Exercise 10: Data Manipulation with Replication Exampl (Answer Key)

Marcy Shieh

12/5/2020

Extend the dataset

Extend the exercise we did in class up to the Senate confirmation of Associate Justice Sonia Sotomayor, following the instructions in the `lecture-10.R` file.

Links to the datasets required to complete this exercise:

- Harvard Dataverse entry for the Epstein et al. (2006) article (includes PDF of article for reference): https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/W5CV8E&version=1.0
- Poole & Rosenthal scores from 2016: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/W5CV8E&version=1.0
- 111th Congress roll call votes and dictionary: https://legacy.voteview.com/senate111.htm
- Segal-Cover scores: https://en.wikipedia.org/wiki/Segal%E2%80%93Cover_score

```
# download 111th congress roll call votes
congress111 <- read_dta(here::here("data", "sen111kh.dta"))

# V263 = Sotomayor senate vote
congress111_sotomayor <- congress111 %>%
  select(cong:name, V262) %>% # it is on row 263 meaning it is V262 (consider the header)
  filter(name != "OBAMA") # exclude Obama (president)

# 1 = YES, 6 = NAY, 0 = not a member
# you want to change 6 to 0
# you want to change 0 to NA
congress111_sotomayor <- congress111_sotomayor %>%
  mutate(
    vote = ifelse(V262 == 6,
                  0,
                  ifelse(V262 == 0,
                         NA,
                         V262)))

# you can confirm that this is the correct vote column by looking up "sotomayor confirmation" and see t
table(congress111_sotomayor$vote)
```

```
##
##  0  1  9
## 31 68  1
```

```r
# rename "cong" variable to "congress"
congress111_sotomayor <- rename(congress111_sotomayor, congress = cong)

# select the relevant IDs
sotomayor <- congress111_sotomayor %>%
  select(congress, id, state, name, vote)

# select relevant variables to the poole & rosenthal CS scores
poole_select <- poole %>%
  select(congress, idno, statenm, name, dwnom1, party)

# merge the senate votes with their CS scores
sotomayor_poole <- left_join(sotomayor, poole_select, by=c("id" = "idno", "congress", "name"))

# rename "dwnom1" to "cs1"
sotomayor_poole <- rename(sotomayor_poole, cs1 = dwnom1)

# now you want to build a dataframe with the following information:
# nominee, presprty, sameprty, strngprs, nomid, regime_dummy, lackqual, csnom, eucldist, interaction
sotomayor_poole$nominee <- rep("SOTOMAYOR") # name of nominee
sotomayor_poole$presprty <- rep(1) # president held party majority in Senate

# whether senator is the same party as the president
sotomayor_poole %>%
  mutate(sameprty = ifelse(party == 100,
                           1,
                           0))
```

```
## # A tibble: 111 x 11
##    congress    id state name   vote statenm    cs1 party nominee presprty
##       <dbl> <dbl> <dbl> <chr> <dbl> <chr>    <dbl> <dbl> <chr>      <dbl>
##  1      111 49700    41 SESS~     0 " ALAB~  0.551   200 SOTOMA~        1
##  2      111 94659    41 SHEL~     0 " ALAB~  0.437   200 SOTOMA~        1
##  3      111 40300    81 MURK~     0 " ALAS~  0.207   200 SOTOMA~        1
##  4      111 40900    81 BEGI~     1 " ALAS~ -0.235   100 SOTOMA~        1
##  5      111 15429    61 KYL       0 " ARIZ~  0.548   200 SOTOMA~        1
##  6      111 15039    61 MCCA~     0 " ARIZ~  0.382   200 SOTOMA~        1
##  7      111 40301    42 PRYOR     1 " ARKA~ -0.199   100 SOTOMA~        1
##  8      111 29305    42 LINC~     1 " ARKA~ -0.166   100 SOTOMA~        1
##  9      111 15011    71 BOXER     1 " CALI~ -0.450   100 SOTOMA~        1
## 10      111 49300    71 FEIN~     1 " CALI~ -0.263   100 SOTOMA~        1
## # ... with 101 more rows, and 1 more variable: sameprty <dbl>
```

```r
sotomayor_poole$nomid <- rep(0.780) # ideology score of the judicial nominee: https://en.wikipedia.org/

sotomayor_poole$regime_dummy <- rep(1) # whether the judicial nominee was pre- or post-bork (1 is post-

sotomayor_poole$lackqual <- rep(1-0.810) # you take how qualified someone is based on newspaper editori

# you will need to calculate the following (later)

sotomayor_poole$csnom <- NA # this is going to be the Segal-Cover score for each nominee "transformed"

sotomayor_poole$eucldist <- NA # this is going to be the Euclidean distance between the senator and the
```

```
sotomayor_poole$interaction <-NA # this is the interaction between ideological distance and qualificati

# bind the dataset to the bork dataset
all_data <- full_join(bork_alito_poole, sotomayor_poole)
```

```
## Joining, by = c("congress", "statenm", "name", "cs1", "nominee", "vote", "presprty", "nomid", "regime
```

```
# the last two columns are just for binding datasets before
# you can just keep them there if they don't interfere with your analysis going forward

# from reading the article, you know that you need the common space scores for each president as well
# the president is coded with a congressional district (cd) number of 0, and a state number (state) of
poole_filter <- poole %>%
  filter(cd == 0 & state == 99) %>%
  select(congress, name, dwnom1)

# rename name
poole_filter <- rename(poole_filter, pres = name)

# rename dwnom1
poole_filter <- rename(poole_filter, cspres = dwnom1)

# join together the cspres scores
all_data_full <- full_join(all_data, poole_filter)
```

```
## Joining, by = "congress"
```

```
sotomayor_poole_coeffs <- lm(cspres ~ nomid, data = all_data_full)
sotomayor_poole_coeffs$coefficients
```

```
## (Intercept)       nomid
##   0.5832055  -0.9999904
```

```
# result:
# (Intercept)       nomid
# 0.5832055  -0.9999904

# scale the common space scores for sotomayor
all_data_full$csnom <- 0.5832055 - 0.9999904*all_data_full$nomid

# do the calculations for ideological distance
all_data_full$eucldist <- (all_data_full$csnom - all_data_full$cs1)^2

# do the calculations for the interaction between ideological distance and qualifications
all_data_full$interaction <- all_data_full$eucldist*sotomayor_poole$lackqual
```

```
## Warning in all_data_full$eucldist * sotomayor_poole$lackqual: longer object
## length is not a multiple of shorter object length
```

```
# now you have the full dataset so you can run the analysis again from black to sotomayor (and not just
```

```
datasummary((`Senate Vote` = vote) +
              (`Lack of Qualifications` = lackqual) +
              (`Ideological Distance` = eucldist) +
              (`Strong President` = strngprs) +
              (`Same Party` = sameprty) ~
```

```
            Mean + SD + Min + Max,
          data = all_data_full,
          output = 'markdown')
```

	Mean	SD	Min	Max
Senate Vote	0.87	0.36	0.00	9.00
Lack of Qualifications	0.22	0.26	0.00	0.89
Ideological Distance	0.21	0.25	0.00	1.21
Strong President	0.57	0.49	0.00	1.00
Same Party	0.56	0.50	0.00	1.00

```
stargazer(table2_ccs_nominees_model, table2_add_nominees, table2_newadditions,
          covariate.labels = c("Lack of Qualifications",
                               "Ideological Distance",
                               "Strong President",
                               "Same Party"),
          column.labels = c("CCS Nominees",
                            "Additional Nominees",
                            "Add Alito and Sotomayor"),
          dep.var.caption = "Senate Vote",
          header = FALSE,
          dep.var.labels.include = FALSE,
          type = "latex")
```

Table 2:

	Senate Vote		
	CCS Nominees	Additional Nominees	Add Alito and Sotomayor
	(1)	(2)	(3)
Lack of Qualifications	-2.518^{***}	-2.320^{***}	-2.250^{***}
	(0.155)	(0.122)	(0.121)
Ideological Distance	-2.128^{***}	-2.242^{***}	-2.029^{***}
	(0.177)	(0.151)	(0.133)
Strong President	0.882^{***}	0.767^{***}	0.649^{***}
	(0.104)	(0.071)	(0.071)
Same Party	0.477^{***}	0.707^{***}	0.625^{***}
	(0.093)	(0.078)	(0.080)
Constant	1.840^{***}	1.816^{***}	1.909^{***}
	(0.094)	(0.082)	(0.086)
Observations	2,062	3,709	3,709
Log Likelihood	-550.328	-846.606	-833.681
Akaike Inf. Crit.	1,110.656	1,703.213	1,677.361
Note:			*p<0.1; **p<0.05; ***p<0.01

Replication

Find replication materials for a paper and replicate a table within the paper.

- There are many political science papers with data stored in the Stata (`.dta`) file format accompanied by a Stata file (`.do`) file. In those cases, you will need to read the Stata file into R using the `haven` package and recreate your table of choice using R code. In other words, please do not turn in Stata code!
- If you do find replication materials in R and would like to run the entire R script, please feel free to do so! Reviewing other people's replication materials is the best way to learn how scholars assemble their results.

Here are some great resources for finding articles with reproducible data.

- Political Science Replication (https://politicalsciencereplication.wordpress.com/)
- Harvard Dataverse (https://dataverse.harvard.edu/dataverse/harvard?q=&fq0=subject_ss%3A%22Social%20Sciences%22&types=dataverses%3Adatasets&sort=dateSort&order=desc)
- Recent articles from *Political Science Research and Methods*, *American Journal of Political Science*, *American Political Science Review*, and *Political Analysis*
  - You can access articles in the top journals by signing into the UW-Madison Library website (https://library.wisc.edu) with your UW netID

Submit

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