

Problem Set #4

EH6105 - Quantitative Methods

Steven V. Miller

This homework makes use of data available in `{stevedata}` and implies the use of `{tidyverse}` to answer the questions. `{tidyverse}` is not necessary to answer these questions though it will assuredly make the process easier. Load these two libraries to get started answering these questions.

```
library(tidyverse)
library(stevedata)
```

State-Level Education and (U.S.) Voter Turnout in 2016

This homework will refer to the `election_turnout` data set that is available in `{stevedata}`. I created this data set in early 2017 largely for pedagogical purposes like this, allowing students to fit more complex OLS models with a still simple and intuitive application. The topic here is about Donald Trump's share of the vote in the 2016 presidential election for all 51 voting units in the United States (i.e. all 50 states + the District of Columbia). You can find out more information about the data by visiting [this part of the package's website](#), or with the following command.¹

```
?election_turnout
```

Here's a little preview of these data.

```
election_turnout
```

```
## # A tibble: 51 x 14
##   year state region division turnoutho perhsed percoled gdppercap  ss trumpw
##   <dbl> <chr> <chr> <chr>      <dbl> <dbl> <dbl> <int> <int> <int>
## 1 2016 Alab~ South East So~    59    84.3  23.5  42663  0     1
## 2 2016 Alas~ West  Pacific    61.3  92.1  28    81801  0     1
## 3 2016 Ariz~ West  Mountain    55    86    27.5  43269  0     1
## 4 2016 Arka~ South West So~    52.8  84.8  21.1  41129  0     1
## 5 2016 Cali~ West  Pacific    56.7  81.8  31.4  61924  0     0
## 6 2016 Colo~ West  Mountain    70.1  90.7  38.1  58009  1     0
## 7 2016 Conn~ North~ New Eng~    65.2  89.9  37.6  72331  0     0
## 8 2016 Dela~ South South A~    64.4  88.4  30    69930  0     0
```

¹The codebook isn't as clear as it should be about the `sunempr12md` variable, but I'll communicate here that higher values of this variable indicate there was more state-level unemployment relative to the previous point 12 months prior. For example, the state-level unemployment rate in Ohio for Nov. 2015 was 4.8%. In Nov. 2016: 5.2%. That is an increase of .4, which is what you see for `sunempr12md` for that observation.

```
## 9 2016 Dist~ South South A~      60.9   89.3   54.6   181185   0   0
## 10 2016 Flor~ South South A~      64.6   86.9   27.3   42595   1   1
## # ... with 41 more rows, and 4 more variables: trumpshare <dbl>, sunempr <dbl>,
## #   sunempr12md <dbl>, gdp <dbl>
```

Answer these questions. A successful answer of these question must include the R code you used to help you answer the question.

1. This is less of a question and more of a command. The `trumpshare` variable is a proportion and not a percent. How would you create a new variable—call it `trumpshareperc`— that converts this proportion to a percent as a new column in this data set? Make sure you do that before proceeding.
2. Donald Trump once (in)famously boasted “I love the poorly educated”, which suggested he may have done better in states with lower levels of educational attainment. Regress the `trumpshareperc` variable you just created as a dependent variable on the `percol` variable in the data set. Describe what you see in substantive terms.
3. One common refrain about the 2016 presidential election is that “economic anxiety” can account for why enough Americans rejected the incumbent Democrats in favor of the demagoguery of Donald Trump. We can perhaps estimate that concept with the `sunempr12md` variable in this data set. Re-run the regression you just ran above, but add this `sunempr12md` to the right-hand side of the formula. Explain the results to me.
4. There are four Census regions in the United States in which states are located. There is the North Central (Midwest) (e.g. Ohio, Illinois), the Northeast (e.g. New York, Massachusetts), the South (e.g. Florida, Georgia), and the more diverse and varied West (e.g. California, Utah). Run a “fixed effects” regression that regresses the `trumpshareperc` variable you created on just the `region` variable included in the data. Interpret the results to me, and *be careful* in how you describe them.
5. (2 POINTS) Take the same regression model you ran in the previous question and add in the `percol` and `sunempr12md` variables. Interpret the results to me, and *be careful* in how you describe them.
6. (2 POINTS) Time for a teachable moment on regression intercepts. In the above regression, the intercept is “statistically significant.” What is that value for the intercept actually communicating? Does it make sense?