Additional links:

- Link to Rmd file
  - https://github.com/madisonthantu/Behavioral-RP-recruiting-Fall-2023/blob/main/data_analysis_task.html
- Link to html file
  - https://github.com/madisonthantu/Behavioral-RP-recruiting-Fall-2023/blob/main/data_analysis_task.html

Data Analysis Task 1

Clean the data. Consider what variables you might use to exclude observations.

- Variables used for exclusion: If variable x != y, then exclude sample
  - Progress = 100%
  - Finished = True
  - consent = Agree
  - QID54_First Click, QID54_Last Click, QID54_Page Submit, QID54_Click Count != nan
  - passedattn = 'yes'

```
library(readxl)
# REF: https://stackoverflow.com/questions/2631780/set-the-plot-title-to-wrap-around-and-shrink-the-tex
wrapper <- function(x, ...)
{
  paste(strwrap(x, ...), collapse = "\n")
}
# Read in data
# REF: https://stackoverflow.com/questions/47066336/trying-to-knit-but-cant-because-read-excel-function
data <- read_excel("Data - Fall 2023.xlsx", sheet = "Data")
df <- data.frame(data)
# unique(df[["UserLanguage"]])

# REF: https://chryswoods.com/beginning_r/dictionaries.html
col_val_constraints <- c("Progress"=100, "Finished"=TRUE, "consent"="AGREE", "passedattn"="yes")

# colnames(df)
# Removing samples based on exclusion criteria
df_cleaned <- subset(x = df,
    subset = Progress == 100 &
             Finished == TRUE &
             consent == "AGREE" &
             passedattn == "yes" &
             !is.na(QID54_First.Click) &
             !is.na(QID54_Last.Click) &
             !is.na(QID54_Page.Submit) &
             !is.na(QID54_Click.Count))

# unique(df[["UserLanguage"]])
summary(df_cleaned)
```

```
##    StartDate                       EndDate
##  Min.   :2021-01-27 12:12:35.00   Min.   :2021-01-27 12:16:04.00
##  1st Qu.:2021-01-27 12:12:42.00   1st Qu.:2021-01-27 12:17:22.75
##  Median :2021-01-27 12:12:46.00   Median :2021-01-27 12:18:14.00
##  Mean   :2021-01-27 12:12:54.20   Mean   :2021-01-27 12:18:28.82
```

```
##  3rd Qu.:2021-01-27 12:12:52.75   3rd Qu.:2021-01-27 12:19:23.00
##  Max.   :2021-01-27 12:15:09.00   Max.   :2021-01-27 12:22:03.00
##
##     Status            IPAddress            Progress   Duration..in.seconds.
##  Length:46          Length:46          Min.   :100   Min.   :187.0
##  Class :character   Class :character   1st Qu.:100   1st Qu.:271.5
##  Mode  :character   Mode  :character   Median :100   Median :318.5
##                                        Mean   :100   Mean   :334.0
##                                        3rd Qu.:100   3rd Qu.:395.0
##                                        Max.   :100   Max.   :505.0
##
##  Finished        RecordedDate                      ResponseId
##  Mode:logical   Min.   :2021-01-27 12:16:04.00   Length:46
##  TRUE:46        1st Qu.:2021-01-27 12:17:22.75   Class :character
##                 Median :2021-01-27 12:18:15.00   Mode  :character
##                 Mean   :2021-01-27 12:18:29.39
##                 3rd Qu.:2021-01-27 12:19:24.75
##                 Max.   :2021-01-27 12:22:04.00
##
##  RecipientLastName RecipientFirstName RecipientEmail ExternalReference
##  Mode:logical      Mode:logical       Mode:logical   Mode:logical
##  NA's:46           NA's:46            NA's:46        NA's:46
##
##
##
##
##
##  LocationLatitude LocationLongitude DistributionChannel UserLanguage
##  Min.   :23.75    Min.   :-121.95   Length:46           Length:46
##  1st Qu.:41.80    1st Qu.: -87.67   Class :character    Class :character
##  Median :41.89    Median : -87.62   Mode  :character    Mode  :character
##  Mean   :41.08    Mean   : -83.43
##  3rd Qu.:41.92    3rd Qu.: -87.60
##  Max.   :50.72    Max.   :  90.38
##
##    consent          Q26_Browser        Q26_Version        Q26_Operating.System
##  Length:46          Length:46          Length:46          Length:46
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Q26_Resolution     QID54_First.Click QID54_Last.Click QID54_Page.Submit
##  Length:46          Min.   :  0.149   Min.   : 12.72   Min.   : 13.60
##  Class :character   1st Qu.: 20.130   1st Qu.: 29.07   1st Qu.: 34.43
##  Mode  :character   Median : 32.813   Median : 37.73   Median : 39.55
##                     Mean   : 35.799   Mean   : 42.04   Mean   : 44.98
##                     3rd Qu.: 46.946   3rd Qu.: 49.28   3rd Qu.: 51.81
##                     Max.   :101.926   Max.   :101.93   Max.   :103.91
##
##  QID54_Click.Count real_imaginary     initials_box         describe
##  Min.   : 1.000    Length:46          Length:46          Length:46
##  1st Qu.: 1.000    Class :character   Class :character   Class :character
```

```
##  Median : 1.000    Mode  :character   Mode  :character   Mode  :character
##  Mean   : 2.087
##  3rd Qu.: 2.000
##  Max.   :16.000
##
##  feelings_youalone feelings_bothyoufirst feelings_themalone
##  Min.   :-30.00    Min.   :-30.000       Min.   :-30.000
##  1st Qu.:-30.00    1st Qu.:  1.000       1st Qu.:-20.000
##  Median :-20.00    Median : 10.000       Median : -9.000
##  Mean   :-18.61    Mean   :  7.848       Mean   : -5.391
##  3rd Qu.:-10.00    3rd Qu.: 20.000       3rd Qu.:  9.500
##  Max.   : 10.00    Max.   : 30.000       Max.   : 30.000
##
##  feelings_boththemfirst feelings_neither feelings_youaloneforgiven
##  Min.   :-20.00         Min.   :-30.0    Min.   :-30.00
##  1st Qu.: 10.50         1st Qu.:-22.5    1st Qu.:-30.00
##  Median : 20.00         Median :-16.5    Median :-19.50
##  Mean   : 17.48         Mean   :-14.5    Mean   :-14.26
##  3rd Qu.: 25.75         3rd Qu.: -8.0    3rd Qu.: -5.00
##  Max.   : 30.00         Max.   : 20.0    Max.   : 22.00
##
##  feelings_DO_1  feelings_DO_2  feelings_DO_3   feelings_DO_4 feelings_DO_5
##  Min.   :1.00   Min.   :1.00   Min.   :1.000   Min.   :1     Min.   :1.000
##  1st Qu.:1.25   1st Qu.:1.00   1st Qu.:2.000   1st Qu.:2     1st Qu.:2.000
##  Median :3.00   Median :2.00   Median :3.000   Median :3     Median :4.000
##  Mean   :2.87   Mean   :2.63   Mean   :3.196   Mean   :3     Mean   :3.304
##  3rd Qu.:4.00   3rd Qu.:4.00   3rd Qu.:4.750   3rd Qu.:4     3rd Qu.:4.000
##  Max.   :5.00   Max.   :5.00   Max.   :5.000   Max.   :5     Max.   :5.000
##
##  feelings_DO_6 feelings_exp       outcome_binary1    outcome_binary1_DO_1
##  Min.   :6     Length:46          Length:46          Min.   :1.000
##  1st Qu.:6     Class :character   Class :character   1st Qu.:1.000
##  Median :6     Mode  :character   Mode  :character   Median :2.000
##  Mean   :6                                           Mean   :1.522
##  3rd Qu.:6                                           3rd Qu.:2.000
##  Max.   :6                                           Max.   :2.000
##
##  outcome_binary1_DO_2 outcome_binary2    outcome_binary2_DO_1
##  Min.   :1.000        Length:46          Min.   :1.000
##  1st Qu.:1.000        Class :character   1st Qu.:1.000
##  Median :1.000        Mode  :character   Median :1.000
##  Mean   :1.478                           Mean   :1.304
##  3rd Qu.:2.000                           3rd Qu.:2.000
##  Max.   :2.000                           Max.   :2.000
##
##  outcome_binary2_DO_2    blame_1       attention_1    attention_1_TEXT
##  Min.   :1.000        Min.   : 0.00   Min.   :4.000   Length:46
##  1st Qu.:1.000        1st Qu.:30.00   1st Qu.:4.500   Class :character
##  Median :2.000        Median :39.50   Median :5.000   Mode  :character
##  Mean   :1.696        Mean   :37.93   Mean   :4.667
##  3rd Qu.:2.000        3rd Qu.:50.00   3rd Qu.:5.000
##  Max.   :2.000        Max.   :75.00   Max.   :5.000
##                                       NA's   :43
##   attention_2    attention_2_TEXT    attention_3 attention_3_TEXT
```

```
##  Min.   :3.000   Length:46          Min.   :5    Length:46
##  1st Qu.:3.000   Class :character   1st Qu.:5    Class :character
##  Median :3.000   Mode  :character   Median :5    Mode  :character
##  Mean   :3.333                      Mean   :5
##  3rd Qu.:3.500                      3rd Qu.:5
##  Max.   :4.000                      Max.   :5
##  NA's   :43                         NA's   :43
##   target_sex        target_sex_3_TEXT     sex            sex_3_TEXT
##  Length:46          Mode:logical      Length:46          Mode:logical
##  Class :character   NA's:46           Class :character   NA's:46
##  Mode  :character                     Mode  :character
##
##
##
##
##       age           comments         mainControl         passedattn
##  Min.   : 18.00   Length:46          Length:46          Length:46
##  1st Qu.: 26.25   Class :character   Class :character   Class :character
##  Median : 29.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 33.63
##  3rd Qu.: 32.75
##  Max.   :149.00
##
##   lottery_draw       winner          initials         initiator_type
##  Min.   : 1.00   Min.   :0.00000   Length:46          Length:46
##  1st Qu.:10.50   1st Qu.:0.00000   Class :character   Class :character
##  Median :22.50   Median :0.00000   Mode  :character   Mode  :character
##  Mean   :23.35   Mean   :0.04348
##  3rd Qu.:37.00   3rd Qu.:0.00000
##  Max.   :47.00   Max.   :1.00000
##
##  InitiatorType.binarychoice_DO_outcome_binary2
##  Min.   :2.000
##  1st Qu.:2.000
##  Median :2.000
##  Mean   :2.435
##  3rd Qu.:3.000
##  Max.   :3.000
##
##  InitiatorType.binarychoice_DO_outcome_binary1
##  Min.   :2.000
##  1st Qu.:2.000
##  Median :3.000
##  Mean   :2.565
##  3rd Qu.:3.000
##  Max.   :3.000
##
##  InitiatorType.binarychoice_DO_Q30
##  Min.   :1
##  1st Qu.:1
##  Median :1
##  Mean   :1
##  3rd Qu.:1
##  Max.   :1
```

##

```
# For task 2, question (c)
df_c <- subset(df_cleaned, select = c(
    ResponseId,
    outcome_binary1
    ))
# Keeping only relevant columns
df_cleaned <- subset(df_cleaned, select = c(
    Progress,
    ResponseId,
    feelings_youalone,
    feelings_bothyoufirst,
    feelings_themalone,
    feelings_boththemfirst,
    feelings_neither,
    feelings_youaloneforgiven,
    initiator_type
    ))

# colnames(df_cleaned)
# df_cleaned
```

Restructure the data as needed. (Hint: Within-subjects variables require making a "long" version of the dataset.)

```
library(reshape)

# Creating a long version of the cleaned dataset
# REF: https://www.digitalocean.com/community/tutorials/r-melt-and-cast-function
df_long <- melt(df_cleaned, id = c("ResponseId","initiator_type"))

dim(df_long)
```

```
## [1] 322   4
```

Include significance testing. Include all code for reproducing your analyses and figures.

1. Whether people care about getting a return apology after being the first to apologize;

- Conduct significance testing on variables of: 'feelings_youalone' v. 'feelings_bothyoufirst'

```
library(stats)
library(ggplot2)
library(jtools)

df_q1 <- subset(df_cleaned, select = c(
    ResponseId,
    feelings_youalone,
    feelings_bothyoufirst
    ))
# REF: https://www.r-bloggers.com/2021/10/paired-sample-t-test-using-r/
df_q1_long <- melt(df_q1, id="ResponseId")
# attach(df_q1_long)
# Summary statistics of `feelings_youalone` v. `feelings_bothyoufirst`
by(data = df_q1_long,
```

```
   INDICES = df_q1_long[["variable"]],
   FUN = summary)
```

```
## df_q1_long[["variable"]]: feelings_youalone
##   ResponseId                        variable      value
##  Length:46          feelings_youalone    :46   Min.   :-30.00
##  Class :character   feelings_bothyoufirst: 0   1st Qu.:-30.00
##  Mode  :character                              Median :-20.00
##                                                Mean   :-18.61
##                                                3rd Qu.:-10.00
##                                                Max.   : 10.00
## ------------------------------------------------------------
## df_q1_long[["variable"]]: feelings_bothyoufirst
##   ResponseId                        variable      value
##  Length:46          feelings_youalone    : 0   Min.   :-30.000
##  Class :character   feelings_bothyoufirst:46   1st Qu.:  1.000
##  Mode  :character                              Median : 10.000
##                                                Mean   :  7.848
##                                                3rd Qu.: 20.000
##                                                Max.   : 30.000
```

```
lapply(df_q1[,c("feelings_bothyoufirst", "feelings_youalone")], sd)
```

```
## $feelings_bothyoufirst
## [1] 14.29058
##
## $feelings_youalone
## [1] 12.01569
```

```
# Correlation test
feelings_youalone <- df_q1[["feelings_youalone"]]
feelings_bothyoufirst <- df_q1[["feelings_bothyoufirst"]]
cor.test(x = feelings_youalone, y = feelings_bothyoufirst,
         method = c("pearson"),
         conf.level = 0.95)
```

```
##
##  Pearson's product-moment correlation
##
## data:  feelings_youalone and feelings_bothyoufirst
## t = 3.3844, df = 44, p-value = 0.001509
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1891307 0.6579663
## sample estimates:
##       cor
## 0.4544761
```

```
t.test(feelings_youalone,
    feelings_bothyoufirst,
    paired=TRUE)
```

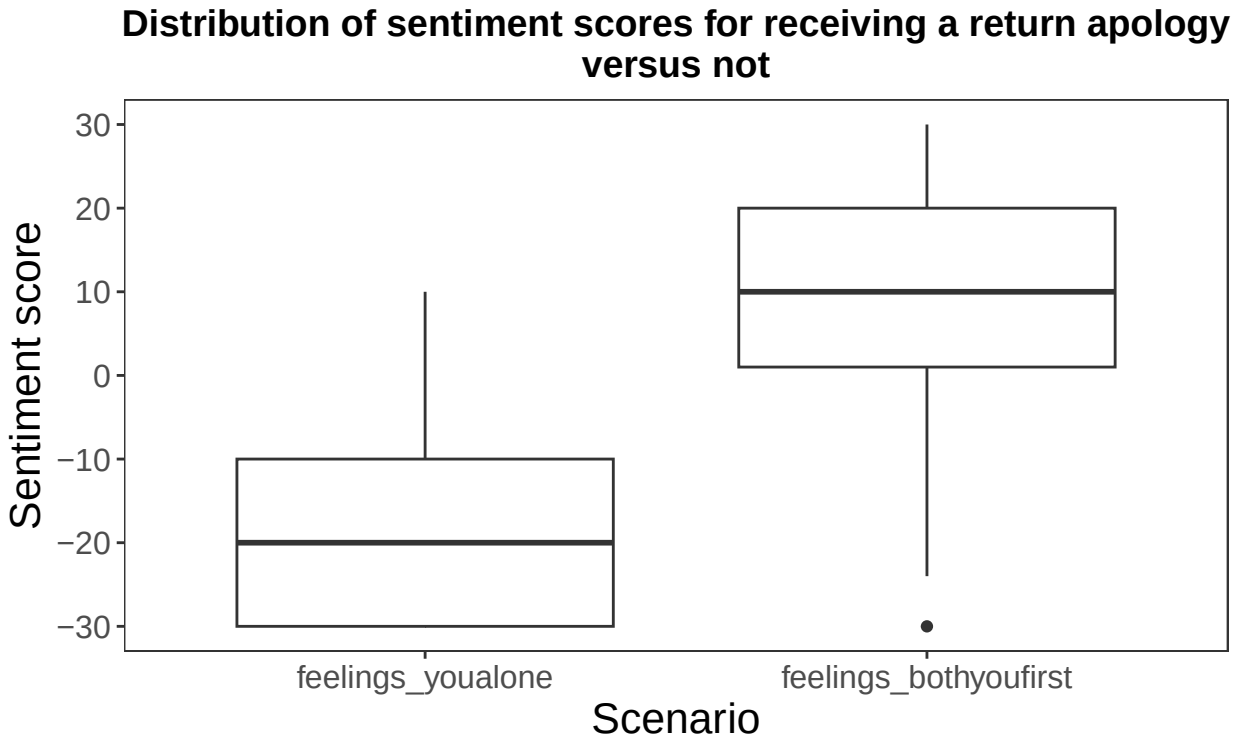
```
##
##  Paired t-test
##
## data:  feelings_youalone and feelings_bothyoufirst
```

```
## t = -12.932, df = 45, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -30.57694 -22.33611
## sample estimates:
## mean difference
##       -26.45652
```

```
# REF: http://www.sthda.com/english/wiki/ggplot2-box-plot-quick-start-guide-r-software-and-data-visuali
png("q1_box_plot.png")
p <- ggplot(df_q1_long, aes(x=variable, y=value)) +
    geom_boxplot() +
    xlab("Scenario") + ylab("Sentiment score") +
    theme_apa(legend.font.size = 16,
      x.font.size = 16,
      y.font.size = 16,
      facet.title.size = 16)  +
    theme(aspect.ratio = 1/2,
      legend.title = element_text(size = 12),
      axis.text.y = element_text(size = 12),
      axis.text.x = element_text(size = 12),
      plot.title = element_text(hjust=0.5)) +
    scale_y_continuous(breaks=seq(-30,30,10)) +
    ggtitle(wrapper("Distribution of sentiment scores for receiving a return apology versus not", width
print(p)
dev.off()
```

```
## pdf
##   2
```

```
p
```



```r
# Homogeneity in variances - above REF
bartlett.test(df_q1_long[["value"]] ~ df_q1_long[["variable"]])
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  df_q1_long[["value"]] by df_q1_long[["variable"]]
## Bartlett's K-squared = 1.3313, df = 1, p-value = 0.2486
```

```r
# Paired samples t-test - REF: https://www.statmethods.net/stats/ttest.html
q1.t_test <- t.test(
    formula=df_q1_long[["value"]] ~ df_q1_long[["variable"]],
    paired=TRUE)
q1.t_test
```

```
##
##  Paired t-test
##
## data:  df_q1_long[["value"]] by df_q1_long[["variable"]]
## t = -12.932, df = 45, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -30.57694 -22.33611
## sample estimates:
## mean difference
##       -26.45652
```

```r
# detach(df_q1_long)
```

2. Whether this varies as function of individual differences in "initiator type"; and

- 'initiator_type' variable

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following object is masked from 'package:reshape':
##
##     rename
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
```

```
##     recode
```

```
# REF: https://www.scribbr.com/statistics/anova-in-r/
df_q2 <- subset(df_cleaned, select = c(
    ResponseId,
    feelings_youalone,
    feelings_bothyoufirst,
    initiator_type
    ))

df_q2_long <- melt(df_q2, id=c("ResponseId", "initiator_type") )
# attach(df_q2_long)

group_by(df_q2_long, initiator_type, variable) %>%
  summarise(
    mean = round(mean(value, na.rm = TRUE), 4),
    sd = round(sd(value, na.rm = TRUE), 4),
    count = length(value)
  )
```

```
## `summarise()` has grouped output by 'initiator_type'. You can override using
## the `.groups` argument.
```

```
## # A tibble: 6 x 5
## # Groups:   initiator_type [3]
##   initiator_type variable               mean    sd count
##   <chr>          <fct>                 <dbl> <dbl> <int>
## 1 always         feelings_youalone     -13.5  11.9     21
## 2 always         feelings_bothyoufirst  12.8  11.1     21
## 3 conditional    feelings_youalone     -24.5  11.3     15
## 4 conditional    feelings_bothyoufirst   5.07 16.4     15
## 5 never          feelings_youalone     -20.6   9.20    10
## 6 never          feelings_bothyoufirst   1.7  14.9     10
```

```
aov1.results <- aov(value ~ variable + initiator_type + variable:initiator_type, data=df_q2_long)
print("Type I ANOVA, assuming no significant interaction")
```

```
## [1] "Type I ANOVA, assuming no significant interaction"
```

```
summary(aov1.results)
```

```
##                          Df Sum Sq Mean Sq F value   Pr(>F)
## variable                  1  16099   16099 101.965 2.89e-16 ***
## initiator_type            2   1951     975   6.178  0.00311 **
## variable:initiator_type   2    158      79   0.500  0.60828
## Residuals                86  13578     158
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov_mod <- lm(value ~ variable*initiator_type, data=df_q2_long)
# print("Type II ANOVA, assuming no significant interaction")
# Anova(aov_mod.results, type=2)
print("Type III ANOVA, assuming significant interaction")
```

```
## [1] "Type III ANOVA, assuming significant interaction"
```

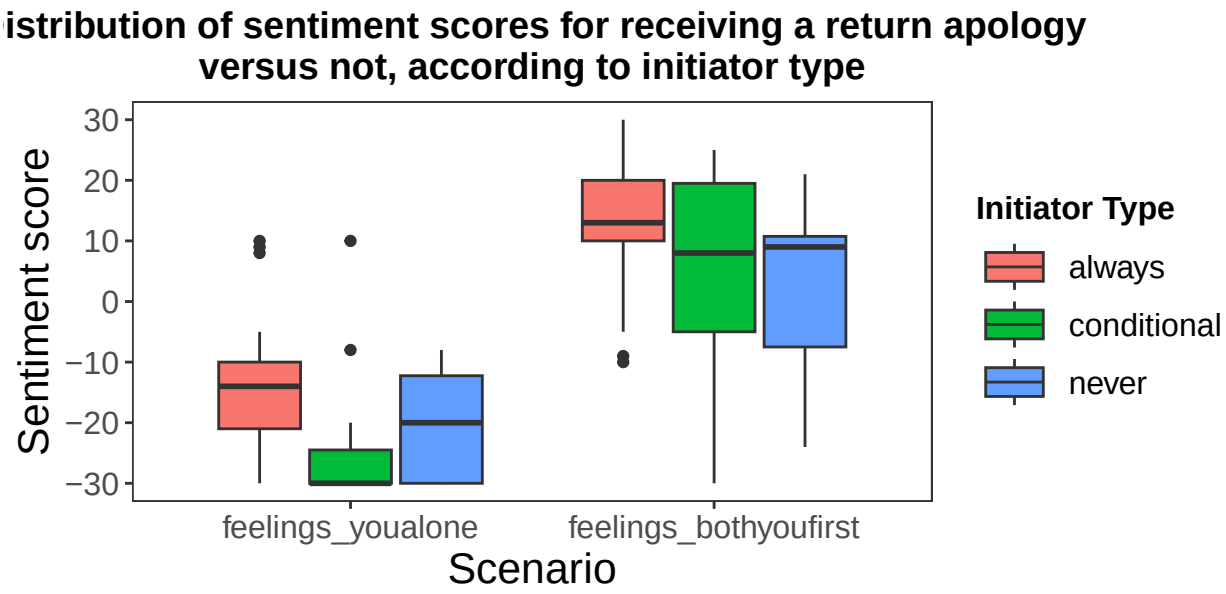
```
Anova(aov_mod, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: value
##                            Sum Sq Df F value    Pr(>F)
## (Intercept)                3813.8  1 24.1551 4.208e-06 ***
## variable                   7228.6  1 45.7836 1.526e-09 ***
## initiator_type             1107.6  2  3.5075   0.03433 *
## variable:initiator_type     157.9  2  0.5000   0.60828
## Residuals                 13578.2 86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
png("q2_box_plot.png")
p <- ggplot(df_q2_long, aes(x=variable, y=value, fill=initiator_type)) +
    geom_boxplot() +
    xlab("Scenario") + ylab("Sentiment score") +
    labs(fill='Initiator Type') +
    theme_apa(legend.font.size = 12,
      x.font.size = 16,
      y.font.size = 16,
      facet.title.size = 16,
      legend.use.title = TRUE) +
    theme(aspect.ratio = 1/2,
      legend.title = element_text(size = 12),
      axis.text.y = element_text(size = 12),
      axis.text.x = element_text(size = 12),
      plot.title = element_text(hjust=0.5)) +
    scale_y_continuous(breaks=seq(-30,30,10)) +
    ggtitle(wrapper("Distribution of sentiment scores for receiving a return apology versus not, accordi
print(p)
dev.off()
```

```
## pdf
##   2
```

```
p
```



```
# detach(df_q2_long)
```

3. Whether a return apology is simply viewed as a form of forgiveness.

- Conduct significance testing on variables of: 'feelings_bothyoufirst' v. 'feelings_youaloneforgiven'

```
df_q3 <- subset(df_cleaned, select = c(
    ResponseId,
    feelings_youalone,
    feelings_youaloneforgiven,
    feelings_bothyoufirst
    ))
# Correlation test
cor.test(x = df_q3[["feelings_youaloneforgiven"]], y = df_q3[["feelings_bothyoufirst"]],
         method = c("pearson"),
         conf.level = 0.95)
```

```
##
##  Pearson's product-moment correlation
##
## data:  df_q3[["feelings_youaloneforgiven"]] and df_q3[["feelings_bothyoufirst"]]
## t = 2.3261, df = 44, p-value = 0.02468
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.04493746 0.56676988
## sample estimates:
##       cor
## 0.3309186
```

```
# REF: https://www.r-bloggers.com/2021/10/paired-sample-t-test-using-r/
df_q3_long <- melt(df_q3, id="ResponseId")
# attach(df_q3_long)
# Summary statistics of `feelings_youalone` v. `feelings_bothyoufirst`
group_by(df_q3_long, variable) %>%
  summarise(
    mean = round(mean(value, na.rm = TRUE), 6),
    sd = round(sd(value, na.rm = TRUE), 6),
    count = length(value)
  )
```

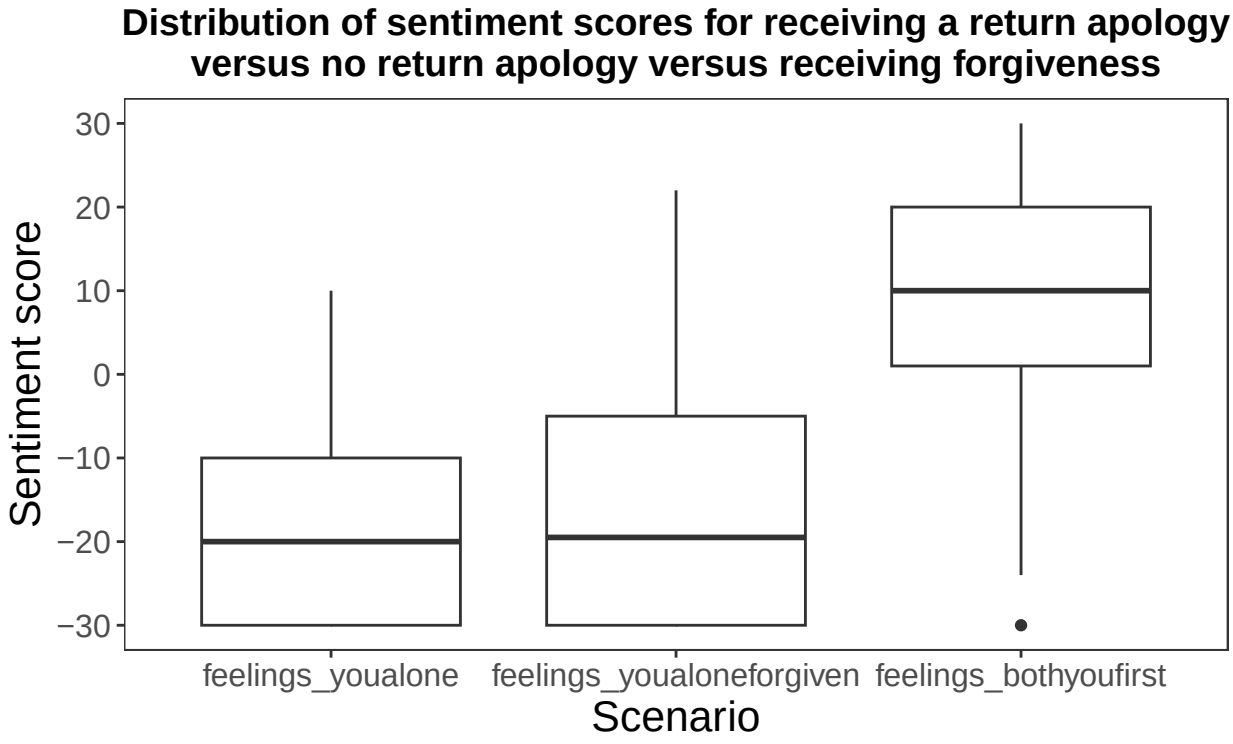
```
## # A tibble: 3 x 4
##   variable                    mean    sd count
##   <fct>                      <dbl> <dbl> <int>
## 1 feelings_youalone         -18.6   12.0    46
## 2 feelings_youaloneforgiven -14.3   15.6    46
## 3 feelings_bothyoufirst       7.85  14.3    46
```

```
# Plotting - above REF
png("q3_box_plot.png")
p <- ggplot(df_q3_long, aes(x=variable, y=value)) +
    geom_boxplot() +
    xlab("Scenario") + ylab("Sentiment score") +
    theme_apa(legend.font.size = 16,
      x.font.size = 16,
      y.font.size = 16,
      facet.title.size = 16)  +
```

```
    theme(aspect.ratio = 1/2,
      legend.title = element_text(size = 12),
      axis.text.y = element_text(size = 12),
      axis.text.x = element_text(size = 12),
      plot.title = element_text(hjust=0.5)) +
    scale_y_continuous(breaks=seq(-30,30,10)) +
    ggtitle(wrapper("Distribution of sentiment scores for receiving a return apology versus no return ap
print(p)
dev.off()
```

```
## pdf
##   2
```

```
p
```




```
# Homogeneity in variances - above REF
bartlett.test(value ~ variable, data = df_q3_long)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  value by variable
## Bartlett's K-squared = 3.0773, df = 2, p-value = 0.2147
```

```
t.test(df_q3[,"feelings_youaloneforgiven"],
    df_q3[,"feelings_bothyoufirst"],
    paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  df_q3[, "feelings_youaloneforgiven"] and df_q3[, "feelings_bothyoufirst"]
```

```
## t = -8.6488, df = 45, p-value = 3.996e-11
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -27.25729 -16.96010
## sample estimates:
## mean difference
##        -22.1087
```

```
pairwise.t.test(df_q3_long[["value"]], df_q3_long[["variable"]], p.adjust.method = "bonferroni", paired
```

```
##
##  Pairwise comparisons using paired t tests
##
## data:  df_q3_long[["value"]] and df_q3_long[["variable"]]
##
##                           feelings_youalone feelings_youaloneforgiven
## feelings_youaloneforgiven 0.11              -
## feelings_bothyoufirst     2.8e-16           1.2e-10
##
## P value adjustment method: bonferroni
```

```
# detach(df_q3_long)
```

Data Analysis Task 2

(a) Produce a single bar graph that shows the average of the "feelings" variable for all six scenarios, in order of decreasing value. Include error bars (standard errors or confidence intervals). Label fully. Describe your observations in 1-2 sentences.

Describe your observations in 1-2 sentences.

In general, participants associated the highest positive sentiments with scenarios in which both them and the other person apologizes. Additionally, the scenarios in which no apoogy is received from the other person are associated with negative sentiments, even if the other person explicitly provides their forgiveness.

```
df_a <- subset(df_cleaned, select = c(
    ResponseId,
    feelings_youalone,
    feelings_bothyoufirst,
    feelings_themalone,
    feelings_boththemfirst,
    feelings_neither,
    feelings_youaloneforgiven
    ))
df_a_long <- melt(df_a, value.name = "scenario", id = 'ResponseId')
colnames(df_a_long)[2] ="scenario"
# attach(df_a_long)

# REF: http://www.sthda.com/english/wiki/ggplot2-error-bars-quick-start-guide-r-software-and-data-visua
data_summary <- function(data, varname, groupnames){
  require(plyr)
  summary_func <- function(x, col){
    c(mean = mean(x[[col]], na.rm=TRUE),
      sd = sd(x[[col]], na.rm=TRUE))
  }
  data_sum<-ddply(data, groupnames, .fun=summary_func,
```

```
                  varname)
 return(data_sum)
}
data_summ <- data_summary(df_a_long, varname="value",
                    groupnames=c("scenario"))
```

```
## Loading required package: plyr
```

```
## ------------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## ------------------------------------------------------------------------------
```

```
##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following objects are masked from 'package:reshape':
##
##     rename, round_any
```

```
# REFs:
# - http://www.sthda.com/english/wiki/ggplot2-barplots-quick-start-guide-r-software-and-data-visualizat
# - https://stackoverflow.com/questions/30183199/ggplot2-plot-mean-with-geom-bar
png("task2_qA_bar_plot.png")
p<-ggplot(data=data_summ) +
    geom_bar(
        aes(x=reorder(scenario, -mean), y=mean),
            position = "dodge", stat = "identity", width=0.9) +
    xlab("Scenario") + ylab("Mean sentiment score") +
    theme_apa(legend.font.size = 12,
      x.font.size = 16,
      y.font.size = 16,
      facet.title.size = 16,
      legend.use.title = TRUE) +
    theme(aspect.ratio = 1/2,
      legend.title = element_text(size = 12),
      axis.text.y = element_text(size = 12),
      axis.text.x = element_text(size = 12, angle = 90),
      plot.title = element_text(hjust=0.5)) +
    scale_y_continuous(breaks=seq(-30,30,5)) +
    ggtitle("Avg. `feelings` score by scenario")
    geom_errorbar(
        aes(x=scenario, ymin=mean-sd, ymax=mean+sd),
        width=0.4, colour="orange", alpha=0.9, size=1.3)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
```

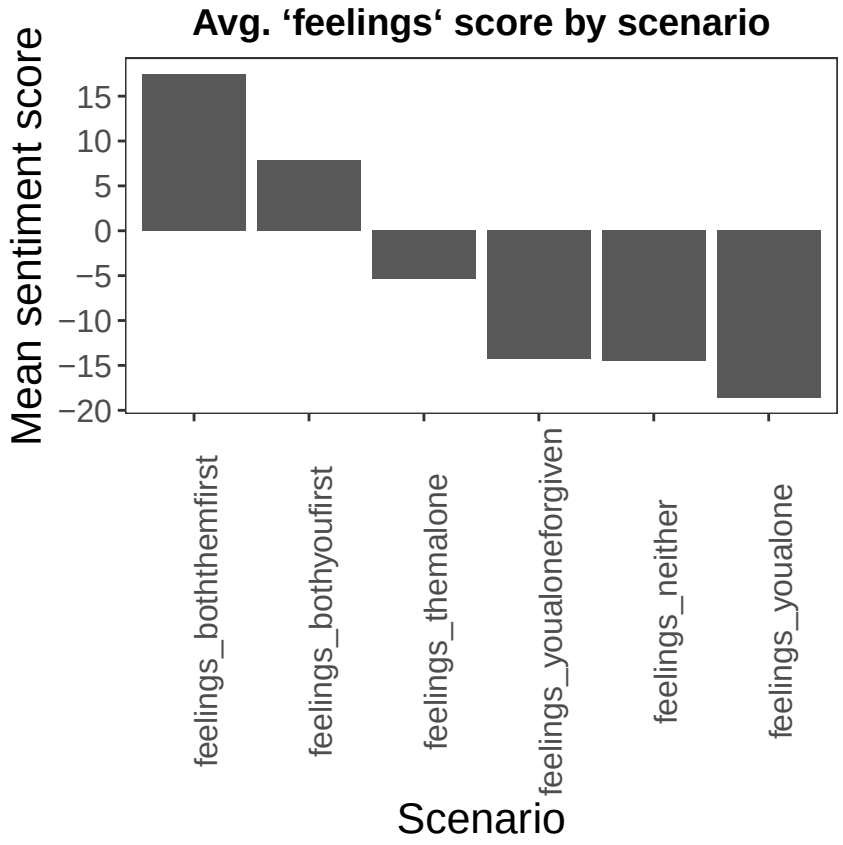
```
## generated.
```

```
## mapping: x = ~scenario, ymin = ~mean - sd, ymax = ~mean + sd
## geom_errorbar: na.rm = FALSE, orientation = NA, width = 0.4
## stat_identity: na.rm = FALSE
## position_identity
```

```
print(p)
dev.off()
```

```
## pdf
##   2
```

```
p
```




```
# detach(df_a_long)
```

(b) Conduct a one way ANOVA to determine if there are differences in feelings across the six scenarios.

```
# REF: http://www.sthda.com/english/wiki/one-way-anova-test-in-r
# Compute the analysis of variance
res.aov <- aov(value ~ scenario, data = df_a_long)
# Summary of the analysis
summary(res.aov)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## scenario      5  47408    9482   50.71 <2e-16 ***
## Residuals   270  50482     187
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Then perform pairwise t-tests to compare "feelings_youalone" to the other five scenarios.

```
pairwise.t.test(df_a_long[["value"]], df_a_long[["scenario"]], p.adjust.method = "bonferroni", paired =
```

```
##
##  Pairwise comparisons using paired t tests
##
## data:  df_a_long[["value"]] and df_a_long[["scenario"]]
##
##                           feelings_youalone feelings_bothyoufirst
## feelings_bothyoufirst     1.4e-15           -
## feelings_themalone        0.00135           0.00803
## feelings_boththemfirst    < 2e-16           0.00052
## feelings_neither          0.63127           1.1e-09
## feelings_youaloneforgiven 0.56652           6.0e-10
##                           feelings_themalone feelings_boththemfirst
## feelings_bothyoufirst     -                  -
## feelings_themalone        -                  -
## feelings_boththemfirst    1.9e-08            -
## feelings_neither          0.03896            9.1e-14
## feelings_youaloneforgiven 0.17833            8.2e-15
##                           feelings_neither
## feelings_bothyoufirst     -
## feelings_themalone        -
## feelings_boththemfirst    -
## feelings_neither          -
## feelings_youaloneforgiven 1.00000
##
## P value adjustment method: bonferroni
```

Describe your observations in 1-2 sentences.

The results of these pairwise t-test show that the differences of the `feelings_youalone` scenario with the scenarios of `feelings_bothyoufirst`, `feelings_themalone`, `feelings_boththemfirst` are statistically significant, which supports the observations made above that scenarios in which the other person apologizes are significantly different from the case in which the other person does not apologize.

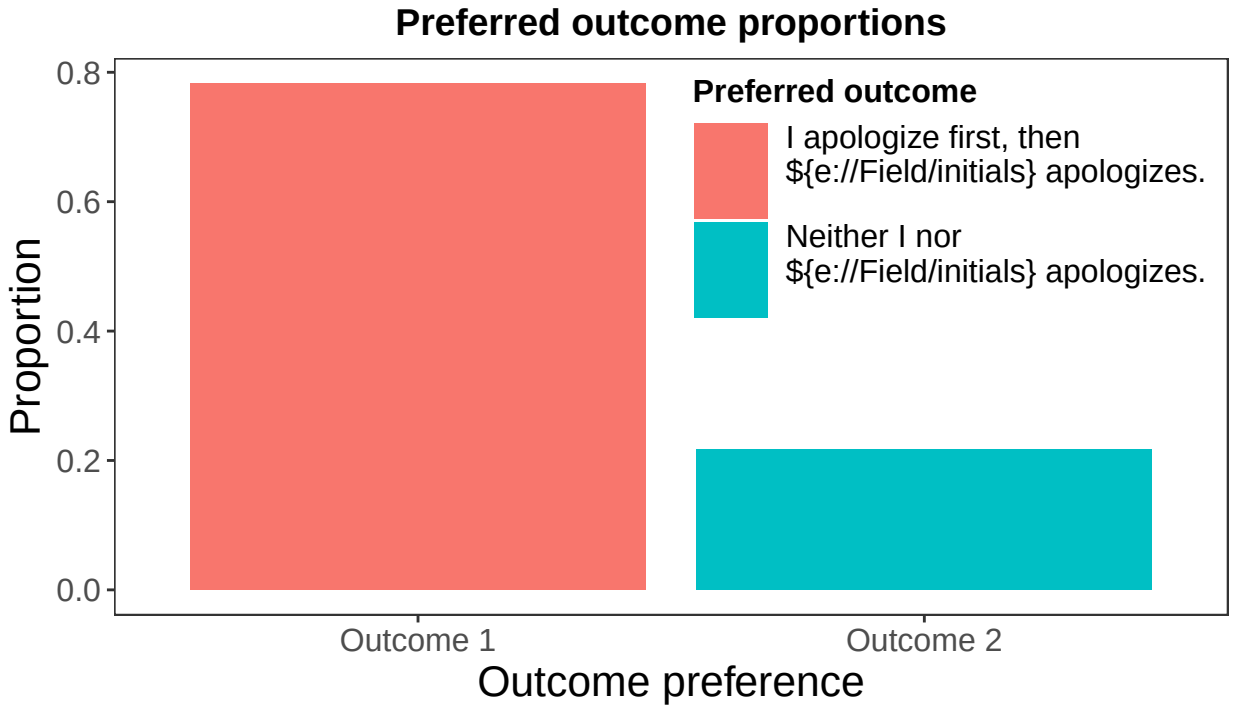
(c) Create a graph showing the proportion of people choosing each of the different options for the following variable: outcome_binary1.

```
bin_opts <- c("I apologize first, then \n${e://Field/initials} apologizes.\n", "Neither I nor \n${e://F
t <- as.data.frame(table(df_c[["outcome_binary1"]], dnn = list("preferred_outcome")), responseName = "f
t$total <- sum(t$freq)
t$bin_label <- c("Outcome 1", "Outcome 2")

plot <- ggplot(data=t, aes(x=bin_label, y=freq/total, fill=preferred_outcome)) +
    geom_bar(stat="identity") +
    ggtitle("Preferred outcome proportions") +
    xlab("Outcome preference") + ylab("Proportion") +
    scale_fill_discrete(name = "Preferred outcome", labels = bin_opts) +
    theme_apa(legend.font.size = 12,
      x.font.size = 16,
```

```
      y.font.size = 16,
      facet.title.size = 16,
      legend.use.title = TRUE) +
    theme(aspect.ratio = 1/2,
      legend.title = element_text(size = 12),
      axis.text.y = element_text(size = 12),
      axis.text.x = element_text(size = 12),
      plot.title = element_text(hjust=0.5),
      legend.position = c(0.75, 0.75))

plot
```



Conduct a test to determine if the proportion differences across the answers are significantly different from one another.

```
# REF: https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/prop.test
prop.test(x = t$freq, n = t$total)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  t$freq out of t$total
## X-squared = 27.174, df = 1, p-value = 1.86e-07
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.3749093 0.7555255
## sample estimates:
##    prop 1    prop 2
## 0.7826087 0.2173913
```