Evidence of Data Fabrication (with Solutions)

In 2014, Michael LaCour and Donald Green published in *Science* an article titled "When Contact Changes Minds: An Experiment on Transmission of Support for Gay Equality." presenting the results of a randomized experiment allegedly conducted in Los Angeles county. The study found that a 20-minute conversation with a gay canvasser produced a large positive shift in feelings towards gay people as well as in support of same-sex marriage. Furthermore, the change in attitudes appeared to be long lasting as it was still present 9 months after the initial conversation.

In this exercise, we are going to explore the evidence put forth by David Broockman, Joshua Kalla, and Peter Aronow in their manuscript "Irregularities in LaCour (2014)", who argued that the data from LaCour and Green (2014) were likely to have been fabricated by LaCour and not collected as described in the article by asking respondents to fill out online surveys before and after sending canvassers door to door to converse with residents of the Los Angeles county. These allegations ultimately led to the retraction of the original article.

The data from the gay marriage experiment is in a file called "gay.csv" and Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is respondents.

variable	description
treatment	respondent's treatment assignment: 'No Contact', 'Same-Sex Marriage Script by Gay Canvasser', 'Same-Sex Marriage Script by Straight Canvasser', 'Recycling Script by Gay Canvasser', and 'Recycling Script by Straight Canvasser'
feel1	respondent's feeling towards gay men and lesbian women before canvassers went into the field (on a scale from 0 to 100)
feel2	respondent's feeling towards gay men and lesbian women 3 weeks after canvassers went into the field (on a scale from 0 to 100)
feel3	respondent's feeling towards gay men and lesbian women 6 weeks after canvassers went into the field (on a scale from 0 to 100)
feel4	respondent's feeling towards gay men and lesbian women 9 months after canvassers went into the field (on a scale from 0 to 100)

Table 1: Variables in "gay.csv"

In this experiment, households were randomly assigned to either: (1) not be contacted by any canvasser, (2) be contacted by a gay canvasser who would deliver a same-sex marriage script and reveal their sex-orientation over the course of the conversation, (3) be contacted by a straight canvasser who would deliver a same-sex marriage script, (4) be contacted by a gay canvasser who would deliver a recycling script, (5) be contacted by a straight canvasser who would deliver a recycling script.

Participants were asked to provide their feelings towards gay men and lesbian women in a scale of 0 to 100, where higher numbers reflect more positive feelings. Participants were asked to provide this same information, via an online survey, at four different points in time. The variable *feel1* captures the baseline data, that is, the feelings of the respondents before canvassers went into the field. The

data in the variables *feel2*, *feel3*, and *feel4* were collected 3 weeks, 6 weeks, and 9 months after the households received the different treatments. (The original dataset from LaCour and Green (2014) contains information about two independent studies. For the sake of simplicity, here we focus on the data from study 2. Conclusions do not change, however, when study 1 is also taken into consideration.)

Among other irregularities, Broockman et al. found that the baseline data from the gay marriage experiment were statistically indistinguishable from the 2012 Cooperative Campaign Analysis Project (CCAP) data based on a national sample of U.S. voters (something unlikely to occur had the data from the gay marriage experiment been truly collected from Los Angeles county residents as claimed by LaClour). The CCAP data is in a file called "CCAP.csv" and Table 2 shows the name and description of the one variable in this dataset, where the unit of observation is voters. (Note: Many respondents in the CCAP dataset did not provide their feelings towards gay men and lesbian women. These missing data were initially coded as NAs, but were later re-coded with a value of 50. This is common practice in feeling thermometer questions since 50 is the midpoint between 0 and 100, and thus, conveys neutrality.)

variable	description
feel	voter's feeling towards gay men and lesbian women (on a scale from 0 to 100)

Table 2: Variables in "CCAP.csv"

In this problem set, we practice creating histograms, computing descriptive statistics, creating scatter plots, and computing correlation coefficients. We also practice using optional arguments in R functions.

As always, we start by loading and looking at the data:

```
## load and look at the data
gay <- read.csv("gay.csv") # reads and stores data
head(gay, n=4) #shows first observations
##     treatment feel1 feel2 feel3 feel4
## 1 No Contact     2     6     0     6
## 2 No Contact    91    90   100    93
## 3 No Contact    46    67    38    46
## 4 No Contact     6     0    NA    16
```

```
## load and look at the data
ccap <- read.csv("CCAP.csv") #reads and stores data
head(ccap, n=4) #shows first observations
##    feel
## 1  100
## 2    6
## 3   61
## 4   50
```

Note that in the code above we only asked R to show us the first four observations since, in the head() functions, we set the optional argument n to equal 4. Recall: Whenever we specify multiple arguments inside a function, we separate them with a comma.

Now, to figure out how many observations are in each of the two datasets, we run:

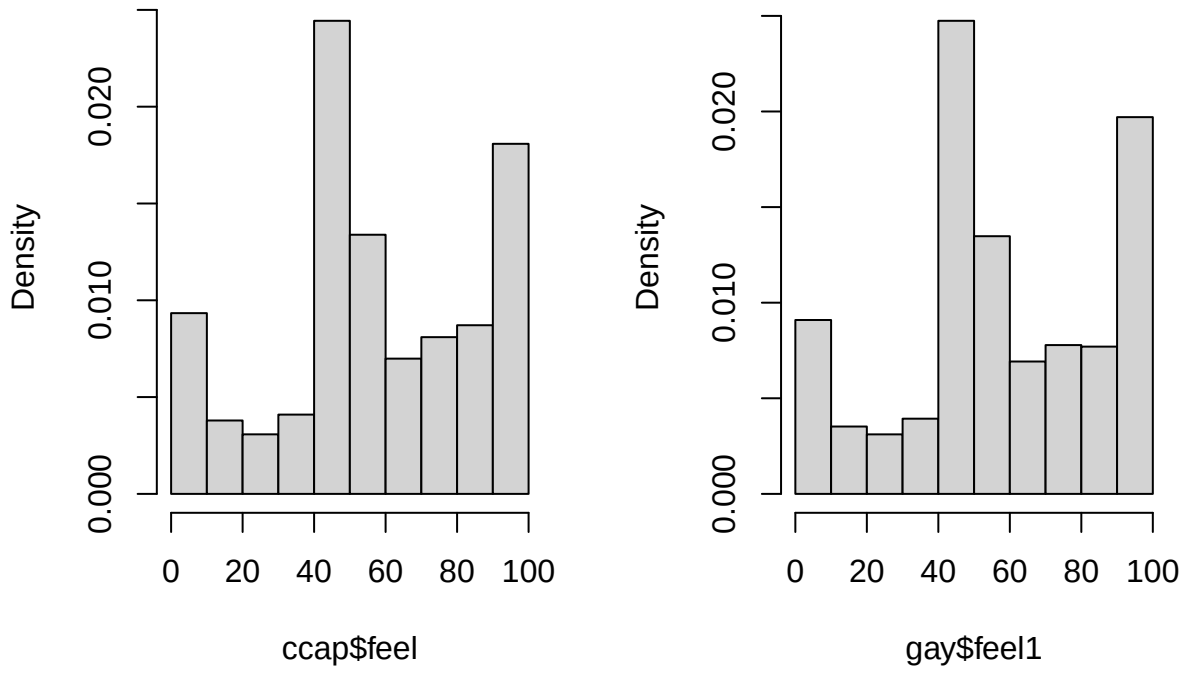
```
## provide dimensions of dataframes: rows, columns
dim(gay) # experiment
## [1] 2441    5
dim(ccap) # ccap
## [1] 43998     1
```

Based on the outputs above, we learn that in the gay marriage experiment dataset there is information about 2,441 respondents and that in the CCAP dataset there is information about 43,998 voters. The two datasets are quite different in size, which makes sense since the gay marriage experiment aimed to collect data from a representative sample of residents in the Los Angles county and the CCAP aimed to collect data from a representative sample of voters in the whole of the United States.

1. Let's compare the distribution of *feel* in the CCAP dataset to the distribution of *feel1* in the gay marriage experiment and figure out whether Broockman et al. were right when describing the two distributions as indistinguishable from each other. As you may recall, when comparing distributions with different number of observations, it is best to create density histograms, so go ahead and create (1) the density histogram of *feel* in the CCAP dataset and (2) the density histogram of *feel1* in the gay marriage experiment dataset. (Optional: If you want to ensure that both histograms have the same number of bins (let's say 10) so that it is easier to compare them to each other, we can set the optional argument breaks to equal 10 inside the function hist().) Do these two distributions look similar to each other? A yes/no answer will suffice. (10 points)

   R code:

```
## create density histograms
hist(ccap$feel, freq=FALSE, breaks=10) #of feel in ccap dataset
hist(gay$feel1, freq=FALSE, breaks=10) #of feel1 in experiment dataset
```



(Recall: The histogram of a variable is the visual representation of its distribution. The function in R to create a histogram is hist(). The only required argument is the code identifying the variable. To create the density histogram, we need to set the optional argument freq to equal FALSE inside the hist() function.)

Answer: Yes, despite the large difference in the number of observations in both datasets, the two distributions look very similar to each other.

2. To further compare the two distributions above, compute the mean, median, and standard deviation of both. Do the two distributions have similar descriptive statistics? A yes/no answer will suffice. (10 points)

   R code:

```
## compute means of feelings
mean(ccap$feel) # in ccap dataset
## [1] 58.09241
mean(gay$feel1) # in experiment dataset
## [1] 58.62884
```

```
## compute medians of feelings
median(ccap$feel) # in ccap dataset
## [1] 52
median(gay$feel1) # in experiment dataset
## [1] 52
```

```
## compute standard deviations of feelings
sd(ccap$feel) # in ccap dataset
## [1] 28.40383
sd(gay$feel1) # in experiment dataset
## [1] 28.55804
```

   Answers: Yes, the two distributions have very similar descriptive statistics. (Means: 58 vs. 58.6. Medians: 52 vs. 52. Standard deviations: 28.4 vs. 28.6.)

3. Given that the CCAP data is based on a national sample of U.S. voters and that the gay marriage experiment data was supposedly collected from a sample of residents from the Los Angeles county, would you have expected that the two distributions of feelings towards gay men and lesbian women were similar to each other? (5 points)

   Answer: No, I would have not. These two populations are very different from each other so I would have expected the distributions to be somewhat different. For example, I would have expected the residents of the Los Angeles county to be more likely to have positive feelings towards gay men and lesbian women than a national sample.

4. As evidence that the data from the gay marriage experiment was not collected as claimed by LaCour but rather fabricated, Broockman et al. pointed out that it could have been easily generated by a computer by:

   (a) drawing a random sample of observations from the CCAP dataset to produce the baseline feelings (*feel1*), which would explain the similarities between the two distributions observed in questions 1 and 2 above, and then

   (b) adding random noise to *feel1* to produce the follow-up waves (*feel2*, *feel3*, and *feel4*), which would explain the unusual high levels of stability in respondent's feelings towards gay men and lesbian women over time found in the gay marriage experiment.

   Background information: Most panel survey data (that is, data from surveys that ask the same group of respondents for their opinions several times over the course of a few weeks/months/years) show a fair amount of variation in respondent's opinions over time, especially in questions that use large scales such as 0 to 100. First, respondents are not very good at being precise when providing their feelings in a 0 to 100 scale, thus, even if their feelings have not changed they will typically provide different numbers when asked the same question at different points in time. Second, there is also some people who do change their minds over time.

   To explore the stability of respondent's feelings in the gay marriage experiment data over time, create the following three visualizations:

   (a) the scatter plot of *feel1* and *feel2* (to show the relationship between respondents feelings before canvassers went into the field and three weeks after),

   (b) the scatter plot of *feel1* and *feel3* (to show the relationship between respondents feelings before canvassers went into the field and six weeks after),

   (c) the scatter plot of *feel1* and *feel4* (to show the relationship between respondents feelings before canvassers went into the field and 9 months after).

   To simplify the comparisons, include only the feelings from respondents who were in the control group, that is, who were not contacted by any canvasser and, thus, should not experience changes in their feelings due to exposure to any of the treatments. (Hint: The `[]` and `==` operators might be helpful when subsetting the variables.)

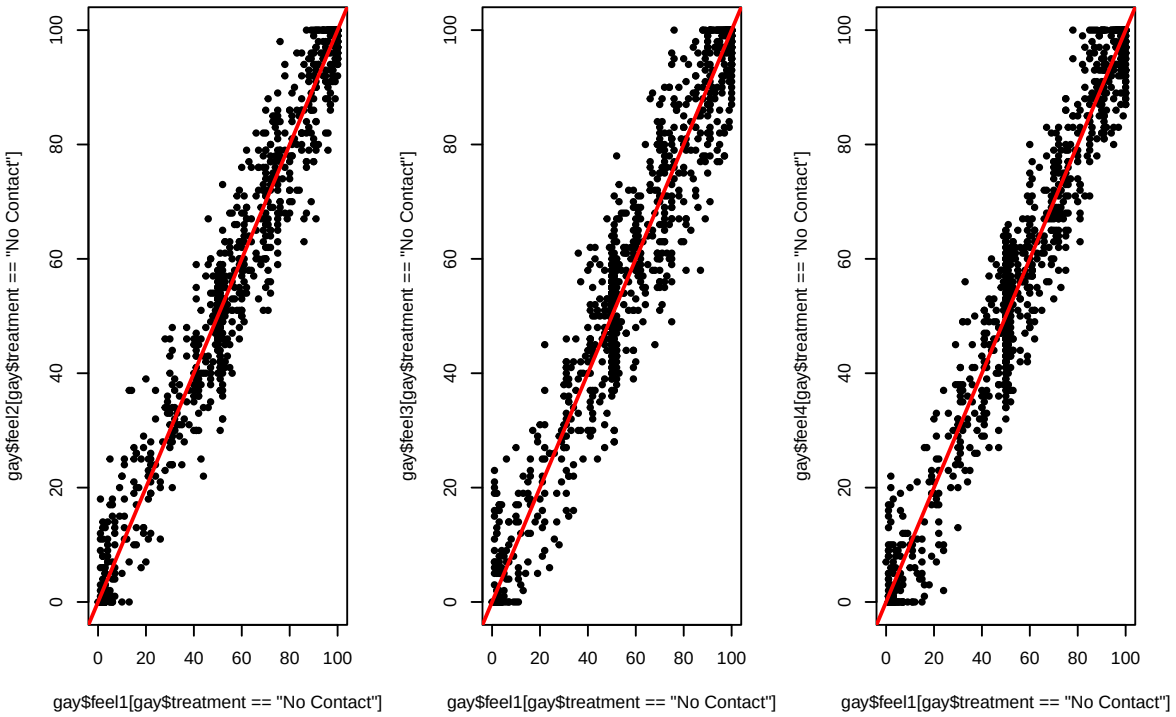
   Does it look that the respondents in the gay marriage experiment dataset have relatively stable feelings over the whole period of time? A yes/no answer will suffice. Note that if all respondents' feelings remained *exactly* the same, all the observations in the scatter plots would be on top of the 45 degree angle line. (10 points)

   R code:

```
## create scatter plots for respondents in the control group
plot(gay$feel1[gay$treatment=="No Contact"],
     gay$feel2[gay$treatment=="No Contact"]) #between feel1 and feel 2
```

```
plot(gay$feel1[gay$treatment=="No Contact"],
     gay$feel3[gay$treatment=="No Contact"]) #between feel1 and feel 2
```

```
plot(gay$feel1[gay$treatment=="No Contact"],
     gay$feel4[gay$treatment=="No Contact"]) #between feel1 and feel 2
```



(Recall: the scatter plot of two variables shows how the two variables relate to each other. The function in R to create a scatter plot is plot(). The only two required arguments are the code identifying the two variables (first the X variable and then the Y variable). In this case, however, it doesn't matter which variable is displayed on the X axis and which is on the Y axis. Also, recall: We use [] to subset a variable; inside the square brackets, we specify the criterion of selection. For example, we can use the relational operator == to set a logical test; only the observations for which the logical test are true will be extracted. In the cases above, we extract from *feel1*, *feel2*, *feel3*, and *feel4* the observations that belong to the respondents that were assigned to not be contacted by any canvasser (for which *treatment* equals "No Contact' ') and then plot those observations.)

Answer: Yes, respondent's feelings look very stable, plus or minus some random noise. In neither of the three scatter plots, there are any major outliers, that is, extreme values, away from the 45 degree angle line. (Note: Although we didn't ask you to do this, in the scatter plots above, we added a 45 degree angle line in red by running abline(0,1, col="red") after each plot() function. (You can search abline in the help tab of RStudio to learn more about how this function works and the different optional arguments it can take.) Then, we can see more clearly that although respondent's feelings vary ever so slightly from one survey to the next, they never do so in any major way.)

5. To further explore the stability of respondent's feelings over time, calculate the following correlations coefficients:

   (a) the correlation between *feel1* and *feel2* (to measure the direction and strength of the linear association between respondents feelings before canvassers went into the field and three weeks after),

   (b) the correlation between *feel1* and *feel3* (to measure the direction and strength of the linear association between respondents feelings before canvassers went into the field and six weeks after),

   (c) the correlation between *feel1* and *feel4* (to measure the direction and strength of the linear association between respondents feelings before canvassers went into the field and 9 months after).

   Again, to simplify the comparisons, include only the feelings from respondents who were in the control group. Are these variables highly correlated with each other? A yes/no answer will suffice. (10 points)

   (Important note: Some of the data in the gay marriage experiment is missing. See, for example, the value of *feel3* of the fourth observation shown by the function head() on page 2. To ask R to take into consideration only non-missing information when computing a correlation coefficient, we can specify the optional argument use and set it to equal "complete.obs". Without this optional argument, R will return an NA whenever any of the two specified variables contain any missing information.)

   R code:

```
## compute correlations
cor(gay$feel1[gay$treatment=="No Contact"], gay$feel2[gay$treatment=="No Contact"],
    use="complete.obs") # between feel1 and feel2
##  [1] 0.9734449

cor(gay$feel1[gay$treatment=="No Contact"], gay$feel3[gay$treatment=="No Contact"],
    use="complete.obs") # between feel1 and feel3
##  [1] 0.9594085

cor(gay$feel1[gay$treatment=="No Contact"], gay$feel4[gay$treatment=="No Contact"],
    use="complete.obs") # between feel1 and feel4
##  [1] 0.9709017
```

   (Recall: The function in R to compute a correlation coefficient is cor(). The only two required arguments are the code identifying the two variables. The order of the variables does not matter since the correlation between X and Y is the same as the correlation between Y and X.)

   Answer: Yes, these variables are highly correlated with each other. The correlation coefficients are remarkably close to 1. They are 0.97, 0.96, and 0.97. This further corroborates that respondent's feelings in the gay marriage experiment data are incredibly stable.

6. Given your answers to questions 1 through 5, would you conclude that it is likely that the gay marriage experiment data was fabricated and not collected as described in the article? A yes/no answer will suffice. (5 points)

   Answer: Yes, given everything we have observed in questions 1 through 5, it seems likely that the data was fabricated.