Math 42 Final Project

Noor Benny, Samantha Alejandre, Han Qiang, Judy Zhu, and Bo He
Math 42

November 2022

Contents

1 Abstract

The purpose of this project is to examine the distribution of voter turnout rates in Congressional Elections of the population of different age groups and the distribution of voter turnout rates of different races. Using the data of voter turnout rates in Congressional Elections from 1966 to 2018, we develop a linear regression model on the growth and decay of voter turnout rates of various age groups and races through curve fitting. In this project, we provide an insight on how to resolve the lower voter turnout rates in Congressional Elections for young people and racial minority groups. We propose methods to increase the voter turnout rates in the younger population and racial minority groups.

2 Problem description

The United States lags behind other developed countries in the percentage of the voting-age population and equal ethnic minority population that votes. Age does have an impact on voter registration methods, so there should be equivalent voter turnout for each age category. Meanwhile, older voters chose places like town halls, and polling stations; methods like online and college campus registration were chosen by the younger voters. The problem we are attempting to solve is increasing the turnout rate concerning the population of 18-24 years old. From the census governmental data in the years from 1966 to 2018, the age population category between 18-24 consistently had the lowest turnout, which indicates how young people are not as aware/informed about elections as the older population. Therefore, it implies that our country is failing at promoting the importance of voting to the younger generation. Similarly, there is a large discrepancy in voting between different ethnic populations. The ethnic category that far exceeds other categories in voter turnout rates is the white population, meanwhile, the other categories such as Asian, Hispanic, and Black don't have the same amount of turnout rates. This further suggests how there is a need to take action to increase the voter turnout for these categories. The implication of diverse categories will allow a better representation of our democracy, and further help to avoid a skewed perspective.

The turnout rate of Congressional elections is influenced based on age and race. The turnout rate of Congressional elections is significantly lower than the voter turnout rate of General elections, especially in the younger population and in racially minority groups. For example, the voter turnout rate of the 2018 Congressional elections is 30.1% for people 18-24 years old. However, the voter turnout rate for 65 years and over is 63.8%. The voter turnout rate is essential because it may determine the policies of the candidates. Policymakers may lean towards populations who vote more to gain votes. As a result, our project aims to raise awareness of the importance of Congressional elections, which can lead to an election result that can represent citizens of all groups.
Some of the questions that we would like to answer are:

1. How does the percentage of voter turnout differ in age categories from 1966 to 2018? Which category has the highest/lowest turnout?

2. What is the linear relationship between voter turnout in different age categories? Why is this important, what does it tell us about voting in the United States?

3. How does the percentage of voter turnout differ in racial categories from 1966 to 2018? Which category has the highest/lowest turnout?

4. What is the linear relationship between voter turnout in different racial categories? Why is this important, and what does it tell us about voting in the United States?

5. What solution do we propose to increase the voter turnout for the minority age and race category? How do we expect the solution to increase voter turnout?

3 Simplifications

Voter turnout statistics for several demographic groups were presented to us in the data, so to simplify the model, we looked at voter turnout statistics for different race groups and age groups. We extracted and reorganized the data we needed onto a new spreadsheet with statistics for the percent turnout for race

groups and age groups. We also used data for the voting years with available statistics. For the race groups, data regarding voter turnout was available from 1966 to 2018. For the age groups, data regarding voter turnout was available from 1994 to 2018. By cleaning the data, we were able to focus on graphing the data and finding a model that best fit the data points.

The model was simplified to race groups and age groups because of the following assumptions:

- Race and ethnic groups play a significant role in political attitudes. The assumption is that several race based issues exist in the United States; so, race is a major, if not one of the most important, categories to analyze when analyzing voter turnout.
- Age groups also play an important role in political attitudes. The assumption is that several major issues in the United States exist that directly impact specific age groups, including education, retirement, and healthcare.

Although other demographics and issues exist in the United States, race and age related issues are among some of the most important subgroups to analyze in regards to voter turnout. Therefore, for our model, we simplified our extracted data to the following variables:

- Percent voter turnout for age groups
  - 18 to 24 year olds
  - 25 to 44 year olds
  - 45 to 64 year olds
  - 65+ year olds
- Percent voter turnout for race groups
  - White
  - White non-hispanic
  - Black

  - Asian
  - Hispanic
- Congressional election years
  - 1966 to 2018 for age groups
  - 1994 to 2018 for race groups

We plotted this data using R, and found that a linear regression model would best fit the data. We plotted the regression line on the graph, and found the statistics corresponding with the linear model, including the equation for the regression line.

4 Mathematical model

In this project, we utilize a linear regression model to turn the simplified problem into a mathematical model. Curve fitting is a process to find a curve that best describes the data. We created graphs for the voter turnout rates of Congressional elections from 1994 to 2018 for races including, white, non-white Hispanic, black, Asian, and Hispanic, and for age groups including 18-24 years old, 25-44 years old, 45-64 years old, 65 years old and over. We observed the patterns of the graphs of voter turnout rates. We determined that our outcome is supposed to be continuous and the patterns of the graphs appear to be a straight line. We thus concluded that we can best fit the data into a linear regression model. We used a simple linear regression model because our only independent variable is time. Our linear regression model describes the relationship between the voter turnout rates and various groups of citizens. The linear regression model allows us to study the variations in voter turnout rates among the different population groups. Therefore, we compared the linear relationship between ages and voter turnout during the duration of 1966 to 2018.

Slopes for racial voter turnout:

- The slope of the voter turnout rate of white is 0.02142857.
- The slope of the voter turnout rate of white Non Hispanic is 0.10625.
- The slope of the voter turnout rate of black is 0.06517857.

- The slope of the voter turnout rate of Asian is 0.1723214.
- The slope of the voter turnout rate of Hispanic is 0.2410714.

Slopes for age voter turnout:

- The slope of the voter turnout for 18-24 year olds is -0.1591758.
- The slope of the voter turnout for 25-44 year olds is -0.3589011.
- The slope of the voter turnout for 45-64 year olds is -0.2556044.
- The slope of the voter turnout for 65 year olds and older is 0.1140659.

It also allows us to predict future turnout rates based on past data.

Our independent variable is time, which are the years of every Congressional election since 1966 to 2018. Our dependent variables are voter turnout for the white population, voter turnout for white non-Hispanic population, voter turnout for the black population, voter turnout for the Asian population, voter turnout for the Hispanic population, voter turnout for 18-24 years, voter turnout for 25-44 years, voter turnout for 45-66 years, voter turnout for 65+ years.

5 Solution of the mathematical problem

In this project, we use computer programming in R to develop the mathematical solution for increasing voter turnout over the past years in congressional elections. We are able to apply standard practices of linear regression models for our study, by focusing on conditional probabilities distribution of the voter turnout given the values of predictors (age groups and race groups). We, therefore, assume a normal distribution of voting rates collected due to the great population in the states. We analyze the correlation coefficients in both predictor groups (age groups: 18-24, 25-44, 45-65, 65+; race groups: White, White Non-Hispanic, African, Asian) accordingly.

Degan and Merlo (2011) found that voter turnout rate had a positive correlation with the cost of each individual vote, which is also defined as one's " ideological positions, knowledge and civil duty"(Degan Merlo, 2011). However, Degan and Merlo were not able to observe these characteristics with their

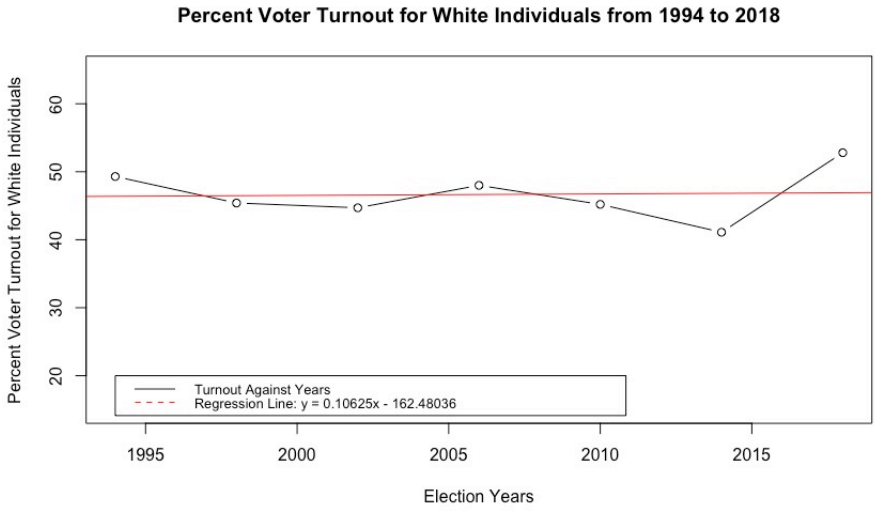
data collection. We improve this deflect by taking a database with a more precise description of the U.S. demographics. Thus, we could relate voters' cost of congressional elections with trace groups and age groups, for these represent their social classifications and experiences. Correlation coefficients of these models generated play a significant role to describe the relationship between voter turnout proportions of the voting population and these predictor markers mentioned, for the fact that linear regressions are applied. Therefore, by observing positive increases of voter turnout rates over time in each age group as well as race groups, we are able to identify voting activity that was positively impacted by the factor of time, thus, election education and efforts on social equity over the last few decades. Although there are outliers in few of these models, they were cast out by incidental policy changes.

6 Results

6.1 White individual voter turnout

```
# white individual voter turnout
# reading data from voterInfo table
whiteNEW <- voterInfoNEW$white
whiteNEW <- whiteNEW[1:7]
whiteNEW
typeof(whiteNEW)

votingYearsRaceNEW <- voterInfoNEW$yearsRace
votingYearsRaceNEW <- votingYearsRaceNEW[1:7]
votingYearsRaceNEW
typeof(votingYearsRaceNEW)

whiteNEWT <- t(whiteNEW)
votingYearsRaceNEWT <- t(votingYearsRaceNEW)
typeof(whiteNEWT)
typeof(votingYearsRaceNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$white ~ voterInfoNEW$yearsRace)

# plotting the data
plot(votingYearsRaceNEWT, whiteNEWT, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for White Individuals",
     main = "Percent Voter Turnout for White Individuals from 1994 to 2018",
     xlim = c(1994, 2018),
```

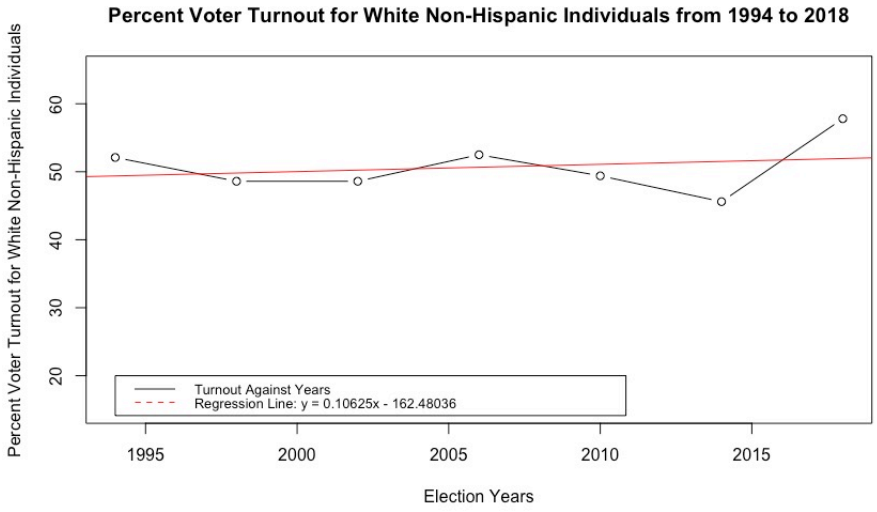
```
     ylim = c(15, 65))

# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1994, 20, legend = c("Turnout Against Years", "Regression Line:

y = 0.10625x - 162.48036"), col = c("black", "red"), lty = 1:6, cex = 0.8)
```



6.2 White Non-Hispanic individual voter turnout

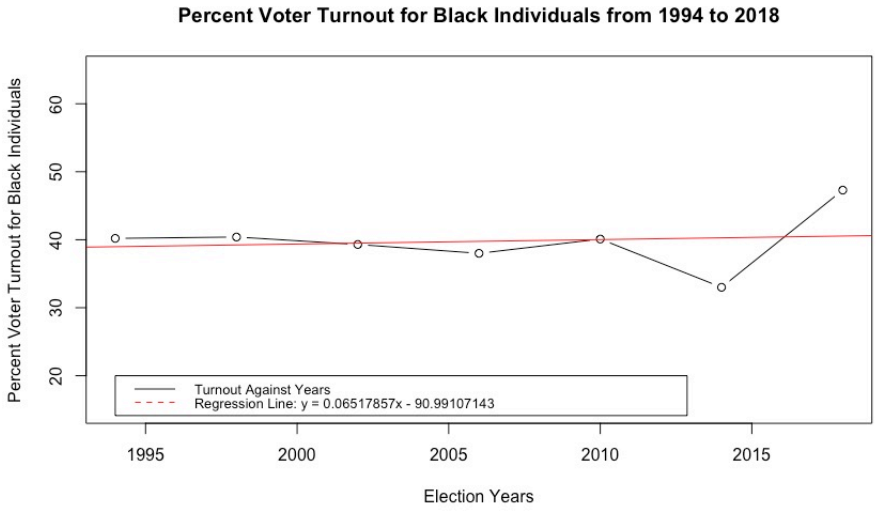
```
# white non-hispanic individual voter turnout

# reading data from voterInfo table
whiteNHNEW <- voterInfoNEW$whiteNonHispanic
whiteNHNEW <- whiteNHNEW[1:7]
whiteNHNEW
typeof(whiteNHNEW)

votingYearsRaceNEW <- voterInfoNEW$yearsRace
votingYearsRaceNEW <- votingYearsRaceNEW[1:7]
votingYearsRaceNEW
typeof(votingYearsRaceNEW)
```

```
whiteNHNEWT <- t(whiteNHNEW)
votingYearsRaceNEWT <- t(votingYearsRaceNEW)
typeof(whiteNHNEWT)
typeof(votingYearsRaceNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$whiteNonHispanic ~ voterInfoNEW$yearsRace)

# plotting the data
plot(votingYearsRaceNEWT, whiteNHNEWT, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for White Non-Hispanic Individuals",
     main = "Percent Voter Turnout for White Non-Hispanic Individuals
     from 1994 to 2018",
     xlim = c(1994, 2018),
     ylim = c(15, 65))

# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1994, 20, legend = c("Turnout Against Years", "Regression Line:
y = 0.10625x - 162.48036"), col = c("black", "red"), lty = 1:6, cex = 0.8)
```



6.3 Black individual voter turnout

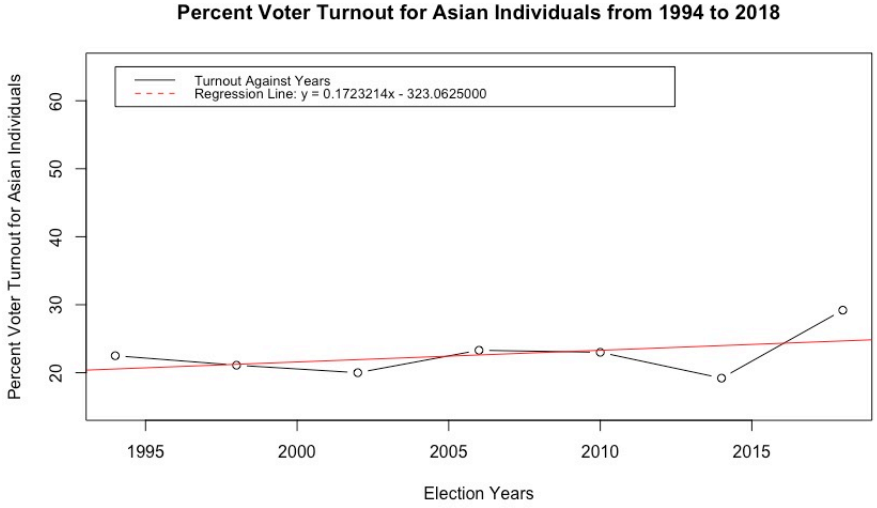
```
# black individual voter turnout

# reading data from voterInfo table
blackNEW <- voterInfoNEW$black
blackNEW <- blackNEW[1:7]
blackNEW
typeof(blackNEW)

votingYearsRaceNEW <- voterInfoNEW$yearsRace
votingYearsRaceNEW <- votingYearsRaceNEW[1:7]
votingYearsRaceNEW
typeof(votingYearsRaceNEW)

blackNEWT <- t(blackNEW)
votingYearsRaceNEWT <- t(votingYearsRaceNEW)
typeof(blackNEWT)
typeof(votingYearsRaceNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$black ~ voterInfoNEW$yearsRace)

# plotting the data
plot(votingYearsRaceNEWT, blackNEWT, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for Black Individuals",
     main = "Percent Voter Turnout for Black Individuals from 1994 to 2018",
     xlim = c(1994, 2018),
     ylim = c(15, 65))

# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1994, 20, legend = c("Turnout Against Years", "Regression Line:
y = 0.06517857x - 90.99107143"), col = c("black", "red"), lty = 1:6, cex = 0.8)
```



6.4 Asian individual voter turnout

```
# asian individual voter turnout

# reading data from voterInfo table
asianNEW <- voterInfoNEW$asian
asianNEW <- asianNEW[1:7]
asianNEW
typeof(asianNEW)

votingYearsRaceNEW <- voterInfoNEW$yearsRace
votingYearsRaceNEW <- votingYearsRaceNEW[1:7]
votingYearsRaceNEW
typeof(votingYearsRaceNEW)

asianNEWT <- t(asianNEW)
votingYearsRaceNEWT <- t(votingYearsRaceNEW)
typeof(asianNEWT)
typeof(votingYearsRaceNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$asian ~ voterInfoNEW$yearsRace)

# plotting the data
plot(votingYearsRaceNEWT, asianNEWT, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for Asian Individuals",
     main = "Percent Voter Turnout for Asian Individuals from 1994 to 2018",
     xlim = c(1994, 2018),
     ylim = c(15, 65))
```

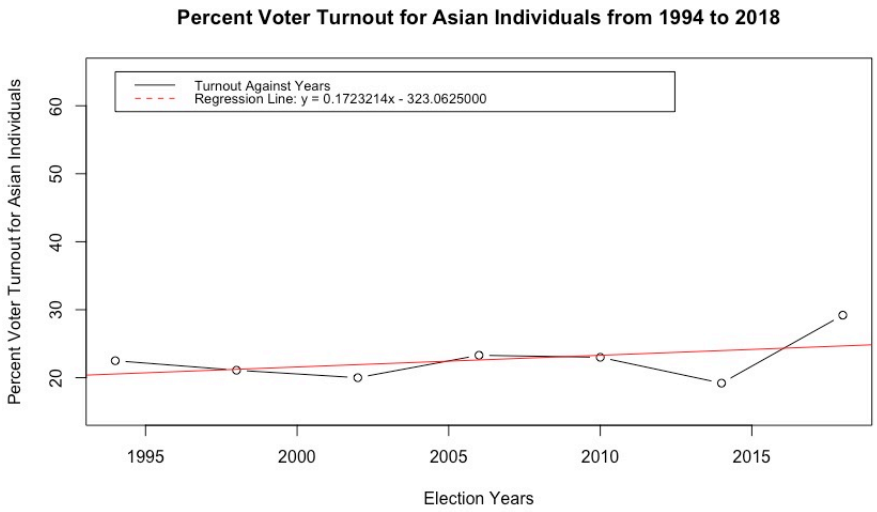
```
# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1994, 65, legend = c("Turnout Against Years", "Regression Line:
y = 0.1723214x - 323.0625000"), col = c("black", "red"), lty = 1:6, cex = 0.8)
```



6.5 Hispanic individual voter turnout

```
# hispanic individual voter turnout

# reading data from voterInfo table
hispanicNEW <- voterInfoNEW$hispanic
hispanicNEW <- hispanicNEW[1:7]
hispanicNEW
typeof(hispanicNEW)

votingYearsRaceNEW <- voterInfoNEW$yearsRace
votingYearsRaceNEW <- votingYearsRaceNEW[1:7]
votingYearsRaceNEW
typeof(votingYearsRaceNEW)

hispanicNEWT <- t(hispanicNEW)
votingYearsRaceNEWT <- t(votingYearsRaceNEW)
typeof(hispanicNEWT)
```

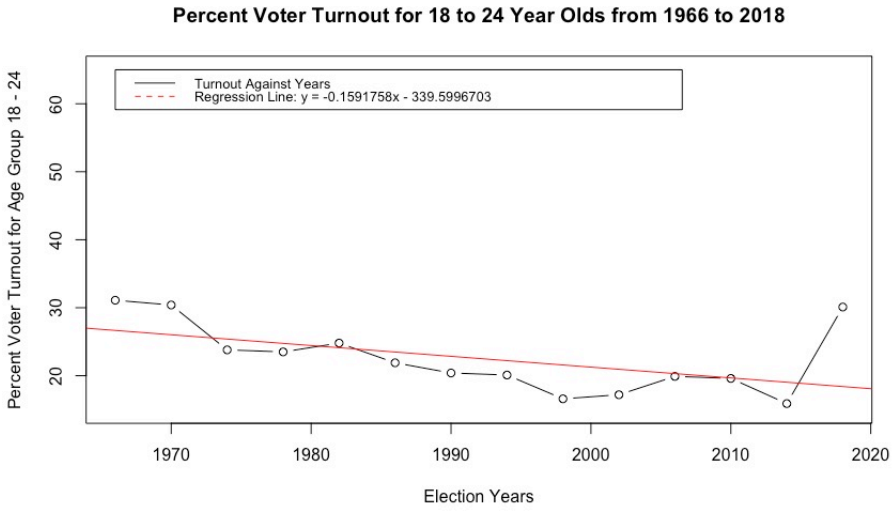
```
typeof(votingYearsRaceNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$hispanic ~ voterInfoNEW$yearsRace)

# plotting the data
plot(votingYearsRaceNEWT, hispanicNEWT, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for Hispanic Individuals",
     main = "Percent Voter Turnout for Hispanic Individuals from 1994 to 2018",
     xlim = c(1994, 2018),
     ylim = c(15, 65))

# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1994, 65, legend = c("Turnout Against Years", "Regression Line:
y = 0.2410714x - 461.7892857"), col = c("black", "red"), lty = 1:6, cex = 0.8)
```



6.6 Ages 18 to 24 voter turnout

```
# ages 18 to 24 voter turnout

# reading data from voterInfo table
```

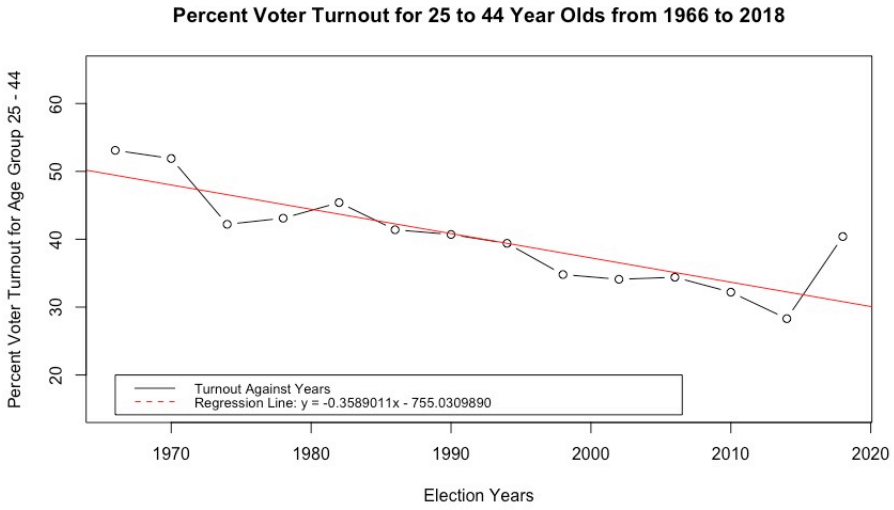
```
ages18to24NEW <- voterInfoNEW$ages18to24
ages18to24NEW <- ages18to24NEW[1:14]
ages18to24NEW
typeof(ages18to24NEW)

votingYearsAgeNEW <- voterInfoNEW$yearsAge
votingYearsAgeNEW <- votingYearsAgeNEW[1:14]
votingYearsAgeNEW
typeof(votingYearsAgeNEW)

ages18to24NEWT <- t(ages18to24NEW)
votingYearsAgeNEWT <- t(votingYearsAgeNEW)
typeof(ages18to24NEWT)
typeof(votingYearsAgeNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$ages18to24 ~ voterInfoNEW$yearsAge)

# plotting the data
plot(votingYearsAgeNEWT, ages18to24NEWT, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for Age Group 18 - 24",
     main = "Percent Voter Turnout for 18 to 24 Year Olds
     from 1966 to 2018",
     xlim = c(1966, 2018),
     ylim = c(15, 65))

# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1966, 65, legend = c("Turnout Against Years", "Regression Line:
y = -0.1591758x - 339.5996703"), col = c("black", "red"), lty = 1:6, cex = 0.8)
```



6.7 Ages 25 to 44 voter turnout

```
# ages 25 to 44 voter turnout

# reading data from voterInfo table
ages25to44NEW <- voterInfoNEW$ages25to44
ages25to44NEW <- ages25to44NEW[1:14]
ages25to44NEW
typeof(ages25to44NEW)

votingYearsAgeNEW <- voterInfoNEW$yearsAge
votingYearsAgeNEW <- votingYearsAgeNEW[1:14]
votingYearsAgeNEW
typeof(votingYearsAgeNEW)

ages25to44NEWT <- t(ages25to44NEW)
votingYearsAgeNEWT <- t(votingYearsAgeNEW)
typeof(ages25to44NEWT)
typeof(votingYearsAgeNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$ages25to44 ~ voterInfoNEW$yearsAge)

# plotting the data
plot(votingYearsAgeNEWT, ages25to44NEWT, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for Age Group 25 - 44",
     main = "Percent Voter Turnout for 25 to 44 Year Olds
     from 1966 to 2018",
     xlim = c(1966, 2018),
     ylim = c(15, 65))
```

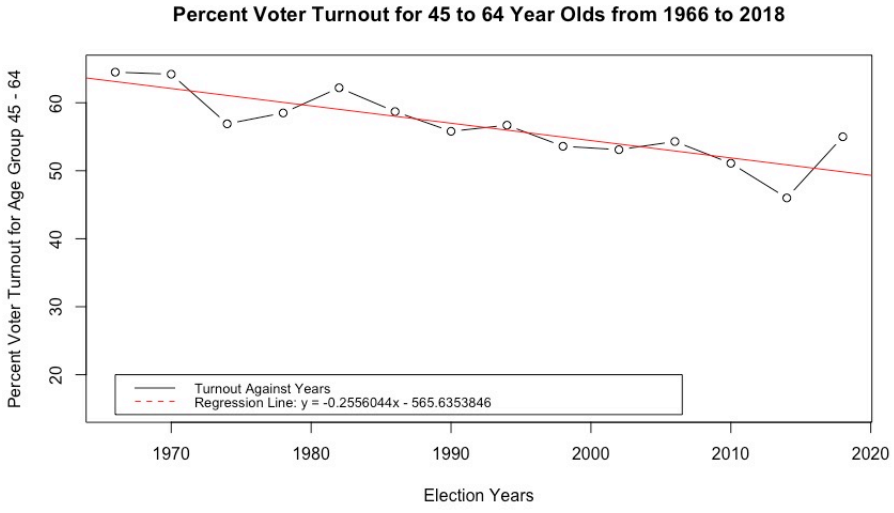
```
# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1966, 20, legend = c("Turnout Against Years", "Regression Line:
y = -0.3589011x - 755.0309890"), col = c("black", "red"), lty = 1:6, cex = 0.8)
```



6.8 Ages 45 to 64 voter turnout

```
# ages 45 to 64 voter turnout

# reading data from voterInfo table
ages45to64NEW <- voterInfoNEW$ages45to64
ages45to64NEW <- ages45to64NEW[1:14]
ages45to64NEW
typeof(ages45to64NEW)

votingYearsAgeNEW <- voterInfoNEW$yearsAge
votingYearsAgeNEW <- votingYearsAgeNEW[1:14]
votingYearsAgeNEW
typeof(votingYearsAgeNEW)

ages45to64NEWT <- t(ages45to64NEW)
votingYearsAgeNEWT <- t(votingYearsAgeNEW)
```

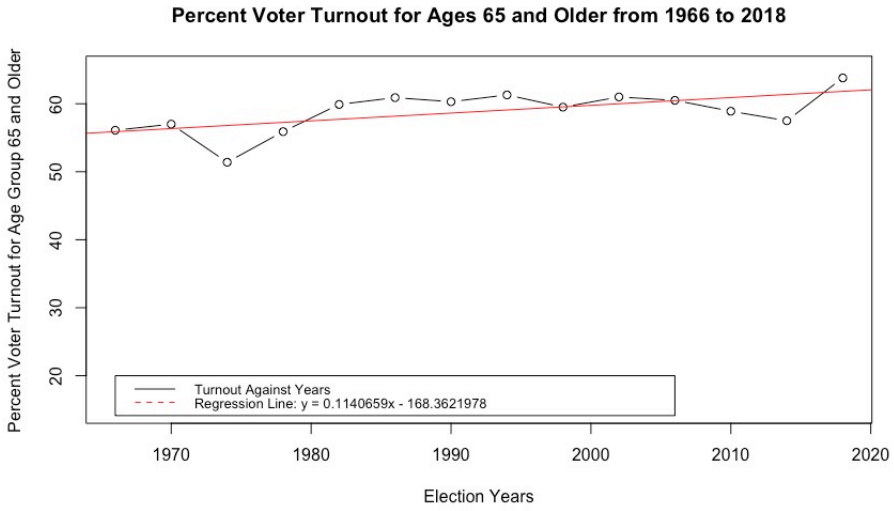
```
typeof(ages45to64NEWT)
typeof(votingYearsAgeNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$ages45to64 ~ voterInfoNEW$yearsAge)

# plotting the data
plot(votingYearsAgeNEWT, ages45to64NEWT, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for Age Group 45 - 64",
     main = "Percent Voter Turnout for 45 to 64 Year Olds
     from 1966 to 2018",
     xlim = c(1966, 2018),
     ylim = c(15, 65))

# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1966, 20, legend = c("Turnout Against Years", "Regression Line:
y = -0.2556044x - 565.6353846"), col = c("black", "red"), lty = 1:6, cex = 0.8)
```



6.9 Ages 65+ voter turnout

```
# ages 65+ voter turnout
```

```
# reading data from voterInfo table
agesAbove65NEW <- voterInfoNEW$agesAbove65
agesAbove65NEW <- agesAbove65NEW[1:14]
agesAbove65NEW
typeof(agesAbove65NEW)

votingYearsAgeNEW <- voterInfoNEW$yearsAge
votingYearsAgeNEW <- votingYearsAgeNEW[1:14]
votingYearsAgeNEW
typeof(votingYearsAgeNEW)

agesAbove65NEWT <- t(agesAbove65NEW)
votingYearsAgeNEWT <- t(votingYearsAgeNEW)
typeof(agesAbove65NEW)
typeof(votingYearsAgeNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$agesAbove65 ~ voterInfoNEW$yearsAge)

# plotting the data
plot(votingYearsAgeNEWT, agesAbove65NEW, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for Age Group 65 and Older",
     main = "Percent Voter Turnout for Ages 65 and Older
     from 1966 to 2018",
     xlim = c(1966, 2018),
     ylim = c(15, 65))

# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1966, 20, legend = c("Turnout Against Years", "Regression Line:
y = 0.1140659x - 168.3621978"), col = c("black", "red"), lty = 1:6, cex = 0.8)
```




```
# creating a data table of correlation coefficients for regression lines
install.packages("data.table")
library("data.table")

x1 <- c("White", "WhiteNH", "Black", "Asian", "Hispanic",
        "Ages18-24", "Ages25-44", "Ages45-64", "Ages65+")
x2 <- c(0.002422, 0.0549, 0.01786, 0.2053, 0.2739,
        0.2759, 0.7207, 0.7168, 0.386)
# Create data.table
dt_1 <- data.table(" " = x1, "Correlation Coefficent" = x2)
dt_1
```

```
            Correlation Coefficent
1:     White               0.002422
2:   WhiteNH               0.054900
3:     Black               0.017860
4:     Asian               0.205300
5:  Hispanic               0.273900
6: Ages18-24               0.275900
7: Ages25-44               0.720700
8: Ages45-64               0.716800
9:   Ages65+               0.386000
```

Analyzing the graphs:

- The age groups with the highest consistent turnout rates from the years 1966 to 2018 are voters who are between 45-65 years old and 65 or older. While the age group who have consistently the lowest turnout rate from the years 1966 to 2018 are people who are between the ages of 18-25.
- The race group with the highest consistent turnout rates from years 1966 to 2018 are voters who are white non-Hispanic. While the race group who

have consistently been at the lowest turnout rate from years 1966 to 2018 are voters who are Hispanic and Asian.

Analyzing the linear relationship between voter turnout and years:

- The age group with the greatest slope from the years 1966 to 2018 is voters who are 65 years old or older. Meanwhile, the other age groups have a negative slope, with voters 25-44 years old having a decreasing voter turnout.
- The race category with the greatest slope from the years 1966 to 2018 are voters who are Hispanic, while the group with the smallest increase is the white population. However, if we observe the graphs, the white population has a consistent voter turnout throughout those years, along with outwinning any other category when it comes to total percent turnout.

These results are interpretable in terms of the original problem by providing a visual and mathematically giving us evidence that there is low turnout rates for the younger population and racial minority. They are reasonable because we expected these results based on other historical governmental elections, however we are content to say that there has been an increase in groups who are the racial minority and continues to have an increase. On the other hand, this raises awareness that the younger age categories all have a decrease in voter turnout, especially the age group who are 18-25 years old. These results were not expected since there has been an increased number of ways you could vote with technology. In order to solve this problem, we will be proposing our own solutions in hope to see a positive slope for the younger age categories in the next few years.

Our propose solutions:

- Schools promoting voting and raising awareness on the importance in early high school
- Making it easily accessible by giving online option and having lots of stations open
- Sending voting information by mail to younger generation
- Incentivize colleges/universities workplaces to have their students/employees vote

- Could be a tax deductible per vote
- Make "Election Day" more than one day
  - Increase early voting initiatives in states
  - Allow in-person voting on more than one day
  - Make election day a holiday for those in the labor force and college students
  - To alleviate the problem of voters being busy
- Universal Absentee voting
  - Voters have two options
  - Mail-in
  - Drop off ballot

Throughout the past 4 years, absentee/mail voting in the US has been consistently trending upwards, and with COVID in 2020, the percentage of total votes cast absentee/by mail has gone up by 83.58 percent as of October 13th. Some states require an excuse for voting by mail. Due to the coronavirus, some states are giving all voters an excuse to vote by mail for certain elections. 83.8 million absentee ballots have been requested or sent to voters in 47 states and the District of Columbia for the 2020 general election, an indication that the number of Americans who plan to vote early or by mail will set records (DeSilver, 2020).

7 Improvement

Overall, we have some improvements about our model. First, the population under consideration is not the citizen voting age population analyzed in most recent reports, but rather the population of those of voting age. Furthermore, there is not much information available for race groups based on our dataset. In order to make our model a more accurate representation of the turnout rates for earlier years is by having more data on the minority race categories. Many of the data before 1966 had null values so they were not used, which decreases our knowledge of how voter turnout was in earlier years. Additionally, our model is based on a population of the United States that is representative of the entire population of the country. There are hence diverse circumstances according to state. For example, while Texas has lower turnout because less stations available and no online option, other states perform much better. For our model's graphs,

we have slightly differing R^2 statistics across our graphs, and for some data points our regression predictions are not the best match to the actual value. For example, on voter turnout for Black individuals from 1994 to 2018. In 2014, the percent voter turnout was 33% and the regression line is around 40%. So in order to improve our model, getting data after 2018 and before 1966 will produce more accurate predictions and results, along with getting specific data for certain groups in other elections to see if there is a similar outcome (we can add these lines to our current graphs to see how similar or different it is to other elections). It is definitely not impossible, but it will take time to find the available data and try to construct all the graphs.

8 Conclusions

In this paper, we have done a developed linear regression model about the growth and decay of voter turnout rates of ages and races groups. We used the data from Governmental from 1966 to 2018 to resolve how to increase young voter and racially minority groups' vote rates. We used R language to solve our mathematical model and give graphs to show how age and races group voter turnout rates are increasing or decreasing. We compared our data through the table of correlations of coefficients. We would like to use our linear regression model as a predictor for the 2022 Congressional Election. At this time, data is not yet available to the 2022 Congressional Election, so we cannot yet predict the data for future Congressional Elections.

References

[1] Bureau, U.S.C. 2021, Historical reported voting rates, Census.gov. Table A-7

[2] DeSilver, D. (2020, October 13). Mail-in voting became much more common in 2020 primaries as COVID-19 spread. Pew Research Center.

[3] Degan, A., Merlo, A. (2011). A STRUCTURAL MODEL OF TURNOUT AND VOTING IN MULTIPLE ELECTIONS. Journal of the European Economic Association, 9(2), 209–245.

Mathematical Modeling of U.S. Congressional Electoral Voters and the Solutions

Noor Benny, Samantha Alejandre, Han Qiang, Judy Zhu, and Bo He

December 2022

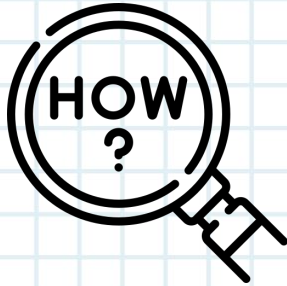
About our Model

The purpose of this project is to examine the distribution of voter turnout rates in Congressional Elections of the population of different ***age groups*** and the distribution of voter turnout rates of different ***races***. Using the data of voter turnout rates in Congressional Elections from 1966 to 2018, we develop a linear regression model on the growth and decay of voter turnout rates of various age groups and races through curve fitting. In this project, we provide an insight on how to resolve the lower voter turnout rates in Congressional Elections for ***young people and racial minority groups***. We propose methods to increase the voter turnout rates in the younger population and racial minority groups.

Problem Description

1. The United States lags behind other developed countries in the percentage of the voting-age population and equal ethnic minority population that votes.
2. The problem we are attempting to solve is increasing the turnout rate concerning the population of 18-24 years old.
3. The age population category between 18-24 consistently had the lowest turnout.
4. There is a large discrepancy in voting between different ethnic populations.

Questions



- How does the percentage of voter turnout differ in age categories from 1966 to 2018? Which category has the highest/lowest turnout?
- What is the linear relationship between voter turnout in different age categories? Why is this important, what does it tell us about voting in the United States?
- How does the percentage of voter turnout differ in racial categories from 1966 to 2018? Which category has the highest/lowest turnout?
- What is the linear relationship between voter turnout in different racial categories? Why is this important, and what does it tell us about voting in the United States?
- What solution do we propose to increase the voter turnout for the minority age and race category? How do we expect the solution to increase voter turnout?

Simplifications

- Voter turnout statistics for several demographic groups were presented to us in the data
  a. we looked at voter turnout statistics for different race groups and age groups.
     - For the race groups, data regarding voter turnout was available from 1966 to 2018.
     - For the age groups, data regarding voter turnout was available from 1994 to 2018.
- The model was simplified to race groups and age groups because of the following assumptions:
  a. Race and ethnic groups play a significant role in political attitudes. The assumption is that several race based issues exist in the United States; so, race is a major, if not one of the most important, categories to analyze when analyzing voter turnout.
  b. Age groups also play an important role in political attitudes. The assumption is that several major issues in the United States exist that directly impact specific age groups, including education, retirement, and healthcare.

Simplifications cont.

Race and age groups voter turnout:

- Percent voter turnout for age groups
  - 18 to 24 year olds
  - 25 to 44 year olds
  - 45 to 64 year olds
  - 65+ year olds
- Percent voter turnout for race groups
  - White
  - White non-hispanic
  - Black
  - Asian
  - Hispanic
- Congressional election years
  - 1966 to 2018 for age groups
  - 1994 to 2018 for race groups

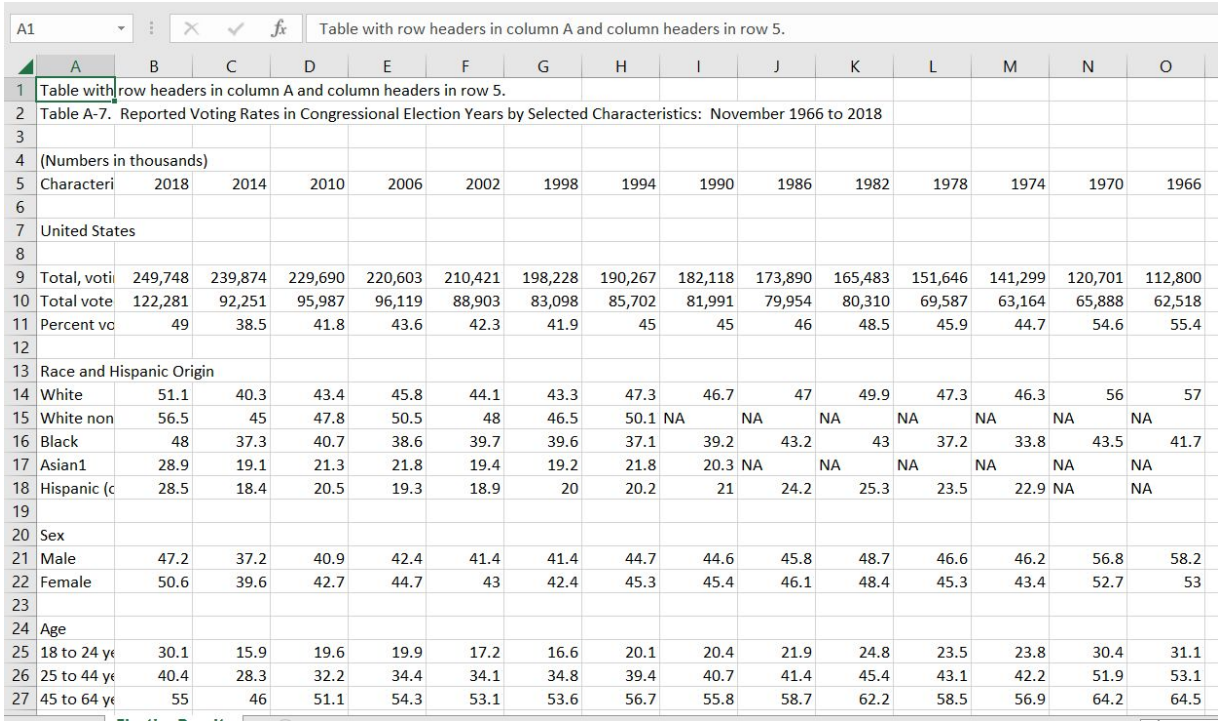
Mathematical Model

- We plotted this data using R, and found that a linear regression model would best fit the data.
- We utilize a linear regression model to turn the simplified problem into a mathematical model.
- Curve fitting is a process to find a curve that best describes the data.
    - We created graphs for the voter turnout rates of Congressional elections from 1994 to 2018 for races including, white, non-white Hispanic, black, Asian, and Hispanic, and for age groups including 18-24 years old, 25-44 years old, 45-64 years old, 65 years old and over.
    - We determined that our outcome is supposed to be continuous and the patterns of the graphs appear to be a straight line.
- Our linear regression model describes the relationship between the voter turnout rates and various groups of citizens.
- The linear regression model allows us to study the variations in voter turnout rates among the different population groups.

Mathematical Model cont.

- Our independent variable is time, which are the years of every Congressional election since 1966 to 2018. Our dependent variables are voter turnout for the white population, voter turnout for white non-Hispanic population, voter turnout for the black population, voter turnout for the Asian population, voter turnout for the Hispanic population, voter turnout for 18-24 years, voter turnout for 25-44 years, voter turnout for 45-64 years, voter turnout for 65+ years.

Sample of Original Data Set



A1 | Table with row headers in column A and column headers in row 5.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Table with row headers in column A and column headers in row 5.														
2	Table A-7. Reported Voting Rates in Congressional Election Years by Selected Characteristics: November 1966 to 2018														
3															
4	(Numbers in thousands)														
5	Characteri	2018	2014	2010	2006	2002	1998	1994	1990	1986	1982	1978	1974	1970	1966
6															
7	United States														
8															
9	Total, voti	249,748	239,874	229,690	220,603	210,421	198,228	190,267	182,118	173,890	165,483	151,646	141,299	120,701	112,800
10	Total vote	122,281	92,251	95,987	96,119	88,903	83,098	85,702	81,991	79,954	80,310	69,587	63,164	65,888	62,518
11	Percent vo	49	38.5	41.8	43.6	42.3	41.9	45	45	46	48.5	45.9	44.7	54.6	55.4
12															
13	Race and Hispanic Origin														
14	White	51.1	40.3	43.4	45.8	44.1	43.3	47.3	46.7	47	49.9	47.3	46.3	56	57
15	White non	56.5	45	47.8	50.5	48	46.5	50.1	NA	NA	NA	NA	NA	NA	NA
16	Black	48	37.3	40.7	38.6	39.7	39.6	37.1	39.2	43.2	43	37.2	33.8	43.5	41.7
17	Asian1	28.9	19.1	21.3	21.8	19.4	19.2	21.8	20.3	NA	NA	NA	NA	NA	NA
18	Hispanic (c	28.5	18.4	20.5	19.3	18.9	20	20.2	21	24.2	25.3	23.5	22.9	NA	NA
19															
20	Sex														
21	Male	47.2	37.2	40.9	42.4	41.4	41.4	44.7	44.6	45.8	48.7	46.6	46.2	56.8	58.2
22	Female	50.6	39.6	42.7	44.7	43	42.4	45.3	45.4	46.1	48.4	45.3	43.4	52.7	53
23															
24	Age														
25	18 to 24 ye	30.1	15.9	19.6	19.9	17.2	16.6	20.1	20.4	21.9	24.8	23.5	23.8	30.4	31.1
26	25 to 44 ye	40.4	28.3	32.2	34.4	34.1	34.8	39.4	40.7	41.4	45.4	43.1	42.2	51.9	53.1
27	45 to 64 ye	55	46	51.1	54.3	53.1	53.6	56.7	55.8	58.7	62.2	58.5	56.9	64.2	64.5

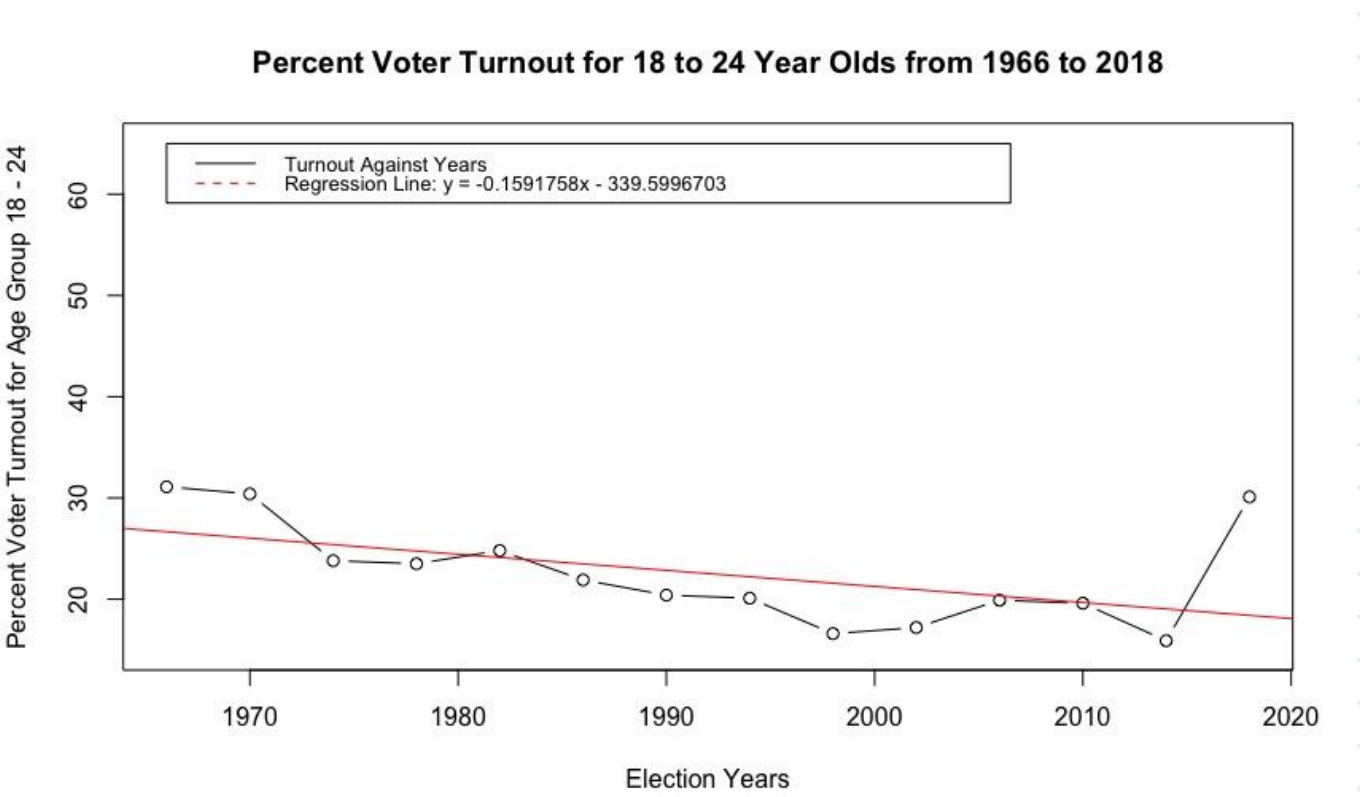
Results cont.

```r
# ages 18 to 24 voter turnout

# reading data from voterInfo table
ages18to24NEW <- voterInfoNEW$ages18to24
ages18to24NEW <- ages18to24NEW[1:14]
ages18to24NEW
typeof(ages18to24NEW)

votingYearsAgeNEW <- voterInfoNEW$yearsAge
votingYearsAgeNEW <- votingYearsAgeNEW[1:14]
votingYearsAgeNEW
typeof(votingYearsAgeNEW)

ages18to24NEWT <- t(ages18to24NEW)
votingYearsAgeNEWT <- t(votingYearsAgeNEW)
typeof(ages18to24NEWT)
typeof(votingYearsAgeNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$ages18to24 ~ voterInfoNEW$yearsAge)
```

```r
# plotting the data
plot(votingYearsAgeNEWT, ages18to24NEWT, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for Age Group 18 - 24",
     main = "Percent Voter Turnout for 18 to 24 Year Olds from 1966 to 2018",
     xlim = c(1966, 2018),
     ylim = c(15, 65))

# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1966, 65, legend = c(
  "Turnout Against Years", "Regression Line: y = -0.1591758x - 339.5996703"),
  col = c("black", "red"),
  lty = 1:6, cex = 0.8)
```

Results cont.

Ages 18 to 24 voter turnout



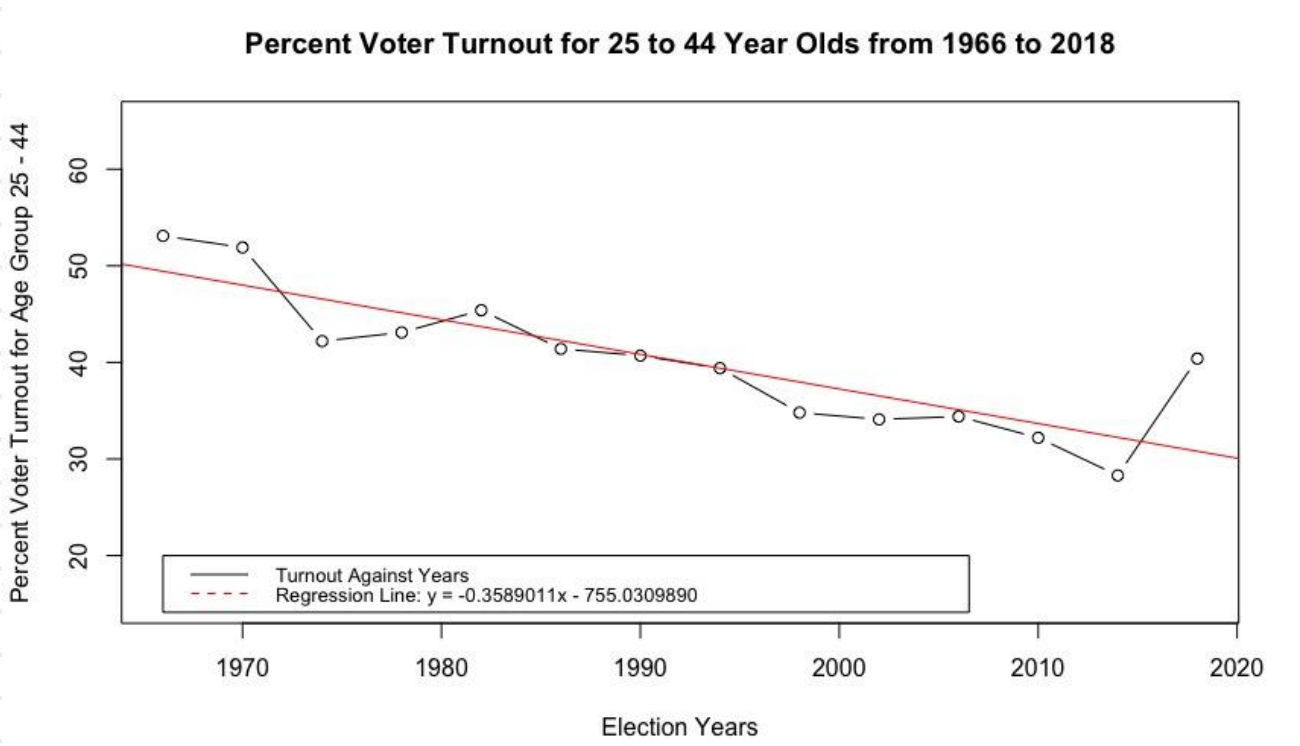
Results cont.

```r
# ages 25 to 44 voter turnout

# reading data from voterInfo table
ages25to44NEW <- voterInfoNEW$ages25to44
ages25to44NEW <- ages25to44NEW[1:14]
ages25to44NEW
typeof(ages25to44NEW)

votingYearsAgeNEW <- voterInfoNEW$yearsAge
votingYearsAgeNEW <- votingYearsAgeNEW[1:14]
votingYearsAgeNEW
typeof(votingYearsAgeNEW)

ages25to44NEWT <- t(ages25to44NEW)
votingYearsAgeNEWT <- t(votingYearsAgeNEW)
typeof(ages25to44NEWT)
typeof(votingYearsAgeNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$ages25to44 ~ voterInfoNEW$yearsAge)
```

```r
# plotting the data
plot(votingYearsAgeNEWT, ages25to44NEWT, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for Age Group 25 - 44",
     main = "Percent Voter Turnout for 25 to 44 Year Olds from 1966 to 2018",
     xlim = c(1966, 2018),
     ylim = c(15, 65))

# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1966, 20, legend = c(
  "Turnout Against Years", "Regression Line: y = -0.3589011x - 755.0309890"),
  col = c("black", "red"), lty = 1:6, cex = 0.8)
```

Results cont.

Ages 25 to 44 voter turnout



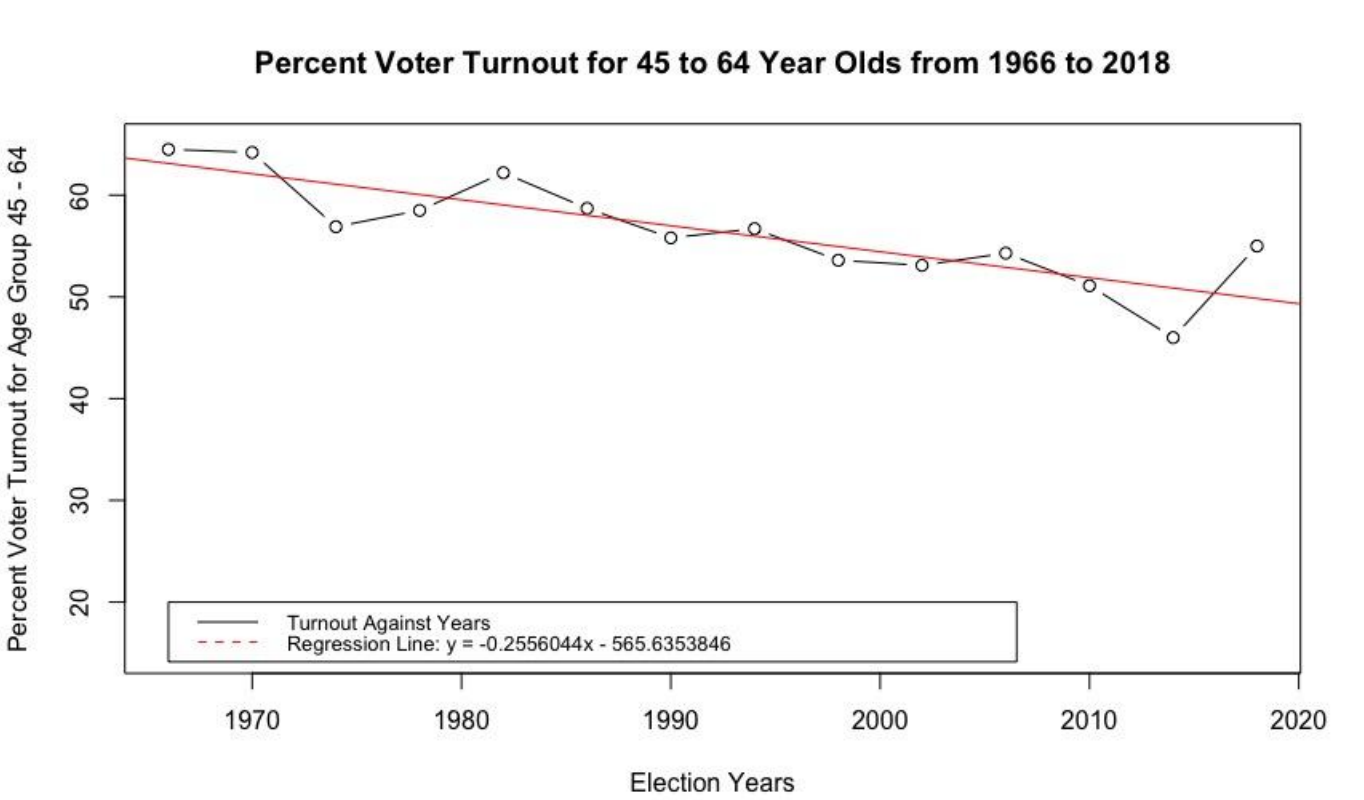
Results cont.

```r
# ages 45 to 64 voter turnout

# reading data from voterInfo table
ages45to64NEW <- voterInfoNEW$ages45to64
ages45to64NEW <- ages45to64NEW[1:14]
ages45to64NEW
typeof(ages45to64NEW)

votingYearsAgeNEW <- voterInfoNEW$yearsAge
votingYearsAgeNEW <- votingYearsAgeNEW[1:14]
votingYearsAgeNEW
typeof(votingYearsAgeNEW)

ages45to64NEWT <- t(ages45to64NEW)
votingYearsAgeNEWT <- t(votingYearsAgeNEW)
typeof(ages45to64NEWT)
typeof(votingYearsAgeNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$ages45to64 ~ voterInfoNEW$yearsAge)
```

```r
# plotting the data
plot(votingYearsAgeNEWT, ages45to64NEWT, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for Age Group 45 - 64",
     main = "Percent Voter Turnout for 45 to 64 Year Olds from 1966 to 2018",
     xlim = c(1966, 2018),
     ylim = c(15, 65))

# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1966, 20, legend = c(
  "Turnout Against Years", "Regression Line: y = -0.2556044x - 565.6353846"),
  col = c("black", "red"),
  lty = 1:6, cex = 0.8)
```

Results cont.

Ages 45 to 64 voter turnout



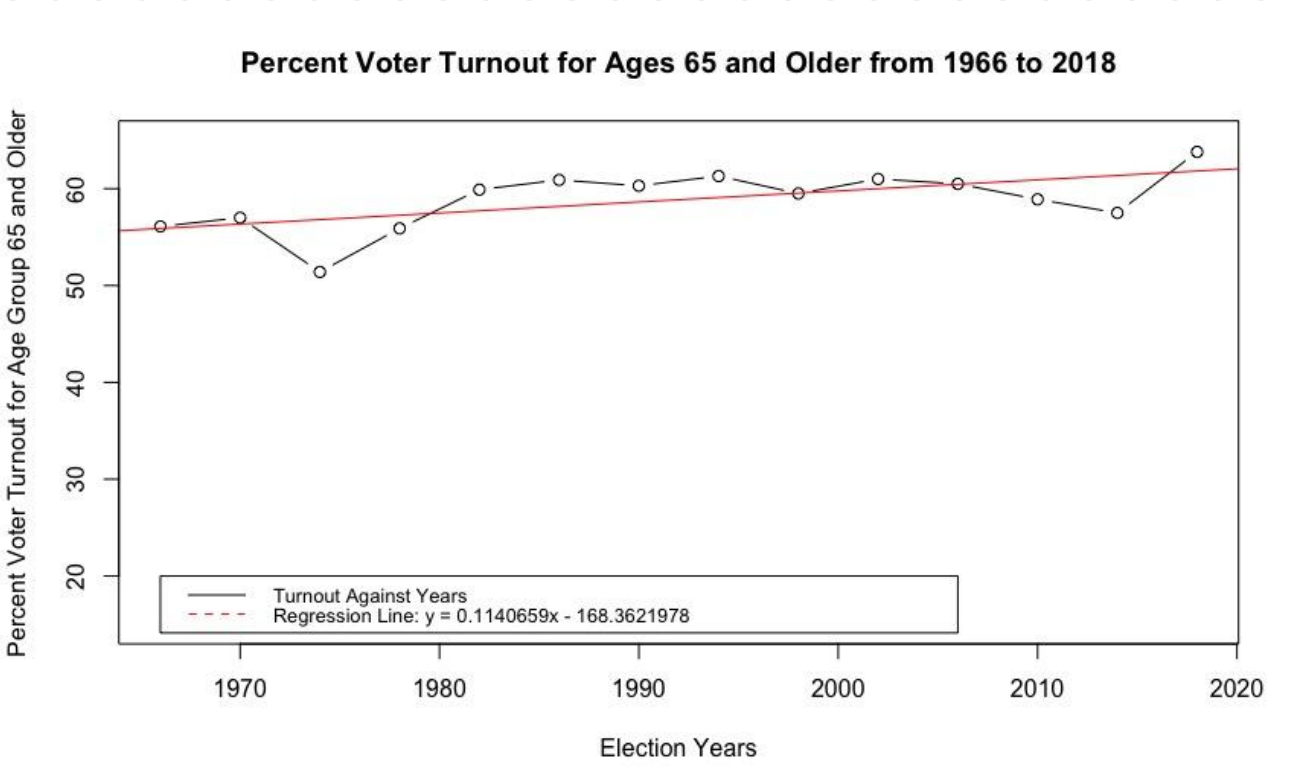
Results cont.

```r
# ages 65+ voter turnout

# reading data from voterInfo table
agesAbove65NEW <- voterInfoNEW$agesAbove65
agesAbove65NEW <- agesAbove65NEW[1:14]
agesAbove65NEW
typeof(agesAbove65NEW)

votingYearsAgeNEW <- voterInfoNEW$yearsAge
votingYearsAgeNEW <- votingYearsAgeNEW[1:14]
votingYearsAgeNEW
typeof(votingYearsAgeNEW)

agesAbove65NEWT <- t(agesAbove65NEW)
votingYearsAgeNEWT <- t(votingYearsAgeNEW)
typeof(agesAbove65NEW)
typeof(votingYearsAgeNEWT)

# creating a linear model of the data
model <- lm(voterInfoNEW$agesAbove65 ~ voterInfoNEW$yearsAge)
```

```r
# plotting the data
plot(votingYearsAgeNEWT, agesAbove65NEW, type = "b",
     xlab = "Election Years",
     ylab = "Percent Voter Turnout for Age Group 65 and Older",
     main = "Percent Voter Turnout for Ages 65 and Older from 1966 to 2018",
     xlim = c(1966, 2018),
     ylim = c(15, 65))

# graphing the regression line
abline(model, col = "red", lwd = 1)

# finding the regression line
model$coefficients

# summary of data statistics
summary(model)

# creating a legend for the graph
legend(1966, 20, legend = c(
  "Turnout Against Years", "Regression Line: y = 0.1140659x - 168.3621978"),
  col = c("black", "red"), lty = 1:6, cex = 0.8)
```

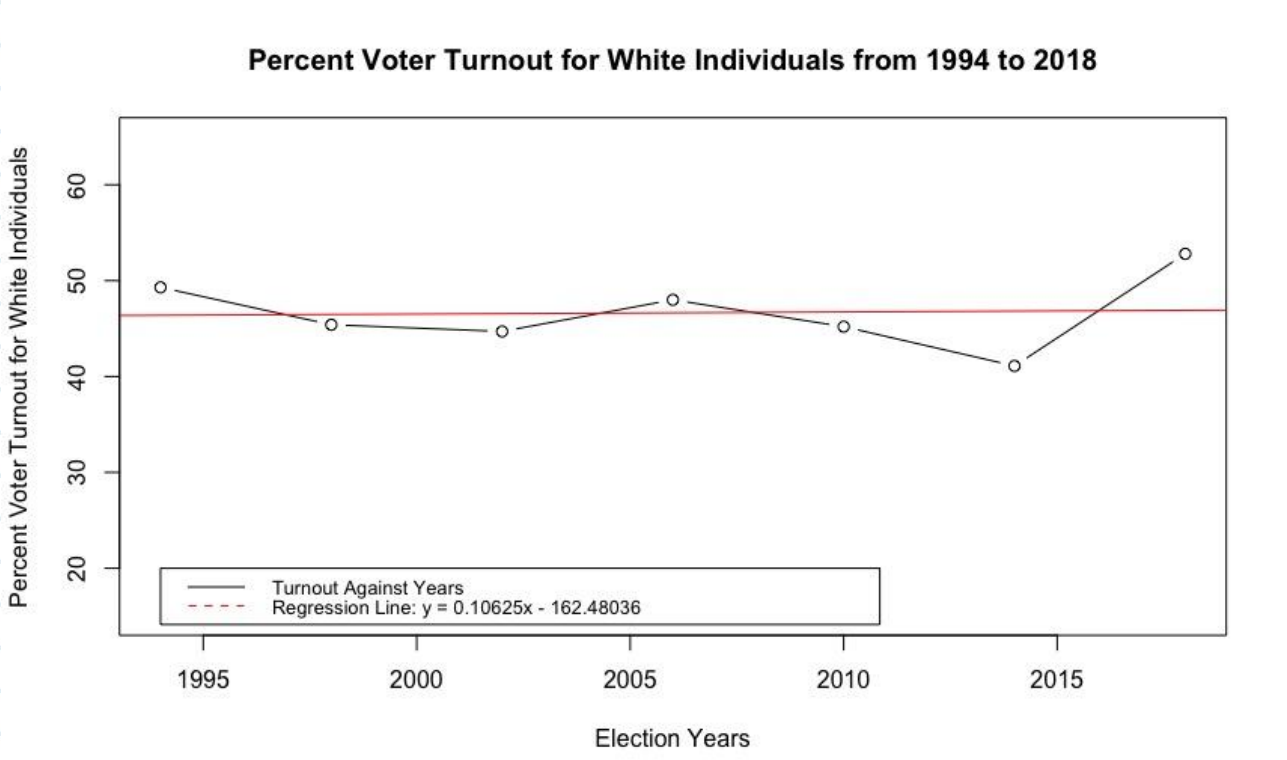
Results cont.

Ages 65+ voter turnout



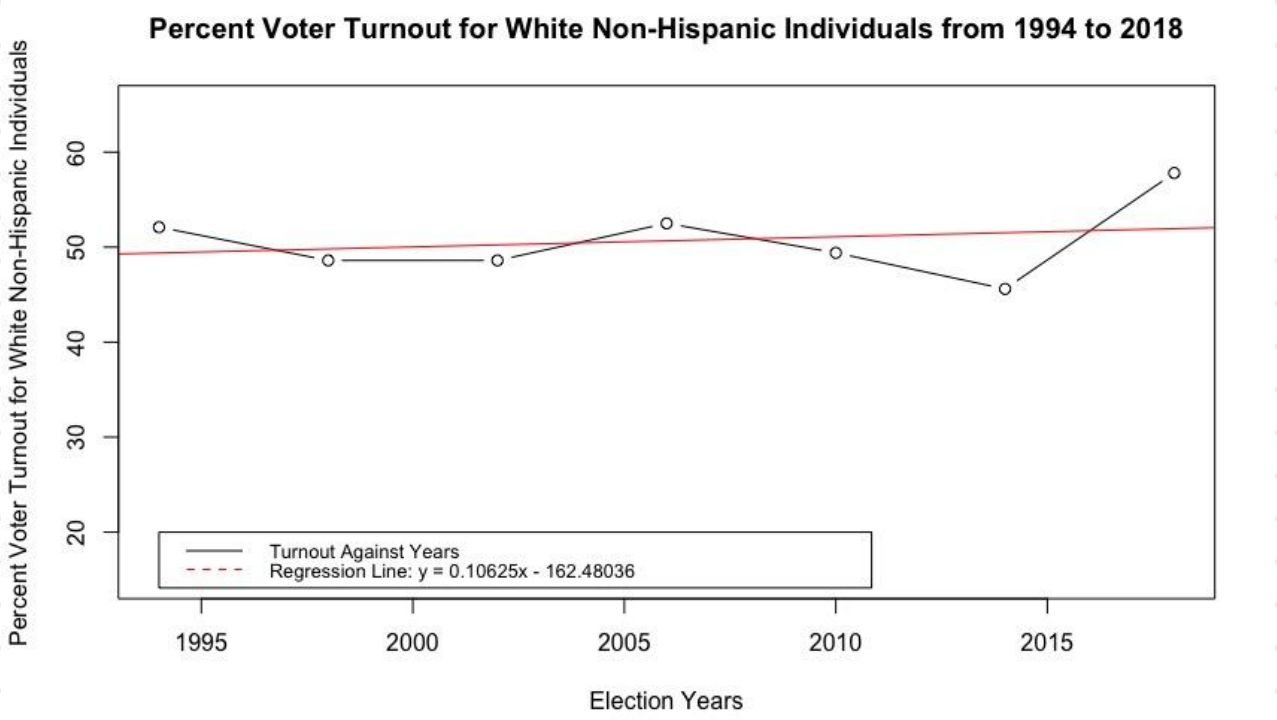
Results

White individual voter turnout



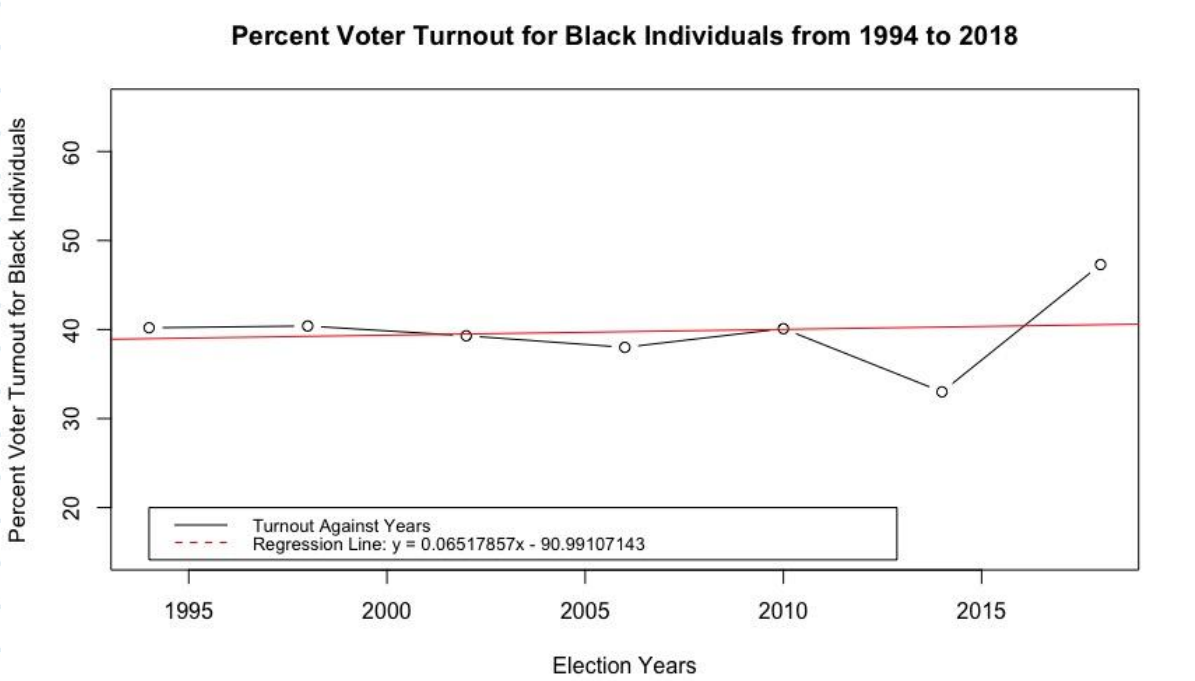
Results cont.

White Non-Hispanic individual voter turnout



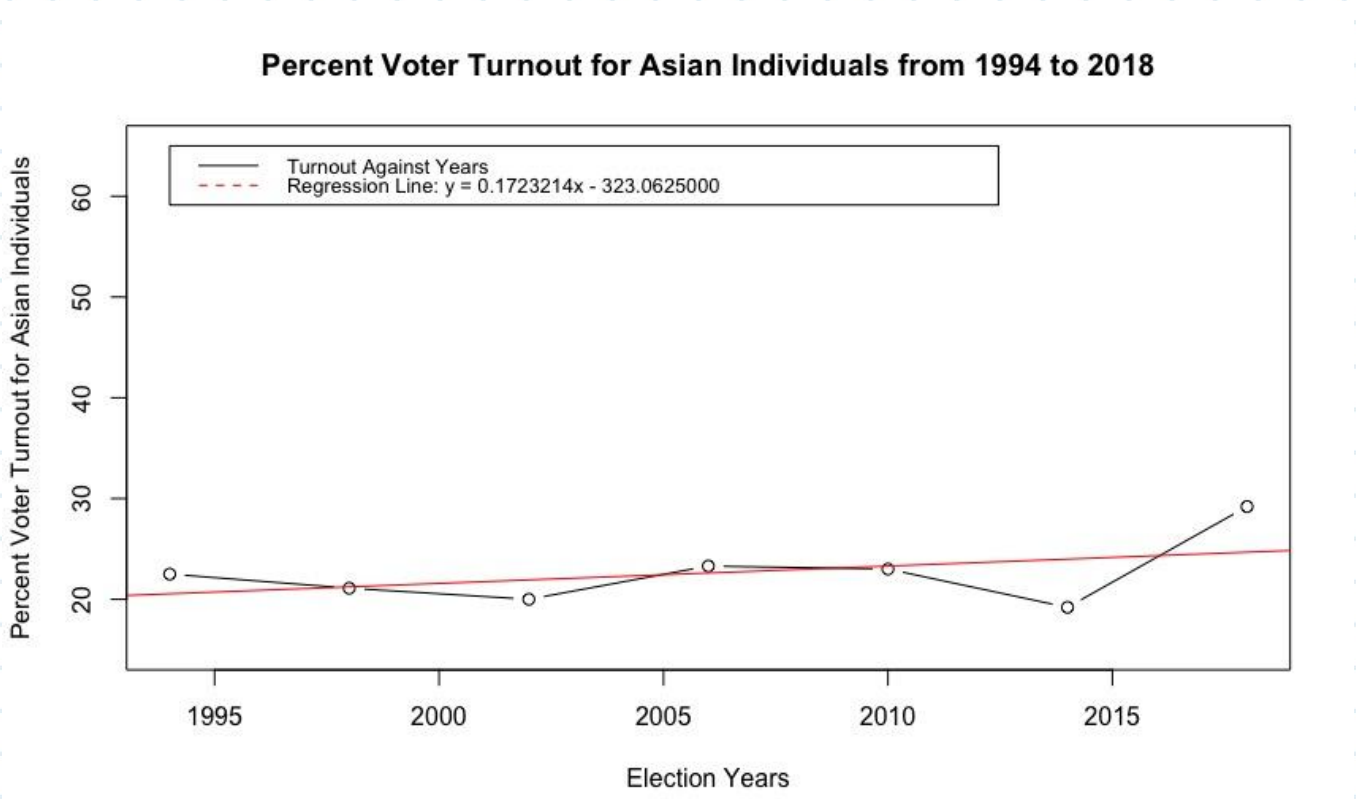
Results cont.

Black individual voter turnout



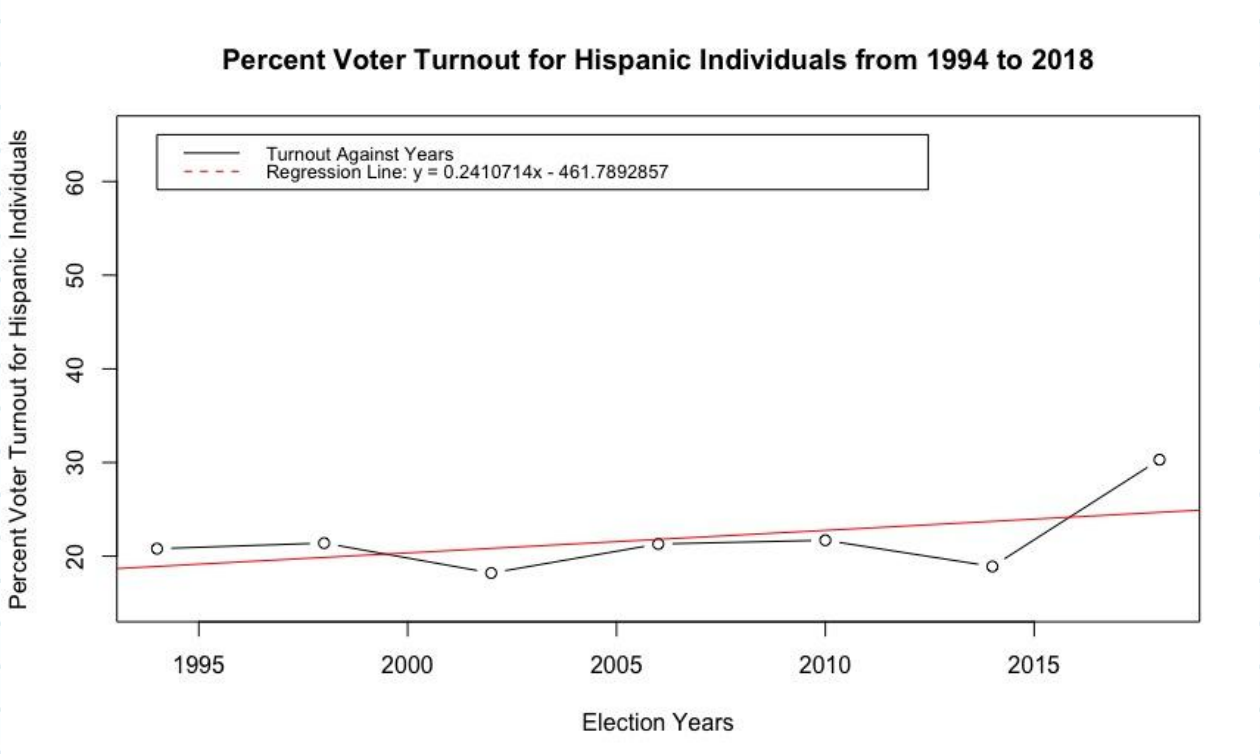
Results cont.

Asian individual voter turnout



Results cont.

Hispanic individual voter turnout



Results cont.

```
# creating a data table of correlation coefficients for regression lines
install.packages("data.table")
library("data.table")

x1 <- c("White", "WhiteNH", "Black", "Asian", "Hispanic",
        "Ages18-24", "Ages25-44", "Ages45-64", "Ages65+")
x2 <- c(0.002422, 0.0549, 0.01786, 0.2053, 0.2739,
        0.2759, 0.7207, 0.7168, 0.386)
# Create data.table
dt_1 <- data.table(" " = x1, "Correlation Coefficent" = x2)
dt_1
```

Results cont.

		Correlation Coefficent
1:	White	0.002422
2:	WhiteNH	0.054900
3:	Black	0.017860
4:	Asian	0.205300
5:	Hispanic	0.273900
6:	Ages18-24	0.275900
7:	Ages25-44	0.720700
8:	Ages45-64	0.716800
9:	Ages65+	0.386000

Results cont.

- We compared the linear relationship between voter turnout rates and age groups, or races during the duration of 1966 to 2018.
- Slopes for racial voter turnout:
  - The slope of the voter turnout rate of white is 0.02142857.
  - The slope of the voter turnout rate of white non hispanic is 0.10625.
  - The slope of the voter turnout rate of black is 0.06517857.
  - The slope of the voter turnout rate of Asian is 0.1723214.
  - The slope of the voter turnout rate of Hispanic is 0.2410714.
- Slopes for age voter turnout:
  - The slope of the voter turnout for 18-24 year olds is -0.1591758
  - The slope of the voter turnout for 25-44 year olds is -0.3589011
  - The slope of the voter turnout for 45-64 year olds is -0.2556044
  - The slope of the voter turnout for 65 year olds and older is 0.1140659
- It also allows us to predict future turnout rates based on past data.

Results (Age Groups)

- The age groups **with the highest consistent turnout rates** from the years 1966 to 2018 are voters who are between **45-65 years old and 65 or older.**
- While the age group who have consistently the **lowest** turnout rate from the years 1966 to 2018 are people who are **between the ages of 18-25.**
- The age group with the greatest slope from the years 1966 to 2018 is voters who are 65 years old or older.
- Meanwhile, the other age groups have a negative slope, with voters 25-44 years old having a decreasing voter turnout.

Results (Race Groups)

- The race group with the highest consistent turnout rates from years 1966 to 2018 are voters who are **white non-hispanic.**
- While the race group who have consistently been at the lowest turnout rate from years 1966 to 2018 are voters who are **hispanic and asian.**
- The race category **with the greatest slope** from the years 1966 to 2018 are voters who are **Hispanic**, while the group with the smallest increase is the white population.
- The white population has a **consistent voter turnou**t throughout those years, along with outwinning any other category when it comes to total percent turnout.

Proposed Solutions

1. **Schools promoting voting and raising awareness on the importance in early high school**
- Making it easily accessible by giving online option and having lots of stations open
- Sending voting information by mail to younger generation
- Incentivize colleges/universities & workplaces to have their students/employees vote
    - Could be a tax deductible per vote
2. **Make “Election Day” more than one day**
    - Increase early voting initiatives in states
    - Allow in-person voting on more than one day
    - Make election day a holiday for those in the labor force and college students
    - To alleviate the problem of voters being busy
3. **Universal Absentee voting**
    - Voters have two options
        - Mail-in
        - Drop off ballot

Improvements

1. The population considered is the voting age population, rather than the citizen voting age population examined in most recent reports (could have possibly had a dramatic change).
2. Data from pre-pandemic (Early Voting-Absentee Ballots)
3. Various situations by state
4. Different Groups
5. Little data available for race groups

Conclusions

We have developed a linear regression model about the growth and decay of voter turnout rates of ages and races groups. We used the data from Governmental from 1966 to 2018 to resolve how to increase young voter and racially minority groups' vote rates. We used R language to solve our mathematical model and give graphs to show how age and races group voter turnout rates are increasing or decreasing. We compared our data through the table of correlations of coefficients. We would like to use our linear regression model as a predictor for the 2022 Congressional Election. At this time, data is not yet available to the 2022 Congressional Election, so we cannot yet predict the data for future Congressional Elections.

Bibliographies

DeSilver, D. (2020, October 13). *Mail-in voting became much more common in 2020 primaries as COVID-19 spread*. Pew Research Center. Retrieved December 1, 2022, from https://pewrsr.ch/2GYaEGS

Degan, A., & Merlo, A. (2011). A STRUCTURAL MODEL OF TURNOUT AND VOTING IN MULTIPLE ELECTIONS. *Journal of the European Economic Association*, *9*(2), 209–245. http://www.jstor.org/stable/25836065

Bureau, U.S.C. 2021, *Historical reported voting rates*, Census.gov. Table A-7


FIND REGRESSION LINE
FROM THE NUMBERS FROM R


t-test
ANOVA
Correlation
Regression
etc.
The general linear model

Thank You!