The Correlation Between College Attainment in Presidential Vote Swings: An Analysis Using New England

John Curran

9 October 2020

Introduction

For this study, the swing in a vote preference between two presidential elections is compared using data from New England counties. *Vote swing* is defined as the difference between a party's margin of victory/loss in two distinct presidential elections. For example, if a Democrat wins a certain county by 5 in one election, and then loses by 3 in the next, the vote swing between the two elections was 8 points towards the Republican. In this study, a negative swing is a swing towards the Republican candidate, while a positive swing is a swing towards the Democratic candidate.

This small study breaks down the vote swings between the 2012 and 2016 elections and the vote swings between the 1992 and 1996 elections. Swings are broken down by county.

Dependent Variable: vote swings by county
Independent Variable: college attainment rates by county

The purpose of this study is to examine how the correlation between education and vote swing has changed in the past several decades. Two linear models will be created, one modeling education and vote swings in the 1990s and the other modeling education and vote swings in the 2010s.

Hypothesis: college attainment has become more correlated with vote preference swings in United States presidential elections since the 1990s.

Establishing Models

First, we will create linear models of the correlation between vote swings from 2012 to 2016 and 1992 to 1996, and college attainment rates in each New England county at the time of the given elections.

```
library(readxl)
statedata <- read_excel("~/NESwings.xlsx", range='A1:E68')

model2016 <- lm(SWING2016~I(COLLEGE2018*100), data=statedata)
model1996 <- lm(SWING1996~I(COLLEGE1990*100), data=statedata)
```

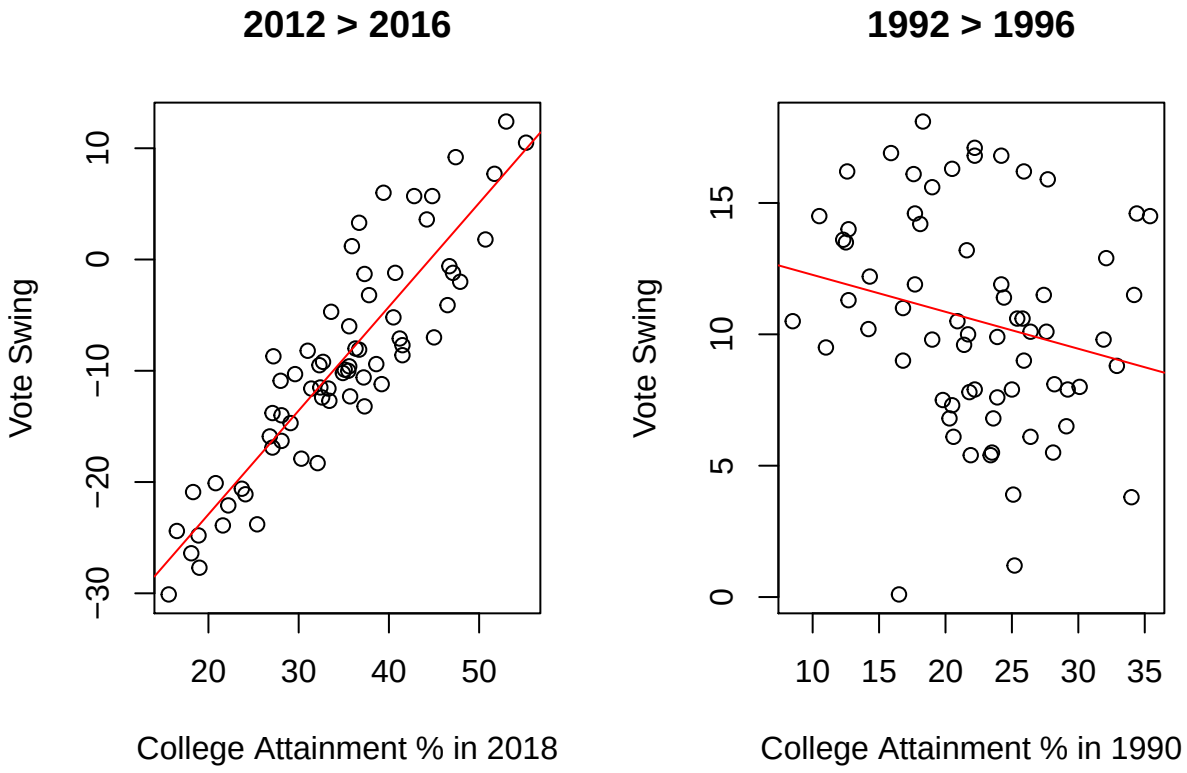
Next, we will plot and compare the results.

```
par(mfrow=c(1,2))

plot(statedata$COLLEGE2018*100, statedata$SWING2016, main="2012 > 2016",
     xlab="College Attainment % in 2018", ylab="Vote Swing")
abline(model2016, col="red")

plot(statedata$COLLEGE1990*100, statedata$SWING1996, main="1992 > 1996",
```

```
     xlab="College Attainment % in 1990", ylab="Vote Swing")
abline(model1996, col="red")
```



Analysis of the 2016 Model

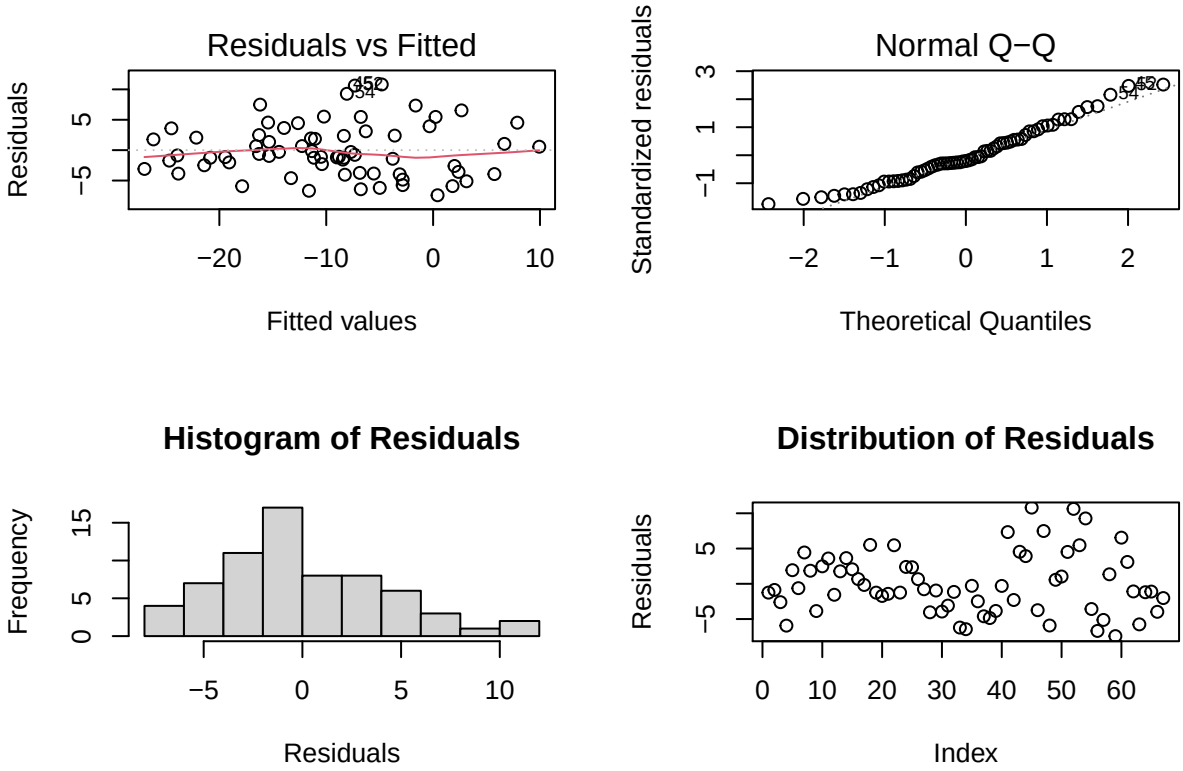
From a preliminary analysis, it appears that there is a far stronger correlation in the more recent election swing. We will now look at the summary statistics and diagnostic plots for the 2016 model.

```
summary(model2016)
```

```
##
## Call:
## lm(formula = SWING2016 ~ I(COLLEGE2018 * 100), data = statedata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.4187 -3.3349 -0.8705  2.4363 10.8055
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -41.56090    2.02130  -20.56   <2e-16 ***
## I(COLLEGE2018 * 100)   0.93288    0.05674   16.44   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.326 on 65 degrees of freedom
```

```
## Multiple R-squared:  0.8061, Adjusted R-squared:  0.8031
## F-statistic: 270.3 on 1 and 65 DF,  p-value: < 2.2e-16
```

```r
par(mfrow=c(2,2))
plot(model2016, 1)
plot(model2016, 2)
hist(model2016$residuals, main="Histogram of Residuals", xlab="Residuals")
plot(model2016$residuals, main="Distribution of Residuals", ylab="Residuals")
```



From an early glance at the summary statistics and diagnostic plots, there appears to be a slight correlation between college attainment and vote swing between the 2012 and 2016 presidential elections in New England. Specifically, the expected formula appears to be the following:

$$SWING = -41.5609 + 0.93288(EDUCATION)$$

This means that for every 1% increase in college attainment in a given county, it is expected that vote swing towards Hillary Clinton from Barack Obama would increase by 0.93288 points. Approximately 80% of the change in vote swing can be attributed to education, according to this model.

From this simple linear regression, the expected swing for a county with 0% college attainment would be 41.5609 points towards Donald Trump, and the expected swing for a county with 100% college attainment would be 51.7271 points towards Hillary Clinton. This is, of course, just an expected value and not a precise estimation.

Hypothesis Test

We will now test whether a correlation from 2012 to 2016 can be concluded, using a 5% level of significance.

Null Hypothesis: there is no correlation between education and vote swing
Alternative Hypothesis: there is a correlation between education and vote swing

Test Statistic: 16.44

We will be testing whether the test statistic is the critical region to the left of $t_{0.025,65}$ and right of $t_{0.975,65}$ in order to reject the null hypothesis.

```
qt(0.975, 65)
```

```
## [1] 1.997138
```

Given that t > 1.997138, and thus to the right of $t_{0.975,65}$, we are able to *reject* the null hypothesis and claim with 95% certainty that there exists a correlation between college attainment and vote swing from the 2012 to 2016 presidential elections.

Confidence Interval

After concluding that a correlation exists, we will now develop a 95% confidence interval for what that correlation is.

```
confint(model2016)
```

```
##                               2.5 %     97.5 %
## (Intercept)           -45.5977120 -37.524079
## I(COLLEGE2018 * 100)    0.8195535   1.046204
```

It can now be concluded with 95% confidence that the expected increase of vote swing towards Hillary Clinton from Barack Obama is between 0.8195535 and 1.046204 points for every 1% increase in college attainment.

Overall, college attainment can explain approximately 80% of the vote swing from the 2012 to 2016 presidential election in each county due to this model's R^2 value. It can also be concluded with 95% confidence that there exists a correlation between college attainment and vote swing, and that this swing is between 0.8195535 and 1.046204 points for every 1% increase in college attainment. Thus, there existed a correlation between vote swing and college attainment in the 2016 presidential election.

Analysis of the 1996 Model

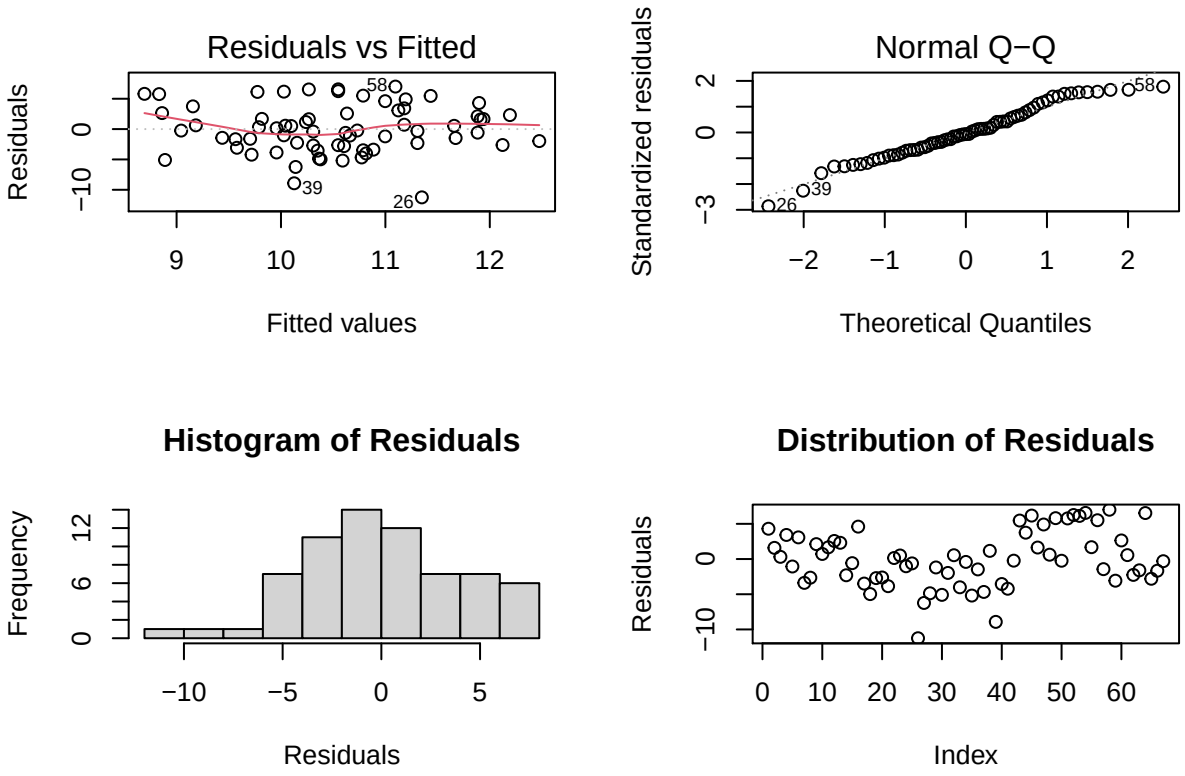
We will now move on to the 1992 > 1996 model in order to compare its findings with the 2012 > 2016 model. First, we will retrieve summary statistics and diagnostic plots for this model.

```
summary(model1996)
```

```
##
## Call:
## lm(formula = SWING1996 ~ I(COLLEGE1990 * 100), data = statedata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2494  -2.6786  -0.2438   2.6033   7.0036
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           13.66910    1.79157   7.630 1.33e-10 ***
## I(COLLEGE1990 * 100) -0.14059    0.07717  -1.822   0.0731 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.987 on 65 degrees of freedom
## Multiple R-squared:  0.04858,    Adjusted R-squared:  0.03394
```

```
## F-statistic: 3.319 on 1 and 65 DF,  p-value: 0.0731
```

```
par(mfrow=c(2,2))
plot(model1996, 1)
plot(model1996, 2)
hist(model1996$residuals, main="Histogram of Residuals", xlab="Residuals")
plot(model1996$residuals, main="Distribution of Residuals", ylab="Residuals")
```



From an early glance at the summary statistics and diagnostic plots, there appears to be very little correlation between college attainment and vote swing between the 1992 and 1996 presidential elections in New England. However, if there were a correlation, the expected formula appears to be the following:

$$SWING = 13.6691 - 0.14059(EDUCATION)$$

This means that for every 1% increase in college attainment in a given county, it is expected that vote swing towards Bob Dole from George HW Bush would increase by 0.14059. Approximately only 5% of the change in vote swing can be attributed to education, according to this model. This very small R^2 value points to no correlation, or at least an incredibly weak correlation.

From this simple linear regression, the expected swing for a county with 0% college attainment would be 13.6691 points towards Bill Clinton, and the expected swing for a county with 100% college attainment would be 14.059 points towards Bob Dole. This is, of course, just an expected value and not a precise estimation.

Hypothesis Test

We will now test whether a correlation from 1992 to 1996 can be concluded, using a 5% level of significance.

Null Hypothesis: there is no correlation between education and vote swing
Alternative Hypothesis: there is a correlation between education and vote swing

Test Statistic: -1.822

We will be testing whether the test statistic is the critical region to the left of $t_{0.025,65}$ and right of $t_{0.975,65}$ in order to reject the null hypothesis.

```
qt(0.025, 65)
```

```
## [1] -1.997138
```

Given that t > -1.997138, and thus to the right of $t_{0.025,65}$, we are *not* able to reject the null hypothesis. This means we *cannot* conclude any correlation between college attainment and vote swing from the 1992 to 1996 presidential elections.

Conclusion

From the two models, we are able to conclude that a correlation between college attainment and vote swing exists between the 2012 and 2016 presidential elections. However, such a conclusion cannot be made with the swing between the 1992 and 1996 presidential elections. From these conclusions, it can be inferred that the role of education in American voters' presidential party preferences has increased significantly since the 1990s. The role of education in vote swings has increased from ~5% to ~80% in that time.

While education is certainly not the only factor in the voting trends across the country, it does appear significant and increasing in significance. Polling outlets, political prognosticators, and pundits should take these conclusions into account going forward.

Bibliography

USDA, United States Department of Agriculture Economic Research Service, 2018, https://data.ers.usda.gov/reports.aspx?ID=17829

Dave Leip's Atlas of U.S. Presidential Elections, 2020, http://uselectionatlas.org