

Resident Elections at City of Madison

2022/5/1

DIS 315

Group: 6

Team Member Names: SHI KAIWEN; LOUIS NGUYEN; YINGQI ZHAO; COLE MICHAELSEN;

Sec #1: Introduction

Wisconsin is ranked the 25th most accessible state for citizens to vote. As the capital city of Wisconsin, the City of Madison’s residents vote a couple of times a year. We have collected the City of Madison’s voting data from 2008 to 2019. We want to investigate a couple of features in these data, such as which time slots Madison residents prefer to vote (before 11 AM or 11 AM till 4 PM). We would also like to investigate the changes in the number of people who vote each year and any other statistical characteristics show in the data set. We have found that Madison residents prefer to vote between 11 AM to 4 PM than to vote before 11 AM.

Sec #2: Background Information

The first data set (main data set we use) contains approximately 4000 rows of observations and 21 variables. Each observation has the election date, polling place, the number of people turning out at polling places to vote at different time slots (before 11 AM or from 11 AM to 4 PM), etc. The time of these polling place data is from 2008 to 2019. The data was collected via City of Madison Open Data. Our second data also came from the website City of Madison Open Data. It shows the geographical coordinates of all voting locations in the city of Madison.

Our goal for the rest of the report will be to analyze and compare the voter turnout during the year during those respective time frames in Madison. We will want to focus on these variables to understand better when voters show up to vote and if any other trends can be seen.

Here are explanations of some Columns of variables in the data set:

Key Variables from both the “Polling Place” and the “Election Turnout”

Name	Description
x	longitude of polling place
y	latitude of polling space
tvpoll_p_BLDG_NAME	Polling place name
Polling Place	Polling place name (same as above)
Election_Date	Date of the election (when do they vote)
11_AM_Turnout_Total	total number of people vote at the polling place before 11AM
4_PM_Turnout_Total	total number of people vote at the polling space from 11 AM to 4PM

Sec #3: References (Link to Dataset)

We collected these open data from **The City of Madison Collected Dataset** at <https://data-cityofmadison.opendata.arcgis.com/search?collection=Dataset> (<https://data-cityofmadison.opendata.arcgis.com/search?collection=Dataset>). The *City of Madison Election Turnout* provides all election statistics for our research, and the *City of Madison Polling Places* provides the geographic coordinate axis of all polling places.

Here is some citations we have used in our project:

- a. **Madison Population 2022;** (<https://worldpopulationreview.com/us-cities/madison-wi-population>)

Here is the link to the two data sets:

1. **City of Madison Election Turnout**; (<https://data-cityofmadison.opendata.arcgis.com/datasets/cityofmadison::election-turnout/about>)

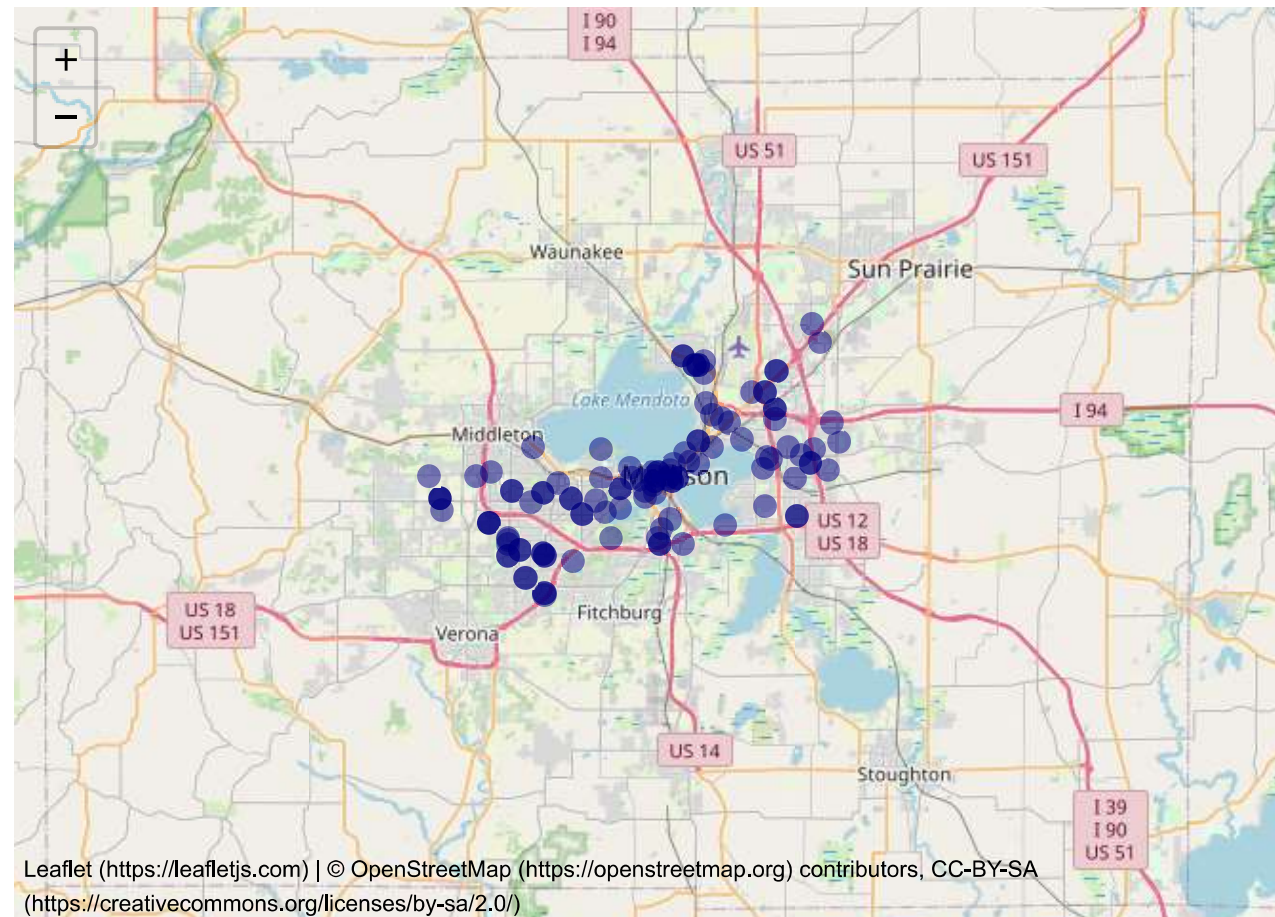
2. **City of Madison Polling Places**; (<https://data-cityofmadison.opendata.arcgis.com/datasets/cityofmadison::polling-places/about>)

Sec #4: Analysis & Graphs

We first tidy up our data so that our dataset fits our statistical research requirements.

Polling Places Map, City of Madison

p.s. Polling places show as **blue dots** on map

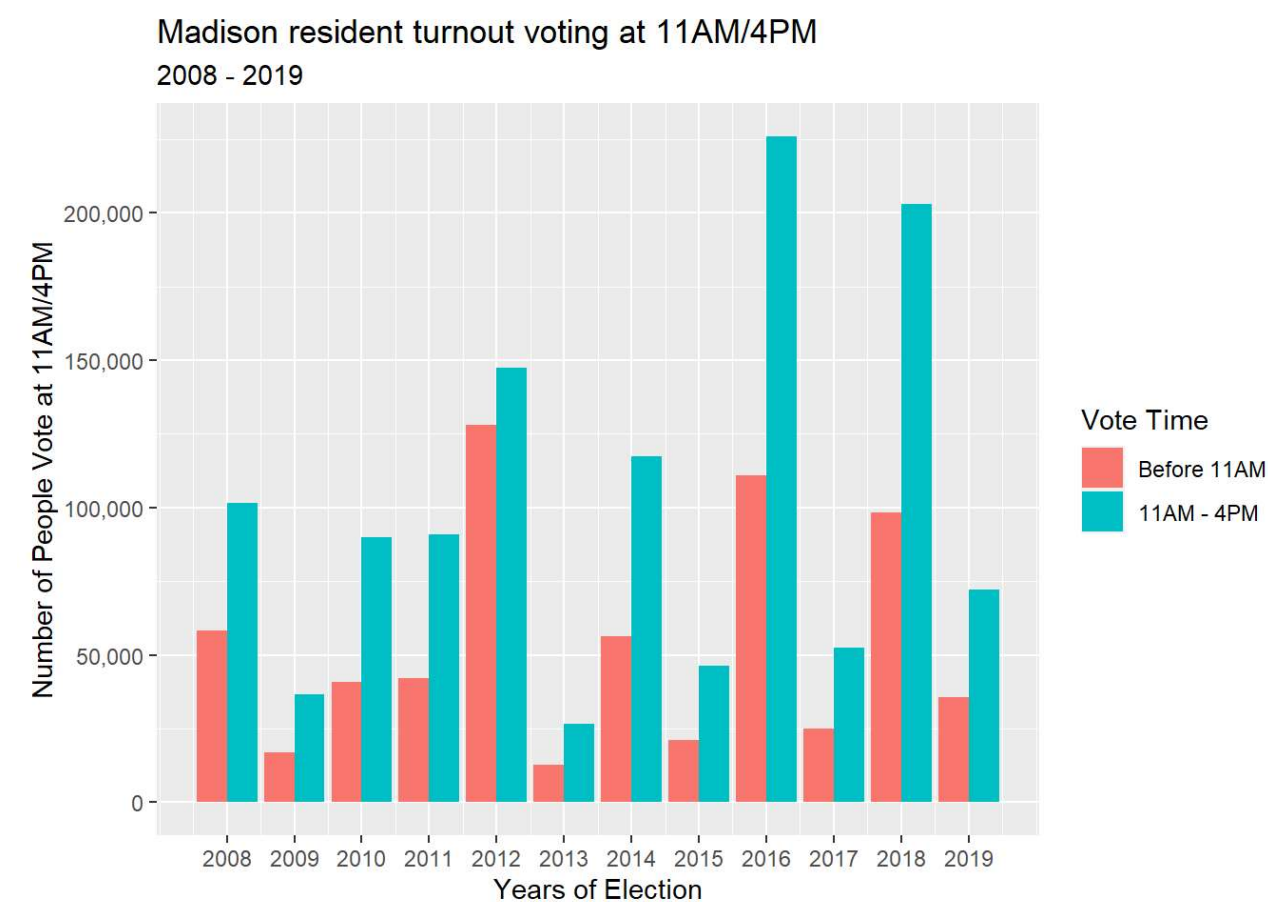


The Map above shows the geographical location of all polling places in the City of Madison (in blue dots).

When do Madison residents prefer to vote? Any preferences?

We want to investigate Madison residents' voting time preferences: do people prefer to turn out before 11 AM at polling places or from 11 AM to 4 PM?

Firstly, we have grouped our data by years (time of the election, the "**Election_Date**"). Then, we added up the number of people who turn out at polling places and vote before 11 AM and the number who turn out between 11 AM to 4 PM separately (the **11_AM_Turnout_Total** and the **4_PM_Turnout_Total**). Based on these data, in a cumulative bar chart with different color bars, we plot to show annually (2008 - 2019) the total number of people who turn out to vote before 11 AM and people who turn out to vote between 11 AM to 4 PM.



From the bar plot above, we can see: that from 2008 to 2019, the number of people who voted from 11 AM to 4 PM is different from people who vote before 11 AM. This could indicate that Madison residents have preferences on when they go to the polling station and vote. They are NOT indifferent between the 11 AM and 4 PM voting time slot. To statistically prove our presumption from the bar chart above, we try to operate a hypothesis test on it.

T-test

Our null hypothesis H_0 is: the number of Madison residents who vote at 11 AM or 4 PM is the **SAME**; and our alternative hypothesis H_A is: the number of people who vote at 11 (denoted as “ N_{11} ”) is **DIFFERENT** as 4 (denoted as “ N_4 ”) in every specific election date.

Therefore, our Hypothesis is:

$$\begin{aligned}
 &Two - tail\ T\ test \\
 &H_0 : N_{11} = N_4 \\
 &H_A : N_{11} \neq N_4 \\
 &\alpha = 0.05
 \end{aligned}$$

We then run our test and get our result.

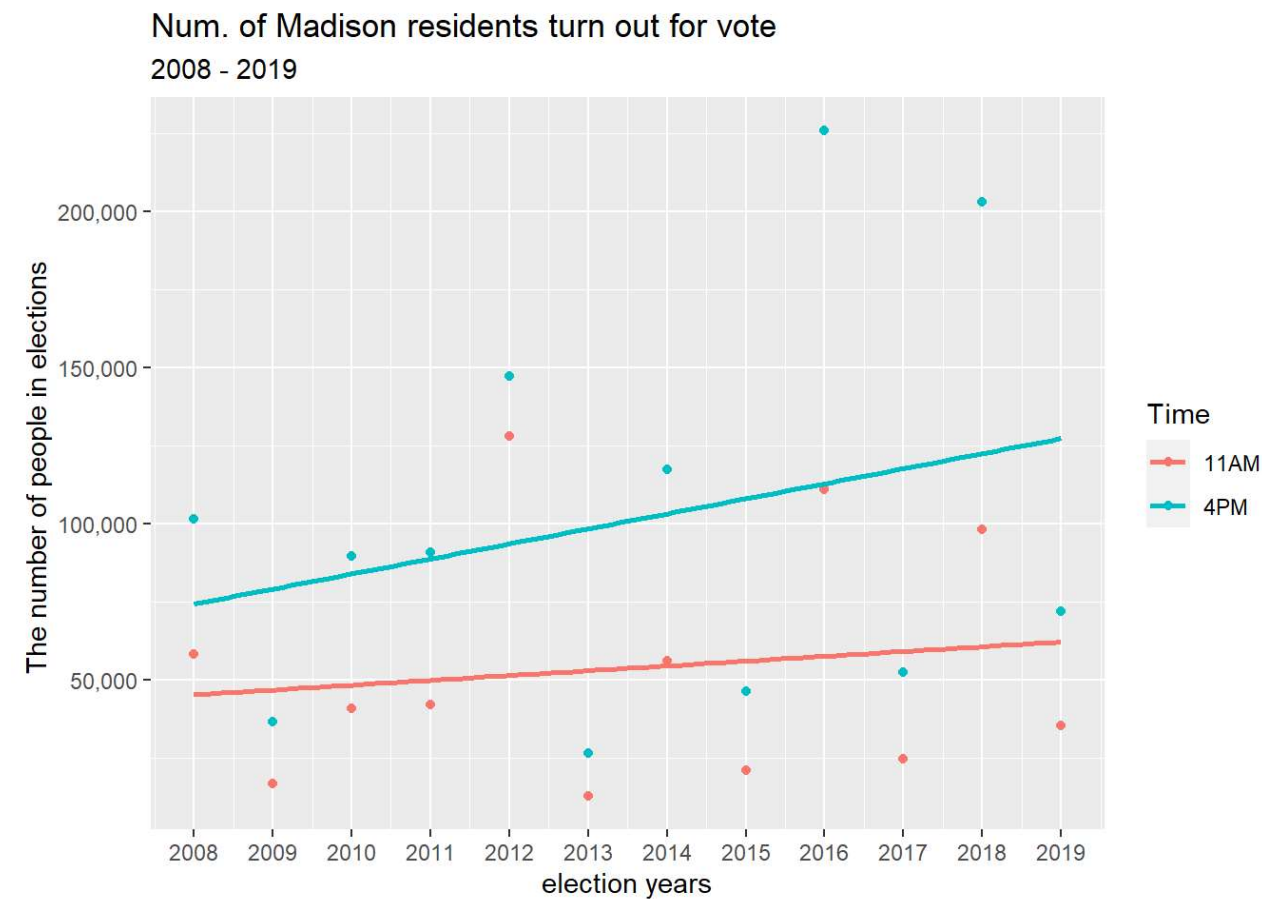
Welch Two Sample t-test: turnout11 and turnout4 (continued below)

Test statistic	df	P value	Alternative hypothesis	mean of x
-20.93	6824	2.467e-94 * * *	two.sided	146.7
mean of y				
275				

Our result clearly shows that the p-value is smaller than the criteria. So, with 95% of confidence, we reject the null hypothesis and claim that Madison resident has preferences among different time slot. The number of people who vote at 11 (denoted as “ N_{11} ”) is DIFFERENT from 4 (denoted as “ N_4 ”) on every specific election date.

More residents turning out to vote in the City of Madison?

We have already known that Madison residents have preferences on the time they turn out at polling places to vote. Aside from this habit, We are also curious about whether Madison residents have become more active to vote. That is, are there any changes on the total number of people who vote in the city of Madison during 2008 - 2019?



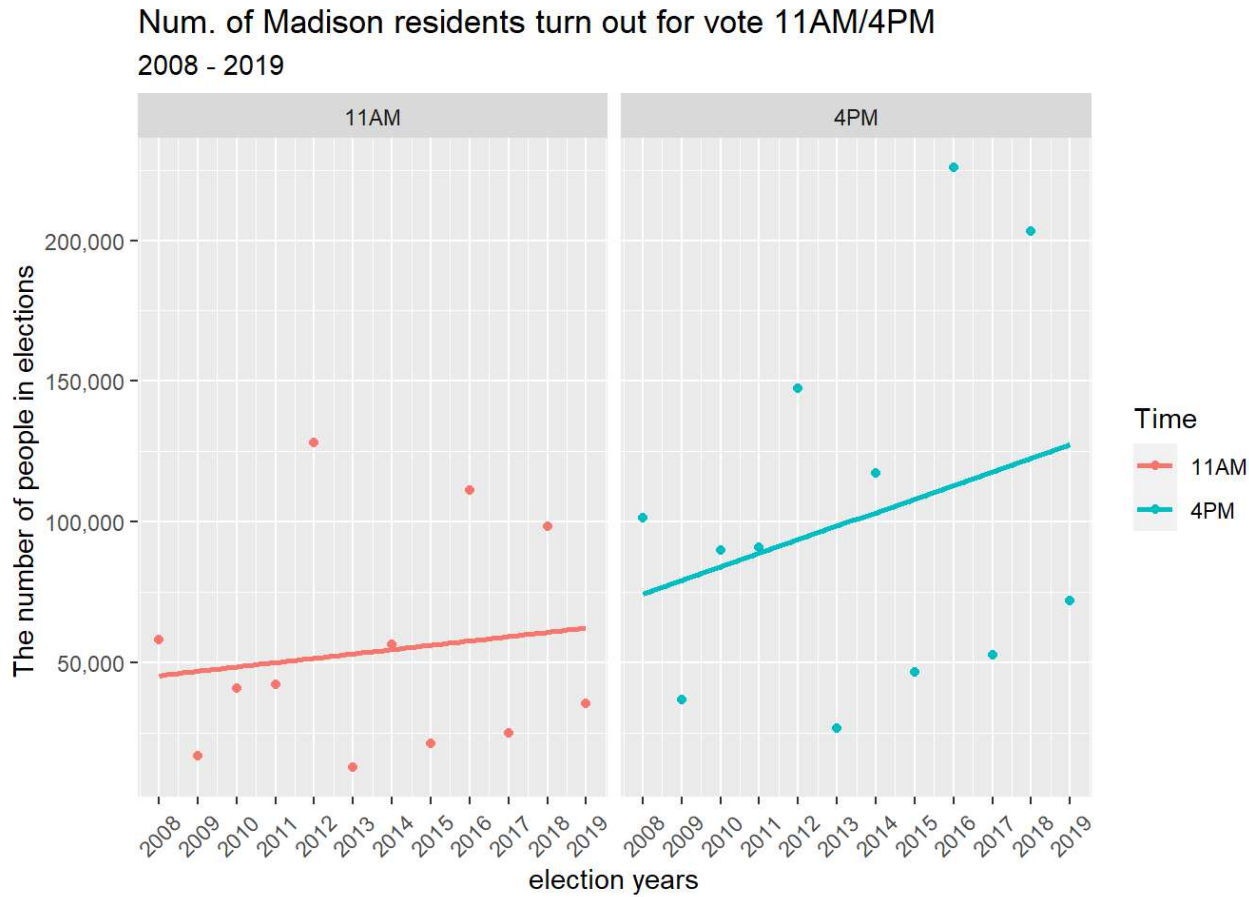
From the scatter plot above, we can see that the number of people turning out at polling places to vote has increased generally from 2008 to 2019. This makes sense as the City of Madison has a relatively 13.7% population increase rate in the past census ¹. This is an interesting discovery, as the general tendency to increase means both the number of people who turn out at polling places before 11 AM and the number of people who turn out at polling places between 11 AM to 4 PM have increased as time goes by.

However, are there any differences in terms of the increasing rate? As more residents turn out at polling places to vote, does it cause a steeper increase rate at any of the two voting time slots? To figure out the answer to this question, we try to run a regression analysis on our voting data set.

Linear Regression

Graphical Visualization

We firstly plot out a graph to see if we can find any visualized differences between `11_AM_Turnout_Total` and `4_PM_Turnout_Total` in terms of change in the number of people turning out for the vote.



Our graph above shows that the increasing rate for the two voting time slot is different. Specifically, the `4_PM_Turnout_Total` seems to have a steeper increase rate. So, what is the relationship between the two different time slots in terms of increasing the rate of total voting residents?

First Linear Regression

The first linear regression runs on the “11_AM_Turnout_Total” data set. It tries to find a linear relationship that tells the change of Madison residents to turn out at polling places before 11 AM to vote from 2008 to 2019.

(Intercept)	index
14185	274.5

Second Linear Regression

The first linear regression runs on the “4_PM_Turnout_Total” data set. It tries to find a linear relationship that tells the change of Madison residents to turn out at polling places from 11 AM to 4 PM to vote from 2008 to 2019.

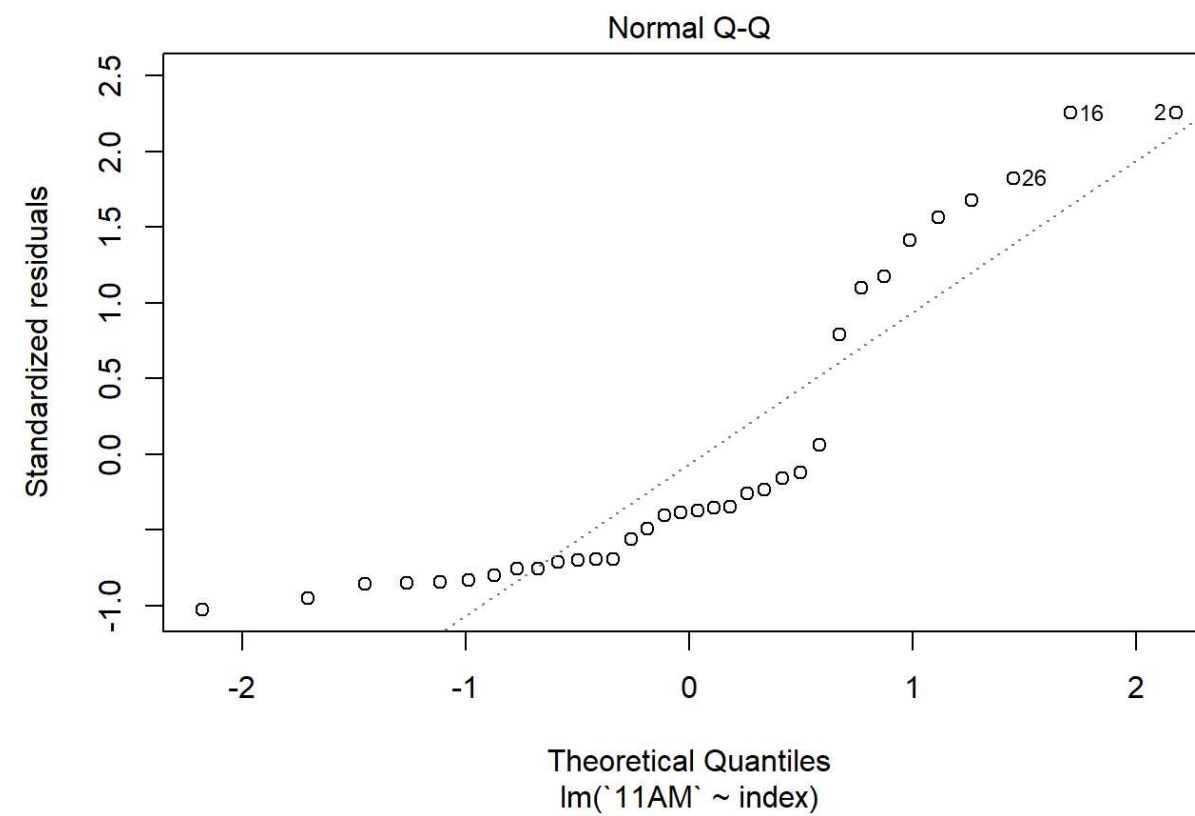
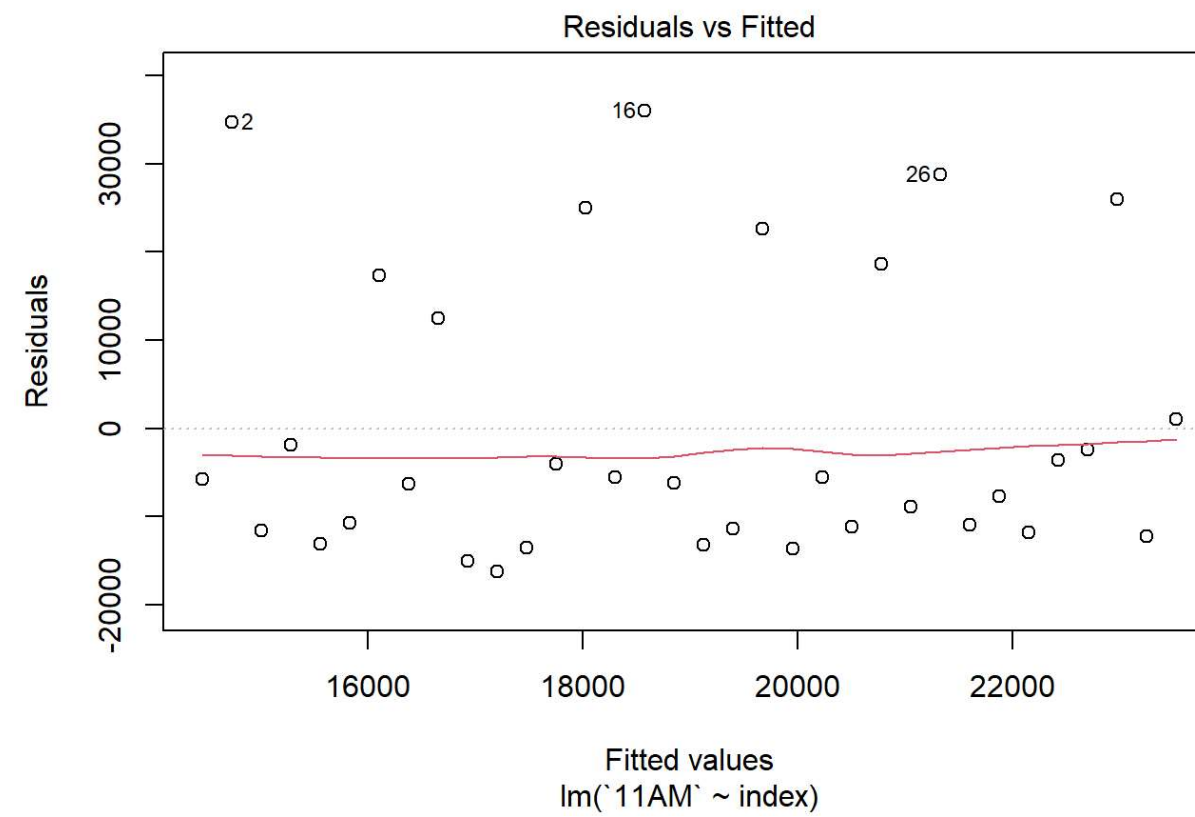
(Intercept)	index
24077	658.1

From the two linear regression results, we have seen that the `4_PM_Turnout_Total` has a steeper increase rate than the `11_AM_Turnout_Total` , as the index we got in the second linear regression model is larger than what we got from the first one. Therefore, we can conclude that generally, in the City of Madison, the number of people turning out at polling places to vote has increased from 2008 to 2019. Specifically, the increasing rate of Madison residents who vote between 11 AM and 4 PM is steeper than the increasing rate of Madison residents who vote before 11 AM.

Check: equal variance and normality

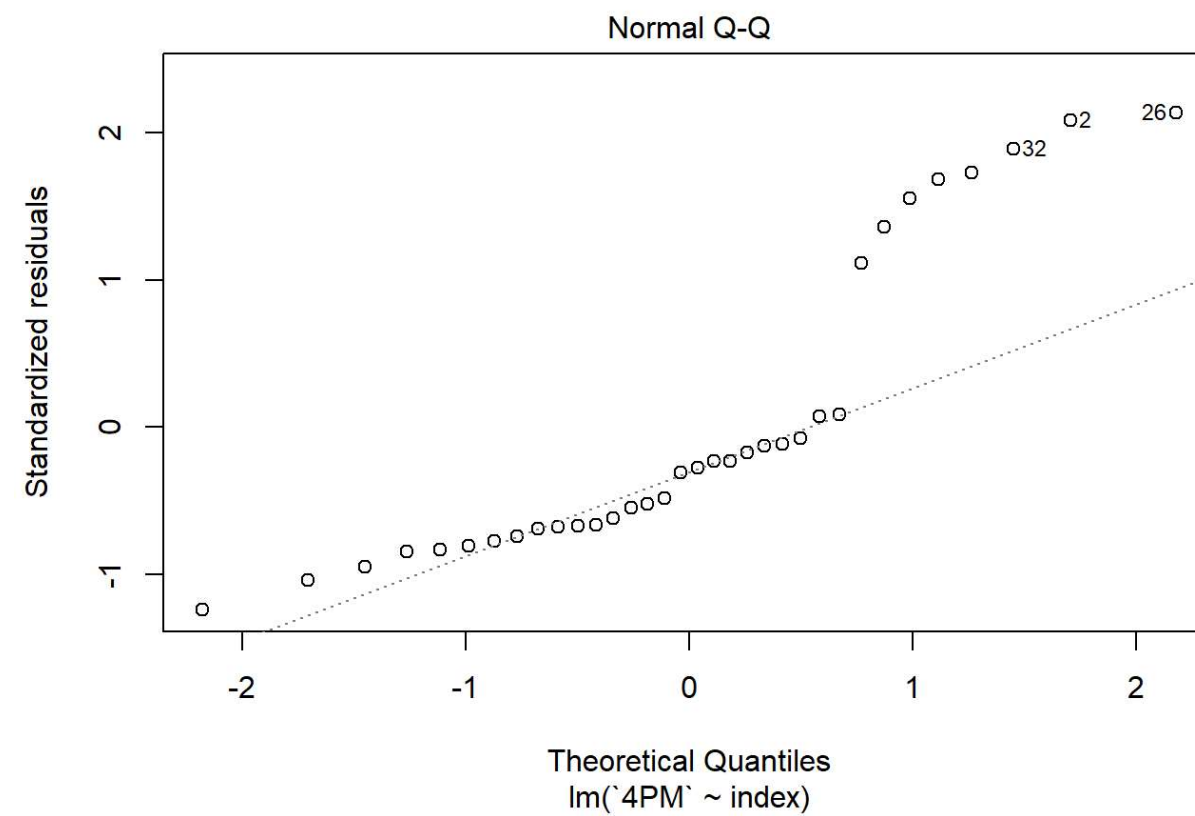
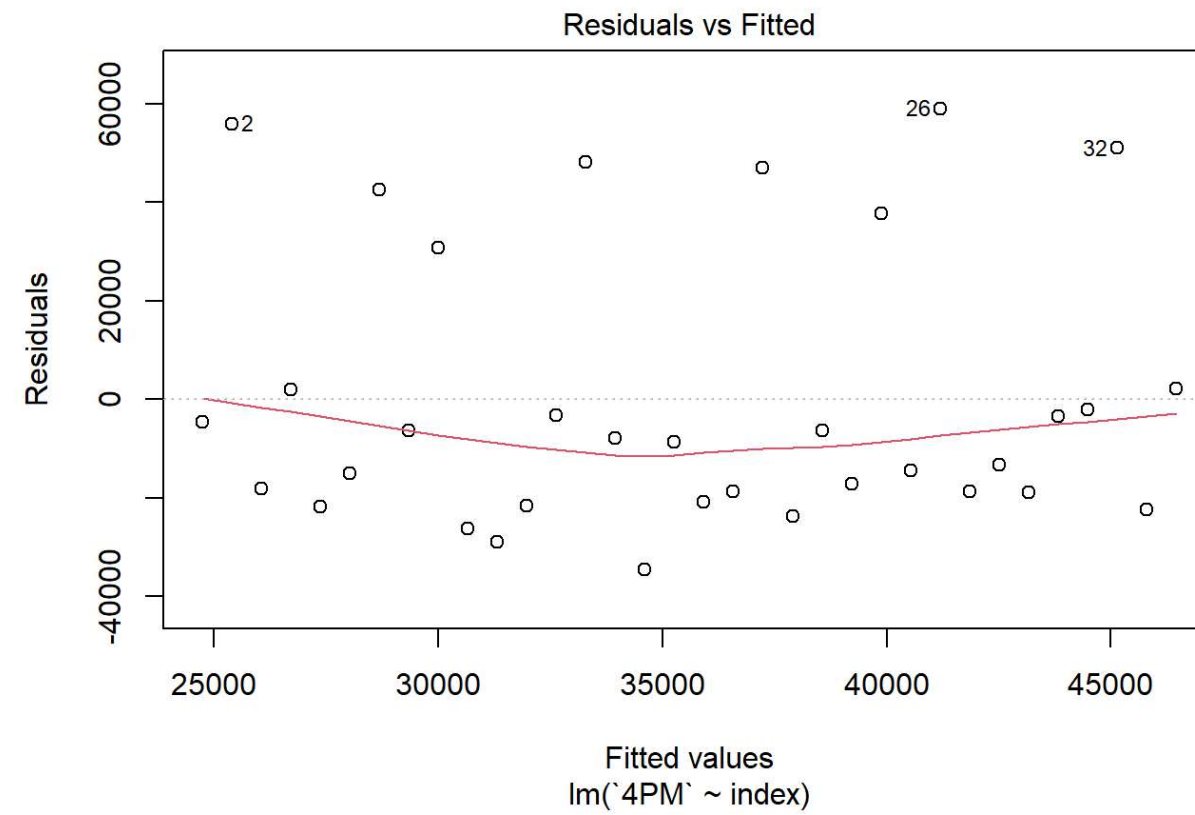
We want to check if our run regressions above are consistent with the regression model’s basic assumptions: Normality and equal variance. We have them plot the related diagram below.

Before 11 AM



The points are equally spread from the horizontal line by checking the equal variance for `11_AM_Turnout_Total` , so the equal variance assumption is not violated in our regression models. By checking the normality of the `11_AM_Turnout_Total` , we see points are mostly on the 45 degrees of line, although there are indeed some outliers which may be because of some political effects (e.g., presidential year election, etc.). Generally, the normality assumption is not violated in our regression models.

Between 11AM to 4 PM



The points are equally spread from the horizontal line by checking the equal variance for `4_PM_Turnout_Total`, so the equal variance assumption is not violated in our regression models. By checking the normality of the `4_PM_Turnout_Total`, we see points are mostly on the 45 degrees of line, although there are indeed some outliers which may be because of some political effects (e.g., presidential year election, etc.). Generally, the normality assumption is not violated in our regression models.

Sec #5 Discussion:

Our data analysis has gathered some interesting facts about polling in the City of Madison in the recent census. Firstly, Madison residents have a preference to turn out at polling places to vote. From the first bar chart, we see that Madison residents are more likely to vote between 11 AM to 4 PM rather than before 11 AM. We have also found that the number of residents turning out at polling places to vote increased generally in the past census. Specifically, most “new people” prefer to turn out at polling places to vote from 11 AM to 4 PM, as shown with a steeper slope in our linear regression model.

The shortcoming of our project is the clarity of the raw data set we used. Because there is no clarification on the original data set, we are not sure if the number of people turning out for voting at 4 PM is cumulative since morning or not. Therefore, we are not sure if we have made the correct utilization of the data set. Further clarification needs to be confirmed with the data set publisher.

Because we can't find a more detailed data set about voting, such as types of the vote and the total number of residents in the City of Madison at each election date, we cannot do some further analysis. In the future, if extra data set is available, we can also analyze the political effect on voting in the City of Madison. Our further question can be: do political affairs, such as types of election (presidential, congressional, or city-level), attracts or distracts Madison residents from voting.