Abstract
This study investigates the connection/correlation between education levels and voter turnout rates in the United States, with additional attention paid to potential outlying factors such as population density and non-citizen demographics. Utilizing publicly available datasets from reputable sources such as the National Center for Education Statistics and the U.S. Census Bureau, we analyze voter turnout data from the 2000 and 2008 general elections alongside education attainment statistics. Our findings suggest a nuanced relationship between education levels and voter engagement, as well as denoting the potential impact of factors such as population density and non-citizen demographics among others on electoral participation.

Introduction

In the United States, voter turnout is a hot button issue, particularly during presidential election years, which occur every four years. Focus is placed on encouraging eligible citizens to vote, in order to ensure the voices of the public be heard, and uphold foundations of democracy. The high visibility of presidential elections brings about common discussions and debates amongst fellow members of the population, and commonly leads to the highest voter turnouts in comparison to midterm and local elections. Voter registration drives, early voting, and mail-in ballots have raised voter turnout significantly in recent years, but even so, voter turnout percentages in the U.S. rank among the lowest in comparison to other developed countries. We began to explore what other factors might contribute to varying voter turnout rates within the U.S. The main factor we settled on researching further was education levels, and how they may contribute to voter turnout rates.

1. Does population density impact voter turnout or education level?
   - Connor's research question context:
When thinking about education levels and voter turnouts, we recognized how varying factors may contribute to a skew in either collection of data. For instance, the distribution of resources may be vastly different from county to county, or even state to state, with a denser or wealthier population receiving access to far greater resources. To investigate this possibility, we sized up our data in voter turnout and education level in comparison to the population sizes of each state. In doing so we hoped it might reveal any correlations or patterns that could shed light on the relationship between resource distribution, education levels, and voter turnout in the United States. The returns from this data could lead to further resources being allocated to address disparities and improve voter engagement in areas where it is most needed. This could inform policy decisions aimed at promoting equitable access to education and civic participation, ultimately strengthening democracy and ensuring that all voices are heard in the electoral process.

2. Do states with above average non-citizen population percentages carry below average voter turnout rates?
    ○ Thuwaybah research question context:

The relationship between non-citizen population percentages and voter turnout rate is primarily grounded in several factors, all of which typically tends to result in lower levels of political engagement among citizens. For non-citizens, there are more barriers immensely preventing them from voting and actively participating in civic engagement. Although the factors may vary by state or region, it is common that language barriers, legal restrictions or legislative limitations on non-citizen voters, and socio-economic conditions are frequently cited as contributing reasons. Understanding the intricate relationships between non-citizen voting, education level, and voter turnout rates is crucial, as they are all interdependent and impact each other. This is better put into perspective when we consider the fact that a state or region with a high non-citizen population and a low education level, for instance, would have a significantly lower voter turnout rate than a state or region with a high non-citizen population and a somewhat higher education level.

3. Do states with above average college education rates carry above average voter turnout rates?
    ○ Haruka research question context:

This research question was born out of a deep interest in the link between educational attainment and political participation. Although voter registration drives, advance voting, and mail-in ballots have significantly increased voter turnout in recent years, the U.S. still has one of the lowest voter turnout rates compared to other developed countries.

Based on the general observation that people with higher education are more politically active, we are motivated to understand specifically how a college education affects democratic participation. Identifying the relationship between education and voting behavior is also important for justifying policies that improve access to education and promote a more inclusive democracy.

4. Do states with more young HS graduates have higher voter turnout?
    ○ Coby research question context:

When considering voter turnout, we should also consider the effect of the education people receive prior to being able to vote. We wondered if high school educated people tended to vote more, as well as if individuals who recently graduated high school made a more noticeable impact on voter turnout.

Data

We chose our datasets because we felt they would be the best for helping us answer our research questions about voter turnout and education rates across each state. Our datasets originate from U.S. Government sources and are publicly available.

Our education attainment datasets come from the National Center for Education Statistics, which is an agency of the U.S. Department of Education. The data was collected by the U.S. Department of Education. This dataset includes important data about education attainment percentages for each state, including columns like % population who are high school graduates, % of population with a bachelor's degree. The population is the country (U.S.) and each row represents the statistics for a state. The datasets are each 52 rows by 7 columns. The dataset can be found at this link: https://nces.ed.gov/programs/digest/d10/tables/dt10_011.asp
The cleaned datasets can be found at these links:
2000 Edu Attainment:
https://docs.google.com/spreadsheets/d/1_D9R2Fr7c9lYAIh0QLawci_ILAza59bHpGKqQyKaZWo/edit?usp=sharing
2008 Edu Attainment:
https://docs.google.com/spreadsheets/d/1RdJ6QnPMmeniO17G4phgqTV5q66x2PyhqvUFG3hwJPY/edit?usp=sharing

Our election statistics were compiled from a 3rd party, The United States Elections Project (run by a professor at the University of Florida), who pulled the data from the U.S. Census Bureau. The data was collected by the U.S. Census Bureau. This dataset includes data about voter turnout by state, including the percentage of ballots received out of the total population, the total number of ballots counted, the total voter eligible population, and more. The population is the country (U.S.) and each row represents the statistics for a state. The datasets are each 52 rows by 12 columns. The website for the compiled data can be found at this link:
https://www.electproject.org/election-data/voter-turnout-data
The cleaned datasets can be found at these links:
2000 General Election:
https://docs.google.com/spreadsheets/d/1nmik6IsElo9aNSbLYyddII5MHKh7mAwyrSDizKXctDo/edit?usp=sharing
2008 General Election:
https://docs.google.com/spreadsheets/d/1V1NLVwNXoO1EIz5sG3SjbUyn7uT4bHGCN8S4K9OSDfg/edit?usp=sharing

We did encounter data quality issues with some missing data entries, but we managed to bypass this by performing our own calculations (e.g. missing percentage of ballots, but we had total ballots and total population, so we can calculate it ourselves). One concern is the age of the

datasets, since we were unable to find education attainment data from recent years. Fortunately, if we were to find more recent attainment data, we could easily improve the project since we have access to recent election data.

- Datasets
  - [2000 General Election](#)
  - [2008 General Election](#)
  - [2000 Edu Attainment](#)
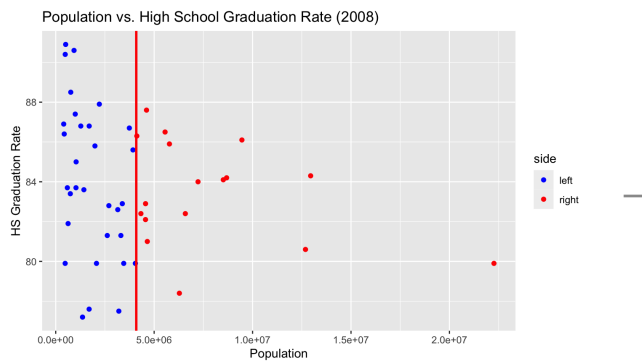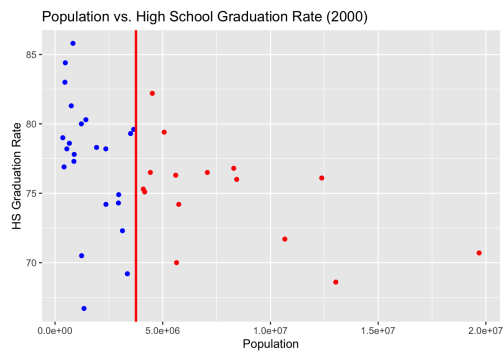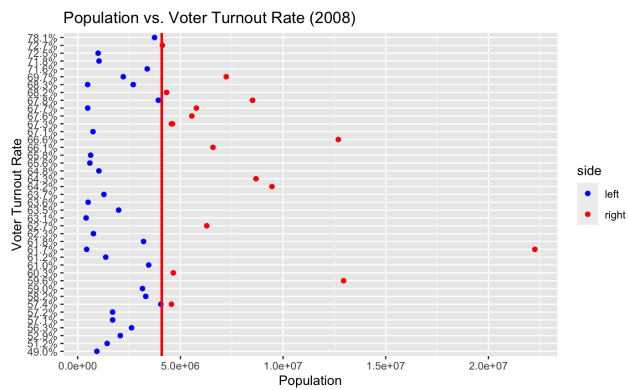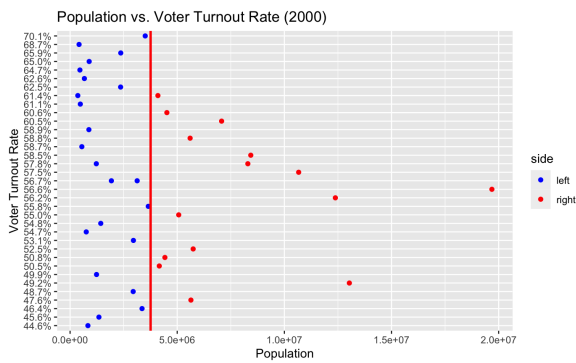 - [2008 Edu Attainment](#)

Methods

Before loading the data into R, we had to remove formatted rows and columns using google sheets. This included things like descriptions, footnotes, and empty columns used for formatting. Once all of our datasets were clean, we were able to load the data into R.

Specifically for the educational attainment dataset, since both the 2000 and 2008 data was on the same dataset, we had to split up the data into two separate datasets.
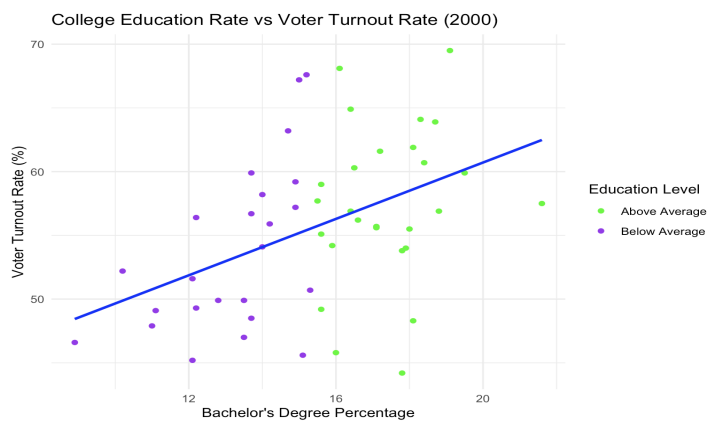
After loading the datasets into R, we created two merged datasets for 2000 and 2008 by the column "State". We chose to include every column of the original datasets when merging and did not choose to replace missing data since we felt the columns with missing data were unnecessary. These merged datasets now had 52 rows and 19 columns, where each row represented a state's voter turnout data and educational attainment data.

Results

1.

Population vs. Voter Turnout Rate (2000)

Population vs. Voter Turnout Rate (2008)

Population vs. High School Graduation Rate (2000)

Population vs. High School Graduation Rate (2008)

2.

College Education Rate vs Voter Turnout Rate (2000)

College Education Rate vs Voter Turnout Rate (2008)

3.

2000 Voter Turnout by % of HS Graduate 18-24 year olds

2008 Voter Turnout by % of HS Graduate 18-24 year olds

State Voter Turnout (% of total Voting Population)

% of 18-24 year olds that Graduated High School

4.

Voter Turnout Rate and Non-Citizen Population in 2000

Percent of Non-Citizens

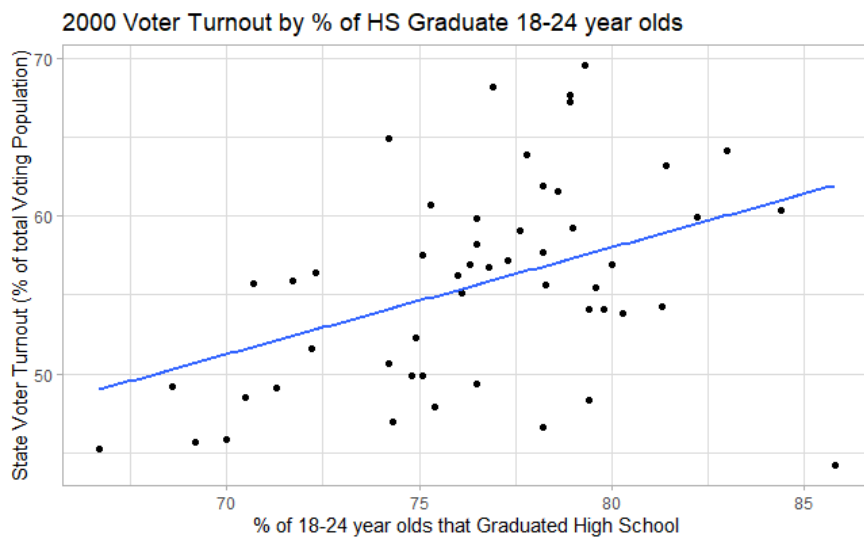Voter Turnout Rate
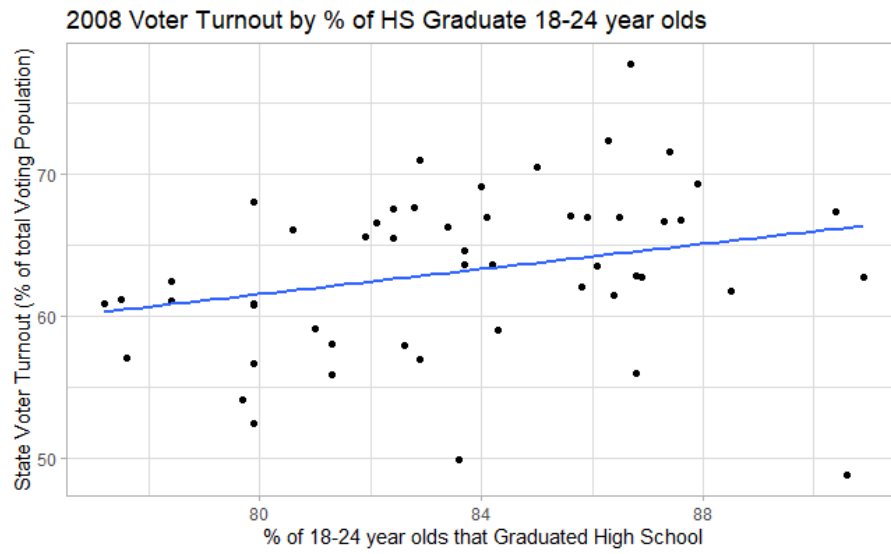
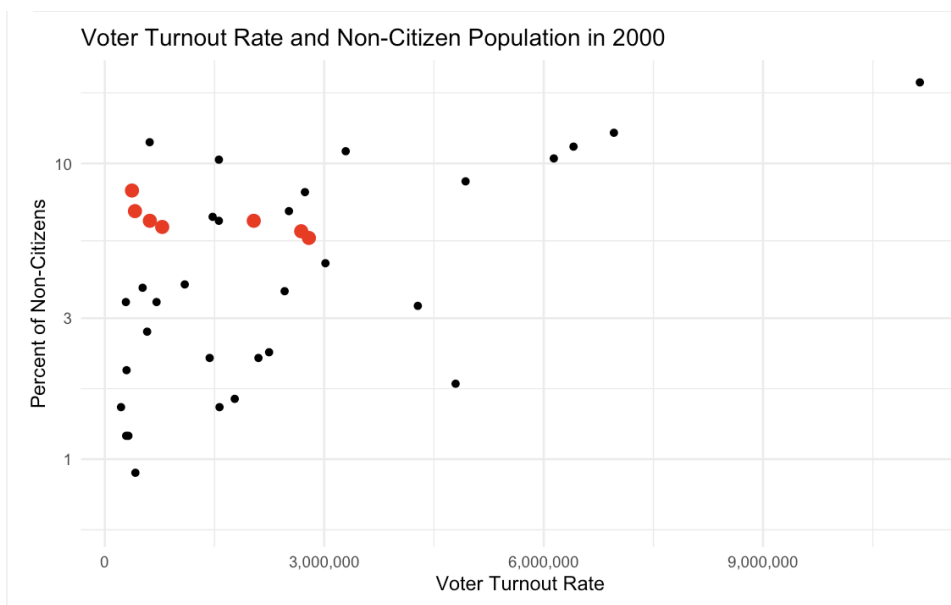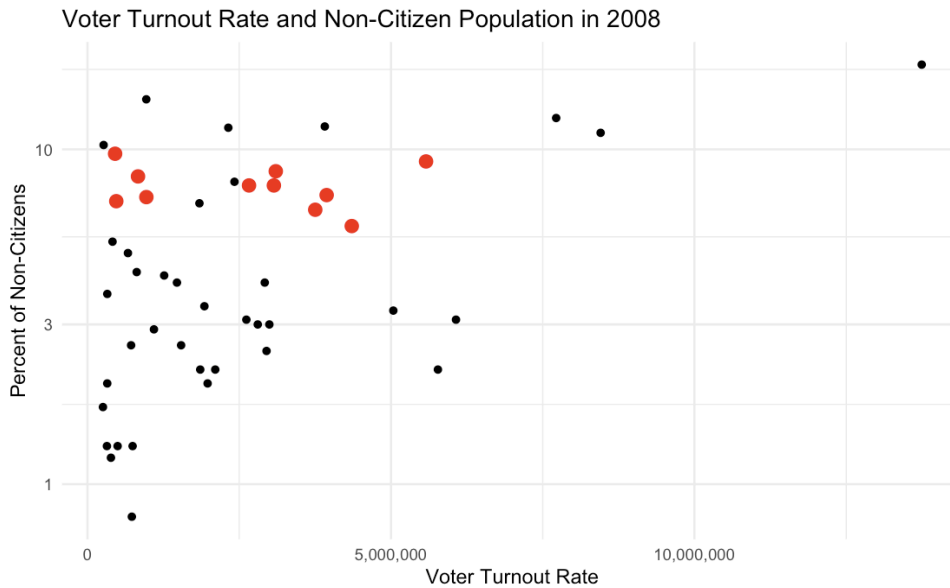**Voter Turnout Rate and Non-Citizen Population in 2008**

Findings

Connor Research Question Findings: I found a slight negative correlation among states ranking above the mean population (colored red on graphs) similarly shown in data from the years 2000 and 2008, as both voter turnout and educational attainment level seemed to decline with larger populations, though even this localized correlation was quite weak, and the overall data showed little to no correlation. This shows that the relationship between population density, education attainment, and voter turnout may not be uniform across all states. Instead, there could be variations based on population size and density. I feel these data sets may be slightly limited in defining these variations however, as the state level may be too broad of a population metric to extract meaningful insights. Looking forward, it might be valuable to conduct the same methods at the county level, or compare additional demographic types, such as socioeconomic status, income level, ethnicity, or urban/rural divide within each state.

Thuwaybah Question Findings:
I found a slightly positive correlation between the percentage of non-citizen per state and the voter turnout rate for that state. In the graphs above, the red dots indicate the states whose percentage of non-citizens are above average and whose voter turnout is below average. In the year 2000, there were seven such states and in 2008, there were 12 states. A deeper analysis of the plots from different years demonstrates that voter turnout is consistently low across all states regardless of the presence of non-citizens. The red dots being clustered in a similar area could potentially suggest similarities in a particular region (i.e., PNW, Midwest, East, South) or within a similar demographic. Because voter turnout seems to be low in general, based on our data, it is evident that one's citizenship status is not a strong factor influencing their voting behavior. The act of voting has many multifaceted layers and while citizenship status is one part of the whole,

other related factors such as voter suppression laws, redlining, registration difficulty, and physical barriers all play a role in the low voter turnout depicted in the plots.

Haruka Research Question Findings:
We found that there was a slight correlation between voter turnout and the percentage of bachelor's degree holders. From the graph, we found that areas with a lower percentage of the degree holders had lower voter turnout and vice versa. This means that college education helps people to actively participate in politics.

Coby Research Question Findings: For all high school graduates, there was a positive correlation between percentage of graduates in the state population and voter turnout. This tells us that high school level education noticeably contributes to whether or not people vote in elections. However, when looking at a population of only 18-24 year old individuals, the correlation becomes weaker, perhaps showing that young people are less likely to vote in elections.

Summary:
Over the course of this project we worked to examine the relationship between education level and voter turnout. We utilized two data sets, from the National Center for Education Statistics and the U.S. Census Bureau to allow us to depict said relationship with the skills we have learned this quarter in Info 201. We sought to answer four research questions to draw meaningful conclusions from our findings, and present opportunities for future research in the field. After carefully examining our data we were able to generate graphs and statistical support to draw from in answering these questions. First, we found a strong positive correlation in the connection between high school graduates and voter turnout at the state level, meaning states with a higher level of education tended to show a stronger voter turnout in data from both the 2000 and 2008 elections. Next, we found the same to be true of the connection between states' bachelor degree attainment rate and voter turnout rate, as states with higher populations of individuals with bachelor's degrees also showed higher voter turnout rates, in both the years 2000 and 2008. For our next two questions, we wanted to recognize the possibility of outlying factors in the collection of these data sets, as factors such as resource availability, socioeconomic status, and others can contribute to an individual's education level as well as their accessibility to voting options, especially in the pre mail-in era. More specifically, we first examined how population density might affect these variables, and though there was a slight negative correlation in above average states in terms of population, the overall data showed little to no correlation. Additionally, we examined whether the percentage of non-citizen members of the state populations had any impact on the voting turnout rate overall. Despite noticing some slight positive correlation in our returns, this was again somewhat insignificant, most likely due to the broad nature of our data, where more localized data might have had more to reveal. Future data would hopefully include data collected at the county level, or introduce further demographics to explore outlying factors' impact on these variables.

Looking forward we would hope our research could be built upon to learn more about voter demographics and the connection between education and overall voter turnout. We hope to see the trend of increased voter turnout continue into the future, and for support to be given to underserved populations through further policy integration and resource allocation.

<u>Coby Code:</u>
- **Data Initialization**

```{r}
library(tidyverse)
GE2000 <- read_delim("2000 November General Election - Turnout Rates.csv")
GE2008 <- read_delim("2008 November General Election - Turnout Rates.csv")
EDAT2000 <- read_delim("EduAttainment2000CLEAN - Sheet1.csv")
EDAT2008 <- read_delim("EduAttainment2008CLEAN - Sheet1.csv")
```

```{r}
vote2000 <- inner_join(GE2000, EDAT2000, by="State")
vote2008 <- inner_join(GE2008, EDAT2008, by="State")
names(vote2000) <- make.names(names(vote2000))
names(vote2008) <- make.names(names(vote2008))
```

```{r}
# New Numerical Variable (% of High schoolers getting a Bachelors)
vote2000$HStoBach <- vote2000$TOTBACH/vote2000$HS * 100
vote2008$HStoBach <- vote2008$TOTBACH/vote2008$HS * 100
# New Summarization with new Categorical Variable
sum2000 <- vote2000 %>%
  group_by(State, HS, YoungHS=(X18.24HSCompleters), HStoBach,
       Turnout=(Highest.Office / Voting.Eligible.Population..VEP.*100)) %>%
  summarize(HSAboveMean = HS > mean(HS))
sum2008 <- vote2008 %>%
  group_by(State, HS, YoungHS=(X18.24HSCompleters), HStoBach,
       Turnout=(Highest.Office / Voting.Eligible.Population..VEP.*100)) %>%
  summarize(HSAboveMean = HS > mean(HS))
```

- **Graphing**
    - **2000 all HS**

```{r}
ggplot(sum2000, aes(HS, Turnout))+
  geom_point() +
```

```
  geom_smooth(method="lm", formula=y~x, se=FALSE) +
  theme_light() +
  labs(
    title="2000 Voter Turnout by % of HS Graduate",
    x="% of population that Graduated High School",
    y="State Voter Turnout (% of total Voting Population)"
  )
```

- **2000 18-24 HS**

```{r}
ggplot(sum2000, aes(YoungHS, Turnout)) +
  geom_point() +
  geom_smooth(method="lm", formula=y~x, se=FALSE) +
  theme_light() +
  labs(
    title="2000 Voter Turnout by % of HS Graduate 18-24 year olds",
    x="% of 18-24 year olds that Graduated High School",
    y="State Voter Turnout (% of total Voting Population)"
  )
```

- **2008 all HS**

```{r}
ggplot(sum2008, aes(HS, Turnout)) +
  geom_point() +
  geom_smooth(method="lm", formula=y~x, se=FALSE) +
  theme_light() +
  labs(
    title="2008 Voter Turnout by % of HS Graduate",
    x="% of population that Graduated High School",
    y="State Voter Turnout (% of total Voting Population)"
  )
```

- **2008 18-24 HS**

```{r}
ggplot(sum2008, aes(YoungHS, Turnout)) +
  geom_point() +
  geom_smooth(method="lm", formula=y~x, se=FALSE) +
  theme_light() +
  labs(
    title="2008 Voter Turnout by % of HS Graduate 18-24 year olds",
```

```
    x="% of 18-24 year olds that Graduated High School",
    y="State Voter Turnout (% of total Voting Population)"
  )
```

Thuwaybah's Code

```{r message=FALSE}
data2000 <- read_delim("../Labs/Group Proj/2000 November General Election - Turnout
Rates.csv")
data2008 <- read_delim("../Labs/Group Proj/2008 November General Election - Turnout
Rates.csv")
```

```{r}
data2000_percent_non_citizen <- parse_number(data2000$`% Non-citizen`)
data2000_percent_turnout <- parse_number(data2000$`VEP Total Ballots Counted`)
```

```{r}
non_citizen_avg2000 <- mean(data2000_percent_non_citizen, na.rm=TRUE)
turnout_avg2000 <- mean(data2000_percent_turnout, na.rm=TRUE)
```

```{r}
non_citizen_above_avg <- data2000$`% Non-citizen` > non_citizen_avg2000
turnout_below_avg <- data2000$`VEP Total Ballots Counted` < turnout_avg2000
```

```{r}
sel <- data2000 %>%
  filter(non_citizen_above_avg & turnout_below_avg)
```

```{r}
sel$`% Non-citizen` <- as.numeric(gsub("%", "", sel$`% Non-citizen`))
```

Plot 1
```{r}
ggplot(data = data2000, aes (x = `Total Ballots Counted`, y = data2000_percent_non_citizen)) +
```

```
  geom_point() +
  geom_point(data = sel, aes(y = `% Non-citizen`), color = "red", size = 3) +
  labs(
    title = "Voter Turnout Rate and Non-Citizen Population in 2000",
    x = "Voter Turnout Rate",
    y = "Percent of Non-Citizens"
  ) +
  theme_minimal() +
  scale_y_log10() +
  scale_x_continuous(labels = scales::comma)
```

```{r}
data2008_percent_non_citizen <- parse_number(data2008$`% Non-citizen`)
data2008_percent_turnout <- parse_number(data2008$`VEP Total Ballots Counted`)
```

```{r}
non_citizen_avg2008 <- mean(data2008_percent_non_citizen, na.rm=TRUE)
turnout_avg2008 <- mean(data2008_percent_turnout, na.rm=TRUE)
```

```{r}
non_citizen_above_avg2008 <- data2008$`% Non-citizen` > non_citizen_avg2008
turnout_below_avg2008 <- data2000$`VEP Total Ballots Counted` < turnout_avg2008
```

```{r}
sel_08 <- data2008 %>%
  filter(non_citizen_above_avg2008 & turnout_below_avg2008)
```

```{r}
sel_08$`% Non-citizen` <- as.numeric(gsub("%", "", sel_08$`% Non-citizen`))
```
```

**Plot 2**
```{r}
ggplot(data = data2008, aes(x = `Total Ballots Counted`, y = data2008_percent_non_citizen)) +
  geom_point() +
```

```
  geom_point(data = sel_08, aes(y = `% Non-citizen`), color = "red", size = 3) +
  labs(
    title = "Voter Turnout Rate and Non-Citizen Population in 2008",
    x = "Voter Turnout Rate",
    y = "Percent of Non-Citizens"
  ) +
  theme_minimal() +
  scale_y_log10() +
  scale_x_continuous(labels = scales::comma)
```

Haruka's code:

```{r}
library(tidyverse)
library(readr)
library(ggplot2)


edu_attainment_2008_path <-
'/Users/sunakawa/Downloads/INFO/INFO201/week9/EduAttainment2008.csv'
turnout_2008_path <-
'/Users/sunakawa/Downloads/INFO/INFO201/week9/2008TurnoutRates.csv'


edu_attainment_2008 <- read_csv(edu_attainment_2008_path)
turnout_2008 <- read_csv(turnout_2008_path)

turnout_2008 <- turnout_2008 %>%
  mutate(`VEP Highest Office` = as.numeric(str_remove(`VEP Highest Office`, "%")))

data_merged <- merge(edu_attainment_2008, turnout_2008, by = "State")

average_bach_rate <- mean(data_merged$BACH, na.rm = TRUE)
average_turnout_rate <- mean(data_merged$`VEP Highest Office`, na.rm = TRUE)

data_merged <- data_merged %>%
  mutate(Above_Avg_Bach = if_else(BACH > average_bach_rate, "Above", "Below"),
```

```
      Above_Avg_Turnout = if_else(`VEP Highest Office` > average_turnout_rate, "Above",
"Below"))

ggplot(data_merged, aes(x = BACH, y = `VEP Highest Office`, color = Above_Avg_Bach)) +
  geom_point() +
  geom_smooth(method = "lm", col = "blue", se = FALSE) +
  labs(title = "College Education Rate vs Voter Turnout Rate (2008)",
       x = "Bachelor's Degree Percentage",
       y = "Voter Turnout Rate (%)") +
  theme_minimal() +
  scale_color_manual(values = c("Above" = "green", "Below" = "purple"),
             name = "Education Level",
             labels = c("Above Average", "Below Average"))


```


```{r}
library(tidyverse)
library(readr)
library(ggplot2)

edu_attainment_2000_path <-
'/Users/sunakawa/Downloads/INFO/INFO201/week9/EduAttainment2000.csv'
turnout_2000_path <-
'/Users/sunakawa/Downloads/INFO/INFO201/week9/2000TurnoutRates.csv'

edu_attainment_2000 <- read_csv(edu_attainment_2000_path)
turnout_2000 <- read_csv(turnout_2000_path)

turnout_2000 <- turnout_2000 %>%
  mutate(`VEP Highest Office` = as.numeric(str_remove(`VEP Highest Office`, "%")))

data_merged <- merge(edu_attainment_2000, turnout_2000, by = "State")

average_bach_rate <- mean(data_merged$BACH, na.rm = TRUE)
average_turnout_rate <- mean(data_merged$`VEP Highest Office`, na.rm = TRUE)

data_merged <- data_merged %>%
```

```r
  mutate(Above_Avg_Bach = if_else(BACH > average_bach_rate, "Above", "Below"),
         Above_Avg_Turnout = if_else(`VEP Highest Office` > average_turnout_rate, "Above",
"Below"))

ggplot(data_merged, aes(x = BACH, y = `VEP Highest Office`, color = Above_Avg_Bach)) +
  geom_point() +
  geom_smooth(method = "lm", col = "blue", se = FALSE) +
  labs(title = "College Education Rate vs Voter Turnout Rate (2000)",
       x = "Bachelor's Degree Percentage",
       y = "Voter Turnout Rate (%)") +
  theme_minimal() +
  scale_color_manual(values = c("Above" = "green", "Below" = "purple"),
              name = "Education Level",
              labels = c("Above Average", "Below Average"))

```
```