Feedback MTurk Study

Introduction

Load Data

```
{{r, include=FALSE}} # d <- fread('Lungs_November+14,+2020_17.33.csv') d <- fread('../check-valid-respon
#head(d)
```

“‘{{r, include=FALSE}} d_respondents_only <- d[(Status == “IP Address”) & (Finished == ‘True’),]

Remove these survey responses because they were from people who did the survey again. Double check that they are removed:

d_respondents_only <- d_respondents_only[!ResponseId %in% c(‘R_1eRkKqfVAmkVzj2’, ‘R_3FR03xu5zyOsRSU’, ‘R_3HBQsMSMCgXPpKf’, ‘R_dbzictBknL9jG3T’),]

These WorkerId put in all 1 response (all Normal or all Pneumonia)

d_respondents_only <- d_respondents_only[!Q80 %in% c(“A119EX2L0DNN1B”, “A12NQJV6TA5OWB”, “A18WFPSLFV4FKY”, “A1BUYK6LXYWMLL”, “A1FHRZXSE7XNJ4”, “A1GMYDH5MKN105”, “A2GSZ3D2XXC533”, “A2IGIOD74EPOEF”, “A2J016DRTOBXWO”, “A2NGFU82LMJ80X”, “A32K1M0A36EAK5”, “A371SNJNNUY9Z6”, “A3BPENSX5EVJ2H”, “A3EPIT2P3ISA3K”, “A3NYIJYBHAJ74V”, “AUFLTHQAXWLH1”, “AVINXZZV3FNG7” , “A1CD7O60QAQQRT”, “A1CF1W8CP0DHB0”, “A1PGY59BR6C5BX”, “A1YSYI926BBOHW”, “A1Z3GFH6MNSU46”, “A211KGJ94WNFLN”, “A26RPQDD0RQEHL”, “A2BUHMLNE3LUU0”, “A2J5BRQ88W745H”, “A2XIHO2W7EEP32”, “A3EZ0H07TSDAPW”, “A3FLBC6LC5GJ3W”, “A3QLKLIQW1B1FR”, “A8F6JFG0WSELT”, “A9K6IVBA0J1CX”, “ADLZLGHKOAEE6”, “AE7NJG0KOVZYJ” , “AG5RF4UGQJ7A7” , “AQ9Y6WD8O72ZC” , “tuturtu”),]

These people just gave alternating responses (Normal, Pneumonia, Normal,. . . ,Pneumonia)

d_respondents_only <- d_respondents_only[!Q80 %in% c(‘A1W05TSPORJPXR’ ,‘A3SUWCLD1GEGM7’ ,‘A3A09JB9X1RBXW’ ,‘A7VQQEIBSM9IU’ ,‘A8DER1QY96C5X’ ,‘A1M8MNKK8H5ZGW’ ,‘A34D5D6PU193AR’),]

#head(d_respondents_only) “‘

“‘{{r, include=FALSE}} #rename task phase questions setnames(d_respondents_only, old = c(‘Q2’, ‘Q42’), new = c(‘Self_Reflect_Q1’, ‘Self_Reflect_Q2’))

setnames(d_respondents_only, old = c(‘Q69’, ‘Q89’), new = c(‘Control_Q1’, ‘Control_Q2’))

setnames(d_respondents_only, old = c(‘Q80’, ‘Q82’, ‘Q83’, ‘Q84’, ‘SC0’, ‘FL_6_DO’), new = c(‘Amazon_Turk_ID’, ‘Gender’, ‘Age_Range’, ‘Education_Level’, ‘Total_Score’, ‘Assignment’)) setnames(d_respondents_only, old = c(‘Q1’, ‘Q5’, ‘Q6’, ‘Q7’, ‘Q16’, ‘Q17’, ‘Q18’, ‘Q19’, ‘Q20’, ‘Q21’, ‘Q8’, ‘Q9’, ‘Q10’, ‘Q11’, ‘Q22’, ‘Q23’, ‘Q24’, ‘Q25’, ‘Q26’, ‘Q27’, ‘Q12’, ‘Q13’, ‘Q14’, ‘Q15’, ‘Q28’, ‘Q29’, ‘Q30’, ‘Q31’, ‘Q32’, ‘Q33’), new = c(‘Q1’, ‘Q2’, ‘Q3’, ‘Q4’, ‘Q5’, ‘Q6’, ‘Q7’, ‘Q8’, ‘Q9’, ‘Q10’, ‘Q11’, ‘Q12’, ‘Q13’, ‘Q14’, ‘Q15’, ‘Q16’, ‘Q17’, ‘Q18’, ‘Q19’, ‘Q20’, ‘Q21’, ‘Q22’, ‘Q23’, ‘Q24’, ‘Q25’, ‘Q26’, ‘Q27’, ‘Q28’, ‘Q29’, ‘Q30’))

d_respondents_only[, c(“Q1_Score”, “Q2_Score”, “Q3_Score”, “Q4_Score”, “Q5_Score”, “Q6_Score”, “Q7_Score”, “Q8_Score”, “Q9_Score”, “Q10_Score”, “Q11_Score”, “Q12_Score”, “Q13_Score”, “Q14_Score”, “Q15_Score”, “Q16_Score”, “Q17_Score”, “Q18_Score”, “Q19_Score”, “Q20_Score”, “Q21_Score”, “Q22_Score”, “Q23_Score”, “Q24_Score”, “Q25_Score”, “Q26_Score”, “Q27_Score”, “Q28_Score”, “Q29_Score”, “Q30_Score”) := list(ifelse(Q1 == “Normal”, 1, 0), ifelse(Q2 == “Normal”, 1, 0), ifelse(Q3 == “Pneumonia”, 1, 0), ifelse(Q4 == “Pneumonia”, 1, 0), ifelse(Q5 == “Normal”, 1, 0), ifelse(Q6 == “Pneumonia”, 1, 0), ifelse(Q7 == “Pneumonia”, 1, 0), ifelse(Q8 == “Normal”, 1, 0), ifelse(Q9 == “Pneumonia”, 1, 0), ifelse(Q10 == “Normal”, 1, 0), ifelse(Q11 == “Pneumonia”, 1, 0), ifelse(Q12 == “Normal”, 1, 0), ifelse(Q13 == “Pneumonia”, 1, 0), ifelse(Q14 == “Pneumonia”, 1, 0), ifelse(Q15 == “Normal”, 1, 0), ifelse(Q16 == “Normal”, 1, 0), ifelse(Q17 == “Pneumonia”, 1, 0), ifelse(Q18 == “Normal”, 1, 0), ifelse(Q19 == “Pneumonia”, 1, 0), ifelse(Q20 == “Normal”, 1, 0), ifelse(Q21 == “Normal”, 1, 0), ifelse(Q22 == “Normal”, 1, 0), ifelse(Q23 == “Pneumonia”, 1, 0), ifelse(Q24 == “Normal”, 1, 0), ifelse(Q25 == “Pneumonia”, 1, 0), ifelse(Q26 == “Pneumonia”, 1, 0), ifelse(Q27 == “Pneumonia”, 1, 0), ifelse(Q28 == “Pneumonia”, 1, 0), ifelse(Q29 == “Normal”, 1, 0), ifelse(Q30 == “Normal”, 1, 0))]

d_respondents_only[, Assignment_Group := ifelse(Assignment == “FL_17”, “Control”, ifelse(Assignment == “FL_14”, “Self-Reflect”, ifelse(Assignment == “FL_15”, “Medical Feedback”, ifelse(Assignment == “FL_16”, “Positive Images”, “Negative Images”))))]

d_respondents_only[, c(“TaskPhase1_Score”, “TaskPhase2_Score”, “TaskPhase3_Score”) := list(sum(Q1_Score, Q2_Score, Q3_Score, Q4_Score, Q5_Score, Q6_Score, Q7_Score, Q8_Score, Q9_Score, Q10_Score)/10, sum(Q11_Score, Q12_Score, Q13_Score, Q14_Score, Q15_Score, Q16_Score, Q17_Score, Q18_Score, Q19_Score, Q20_Score)/10, sum(Q21_Score, Q22_Score, Q23_Score, Q24_Score, Q25_Score, Q26_Score, Q27_Score, Q28_Score, Q29_Score, Q30_Score)/10), by = Amazon_Turk_ID]

#head(d_respondents_only)

“‘

```
# ?register_google
# register_google(key = "AIzaSyCTk2a5vIEqcvgz9KmQmItoNF7J8_hiMMk")
#
# #uses Google API to obtain location data based on longitude and latitude....dont use unless necessary
# d_respondents_only[ , c("housenumber", "street", "city", "county", "state", "zip", "country") := revg
# #
# head(d_respondents_only)
# #
# #
# fwrite(d_respondents_only, file='datatable_clean_survey_responses_v2.dta')

d_respondents <- fread('datatable_clean_survey_responses_v2.dta')

setnames(d_respondents,
         old = c('Duration (in seconds)'),
         new = c('Survey_Duration'))
head(d_respondents)
```

```
##              StartDate             EndDate     Status      IPAddress Progress
## 1: 2020-11-09 20:46:55 2020-11-09 20:50:39 IP Address 174.88.123.135      100
## 2: 2020-11-09 20:47:33 2020-11-09 20:51:24 IP Address  172.93.166.91      100
## 3: 2020-11-09 20:47:23 2020-11-09 20:51:35 IP Address  68.36.215.223      100
## 4: 2020-11-09 20:46:32 2020-11-09 20:51:43 IP Address   99.75.53.174      100
## 5: 2020-11-09 20:47:44 2020-11-09 20:52:08 IP Address   24.35.119.43      100
## 6: 2020-11-09 20:46:47 2020-11-09 20:52:39 IP Address  98.212.214.93      100
##    Survey_Duration Finished        RecordedDate        ResponseId
## 1:             223     TRUE 2020-11-09 20:50:39 R_VLuUQ4C82PP9HEd
## 2:             231     TRUE 2020-11-09 20:51:25 R_29cCZD1XK1dpmdY
## 3:             251     TRUE 2020-11-09 20:51:35 R_3lVN8EncJofnqnV
## 4:             310     TRUE 2020-11-09 20:51:43 R_50vJlfmoFTK1IeB
## 5:             264     TRUE 2020-11-09 20:52:08 R_1dFaKMSjyE3FJHg
## 6:             351     TRUE 2020-11-09 20:52:39 R_25vjj4Ik4Dkm2UN
##    RecipientLastName RecipientFirstName RecipientEmail ExternalReference
## 1:                NA                 NA             NA                NA
## 2:                NA                 NA             NA                NA
## 3:                NA                 NA             NA                NA
## 4:                NA                 NA             NA                NA
## 5:                NA                 NA             NA                NA
## 6:                NA                 NA             NA                NA
##    LocationLatitude LocationLongitude DistributionChannel UserLanguage
## 1:         43.67850          -79.2935           anonymous           EN
## 2:         33.74850          -84.3871           anonymous           EN
## 3:         42.65630          -83.1231           anonymous           EN
## 4:         42.00031          -88.1422           anonymous           EN
## 5:         40.08180          -82.9665           anonymous           EN
## 6:         42.01280          -88.0967           anonymous           EN
##    Amazon_Turk_ID Gender Q82_3_TEXT Age_Range           Education_Level
## 1:  A4D99Y82KOLC8   Male         NA     35-44              Trade school
## 2: A1AC47WJLNW4G7   Male         NA     25-34 Master's degree and above
## 3:  A77K8W55MJEKX Female         NA     45-54         Bachelor's degree
## 4: A17TKHT8FEVHOR   Male         NA     25-34        Associate's degree
## 5: A1A0WMOOJMOF7Z Female         NA     25-34              Trade school
## 6: A2VO8C41JJIQY9   Male         NA     25-34 Master's degree and above
##           Q1        Q2        Q3        Q4        Q5        Q6        Q7
## 1: Pneumonia    Normal    Normal Pneumonia    Normal Pneumonia Pneumonia
## 2: Pneumonia    Normal    Normal Pneumonia    Normal    Normal    Normal
## 3: Pneumonia Pneumonia Pneumonia Pneumonia Pneumonia Pneumonia    Normal
## 4:    Normal Pneumonia Pneumonia    Normal Pneumonia Pneumonia Pneumonia
## 5:    Normal Pneumonia Pneumonia    Normal Pneumonia Pneumonia Pneumonia
## 6: Pneumonia    Normal    Normal Pneumonia Pneumonia Pneumonia    Normal
##           Q8        Q9       Q10
## 1:    Normal Pneumonia Pneumonia
## 2:    Normal    Normal Pneumonia
## 3:    Normal Pneumonia Pneumonia
## 4:    Normal Pneumonia    Normal
## 5:    Normal Pneumonia    Normal
## 6: Pneumonia Pneumonia Pneumonia
##
## 1:
## 2:
## 3:
## 4:
## 5: The sentiment that this place brings, and how much hospitality means to them. How open, diverse ar
## 6:
##    Q70_First Click Q70_Last Click Q70_Page Submit Q70_Click Count       Q11
## 1:              NA             NA              NA              NA    Normal
## 2:              NA             NA               NA              NA    Normal
## 3:              NA             NA               NA              NA Pneumonia
## 4:              NA             NA               NA              NA Pneumonia
## 5:           31.08          31.08           77.394               1    Normal
## 6:              NA             NA               NA              NA    Normal
##          Q12       Q13       Q14       Q15       Q16       Q17       Q18
## 1:    Normal Pnuemonia Pneumonia Pneumonia    Normal    Normal Pneumonia
## 2:    Normal Pnuemonia    Normal    Normal Pneumonia    Normal Pneumonia
## 3:    Normal Pnuemonia Pneumonia Pneumonia Pneumonia Pneumonia Pneumonia
## 4:    Normal Pnuemonia    Normal    Normal    Normal Pneumonia    Normal
## 5: Pneumonia    Normal Pneumonia Pneumonia    Normal Pneumonia Pneumonia
## 6:    Normal Pnuemonia Pneumonia    Normal    Normal Pneumonia Pneumonia
##          Q19       Q20
## 1: Pneumonia    Normal
## 2:    Normal Pneumonia
## 3: Pneumonia Pneumonia
## 4: Pneumonia Pneumonia
## 5:    Normal Pneumonia
## 6: Pneumonia Pneumonia
##
## 1:
## 2:
## 3:
## 4:
## 5: It brings awareness to a serious issue that can harm people. It's an advertisement to bring peopl
## 6:
##    Q90_First Click Q90_Last Click Q90_Page Submit Q90_Click Count       Q21
## 1:              NA             NA               NA              NA Pneumonia
## 2:              NA             NA               NA              NA    Normal
## 3:              NA             NA               NA              NA Pneumonia
## 4:              NA             NA               NA              NA    Normal
## 5:          10.128         68.736           70.972               2 Pneumonia
## 6:              NA             NA               NA              NA    Normal
##          Q22       Q23       Q24       Q25       Q26       Q27       Q28
## 1: Pneumonia    Normal Pneumonia    Normal Pneumonia Pneumonia    Normal
## 2: Pneumonia Pneumonia    Normal    Normal    Normal Pneumonia    Normal
## 3: Pneumonia    Normal Pneumonia Pneumonia Pneumonia Pneumonia Pneumonia
## 4: Pneumonia Pneumonia Pneumonia Pneumonia Pneumonia    Normal Pneumonia
## 5: Pneumonia Pneumonia    Normal Pneumonia    Normal    Normal Pneumonia
## 6: Pneumonia Pneumonia    Normal Pneumonia    Normal    Normal Pneumonia
##          Q29       Q30 Q36
## 1:    Normal Pneumonia
## 2: Pneumonia    Normal
## 3: Pneumonia Pneumonia
## 4: Pneumonia    Normal
## 5: Pneumonia Pneumonia
## 6:    Normal    Normal
##                                                            Self_Reflect_Q1
## 1:
## 2:
## 3:
## 4: I think I did pretty good. I was not expecting to do as well as I did.
## 5:
## 6:
```

```
##    Q61_First Click Q61_Last Click Q61_Page Submit Q61_Click Count Q41
## 1:              NA             NA              NA              NA    
## 2:              NA             NA              NA              NA    
## 3:              NA             NA              NA              NA    
## 4:          32.361         43.794         111.584               3    
## 5:              NA             NA              NA              NA    
## 6:              NA             NA              NA              NA    
##                                                                           Self_Reflect_Q2
## 1:                                                                                       
## 2:                                                                                       
## 3:                                                                                       
## 4: I think I did incredible. I only got 2 wrong. This was harder than the previous page.
## 5:                                                                                       
## 6:                                                                                       
##    Q62_First Click Q62_Last Click Q62_Page Submit Q62_Click Count
## 1:              NA             NA              NA              NA
## 2:              NA             NA              NA              NA
## 3:              NA             NA              NA              NA
## 4:          10.059         10.059         100.308               1
## 5:              NA             NA              NA              NA
## 6:              NA             NA              NA              NA
##                                                                                          Q38
## 1:                                                                                          
## 2:                                                                                          
## 3:                                                                                          
## 4:                                                                                          
## 5:                                                                                          
## 6: Image 2Correct diagnosis: Normal\nYou chose: ${q://QID5/ChoiceGroup/SelectedChoices}\n
##    Q63_First Click Q63_Last Click Q63_Page Submit Q63_Click Count Q43
## 1:              NA             NA              NA              NA    
## 2:              NA             NA              NA              NA    
## 3:              NA             NA              NA              NA    
## 4:              NA             NA              NA              NA    
## 5:              NA             NA              NA              NA    
## 6:           1.205        100.696         107.951              16    
##    Q64_First Click Q64_Last Click Q64_Page Submit Q64_Click Count Q45
## 1:              NA             NA              NA              NA  NA
## 2:              NA             NA              NA              NA  NA
## 3:              NA             NA              NA              NA  NA
## 4:              NA             NA              NA              NA  NA
## 5:              NA             NA              NA              NA  NA
## 6:          76.034        101.185         102.397               3  NA
##    Q65_First Click Q65_Last Click Q65_Page Submit Q65_Click Count Q47
## 1:              NA             NA              NA              NA  NA
## 2:          14.693         16.235           47.68               2  NA
## 3:              NA             NA              NA              NA  NA
## 4:              NA             NA              NA              NA  NA
## 5:              NA             NA              NA              NA  NA
## 6:              NA             NA              NA              NA  NA
##    Q66_First Click Q66_Last Click Q66_Page Submit Q66_Click Count Q46
## 1:              NA             NA              NA              NA  NA
## 2:            5.75         18.281           46.59               2  NA
## 3:              NA             NA              NA              NA  NA
## 4:              NA             NA              NA              NA  NA
## 5:              NA             NA              NA             NA  NA
## 6:              NA             NA              NA             NA  NA
##    Q67_First Click Q67_Last Click Q67_Page Submit Q67_Click Count Q48
## 1:           0.855         57.413          58.201             29  NA
## 2:              NA             NA              NA             NA  NA
## 3:          16.263         16.263          50.056              1  NA
## 4:              NA             NA              NA             NA  NA
## 5:              NA             NA              NA             NA  NA
## 6:              NA             NA              NA             NA  NA
##    Q68_First Click Q68_Last Click Q68_Page Submit Q68_Click Count Total_Score
## 1:           0.530         60.605          61.222             15          16
## 2:              NA             NA              NA             NA          12
## 3:           9.427          9.427          49.629              1          15
## 4:              NA             NA              NA             NA          21
## 5:              NA             NA              NA             NA          14
## 6:              NA             NA              NA             NA          18
##    Random ID Assignment Q1_Score Q2_Score Q3_Score Q4_Score Q5_Score Q6_Score
## 1:     14409      FL_41        0        1        0        1        1        1
## 2:     58508      FL_16        0        1        0        1        1        0
## 3:     96075      FL_41        0        0        1        1        0        1
## 4:     74553      FL_14        1        0        1        0        0        1
## 5:     35543      FL_17        1        0        1        0        0        1
## 6:     84565      FL_15        0        1        0        1        0        1
##    Q7_Score Q8_Score Q9_Score Q10_Score Q11_Score Q12_Score Q13_Score Q14_Score
## 1:        1        1        1         0         0         1         0         1
## 2:        0        1        0         0         0         1         0         0
## 3:        0        1        1         0         1         1         0         1
## 4:        1        1        1         1         1         1         0         0
## 5:        1        1        1         1         0         0         0         1
## 6:        0        0        1         0         0         1         0         1
##    Q15_Score Q16_Score Q17_Score Q18_Score Q19_Score Q20_Score Q21_Score
## 1:         0         1         0         0         1         1         0
## 2:         1         0         0         0         0         0         1
## 3:         0         0         1         0         1         0         0
## 4:         1         1         1         1         1         0         1
## 5:         0         1         1         0         0         0         0
## 6:         1         1         1         0         1         0         1
##    Q22_Score Q23_Score Q24_Score Q25_Score Q26_Score Q27_Score Q28_Score
## 1:         0         0         0         0         1         1         0
## 2:         0         1         1         0         0         1         0
## 3:         0         0         0         1         1         1         1
## 4:         0         1         0         1         1         0         1
## 5:         0         1         1         1         0         0         1
## 6:         0         1         1         1         0         0         1
##    Q29_Score Q30_Score Assignment_Group TaskPhase1_Score TaskPhase2_Score
## 1:         1         0  Negative Images              0.7              0.5
## 2:         0         1  Positive Images              0.4              0.2
## 3:         0         0  Negative Images              0.5              0.5
## 4:         0         1     Self-Reflect              0.7              0.7
## 5:         0         0          Control              0.7              0.3
## 6:         1         1 Medical Feedback              0.4              0.6
##    TaskPhase3_Score housenumber                   street            city
## 1:              0.3         351         Glen Manor Drive         Toronto
## 2:              0.5         262 Capitol Avenue Southeast         Atlanta
## 3:              0.4         440         Bedlington Drive Rochester Hills
## 4:              0.6        1200          Sycamore Avenue    Hanover Park
## 5:              0.4        1913          Brookfield Road        Columbus
## 6:              0.7         617            Boxwood Drive      Schaumburg
##           county    state     zip        country
## 1:        Canada  Ontario M4E 2X8         Canada
## 2: United States  Georgia   30312 United States
## 3: United States Michigan   48307 United States
## 4: United States Illinois   60133 United States
## 5: United States     Ohio   43229 United States
## 6: United States Illinois   60193 United States
```

```
nrow(d_respondents)
```

```
## [1] 350
```

```
#skip

# ?register_google
# register_google(key = "AIzaSyCTk2a5vIEqcvgz9KmQmItoNF7J8_hiMMk")
# ggmap_show_api_key()
#
#
# revgeocode(c(df$lon[1], df$lat[1]))
#
# d_respondents_only[ Q80 == "A1AC47WJLNW4G7", revgeocode(c(as.numeric(LocationLongitude)[1], as.numeri
# ?revgeocode
```

```
#remove duplicate Amazon Turk IDs
nrow(d_respondents)  #350 rows
```

```
## [1] 350
```

```
d_respondents <- d_respondents[ !duplicated(d_respondents$Amazon_Turk_ID) , ]  #350 rows
```

EDA

```
#some EDA

#d_respondents[ , table(state, country)]


table(d_respondents$state, d_respondents$country) %>%
        as.data.frame() %>%
        arrange(desc(Freq)) %>%
        filter(Freq>0)
```

```
##                          Var1            Var2 Freq
## 1                  Tamil Nadu           India  107
## 2                  California   United States   72
## 3                    New York   United States   22
## 4                      Kansas   United States   21
## 5                       Texas   United States   15
## 6                     Florida   United States    9
## 7               Massachusetts   United States    7
## 8                    Missouri  United States    6
## 9                 Connecticut  United States    5
## 10                    Georgia  United States    5
## 11                    Indiana  United States    5
## 12                   Michigan  United States    5
## 13                 New Jersey  United States    5
## 14                   Illinois  United States    4
## 15                   Virginia  United States    4
## 16                     Kerala          India    3
## 17                Maharashtra          India    3
## 18                   Colorado  United States    3
## 19                   Kentucky  United States    3
## 20                   Maryland  United States    3
## 21             North Carolina  United States    3
## 22                     Oregon  United States    3
## 23                    Ontario         Canada    2
## 24                    Alabama  United States    2
## 25                      Idaho  United States    2
## 26                  Minnesota  United States    2
## 27                Mississippi  United States    2
## 28                     Nevada  United States    2
## 29                       Ohio  United States    2
## 30               Pennsylvania  United States    2
## 31                 Washington  United States    2
## 32            Qarku i Tiranës        Albania    1
## 33            Khulna Division     Bangladesh    1
## 34                      Bahia         Brazil    1
## 35                    Atacama          Chile    1
## 36 Provence-Alpes-Côte d'Azur         France    1
## 37    Departamento de Olancho       Honduras    1
## 38             Andhra Pradesh          India    1
## 39                  Karnataka          India    1
## 40                   Sardegna          Italy    1
## 41                    England United Kingdom    1
## 42                    Arizona  United States    1
## 43                       Iowa  United States    1
## 44                  Louisiana  United States    1
## 45                      Maine  United States    1
## 46                   Nebraska  United States    1
## 47                   Oklahoma  United States    1
## 48             South Carolina  United States    1
## 49               South Dakota  United States    1
## 50                  Tennessee  United States    1
```

```
table(d_respondents$country) %>%
        as.data.frame() %>%
        arrange(desc(Freq))
```

```
##              Var1 Freq
## 1   United States  225
## 2           India  115
## 3          Canada    2
## 4         Albania    1
## 5      Bangladesh    1
## 6          Brazil    1
## 7           Chile    1
## 8          France    1
## 9        Honduras    1
## 10          Italy    1
## 11 United Kingdom    1
```

```r
table(d_respondents$Total_Score) %>%
  as.data.frame() %>%
  arrange(desc(Var1))
```

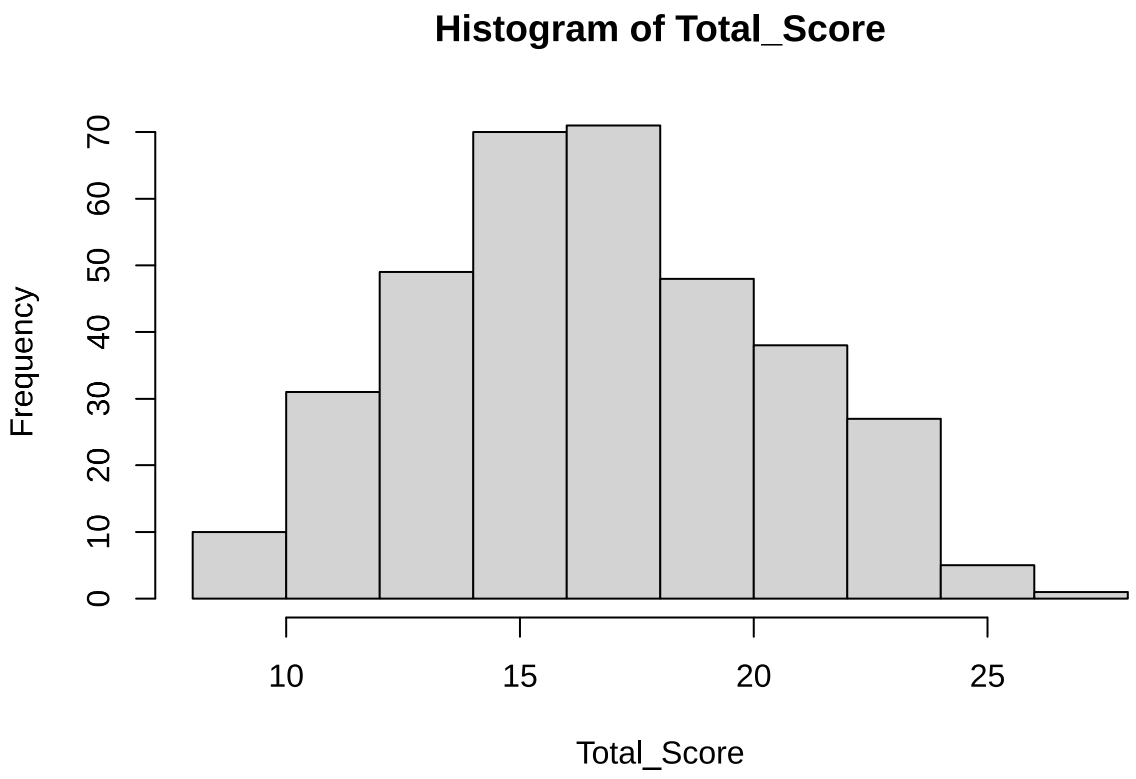
```
##    Var1 Freq
## 1    27    1
## 2    26    1
## 3    25    4
## 4    24   12
## 5    23   15
## 6    22   16
## 7    21   22
## 8    20   27
## 9    19   21
## 10   18   31
## 11   17   40
## 12   16   40
## 13   15   30
## 14   14   30
## 15   13   19
## 16   12   18
## 17   11   13
## 18   10    6
## 19    9    3
## 20    8    1
```

```r
d_respondents %>%
  group_by(Assignment_Group) %>%
  summarise(mean = mean(Total_Score),
            count = n(),
            time_duration = mean(Survey_Duration))
```

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

```
## # A tibble: 5 x 4
##   Assignment_Group  mean count time_duration
##   <chr>            <dbl> <int>         <dbl>
## 1 Control           16.7    69          638.
## 2 Medical Feedback  17.8    70          656.
## 3 Negative Images   16.5    72          783 
## 4 Positive Images   17.3    70          505.
## 5 Self-Reflect      17.2    69          612.
```

```r
#d_respondents[ , .(count = .N, avg = mean(Total_Score)), by=Assignment_Group] #same thing

d_respondents[ , hist(Total_Score)]
```



```
## $breaks
##  [1]  8 10 12 14 16 18 20 22 24 26 28
##
## $counts
##  [1] 10 31 49 70 71 48 38 27  5  1
##
## $density
##  [1] 0.014285714 0.044285714 0.070000000 0.100000000 0.101428571 0.068571429
##  [7] 0.054285714 0.038571429 0.007142857 0.001428571
##
## $mids
##  [1]  9 11 13 15 17 19 21 23 25 27
##
## $xname
## [1] "Total_Score"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

```
tapply(d_respondents$Total_Score, d_respondents$Assignment_Group, summary)
```

```
## $Control
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     8.0    14.0    16.0    16.7    19.0    24.0
##
## $`Medical Feedback`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.00   16.00   17.50   17.79   20.00   24.00
##
## $`Negative Images`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.00   13.00   16.00   16.51   19.25   25.00
##
## $`Positive Images`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.00   15.00   17.00   17.31   20.00   27.00
##
## $`Self-Reflect`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.00   14.00   17.00   17.25   20.00   25.00
```

```
tapply(d_respondents$Total_Score, d_respondents$Assignment_Group, sd)
```

```
##         Control Medical Feedback  Negative Images  Positive Images
##        3.659413         3.278798         3.996453         3.816603
##    Self-Reflect
##        3.882108
```

```
d_respondents[ , sd(Total_Score)]
```

```
## [1] 3.743141
```

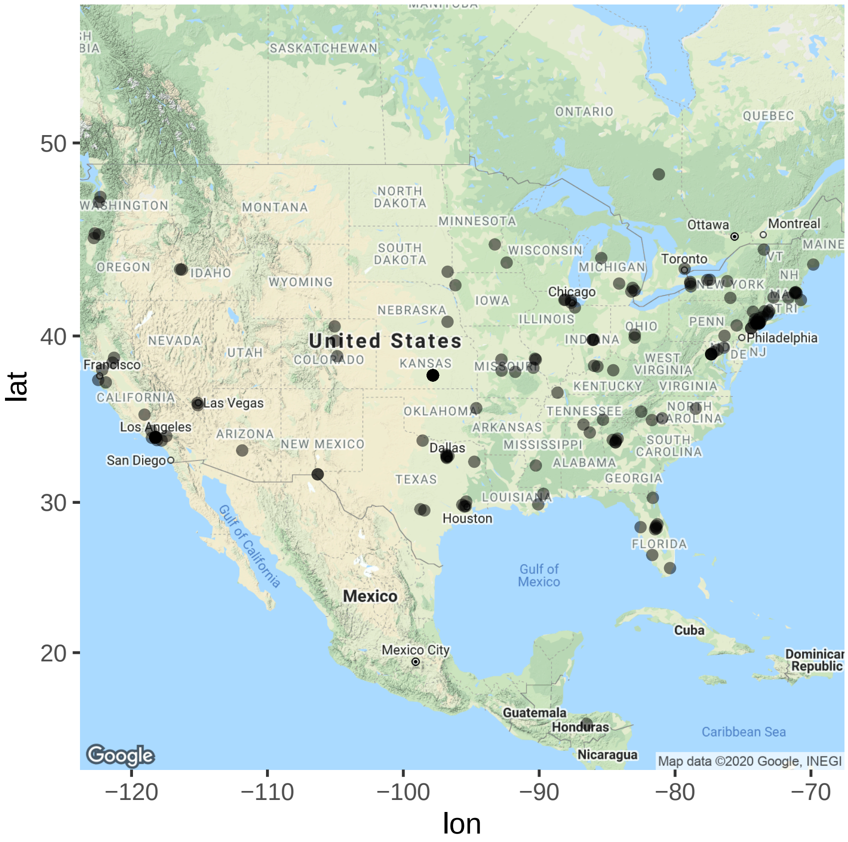
```
library(ggmap)
?register_google
register_google(key = "AIzaSyCTk2a5vIEqcvgz9KmQmItoNF7J8_hiMMk")
#ggmap_show_api_key()

us_map<-get_map(location='united states', zoom=4, maptype = "terrain",
             source='google',color='color')
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=united%20states&zoom=4&size=640x640&s
```

```
## Source : https://maps.googleapis.com/maps/api/geocode/json?address=united+states&key=xxx
```

```
ggmap(us_map) + geom_point(x=d_respondents$LocationLongitude, y = d_respondents$LocationLatitude, show_g
```

```
## Warning: `show_guide` has been deprecated. Please use `show.legend` instead.
```



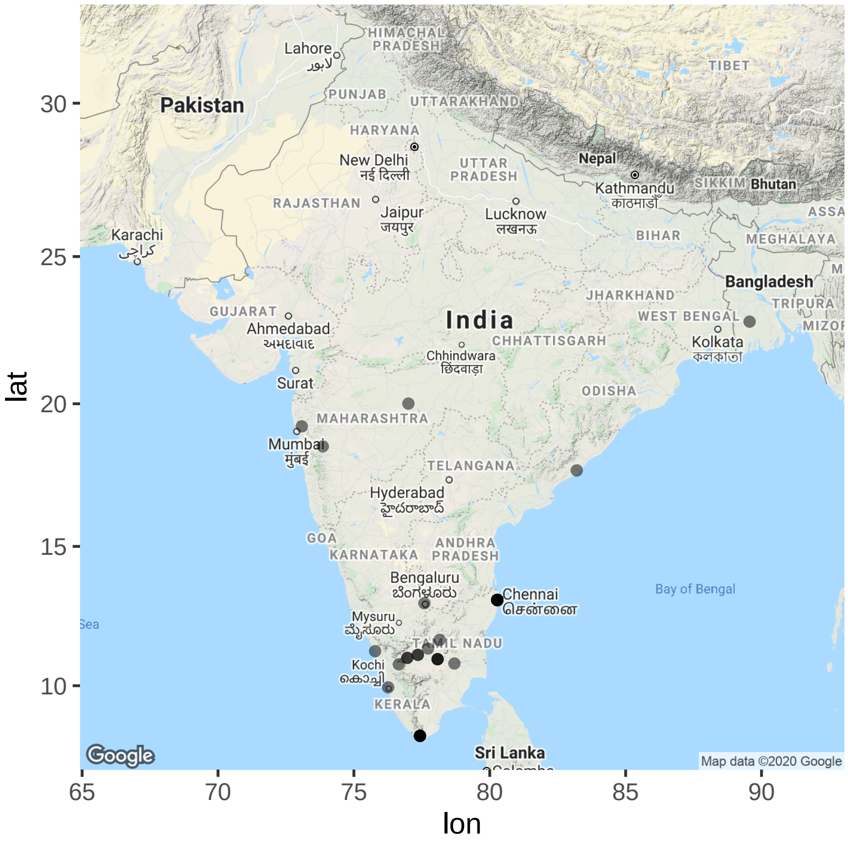
```
india_map<-get_map(location='india', zoom=5, maptype = "terrain",
             source='google',color='color')
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=india&zoom=5&size=640x640&scale=2&map
```

```
## Source : https://maps.googleapis.com/maps/api/geocode/json?address=india&key=xxx
```

```
ggmap(india_map) + geom_point(x=d_respondents$LocationLongitude, y = d_respondents$LocationLatitude, sho
```

```
## Warning: `show_guide` has been deprecated. Please use `show.legend` instead.
```



Randomization Check

```
#http://www.sthda.com/english/wiki/chi-square-goodness-of-fit-test-in-r

respondent_counts <- d_respondents[ , .(.N), keyby=Assignment_Group][,2]

respondent_counts_chisq_test <- chisq.test(respondent_counts, p=c(1/5, 1/5, 1/5, 1/5, 1/5))

respondent_counts_chisq_test
```

```
##
##  Chi-squared test for given probabilities
##
## data:  respondent_counts
## X-squared = 0.085714, df = 4, p-value = 0.9991
```

```
#p-value = 0.9991, which is greater than significance level of 0.05.
#We can conclude that the observed proportions are not significantly different from the expected propor
```

Covariate Balance Check

```
#let's consider adding age bins and education bins

d_respondents[ Age_Range == "18-24", age_bin := 1]
d_respondents[ Age_Range == "25-34", age_bin := 2]
d_respondents[ Age_Range == "35-44", age_bin := 3]
d_respondents[ Age_Range == "45-54", age_bin := 4]
```

```r
d_respondents[ Age_Range == "55-64", age_bin := 5]
d_respondents[ Age_Range == "Above 65", age_bin := 6]

d_respondents[ Education_Level == "Associate's degree", edu_bin := 1]
d_respondents[ Education_Level == "Bachelor's degree", edu_bin := 2]
d_respondents[ Education_Level == "High school", edu_bin := 3]
d_respondents[ Education_Level == "Master's degree and above", edu_bin := 4]
d_respondents[ Education_Level == "Some high school", edu_bin := 5]
d_respondents[ Education_Level == "Trade school", edu_bin := 6]

d_respondents[ Assignment_Group == "Control", assign_bin := 1]
d_respondents[ Assignment_Group == "Medical Feedback", assign_bin := 2]
d_respondents[ Assignment_Group == "Negative Images", assign_bin := 3]
d_respondents[ Assignment_Group == "Positive Images", assign_bin := 4]
d_respondents[ Assignment_Group == "Self-Reflect", assign_bin := 5]

d_respondents[ , US_Dummy := ifelse(country == "United States", 1, 0)]

d_respondents[ , Male_Dummy := ifelse(Gender == "Male", 1, 0)]

#add treatment dummy

d_respondents[ , Treatment_Dummy := ifelse(Assignment_Group != "Control", 1, 0)]

#head(d_respondents)
```

```r
d_respondents %>%
  group_by(Assignment_Group) %>%
  summarise(num_respondents = n(),
            pre_treatment_avg = mean(TaskPhase1_Score),
            taskphase2_avg = mean(TaskPhase2_Score),
            taskphase3_avg = mean(TaskPhase3_Score))
```

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

```
## # A tibble: 5 x 5
##   Assignment_Group num_respondents pre_treatment_a~ taskphase2_avg
##   <chr>                      <int>            <dbl>          <dbl>
## 1 Control                       69            0.607          0.461
## 2 Medical Feedback              70            0.634          0.523
## 3 Negative Images               72            0.578          0.494
## 4 Positive Images               70            0.614          0.514
## 5 Self-Reflect                  69            0.599          0.526
## # ... with 1 more variable: taskphase3_avg <dbl>
```

```r
d_respondents %>%
  group_by(Assignment_Group) %>%
  summarise(num_respondents = n(),
            avg_age_bin = mean(age_bin),
            avg_edu_bin = mean(edu_bin),
            male = mean(Male_Dummy),
            US = mean(US_Dummy))
```

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

```
## # A tibble: 5 x 6
##   Assignment_Group num_respondents avg_age_bin avg_edu_bin  male    US
##   <chr>                      <int>       <dbl>       <dbl> <dbl> <dbl>
## 1 Control                       69        2.68        2.61 0.609 0.652
## 2 Medical Feedback              70        2.63        2.47 0.586 0.529
## 3 Negative Images               72        2.62        2.58 0.583 0.625
## 4 Positive Images               70        2.86        2.6  0.586 0.714
## 5 Self-Reflect                  69        2.83        2.42 0.594 0.696
```

```
d_respondents %>%
  group_by(Assignment_Group) %>%
  summarise(num_respondents = n(),
            )
```

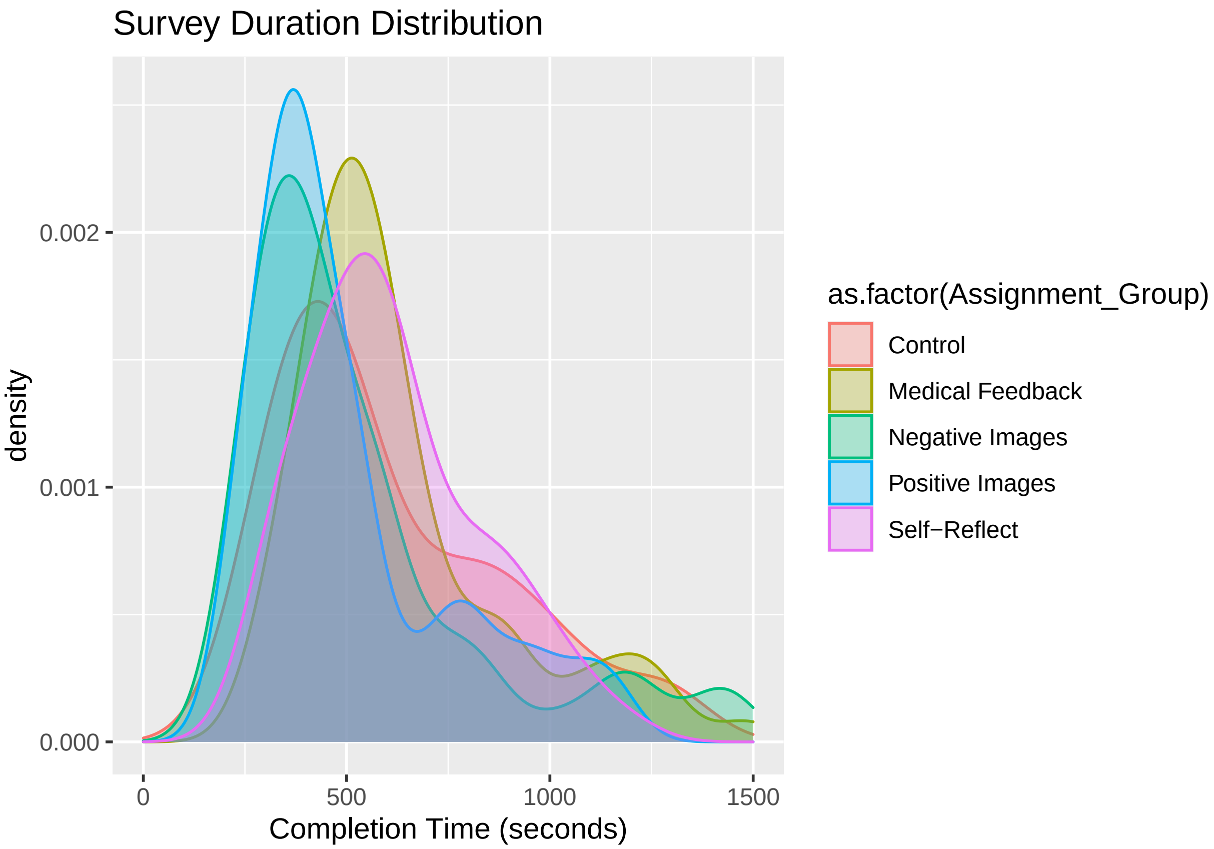
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 5 x 2
##   Assignment_Group num_respondents
##   <chr>                      <int>
## 1 Control                       69
## 2 Medical Feedback              70
## 3 Negative Images               72
## 4 Positive Images               70
## 5 Self-Reflect                  69
```

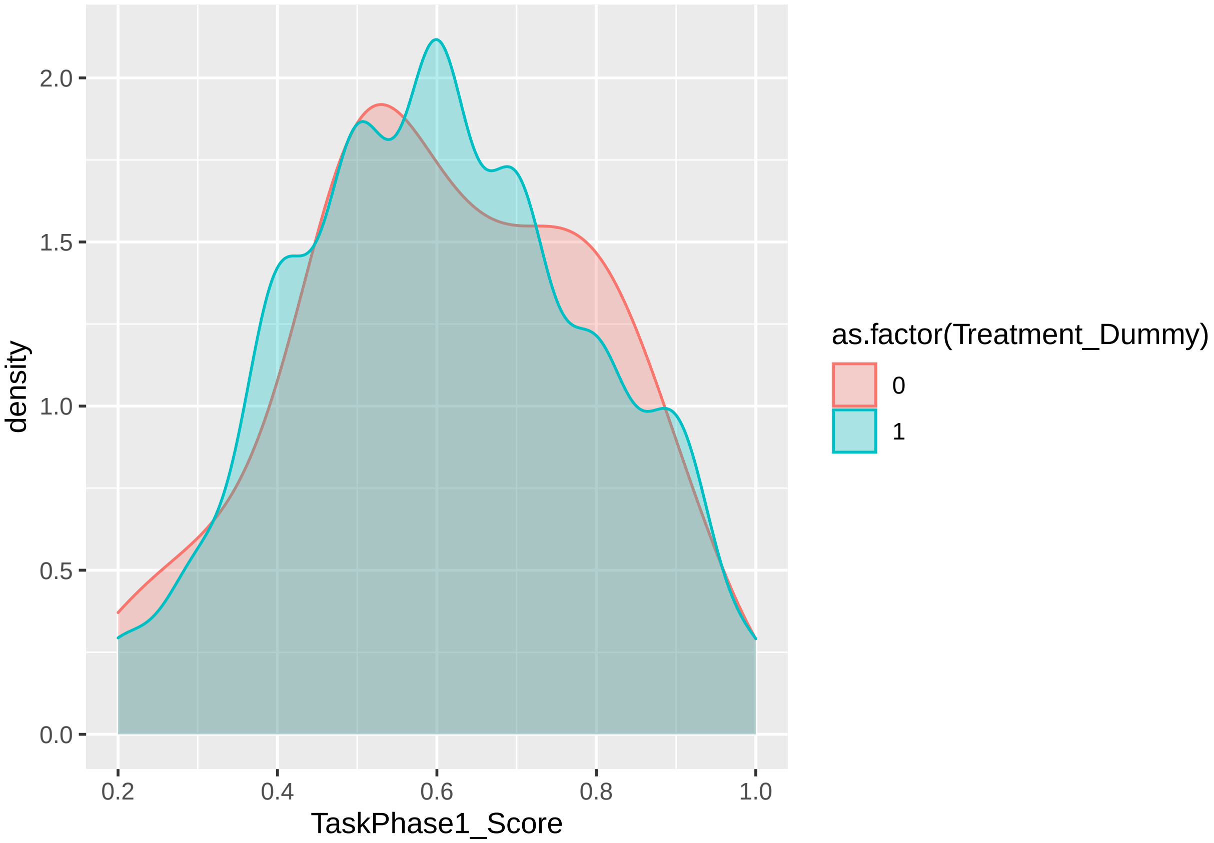
Visuals

```
#Density distribution of Survey Duration
ggplot(d_respondents, aes(x=Survey_Duration, colour=as.factor(Assignment_Group), fill = as.factor(Assig
  geom_density(alpha = 0.3) +
  xlim(0, 1500) +
  xlab("Completion Time (seconds)") +
  ggtitle("Survey Duration Distribution")
```

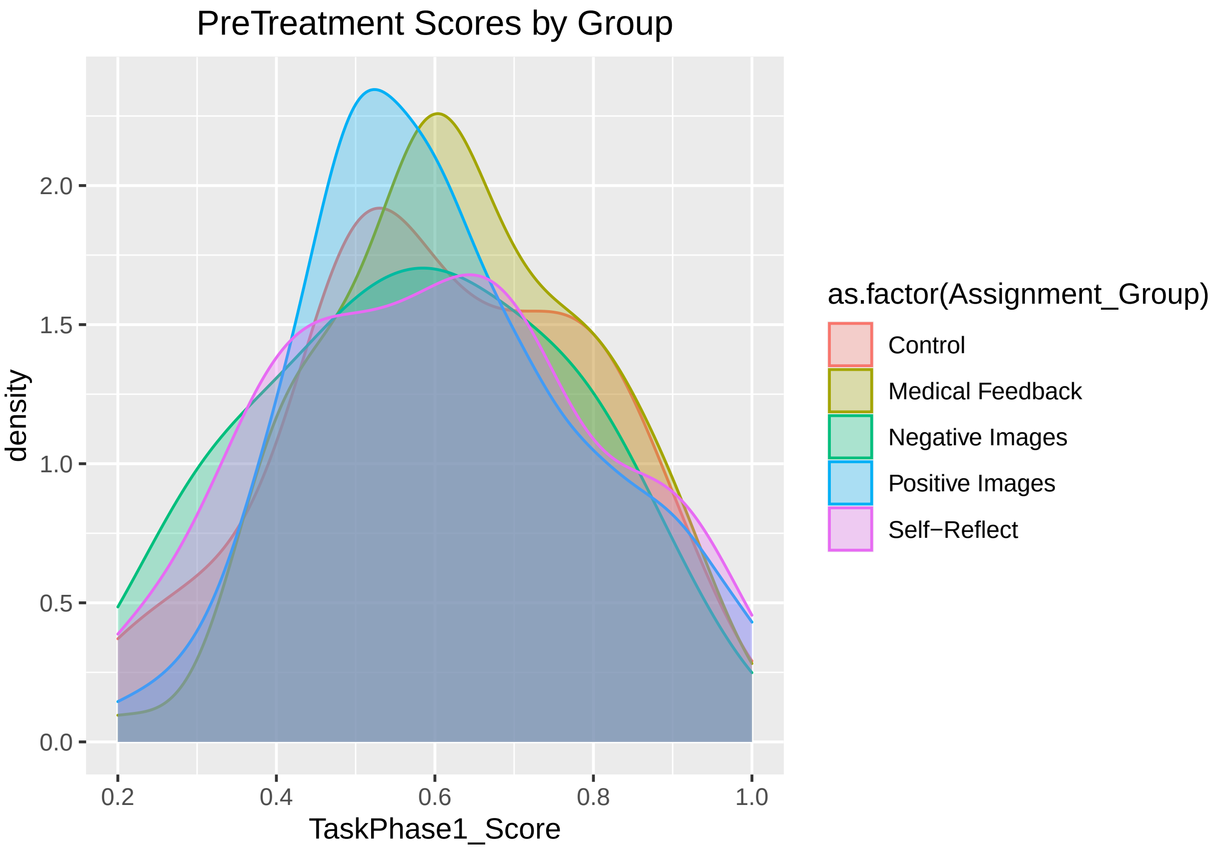
```
## Warning: Removed 6 rows containing non-finite values (stat_density).
```



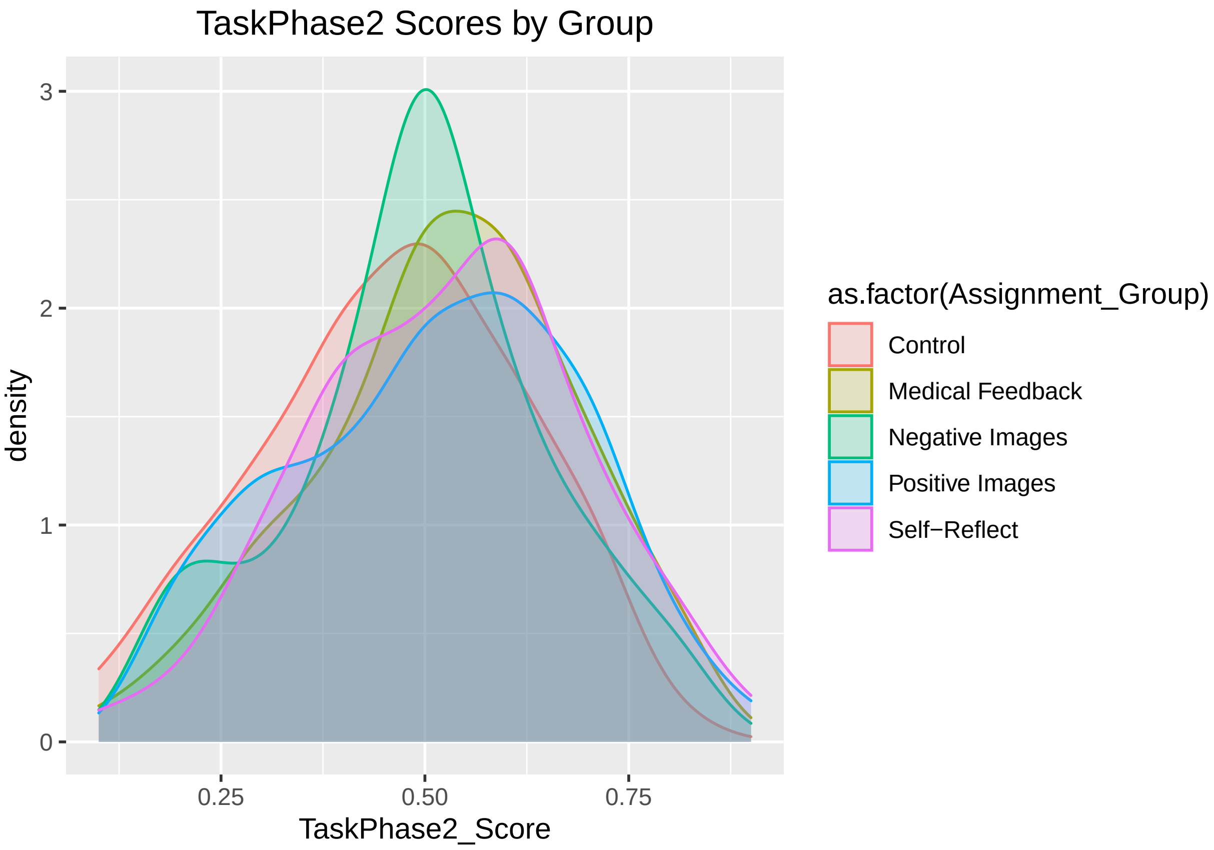
```
#Comparing pretreatment values
ggplot(d_respondents, aes(x=TaskPhase1_Score, fill = as.factor(Treatment_Dummy), colour=as.factor(Treatm
```



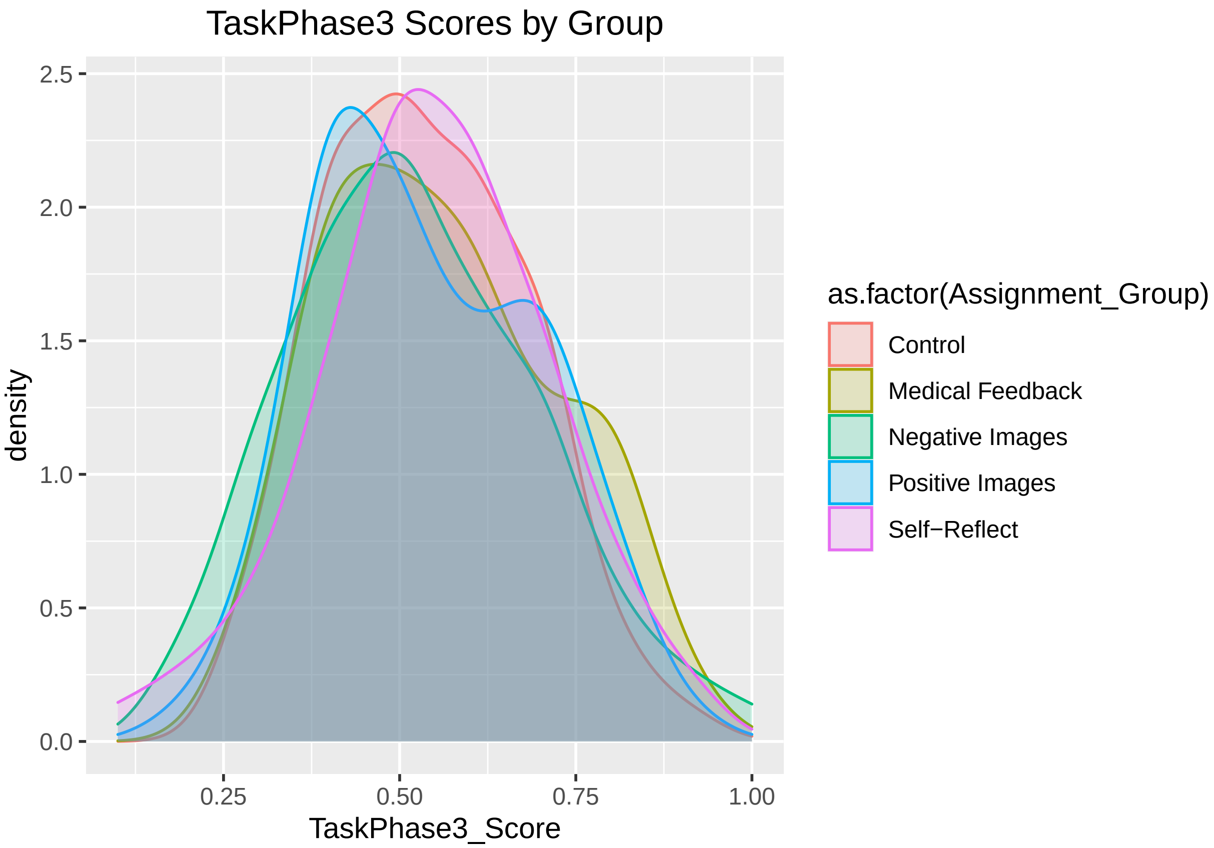
```
ggplot(d_respondents, aes(x=TaskPhase1_Score, fill = as.factor(Assignment_Group), colour=as.factor(Assig
```



```r
#Comparing taskphase2 values
ggplot(d_respondents, aes(x=TaskPhase2_Score, fill = as.factor(Assignment_Group), colour=as.factor(Assig
```

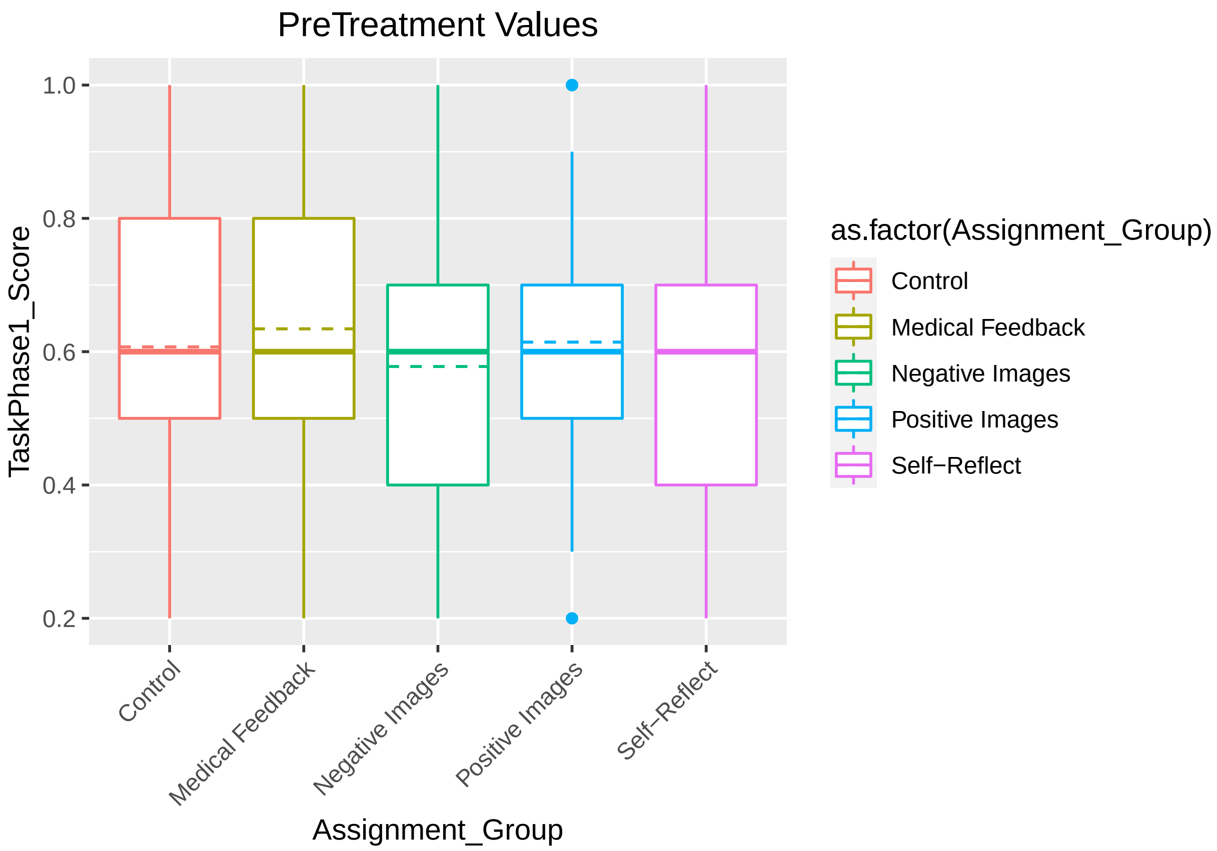


```
#Comparing taskphase3 values
ggplot(d_respondents, aes(x=TaskPhase3_Score, fill = as.factor(Assignment_Group), colour=as.factor(Assi
```



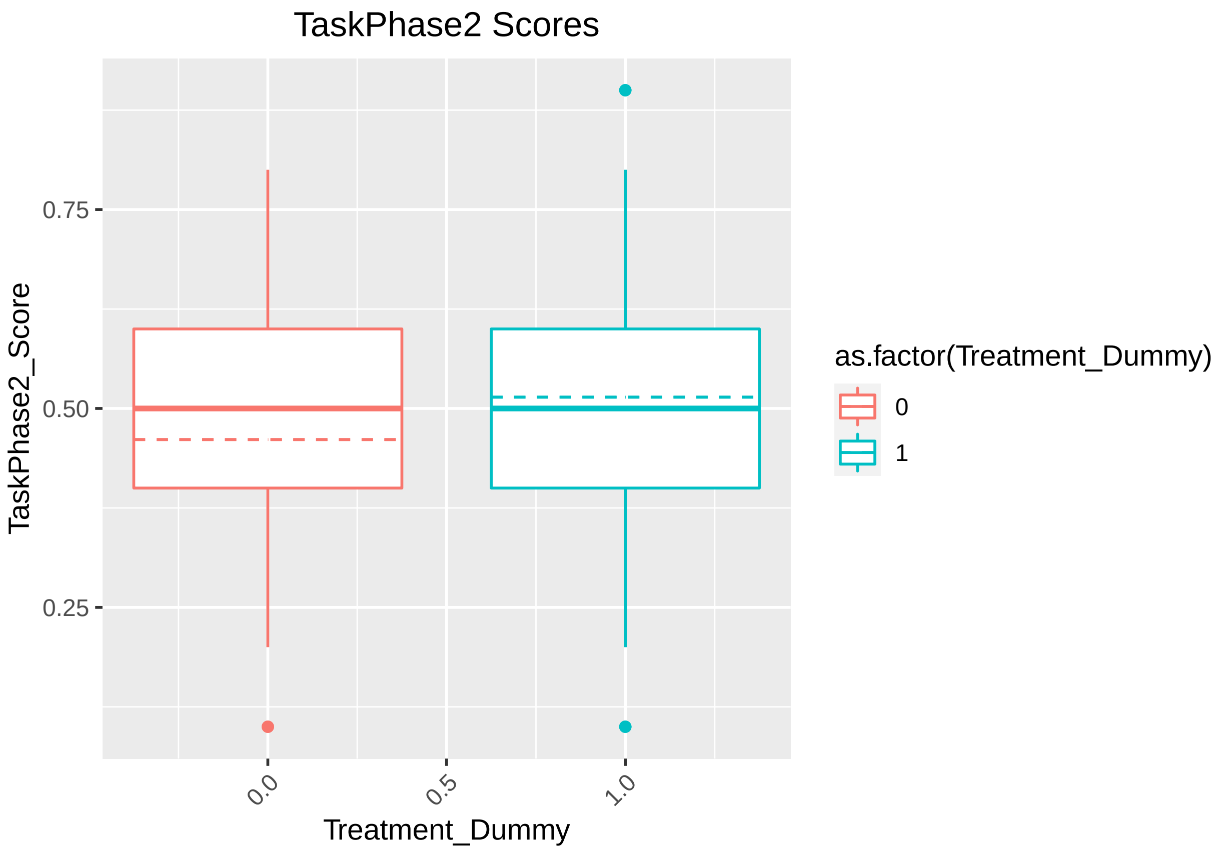
```
#boxplots for pretreatment values
ggplot(d_respondents, aes(x = Assignment_Group, y=TaskPhase1_Score, colour=as.factor(Assignment_Group))
                 width = .75, linetype = "dashed") + ggtitle("PreTreatment Values") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) + theme(plot.title = element_text(h
```

```
## Warning: `fun.y` is deprecated. Use `fun` instead.
```



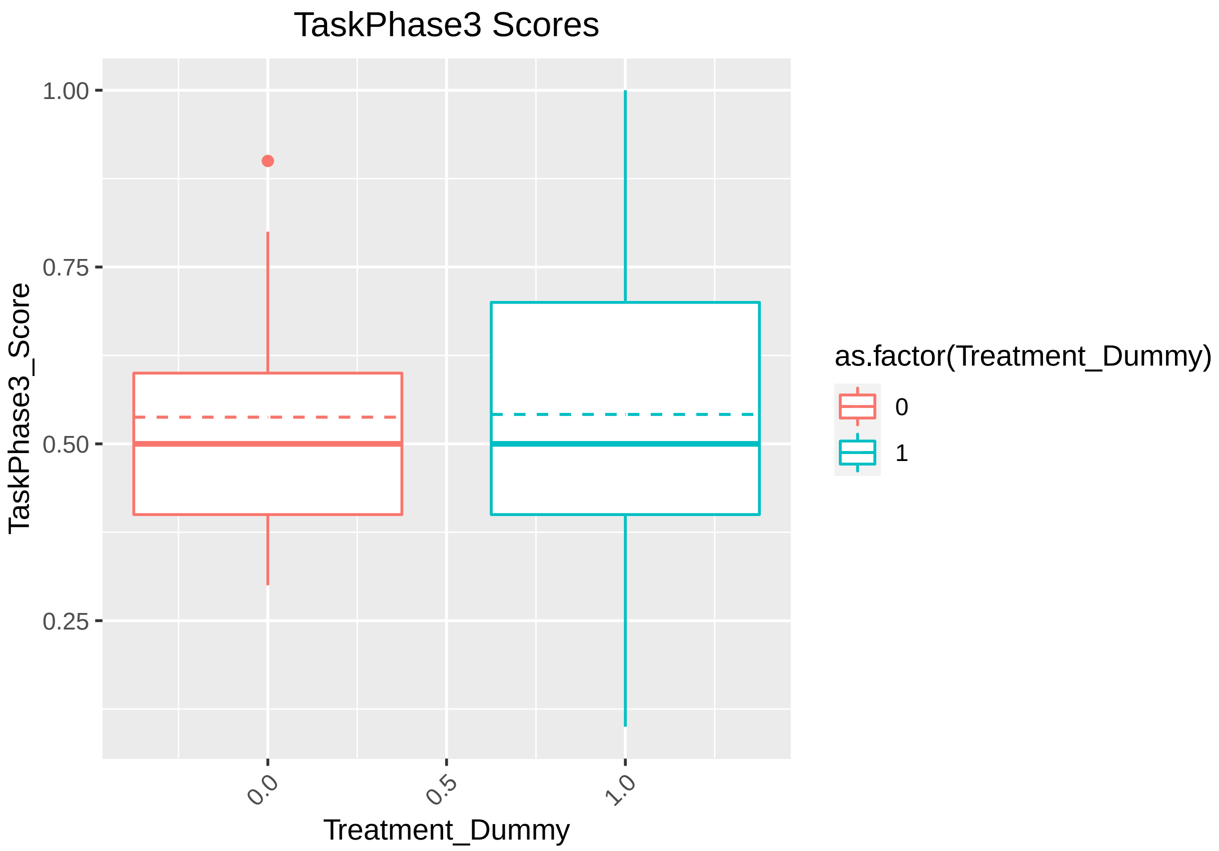
```
#boxplots for TaskPhase2 values
ggplot(d_respondents, aes(x = Treatment_Dummy, y=TaskPhase2_Score, colour=as.factor(Treatment_Dummy))) +
                 width = .75, linetype = "dashed") + ggtitle("TaskPhase2 Scores") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) + theme(plot.title = element_text(h
```

```
## Warning: `fun.y` is deprecated. Use `fun` instead.
```



```
#boxplots for TaskPhase3 values
ggplot(d_respondents, aes(x = Treatment_Dummy, y=TaskPhase3_Score, colour=as.factor(Treatment_Dummy))) +
                 width = .75, linetype = "dashed") + ggtitle("TaskPhase3 Scores") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) + theme(plot.title = element_text(h
```

```
## Warning: `fun.y` is deprecated. Use `fun` instead.
```



```
#check balance between age-range, education, age
d_respondents[ , table(Assignment_Group, Gender)]
```

```
##                     Gender
## Assignment_Group   Female Male
##   Control              27   42
##   Medical Feedback     29   41
##   Negative Images      30   42
##   Positive Images      29   41
##   Self-Reflect         28   41
```

```
cro(d_respondents$Assignment_Group, d_respondents$Gender)
```

d_respondents$Gender

Female

Male

d_respondents$Assignment_Group

Control

27

42

Medical Feedback

29

41

Negative Images

30

42

Positive Images

29

41

Self-Reflect

28

41

#Total cases

143

207

```
chisq.test(d_respondents[ , table(Assignment_Group, Gender)])
```

```
##
##  Pearson's Chi-squared test
##
## data:  d_respondents[, table(Assignment_Group, Gender)]
## X-squared = 0.12578, df = 4, p-value = 0.9981
```

```
(d_respondents[ , table(Assignment_Group, Age_Range)])
```

```
##                     Age_Range
## Assignment_Group   18-24 25-34 35-44 45-54 55-64 Above 65
##   Control              5    37    11     7     9        0
##   Medical Feedback     5    38    15     4     6        2
##   Negative Images      4    38    16     9     5        0
##   Positive Images      3    31    20     5    11        0
##   Self-Reflect         3    36    10    11     8        1
```

```
# expected frequency count for each cell of the contingency table should be at least 5. Since this is n
# https://stats.stackexchange.com/questions/81483/warning-in-r-chi-squared-approximation-may-be-incorre
chisq.test(d_respondents[ , table(Assignment_Group, Age_Range)],simulate.p.value = TRUE)
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  d_respondents[, table(Assignment_Group, Age_Range)]
## X-squared = 19.218, df = NA, p-value = 0.5177
```

```
(d_respondents[ , table(Assignment_Group, Education_Level)])
```

```
##                     Education_Level
## Assignment_Group   Associate's degree Bachelor's degree High school
##   Control                           3                44           1
##   Medical Feedback                  0                54           1
##   Negative Images                   2                50           3
##   Positive Images                   4                45           0
##   Self-Reflect                      4                46           7
##                     Education_Level
## Assignment_Group   Master's degree and above Some high school Trade school
##   Control                                 20                0            1
##   Medical Feedback                        14                0            1
##   Negative Images                         13                1            3
##   Positive Images                         19                0            2
##   Self-Reflect                            11                0            1
```

```
chisq.test(d_respondents[ , table(Assignment_Group, Education_Level)],simulate.p.value = TRUE)
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  d_respondents[, table(Assignment_Group, Education_Level)]
## X-squared = 28.7, df = NA, p-value = 0.07296
```

```
(d_respondents[ , table(Assignment_Group, US_Dummy)])
```

```
##                     US_Dummy
## Assignment_Group    0  1
##   Control          24 45
##   Medical Feedback 33 37
##   Negative Images  27 45
##   Positive Images  20 50
##   Self-Reflect     21 48
```

```
chisq.test(d_respondents[ , table(Assignment_Group, US_Dummy)],simulate.p.value = TRUE)
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  d_respondents[, table(Assignment_Group, US_Dummy)]
## X-squared = 6.5015, df = NA, p-value = 0.1694
```

```
# ATE of treatment on Total Score

d_respondents[ Treatment_Dummy == 1, mean(Total_Score)] - d_respondents[ Treatment_Dummy == 0, mean(Tota
```

```
## [1] 0.5143122
```

```
sd(d_respondents$Total_Score)
```

```
## [1] 3.743141
```

```
# ATE of treatment on TaskPhase2 Score

d_respondents[ Treatment_Dummy == 1, mean(TaskPhase2_Score)] - d_respondents[ Treatment_Dummy == 0, mea
```

```
## [1] 0.05336531
```

```
sd(d_respondents$TaskPhase2_Score)
```

```
## [1] 0.16451
```

```
#trying 2SLS...but dont think it applies here

# d_respondents[ , lm(Total_Score ~ Education_Level)]
# d_respondents[ , ivreg(Total_Score ~ Education_Level | Assignment_Group)]
```

```
power.t.test( delta = .05, sd=.16, sig.level = 0.05, power=0.8)
```

```
##
##      Two-sample t test power calculation
##
##               n = 161.711
##           delta = 0.05
##              sd = 0.16
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Analysis

Helper Functions

```
get_robust_se <- function(model){
  # Get robust SE for use in stargazer
  vcov <- vcovHC(model,type = "HC1")
  return(sqrt(diag(vcov)))
}
```

Task Phase 2 Analysis

```
# does any treatment have an effect on task phase 2 score?
mod_task2_a <- d_respondents[, lm(TaskPhase2_Score ~ Treatment_Dummy)]

mod_task2_b <- d_respondents[, lm(TaskPhase2_Score ~ Treatment_Dummy +
                                                     TaskPhase1_Score +
                                                     as.factor(Gender) +
                                                     as.factor(Education_Level) +
                                                     as.factor(Age_Range))]

stargazer(mod_task2_a,
          mod_task2_b,
          se = list(get_robust_se(mod_task2_a),get_robust_se(mod_task2_b)),
          omit = c("Education_Level","Age_Range"),
          add.lines = list(c('Education Fixed Effects', 'No','Yes'),
                           c('Age Fixed Effects','No','Yes')),
          type='text')
```

```
##
## ====================================================================
##                                       Dependent variable:
##                         ---------------------------------------------
##                                        TaskPhase2_Score
##                                  (1)                   (2)
## --------------------------------------------------------------------
## Treatment_Dummy                0.053**               0.051**
##                                (0.022)               (0.022)
```

```
##
## TaskPhase1_Score                                     0.240***
##                                                      (0.047)
##
## as.factor(Gender)Male                                 -0.010
##                                                      (0.017)
##
## Constant                      0.461***               0.281***
##                                (0.019)               (0.072)
##
## ---------------------------------------------------------------------
## Education Fixed Effects          No                     Yes
## Age Fixed Effects                No                     Yes
## Observations                     350                    350
## R2                              0.017                  0.117
## Adjusted R2                     0.014                  0.083
## Residual Std. Error       0.163 (df = 348)       0.158 (df = 336)
## F Statistic             5.911** (df = 1; 348) 3.433*** (df = 13; 336)
## =====================================================================
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

```r
#add an F test to compare
anova(mod_task2_a, mod_task2_b, test='F')
```

```
## Analysis of Variance Table
##
## Model 1: TaskPhase2_Score ~ Treatment_Dummy
## Model 2: TaskPhase2_Score ~ Treatment_Dummy + TaskPhase1_Score + as.factor(Gender) +
##     as.factor(Education_Level) + as.factor(Age_Range)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    348 9.2874
## 2    336 8.3377 12   0.94975 3.1895 0.0002426 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#does the specific treatment group have an effect on task phase 2 score?
mod_task2_c <- d_respondents[, lm(TaskPhase2_Score ~ as.factor(Assignment_Group))]

mod_task2_d <- d_respondents[, lm(TaskPhase2_Score ~ as.factor(Assignment_Group) +
                                                     TaskPhase1_Score +
                                                     as.factor(Gender) +
                                                     as.factor(Education_Level) +
                                                     as.factor(Age_Range))]

# Do you think that there are features of the data that might systematically predict that people will r
# TODO update this heterogeneity issue. I'm not quite sure this applies because they're both considered
# mod5 <- d_respondents[, lm(TaskPhase2_Score ~ Treatment_Dummy + as.factor(assign_bin) +
#                              Treatment_Dummy * as.factor(assign_bin))]
stargazer(mod_task2_c,
          mod_task2_d,
          se = list(get_robust_se(mod_task2_c),get_robust_se(mod_task2_d)),
          omit = c("Education_Level","Age_Range"),
          add.lines = list(c('Education Fixed Effects', 'No','Yes'),
                           c('Age Fixed Effects','No','Yes')),
          type='text')
```

```
##
## =======================================================================================
##                                                          Dependent variable:
##                                             -------------------------------------------
##                                                           TaskPhase2_Score
##                                                     (1)                   (2)
## ---------------------------------------------------------------------------------------
## as.factor(Assignment_Group)Medical Feedback       0.062**               0.055*
##                                                   (0.027)               (0.029)
##
## as.factor(Assignment_Group)Negative Images         0.034                 0.039
##                                                   (0.027)               (0.027)
##
## as.factor(Assignment_Group)Positive Images        0.053*                0.050*
##                                                   (0.029)               (0.027)
##
## as.factor(Assignment_Group)Self-Reflect           0.065**               0.058**
##                                                   (0.028)               (0.029)
##
## TaskPhase1_Score                                                       0.238***
##                                                                         (0.048)
##
## as.factor(Gender)Male                                                   -0.010
##                                                                         (0.017)
##
## Constant                                         0.461***              0.282***
##                                                   (0.019)               (0.073)
##
## ---------------------------------------------------------------------------------------
## Education Fixed Effects                             No                    Yes
## Age Fixed Effects                                   No                    Yes
## Observations                                        350                   350
## R2                                                 0.021                 0.119
## Adjusted R2                                        0.010                 0.076
## Residual Std. Error                            0.164 (df = 345)      0.158 (df = 333)
## F Statistic                                   1.874 (df = 4; 345) 2.805*** (df = 16; 333)
## =======================================================================================
## Note:                                                       *p<0.1; **p<0.05; ***p<0.01
```

```r
anova(mod_task2_c, mod_task2_d, test='F')
```

```
## Analysis of Variance Table
##
## Model 1: TaskPhase2_Score ~ as.factor(Assignment_Group)
## Model 2: TaskPhase2_Score ~ as.factor(Assignment_Group) + TaskPhase1_Score +
##     as.factor(Gender) + as.factor(Education_Level) + as.factor(Age_Range)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    345 9.2443
## 2    333 8.3233 12   0.92104 3.0708 0.0003943 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Task Phase 3 Analysis

```
# test final task and any treatment
mod_task3_a <- d_respondents[, lm(TaskPhase3_Score ~ Treatment_Dummy)]
mod_task3_b <- d_respondents[, lm(TaskPhase3_Score ~ Treatment_Dummy +
                                                     TaskPhase1_Score +
                                                     as.factor(Gender) +
                                                     as.factor(Education_Level) +
                                                     as.factor(Age_Range))]

stargazer(mod_task3_a,
          mod_task3_b,
          se = list(get_robust_se(mod_task3_a),get_robust_se(mod_task3_b)),
          omit = c("Education_Level","Age_Range"),
          add.lines = list(c('Education Fixed Effects', 'No','Yes'),
                           c('Age Fixed Effects','No','Yes')),
          type='text')
```

```
##
## ===================================================================
##                                      Dependent variable:
##                          -------------------------------------------
##                                       TaskPhase3_Score
##                                 (1)                   (2)
## -------------------------------------------------------------------
## Treatment_Dummy                0.004                 0.002
##                               (0.019)               (0.019)
##
## TaskPhase1_Score                                    0.161***
##                                                     (0.047)
##
## as.factor(Gender)Male                               -0.004
##                                                     (0.017)
##
## Constant                     0.538***              0.515***
##                               (0.017)               (0.064)
##
## -------------------------------------------------------------------
## Education Fixed Effects         No                    Yes
## Age Fixed Effects               No                    Yes
## Observations                    350                   350
## R2                            0.0001                 0.084
## Adjusted R2                   -0.003                 0.049
## Residual Std. Error      0.160 (df = 348)      0.155 (df = 336)
## F Statistic             0.034 (df = 1; 348) 2.384*** (df = 13; 336)
## ===================================================================
## Note:                                   *p<0.1; **p<0.05; ***p<0.01
```

```
anova(mod_task3_a, mod_task3_b, test='F')
```

```
## Analysis of Variance Table
##
## Model 1: TaskPhase3_Score ~ Treatment_Dummy
## Model 2: TaskPhase3_Score ~ Treatment_Dummy + TaskPhase1_Score + as.factor(Gender) +
##     as.factor(Education_Level) + as.factor(Age_Range)
##   Res.Df    RSS Df Sum of Sq    F   Pr(>F)   
## 1    348 8.8649                              
## 2    336 8.1170 12   0.74791 2.58 0.002743 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# test final task and specific treatment
mod_task3_c <- d_respondents[, lm(TaskPhase3_Score ~ as.factor(Assignment_Group))]
mod_task3_d <- d_respondents[, lm(TaskPhase3_Score ~ as.factor(Assignment_Group) +
                                                     TaskPhase1_Score +
                                                     as.factor(Gender) +
                                                     as.factor(Education_Level) +
                                                     as.factor(Age_Range))]

stargazer(mod_task3_c,
          mod_task3_d,
          se = list(get_robust_se(mod_task3_c),get_robust_se(mod_task3_d)),
          omit = c("Education_Level","Age_Range"),
          add.lines = list(c('Education Fixed Effects', 'No','Yes'),
                           c('Age Fixed Effects','No','Yes')),
          type='text')
```

```
## 
## ======================================================================================
##                                                         Dependent variable:           
##                                              ------------------------------------------
##                                                           TaskPhase3_Score            
##                                                     (1)                  (2)          
## --------------------------------------------------------------------------------------
## as.factor(Assignment_Group)Medical Feedback        0.022                0.011         
##                                                   (0.026)              (0.026)        
##                                                                                       
## as.factor(Assignment_Group)Negative Images        -0.015                -0.011        
##                                                   (0.027)              (0.026)        
##                                                                                       
## as.factor(Assignment_Group)Positive Images        -0.001                0.004         
##                                                   (0.025)              (0.025)        
##                                                                                       
## as.factor(Assignment_Group)Self-Reflect            0.010                0.005         
##                                                   (0.026)              (0.026)        
##                                                                                       
## TaskPhase1_Score                                                       0.157***       
##                                                                        (0.047)        
##                                                                                       
## as.factor(Gender)Male                                                   -0.004        
##                                                                        (0.017)        
##                                                                                       
## Constant                                         0.538***              0.518***       
##                                                   (0.017)              (0.064)        
##                                                                                       
## --------------------------------------------------------------------------------------
## Education Fixed Effects                             No                   Yes          
## Age Fixed Effects                                   No                   Yes          
## Observations                                        350                  350          
## R2                                                 0.006                0.087         
## Adjusted R2                                       -0.005                0.043
## Residual Std. Error                          0.160 (df = 345)      0.156 (df = 333)
## F Statistic                                 0.545 (df = 4; 345) 1.971** (df = 16; 333)
## ======================================================================================
## Note:                                                      *p<0.1; **p<0.05; ***p<0.01
```

```
anova(mod_task3_c, mod_task3_d, test='F')
```

```
## Analysis of Variance Table
##
## Model 1: TaskPhase3_Score ~ as.factor(Assignment_Group)
## Model 2: TaskPhase3_Score ~ as.factor(Assignment_Group) + TaskPhase1_Score +
##     as.factor(Gender) + as.factor(Education_Level) + as.factor(Age_Range)
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1    345 8.8101
## 2    333 8.0988 12   0.71132 2.4373 0.004751 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wearing Off Effects

```
mod_task3_e <- d_respondents[ , lm(TaskPhase3_Score ~ TaskPhase2_Score)]
mod_task3_f <- d_respondents[ , lm(TaskPhase3_Score ~ TaskPhase2_Score + Treatment_Dummy)]
mod_task3_g <- d_respondents[ , lm(TaskPhase3_Score ~ TaskPhase2_Score + as.factor(Assignment_Group))]
mod_task3_h <- d_respondents[ , lm(TaskPhase3_Score ~ TaskPhase2_Score +
                                                      as.factor(Assignment_Group) +
                                                      as.factor(Gender) +
                                                      as.factor(Education_Level) +
                                                      as.factor(Age_Range))]


stargazer(mod_task3_e,
          mod_task3_f,
          mod_task3_g,
          mod_task3_h,
          se = list(get_robust_se(mod_task3_e),
                    get_robust_se(mod_task3_f),
                    get_robust_se(mod_task3_g),
                    get_robust_se(mod_task3_h)),
          type='text')
```

```
##
## ====================================================================================================
##                                                                                          Dependent va
##                                                     ------------------------------------------------
##                                                                                            TaskPhase3
##                                                               (1)                     (2)
## ----------------------------------------------------------------------------------------------------
## TaskPhase2_Score                                           0.239***                0.242***
##                                                             (0.050)                 (0.051)
##
## Treatment_Dummy                                                                     -0.009
##                                                                                     (0.019)
##
## as.factor(Assignment_Group)Medical Feedback
##
```

```
##
## as.factor(Assignment_Group)Negative Images
##
##
## as.factor(Assignment_Group)Positive Images
##
##
## as.factor(Assignment_Group)Self-Reflect
##
##
## as.factor(Gender)Male
##
##
## as.factor(Education_Level)Bachelor's degree
##
##
## as.factor(Education_Level)High school
##
##
## as.factor(Education_Level)Master's degree and above
##
##
## as.factor(Education_Level)Some high school
##
##
## as.factor(Education_Level)Trade school
##
##
## as.factor(Age_Range)25-34
##
##
## as.factor(Age_Range)35-44
##
##
## as.factor(Age_Range)45-54
##
##
## as.factor(Age_Range)55-64
##
##
## as.factor(Age_Range)Above 65
##
##
## Constant                                                   0.420***                0.426***
##                                                             (0.026)                 (0.028)
##
## ----------------------------------------------------------------------------------------------------
## Observations                                                  350                     350
## R2                                                           0.061                   0.061
## Adjusted R2                                                  0.058                   0.056
## Residual Std. Error                                    0.155 (df = 348)        0.155 (df = 347)
## F Statistic                                         22.539*** (df = 1; 348) 11.334*** (df = 2; 347) 4
## ====================================================================================================
## Note:
```

```r
anova(mod_task3_e, mod_task3_f, test='F')
```

```
## Analysis of Variance Table
##
## Model 1: TaskPhase3_Score ~ TaskPhase2_Score
## Model 2: TaskPhase3_Score ~ TaskPhase2_Score + Treatment_Dummy
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    348 8.3265
## 2    347 8.3221  1 0.0043584 0.1817 0.6702
```

```r
anova(mod_task3_g, mod_task3_h, test='F')
```

```
## Analysis of Variance Table
##
## Model 1: TaskPhase3_Score ~ TaskPhase2_Score + as.factor(Assignment_Group)
## Model 2: TaskPhase3_Score ~ TaskPhase2_Score + as.factor(Assignment_Group) +
##     as.factor(Gender) + as.factor(Education_Level) + as.factor(Age_Range)
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    344 8.2858
## 2    333 7.8631 11    0.4227 1.6274 0.08955 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
##
mod_task3_i <- d_respondents[ , lm(TaskPhase3_Score ~ TaskPhase1_Score + TaskPhase2_Score)]
mod_task3_j <- d_respondents[ , lm(TaskPhase3_Score ~ TaskPhase1_Score + TaskPhase2_Score + Treatment_Du
mod_task3_k <- d_respondents[ , lm(TaskPhase3_Score ~ TaskPhase1_Score + TaskPhase2_Score + as.factor(A
mod_task3_l <- d_respondents[, lm(TaskPhase3_Score ~ TaskPhase1_Score + TaskPhase2_Score +
                                                      as.factor(Assignment_Group) +
                                                      as.factor(Gender) +
                                                      as.factor(Education_Level) +
                                                      as.factor(Age_Range))]

stargazer(mod_task3_i,
          mod_task3_j,
          mod_task3_k,
          mod_task3_l,
          se = list(get_robust_se(mod_task3_i),
                    get_robust_se(mod_task3_j),
                    get_robust_se(mod_task3_k),
                    get_robust_se(mod_task3_l)),
          type='text')
```

```
##
## ====================================================================================================
##                                                                                          Dependent va
##                                                      ------------------------------------------------
##                                                                                           TaskPhase3_
##                                                                (1)                    (2)
## ----------------------------------------------------------------------------------------------------
## TaskPhase1_Score                                             0.113**                0.113**
##                                                              (0.046)                (0.046)
##
## TaskPhase2_Score                                            0.202***               0.204***
##                                                              (0.053)                (0.054)
```

```
##
## Treatment_Dummy                                                                     -0.007
##                                                                                    (0.019)
##
## as.factor(Assignment_Group)Medical Feedback
##
##
## as.factor(Assignment_Group)Negative Images
##
##
## as.factor(Assignment_Group)Positive Images
##
##
## as.factor(Assignment_Group)Self-Reflect
##
##
## as.factor(Gender)Male
##
##
## as.factor(Education_Level)Bachelor's degree
##
##
## as.factor(Education_Level)High school
##
##
## as.factor(Education_Level)Master's degree and above
##
##
## as.factor(Education_Level)Some high school
##
##
## as.factor(Education_Level)Trade school
##
##
## as.factor(Age_Range)25-34
##
##
## as.factor(Age_Range)35-44
##
##
## as.factor(Age_Range)45-54
##
##
## as.factor(Age_Range)55-64
##
##
## as.factor(Age_Range)Above 65
##
##
## Constant                                                   0.370***                0.375***
##                                                             (0.032)                (0.034)
##
## ----------------------------------------------------------------------------------------------------
## Observations                                                  350                    350
## R2                                                           0.077                  0.078
## Adjusted R2                                                  0.072                  0.070
## Residual Std. Error                                    0.154 (df = 347)        0.154 (df = 346)
## F Statistic                                         14.523*** (df = 2; 347) 9.693*** (df = 3; 346) 5
## ====================================================================================================
## Note:
```

```
anova(mod_task3_i, mod_task3_j, test = 'F')
```

```
## Analysis of Variance Table
##
## Model 1: TaskPhase3_Score ~ TaskPhase1_Score + TaskPhase2_Score
## Model 2: TaskPhase3_Score ~ TaskPhase1_Score + TaskPhase2_Score + Treatment_Dummy
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    347 8.1810
## 2    346 8.1784  1 0.0025307 0.1071 0.7437
```

```
anova(mod_task3_k, mod_task3_l, test = 'F')
```

```
## Analysis of Variance Table
##
## Model 1: TaskPhase3_Score ~ TaskPhase1_Score + TaskPhase2_Score + as.factor(Assignment_Group)
## Model 2: TaskPhase3_Score ~ TaskPhase1_Score + TaskPhase2_Score + as.factor(Assignment_Group) +
##     as.factor(Gender) + as.factor(Education_Level) + as.factor(Age_Range)
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    343 8.1514
## 2    332 7.7386 11   0.41276 1.6098 0.09435 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# lm(TaskPhase3_Score ~ TaskPhase2_Score) vs lm(TaskPhase3_Score ~ TaskPhase1_Score + TaskPhase2_Score)
anova(mod_task3_e, mod_task3_i, test = 'F')
```

```
## Analysis of Variance Table
##
## Model 1: TaskPhase3_Score ~ TaskPhase2_Score
## Model 2: TaskPhase3_Score ~ TaskPhase1_Score + TaskPhase2_Score
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    348 8.3265
## 2    347 8.1810  1   0.14552 6.1721 0.01345 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Playground

```
# compare self-reflect against medical feedback groups?
#make dummies
d_respondents[ , Self_Reflect_Dummy := ifelse(Assignment_Group == "Self-Reflect", 1, 0)]
d_respondents[ , Med_Feedback_Dummy := ifelse(Assignment_Group == "Medical Feedback", 1, 0)]

mod_test_dummies1 <- d_respondents[ , lm(TaskPhase2_Score ~ Treatment_Dummy + Self_Reflect_Dummy)]
mod_test_dummies2 <- d_respondents[ , lm(TaskPhase2_Score ~ Treatment_Dummy + Med_Feedback_Dummy)]

stargazer(mod_test_dummies1,
          mod_test_dummies2,
          se = list(get_robust_se(mod_test_dummies1),
                    get_robust_se(mod_test_dummies2)),
          type = 'text')
```

```
##
## ===========================================================
##                                   Dependent variable:
##                               ---------------------------
##                                     TaskPhase2_Score
##                                    (1)           (2)
## -----------------------------------------------------------
## Treatment_Dummy                   0.050**       0.051**
##                                   (0.022)       (0.023)
##
## Self_Reflect_Dummy                 0.016
##                                   (0.023)
##
## Med_Feedback_Dummy                               0.011
##                                                 (0.022)
##
## Constant                          0.461***      0.461***
##                                   (0.019)       (0.019)
##
## -----------------------------------------------------------
## Observations                        350           350
## R2                                 0.018         0.017
## Adjusted R2                        0.012         0.012
## Residual Std. Error (df = 347)     0.163         0.164
## F Statistic (df = 2; 347)         3.192**       3.079**
## ===========================================================
## Note:                           *p<0.1; **p<0.05; ***p<0.01
```

Playground 2

```
### linear model playground
d_test <- d_respondents[,c("Assignment_Group","TaskPhase1_Score","TaskPhase2_Score","TaskPhase3_Score",

#does treatment have an effect on total score?
mod_test1 <- d_test[ , lm(TaskPhase2_Score ~ TaskPhase1_Score + Treatment_Dummy)]


mod_test2 <- d_test[, lm(TaskPhase2_Score ~ TaskPhase1_Score + Treatment_Dummy + (TaskPhase1_Score * Tre

#does treatment and pretreatment score have an effect on total score?


###
# seems that if i ad in TaskPhase1 to the linear model, the RSEs disappear...
mod_test3 <- d_test[, lm(TaskPhase2_Score ~  Treatment_Dummy  +
                                  TaskPhase1_Score +
                                  as.factor(Education_Level) +
                                  as.factor(Gender) +
                                  as.factor(Age_Range)
                         )]
```

```r
coeftest(mod_test3, vcov = vcovHC(mod_test3,"HC1"))
```

```
## 
## t test of coefficients:
## 
##                                                       Estimate Std. Error
## (Intercept)                                          0.2810773  0.0721331
## Treatment_Dummy                                      0.0507118  0.0221728
## TaskPhase1_Score                                     0.2402704  0.0468151
## as.factor(Education_Level)Bachelor's degree         -0.0068270  0.0485557
## as.factor(Education_Level)High school                0.0406800  0.0561886
## as.factor(Education_Level)Master's degree and above -0.0169778  0.0512819
## as.factor(Education_Level)Some high school          -0.1206457  0.0510769
## as.factor(Education_Level)Trade school               0.0286671  0.0692626
## as.factor(Gender)Male                               -0.0099516  0.0173475
## as.factor(Age_Range)25-34                            0.0446944  0.0376821
## as.factor(Age_Range)35-44                            0.0419834  0.0395245
## as.factor(Age_Range)45-54                            0.0697483  0.0417760
## as.factor(Age_Range)55-64                            0.0803500  0.0425233
## as.factor(Age_Range)Above 65                         0.1257451  0.0517169
##                                                     t value  Pr(>|t|)    
## (Intercept)                                          3.8966 0.0001177 ***
## Treatment_Dummy                                      2.2871 0.0228105 *  
## TaskPhase1_Score                                     5.1323 4.852e-07 ***
## as.factor(Education_Level)Bachelor's degree         -0.1406 0.8882698    
## as.factor(Education_Level)High school                0.7240 0.4695746    
## as.factor(Education_Level)Master's degree and above -0.3311 0.7407994    
## as.factor(Education_Level)Some high school          -2.3620 0.0187444 *  
## as.factor(Education_Level)Trade school               0.4139 0.6792187    
## as.factor(Gender)Male                               -0.5737 0.5665811    
## as.factor(Age_Range)25-34                            1.1861 0.2364245    
## as.factor(Age_Range)35-44                            1.0622 0.2889014    
## as.factor(Age_Range)45-54                            1.6696 0.0959341 .  
## as.factor(Age_Range)55-64                            1.8896 0.0596782 .  
## as.factor(Age_Range)Above 65                         2.4314 0.0155623 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(d_respondents$TaskPhase1_Score)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.2000  0.5000  0.6000  0.6063  0.7000  1.0000
```

```r
stargazer(mod_test1,
          mod_test2,
          mod_test3,
          se = list(get_robust_se(mod_test1),get_robust_se(mod_test2), get_robust_se(mod_test3)),
          type='text')
```

```
## 
## ====================================================================================================
##                                                                                Dependent variable:
##                                                      ------------------------------------------------
##                                                                                 TaskPhase2_Score
##                                                               (1)                     (2)
## ----------------------------------------------------------------------------------------------------
## TaskPhase1_Score                                           0.249***                  0.153
##                                                             (0.044)                 (0.095)
##
## as.factor(Education_Level)Bachelor's degree
##
##
## as.factor(Education_Level)High school
##
##
## as.factor(Education_Level)Master's degree and above
##
##
## as.factor(Education_Level)Some high school
##
##
## as.factor(Education_Level)Trade school
##
##
## as.factor(Gender)Male
##
##
## as.factor(Age_Range)25-34
##
##
## as.factor(Age_Range)35-44
##
##
## as.factor(Age_Range)45-54
##
##
## as.factor(Age_Range)55-64
##
##
## as.factor(Age_Range)Above 65
##
##
## Treatment_Dummy                                             0.054**                 -0.019
##                                                             (0.021)                 (0.065)
##
## TaskPhase1_Score:Treatment_Dummy                                                     0.120
##                                                                                     (0.107)
##
## Constant                                                   0.310***                0.368***
##                                                             (0.032)                 (0.057)
##
## ----------------------------------------------------------------------------------------------------
## Observations                                                 350                     350
## R2                                                          0.098                   0.101
## Adjusted R2                                                 0.092                   0.093
## Residual Std. Error                                    0.157 (df = 347)        0.157 (df = 346)
## F Statistic                                         18.782*** (df = 2; 347) 12.917*** (df = 3; 346) 3
## ====================================================================================================
## Note:                                                                                            *p<0
```

```r
mod_test4 <- d_test[ , lm(TaskPhase3_Score ~ TaskPhase2_Score)]
coeftest(mod_test4)
```

```
## 
## t test of coefficients:
## 
##                  Estimate Std. Error t value  Pr(>|t|)    
## (Intercept)      0.420497   0.026667 15.7686 < 2.2e-16 ***
## TaskPhase2_Score 0.238946   0.050331  4.7475 3.014e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# use Robust SE
mod_test2 <- d_respondents[, lm(TaskPhase2_Score ~ Treatment_Dummy  + as.factor(Education_Level) + (Trea
mod_test2$vcovHC_ <- vcovHC(mod_test2)
coeftest(mod_test2, vcov = mod_test2$vcovHC_)
```

```
## 
## t test of coefficients:
## 
##                                                                        Estimate
## (Intercept)                                                           0.5333333
## Treatment_Dummy                                                       0.0066667
## as.factor(Education_Level)Bachelor's degree                          -0.0742424
## as.factor(Education_Level)High school                                -0.0333333
## as.factor(Education_Level)Master's degree and above                  -0.0733333
## as.factor(Education_Level)Some high school                           -0.1400000
## as.factor(Education_Level)Trade school                               -0.2333333
## Treatment_Dummy:as.factor(Education_Level)Bachelor's degree           0.0414219
## Treatment_Dummy:as.factor(Education_Level)High school                 0.0751515
## Treatment_Dummy:as.factor(Education_Level)Master's degree and above  0.0438596
## Treatment_Dummy:as.factor(Education_Level)Trade school                0.3076190
##                                                                      Std. Error
## (Intercept)                                                                  NA
## Treatment_Dummy                                                              NA
## as.factor(Education_Level)Bachelor's degree                                  NA
## as.factor(Education_Level)High school                                        NA
## as.factor(Education_Level)Master's degree and above                          NA
## as.factor(Education_Level)Some high school                                   NA
## as.factor(Education_Level)Trade school                                       NA
## Treatment_Dummy:as.factor(Education_Level)Bachelor's degree                  NA
## Treatment_Dummy:as.factor(Education_Level)High school                        NA
## Treatment_Dummy:as.factor(Education_Level)Master's degree and above          NA
## Treatment_Dummy:as.factor(Education_Level)Trade school                       NA
##                                                                      t value
## (Intercept)                                                               NA
## Treatment_Dummy                                                           NA
## as.factor(Education_Level)Bachelor's degree                               NA
## as.factor(Education_Level)High school                                     NA
## as.factor(Education_Level)Master's degree and above                       NA
## as.factor(Education_Level)Some high school                                NA
## as.factor(Education_Level)Trade school                                    NA
## Treatment_Dummy:as.factor(Education_Level)Bachelor's degree               NA
## Treatment_Dummy:as.factor(Education_Level)High school                     NA
## Treatment_Dummy:as.factor(Education_Level)Master's degree and above       NA
## Treatment_Dummy:as.factor(Education_Level)Trade school                   NA
##                                                                     Pr(>|t|)
## (Intercept)                                                               NA
## Treatment_Dummy                                                           NA
## as.factor(Education_Level)Bachelor's degree                               NA
## as.factor(Education_Level)High school                                     NA
## as.factor(Education_Level)Master's degree and above                       NA
## as.factor(Education_Level)Some high school                                NA
## as.factor(Education_Level)Trade school                                    NA
## Treatment_Dummy:as.factor(Education_Level)Bachelor's degree               NA
## Treatment_Dummy:as.factor(Education_Level)High school                     NA
## Treatment_Dummy:as.factor(Education_Level)Master's degree and above       NA
## Treatment_Dummy:as.factor(Education_Level)Trade school                    NA
```