

Beauty Ad Causal Experiment

Lisa Minas, Ashton Chevallier & Olivier Zimmer

```
library(dplyr)

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

library(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

library(ggplot2)
library(car)

## Warning: package 'car' was built under R version 3.4.1
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##   recode
```

Load & Clean data

```
setwd('/Users/ozimmer/GoogleDrive/berkeley/w241/BeautyAd_CausalExperiment/Data')
d <- read.csv('BeautyAds_July 19, 2017_22.18.csv')

# Filter out irrelevant entries
d <- d[-c(1,2),] #Remove the first 2 lines
d <- d[d$Status == 'IP Address',] #Remove survey preview
d <- d[d$Welcome == 'I agree',] #Remove users who didn't agree to participate
d <- d[d$Finished == 'True',] #Remove users who didn't finish the survey
```

```

d <- d[d$Group %in% c('Treatment', 'Control'),]
d <- d[d$AudioCheck == 'Pineapple', ] #Should capture people after the pineapple test

length(unique(d$IPAddress))

## [1] 375

# Recoding of values
recode_values <- function(d, column){
  d[[column]] <- as.character(d[[column]])
  d[[column]] <- dplyr::recode(d[[column]], 'Strongly disagree' = -2,
                                'Disagree' = -1, 'Agree' = 1, 'Strongly agree' = 2,
                                .missing = 0, .default = 0)
}

columns_to_analyse <- c('Personal_Views_Confident', 'Personal_Views_Beautiful',
                       'Personal_Views_Beauty_Importance', 'Personal_Views_Relate_To_Model')
for (column in columns_to_analyse){
  d[[column]] <- recode_values(d, column)
}

# Correct column names misspellings
d <- dplyr::rename(d, Coffee_validate_1 = Coffe_validate_1,
                  Fit_i_identify_1 = FIit_i_identify_1,
                  Work_i_identify_2 = work_i_identify_2)

# Combine and recode randomization 1 & 2 for the images
images <- c('Passion', 'Coffee', 'Couple', 'Work', 'Fit')
questions <- c('_i_identify_', '_i_prefer_', '_o_prefer_', '_validate_')
randomization <- c('1', '2')

for (image in images){
  for (question in questions){
    column1 <- paste(image, question, '1', sep = "")
    d[[column1]] <- ifelse(d[[column1]] == 'Ad 1', 2, ifelse(d[[column1]] == 'Ad 2', 1, 0))
    column2 <- paste(image, question, '2', sep = "")
    d[[column2]] <- ifelse(d[[column2]] == 'Ad 2', 2, ifelse(d[[column2]] == 'Ad 1', 1, 0))
    new_column <- paste(image, question, sep = "")
    d[[new_column]] <- d[[column1]] + d[[column2]] - 1
    d[[new_column]] <- ifelse(d[[new_column]] == -1, NA, d[[new_column]])
  }
}

#Removing the NAs

#summary(d)

```

ATE for text questions

```

columns_to_analyse <- c('Personal_Views_Confident', 'Personal_Views_Beautiful',
                       'Personal_Views_Beauty_Importance', 'Personal_Views_Relate_To_Model')
for (column in columns_to_analyse){
  l1 <- lm(d[[column]] ~ d[['Group']])

```

```
print(column)
print(coef(summary(l1))[2,])
}
```

```
## [1] "Personal_Views_Confident"
##      Estimate Std. Error    t value    Pr(>|t|)
## 0.02564103 0.09016216 0.28438788 0.77626486
## [1] "Personal_Views_Beautiful"
##      Estimate Std. Error    t value    Pr(>|t|)
## 0.005128205 0.095368810 0.053772351 0.957144214
## [1] "Personal_Views_Beauty_Importance"
##      Estimate Std. Error    t value    Pr(>|t|)
## -0.05128205 0.10227822 -0.50139759 0.61637569
## [1] "Personal_Views_Relate_To_Model"
##      Estimate Std. Error    t value    Pr(>|t|)
## 0.24102564 0.11254984 2.14150141 0.03285591
```

Only Personal views relate to model has significant results.

Adding covariates

```
#str(d)
columns_to_analyse <- c('Personal_Views_Confident', 'Personal_Views_Beautiful',
                        'Personal_Views_Beauty_Importance', 'Personal_Views_Relate_To_Model')
for (column in columns_to_analyse){
  #l1 <- lm(d[[column]] ~ d[['Group']] + d$Age + d$Gender + d$Race + d$Location)
  l1 <- lm(d[[column]] ~ d[['Group']] + d$Gender)
  print(column)
  print(summary(l1))
}
```

```
## [1] "Personal_Views_Confident"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]] + d$Gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2938 -0.2938 -0.1532  0.7228  0.8468
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.15319    0.07865   1.948  0.0521 .
## d[["Group"]]Treatment 0.01655    0.09010   0.184  0.8544
## d$GenderMale      0.12404    0.09018   1.375  0.1698
## d$GenderOther    -1.15319    0.89174  -1.293  0.1967
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8883 on 386 degrees of freedom
## Multiple R-squared:  0.009848, Adjusted R-squared:  0.002152
## F-statistic: 1.28 on 3 and 386 DF, p-value: 0.281
##
```

```

## [1] "Personal_Views_Beautiful"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]] + d$Gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3016 -1.0832  0.6984  0.9157  0.9168
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.301615   0.082447   3.658 0.000289 ***
## d[["Group"]]Treatment -0.001103   0.094453  -0.012 0.990689
## d$GenderMale      -0.217306   0.094543  -2.298 0.022068 *
## d$GenderOther     -2.301615   0.934846  -2.462 0.014252 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9312 on 386 degrees of freedom
## Multiple R-squared:  0.0272, Adjusted R-squared:  0.01964
## F-statistic: 3.597 on 3 and 386 DF, p-value: 0.01373
##
## [1] "Personal_Views_Beauty_Importance"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]] + d$Gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.20039 -1.01983  0.03811  0.85755  1.03811
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.20039   0.08878   2.257  0.0246 *
## d[["Group"]]Treatment -0.05794   0.10171  -0.570  0.5693
## d$GenderMale      -0.18056   0.10181  -1.773  0.0769 .
## d$GenderOther     -2.20039   1.00670  -2.186  0.0294 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.003 on 386 degrees of freedom
## Multiple R-squared:  0.0198, Adjusted R-squared:  0.01218
## F-statistic: 2.599 on 3 and 386 DF, p-value: 0.05194
##
## [1] "Personal_Views_Relate_To_Model"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]] + d$Gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9472 -0.7032 -0.2444  1.0528  1.7556
##
## Coefficients:

```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.2968    0.0963  -3.082   0.0022 **
## d[["Group"]]Treatment  0.2440    0.1103   2.212   0.0275 *
## d$GenderMale       -0.4587    0.1104  -4.154 4.02e-05 ***
## d$GenderOther      -1.7032    1.0919  -1.560   0.1196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.088 on 386 degrees of freedom
## Multiple R-squared:  0.05822,    Adjusted R-squared:  0.0509
## F-statistic: 7.954 on 3 and 386 DF,  p-value: 3.7e-05
```

Differences in Mean for Text question

```
get_ATE <- function(d, column){
  return(mean(d[d$Group == 'Treatment',][[column]], na.rm = TRUE)- mean(d[d$Group == 'Control',][[column]]))
}
for (column in columns_to_analyse){
  print(column)
  print(get_ATE(d, column))
}

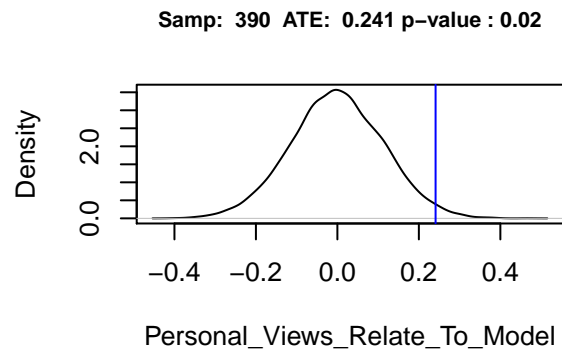
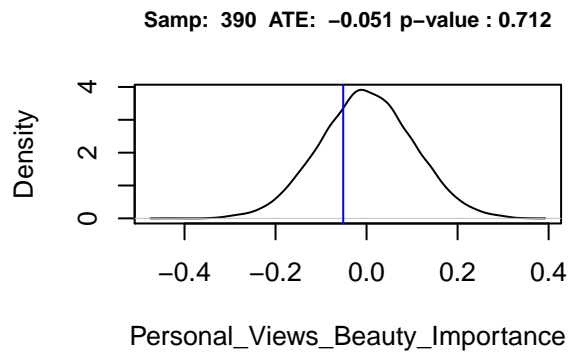
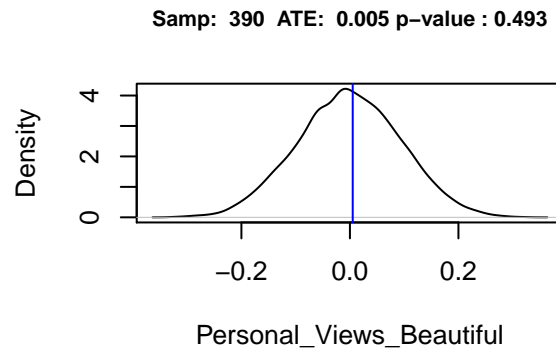
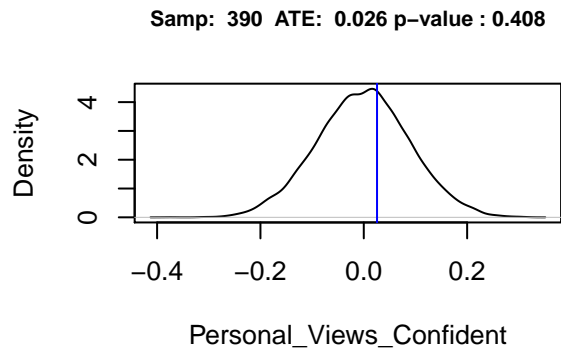
## [1] "Personal_Views_Confident"
## [1] 0.02564103
## [1] "Personal_Views_Beautiful"
## [1] 0.005128205
## [1] "Personal_Views_Beauty_Importance"
## [1] -0.05128205
## [1] "Personal_Views_Relate_To_Model"
## [1] 0.2410256

column <- 'Personal_Views_Confident'

control_treatment <- c(rep(1, length(d[[column]]) * (1/2)), rep(0, length(d[[column]]) * (1/2)))
sample_size <- length(d[[column]])

get_null_ATE_from_current_sample <- function(d, control_treatment, column){
  assignment <- sample(control_treatment, length(d[[column]]))
  dt <- data.frame(outcome = d[[column]], assignment = assignment)
  null.ATE <- mean(dt[assignment == 1,]$outcome) - mean(dt[assignment == 0,]$outcome)
  return(null.ATE)
}

par(mfrow=c(2,2))
for (column in columns_to_analyse){
  sharp.null.hypothesis <- replicate(10000, get_null_ATE_from_current_sample(d, control_treatment, column))
  ATE <- get_ATE(d, column)
  p_value <- mean(ATE <= sharp.null.hypothesis)
  plot(density(sharp.null.hypothesis), main = paste('Samp: ', sample_size, ' ATE: ', round(ATE, 3),
                                                    'p-value: ', round(p_value, 3)), cex.main= 0.8,
        xlab=column)
  abline(v = ATE, col = "blue")
}
```



Similar results than yielded by linear regression

Getting the ATE for IMAGES questions

```
images <- c('Passion', 'Coffee', 'Couple', 'Work', 'Fit')
questions <- c('_i_identify_', '_i_prefer_', '_o_prefer_', '_validate_')

for (image in images){
  for (question in questions){
    column <- paste(image, question, sep = "_")
    l1 <- lm(d[[column]] ~ d[['Group']])
    print(column)
    print(summary(l1))
  }
}
```

```
## [1] "Passion_i_identify_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3529 -0.3529 -0.2895  0.6471  0.7105
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.35294    0.03415  10.335  <2e-16 ***
```

```

## d[["Group"]]Treatment -0.06347    0.04811  -1.319    0.188
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.467 on 375 degrees of freedom
## (13 observations deleted due to missingness)
## Multiple R-squared:  0.00462, Adjusted R-squared:  0.001966
## F-statistic: 1.741 on 1 and 375 DF, p-value: 0.1879
##
## [1] "Passion_i_prefer_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3936 -0.3936 -0.2872  0.6064  0.7128
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.39362    0.03443  11.432 <2e-16 ***
## d[["Group"]]Treatment -0.10638    0.04869  -2.185  0.0295 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4721 on 374 degrees of freedom
## (14 observations deleted due to missingness)
## Multiple R-squared:  0.0126, Adjusted R-squared:  0.009961
## F-statistic: 4.773 on 1 and 374 DF, p-value: 0.02953
##
## [1] "Passion_o_prefer_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6064 -0.5895  0.3936  0.4105  0.4105
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.60638    0.03585  16.914 <2e-16 ***
## d[["Group"]]Treatment -0.01691    0.05057  -0.334  0.738
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4916 on 376 degrees of freedom
## (12 observations deleted due to missingness)
## Multiple R-squared:  0.0002973, Adjusted R-squared:  -0.002361
## F-statistic: 0.1118 on 1 and 376 DF, p-value: 0.7383
##
## [1] "Passion_validate_"
##
## Call:

```

```

## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.98936  0.01064  0.01064  0.02632  0.02632
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.989362   0.009842 100.525  <2e-16 ***
## d[["Group"]]Treatment -0.015677   0.013882  -1.129   0.259
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1349 on 376 degrees of freedom
## (12 observations deleted due to missingness)
## Multiple R-squared:  0.003381, Adjusted R-squared:  0.00073
## F-statistic: 1.275 on 1 and 376 DF, p-value: 0.2595
##
## [1] "Coffee_i_identify_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1848 -0.1848 -0.1482 -0.1482  0.8518
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.18478   0.02749   6.723 6.74e-11 ***
## d[["Group"]]Treatment -0.03663   0.03861  -0.949   0.343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3728 on 371 degrees of freedom
## (17 observations deleted due to missingness)
## Multiple R-squared:  0.002421, Adjusted R-squared: -0.0002684
## F-statistic: 0.9002 on 1 and 371 DF, p-value: 0.3433
##
## [1] "Coffee_i_prefer_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3027 -0.3027 -0.2526  0.6973  0.7474
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.30270   0.03295   9.187  <2e-16 ***
## d[["Group"]]Treatment -0.05007   0.04629  -1.082   0.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```



```

##
## Residual standard error: 0.4482 on 373 degrees of freedom
## (15 observations deleted due to missingness)
## Multiple R-squared: 0.003127, Adjusted R-squared: 0.0004542
## F-statistic: 1.17 on 1 and 373 DF, p-value: 0.2801
##
## [1] "Coffee_o_prefer_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4842 -0.4842 -0.4216  0.5158  0.5784
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.42162    0.03663   11.512 <2e-16 ***
## d[["Group"]]Treatment 0.06259    0.05145    1.216  0.225
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4982 on 373 degrees of freedom
## (15 observations deleted due to missingness)
## Multiple R-squared: 0.003951, Adjusted R-squared: 0.001281
## F-statistic: 1.48 on 1 and 373 DF, p-value: 0.2246
##
## [1] "Coffee_validate_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.98919  0.01081  0.01081  0.02105  0.02105
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.989189    0.009242  107.031 <2e-16 ***
## d[["Group"]]Treatment -0.010242    0.012984   -0.789  0.431
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1257 on 373 degrees of freedom
## (15 observations deleted due to missingness)
## Multiple R-squared: 0.001665, Adjusted R-squared: -0.001011
## F-statistic: 0.6222 on 1 and 373 DF, p-value: 0.4307
##
## [1] "Couple_i_identify_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:

```

```

##      Min      1Q  Median      3Q      Max
## -0.2234 -0.2234 -0.2064 -0.2064  0.7936
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.20635    0.02995   6.890 2.37e-11 ***
## d[["Group"]]Treatment 0.01706    0.04241   0.402   0.688
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4117 on 375 degrees of freedom
## (13 observations deleted due to missingness)
## Multiple R-squared:  0.0004311, Adjusted R-squared:  -0.002234
## F-statistic: 0.1617 on 1 and 375 DF,  p-value: 0.6878
##
## [1] "Couple_i_prefer_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -0.3191 -0.3191 -0.3175  0.6808  0.6825
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.317460    0.033973   9.344 <2e-16 ***
## d[["Group"]]Treatment 0.001689    0.048110   0.035   0.972
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4671 on 375 degrees of freedom
## (13 observations deleted due to missingness)
## Multiple R-squared:  3.285e-06, Adjusted R-squared:  -0.002663
## F-statistic: 0.001232 on 1 and 375 DF,  p-value: 0.972
##
## [1] "Couple_o_prefer_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -0.5904 -0.5106  0.4096  0.4894  0.4894
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.51064    0.03626  14.083 <2e-16 ***
## d[["Group"]]Treatment 0.07979    0.05128   1.556   0.121
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4972 on 374 degrees of freedom
## (14 observations deleted due to missingness)

```

```

## Multiple R-squared:  0.006432,   Adjusted R-squared:  0.003775
## F-statistic: 2.421 on 1 and 374 DF,  p-value: 0.1206
##
## [1] "Couple_validate_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.98930  0.01070  0.01070  0.02646  0.02646
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.973545   0.009842  98.922  <2e-16 ***
## d[["Group"]]Treatment 0.015760   0.013955   1.129    0.259
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1353 on 374 degrees of freedom
## (14 observations deleted due to missingness)
## Multiple R-squared:  0.003398,   Adjusted R-squared:  0.0007338
## F-statistic: 1.275 on 1 and 374 DF,  p-value: 0.2595
##
## [1] "Work_i_identify_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5027 -0.4628 -0.4628  0.4974  0.5372
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.46277   0.03651  12.674  <2e-16 ***
## d[["Group"]]Treatment 0.03988   0.05157   0.773    0.44
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5006 on 375 degrees of freedom
## (13 observations deleted due to missingness)
## Multiple R-squared:  0.001592,   Adjusted R-squared:  -0.00107
## F-statistic: 0.598 on 1 and 375 DF,  p-value: 0.4398
##
## [1] "Work_i_prefer_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6738 -0.6649  0.3262  0.3351  0.3351
##

```

```

## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.664894   0.034402  19.327  <2e-16 ***
## d[["Group"]]Treatment 0.008903   0.048716   0.183    0.855
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4717 on 373 degrees of freedom
## (15 observations deleted due to missingness)
## Multiple R-squared:  8.954e-05, Adjusted R-squared:  -0.002591
## F-statistic: 0.0334 on 1 and 373 DF,  p-value: 0.8551
##
## [1] "Work_o_prefer_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7884  0.2116  0.2116  0.2394  0.2394
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.76064   0.03054  24.905  <2e-16 ***
## d[["Group"]]Treatment 0.02772   0.04314   0.643    0.521
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4188 on 375 degrees of freedom
## (13 observations deleted due to missingness)
## Multiple R-squared:  0.0011, Adjusted R-squared:  -0.001564
## F-statistic: 0.413 on 1 and 375 DF,  p-value: 0.5208
##
## [1] "Work_validate_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.98413  0.01587  0.01587  0.02128  0.02128
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.978723   0.009870  99.166  <2e-16 ***
## d[["Group"]]Treatment 0.005404   0.013939   0.388    0.698
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1353 on 375 degrees of freedom
## (13 observations deleted due to missingness)
## Multiple R-squared:  0.0004006, Adjusted R-squared:  -0.002265
## F-statistic: 0.1503 on 1 and 375 DF,  p-value: 0.6985
##

```

```

## [1] "Fit_i_identify_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4032 -0.4032 -0.3723  0.5968  0.6277
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.40323    0.03580  11.262  <2e-16 ***
## d[["Group"]]Treatment -0.03089    0.05050  -0.612    0.541
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4883 on 372 degrees of freedom
## (16 observations deleted due to missingness)
## Multiple R-squared:  0.001005, Adjusted R-squared:  -0.001681
## F-statistic: 0.3741 on 1 and 372 DF, p-value: 0.5412
##
## [1] "Fit_i_prefer_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3936 -0.3936 -0.3817  0.6064  0.6183
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.38172    0.03582  10.657  <2e-16 ***
## d[["Group"]]Treatment 0.01190    0.05052   0.235    0.814
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4885 on 372 degrees of freedom
## (16 observations deleted due to missingness)
## Multiple R-squared:  0.000149, Adjusted R-squared:  -0.002539
## F-statistic: 0.05545 on 1 and 372 DF, p-value: 0.814
##
## [1] "Fit_o_prefer_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5957 -0.5806  0.4043  0.4194  0.4194
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.58065    0.03618  16.049  <2e-16 ***

```

```
## d[["Group"]]Treatment 0.01510 0.05103 0.296 0.767
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4934 on 372 degrees of freedom
## (16 observations deleted due to missingness)
## Multiple R-squared: 0.0002353, Adjusted R-squared: -0.002452
## F-statistic: 0.08756 on 1 and 372 DF, p-value: 0.7675
##
## [1] "Fit_validate_"
##
## Call:
## lm(formula = d[[column]] ~ d[["Group"]])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.98925  0.00000  0.00000  0.01075  0.01075
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.989247   0.005347 184.997 <2e-16 ***
## d[["Group"]]Treatment 0.010753   0.007542   1.426  0.155
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07293 on 372 degrees of freedom
## (16 observations deleted due to missingness)
## Multiple R-squared: 0.005434, Adjusted R-squared: 0.002761
## F-statistic: 2.033 on 1 and 372 DF, p-value: 0.1548
```

Taking the total and average by questions for all images

```
#str(d)
#d$total <- rep(0, nrow(d))

questions <- c('_i_identify_', '_i_prefer_', '_o_prefer_', '_validate_')
images <- c('Passion', 'Coffee', 'Couple', 'Work', 'Fit')

for (question in questions){
  question_average <- paste(question, 'average', sep = "")
  #print(question_average)
  d[[question_average]] <- rep(0, nrow(d))
  for (image in images){
    column <- paste(image, question, sep = "")
    d[[question_average]] <- d[[question_average]] + d[[column]]
  }
}

for (question in questions){
  question_average <- paste(question, 'average', sep = "")
  print(question_average)
  print(summary(d[[question_average]]))
}
```

```

l1 <- lm(d[[question_average]] ~ d[['Group']] + d$Gender)
print(summary(l1))
}

## [1] "_i_identify_average"
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.      NA's
##      0.000   0.000    1.000    1.536   2.000    5.000       30
##
## Call:
## lm(formula = d[[question_average]] ~ d[["Group"]] + d$Gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7431 -1.3108 -0.3108  0.6892  3.6892
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.39162    0.12294   11.320 <2e-16 ***
## d[["Group"]]Treatment -0.08085    0.13884   -0.582  0.561
## d$GenderMale      0.35144    0.13911    2.526  0.012 *
## d$GenderOther     -0.39162    1.32082   -0.296  0.767
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.315 on 356 degrees of freedom
## (30 observations deleted due to missingness)
## Multiple R-squared:  0.01888,    Adjusted R-squared:  0.01061
## F-statistic: 2.284 on 3 and 356 DF,  p-value: 0.07873
##
## [1] "_i_prefer_average"
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.      NA's
##      0.000   1.000    2.000    1.969   3.000    5.000       30
##
## Call:
## lm(formula = d[[question_average]] ~ d[["Group"]] + d$Gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3662 -0.6892 -0.2122  0.7878  3.4648
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.6892    0.1158   14.581 < 2e-16 ***
## d[["Group"]]Treatment -0.1540    0.1313   -1.173  0.242
## d$GenderMale      0.6770    0.1316    5.145 4.44e-07 ***
## d$GenderOther     -0.6892    1.2493   -0.552  0.582
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.244 on 356 degrees of freedom
## (30 observations deleted due to missingness)
## Multiple R-squared:  0.07362,    Adjusted R-squared:  0.06581
## F-statistic: 9.43 on 3 and 356 DF,  p-value: 5.19e-06
##

```

```
## [1] "_o_prefer_average"
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.      NA's
##      0.000   2.000   3.000   2.981   4.000   5.000       28
##
## Call:
## lm(formula = d[[question_average]] ~ d[["Group"]] + d$Gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3316 -0.7471  0.2529  0.8333  2.4177
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.5823     0.1298  19.889 < 2e-16 ***
## d[["Group"]]Treatment  0.1648     0.1469   1.122  0.2627
## d$GenderMale      0.5844     0.1472   3.971 8.65e-05 ***
## d$GenderOther      2.4177     1.4017   1.725  0.0854 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.396 on 358 degrees of freedom
## (28 observations deleted due to missingness)
## Multiple R-squared:  0.05102,    Adjusted R-squared:  0.04306
## F-statistic: 6.415 on 3 and 358 DF,  p-value: 0.0003047
##
## [1] "_validate_average"
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.      NA's
##      4.000   5.000   5.000   4.994   5.000   5.000       28
##
## Call:
## lm(formula = d[[question_average]] ~ d[["Group"]] + d$Gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.98941 -0.00041  0.00045  0.01059  0.01145
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.9995505  0.0069152  722.977 <2e-16 ***
## d[["Group"]]Treatment -0.0110004  0.0078252  -1.406   0.161
## d$GenderMale      0.0008560  0.0078409   0.109   0.913
## d$GenderOther      0.0004495  0.0746514   0.006   0.995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07433 on 358 degrees of freedom
## (28 observations deleted due to missingness)
## Multiple R-squared:  0.005528,    Adjusted R-squared:  -0.002806
## F-statistic: 0.6633 on 3 and 358 DF,  p-value: 0.5751
```