Writing Assignment 2

1 Instructions

- This writing assignment is due on October 15 by the beginning of class.
- Hopefully, you can have the computation (Part 1) done by October 8, so that we can spend a couple of class periods talking about the writing.
- Plagiarizing prose or code is not acceptable.

2 Background

Note: This question is based on Section 3.9 of *QSS*.

2.1 The Original Study

For the background, read the original article by Lacour and Green (2014). Notice that this study was covered extensively in the news. Read this short Buzzfeed Report and listen to this 15-minute SCPR interview.

2.2 Irregularities

In May 2015, though, three scholars reported several irregularities in the dataset used to produce the results in the study. They found that the gay marriage experimental data were statistically indistinguishable from data in the 2012 Cooperative Campaign Analysis Project (CCAP). The scholars suggested that the CCAP survey data–and not the original data alleged to have been collected in the experiment–were used to produce the results reported in the gay marriage study. The release of a report on these irregularities ultimately led to the retraction of the original article. In this exercise, we will use several measurement strategies to reproduce the irregularities observed in the gay marriage dataset.

For this project, we'll use three data sets:

1. Four feeling thermometers from the pre- and post-election waves of the 2012 ANES, `anes-2012-therms.csv`. This data set is available here. See Table 1.
2. The 2012 CCAP dataset alleged to have been used as the basis for the gay marriage study results, `ccap2012.csv`. This data set is available here. See Table 2.
3. The data used in Lacour and Green (2014), `gayreshaped.csv`. This data set is available here. See Table 3.

2.3 Feeling Thermometers

Many variables for this project are "feeling thermometers," which offer one way to measure a respondent's feelings toward a person or a group.

From the supplementary materials to Lacour and Green (2014):

Wording and format followed the Cooperative Campaign Analysis Project 2012 (Jackman et al 2012): "We would like to get your feelings toward a series of demographic groups. We will display the name of a group, and we would like you to rate the group using a 'feeling thermometer.' Ratings between 50 degrees and 100 degrees indicate that you feel favorable and warm toward the group. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the group and that you don't care too much for that group. You would rate the group at the 50 degree mark if you don't feel particularly warm or cold toward the group. Your rating will appear at the end of the slider." In Study 1, subjects were presented a virtual slider button that was pre-set to their rating in the previous wave. Concerned that this method may distort our assessment of how effects persist over time, we presented no slider button or pre-set rating in Wave 7 or in Study 2.

The ANES uses the following wording:

I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person. If we come to a person whose name you don't recognize, you don't need to rate that person. Just tell me and we'll move on to the next one.

Data and Variables

Variable	Description
`ft_barack_obama_pre`	pre-election feeling thermometer for Barack Obama
`ft_barack_obama_post`	post-election feeling thermometer for Barack Obama
`ft_mitt_romney_pre`	pre-election feeling thermometer for Mitt Romney
`ft_mitt_romney_post`	post-election feeling thermometer for Mitt Romney
`ft_joe_biden_pre`	pre-election feeling thermometer for Joe Biden
`ft_joe_biden_post`	post-election feeling thermometer for Joe Biden
`ft_paul_ryan_pre`	pre-election feeling thermometer for Paul Ryan
`ft_paul_ryan_post`	post-election feeling thermometer for Paul Ryan
`ft_gay`	(post-election) feeling thermometer for gay men and lesbians

Table 1: Names and Descriptions of the variables in `anes-2012-therms.csv`.

Variable	Description
`caseid`	unique respondent ID
`gaytherm`	Feeling thermometer for gay couples

Table 2: Names and Descriptions of the variables in `ccap2012.csv`.

Variable	Description
`study`	Which study is the data from (1 = Study1, 2 = Study2)
`treatment`	Five possible treatment assignment options
`therm1`	Feeling thermometer for gay couples in wave 1
`therm2`	Feeling thermometer for gay couples in wave 2
`therm3`	Feeling thermometer for gay couples in wave 3
`therm4`	Feeling thermometer for gay couples in wave 4

Table 3: Names and Descriptions of the variables in `gayreshaped2.csv`.

2.4 Part 1: R Script

Create a new project directory for this assignment. Call it `writing-assignment-2`. You'll want to set your working directory to this folder when working on this assignment.

```
writing-assignment-2
|-- data
   |-- anes-2012-therms.csv
   |-- ccap.csv
   |-- gayreshaped.csv
|-- doc
   |-- figs
|-- R
```

Create an R script called `irregularities.R` and save it to the `R` subdirectory. This script should automatically save all the figures to the subdirectory `doc/figs`–make sure to choose the filenames of the figures carefully. The script `irregularities.R` should do the following:

1. Load the ANES, CCAP, and LaCour and Green's data. Using the `subset()` function split LaCour and Green's data into two data sets: Study 1 (i.e., `study == 1`) and Study 2 (i.e., `study == 2`). Assign these two smaller data sets to their own objects.

2. Now create four histograms: a histogram of the gay thermometer variable from the CCAP, a histogram of the gay thermometer variable from the ANES, a histogram of the gay feeling thermometer in LaCour and Green's Study 1, and a histogram a histogram of the gay feeling thermometer in LaCour and Green's Study 2. For each histogram, set the breaks argument to some fixed, large value, say 50 (this makes it easier to compare). Notice the histograms that look similar and dissimilar. How many observations (i.e., rows) in the CCAP data, the ANES data, Study 1 of LaCour and Green's data, and Study 2 of LaCour and Green's data.

3. Comparing histograms by glancing back and forth is hard. A more direct way to compare the distributions is to use a quantile-quantile plot or QQ-plot (see section 3.6 of *QSS*). Use a QQ-plot to compare the four feeling thermometers: the one from the CCAP data, the one from the ANES data, the one from wave 1 of Study 1 of LaCour and Green's data, and the one from wave 1 of Study 2 of LaCour and Green's data. Be sure to add a 45° line.

4. Using the ANES data, create four scatterplots of the pre- and post-election feeling thermometer ratings–one scatter plot for each candidate. Place the pre-election rating along the x-axis (horizontal axis) and the post-election rating along the y-axis. Be sure to give each plot nice axis labels. Notice that some respondents give very different rating pre- and post-election.

5. Using LaCour and Green's data from Study 2's control group, create a scatterplot of the feeling thermometer ratings from wave 1 against each subsequent wave. You'll have three scatterplots: wave 1 v. wave 2, wave 1 v. wave 3, and wave 1 v. wave 4. Notice whether these scatterplots look similar to those from the pre- and post-election ANES surveys. Think about what might explain this difference. I recommend using the `subset()` function to create a new data set (from the Study 2 subset of LaCour and Green's data that you created in Question 1) that has only cases from the control group (i.e., `treatment == "No Contact"`).

6. Using the data set containing only the control group from Study 2 created above, create a three histograms: one of the difference between the gay feeling thermometer in waves 1 and 2, waves 1 and 3, and waves 1 and 4. Compare those distributions to each other. Notice whether these look like a distribution you might have seen before. Now create four histograms of the difference between the ANES pre- and post-election feeling thermometers for Obama, Romney, Biden, and Ryan. Notice whether these look similar to the differences across waves of LaCour and Green's data. For these seven histograms, make sure to set a large number of breaks (e.g., `breaks = 25`) and set the limits of the x-axis to their logical limits (i.e., `xlim = c(-100, 100)`)–that will make the comparison easier.

2.5 Part 2: Essay

Clearly divide the essay into five numbered sections (as this assignment is divided into three sections). Give each section a nice name. For example, don't call the first section "Section 1." Instead, call it something like "1 An Overview of LaCour and Green (2014)." Section 1 should correspond to the first question, section 2 should correspond to the second question, and so on. While you should be creative with the style of the title, headers, etc, make the paper look professional.

Length is not important. You should, though, try to address the questions completely but compactly. Focus on the quality of the words, not the quantity. I think 750-1,000 words should be sufficient.

Although there is not a section specifically devoted to the data, be sure to carefully explain each variable as it comes up in the text.

1. Give a brief overview of LaCour and Green's study. Why was it important?

2. Discuss the results from Questions 2 and 3 in the R component. Feel free to use any combination of histograms and QQ-plots to make your point. You probably don't need or want to include all of them.

   (a) Do the initial distributions of the CCAP and LaCour and green's two studies look similar? Do they look like they might be the same data?

   (b) Should they look alike? Think about how the respondents were chosen. Where are the respondents from? Should we expect the two sets of respondents to have similar feeling thermometer ratings?

   (c) Are the sample sizes the same? How might a researcher produce a smaller data set from a larger data set? Speculate about what might have happened.

3. Discuss the results from Questions 4 and 5 of the R component.

   (a) Use the scatterplots of the politician feeling thermometers from the pre- and post-election waves of the ANES to show what a typical pre- and post-feeling thermometer looks like. In particular, draw readers' attention to the number of respondents who take extreme and opposite opinions in each wave.

   (b) Compare scatterplots of the waves from LaCour and Green's data. Do they look similar? Speculate about what might have happened.

4. Discuss the results from Question 6. If the initial wave comes directly from the CCAP, where did the subsequent waves come from?

5. Conclude. Briefly highlight your main points and remind your readers what they should take away from your analysis.

Any sources that you rely on should be cited. See the APSA style guide (pdf) for the details on how citations typically work in political science. You can see this paper of mine (pdf) for an example. There is no need to include an abstract. For this particular, project you'll want to at least cite LaCour and Green's paper in *Science*.

3 Turning in Your Assignment

Compress the project folder `writing-assignment-2`. It should now be `writing-assignment-2.zip`. Rename this file `lastname-firstinitial-wa2.zip`. E-mail this compressed file to Abhi (abhisekh@exchange.tamu.edu) and me (crainey@tamu.edu) by e-mail. Also, submit the paper to through Turnitin. You can find the link on the eCampus sidebar. Lastly, bring a hard copy of the paper to class with a print-out of the R code (i.e., File → Print in RStudio) stapled to the back.

4 Grading

- R Code: 50%
  - The code runs and computes the correct quantities: 40%
  - The code is neatly written and thoroughly commented: 10% (Remember that I have to read it! See Hadley's style guide for some suggestions.)
- Paper: 50%
  - Analysis: 20%
  - Grammar and usage: 20%
  - Spelling and punctuation: 5%
  - Organization and format: 5%