Homework #4

Lena Stone

10/25/2020

There are a number of different variables we could look at in the COVID responses dataset. For my first attempt at a bivariate correlation, I want to examine the following variables against each other, each on a 0-100 scale:

- The first question asks people to rate whether or not they attended social gatherings in the past week, with "0" meaning they attended gatherings as normal and "100" meaning they heeded advice and didn't attend
- The second question asks people to rate whether or not they think *other* people believe social gatherings should be cancelled, with 0-100 expressed as percentages of the population

The distinction is a fine one, but it begs an interesting question: does the actual self-reported behavior of an individual correlate with their *perceptions* of what other people are thinking? In other words, do people who skip out on social gatherings believe that most of the population is in agreement with them? I'm going to hypothesize that for the most part, people's actions will correlate with their perceptions of what other people are doing.

Let's briefly see how much data we'll be handling:

```
## # A tibble: 3,460 x 2
##    SelfReported_Behavio_2 SOB_1
##                     <dbl> <dbl>
##  1                     74    NA
##  2                     20    75
##  3                     NA    NA
##  4                     28    42
##  5                     77    56
##  6                      1     1
##  7                     42    55
##  8                     64    60
##  9                     29     7
## 10                     55    45
## # ... with 3,450 more rows
```

Now we'll do some basic descriptive statistics. I'm calling people's self-reported behavior as "self" for simplicity and their perceptions of the public as "other."

For the "self" responses, we have the following rough descriptions:

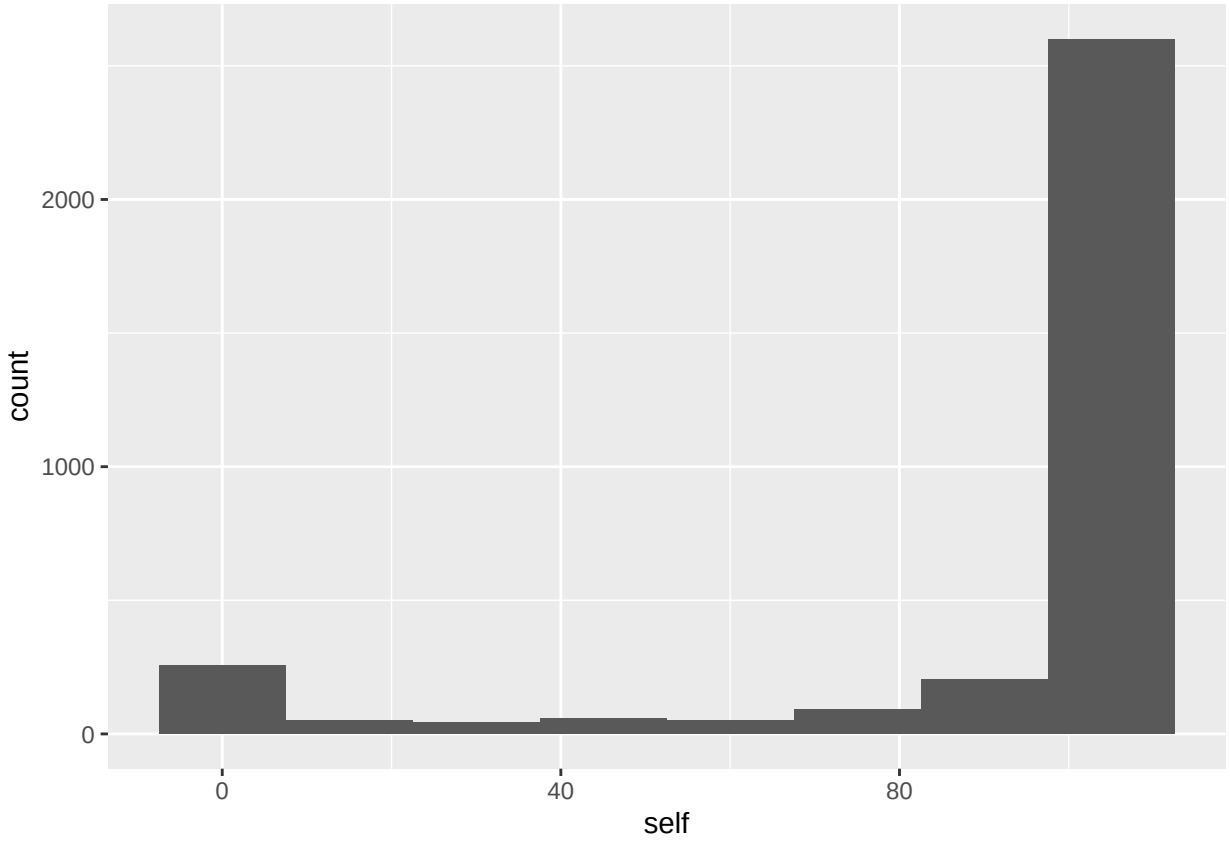
```
mean(self, na.rm=TRUE)
```

```
## [1] 87.47359
```

```
sd(self, na.rm=TRUE)
```

```
## [1] 29.45557
```

A basic graph of this first variable shows us that most people did NOT attend social gatherings. (A score of "0" means that they don't feel the statement "I did not attend social gatherings" applies to them at all, and "100" means they feel this statement applies very much.) The data is very concentrated on the right side, but there is a notable minor peak all the way on the left, which tells us that at least some people were gathering outside the home.



And now for the "other" variable:

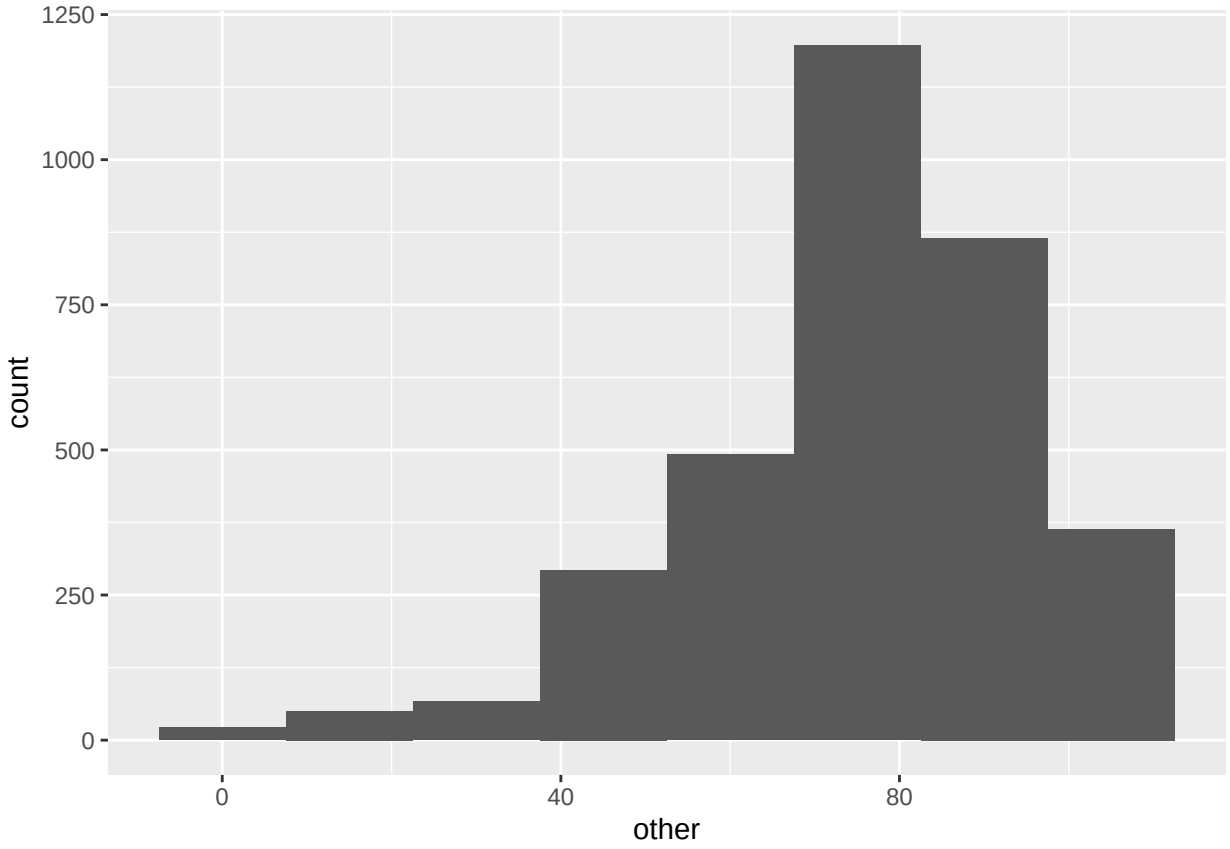
```
mean(other, na.rm=TRUE)
```

```
## [1] 75.15642
```

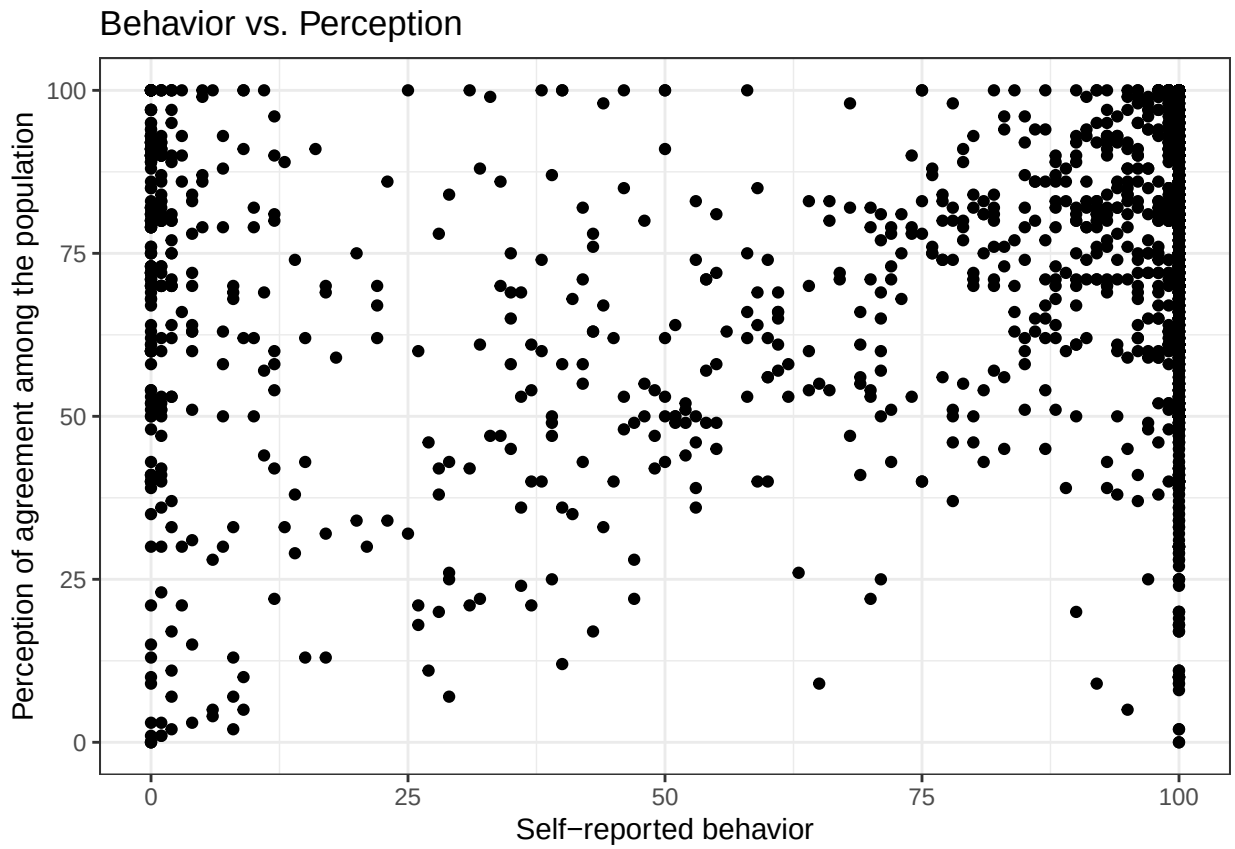
```
sd(other, na.rm=TRUE)
```

```
## [1] 19.19022
```

The responses are graphed below. (Remember, this graph gets at the perception that other people have regarding in-person social events.) We can see from the data that people generally think about 70% of the population is in agreement to stay home away from parties:



Now, let's put them together into a scatterplot.



Let's unpack this information, since the graph looks a little strange in that the data seems to be clustered on either end and wouldn't be well-represented with a line of best fit. For the most part, the data appears to skew towards the top of the graph, echoing our earlier conclusion that people think other citizens are agreeing to stay inside. In a rough visual analysis, it looks like the most dense concentration of data is in the top right–these are people who both *stayed inside themselves* and think *other people agree with them.* Although we haven't performed any inferential statistics up to this point, this seems to suggest a link between how people are acting themselves and their general impressions of others.

However, there was a good percent of the population who still attended social gatherings despite thinking other people would disapprove of their actions.

In general, I don't think this graph was a particularly good representation of the relationship, but I think it was an interesting first attempt at a bivariate correlationship with this dataset. Since I'm also hoping to work with a COVID-19 survey for the final paper, this was a good exercise in parsing out what variables might go together and experimenting a bit with different variable types.