Lab #4: CCES Data

Presidential Elections LAB

August 24th 2017

Part 1: Codebooks

- Today, we will be working with the 2016 CCES data that you will be using in your final paper project. CCES stands for "Cooperative Congressional Election Study". Please download the codebook or "guide" from iLearn now.
- What's a codebook, you ask? No worries! A codebook simply describes the contents, and layout of a data set. Now, if you look at thr CCES codebooks, you will see the basic front matter right away. This includes details like the study title, name of the principal investigators, and a basic introduction to the data.
- It usually goes on to show each variable name, the corresponding question that was asked in the survey, and responde coding or types. Some codebooks even include details like survey weighting and the questionaires used. For instance, on page 26, you will see this:

birthyr	Birth Year
Question:	In what year were you born?
Min.	1917.00
1st Qu.	1955.00
Median	1967.00
Mean	1968.12
3rd Qu.	1983.00
Max.	1998.00

Question 1:

a. Who are the principal investigators and co-investigators listed for the CCES (front page)?

b. When did "data release 2" occur?

c. What is the CCES for? What does this study focus on?

d. What variable would I use to look at a respondent's gender?

Part 2: The Data

- Great job! Now you know how to find out more information about each variable type!
- Please note that if you look at "Birth Year" (which is coded under "birthyr") in the dataset, you won't see much. Why is this!? Well, when this data was collected, each survey respondent was asked, "In what year were you born?". Sometimes, information like this can be sensitive or categorized as that

which may be used to identify anonymous participants. Therefore, it is taken out of the public data. This is probably the case here! Check it out:

```
summary(cces_data$birthyr)
```

```
##   Skipped Not Asked      NA's
##         0         0     64600
```

- This is the case for very few variables. Let's take a closer look at a few with data attached!
- Proceed to Question 2.

Question 2:

a. What is the variable "marstat" about? What question is associated with this variable?

b. What do the numbers look like?

c. What is the name of the variable that asks "OVER THE PAST YEAR the nation's economy has . . . ?"

d. What do the numbers for this variable look like?

Part 3: Analysis

- It's time to get to some data analysis. How exciting!!! Don't worry, we're using the same code that we used in the first few weeks. All of the information and code you need will be included here, and in your LAB 4 starter file.
- Let's say I wanted to see the breakdown of how important the issue of crime is to people who identify with different ideological leanings. In other words– How important is crime to a VERY LIBERAL person compared to a VERY CONSERVATIVE person? We can figure it out by using this code:

```
cces_data %>% group_by(ideo5 ,CC16_301j) %>%
  filter(ideo5 < "Not sure") %>% filter(CC16_301j < "Skipped") %>%
  summarise (n = n()) %>%
  mutate(proportion = n / sum(n)) %>%
  top_n(1)
```

```
## Selecting by proportion
```

```
## Source: local data frame [5 x 4]
## Groups: ideo5 [5]
##
## # A tibble: 5 x 4
##               ideo5                CC16_301j     n proportion
##               <ord>                    <ord> <int>      <dbl>
## 1      Very liberal Somewhat High Importance   455  0.3568627
## 2           Liberal Somewhat High Importance  1012  0.4012688
## 3          Moderate     Very High Importance  1926  0.4352542
## 4      Conservative     Very High Importance  1884  0.5688406
## 5 Very conservative     Very High Importance   735  0.6125000
```

- Intesting! It looks like 61.25% of very conservative people think the issue of crime is of very high importance. Meanwhile, the highest proportion of very liberal people think that it's only somewhat high.

- Wow! Now, let's see how this pans out ONLY for people with 4 year degrees (the code for this is in your starterfile):

```
## Selecting by proportion
```

```
## Source: local data frame [5 x 4]
## Groups: ideo5 [5]
##
## # A tibble: 5 x 4
##               ideo5                CC16_301j     n proportion
##               <ord>                    <ord> <int>      <dbl>
## 1      Very liberal Somewhat High Importance   140  0.3580563
## 2           Liberal Somewhat High Importance   326  0.4250326
## 3          Moderate Somewhat High Importance   498  0.4078624
## 4      Conservative     Very High Importance   467  0.5206243
## 5 Very conservative     Very High Importance   171  0.5327103
```

- Hmmm... Let's see how this pans out ONLY for people who did not graduate from high school (the code for this is in your starterfile):

```
## Selecting by proportion
```

```
## Source: local data frame [6 x 4]
## Groups: ideo5 [5]
##
## # A tibble: 6 x 4
##               ideo5                CC16_301j     n proportion
##               <ord>                    <ord> <int>      <dbl>
## 1      Very liberal     Very High Importance     6  0.3750000
## 2      Very liberal Somewhat High Importance     6  0.3750000
## 3           Liberal     Very High Importance    15  0.5000000
## 4          Moderate     Very High Importance    48  0.6760563
## 5      Conservative     Very High Importance    24  0.6000000
## 6 Very conservative     Very High Importance    15  0.7142857
```

- Okay! Interesting! It's important to realize what we're actually looking at here. "very liberal" comes out twice because responses for "Very High Importance" and "Somewhat High Importance" are equal to eachother. When we "group_by" ideology (second line of code), we are tellingR to give us the highest number of responses for each ideological option (Very liberal, liberal, moderate, conservative, and very conservative). Can we switch this around? Let's look at the how people feel about crime, and ask R to give us the ideological option that was most commonly given as an answer. We will also take out our education filter to get a fuller picture:

```
## Selecting by proportion
```

```
## Source: local data frame [5 x 4]
## Groups: CC16_301j [5]
##
## # A tibble: 5 x 4
##                   CC16_301j    ideo5     n proportion
##                       <ord>    <ord> <int>      <dbl>
## 1     Very High Importance Moderate  1926  0.3416711
## 2 Somewhat High Importance Moderate  1717  0.3659420
## 3  Somewhat Low Importance Moderate   641  0.3309241
## 4      Very Low Importance Moderate   113  0.2973684
## 5     No Importance at All Moderate    28  0.3181818
```

- Now, let's look at some other topics. For this section, we will use some party identification questions.

There are quite a few:

pid7 — 7 point Party ID

pid7text

16251	1	Strong Democrat
8618	2	Not very strong Democrat
8479	7	Strong Republican
6814	6	Not very strong Republican
6270	3	Lean Democrat
5554	5	Lean Republican
10493	4	Independent
2067	8	Not sure
34	98	skipped
20	99	not asked

pid3 — 3 point party ID

Generally speaking, do you think of yourself as a ...?

24881	1	Democrat
15300	2	Republican
18238	3	Independent
2379	4	Other
3782	5	Not sure
20	8	skipped
0	9	not asked

- So, for the following questions, please use "pid3". You can see the details above.

Question 3:

a. Let's think about the retrospective voting, and more specifically, sociotropic voting... What was the majority opinion for Democrats, Republicans, and Independents? ((HINT: USE A VARIABLE YOU HAVE ALREADY USED IN QUESTION 2)). Make sure to remove "Other" from pid3, and "Not sure" from your new variable.

b. What did you find? Why do you think this is the case?

c. Thinking about the retrospective voting, what was the breakdown of voters in my party from a sociotropic perspective? ((HINT: Use "filter(VARIABLE NAME HERE == Your Party") to print out specific information.))

d. Why do you think you got these responses in part c? Does it make sense? Why or why not?

e. Can opinions about the economy be linked to personal income change (variable name: CC16_303)?

f. Do you see evidence of partisanship being divided on economic lines? Why or why not? ((HINT:: You can add another filter command to specify a certain party within the pid3 variable, like before.))

If you finish this lab early, come ask me for some extra credit work!