722 Short Paper

Nat Hill

Our study, which attempts to evaluate the effects that the percentage of high school graduates (ptcthsg), the closing date of registration (close), and whether or not a state is located in the (South) has on overall voter turnout (vote), yields strong support for our proposed hypotheses in addition to providing valuable insights into how voter turnout increases or decreases in US Elections. The model(s)[1] is as follows:

$$Vote = \beta 0 + \beta 1 \times Close + \epsilon$$

$$Vote = \beta 0 + \beta 1 \times South + \epsilon$$

$$Vote = \beta 0 + \beta 1 \times PctHSG + \epsilon$$

$$Vote = \beta 0 + \beta 1 \times Close + \beta 2 \times PctHSG + \beta 3 \times South + \epsilon$$

The tables are highly problematic. Each one needs a number and descriptive label and they need to be professionally formatted – they look like raw output, which is not appropriate for a paper submitted for anything other than casual comments to a colleague and even then it is best to format professionally.

Also, the code should not be included with the tables.

Results

Using bivariate linear regression (See Tables 1, 2 and 3) our results over indicate that the three chosen variables each have a significant association with the potential voter turnout. Holding all else constant and provided there are no confounding factors, the percentage of high school students in a given state has a positive relationship with voter turnout, supporting our initial hypothesis. On average, for every 1% increase in the number of high school graduates, there is estimated to be an approximate .24 increase in voter turnout. However, while these results indicate a strong statistical significance as the p-value associated with the coefficient is less than .001 this does not necessarily indicate a direct causal relationship between the amount of high school graduation in a state and the amount of voter turnout.

This sentence is confusing.

For bivariate models you are not holding anything else constant. It isn't clear if you are going from bivariate to multiple regression. In terms of organization it is best to work through all the bivariates to build the case and then do the regression.

Likewise, our results also support our hypothesis that the average closing date has a negative relationship with voter turnout. On average and holding all else constant, the data suggests that for

---

[1] In this case, the robust standard errors do not seem to significantly change the standard errors of the original model we have used in this study. Therefore, the statistical significance of the variables does not change, and the conclusions of our analysis should remain the same. See Table 5 in the appendix for more specific details.

every day closer to the election that voter registration closes, we can expect the voter turnout to decrease by approximately .267.

> **ℹ Note**
>
> It *increases* by .267 percentage points. `close` increases the further away from the election day registration closes.

Moreover, the p-value of 2.2e-15 indicates a strong statistical significance in our model, however, as with the results from the ptcthsg model we cannot assume that the date of regression has a direct [typo] causation on negative voter turnout. *Maybe note issues of causality once, globally, for efficiency.*

> **ℹ Note**
>
> Please don't use coded variable names in your write up (for example `pcthsg`).

Furthermore, our analysis indicates that being a state located in the South is associated with lower voter turnout compared to non-southern states. On average and holding all else constant, a southern state can predict to have an approximate decrease of -7.11 in voter turnout. More broadly, our results suggest an average voter turnout of 68.18% in non-southern states compared to an average voter turnout of 61.07% in southern states, on the whole showing a lower turnout in the south compared to the rest of the nation. *You also have the scatter plots and box plot and should generally start with them and then do the bivariate OLS.*

Our multivariate regression model (see Table 4) likewise supports the findings of our original theory and shows a statistical and substantive relationship between the predictor variables and voter turnout. The close and ptcthsg variables once again show a positive and negative relationship with voter turnout, respectively, and southern states still possess on average a lower voter turnout. Both the close and south coefficients decreased from the initial bivariate model, and while both are still significant at the .001 level, the decrease could indicate an effect when accounting for the other variables in the model. *This is more clearly the result of the other vars than you suggest.*

Therefore we can imply [infer] that an early closing date registration and location in the south are associated with a lower turnout percentage, holding all else constant. [on average] These findings could potentially have profound implications for policymakers, as targeted campaigns to increase voter turnout could focus their efforts more intently on southern states as well as make an effort to implement closer registration dates to elections across the country. It is also important to account for the magnitude and scale of these findings, for example, while the .21% decrease in voter turnout for every increase in voter turnout percentage might seem small, it could potentially accumulate across a large
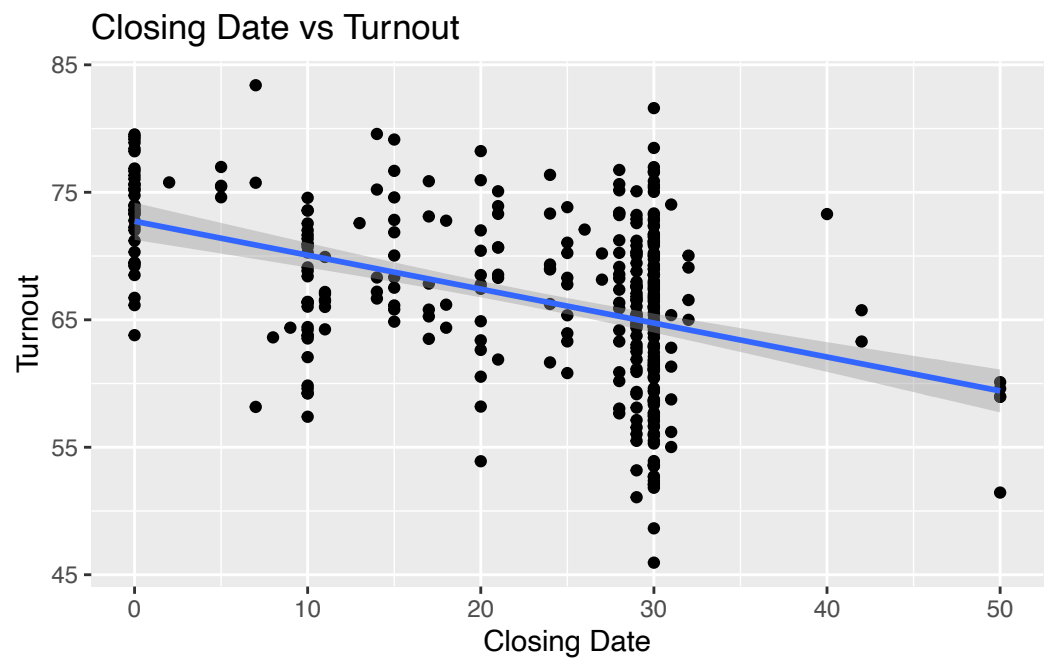
population and have a major impact in an election. Good discussion of sub sig.

The pcthsg variable did see a significant change when accounting for the other variables, as the coefficient decreased from .234 in the bivariate regression to .08 as well as having the significance level decrease (to p = .0252), while the multivariate results still indicate a positive relationship between, the inclusion of other variables could be effecting the impact its has on voter turnout.

You need to discuss sub sig for all the vars. Also, where is the robust SE footnote?

Appendix

Table 1: Closing Date vs Voter Turnout


```
===============================================

                 Dependent variable:

                 --------------------------

                            vote

-----------------------------------------------

close                      -0.266***

                            (0.029)


Constant                   72.728***

                            (0.740)
```

```
--------------------------------------------------
Observations                      357

R2                                0.189

Adjusted R2                       0.187

Residual Std. Error       5.987 (df = 355)

F Statistic           82.704*** (df = 1; 355)

==================================================
Note:                   *p<0.1; **p<0.05; ***p<0.01
```
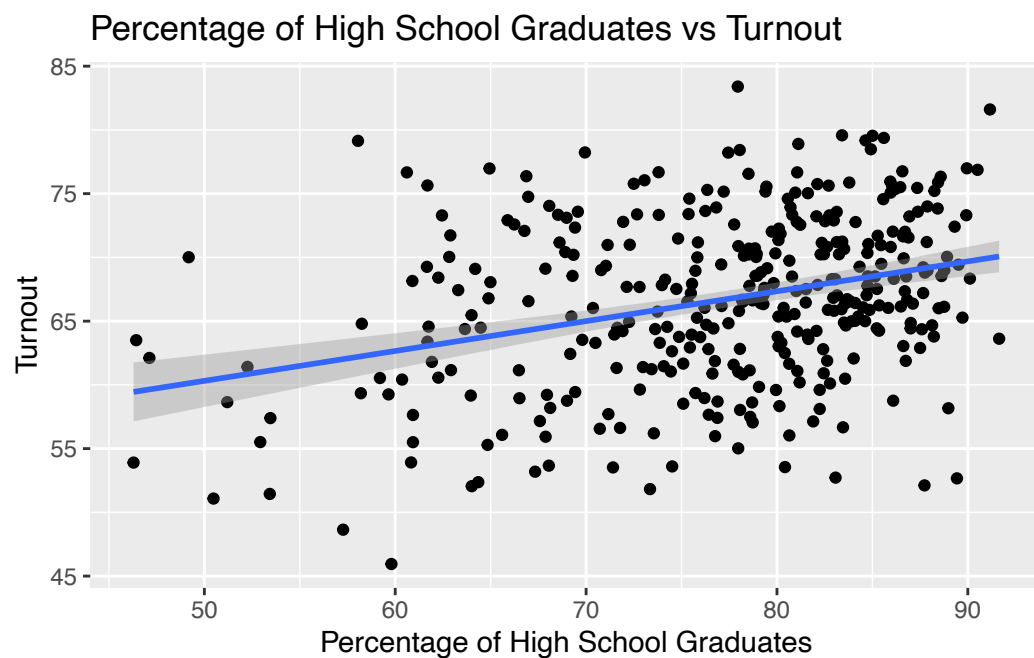
Table 2: Percentage of High School Graduates vs Turnout

Percentage of High School Graduates vs Turnout

```
Call:

lm(formula = vote ~ pcthsg, data = data)


Residuals:

      Min       1Q    Median       3Q      Max

-17.0398   -4.1328   -0.2189   4.6117   16.9383


Coefficients:

             Estimate Std. Error t value Pr(>|t|)

(Intercept) 48.60058    2.84186  17.102  < 2e-16 ***
```

```
pcthsg        0.23428    0.03664    6.394 5.09e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 6.296 on 355 degrees of freedom
Multiple R-squared:  0.1033,     Adjusted R-squared:  0.1008
F-statistic: 40.89 on 1 and 355 DF,  p-value: 5.087e-10
```

```
===============================================
                Dependent variable:
            -----------------------------
                        vote
-----------------------------------------------
pcthsg                 0.234***
                       (0.037)


Constant              48.601***
                       (2.842)


-----------------------------------------------
Observations             357
R2                      0.103
Adjusted R2             0.101
Residual Std. Error    6.296 (df = 355)
F Statistic           40.885*** (df = 1; 355)
===============================================
Note:           *p<0.1; **p<0.05; ***p<0.01
```
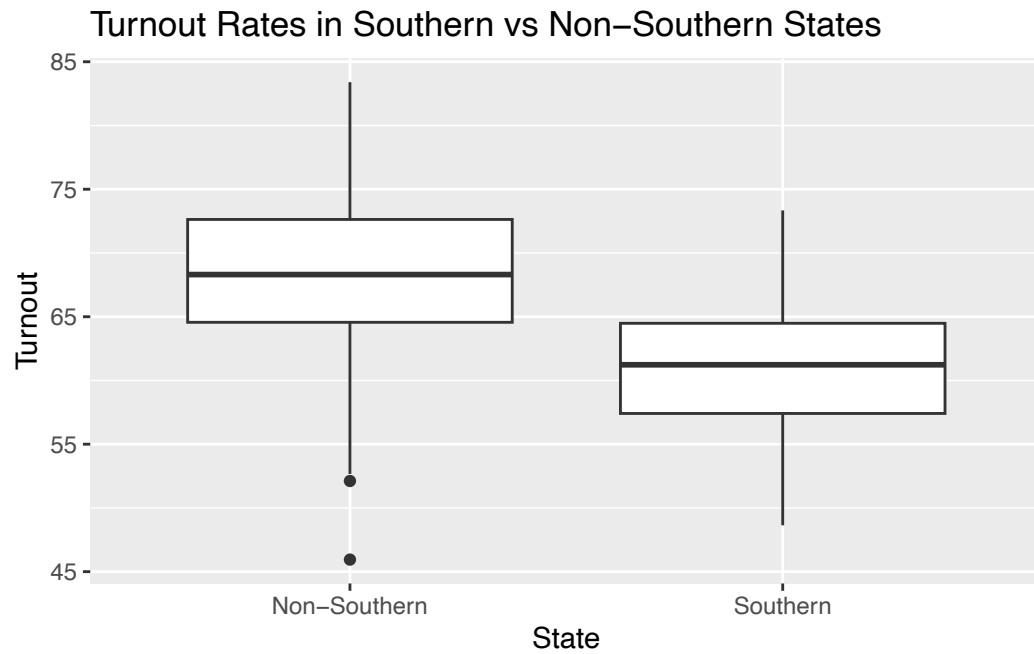
> **i** Note
>
> Never include the raw R model output in your submitted work.

> **i** Note
>
> Always use meaningful variable labels. Not your coded ones.

Table 3: Turnout Rates in Southern vs Non-Southern States

Turnout Rates in Southern vs Non–Southern States

```
Call:

lm(formula = vote ~ south, data = data)


Residuals:

     Min        1Q    Median        3Q       Max

-22.2349   -3.6293    0.1373    4.3761   15.2183


Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept)  68.1809     0.3565 191.239   <2e-16 ***

south        -7.1134     0.7677  -9.266   <2e-16 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 5.966 on 355 degrees of freedom

Multiple R-squared:  0.1948,    Adjusted R-squared:  0.1925

F-statistic: 85.86 on 1 and 355 DF,  p-value: < 2.2e-16


==================================================

                    Dependent variable:
```

```
                        ------------------------
                                 vote
-------------------------------------------------
south                         -7.113***
                               (0.768)


Constant                      68.181***
                               (0.357)



-------------------------------------------------
Observations                     357
R2                              0.195
Adjusted R2                     0.192
Residual Std. Error      5.966 (df = 355)
F Statistic            85.862*** (df = 1; 355)
=================================================
Note:              *p<0.1; **p<0.05; ***p<0.01
```

Table 4: Multivariate Regression Model

```
Call:
lm(formula = vote ~ close + pcthsg + south, data = data)


Residuals:
     Min       1Q   Median       3Q      Max
-18.9845  -3.6174   0.1271   3.6121  14.1879


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 66.41158    2.97056  22.357  < 2e-16 ***
close       -0.20743    0.02783  -7.452 7.13e-13 ***
pcthsg       0.07929    0.03528   2.247   0.0252 *
south       -5.25623    0.77398  -6.791 4.73e-11 ***
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.473 on 353 degrees of freedom

Multiple R-squared: 0.3261,   Adjusted R-squared: 0.3204

F-statistic: 56.94 on 3 and 353 DF, p-value: < 2.2e-16


```
===============================================
                    Dependent variable:
                  ---------------------------
                            vote
-----------------------------------------------
close                     -0.207***
                           (0.028)


pcthsg                     0.079**
                           (0.035)


south                     -5.256***
                           (0.774)


Constant                  66.412***
                           (2.971)


-----------------------------------------------
Observations                 357
R2                          0.326
Adjusted R2                 0.320
Residual Std. Error    5.473 (df = 353)
F Statistic         56.938*** (df = 3; 353)
===============================================
Note:              *p<0.1; **p<0.05; ***p<0.01
```

Table 5: Robust Errors Check

```
(Intercept)       close      pcthsg      south
 3.34662888  0.02508362  0.04049877  0.78317140
```

```
t test of coefficients:


            Estimate Std. Error t value  Pr(>|t|)
(Intercept) 66.411582   3.303048 20.1062 < 2.2e-16 ***
close       -0.207429   0.024890 -8.3339 1.762e-15 ***
pcthsg       0.079293   0.039980  1.9833   0.04811 *
south       -5.256233   0.775157 -6.7809 5.039e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```