among_College_Students_New_Data_and_a_Rethinking_of_Traditional_Theories.

Experiment Design: M.S., E.N., G.M., A.C., A.G Data Cleaning: M.S., A.C, A.G , Modeling and Analysis: A.C., A.G, E.N. Figures and Plots: A.C., Writing: M.S., E.N., G.M.

from? Are you from a swing state?

5) **Socioeconomic Background**: How are you funding your BU tuition?

6) **Parents' Voting Pattern**: Did your parent(s) vote in the 2020 presidential election?

The following null hypotheses were created from these main questions:

Null Hypotheses	**Statement**
Null Hypothesis 1	Students are equally likely to vote regardless if they are from swing states.
Null Hypothesis 2	Students are equally likely to vote regardless if their parents voted in 2020.
Null Hypothesis 3	Students are equally likely to vote regardless if they are over 20.
Null Hypothesis 4	Students are equally likely to vote regardless if their major is social sciences.

Alternative hypotheses were included in this analysis to allow for the exploration of contrary trends that may arise from unexpected data patterns. While the original hypotheses are based on anticipated behaviors and established patterns, alternative hypotheses provide a framework for identifying and interpreting deviations from these expectations. This approach ensures a comprehensive analysis, capturing a broader range of influences on BU students' voting behaviors and highlighting potential nuances or contradictions in the data.

Alternative Hypotheses	**Statement**
Alternative Hypothesis 1	Students from swing states are more likely to vote than those from non-swing states.
Alternative Hypothesis 2	Students whose parents voted in 2020 are more likely to vote than those whose parents did not vote in 2020.
Alternative Hypothesis 3	Students under the age of 20 are more likely to vote than those over the age of 20.
Alternative Hypothesis 4	Students studying social sciences are more likely to vote than students in other fields.

Experimental Methods

This study used an observational approach, collecting data on whether BU undergraduates planned to vote in the 2024 presidential election and their demographics without manipulating variables. The design focused on demographics potentially influencing voting willingness: age, college, major, statehood, socioeconomic background, and parental voting patterns. Leading questions were avoided to reduce response bias, and the survey began by asking if the respondent intended to vote, followed by straightforward demographic and background questions derived from the original variables of interest.

The population of interest was American BU students over 18. Potential biases included nonresponse and social desirability bias, but the most significant issue was sampling bias. The pilot study showed that reliance on social media platforms, like Snapchat and Instagram, skewed the sample. So, to mitigate this, more data was gathered in person and through broader public platforms to better represent the target population. Redundancy in questions also risked respondent fatigue. In the pilot, multiple questions overlapped on similar factors (e.g., age, college, major, state, swing state status). This overlap could have uninterested participants and so consequently, the final survey streamlined these questions to maintain respondent interest and data quality.

Another challenge involved how funding sources were classified. The pilot survey did not distinguish among types of scholarships (need-based, athletic, academic merit), limiting the ability to analyze correlations between financial aid type and voting behavior. The final survey addressed this by providing more detailed funding source options, thus enhancing the precision of socioeconomic analysis.

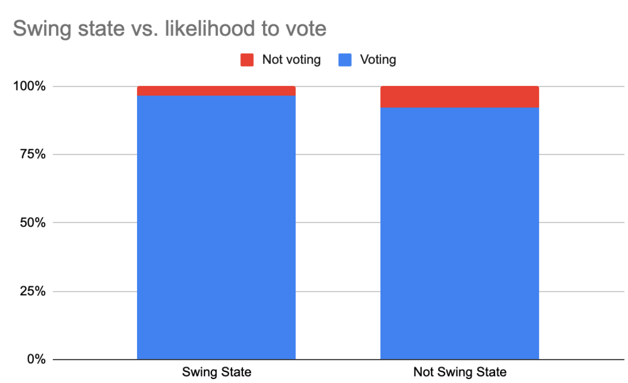
Despite these adjustments, achieving the target of 500 responses proved difficult as only 186 responses (N=186) were collected, likely due to the limited initial outreach methods. Early efforts relied too heavily on personal social media accounts and word of mouth, giving few responses. To broaden outreach, the survey was distributed via BU's public Snapchat story, Facebook groups, and through BU clubs and student organizations and although this increased diversity and engagement, it still fell short of the desired sample size.

In summary, the study's methodology evolved from a pilot phase to a refined final survey, aiming to minimize bias, streamline questions, and clarify funding categories. While outreach methods expanded, the final sample size remained lower than targeted, potentially affecting the representativeness and overall robustness of the findings.

Results

95% Confidence Intervals

Swing State For the graphing of this data, the data was corrected for the 13 unsure responses (3 of which were from swing states) and the 5 incorrect responses (1 of which believed they were from a swing state) based on 2020 presidential election data. It is important to note that the 5 incorrect responses do believe that they are or are not from a swing state, and this may impact their voting habits. Overall, the amount of people not from a swing state who are not voting is over double the percentage of those who are from a swing state (3% and 7% respectively).



	Swing State	Not Swing State	Total
Voting	29	144	173
Not voting	1	12	13
Total	30	156	186

H_0: $p_{voting|swing\ state} = p_{voting|not\ swing\ state}$
H_a: $p_{voting|swing\ state} > p_{voting|not\ swing\ state}$

1. Likelihood to vote given that population member is from a swing state:

$\hat{p}$ is the probability that a population member is from a swing state, and $\hat{q}$ is 1-$\hat{p}$.
$n = 30$ but $n\hat{q} < 15$, so the sample size is not large enough to assume the normality of the population proportion.
$\alpha = .05$
$\rightarrow \frac{\alpha}{2} = .025$
$\rightarrow Z_{\frac{\alpha}{2}} = 1.960$
$n\hat{q} < 15$
$\rightarrow \tilde{p} = \frac{x+2}{n+4} = \frac{31}{34} = 0.912$
$\rightarrow \tilde{q} = 0.088$
Small-sample confidence interval using Wilson's adjustment:
$\tilde{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\tilde{p}\tilde{q}}{n+4}} = 0.912 \pm 1.960\sqrt{\frac{0.912\times0.088}{34}} = (0.817, 1.02)$
The upper bound of the confidence interval exceeds 1 which can be attributed to a high $\hat{p}$ and low n value, further showing that normality cannot be assumed making the confidence interval unreliable.

2. Likelihood to vote given that population member is not from a swing state:

$\hat{p}$ is the probability that a population member is not from a swing state, and $\hat{q}$ is 1-$\hat{p}$.
The sample size is large ($n = 156$, $n\hat{p} \geq 15$, $n\hat{q} \geq 15$) and is sufficient to assume the normality of the population proportion.
The true $p_{voting|not\ swing\ state}$ is estimated to lie between 0.881 and 0.965 with 95% confidence.

Parents' Voting Pattern There is a clear connection between the likelihood of voting and those who are certain that both of their parents voted in the 2020 election. With only 1% of those whose parents both voted (which was the largest category of responses for this question) stating that they will not vote, it is clear that having two parents who vote and tell their children that they are voting (or at least are transparent about their voting habits) increases the likelihood of voting for American BU students over the age of 18.

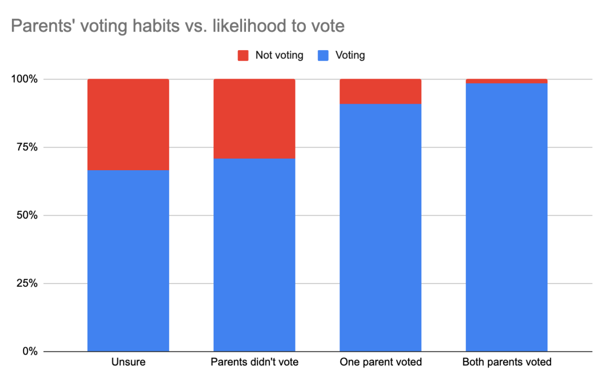


Fig. 1. Parents' Voting Habits vs. Likelihood to Vote

H_0: $p_{voting|at\ least\ 1\ parent\ voted} = p_{voting|no\ parents\ voted}$
H_a: $p_{voting|at\ least\ 1\ parent\ voted} > p_{voting|no\ parents\ voted}$

	Unsure	Parents didn't vote	One parent voted	Both parents voted
Voting	4	17	20	132

	Unsure	Parents didn't vote	One parent voted	Both parents voted
Not voting	2	7	2	2
Total	6	24	22	134

1. Likelihood to vote given that population member is unsure if their parents voted:

p̂ is the probability that a population member is unsure if their parents voted, and q̂ is 1-p̂. $n\hat{q} < 15$ and $n\hat{p} < 15$ so the sample size is not large enough to assume the normality of the population proportion.
The true $p_{\text{voting|unsure}}$ is estimated to lie between 0.903 and 0.765 with 95% confidence.

2. Likelihood to vote given that parents did not vote:

p̂ is the probability that both of a population member's parents did not vote, and q̂ is 1-p̂. $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion.
The true $p_{\text{voting|parents did not vote}}$ is estimated to lie between 0.506 and 0.852 with 95% confidence.

3. Likelihood to vote given that one parent voted:

p̂ is the probability that one of a population member's parents did vote, and q̂ is 1-p̂. $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion.
The true $p_{\text{voting|one parent voted}}$ is estimated to lie between 0.707 and 0.985 with 95% confidence.

4. Likelihood to vote given that both parents voted:

p̂ is the probability that both of a population member's parents did vote, and q̂ is 1-p̂. $n = 134$ and $n\hat{p} > 15$ but $n\hat{q} < 15$ so the sample size may not be large enough to assume the normality of the population proportion.
The true $p_{\text{voting|both parents voted}}$ is estimated to lie between 0.9706 and 0.9714 with 95% confidence.

Age Of the results with substantial data counts (18-21), the trend was that the older the student, the less likely they are to vote, with 20 and 21 year olds having a higher percentage of non-voters compared to 18 and 19 year olds. Two thirds of 22 year olds and both the single 23 and the single 24+ students stated they are voting, but the data size for these groups was not substantial enough to make any claims.

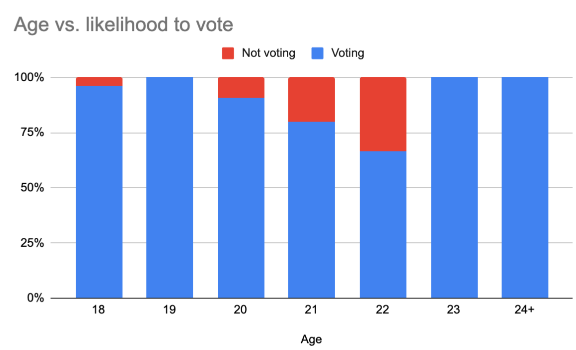


Fig. 2. Age vs. Likelihood to Vote

H_0: $p_{\text{voting|20 or under}} = p_{\text{voting|21+}}$
H_a: $p_{\text{voting|20 or under}} > p_{\text{voting|21+}}$

Age	18	19	20	21	22	23	24+
Voting	72	59	30	12	2	1	1
Not voting	3	0	3	3	1	0	0
Total	75	59	33	15	3	1	1

<u>95% confidence intervals</u>

1. Likelihood to vote given that population member is 18 years old

p̂ is the probability that a population member is 18 years old, and q̂ is 1-p̂. $n\hat{p} > 15$ but $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion.
The true $p_{\text{voting}|18}$ is estimated to lie between 0.935 and 0.938 with 95% confidence.

2. Likelihood to vote given that population member is 19 years old

p̂ is the probability that a population member is 19 years old, and q̂ is 1-p̂. $n\hat{p} > 15$ but $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion.
The true $p_{\text{voting}|19}$ is estimated to lie between 0.967 and 0.969 with 95% confidence.

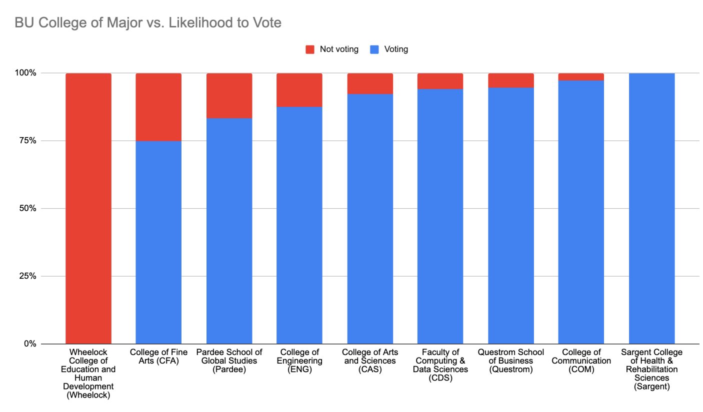
3. Likelihood to vote given that population member is 20 years old

p̂ is the probability that a population member is 20 years old, and q̂ is 1-p̂. $n\hat{p} > 15$ but $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion.
The true $p_{\text{voting}|20}$ is estimated to lie between 0.859 and 0.871 with 95% confidence.

4. Likelihood to vote given that population member is 21 years old

p̂ is the probability that a population member is 21 years old, and q̂ is 1-p̂. $n\hat{p} < 15$ and $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion.
The true $p_{\text{voting}|21}$ is estimated to lie between 0.717 and 0.757 with 95% confidence.

College Within Boston University Of the 17[3] schools and Colleges at Boston University, students from 9 of them responded. Of those 9, the most responses were collected from ENG (College of Engineering), CAS (College of Arts and Sciences), CDS (Faculty of Computing & Data Sciences), Questrom (Questrom School of Business), COM (College of Communication), and Sargent (Sargent College College of Health & Rehabilitation Sciences). The results from Pardee (Frederick S. Pardee School of Global Studies), CFA (College of Fine Arts), and Wheelock (Wheelock College of Education & Human Development) were lesser in numbers and did not have enough data to justify claims. COM and CAS majors were the highest groups, which fits with these being the highest listed and most popular colleges at BU[4]. Of the colleges, Engineering students were by far the least likely to vote, and COM students were the most likely, with only one of the 38 COM responses stating that they will not vote.



H_0: $p_{\text{voting}|\text{Social Sciences}} = p_{\text{voting}|\text{Other Majors}}$
H_a: $p_{\text{voting}|\text{Social Sciences}} > p_{\text{voting}|\text{Other Majors}}$

Fig. 3. BU College of Major vs. Likelihood to Vote

[3] “Bu Facts & Stats.” *BU Facts & Stats | Office of the President*, www.bu.edu/president/boston-university-facts-stats/. Accessed 31 Oct. 2024.

[4] “About Our Schools & Colleges.” *About Our Schools & Colleges | Admissions*, www.bu.edu/admissions/why-bu/academics/schools-and-colleges/#PAR. Accessed 31 Oct. 2024.

BU College	Wheelock	CFA	Pardee	ENG	CAS	CDS	Questrom	COM	Sargent
Voting	0	3	5	14	48	16	18	37	19
Not voting	1	1	1	2	4	1	1	1	0
Total	1	4	6	16	52	17	19	38	19

1. Likelihood to vote given that population member is a student of Wheelock:

$\hat{p}$ is the probability that a population member is a student of Wheelock, and $\hat{q}$ is 1-$\hat{p}$. Since the sample size is so small (only 1 person), a valid confidence interval cannot be calculated, as it would require a larger sample size for reliable estimation.

2. Likelihood to vote given that population member is a student of CFA:

$\hat{p}$ is the probability that a population member is a student of CFA, and $\hat{q}$ is 1-$\hat{p}$. $n\hat{p} < 15$ and $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion. The 95% confidence interval $(0.450, 1.050)$ is unreliable as its upper bound exceeds 1 which can be attributed to a high $\hat{p}$ and low n value, further showing that normality cannot be assumed making the confidence interval unreliable.

3. Likelihood to vote given that population member is a student of Pardee:

$\hat{p}$ is the probability that a population member is a student of Pardee, and $\hat{q}$ is 1-$\hat{p}$. $n\hat{p} < 15$ and $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion. The 95% confidence interval $(0.602, 1.064)$ is unreliable as its upper bound exceeds 1 which can be attributed to a high $\hat{p}$ and low n value, further showing that normality cannot be assumed.

4. Likelihood to vote given that population member is a student of ENG:

$\hat{p}$ is the probability that a population member is a student of ENG, and $\hat{q}$ is 1-$\hat{p}$. $n\hat{p} < 15$ and $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion. The upper bound of the 95% confidence interval $(0.730, 1.020)$ exceeds 1 which can be attributed to a high $\hat{p}$ and low n value, further showing that normality cannot be assumed making the confidence interval unreliable.

5. Likelihood to vote given that population member is a student of CAS:

$\hat{p}$ is the probability that a population member is a student of CAS, and $\hat{q}$ is 1-$\hat{p}$. $n\hat{p} > 15$ but $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion.
The true $p_{\text{voting|CAS}}$ is estimated to lie between 0.8507 and 0.9955 with 95% confidence.

6. Likelihood to vote given that population member is a student of CDS

$\hat{p}$ is the probability that a population member is a student of CDS, and $\hat{q}$ is 1-$\hat{p}$. $n\hat{p} > 15$ but $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion. The upper bound of the 95% confidence interval $(0.840, 1.042)$ exceeds 1 which can be attributed to a high $\hat{p}$ and low n value, further showing that normality cannot be assumed making the confidence interval unreliable.

7. Likelihood to vote given that population member is a student of Questrom:

$\hat{p}$ is the probability that a population member is a student of Questrom, and $\hat{q}$ is 1-$\hat{p}$. $n\hat{p} > 15$ but $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion. The upper bound of the 95% confidence interval $(0.855, 1.039)$ exceeds 1 which can be attributed to a high $\hat{p}$ and low n value, further showing that normality cannot be assumed making the confidence interval unreliable.

8. Likelihood to vote given that population member is a student of COM:

$\hat{p}$ is the probability that a population member is a student of COM, and $\hat{q}$ is 1-$\hat{p}$. $n\hat{p} > 15$ but $n\hat{q} < 15$ so the sample size is not large enough to assume the normality of the population proportion. The upper bound of the 95% confidence interval $(0.926, 1.022)$ exceeds 1 which can be attributed to a high $\hat{p}$ and low $\hat{q}$ value, further showing that normality cannot be assumed making the confidence interval unreliable.

9. Likelihood to vote given that the population member is a student of Sargent:

Because the proportion of those voting is 19/19, which equals 1, the confidence interval will just be (1,1)

From these two hypotheses, further methods were attempted to reject or fail to reject the null hypothesis.

Next Steps: Fisher's Exact Test

Normality could not be assumed for any of our four hypotheses, therefore a Fisher's Exact Test had to be used to compare the p-value of our Swing state and Parent's Voting Pattern data to our significance level. Two assumptions were made before proceeding with the Fisher's Exact Test: data was independently sampled, one or more values of [a, b, c, d] are less than 5 (these variables will be defined below).

Swing States

Null and alternative hypotheses for the swing states dataset:

H_0: $p_{\text{voting|swing state}} = p_{\text{voting|not swing state}}$

H_a: $p_{\text{voting|swing state}} > p_{\text{voting|not swing state}}$

Rejection Region: α=0.05 significance level (One-tailed)

Probability of observing Extreme Table 1:

$$P_1 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^{N}} = \frac{C_{29}^{30} * C_{144}^{156}}{C_{173}^{186}} \approx 0.2528$$

Probability of observing Extreme Table 2:

$$P_2 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^{N}} = \frac{C_{30}^{30} * C_{143}^{156}}{C_{173}^{186}} \approx 0.0933$$

Total Probability of observing results as extreme as the observed values under the null hypothesis:

$P = P_1 + P_2 \approx 0.3462$

Conclusion We fail to reject the null hypothesis since there is insufficient evidence (p-value = 0.3462 > α = 0.05). Therefore, there is insufficient evidence that students from swing states are more likely to vote than those who are not from swing states.

Parent's Voting Patterns

Null and alternative hypotheses for parent's voting patterns:

H_0: $p_{\text{voting|at least 1 parent voted}} = p_{\text{voting|no parents voted}}$

H_a: $p_{\text{voting|at least 1 parent voted}} > p_{\text{voting|no parents voted}}$

Rejection Region: α=0.05 significance level (One-tailed)

Probability of observing Extreme Table 1:

$$P_1 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^{N}} = \frac{C_{17}^{24} * C_{152}^{156}}{C_{169}^{180}} \approx 6.969 * 10^{-5}$$

Using the same methodology and formula, we calculated the remaining Extreme Table values as such:

$P_2 = 3.872 * 10^{-6}, P_3 = 1.341 * 10^{-7}, P_4 = 2.595 * 10^{-9}, P_5 = 2.117 * 10^{-11}$

Total Probability of observing results as extreme as the observed values under the null hypothesis:

$P = P_1 + P_2 + P_3 + P_4 + P_5 \approx 7.370 * 10^{-5}$

Conclusion We reject the null hypothesis with sufficient evidence (p-value = $7.370 * 10^{-5}$ < α = 0.05) Therefore, we have sufficient evidence to support that the students whose parents voted in the last election are more likely to vote in the 2024 presidential election.

Age

Null and alternative hypotheses for age:

H_0: $p_{voting|20 \text{ or under}} = p_{voting|21+}$

H_a: $p_{voting|20 \text{ or under}} > p_{voting|21+}$

Rejection Region: α=0.05 significance level (One-tailed)

Probability of observing Extreme Table 1:

$$P_1 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^{N}} = \frac{C_{162}^{167} * C_{15}^{20}}{C_{177}^{187}} \approx 0.0014$$

Using the same methodology and formula, we calculated the remaining Extreme Table values as such:

$P_2 = 1.074 * 10^{-4}, P_3 = 5.2389 * 10^{-6}, P_4 = 1.5479 * 10^{-7}, P_5 = 2.4865 * 10^{-9},$
$P_6 = 1.6378 * 10^{-11}$

Total Probability of observing results as extreme as the observed values under the null hypothesis:

$P = P_1 + P_2 + P_3 + P_4 + P_5 + P_6 \approx 0.0015$

Conclusion We reject the null hypothesis with sufficient evidence (p-value = 0.0015 < α = 0.05) Therefore, we have sufficient evidence to support that the students 20 or under are more likely to vote in the 2024 presidential election.

College Majors

Null and alternative hypotheses for college majors:

H_0: $p_{voting|\text{Social Sciences}} = p_{voting|\text{Other Majors}}$

H_a: $p_{voting|\text{Social Sciences}} > p_{voting|\text{Other Majors}}$

Rejection Region: α=0.05 significance level (One-tailed)

Probability of observing Extreme Table 1:

$$P_1 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^{N}} = \frac{C_{85}^{90} * C_{75}^{82}}{C_{150}^{172}} \approx 5.0268 * 10^{-11}$$

Using the same methodology and formula, we calculated the remaining Extreme Table values as such: $P_2 = 2.740 * 10^{-11}$, $P_3 = 1.0358 * 10^{-11}$, $P_4 = 2.5776 * 10^{-12}$, $P_5 = 3.7914 * 10^{-13}$, $P_6 = 2.4925 * 10^{-14}$

Total Probability of observing results as extreme as the observed values under the null hypothesis:
$P = P_1 + P_2 + P_3 + P_4 + P_5 + P_6 \approx 9.1005 * 10^{-11}$

Conclusion We reject the null hypothesis with sufficient evidence (p-value = $9.1005 * 10^{-11} < \alpha = 0.05$) Therefore, we have sufficient evidence to support that the students in social science colleges are more likely to vote in the 2024 presidential election.

Discussion

This study aimed to understand how demographics and background factors influence BU undergraduates' voting habits for the 2024 presidential election. It explored parental voting behavior, state of origin, age, and college affiliation but only fully tested hypotheses for swing-state origin and parents' voting habits, leaving age and college-based hypotheses for future analysis.

Using Fisher's Exact Test on a small BU student sample, the study concluded that coming from a swing state does not influence voting habits, while parental voting habits do. To confirm results, a larger sample would allow different statistical tests (e.g., two independent z-tests) and enable comparison with these initial findings. Applying Fisher's Exact Test to the age and college hypotheses would also refine conclusions.

Key findings show parental voting habits significantly shape student engagement; students whose parents voted in previous elections are more likely to vote. Although initial conclusions about swing-state origins were inconclusive with Fisher's test, the text suggests students from swing states showed heightened voting intentions, reflecting the 2024 election's national trend of increased youth turnout in pivotal states. Age also matters, as younger students (18-19) show stronger voting intentions than older peers, reflecting the excitement of first-time voting. Meanwhile, differences between majors, with social science students more politically engaged than those in technical fields, underscore how academic environments influence civic awareness.

Methodological challenges included sampling bias due to social media recruitment, limiting generalizability. Despite outreach through BU-affiliated channels and public platforms, a sample of 186 respondents fell short of the targeted 500. Future efforts should use university mailing lists and in-person surveys for more representative samples. Survey design also needs refinement as overlapping demographic questions and insufficient category detail caused respondent fatigue and limited depth. More precise demographic distinctions, especially in financial aid categories, would provide stronger data. These findings matter for universities seeking to increase their voting engagement. Strategies include targeted voter registration drives for out-of-state and swing-state students, guidance on absent or early voting, and including civic education into classes. Peer-led initiatives and student organization partnerships could also strengthen this culture of participation.

Going forward, longitudinal studies and broader comparative research across universities and regions could better contextualize BU's findings and adding qualitative methods like focus groups would also capture nuances that surveys may miss.

Supplement

Age The normality of the demographic collection of ages is graphed[5] here, in both a histogram and a qq plot. The histogram shows that the data is skewed left, with more data falling on the left side. The qq plot shows this as well, with the points falling heavily above the red line at both the beginning and end. This suggests that our demographic data for age was not random, as it was not normally distributed. This could be attributed to younger demographics being more likely to fill out a survey such as the one used to collect data.

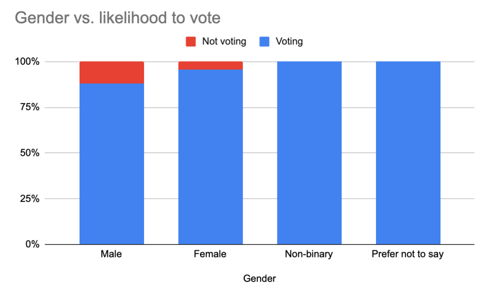


Fig. 4. Gender vs. Likelihood to Vote

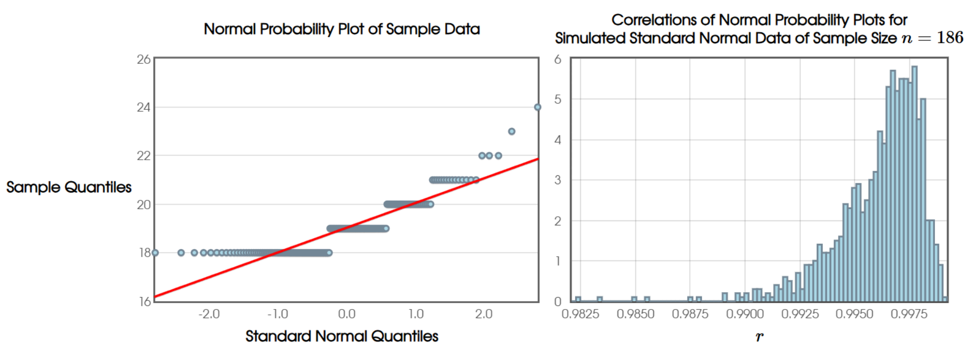


Fig. 5 and 6. QQ plot of Normal Probability Plot of Sample Data and Correlations of Normal Probability Plots for Simulated Standard Normal Data of Sample Size (n = 186)

Gender Of the people surveyed, a vast majority (60.8%) were female. Only four non-binary or non-gender providing responses were collected, and all four stated that they intend to vote. One data point was removed, a joke answer in which they wrote that their gender is “dinosaur”. 4% of females stated that they are not voting, while 11% of males stated that they are not voting. This difference of over double the non-voters could be attributed to the fact that one of the presidential candidates is female, and may draw more female voters.

Gender	Male	Female	Non-binary	Prefer not to say
Voting	60	108	2	2
Not voting	8	5	0	0
Total	68	113	2	2

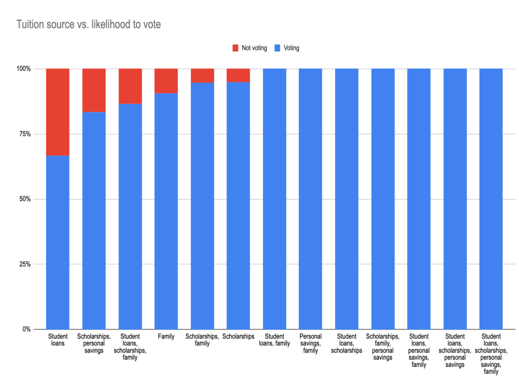


Fig. 7. Tuition Source vs. Likelihood to Vote

Tuition Source There is a large variety of methods that BU student’s use to pay for tuition, which means that there are a lot of specific combinations of responses. Because of this, much of the data is inconclusive when comparing the method of tuition payment and the likelihood of voting. For the sake of graphing, the provided “parents” response and 2 responses that wrote in “grandparents” were combined as the source of tuition into one category titled “family”. The largest group is those who are using only funds provided by their family, and of that group 9% are not voting.

[5] “Probability Plot Correlation Coefficient (PPCC) Test of Normality.” *PPCC Test of Normality*, stats.blue/Stats_Suite/normal_probability_plot.html. Accessed 31 Oct. 2024.

	Student loans	Scholarships, personal savings	Student loans, scholarships, family	Family	Scholarships, family	Scholarships	Student loans, family	Personal savings, family	Student loans, scholarships	Scholarships, family, personal savings	Student loans, personal savings, family	Student loans, scholarships, personal savings	Student loans, scholarships, personal savings, family
Voting	2	5	13	67	18	19	6	8	10	11	4	4	6
Not voting	1	1	2	7	1	1	0	0	0	0	0	0	0

Fisher's Exact Test For the Fisher's Exact Test, the total probability was calculated through observing each extreme table for the alternative hypothesis from each case. For example, for the Age hypotheses, the alternative hypothesis was that the proportion of voters who are between the ages of 18 and 20 were more likely to vote than those who were 21+. The extreme tables where the sample proportion would be greater than usual were summed to get a total probability. The following contingency tables below are examples of what these tables looked like:

Contingency Table for Parent's Voting Habits: Example Extreme Table (Extreme Table 1)

	Swing States	**Not Swing States**	**Total**
Voting	**a = 17**	**b = 152**	**169**
Not Voting	**c = 7**	**d = 4**	**11**
Total	**24**	**156**	**N = 180**

Note: The total sample size was adjusted to remove responses that were "unsure" of their parent's voting habits.

Contingency Table for Swing States: Example Extreme Table (Extreme Table 1)

	Swing States	**Not Swing States**	**Total**
Voting	**a = 29**	**b = 144**	**173**
Not Voting	**c = 1**	**d = 12**	**13**
Total	**30**	**156**	**N = 186**

All the calculations for our Fisher's Exact Test can be found here:
https://drive.google.com/file/d/1ZlvX2hENZuyk2ZTXbbJuUnpXeIo9c1OP/view?usp=sharing

Parent's Voting Patterns Fisher's Exact Test Calculations Below:

Probability of observing Extreme Table 1:

$$P_1 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{17}^{24} * C_{152}^{156}}{C_{169}^{180}} \approx 6.969 * 10^{-5}$$

Probability of observing Extreme Table 2:

$$P_2 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{16}^{24} * C_{153}^{156}}{C_{169}^{180}} \approx 3.872 * 10^{-6}$$

Probability of observing Extreme Table 3:

$$P_3 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{15}^{24} * C_{154}^{156}}{C_{169}^{180}} \approx 1.341 * 10^{-7}$$

Probability of observing Extreme Table 4:

$$P_4 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{14}^{24} * C_{155}^{156}}{C_{169}^{180}} \approx 2.595 * 10^{-9}$$

Probability of observing Extreme Table 5:

$$P_5 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{13}^{24} * C_{156}^{156}}{C_{169}^{180}} \approx 2.117 * 10^{-11}$$

Age Fisher's Exact Test Calculations Below:

Probability of observing Extreme Table 1:

$$P_1 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{162}^{167} * C_{15}^{20}}{C_{177}^{187}} \approx 0.0014$$

Probability of observing Extreme Table 2:

$$P_2 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{163}^{167} * C_{14}^{20}}{C_{177}^{187}} \approx 1.074 * 10^{-4}$$

Probability of observing Extreme Table 3:

$$P_3 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{164}^{167} * C_{13}^{20}}{C_{177}^{187}} \approx 5.2389 * 10^{-6}$$

Probability of observing Extreme Table 4:

$$P_4 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{165}^{167} * C_{12}^{20}}{C_{177}^{187}} \approx 1.5479 * 10^{-7}$$

Probability of observing Extreme Table 5:

$$P_5 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{166}^{167} * C_{11}^{20}}{C_{177}^{187}} \approx 2.4865 * 10^{-9}$$

Probability of observing Extreme Table 6:

$$P_6 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{167}^{167} * C_{10}^{20}}{C_{177}^{187}} \approx 1.6378 * 10^{-11}$$

College Major Fisher's Exact Test Calculations Below:

Probability of observing Extreme Table 1:

$$P_1 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{85}^{90} * C_{75}^{82}}{C_{150}^{172}} \approx 5.0268 * 10^{-11}$$

Probability of observing Extreme Table 2:

$$P_2 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^N} = \frac{C_{86}^{90} * C_{74}^{82}}{C_{150}^{172}} \approx 2.740 * 10^{-11}$$

Probability of observing Extreme Table 3:

Probability of observing Extreme Table 4:

$$P_3 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^{N}} = \frac{C_{87}^{90} * C_{73}^{82}}{C_{150}^{172}} \approx 1.0358 * 10^{-11}$$

$$P_4 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^{N}} = \frac{C_{88}^{90} * C_{72}^{82}}{C_{150}^{172}} \approx 2.5776 * 10^{-12}$$

Probability of observing Extreme Table 5:

$$P_5 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^{N}} = \frac{C_{89}^{90} * C_{71}^{82}}{C_{150}^{172}} \approx 3.7914 * 10^{-13}$$

Probability of observing Extreme Table 6:

$$P_6 = \frac{C_a^{a+c} * C_b^{b+d}}{C_{a+b}^{N}} = \frac{C_{90}^{90} * C_{70}^{82}}{C_{150}^{172}} \approx 2.4925 * 10^{-14}$$

Contributions

Marcos:
Deliverable 1: Helped brainstorm potential topics and questions
Deliverable 2: Created the survey and distributed it to other BU students
Deliverable 3: Wrote introduction and helped brainstorm ideas for other sections
Deliverable 4: Introduction and Abstract
Annabella:
Deliverable 1: Came up with specific target population and helped brainstorm questions
Deliverable 2: Identified independent/dependent and qualitative/quantitative variables, wrote some null and alternative hypotheses
Deliverable 3: Tables and graphs, some primary data analysis
Deliverable 4: Tables and graphs
Final paper: Confidence intervals
Enoch:
Deliverable 1: Suggested the topic of study and contributed to writing the potential questions for Data Collection and the analysis.
Deliverable 2: Contributed to writing the Null and Alternative hypothesis, helped identify independent.dependent and quantitative/quantitative variables.
Deliverable 3: Helped write the Experimental Methods section
Deliverable 4: Helped write the Experimental Methods section
Final Paper: Results and Fisher's Exact Test, Proof-read and proof-calculated the paper, made final corrections for the entire paper from feedback such as creating professional tables and professional wording, formatted the paper
Gui:
Deliverable 1: Helped with brainstorming potential topics for the study and devise initial preliminary questions to be asked in survey
Deliverable 2: Broke down key demographic areas that ultimately played a role when we created the exact questions for the survey that was sent out to collect data, also helped with deciding steps moving forward
Deliverable 3: Focused on Experimental Methods section and helped brainstorm ideas for data visualization
Deliverable 4: Helped write Experimental Methods
Final Paper: Updated experimental methods and discussion section, Proofread and formatted paper throughout revision process
Austin:
Deliverable 1: Organized our brainstorming on paper and helped come up with questions and methods to collect the data.
Deliverable 2: Wrote the action plan and organized our layout to make sure we answered all required questions.
Deliverable 3: Analyzed the data and stated changes that we made for the final survey, stating what we hope to achieve moving forward.
Deliverable 4: Analyzed the final data, showing our process of graphing our findings and citing online sources that are related to our data and analysis

References

"About Our Schools & Colleges." *About Our Schools & Colleges | Admissions*, www.bu.edu/admissions/why-bu/academics/schools-and-colleges/#PAR. Accessed 31 Oct. 2024.

"BU Facts & Stats." *BU Facts & Stats | Office of the President*, www.bu.edu/president/boston-university-facts-stats/. Accessed 31 Oct. 2024.

File, Thom. 2021. "Voting in America: A Look at the 2016 Presidential Election." *Census.Gov*. October 8. https://www.census.gov/newsroom/blogs/random-samplings/2017/05/voting_in_america.html.

Niemi, Richard G., and Michael J. Hanmer. 2010. *Voter Turnout among College Students: New Data and a Rethinking of Traditional Theories | Request PDF*. June. https://www.researchgate.net/publication/227377195_Voter_Turnout_Among_College_Students_New_Data_and_a_Rethinking_of_Traditional_Theories.

"Probability Plot Correlation Coefficient (PPCC) Test of Normality." *PPCC Test of Normality*, stats.blue/Stats_Suite/normal_probability_plot.html. Accessed 31 Oct. 2024.

"What Are the Current Swing States, and How Have They Changed over Time?" *USAFacts*, USAFacts, 7 Aug. 2024, usafacts.org/articles/what-are-the-current-swing-states-and-how-have-they-changed-over-time/.