Data Exploration: Emotional Arousal

Kayla Huang

November 4, 2021

** under other documents: pdf of slides that he used in section ** take a look at those data sets (still good to know what's in those data sets) - survey data from different populations ** survey should be good ** heightened political rhetoric in terms of immigration and covid ** has this caused citizens to become more informed

In this Data Exploration assignment we will explore Clifford and Jerit's (2018) findings about the effects of disgust and anxiety on political learning.

If you have a question about any part of this assignment, please ask! Note that the actionable part of each question is **bolded**.

Emotional Arousal: Disgust and Anxiety

Data Details:

- File Name: `Study1ReplicationData.dta`
- Source: These data are from Study 1 in Clifford and Jerit (2018).

Variable Name	Variable Description
`treat_rand1`	Treatment assignment: 1-Low Anxiety/Low Disgust, 2-High Anxiety/Low Disgust, 3-Low Anxiety/High Disgust, and 4-High Anxiety/High Disgust
`Q11_1`	Self reported feeling of how well DISGUST describes respondent's emotional reaction to virus: 1-Not Well At All, 2- Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well, 8-Skipped
`Q11_2`	Self reported feeling of how well GROSSED OUT describes respondent's emotional reaction to virus: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well, 8-Skipped
`Q11_3`	Self reported feeling of how well REPULSED describes respondent's emotional reaction to virus: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well, 8-Skipped
`Q11_4`	Self reported feeling of how well AFRAID describes respondent's emotional reaction to virus: 1-Not Well At All, 2- Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well, 8-Skipped

Variable Name	Variable Description
Q11_5	Self reported feeling of how well ANXIOUS describes respondent's emotional reaction to virus: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well, 8-Skipped
Q11_6	Self reported feeling of how well WORRIED describes respondent's emotional reaction to virus: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well, 8-Skipped
Q12_1	Identification of FATIGUE as a symptom: 1-Yes, 2- No
Q12_2	Identification of HEADACHES as a symptom: 1-Yes, 2- No
Q12_3	Identification of DIARRHEA as a symptom: 1-Yes, 2- No
Q12_4	Identification of JOINT PAIN as a symptom: 1-Yes, 2- No
Q12_5	Identification of BOILS as a symptom: 1-Yes, 2- No
Q12_6	Identification of WARTS as a symptom: 1-Yes, 2- No
Q12_7	Identification of FEVER as a symptom: 1-Yes, 2- No
Q13	Identification of method of disease transmission: 1-Person to Person Contact, 2- The air, 3- Animals, 4-Insects, 5-Food, 8-Skipped
Q14	Identification of there being a cure for the virus: 1-Yes, 2-No, 8-Skipped
Q15	Requested additional information be sent to them: 1-Yes, 2-No, 8-Skipped
Q16_1	Topic of requested information AFFECTED COUNTRIES: 1-Yes, 2-No, 9-Not Asked
Q16_2	Topic of requested information TRV IN US: 1-Yes, 2-No, 9-Not Asked
Q16_3	Topic of requested information TRV TRANSMISSION: 1-Yes, 2-No, 9-Not Asked
Q16_4	Topic of requested information AT RISK POPULATION: 1-Yes, 2-No, 9-Not Asked
Q16_5	Topic of requested information DEATH TOLL: 1-Yes, 2-No, 9-Not Asked
Q16_6	Topic of requested information PROGRESS ON CURE: 1-Yes, 2-No, 9-Not Asked
Q16_7	Topic of requested information SYMPTOMS: 1-Yes, 2-No, 9-Not Asked
Q17_6	Self-reported likelihood of looking up more info: 1-Not likely at all, 2-Not too likely, 3-Somewhat likely, 4-Very likely, 5-Extremely likely, 8-Skipped
Q17_7	Self-reported likelihood of talking with friends or family about disease in next week: 1-Not likely at all, 2-Not too likely, 3-Somewhat likely, 4-Very likely, 5-Extremely likely, 8-Skipped
page_article_timing	Time spent in seconds on page containing article about disease
Q18	Self-reported gender: 1-Female, 0-Not Female
Q19	Self-reported race/ethnicity: 1-White, 2-Black, 3-Hispanic, 4-Asian, 5-Native American, 6-Mixed Race, 7-Other
Q20	Self-reported education: 1-No HS, 2-High school graduate, 3-Some college, 4-2 year degree, 5-4 year degree, 6-Post-graduate degree
Q21	Self-reported partisanship: 1-Strong Democrat, 2-Not very strong Democrat, 3-Lean Democrat, 4-Independent, 5-Lean Republican, 6-Not very strong Republican, 7- Strong Republican, 8-Not Sure

Variable Name	Variable Description
Q22	Self-reported voter registration status: 1-Yes, 2-No, 3-Don't know
Q23	Self-reported ideology: 1-Very Liberal, 2-Liberal, 3-Moderate, 4-Conservative, 5-Very Conservative

```
#Load the data for Study 1
Study1_preprocess <- read_dta("Study1ReplicationData.dta")
Study1_preprocess
```

```
## # A tibble: 1,000 x 33
##    treat_rand1   Q11_1   Q11_2   Q11_3   Q11_4   Q11_5   Q11_6   Q12_1   Q12_2
##      <dbl+lbl> <dbl+l> <dbl+l> <dbl+l> <dbl+l> <dbl+l> <dbl+l> <dbl+l> <dbl+l>
##  1 1 [Low Anx~ 2 [Not~ 2 [Not~ 2 [Not~ 2 [Not~ 2 [Not~ 2 [Not~ 2 [No]  2 [No]
##  2 4 [High An~ 4 [Ver~ 4 [Ver~ 4 [Ver~ 3 [Som~ 3 [Som~ 4 [Ver~ 1 [Yes] 2 [No]
##  3 1 [Low Anx~ 2 [Not~ 1 [Not~ 1 [Not~ 1 [Not~ 1 [Not~ 1 [Not~ 1 [Yes] 1 [Yes]
##  4 2 [High An~ 4 [Ver~ 2 [Not~ 4 [Ver~ 3 [Som~ 4 [Ver~ 3 [Som~ 2 [No]  1 [Yes]
##  5 1 [Low Anx~ 1 [Not~ 1 [Not~ 1 [Not~ 1 [Not~ 1 [Not~ 3 [Som~ 1 [Yes] 1 [Yes]
##  6 3 [Low Anx~ 2 [Not~ 2 [Not~ 2 [Not~ 3 [Som~ 3 [Som~ 3 [Som~ 2 [No]  2 [No]
##  7 1 [Low Anx~ 1 [Not~ 1 [Not~ 2 [Not~ 2 [Not~ 2 [Not~ 2 [Not~ 1 [Yes] 1 [Yes]
##  8 4 [High An~ 1 [Not~ 1 [Not~ 1 [Not~ 1 [Not~ 1 [Not~ 2 [Not~ 1 [Yes] 2 [No]
##  9 2 [High An~ 3 [Som~ 4 [Ver~ 3 [Som~ 2 [Not~ 2 [Not~ 2 [Not~ 1 [Yes] 1 [Yes]
## 10 1 [Low Anx~ 5 [Ext~ 5 [Ext~ 5 [Ext~ 4 [Ver~ 4 [Ver~ 4 [Ver~ 1 [Yes] 1 [Yes]
## # ... with 990 more rows, and 24 more variables: Q12_3 <dbl+lbl>,
## #   Q12_4 <dbl+lbl>, Q12_5 <dbl+lbl>, Q12_6 <dbl+lbl>, Q12_7 <dbl+lbl>,
## #   Q13 <dbl+lbl>, Q14 <dbl+lbl>, Q15 <dbl+lbl>, Q16_1 <dbl+lbl>,
## #   Q16_2 <dbl+lbl>, Q16_3 <dbl+lbl>, Q16_4 <dbl+lbl>, Q16_5 <dbl+lbl>,
## #   Q16_6 <dbl+lbl>, Q16_7 <dbl+lbl>, Q17_6 <dbl+lbl>, Q17_7 <dbl+lbl>,
## #   page_article_timing <dbl>, Q18 <dbl>, Q19 <dbl+lbl>, Q20 <dbl+lbl>,
## #   Q21 <dbl+lbl>, Q22 <dbl+lbl>, Q23 <dbl+lbl>
```

To date you have read in data from both .csv and .RData files. The data for this week are stored in another common file type, .dta files. This is the format for data exported using Stata, the other most commonly used statistical software package besides R. Reading these files requires the read_dta() function in the haven package, so be sure to install it if you have not already!

##Question 1 DO NOT SKIP Cleaning and organizing data is an important part of any research process. Note: Your blog posts should not address the data cleaning portion of the assignment but rather the content of the material

###Part a The data is not currently in the most usable condition. As currently read in, many of the dataset's variables are somewhat unhelpfully labelled with just the survey question number. In order to make them more intuitive to work with we can rename them easily using the dplyr package. Here we will make use of the rename_with() function. The .cols argument in the rename_with() function specifies which columns should be renamed, which takes either the original name of the variable, the index, or a logical argument. We will explore each of these methods beginning with renaming variables by name. **Modify the code below to rename variables Q13, Q14, and Q15. Make sure the list of replacement variable names has the same number as the number of variables you're renaming. Also be sure to save the modified dataset as a new object to continue working with it.**

```
#Method 1: Renaming by name
Study1_processing1 <- Study1_preprocess %>%
  rename_with(.cols = c(Q13, Q14, Q15), ~c("transmission", "cure", "add_info"))
```

###Part b Many other variables are also unhelpfully labelled. For example the last six variables in the data are demographic characteristics (gender, race, education, party id, voter registration status, and ideology). You can also rename variables by position using the following code. **Modify the code below to rename the demographic variables. Make sure the list of replacement variable names has the same number as the number of variables you're renaming. Also be sure to save the modified dataset as a new object to continue working with it.**

```
#Method 2: Renaming by index/position
#last_col() is just a function that returns the index number of the last variable in the dataset (33 in
Study1_processing2 <- Study1_processing1 %>%
  rename_with(.cols = last_col(offset = 5):last_col(), ~c("gender", "race", "education", "party_id", "v
```

###Part c We can also rename all variables that satisfy a logical condition. Note that every variable relating to emotional reaction of the respondent is labelled as a part of question 11 (Q11). **Modify the code below to rename the demographic variables. Make sure the list of replacement variable names has the same number as the number of variables you're renaming. Also be sure to save the modified dataset as a new object to continue working with it.**

```
#Method 3: Renaming by logical condition
Study1_processing3 <- Study1_processing2 %>%
  rename_with(.cols = contains("Q11"), ~c("disgust", "grossed_out", "repulsed", "afraid", "anxious", "w
```

###Part d We may not need all the variables in the dataset. For instance, the analysis below will not rely on knowing what topic of additional information people were interested in (Q16_1 through Q16_7). **The code below will search for a string in the variable labels and return only the variables that DO NOT include that string. Modify the code and use it to drop the unneeded variables from the dataset. Be sure to save the modified dataset as a new object to continue working with it.**

```
Study1_processing4 <- Study1_processing3 %>%
  select(-contains("Q16"))
```

###Part e Another issue often encountered is that the way missing data is coded can vary across data sources. Here responses are marked with the value of 8 for the respondent skipping the question. The functions mutate() and across() used in tandem are useful for recoding many variables at once in the same manner. across() specifies which columns to mutate and what function to apply to them all, in this case we'll want to use the function na_if() which recodes a variable as NA if it matches the second argument. **Modify the code below to change the relevant data columns to be binary variables with NA for missing data. Hint: For this dataset all values of 8 indicate a kind of missing data.**

```
Study1_processing5 <- Study1_processing4 %>%
  mutate(across(.cols = c(disgust, grossed_out, repulsed, afraid, anxious, worried,
                          transmission, cure, add_info, Q17_6, Q17_7), ~na_if(.,8)))
```

###Part f Some of the binary data are also coded with 1s and 2s as opposed to 0s and 1s. These include the identification of symptoms, the request for additional information, identification that there is a cure, and voter registration status. **Use the mutate function to recode those variables to be binary with 1 for yes and 0 for no. Hint: You can look at the solutions for previous data exercises but you'll want to combine the mutate() and ifelse() functions.**

```
Study1_processing6 <- Study1_processing5 %>%
  mutate(across(.cols = c(Q12_1, Q12_2, Q12_3, Q12_4, Q12_5, Q12_6, Q12_7), ~ifelse(. == 1, 0, 1))) %>%
  mutate(across(.cols = c("transmission", "cure", "add_info"), ~ifelse(. == 1, 0, 1)))

Study1_processing6
```

```
## # A tibble: 1,000 x 26
##    treat_rand1 disgust grossed_out repulsed  afraid anxious worried Q12_1 Q12_2
##      <dbl+lbl> <dbl+l>   <dbl+lbl> <dbl+lb> <dbl+l> <dbl+l> <dbl+l> <dbl> <dbl>
##  1 1 [Low Anx~ 2 [Not~ 2 [Not too~ 2 [Not ~ 2 [Not~ 2 [Not~ 2 [Not~     1     1
##  2 4 [High An~ 4 [Ver~ 4 [Very we~ 4 [Very~ 3 [Som~ 3 [Som~ 4 [Ver~     0     1
##  3 1 [Low Anx~ 2 [Not~ 1 [Not wel~ 1 [Not ~ 1 [Not~ 1 [Not~ 1 [Not~     0     0
##  4 2 [High An~ 4 [Ver~ 2 [Not too~ 4 [Very~ 3 [Som~ 4 [Ver~ 3 [Som~     1     0
##  5 1 [Low Anx~ 1 [Not~ 1 [Not wel~ 1 [Not ~ 1 [Not~ 1 [Not~ 3 [Som~     0     0
##  6 3 [Low Anx~ 2 [Not~ 2 [Not too~ 2 [Not ~ 3 [Som~ 3 [Som~ 3 [Som~     1     1
##  7 1 [Low Anx~ 1 [Not~ 1 [Not wel~ 2 [Not ~ 2 [Not~ 2 [Not~ 2 [Not~     0     0
##  8 4 [High An~ 1 [Not~ 1 [Not wel~ 1 [Not ~ 1 [Not~ 1 [Not~ 2 [Not~     0     1
##  9 2 [High An~ 3 [Som~ 4 [Very we~ 3 [Some~ 2 [Not~ 2 [Not~ 2 [Not~     0     0
## 10 1 [Low Anx~ 5 [Ext~ 5 [Extreme~ 5 [Extr~ 4 [Ver~ 4 [Ver~ 4 [Ver~     0     0
## # ... with 990 more rows, and 17 more variables: Q12_3 <dbl>, Q12_4 <dbl>,
## #   Q12_5 <dbl>, Q12_6 <dbl>, Q12_7 <dbl>, transmission <dbl>, cure <dbl>,
## #   add_info <dbl>, Q17_6 <dbl+lbl>, Q17_7 <dbl+lbl>,
## #   page_article_timing <dbl>, gender <dbl>, race <dbl+lbl>,
## #   education <dbl+lbl>, party_id <dbl+lbl>, voter_reg <dbl+lbl>,
## #   ideo <dbl+lbl>
```

###Part g Lastly we may be interested not just in individual symptom recall but also whether or not the respondent correctly remembered all the symptoms in their treatment with no mistakes. **Use the mutate() function to add a variable capturing whether or not the respondent correctly identified the symptoms in their treatment. Remember that which set of symptoms are correct differs somewhat by treatment! Warning: This one may be a bit tricky, no worries if you don't quite get it. Feel free to skip to the next question.**

```
Study1_processing7 <- Study1_processing6 %>%
  mutate(all_symptoms = ifelse(Q12_1 == 1 & Q12_2 == 1 & Q12_3 == 1 & Q12_4 == 1 & Q12_5 == 1 & Q12_6 ==
Study1_processing7
```

```
## # A tibble: 1,000 x 27
##    treat_rand1 disgust grossed_out repulsed  afraid anxious worried Q12_1 Q12_2
##      <dbl+lbl> <dbl+l>   <dbl+lbl> <dbl+lb> <dbl+l> <dbl+l> <dbl+l> <dbl> <dbl>
##  1 1 [Low Anx~ 2 [Not~ 2 [Not too~ 2 [Not ~ 2 [Not~ 2 [Not~ 2 [Not~     1     1
##  2 4 [High An~ 4 [Ver~ 4 [Very we~ 4 [Very~ 3 [Som~ 3 [Som~ 4 [Ver~     0     1
##  3 1 [Low Anx~ 2 [Not~ 1 [Not wel~ 1 [Not ~ 1 [Not~ 1 [Not~ 1 [Not~     0     0
##  4 2 [High An~ 4 [Ver~ 2 [Not too~ 4 [Very~ 3 [Som~ 4 [Ver~ 3 [Som~     1     0
##  5 1 [Low Anx~ 1 [Not~ 1 [Not wel~ 1 [Not ~ 1 [Not~ 1 [Not~ 3 [Som~     0     0
##  6 3 [Low Anx~ 2 [Not~ 2 [Not too~ 2 [Not ~ 3 [Som~ 3 [Som~ 3 [Som~     1     1
##  7 1 [Low Anx~ 1 [Not~ 1 [Not wel~ 2 [Not ~ 2 [Not~ 2 [Not~ 2 [Not~     0     0
##  8 4 [High An~ 1 [Not~ 1 [Not wel~ 1 [Not ~ 1 [Not~ 1 [Not~ 2 [Not~     0     1
##  9 2 [High An~ 3 [Som~ 4 [Very we~ 3 [Some~ 2 [Not~ 2 [Not~ 2 [Not~     0     0
## 10 1 [Low Anx~ 5 [Ext~ 5 [Extreme~ 5 [Extr~ 4 [Ver~ 4 [Ver~ 4 [Ver~     0     0
## # ... with 990 more rows, and 18 more variables: Q12_3 <dbl>, Q12_4 <dbl>,
## #   Q12_5 <dbl>, Q12_6 <dbl>, Q12_7 <dbl>, transmission <dbl>, cure <dbl>,
## #   add_info <dbl>, Q17_6 <dbl+lbl>, Q17_7 <dbl+lbl>,
## #   page_article_timing <dbl>, gender <dbl>, race <dbl+lbl>,
## #   education <dbl+lbl>, party_id <dbl+lbl>, voter_reg <dbl+lbl>,
## #   ideo <dbl+lbl>, all_symptoms <dbl>
```

###Part h Use the tools above to alter the data in whatever way you see fit. Some examples could be renaming remaining variables, creating new binary variables that identify if respondents are part of a certain

racial or ethnic group (these would be necessary for including race in a regression for example), or any other transformation of the data. **Use the tools above or any others to modify the data to more useful for the following exercises in any way you see fit. Save this final version as the object you'll use for future questions.**

Question 2

###Part a What are the treatments in Study 1? **How many treatment conditions are there in Study 1? What are they? How many respondents were in each condition? Hint: Look at page 269 for the treatment conditions of Study 1**

###Part b The paper lays out three distinct hypotheses concerning the impact of disgust on information uptake and search. **What are the three hypotheses? Which outcome variables in Study 1 speak to which of these hypotheses? Hint: Look at pages 267 and 268 for the hypotheses**

##Question 3

###Part a Often times when running an experiment a researcher will include a 'manipulation check' to confirm that the treatment was noticed and is having some of its intended effect. In this study they ask about the emotional response to the fictitious virus using two three item emotional indices: anxiety (afraid, anxious, worried) and disgust (disgusted, grossed out, repulsed). Take the average of the responses for each emotional index and check to see if the treatments had the desired manipulation. **Compare the average anxiety and disgust responses for each of the four treatments. Are there statistically significant differences in the way one would expect?**

```
q3 <- Study1_processing7 %>%
  mutate(treatment_group = case_when(treat_rand1 == 1 ~ "Low Anxiety/Low Disgust",
                                     treat_rand1 == 2 ~ "High Anxiety/Low Disgust",
                                     treat_rand1 == 3 ~ "Low Anxiety/High Disgust",
                                     treat_rand1 == 4 ~ "High Anxiety/High Disgust"))

q3 <- q3 %>% group_by(treatment_group) %>% summarize(anxious_avg = mean(anxious), disgust_avg = mean(dis
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
q3
```

```
## # A tibble: 4 x 3
##   treatment_group           anxious_avg disgust_avg
##   <chr>                           <dbl>       <dbl>
## 1 High Anxiety/High Disgust          NA          NA
## 2 High Anxiety/Low Disgust           NA          NA
## 3 Low Anxiety/High Disgust           NA          NA
## 4 Low Anxiety/Low Disgust            NA           2.18
```

###Part b Choose one of the three hypotheses identified in Question 2. Compare the responses on one or more of the outcomes relevant to that hypothesis. Is the hypothesis supported by the data? **For one or more outcome variables relevant to one of the three hypotheses check for statistically significant differences. What can we conclude from the data?**

##Question 4: DATA SCIENCE QUESTION ###Part a As we have spoken about in class, using indexes of multiple measures aimed at a single concept is often more reliable than using only one measure. However

how is one to know that all the measures in the index are related to the same concept? Cronbach's alpha is a measure of how internally consistent the answers to multiple questions are. It is given by the formula:

$$\alpha = \frac{N \times \sum c}{\sum v + (N-1) \times \sum c}$$

Where N is the number of items in the index, $\sum c$ is the sum of the covariances for item pairs, and $\sum v$ is the sum of the variance for the items. **Using the formula calculate Cronbach's alpha for the disgust index and the anxiety indexes used in the paper.**

###Part b In R, one can use the function cronbach.alpha() from the ltm package. Generally scales or indices with a Cronbach's alpha below 0.7 are considered insufficiently internally consistent for use. **Calculate Cronbach's alpha for the disgust index and the anxiety indexes using the function in R. Does this match your previous calculation? Are each of the scales sufficiently internally consistent?**

```
install.packages("ltm")
```

```
## Installing package into '/Users/kayla/Library/R/4.0/library'
## (as 'lib' is unspecified)
```

```
## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror
```

```
library(ltm)
```

```
## Error in library(ltm): there is no package called 'ltm'
```

###Part c Create a multivariate regression model for an outcome variable of your choosing. Carefully interpret the results. **Create a multivariate linear regression model for any outcome of your choosing with any covariates of your choice. Justify your choice of models and what the result may indicate. Carefully interpret the coefficients for the model.**

```
m1 <- lm(Q17_6 ~ disgust + anxious, data = Study1_processing7)
summary(m1)
```

```
##
## Call:
## lm(formula = Q17_6 ~ disgust + anxious, data = Study1_processing7)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.64136 -1.03859 -0.03859  0.85352  2.96141
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.63789    0.09431  17.368  < 2e-16 ***
## disgust      0.10790    0.03791   2.846  0.00451 **
## anxious      0.29280    0.04211   6.953 6.51e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.229 on 982 degrees of freedom
##   (15 observations deleted due to missingness)
## Multiple R-squared:  0.1053, Adjusted R-squared:  0.1035
## F-statistic:  57.8 on 2 and 982 DF,  p-value: < 2.2e-16
```

###Part d You can create coefficient plots using regression coefficients as well. Use the results of part c to create a plot of the coefficients in your regression model. **Create a plot of the regression coefficients with their 95% confidence intervals. Be sure to include a line demarcating 0. Hint: the confint() function can take a regression object and return the upper and lower bounds of the confidence intervals.**

Emotional Arousal: Disgust

Data Details:

- File Name: `Study2ReplicationData.dta`
- Source: These data are from Study 2 in Clifford and Jerit (2018).

Variable Name	Variable Description
`treatment`	Treatment assignment: 1-Control, 2-Disgusting Imagery/No Map, 3-Map/No Disgusting Imagery, and 4-Disgusting Imagery and Map
`page_time`	Time spent viewing page with treatment article
`symptpercent`	Belief about percentage of people who contract Dengue Fever but never experience symptoms: 1-0%, 2-20%, 3-40%, 4-60%, 5-80%
`Mexico`	Identify MEXICO to be at risk for spread of Dengue: 0-No, 1-Yes
`SouthAmerica`	Identify SOUTH AMERICA to be at risk for spread of Dengue: 0-No, 1-Yes
`Africa`	Identify AFRICA to be at risk for spread of Dengue: 0-No, 1-Yes
`Canada`	Identify CANADA to be at risk for spread of Dengue: 0-No, 1-Yes
`Russia`	Identify RUSSIA to be at risk for spread of Dengue: 0-No, 1-Yes
`Europe`	Identify EUROPE to be at risk for spread of Dengue: 0-No, 1-Yes
`length`	Belief about how long symptoms of Dengue Fever typically last: 1-A few days, 2-A week, 3-Two to three weeks, 4-A month or more
`fever`	Identification of FEVER as a symptom: 0-No, 1-Yes
`headache`	Identification of HEADACHE as a symptom: 0-No, 1-Yes
`jointpain`	Identification of JOINT PAIN as a symptom: 0-No, 1-Yes
`rash`	Identification of RASH as a symptom: 0-No, 1-Yes
`bleeding`	Identification of BLEEDING FROM EYES, NOSE, AND GUMS as a symptom: 0-No, 1-Yes
`nausea`	Identification of NAUSEA as a symptom: 0-No, 1-Yes
`seizures`	Identification of SEIZURES as a symptom: 0-No, 1-Yes
`breathing`	Identification of DIFFICULTY BREATHING as a symptom: 0-No, 1-Yes
`infosearch`	Self-reported likelihood of looking up more info: 1-Not likely at all, 2-Not too likely, 3-Somewhat likely, 4-Very likely, 5-Extremely likely
`talk`	Self-reported likelihood of talking with friends or family about disease in next week: 1-Not likely at all, 2-Not too likely, 3-Somewhat likely, 4-Very likely, 5-Extremely likely
`infosession`	Request invitation to information session about Dengue Fever: 0-No, 1-Yes

Variable Name	Variable Description
learn	Request additional info about Dengue Fever in survey: 0-No, 1-Yes
infogiveemail	Provided email address to receive invitation to info session: 0-No, 1-Yes
E_disgust	Self reported feeling of how well DISGUSTED describes respondent's emotional reaction to disease: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well
E_gross	Self reported feeling of how well GROSSED OUT describes respondent's emotional reaction to disease: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well
E_resentment	Self reported feeling of how well RESENTFUL describes respondent's emotional reaction to disease: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well
E_revulsion	Self reported feeling of how well REVULSION describes respondent's emotional reaction to disease: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well
E_hateful	Self reported feeling of how well HATEFUL describes respondent's emotional reaction to disease: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well
E_angry	Self reported feeling of how well ANGRY describes respondent's emotional reaction to disease: 1-Not Well At All, 2- Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well
E_anxiety	Self reported feeling of how well ANXIETY describes respondent's emotional reaction to disease: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well
E_nervous	Self reported feeling of how well NERVOUS describes respondent's emotional reaction to disease: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well
E_worry	Self reported feeling of how well WORRIED describes respondent's emotional reaction to disease: 1-Not Well At All, 2-Not Too Well, 3-Somewhat Well, 4-Very Well, 5-Extremely Well

```
Study2_preprocess <- read_dta("Study2ReplicationData.dta")

Study2 <- Study2_preprocess %>%
  #grouping similar treatment conditions and computing time spent viewing treatment
  mutate(treatment = case_when(c_control==1 ~ 1,
                               c_bothd==1|c_bothm==1 ~ 4,
                               c_disguste==1|c_disgustl==1 ~ 2,
                               c_mapl==1|c_mape==1 ~3),
         page_time = rowMeans(select(., contains("t_c")), na.rm = TRUE)) %>%
  #recoding NAs as 0 for country and symptom variables
  mutate(across(contains(c("countries","symptoms")), ~ifelse(is.na(.x),0,.x))) %>%
  #recoding infosession as binary
  mutate(across(c(infosession,learn), ~ifelse(.x==1,1,0))) %>%
  #renaming country variables
  rename_with(.cols= contains("countries"), ~c("Mexico",
                                               "SouthAmerica",
                                               "Africa",
                                               "Canada",
                                               "Russia",
```

```
                                                "Europe")) %>%
  #renaming symptom variables
  rename_with(.cols = contains("ksymptoms"), ~c("fever",
                                                "headache",
                                                "jointpain",
                                                "rash",
                                                "bleeding",
                                                "nausea",
                                                "seizures",
                                                "breathing")) %>%
  #renaming info search variables and other disease knowledge info
  rename_with(.cols = contains("search"), ~c("infosearch",
                                             "talk")) %>%
  rename(symptpercent = kpercent, length = klength) %>%
  #renaming emotion variables
  rename_with(.cols = contains("emotion"), ~paste0("E_", c("disgust",
                                                           "gross",
                                                           "resentment",
                                                           "revulsion",
                                                           "hateful",
                                                           "angry",
                                                           "anxiety",
                                                           "nervous",
                                                           "worry"))) %>%
  #creating indicator variable for correct identification of symptoms and at risk countries
  mutate(country_correct = as.numeric(paste0(Mexico, SouthAmerica, Africa, Canada, Russia, Europe))==11
         symptoms_correct = as.numeric(paste0(fever, headache, jointpain, rash, bleeding, nausea, seizu
  #deleting irrelevant variables
  select(-contains("c_")) %>%
  #reordering treatment as first variable
  relocate(treatment)
```

##Question 5 Above is an example of the kind of data cleaning that sometimes must be done to make the datasets intuitive and easy to work with. **Look through the code above and try to follow what each line is doing. Look at the Study1_preprocess and Study1 datasets and note the differences between them. You do not need to write anything for this question, just get a sense of some of the useful tools when cleaning data!**

##Question 6 ###Part a What are the treatment conditions for Study 2? **What are the treatment conditions for Study 2? How do they differ from Study 1? Hint: Look at page 273 for the treatment conditions of Study 2**

###Part b Study 2 asks about three different categories of emotions: disgust (disgusted, grossed out, revulsion), anxiety (anxious, nervous, worried), and anger (angry, hateful, resentful). Did the treatments succeed in manipulating emotions? Was the impact limited to disgust? **Pool respondents into low disgust and high disgust treatments. Check for statistically significant differences in the average answer to each emotion index.**

###Part c Plot the distribution of disgust, anxiety, and anger for low disgust and high disgust treatment conditions. **Pool respondents into low disgust and high disgust treatments. Plot the distributions for the average item response to each of the three emotional categories.**

##Question 7 ###Part a The researchers tested whether the inclusion of any kind of imagery (the map) would affect informational recall. Does the map impact information recall about the info it shows (affected countries)? What about other information? **Compare the accuracy of affected country recall in map**

and non-map treatment conditions. Then choose one other measure information recall and compare it across map and non-map treatments.

###Part b Choose one of the three hypotheses identified in Question 2. Compare the responses on one or more of the outcomes relevant to that hypothesis using data from Study 2. Is the hypothesis supported by the data? **For one or more outcome variables relevant to one of the three hypotheses check for statistically significant differences. What can we conclude from the data?**

###Part c One argument the paper acknowledges is that those in the disgust treatment may simply be clicking through the treatment quickly to avoid the imagery and this is affecting recall. **Compare mean page_view_length for each group. Do the disgusting images cause people to spend less time viewing the page?**

##Question 8 ###Part a Compare your intepretations of the results of both studies to the paper's interpretation. **What is your interpretation of the findings across both studies? Do your takeaways match or differ those of the authors?**

###Part b What other emotions may be of interest? **What other emotions may affect information uptake or political behavior? How so? How might you test these hypotheses?**

##Question 9 Run a linear regression on an outcome variable of interest (e.g. searching for more information or correctly identifying all symptoms) using any of the variables in the dataset for either Study 1 or Study 2. **Run an OLS model for any outcome variable of interest with your own specification. Carefully interpret the results. What do they tell us?**