

# Data Exploration: Emotional Arousal

Your name here

November 4, 2021

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
<code>treat_rand1</code>	Treatment assignment: 1-Low Anxiety/Low Disgust, 2-High Anxiety/Low Disgust, 3-Low Anxiety/High Disgust, and 4-High Anxiety/High Disgust
<code>Q11_1</code>	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
<code>Q11_2</code>	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
<code>Q11_3</code>	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
<code>Q11_4</code>	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
<code>Q11_5</code>	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

Variable Name	Variable Description
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
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")
```

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", "requested_more"))
```

###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("female", "race", "education", "partisanship",
      "voter_registration", "ideology"))
```

###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", "gross", "repulsed",
    "afraid", "anxious", "worried"))
```

###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 = everything(), ~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_7, "cure", "requested_more",
                          "voter_registration"), ~ifelse(. == 1, 1, 0)))
```

###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.**

*# Create answer key with correct symptom identifications for each treatment*

```
answers <- tibble(labels = c("low anxiety/low disgust",
                           "high anxiety/low disgust",
                           "low anxiety/high disgust",
                           "high anxiety/high disgust"),
  treat_rand1 = seq(1, 4, 1),
  Q12_1 = 1,
  Q12_2 = c(1, 1, 0, 0), # Headaches
  Q12_3 = c(0, 0, 1, 1), # Diarrhea
  Q12_4 = c(1, 1, 0, 0), # Joint Pain
  Q12_5 = c(0, 0, 1, 1), # Boils
  Q12_6 = 0, # Fever
  Q12_7 = 0) # Warts
```

```
answers
```

```
## # A tibble: 4 x 9
```

```
##   labels                                treat_rand1 Q12_1 Q12_2 Q12_3 Q12_4 Q12_5 Q12_6 Q12_7
##   <chr>                                <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 low anxiety/low disgust                1      1      1      0      1      0      0      0
## 2 high anxiety/low disgust               2      1      1      0      1      0      0      0
## 3 low anxiety/high disgust              3      1      0      1      0      1      0      0
## 4 high anxiety/high disgust             4      1      0      1      0      1      0      0
```

*# Check that respondents' answers match correct answers for each treatment group*

```
Study1_processing7 <- Study1_processing6 %>%
  mutate(symptom_check = ifelse(Q12_1 == answers$Q12_1[treat_rand1] &
                                Q12_2 == answers$Q12_2[treat_rand1] &
                                Q12_3 == answers$Q12_3[treat_rand1] &
                                Q12_4 == answers$Q12_4[treat_rand1] &
                                Q12_5 == answers$Q12_5[treat_rand1] &
                                Q12_6 == answers$Q12_6[treat_rand1] &
                                Q12_7 == answers$Q12_7[treat_rand1], 1, 0))
```

####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.**

```
data_clean <- Study1_processing7 %>%
  rename_with(.cols = contains("Q12"), ~c("symptoms_fatigue", "symptoms_headaches",
                                           "symptoms_diarrhea", "symptoms_joint",
                                           "symptoms_boils", "symptoms_warts",
                                           "symptoms_fever"))
```

## 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**

Study 1 was conducted using a national survey of 1,000 adults administered through YouGov over the course of 1 week in 2015. Respondents read an excerpt about a fictitious disease (Tugela River virus, or TRV), and were asked questions gauging their reactions and whether they would seek additional information.

Treatments varied by low to high disgust and low to high anxiety conditions. In the low disgust condition, the excerpt said TRV symptoms included headaches and joint pain. In the high disgust condition, symptoms were boils and diarrhea. In the low anxiety condition, TRV was said to have low transmissibility and patients made full recoveries. In the high anxiety condition, TRV said to have higher transmissibility (infected people could spread the disease before showing symptoms) and infected people might have already taken flights to the US.

Treatment conditions were manipulated in a 2x2 design, so the text respondents read were low anxiety/low disgust, high anxiety/low disgust, low anxiety/high disgust, and high anxiety/high disgust.

The symptoms presented in each condition were as follows:

```

answers %>%
  select(-treat_rand1) %>%
  rename_with(.cols = contains("Q12"), ~c("symptoms_fatigue", "symptoms_headaches",
                                           "symptoms_diarrhea", "symptoms_joint",
                                           "symptoms_boils", "symptoms_warts",
                                           "symptoms_fever"))

```

```

## # A tibble: 4 x 8
##   labels symptoms_fatigue symptoms_headac~ symptoms_diarrh~ symptoms_joint
##   <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 low a~          1            1            0            1
## 2 high ~          1            1            0            1
## 3 low a~          1            0            1            0
## 4 high ~          1            0            1            0
## # ... with 3 more variables: symptoms_boils <dbl>, symptoms_warts <dbl>,
## #   symptoms_fever <dbl>

```

###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**

Disgust is part of the behavioral immune system, and it helps us avoid exposure to pathogens. Things that elicit disgust (bodily fluids, spoiled food, rodents, animals associated with disease, etc.) can trigger the behavioral response of avoidance, which helps humans stay healthy. Based on the existing literature, the authors describe 3 hypotheses on how disgust impacts learning and information uptake:

1. Emotional arousal (ex: disgust) leads to heightened attention to the stimulus (ie the disgusting thing), which leads to greater memory of the stimulus, and worse memory of peripheral objects. The experiment measures memory by asking respondents what the corresponding symptoms were for the anxiety/disgust condition they were presented (Q12 variables).
2. Anticipation of disgust causes people to avoid disgusting situations (ex: when people anticipate medical treatment will be disgusting, they delay seeking treatment). The experiment also measures avoidance through the symptoms questions (Q12; if respondents anticipate the text to be disgusting, they might read it with less attention and remember the symptoms incorrectly) and whether the respondent would seek more information (Q15).
3. Disgust could lead to curiosity if the respondent perceives themselves to be safe from the threat of the disgusting stimuli. Woody and Teachman 2006 find that people enjoy disgust as long as it is “framed as unreal.” The experiment also measures intrigue through the symptoms questions (Q12; if the respondents are curious, they might read the symptoms more carefully and remember the symptoms better) and whether the respondent would seek more information (Q15)

###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?**

```
data_clean %>%
  drop_na() %>%
  group_by(treat_rand1) %>%
  summarize(mean_disgust = mean(disgust), mean_gross = mean(gross),
             mean_repulsed = mean(repulsed), mean_afraid = mean(afraid),
             mean_anxious = mean(anxious), mean_worried = mean(worried),
             .groups = "drop") %>%
  mutate(condition = c("low anxiety/low disgust", "high anxiety/low disgust",
                       "low anxiety/high disgust", "high anxiety/high disgust")) %>%
  select(condition, treat_rand1, mean_disgust, mean_anxious)
```

```
## # A tibble: 4 x 4
##   condition                treat_rand1 mean_disgust mean_anxious
##   <chr>                  <dbl>+<lbl>      <dbl>      <dbl>
## 1 low anxiety/low disgust 1 [Low Anxiety and Low Disg~      2.13      2.14
## 2 high anxiety/low disgu~ 2 [High Anxiety and Low Dis~      2.17      2.36
## 3 low anxiety/high disgu~ 3 [Low Anxiety and High Dis~      2.49      2.17
## 4 high anxiety/high disg~ 4 [High Anxiety and High Di~      2.51      2.18
```

```
# Manipulating anxiety -- differences in disgust not significant
```

```
t.test(data_clean$disgust[data_clean$treat_rand1 == 1], data_clean$disgust[data_clean$treat_rand1 == 2])
```

```
##
## Welch Two Sample t-test
##
## data: data_clean$disgust[data_clean$treat_rand1 == 1] and data_clean$disgust[data_clean$treat_rand1 == 2]
## t = -0.19797, df = 471.39, p-value = 0.8432
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.2427904 0.1983459
## sample estimates:
## mean of x mean of y
## 2.177778 2.200000
```

```
t.test(data_clean$disgust[data_clean$treat_rand1 == 3], data_clean$disgust[data_clean$treat_rand1 == 4])
```

```
##
## Welch Two Sample t-test
##
## data: data_clean$disgust[data_clean$treat_rand1 == 3] and data_clean$disgust[data_clean$treat_rand1 == 4]
## t = -0.18092, df = 503.99, p-value = 0.8565
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.2403215 0.1997934
## sample estimates:
## mean of x mean of y
## 2.521569 2.541833
```

```
# Manipulating disgust -- differences in disgust are significant
```

```
t.test(data_clean$disgust[data_clean$treat_rand1 == 1], data_clean$disgust[data_clean$treat_rand1 == 3])
```

```
##
## Welch Two Sample t-test
##
## data: data_clean$disgust[data_clean$treat_rand1 == 1] and data_clean$disgust[data_clean$treat_rand1 == 2]
## t = -2.986, df = 471.93, p-value = 0.002973
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.5700275 -0.1175542
## sample estimates:
## mean of x mean of y
## 2.177778 2.521569
```

```
t.test(data_clean$disgust[data_clean$treat_rand1 == 2], data_clean$disgust[data_clean$treat_rand1 == 4])
```

```
##
## Welch Two Sample t-test
##
## data: data_clean$disgust[data_clean$treat_rand1 == 2] and data_clean$disgust[data_clean$treat_rand1 == 4]
## t = -3.135, df = 510.82, p-value = 0.001818
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.5560525 -0.1276129
## sample estimates:
## mean of x mean of y
## 2.200000 2.541833
```

```
# Manipulating anxiety -- differences in anxiety not significant
```

```
t.test(data_clean$anxious[data_clean$treat_rand1 == 1], data_clean$anxious[data_clean$treat_rand1 == 2])
```

```
##
## Welch Two Sample t-test
##
## data: data_clean$anxious[data_clean$treat_rand1 == 1] and data_clean$anxious[data_clean$treat_rand1 == 2]
## t = -1.8007, df = 468.68, p-value = 0.0724
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.39560480 0.01726922
## sample estimates:
## mean of x mean of y
## 2.191964 2.381132
```

```
t.test(data_clean$anxious[data_clean$treat_rand1 == 3], data_clean$anxious[data_clean$treat_rand1 == 4])
```

```
##
## Welch Two Sample t-test
##
## data: data_clean$anxious[data_clean$treat_rand1 == 3] and data_clean$anxious[data_clean$treat_rand1 == 4]
## t = -0.18496, df = 500.47, p-value = 0.8533
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.2103695 0.1741679
## sample estimates:
```



```
## mean of x mean of y
## 2.181102 2.199203
```

```
# Manipulating disgust -- differences in anxiety not significant
```

```
t.test(data_clean$anxious[data_clean$treat_rand1 == 1], data_clean$anxious[data_clean$treat_rand1 == 3])
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: data_clean$anxious[data_clean$treat_rand1 == 1] and data_clean$anxious[data_clean$treat_rand1 == 3]
```

```
## t = 0.10541, df = 453.98, p-value = 0.9161
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -0.1916333 0.2133571
```

```
## sample estimates:
```

```
## mean of x mean of y
```

```
## 2.191964 2.181102
```

```
t.test(data_clean$anxious[data_clean$treat_rand1 == 2], data_clean$anxious[data_clean$treat_rand1 == 4])
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: data_clean$anxious[data_clean$treat_rand1 == 2] and data_clean$anxious[data_clean$treat_rand1 == 4]
```

```
## t = 1.8197, df = 512.83, p-value = 0.06939
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -0.0144916 0.3783494
```

```
## sample estimates:
```

```
## mean of x mean of y
```

```
## 2.381132 2.199203
```

The only manipulation that had a 95% statistically significant difference between means was changing the disgust level for both high and low anxiety conditions – manipulating the passage from low to high disgust had a significantly significant impact on how many respondents reported feeling disgust after reading.

###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?**

1. Emotional arousal (ex: disgust) leads to heightened attention to the stimulus (ie the disgusting thing), which leads to greater memory of the stimulus, and worse memory of peripheral objects. The experiment measures memory by asking respondents what the corresponding symptoms were for the anxiety/disgust condition they were presented (Q12 variables).

According to this hypothesis, respondents who report high levels of disgust or anxiety should remember the symptoms better than those who report lower levels of disgust or anxiety.

```
answers_clean <- answers %>%
```

```
select(-labels) %>%
```

```
rename_with(.cols = contains("Q12"), ~c("ans_symptoms_fatigue", "ans_symptoms_headaches",
```

```

      "ans_symptoms_diarrhea", "ans_symptoms_joint",
      "ans_symptoms_boils", "ans_symptoms_warts",
      "ans_symptoms_fever"))

data_h1 <- data_clean %>%
  left_join(answers_clean, by = "treat_rand1") %>%

  # Check which symptoms were identified against the correct symptoms per the
  # treatment. NOTE: The actual study only checks if the 2 symptoms were
  # recalled, not if the respondent indicated more symptoms than were listed.

  mutate(check_fatigue = ifelse(symptoms_fatigue == ans_symptoms_fatigue, 1, 0),
         check_headaches = ifelse(symptoms_headaches == ans_symptoms_headaches, 1, 0),
         check_diarrhea = ifelse(symptoms_diarrhea == ans_symptoms_diarrhea, 1, 0),
         check_joint = ifelse(symptoms_joint == ans_symptoms_joint, 1, 0),
         check_boils = ifelse(symptoms_boils == ans_symptoms_boils, 1, 0),
         check_warts = ifelse(symptoms_warts == ans_symptoms_warts, 1, 0),
         check_fever = ifelse(symptoms_fever == ans_symptoms_fever, 1, 0),
         sum_correct = rowSums(across(check_fatigue:check_fever))) %>%

  # Add disgust indicator variable (low for treatments 1-2, high for treatments
  # 3-4). NOTE: check_symptoms is coded so the analysis matches the study --
  # check_symptoms is 1 if respondent identifies both symptoms present in the
  # condition, 0 if they don't.

  mutate(disgust_condition = ifelse(treat_rand1 < 3, "low", "high"),
         check_symptoms = ifelse(treat_rand1 < 3,
                                ifelse(symptoms_headaches == ans_symptoms_headaches
                                      & symptoms_joint == ans_symptoms_joint, 1, 0),
                                ifelse(symptoms_diarrhea == ans_symptoms_diarrhea
                                      & symptoms_boils == ans_symptoms_boils, 1, 0))) %>%

  # Add mean emotional reaction based on disgust and anxiety indecies

  mutate(mean_disgust = (disgust + gross + repulsed) / 3,
         mean_anxious = (anxious + afraid + worried) / 3)

  # # select(-contains(c("symptom", "check")))) %>%
  # mutate(across(treat_rand1:sum_correct, as.numeric))

# Check that the differences in recall between high and low disgust are
# significant

data_h1 %>%
  group_by(disgust_condition) %>%
  summarize(count = n(),
           check_symptoms = sum(check_symptoms) / n(),
           sum_correct = sum(sum_correct) / (n() * 7))

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

## # A tibble: 2 x 4

```

```
##   disgust_condition count check_symptoms sum_correct
##   <chr>             <int>         <dbl>         <dbl>
## 1 high              509           0.703         0.800
## 2 low               491           0.607         0.801

# Difference between disgust conditions is significant for check_symptoms but
# not sum_correct

t.test(data_h1$check_symptoms[data_h1$disgust_condition == "low"], data_h1$check_symptoms[data_h1$disgust_condition == "high"], var.equal = FALSE)

##
## Welch Two Sample t-test
##
## data: data_h1$check_symptoms[data_h1$disgust_condition == "low"] and data_h1$check_symptoms[data_h1$disgust_condition == "high"]
## t = -3.2181, df = 987.55, p-value = 0.001332
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.15520795 -0.03762253
## sample estimates:
## mean of x mean of y
## 0.6069246 0.7033399

t.test(data_h1$sum_correct[data_h1$disgust_condition == "low"], data_h1$sum_correct[data_h1$disgust_condition == "high"], var.equal = FALSE)

##
## Welch Two Sample t-test
##
## data: data_h1$sum_correct[data_h1$disgust_condition == "low"] and data_h1$sum_correct[data_h1$disgust_condition == "high"]
## t = 0.065407, df = 993.98, p-value = 0.9479
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1645526 0.1759003
## sample estimates:
## mean of x mean of y
## 5.604888 5.599214

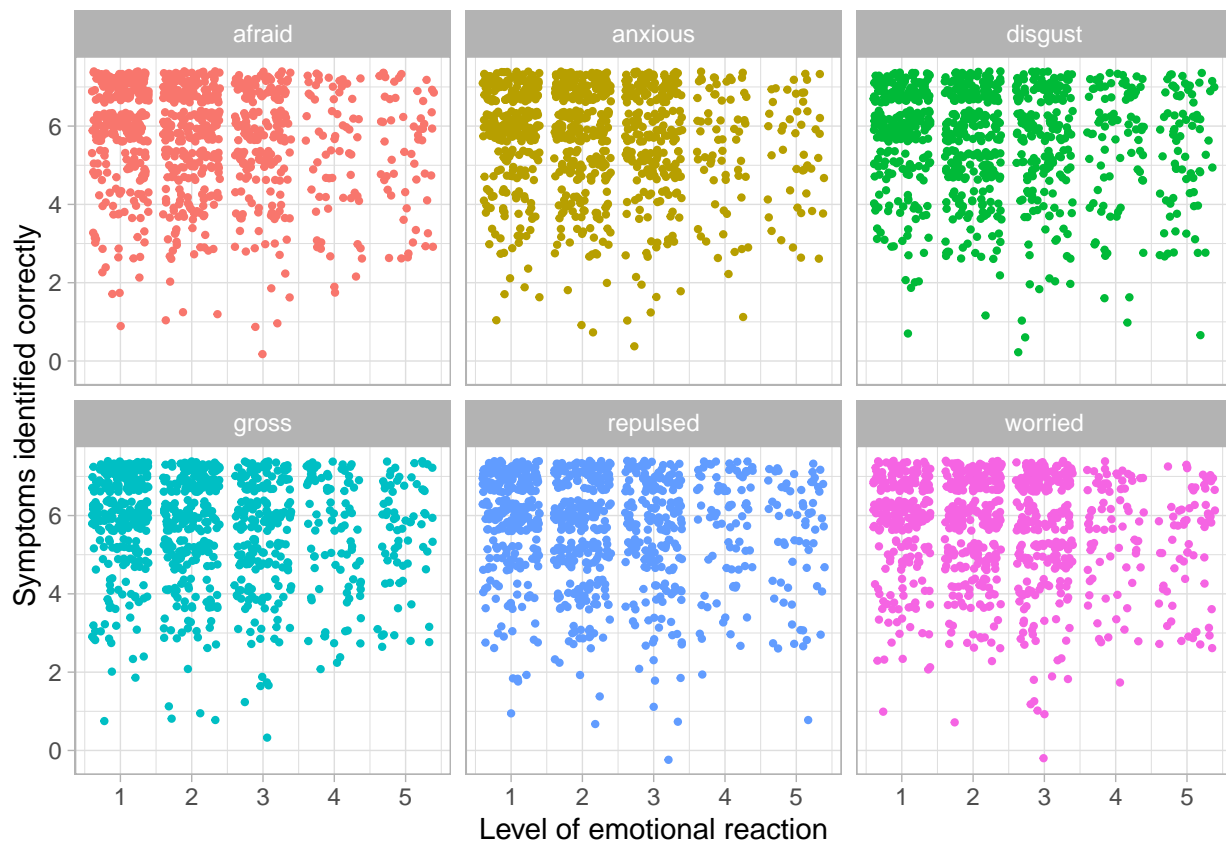
data_h1 %>%
  select(disgust_condition, mean_disgust, mean_anxious, check_symptoms, sum_correct) %>%
  group_by(disgust_condition)

## # A tibble: 1,000 x 5
## # Groups:   disgust_condition [2]
##   disgust_condition mean_disgust mean_anxious check_symptoms sum_correct
##   <chr>             <dbl>         <dbl>         <dbl>         <dbl>
## 1 low              2           2           0           4
## 2 high             4           3.33         1           6
## 3 low             1.33         1           1           7
## 4 low             3.33         3.33         1           5
## 5 low              1           1.67         1           7
## 6 high             2           3           1           5
## 7 low             1.33         2           1           7
## 8 high             1           1.33         1           6
```

```
## 9 low          3.33      2          1          7
## 10 low         5         4          1          6
## # ... with 990 more rows
```

```
data_h1 %>%
  drop_na() %>%
  pivot_longer(cols = disgust:worried, names_to = "type", values_to = "value") %>%
  ggplot(aes(x = value, y = sum_correct, color = type)) +
    geom_jitter(size = 0.8) +
    facet_wrap("type") +
    theme_light() +
    theme(legend.position = "none") +
    labs(x = "Level of emotional reaction", y = "Symptoms identified correctly")
```

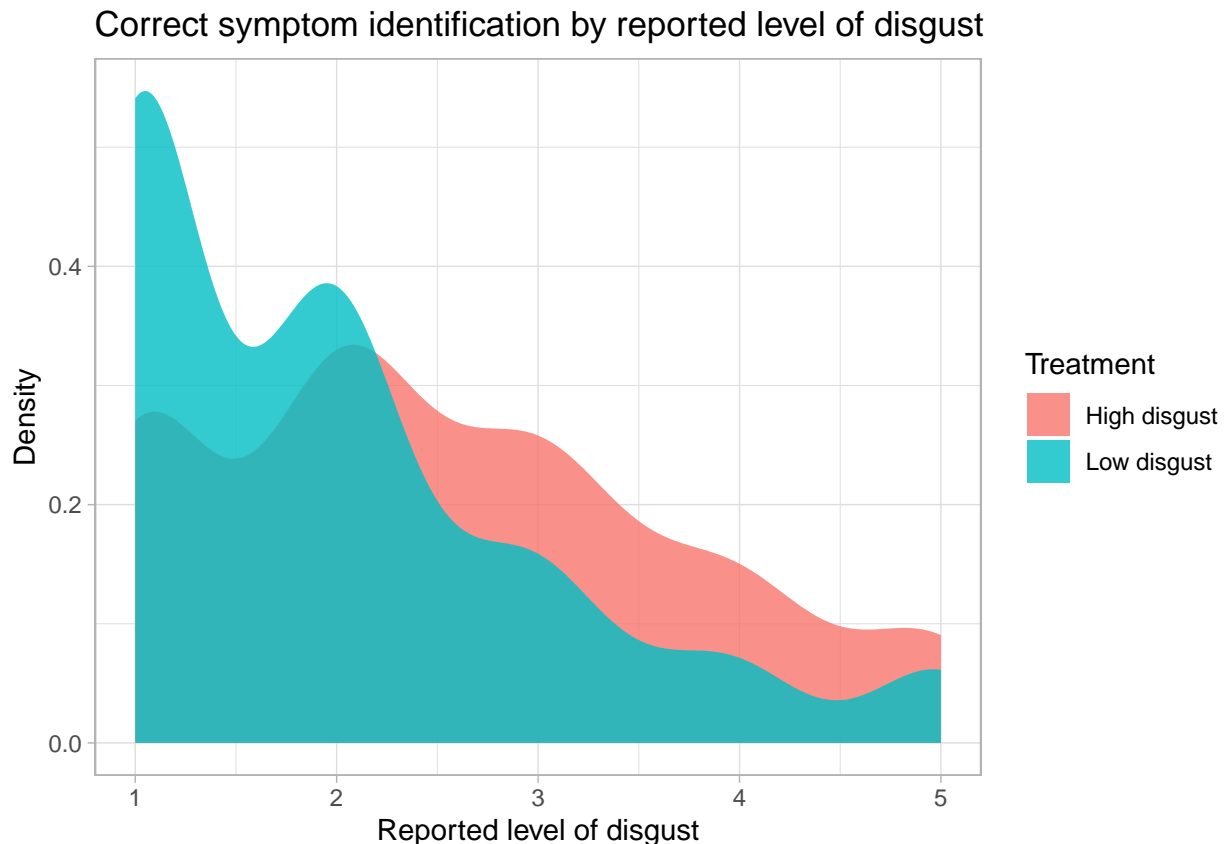
## Don't know how to automatically pick scale for object of type haven\_labelled/vctrs\_vctr/double. Defa



```
data_h1 %>%
  select(disgust_condition, mean_disgust, mean_anxious, check_symptoms, sum_correct) %>%
  rename(disgust = mean_disgust, anxious = mean_anxious) %>%
  pivot_longer(cols = c(disgust, anxious), names_to = "index", values_to = "values") %>%
  filter(check_symptoms == 1, index == "disgust") %>%
  ggplot(aes(x = values, fill = disgust_condition)) +
    geom_density(color = NA, alpha = 0.8) +
    theme_light() +
    labs(x = "Reported level of disgust", y = "Density",
```

```
title = "Correct symptom identification by reported level of disgust") +
scale_fill_discrete(name = "Treatment",
  labels = c("high" = "High disgust", "low" = "Low disgust"))
```

## Warning: Removed 4 rows containing non-finite values (stat\_density).



##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.**

Disgust: disgust, gross, repulsed Anxious: afraid, anxious, worried

```
data_clean4 <- data_clean %>%
  drop_na() %>%
  mutate(across(treat_rand1:symptom_check, as.numeric))

# Alpha for disgust index
```

```

sum_cov <- sum(cov(data_clean4$disgust, data_clean4$gross), cov(data_clean4$disgust, data_clean4$repulsed))
sum_var <- sum(var(data_clean4$disgust), var(data_clean4$gross), var(data_clean4$repulsed))
alpha_disgust <- (3 * sum_cov)/(sum_var + (2 * sum_cov))

# Alpha for anxious index

sum_cov <- sum(cov(data_clean4$afraid, data_clean4$anxious), cov(data_clean4$afraid, data_clean4$worried))
sum_var <- sum(var(data_clean4$afraid), var(data_clean4$anxious), var(data_clean4$worried))
alpha_anxious <- (3 * sum_cov)/(sum_var + (2 * sum_cov))

# Output

alpha_disgust

```

```
## [1] 0.9333667
```

```
alpha_anxious
```

```
## [1] 0.9167872
```

#### 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?**

```

data_clean4 %>%
  select(disgust:repulsed) %>%
  cronbach.alpha()

```

```

##
## Cronbach's alpha for the '.' data-set
##
## Items: 3
## Sample units: 905
## alpha: 0.933

```

```

data_clean4 %>%
  select(afraid:worried) %>%
  cronbach.alpha()

```

```

##
## Cronbach's alpha for the '.' data-set
##
## Items: 3
## Sample units: 905
## alpha: 0.917

```

The alphas calculated using the ltm package were the same as the alphas calculated manually – both are sufficiently high (alpha > 0.7)

###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.**

```
data_4c <- data_h1 %>%
  drop_na() %>%
  # mutate(mean_disgust = (disgust + gross + repulsed) / 3,
  #        mean_anxious = (anxious + afraid + worried) / 3) %>%
  select(treat_rand1, requested_more, sum_correct, check_symptoms, mean_disgust, mean_anxious, page_art,
         rename(disgust = mean_disgust, anxious = mean_anxious)
```

```
data_4c
```

```
## # A tibble: 905 x 29
##   treat_rand1 requested_more sum_correct check_symptoms disgust anxious
##   <dbl+lbl>      <dbl>      <dbl>      <dbl>      <dbl>  <dbl>
## 1 1 [Low Anx~      0          4          0          2          2
## 2 4 [High An~      1          6          1          4         3.33
## 3 1 [Low Anx~      0          7          1         1.33          1
## 4 2 [High An~      0          5          1         3.33         3.33
## 5 1 [Low Anx~      1          7          1          1         1.67
## 6 3 [Low Anx~      0          5          1          2          3
## 7 1 [Low Anx~      0          7          1         1.33          2
## 8 4 [High An~      0          6          1          1         1.33
## 9 2 [High An~      0          7          1         3.33          2
##10 1 [Low Anx~      0          6          1          5          4
## # ... with 895 more rows, and 23 more variables: page_article_timing <dbl>,
## #   female <dbl>, race <dbl+lbl>, education <dbl+lbl>, partisanship <dbl+lbl>,
## #   voter_registration <dbl>, ideology <dbl+lbl>, symptom_check <dbl>,
## #   ans_symptoms_fatigue <dbl>, ans_symptoms_headaches <dbl>,
## #   ans_symptoms_diarrhea <dbl>, ans_symptoms_joint <dbl>,
## #   ans_symptoms_boils <dbl>, ans_symptoms_warts <dbl>,
## #   ans_symptoms_fever <dbl>, check_fatigue <dbl>, check_headaches <dbl>,
## #   check_diarrhea <dbl>, check_joint <dbl>, check_boils <dbl>,
## #   check_warts <dbl>, check_fever <dbl>, disgust_condition <chr>
```

```
m1 <- lm(requested_more ~ disgust + anxious + check_symptoms + female + ideology, data = data_4c)
```

```
# Read time and correct answers are not statistically significant
```

```
m2 <- lm(check_symptoms ~ page_article_timing, data = data_4c)
```

```
m3 <- lm(check_symptoms ~ disgust + anxious, data = data_4c)
```

```
summary(m1)
```

```
##
```

```
## Call:
```

```
## lm(formula = requested_more ~ disgust + anxious + check_symptoms +
##     female + ideology, data = data_4c)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6522 -0.2888 -0.1948  0.4865  0.9428
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.148103   0.056962   2.600  0.00947 **
## disgust        -0.019080   0.016293  -1.171  0.24187
## anxious         0.121908   0.017730   6.876 1.15e-11 ***
## check_symptoms -0.008999   0.031300  -0.287  0.77380
## female          0.024855   0.029049   0.856  0.39244
## ideology       -0.034939   0.010585  -3.301  0.00100 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.431 on 899 degrees of freedom
## Multiple R-squared:  0.07468,    Adjusted R-squared:  0.06953
## F-statistic: 14.51 on 5 and 899 DF,  p-value: 1.079e-13
```

I included the disgust and anxious variables to test the hypotheses the authors presented, but I also added check\_symptoms, female, and ideology. I included check\_symptoms because I thought respondents who recalled less of the correct symptoms might request information to clarify potential misunderstandings. I also included gender and ideology thinking differences in socialization (socialization due to gender and socialization leading to ideology) might impact someone's desire to seek more information.

The model shows that anxiety and ideology are significantly correlated with requesting more information on TRV — anxiety has a coefficient of 0.12 while ideology has a coefficient of -0.035. Higher feelings of anxiety are correlated with more information search, and liberals are more likely to seek additional information.

The results are consistent with Clifford and Jerri, who found that increased anxiety was correlated with more information search.

###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.**

```
confint(m1)
```

```
##              2.5 %      97.5 %
## (Intercept)  0.03630977 0.25989676
## disgust     -0.05105688 0.01289594
## anxious      0.08711170 0.15670384
## check_symptoms -0.07042901 0.05243147
## female       -0.03215702 0.08186690
## ideology     -0.05571382 -0.01416382
```

```
coef_plot <- as_tibble(confint(m1)) %>%
  mutate(variable = c("constant", "disgust", "anxious", "check_symptoms", "female", "ideology"),
         coef = c(0.148, -0.019, 0.122, -0.009, 0.025, -0.035)) %>%
  rename("lower" = '2.5 %',
         "upper" = '97.5 %')
```

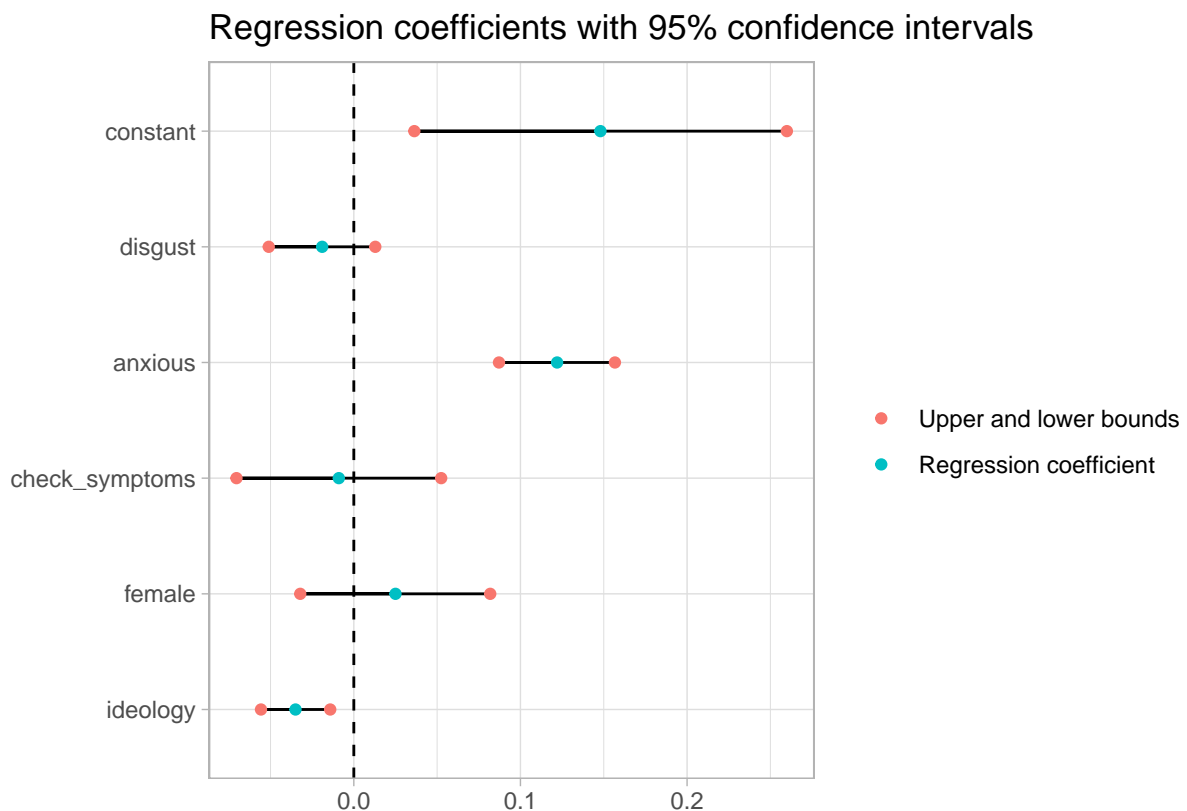


```
coef_plot
```

```
## # A tibble: 6 x 4
##   lower upper variable      coef
##   <dbl> <dbl> <chr>      <dbl>
## 1  0.0363 0.260 constant    0.148
## 2 -0.0511 0.0129 disgust    -0.019
## 3  0.0871 0.157 anxious     0.122
## 4 -0.0704 0.0524 check_symptoms -0.009
## 5 -0.0322 0.0819 female      0.025
## 6 -0.0557 -0.0142 ideology    -0.035
```

```
level_order <- rev(c("constant", "disgust", "anxious", "check_symptoms", "female", "ideology"))
```

```
coef_plot %>%
  pivot_longer(cols = c(upper, lower, coef), names_to = "type", values_to = "values") %>%
  ggplot(aes(x = factor(variable, level = level_order), y = values)) +
  geom_line() +
  geom_point(aes(color = ifelse(type == "coef", "red", "blue"))) +
  geom_hline(yintercept = 0, lty = "dashed") +
  coord_flip() +
  theme_light() +
  labs(x = "", y = "", title = "Regression coefficients with 95% confidence intervals") +
  scale_color_discrete(name = "", labels = c("red" = "Regression coefficient",
                                             "blue" = "Upper and lower bounds"))
```



## 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
symppercent	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
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

Variable Name	Variable Description
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_disgust==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 infoession 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))==1),
        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 interpretations 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?**