Your project title

Approximately Normal: Jenny Huang, Bella Larsen, Jessie Bierschenk, Aidan Gildea

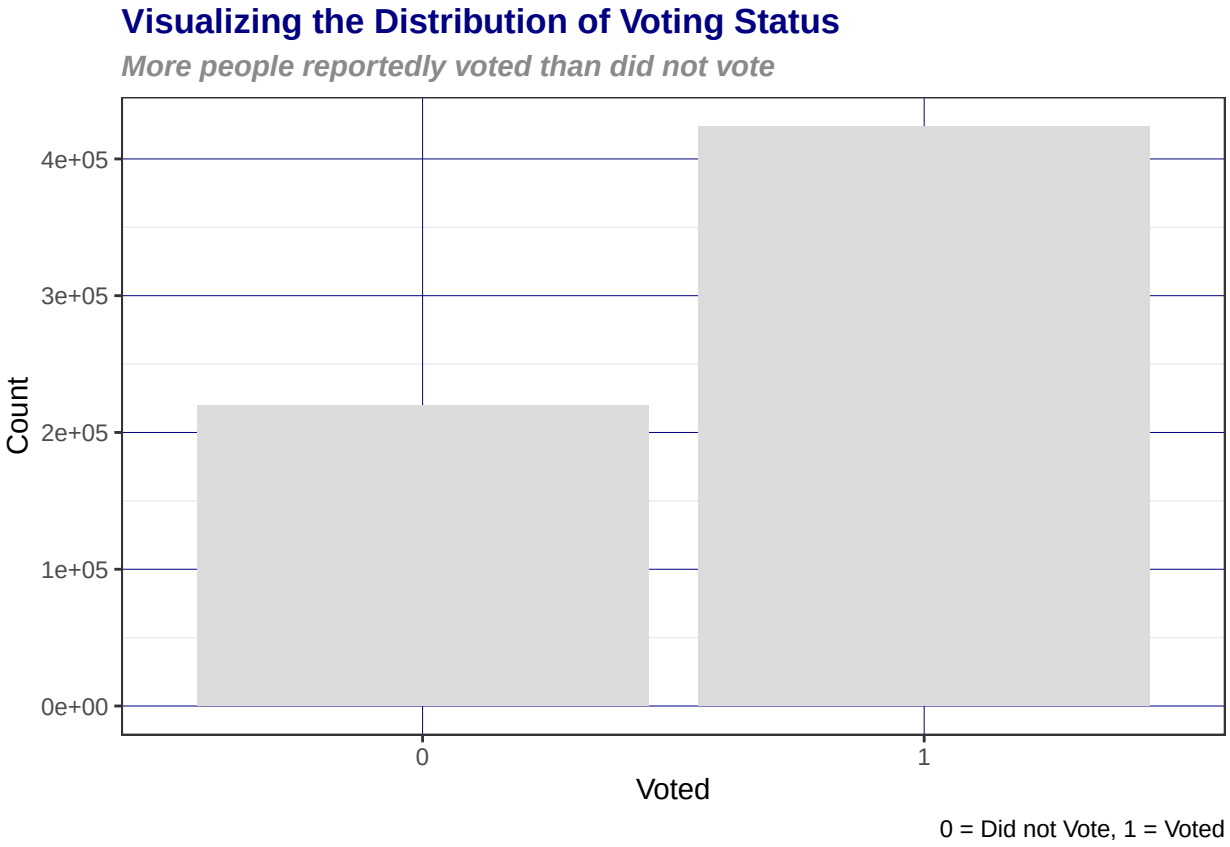
28 October 2020

Your written report goes here! Before you submit, make sure your code chunks are turned off with `echo = FALSE` and there are no warnings or messages with `warning = FALSE` and `message = FALSE`

Introduction

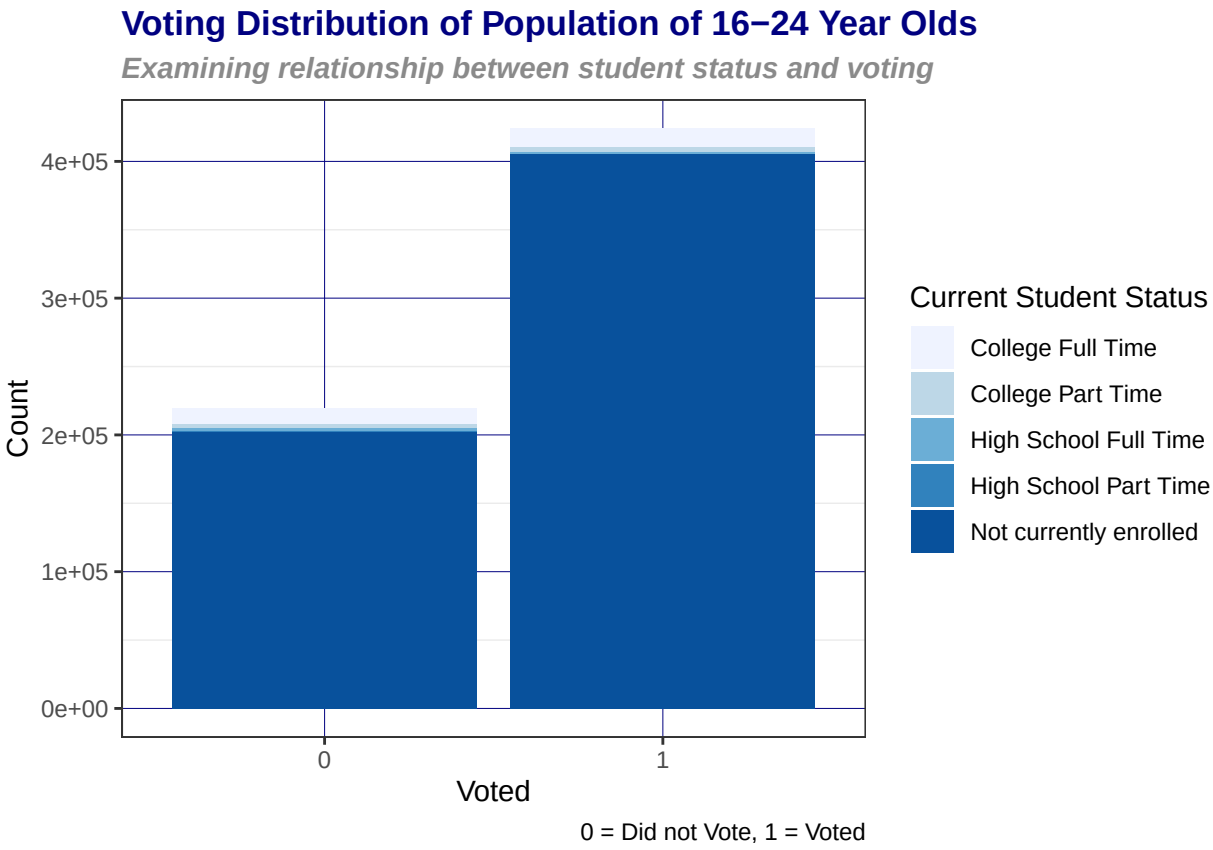
We will begin our EDA by visualizing the relationship between the response variable `voted` and several of the other variables of particular interest.

We will begin by simply looking at the distribution of those who voted throughout the last 8 years of elections.



From the barplot above, it is clear the more individuals in the data set voted (voted = 1) than did not (voted = 0).

As college students ourselves, we want to analyze whether or not being a student influences the frequency of voting. We will explore this preliminarily by visualizing the distribution of if school aged individuals (18-24) voted or not – categorized by their current student level. This is seen in the bar plot below.

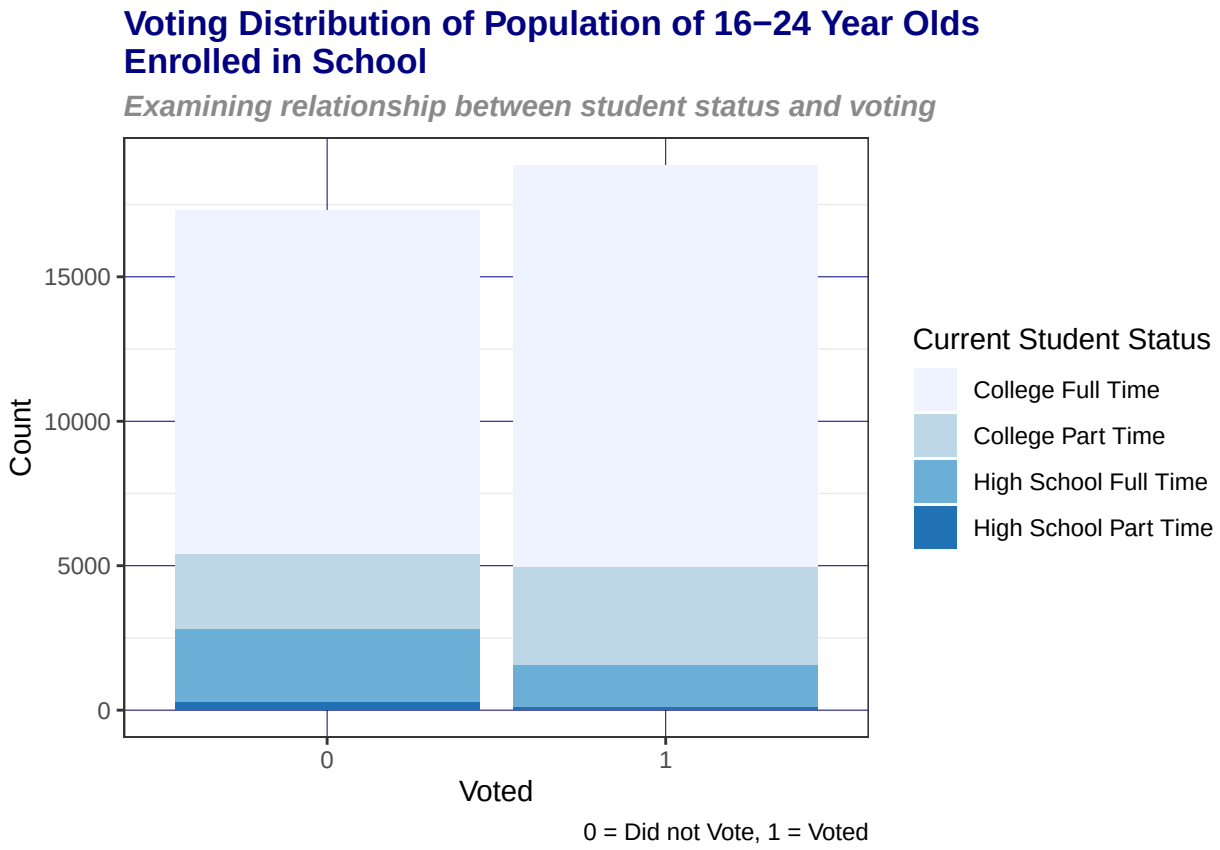


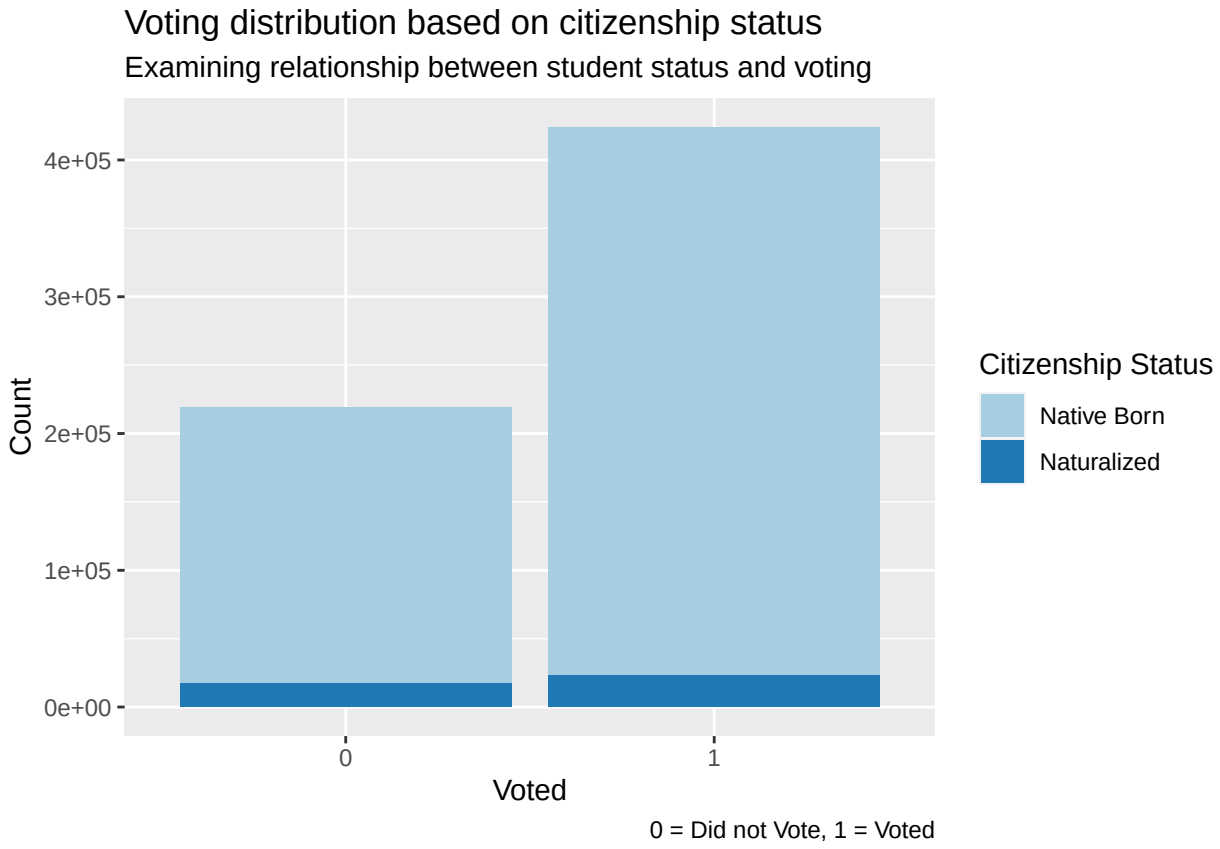
From the bar plot, it is evident that a majority of these individuals were not currently enrolled. This may be a result of a general national trend, but we want to investigate if it is the result of a larger proportion of older individuals within in the range of ages between 16-24. We will investigate this by analyzing those who are not currently enrolled in school within this age range.

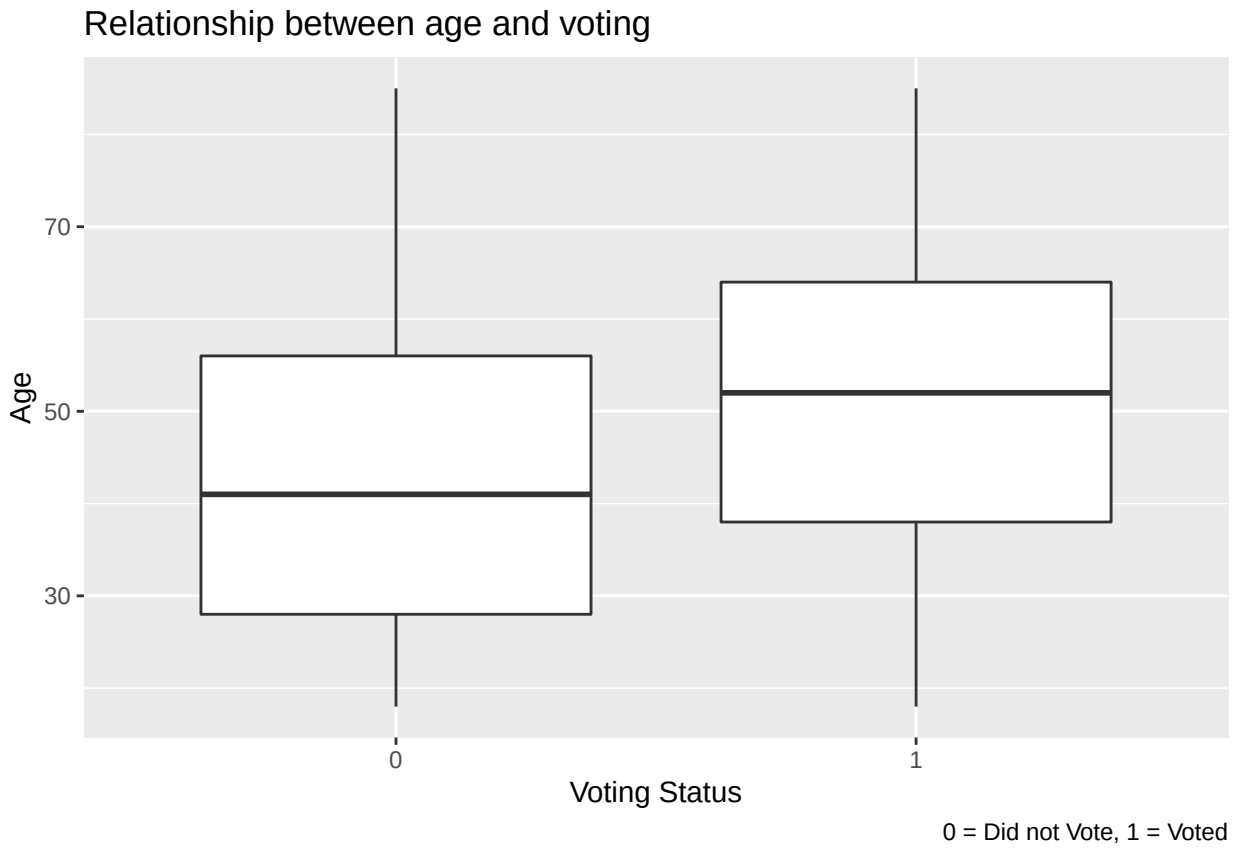
AGE	n	prop
18	2327	0.065
19	3671	0.102
20	4367	0.122
21	4760	0.133
22	5955	0.166
23	6991	0.195
24	7791	0.217

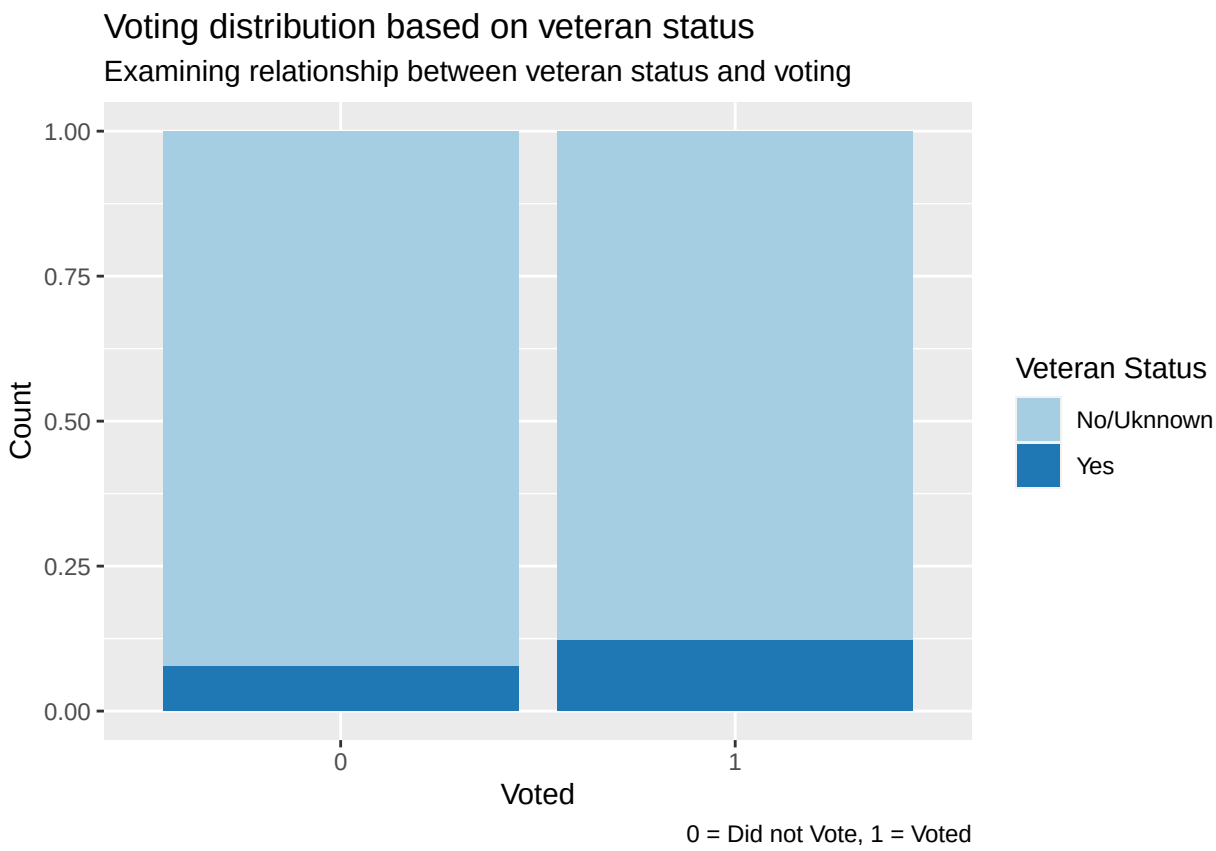
From the kable above, it is apparent that more than 40% of those not currently enrolled in school are 23-24 years old. This could be a potential reason for why this age range includes so many who are not currently enrolled as a student.

To more meaningfully analyze the relationship between being a student and if they vote or not, we adjusted our visualization to only include those currently enrolled in some level of education. This is seen in the visualization below.


Voting Distribution of Population of 16−24 Year Olds Enrolled in School
Examining relationship between student status and voting
15000
10000
5000
0
Count
0
1
Voted
0 = Did not Vote, 1 = Voted
Current Student Status
College Full Time
College Part Time
High School Full Time
High School Part Time


Voting distribution based on citizenship status
Examining relationship between student status and voting
4e+05
3e+05
2e+05
1e+05
0e+00
Count
0
1
Voted
0 = Did not Vote, 1 = Voted
Citizenship Status
Native Born
Naturalized


Relationship between age and voting
70
50
30
Age
0
1
Voting Status
0 = Did not Vote, 1 = Voted


Voting distribution based on veteran status
Examining relationship between veteran status and voting
1.00
0.75
0.50
0.25
0.00
Count
0
1
Voted
0 = Did not Vote, 1 = Voted
Veteran Status
No/Uknnown
Yes

keep in mind: citizenship and registration exclusion for the model