Exercise 9: Analysis

Jess Esplin

11/12/2020

Review

1. Download the .Rmd version of this file from GitHub and change the author to your name.
2. Load packages.
3. Set up your files and folder structure.

Didn't do this. I prefer using my file explorer.

4. Read the ANES .dta data into R using the `here` package.

```
#anes2016 <- read_dta(here("exercise 9/anes_timeseries_2016.dta"))
# this didn't work and I'm not sure why


# set my working directory the way I normally do and was able to read the data in
setwd("C:/Users/jesse/OneDrive - UW-Madison/2020 Fall Semester/811 Intro to R/ps811-exercises/exercise 9
anes <- read_dta("anes_timeseries_2016.dta")
```

5. Download the ANES 2016 codebook (available on the `ps811/data` repository). We will look at the full sample variables.
6. You want to know whether owning a house (pre-election) affects which party the respondent choose to contribute to (post-election). Identify these variables from the codebook and rename the variables to names that are easier to reference.

```
# identify home ownership and party contribution variables
# home ownership is V161334, pg. 547 of codebook; 1=rent, 2=mortgage, 3=no payment homeowner, 4=other,
# party contribution is V162014a, pg. 762 of codebook; 1=democrat, 2=republican, 3=both, 5=other, -1, -

# rename variables
anes$homeowner <- anes$V161334
anes$party_contribution <- anes$V162014a
```

7. Now identify pre-election demographic variables, such as age, gender, and race. Manipulate these variables in ways that you believe would best capture these demographics and explain why you manipulated these variables that way you did. Rename these variables to names that are easier to reference.

```
# identify pre-election demographic variables
# age is V161267, pg. 426 of codebook; numeric, 18-90 or older,
# gender is V161342, pg. 556 of codebook; 1=Male, 2=Female, 3=Other, -9=Refused
# race is V161310x, pg. 512 of codebook; 1=White, 2=Black, 3= Asian/Pacific Islander, 4=Native American,

# rename variables
anes$age <- anes$V161267
```

```
anes$gender <- anes$V161342
anes$race <- anes$V161310x

# manipulate variables

# had a hard time figuring out what I needed to do here - didn't know what I needed to do for the regre

anes2 <- subset(anes, subset = age > 0 &
                               gender > 0 &
                               race > 0 &
                               homeowner > 0 &
                               homeowner < 4 &
                               party_contribution > 0 &
                               party_contribution < 3,
                select = c(homeowner, party_contribution, age, gender, race)
                )

# attempted to select, didn't work
#anes2_select <- c(anes2$homeowner, anes2$party_contribution, anes2$age, anes2$gender, anes2$race)

# turn it into a data frame?
anes2 <- as.data.frame(anes2)


# based on feedback from Marcy, need to revise subset to make all variables binary

# attempted to select only variables I want, didn't work
#anes2_select <- select(anes2, homeowner, party_contribution, age, gender, race)

# went back and revised my subset code to select only desired variables


# turn homeowner variable into binary 0 or 1 for homeowner
# currently: 1=rent, 2=mortgage, 3=no payment homeowner, 4=other, 5=don't know
# we want 0=rent, 1=mortgage or no payment, exclude 4 and 5 entirely
# okay went back and excluded anything > 3
# use mutate function
anes2 <-
  mutate(anes2,
         homeowner_adj = ifelse(homeowner == 1,
                                0,
                                1)
         )

# now do this for party_contribution too
# currently: 1=democrat, 2=republican, 3=both, 5=other, -1, -6, -7=inapplicable or missing
# we want 0=Dem, 2=Rep, exclude all others
# went back and excluded anything > 2
# use mutate function again
anes2 <-
  mutate(anes2,
         party_cont_adj = ifelse(party_contribution == 1,
                                0,
```

```
                                1)
        )
```

I am not sure I did this right because excluding so many answers cut down the data set by a lot.

8. Provide descriptive summaries for each variable.

```
# descriptive stats for each variable
summary(anes2$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   45.00   60.00   57.26   70.00   90.00
```

```
summary(anes2$gender)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.493   2.000   3.000
```

```
summary(anes2$race)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.659   1.000   6.000
```

```
# descriptive stats for our IV and DV
summary(anes2$homeowner_adj)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  1.0000  1.0000  0.7707  1.0000  1.0000
```

```
summary(anes2$party_cont_adj)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.3707  1.0000  1.0000
```

```
# not sure if we need to do CI but going off of Marcy's lecture, yes?
# confidence intervals
group.CI(party_contribution ~ 1, data = anes2)
```

```
##   party_contribution.upper party_contribution.mean party_contribution.lower
## 1                 1.419775                1.370667                 1.321559
```

```
# or grouped by race
group.CI(party_contribution ~ race, data = anes2)
```

```
##   race party_contribution.upper party_contribution.mean
## 1    1                 1.473542                1.416949
## 2    2                 1.122556                1.040000
## 3    3                 1.807682                1.375000
## 4    5                 1.363708                1.206897
## 5    6                 1.574554                1.333333
##   party_contribution.lower
## 1                1.3603564
## 2                0.9574441
## 3                0.9423181
## 4                1.0500847
## 5                1.0921131
```

```
# check correlation
cor(anes2$homeowner, anes2$party_contribution)
```

```
## [1] 0.1528478
```

```r
# not really sure what this tells us tbh
```

9. Run an appropriate regression analysis and insert the table into the R Markdown document.

```r
# ols regression
# lm(anes2$party_contribution ~ anes2$homeowner, data = anes2)
# this doesn't seem to be a good regression since both variables are categorical

# alt regression:
# could use logit regression I think? But has to be binary variables: homeowner or not, Rep or Dem cont
# going back to make my variables binary

# logit regression
logit_reg <- glm(party_cont_adj ~ homeowner_adj + age + gender + race, data = anes2, family = "binomial

# request summary
summary(logit_reg)
```

```
##
## Call:
## glm(formula = party_cont_adj ~ homeowner_adj + age + gender +
##     race, family = "binomial", data = anes2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5239  -0.9687  -0.6338   1.1564   2.1164
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.680778   0.562943  -2.986 0.002829 **
## homeowner_adj  0.623930   0.324417   1.923 0.054451 .
## age            0.028307   0.007402   3.824 0.000131 ***
## gender        -0.562014   0.225660  -2.491 0.012755 *
## race          -0.114855   0.084167  -1.365 0.172373
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 494.48  on 374  degrees of freedom
## Residual deviance: 453.69  on 370  degrees of freedom
## AIC: 463.69
##
## Number of Fisher Scoring iterations: 4
```

```r
# use stargazer to produce a nice table in RMarkdown I guess?
stargazer(logit_reg , type = 'latex',
          title = "Effect of homeownership on party contributions",
          covariate.labels = c("Age", "Gender", "Race"),
          dep.var.labels = "Party Contribution (1=GOP)",
          header = FALSE
          )
```

I don't know why it's putting race in there twice.

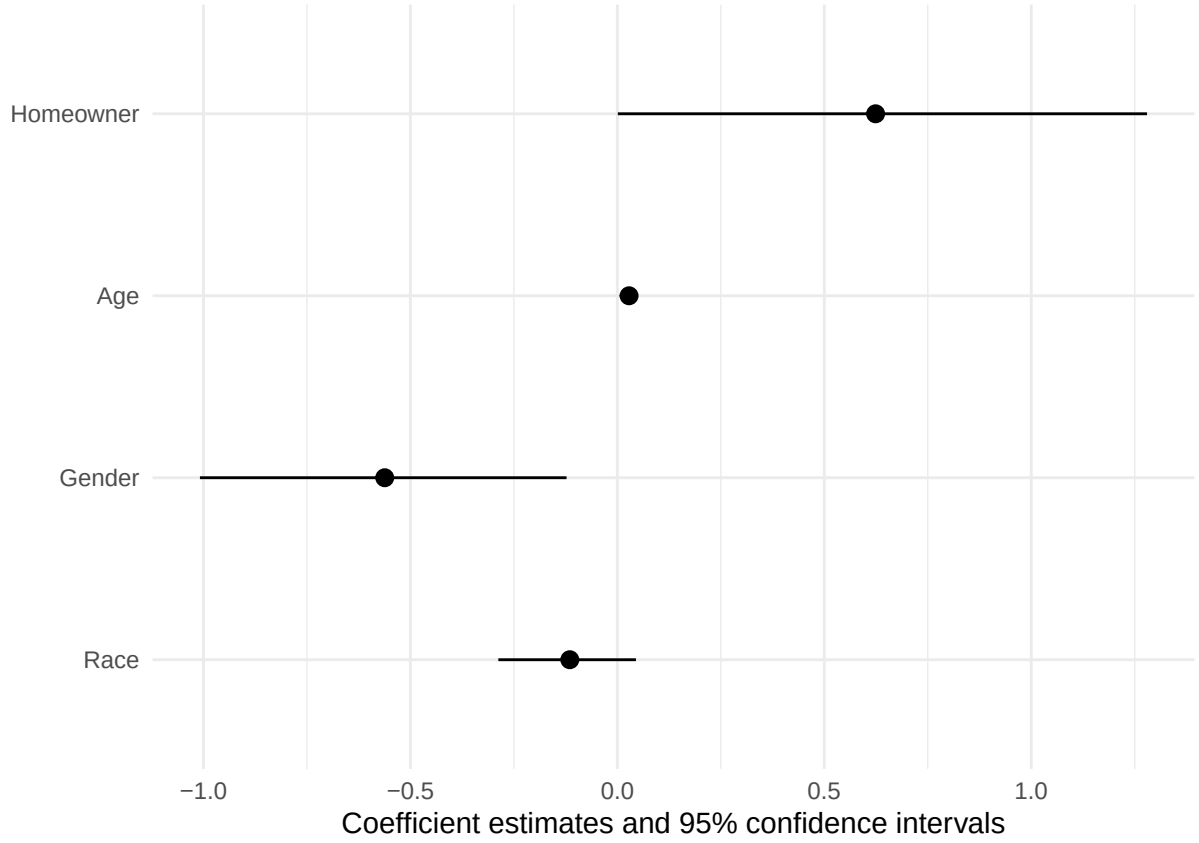
10. Create a coefficient plot based on the above table.

Table 1: Effect of homeownership on party contributions

	Dependent variable:
	Party Contribution (1=GOP)
Age	0.624^{*}
	(0.324)
Gender	0.028^{***}
	(0.007)
Race	-0.562^{**}
	(0.226)
race	-0.115
	(0.084)
Constant	-1.681^{***}
	(0.563)
Observations	375
Log Likelihood	-226.844
Akaike Inf. Crit.	463.687
Note:	*p<0.1; **p<0.05; ***p<0.01

```
# change variable names
vars <- c('party_cont_adj' = 'Party Contribution',
            'homeowner_adj' = 'Homeowner',
            'age' = 'Age',
            'gender' = 'Gender',
            'race' = 'Race')

# coefficient plot
modelplot(logit_reg, coef_map = vars)
```



Your project

Now it's your turn. Use the tools you used today to conduct data analysis for one of your final seminar papers.

Hi Marcy, I'm not going to be able to do this yet. Still trying to figure out if I even have the right/all the data I need. I'll answer these questions as best I can for now.

1. Create a descriptive statistics summary table for your main variables of interest. Note the number of observations.

```
# Here's what I would do:
# Twitter data set = data
#summary(data)
#dim(data)

# it's really hard to know what all my variables are right now. A lot of them are text too so not really
```

2. If you are planning to run a regression, please write out the regression formula. Please take into consideration the dependent variable and its distribution. If you already have the data, you may go ahead and run it. If you do not have the data and is in the process of collecting it, write out the formula. Pre-analysis plans are becoming more common in the discipline, so being able to record what you *plan* to do is becoming increasingly more important.

I do not think I'll be running a regression. I think I'll need to do some kind of text coding/analysis to code instances of org mentioned and somehow also code each post for support/oppose of BLM. I really don't know what this is going to look like yet, but I'm not going to be trying to do extensive analysis on this project.

I'm not interested in this topic at all for my own research projects.

Submit

Email me (mshieh2@wisc.edu) the link to your `ps811-exercises` repository when you are done.