My title*

My subtitle if needed

First author Another author

30 November 2023

First sentence. Second sentence. Third sentence. Fourth sentence.

```
election_data <- read.csv(here("outputs/data/election_data_clean.csv"))
covid_data <- read.csv(here("outputs/data/covid_data_clean.csv"))
acs_data <- read.csv(here("outputs/data/acs_data_clean.csv"))
covid_election_data <- read.csv(here("outputs/data/covid_election.csv"))
data <- read.csv(here("outputs/data/merged_data.csv"))
republican_data <- data %>%
  filter(party == "Republican")
```

1 Introduction

You can and should cross-reference sections and sub-sections.

The remainder of this paper is structured as follows. Section 2....

2 Data

In this paper, the data used mainly consist the voting patterns for 2020 US Federal Election, COVID infection rate and socio-economic variables in each county of US. The voting data is taken from the MIT Election Data Science Lab which includes the voting for each party in county level at each Federal Election from 2000 to 2020. For the COVID data, it is taken from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. The socio-economic data is taken from American Community Survey.

*Code and data are available at: LINK.

2.1 MIT Election Data Science Lab

The MIT Election Data Science Lab is a Lab at MIT which collect and analyzing the election data to supports advances in election science. Their collected data is not limited to the presidential election results, but also covers the midterm or US senate results in either state or county level.

The data abstracted from MITEDSL in this paper is the "County Presidential Election Returns 2000-2020." It contains the number of votes for each party and the total votes at each US county from 2000 to 2020 US Federal Election. In this paper, I will only use the data for the 2020 US Election and focus on the counties on the mainland of US. The Table 1 summarizes the important variables.

Table 1: The summary of voting patterns in 2020 US Federal Presidential Election

Party	Candidate	Total Votes	Mean Pct Vote	Median Pct Vote
Democrat	JOSEPH R BIDEN JR	81264648	0.21	0.17
Republican	DONALD J TRUMP	74218999	0.40	0.40
Libertarian	JO JORGENSEN	1810407	0.01	0.01
Other	OTHER	802586	0.00	0.00
Green	OTHER	381070	0.00	0.00

The Table 1 summarizes the voting patterns for each party during the 2020 US Presidential Election. Undoubtedly, the two most popular parties are the Democrat and the Republican. However, it seems that even though Republican has almost double mean percent vote than the Democrat, the total votes for Republican is less than the Democrat. This pattern may indicate that more people living at the high-population states such as California votes for Biden. This finding aligns with the fact that Biden defeated Trump in 2020. I will explain more latter.

2.2 Center for Systems Science and Engineering at JHU

The Center for Systems Science and Engineering at JHU is the center at the Department of Civils and Engineering to collect the local, national and global multidimensional data including medicine, health care, disaster response etc. During the pandemic, they collected the US and global COVID cases and deaths and report them on their GitHub. Their data is summarized by daily reports, ranging from April 12, 2020 to March 9, 2023.

To illustrate the impact of COVID on the 2020 Election to the greatest extend, the COVID data used in this paper will be the daily report on November 3, 2020 which is the Election

Table 2: **?(caption)**

```
# A tibble: 2 x 5
  winning_party  case deaths   inf  mort
  <chr>         <dbl>  <dbl> <dbl> <dbl>
1 Democrat      9245.  238.  3130.  76.9
2 Republican    1611.   31.9 3061.  62.5
```

Day [citation]. This report contains the aggregated cases and deaths for each each country and counties in US.

2.3 American Community Survey (ACS)

The American Community Survey is the survey conducted by the U.S. Census Bureau. The survey contains a variety of socio-economic variables for each county. Although there are different data tables from ACS, I will take 2021 five years estimate of DP02, DP03 and DP05.

These three tables cover the social, economic and demographic characteristics of each county in US. By referring my previous research where I found the variables which have the highest correlations to the COVID mortality rate, I will use the same variables in this paper. That said, I will extract the data regarding the educational attainment from DP02, especially the proportion of people having at least a bachelor degree. In terms of economic perspective (DP03), I will take the proportion of people with private health

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in **?@sec-model-details**.

3.1 Model set-up

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in **?@tbl-modelresults**.

```
Call:
glm(formula = high_infrate ~ prop_higher_education + pctile +
    no_insurance + private_insurance + males + age_85 + white_pct +
    black_pct, data = republican_data)

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)           -0.1211019  0.1734674  -0.698 0.485132
prop_higher_education -0.0035611  0.0010495  -3.393 0.000697 ***
pctile                -0.0017845  0.0003920  -4.552 5.44e-06 ***
no_insurance           0.0183269  0.0018110  10.120  < 2e-16 ***
private_insurance      0.0043248  0.0012460   3.471 0.000523 ***
males                  0.0097476  0.0027182   3.586 0.000339 ***
age_85                 0.0017823  0.0067744   0.263 0.792492
white_pct             -0.0035343  0.0008663  -4.080 4.58e-05 ***
black_pct              0.0044107  0.0008592   5.134 2.95e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.211391)

    Null deviance: 1234.2  on 4985  degrees of freedom
Residual deviance: 1052.1  on 4977  degrees of freedom
AIC: 6412.2

Number of Fisher Scoring iterations: 2
```

```r
library(Matching)
```

```
Loading required package: MASS


Attaching package: 'MASS'


The following object is masked from 'package:dplyr':

    select


##
##  Matching (Version 4.10-14, Build Date: 2023-09-13)
##  See https://www.jsekhon.com for additional documentation.
##  Please cite software as:
##   Jasjeet S. Sekhon. 2011. ``Multivariate and Propensity Score Matching
##   Software with Automated Balance Optimization: The Matching package for R.''
##   Journal of Statistical Software, 42(7): 1-52.
##
```

```r
# Propensity score matching
# mb <- MatchBalance(high_infrate ~ prop_higher_education + pctile +
#                              no_insurance +private_insurance + males + age_85 +
#                              white_pct + black_pct, data = data)
rr <- Match(Y = republican_data$pct_vote, Tr = republican_data$high_infrate,
            X = republican_data$prop_score, M = 1)
# mb <- MatchBalance(high_infrate ~ prop_higher_education + pctile +
#                               no_insurance +private_insurance + males + age_85 +
#                               white_pct + black_pct, match.out = rr,
#                     data = republican_data, nboots = 10)
summary(rr)
```

```
Estimate...  -0.088284
AI SE......  0.015535
T-stat.....  -5.6829
p.val......  1.3243e-08

Original number of observations..............  4986
```

```
Original number of treated obs...............  2245
Matched number of observations...............  2245
Matched number of observations  (unweighted).  13198
```

```r
# Propensity score regression adjustment
model_adj <- glm(pct_vote ~ high_infrate + prop_score, data = republican_data)
summary(model_adj)
```

```
Call:
glm(formula = pct_vote ~ high_infrate + prop_score, data = republican_data)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.623245   0.010126  61.547   <2e-16 ***
high_infrate  0.012240   0.008612   1.421    0.155    
prop_score   -0.502996   0.022422 -22.433   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.07803473)

    Null deviance: 432.86  on 4985  degrees of freedom
Residual deviance: 388.85  on 4983  degrees of freedom
AIC: 1437.4

Number of Fisher Scoring iterations: 2
```

```r
# Counterfactual Analysis
adjusted_data <- data %>%
  dplyr::select(state, county.x, fips, party, votes, total_votes, pct_vote, high_infrate)

adjust_function <- function(data){
  adjust_votes <- c()
  for (i in 1:nrow(data)){
    if (data[i,"high_infrate"] == 1){
      if (data[i,"party"] == "Republican"){
        adjust_votes[i] <- round((data[i,"pct_vote"] + 0.036928 ) * data[i,"total_votes"],
      }
      else if (data[i,"party"] == "Democrat"){
```

```r
        adjust_votes[i] <- round((data[i,"pct_vote"] - 0.036928 ) * data[i,"total_votes"],
      }
      else {adjust_votes[i] <- data[i,"votes"]}
    }
    else {
      adjust_votes[i] <- data[i,"votes"]
    }
  }
  return(adjust_votes)
}



adjusted_data %>%
  group_by(state, party) %>%
  summarise(votes = sum(votes), .groups = "drop")
```

```
# A tibble: 220 x 3
   state party          votes
   <chr> <chr>          <int>
 1 AK    Democrat       11821
 2 AK    Green            157
 3 AK    Libertarian      671
 4 AK    Other            128
 5 AK    Republican     10551
 6 AL    Democrat      849624
 7 AL    Other          32488
 8 AL    Republican   1441170
 9 AR    Democrat      423932
10 AR    Green           2980
# i 210 more rows
```

```r
adjust_function2 <- function(data) {
  adjust_votes <- numeric(nrow(data))  # Initialize the vector with the correct length
  extra_vote_percentage <- 0.036928

  for (i in 1:nrow(data)) {
    if (data[i, "high_infrate"] == 1) {
      total_votes <- data[i, "total_votes"]
      extra_votes <- round(extra_vote_percentage * total_votes, 0)
```

```r
      if (data[i, "party"] == "Republican") {
        adjust_votes[i] <- data[i, "votes"] + extra_votes
      } else {
        # Calculate total votes for non-Republican parties
        total_non_rep_votes <- sum(data[data$party != "Republican" & data$fips == data[i,

        # Distribute the vote loss proportionally
        if (total_non_rep_votes > 0) {
          party_vote_share <- data[i, "votes"] / total_non_rep_votes
          votes_lost <- round(extra_votes * party_vote_share, 0)
          adjust_votes[i] <- max(0, data[i, "votes"] - votes_lost)
        } else {
          adjust_votes[i] <- data[i, "votes"]
        }
      }
    } else {
      adjust_votes[i] <- data[i, "votes"]
    }
  }

  return(adjust_votes)
}

adjust_vote1 <- adjust_function(adjusted_data)
adjust_vote2 <- adjust_function2(adjusted_data)
adjusted_data$adjust_vote1 <- adjust_vote1
adjusted_data$adjust_vote2 <- adjust_vote2
```

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

6 References