The Effects of Branded Coffee on Preceived Taste Satisfaction

Amy Zhang, Chris John, Jenna Farac, Simran Gill

2024-11-27

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 4.4.1
```

```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.4.1
```

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:gridExtra':
##
##     combine
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(car)
```

```
## Warning: package 'car' was built under R version 4.4.1
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.4.1
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```r
set.seed(123)

control <- read.csv("Coffee Survey Control (Responses) - Form Responses 1.csv")
treatment <- read.csv("Coffee Survey Group II (Responses) - Form Responses 1.csv")

# rename column names for control
colnames(control) <- c('timestamp', 'good_and_gather_score', 'chameleon_score', 'age', 'gender','how_of
# rename column names for treatment
colnames(treatment) <- c('timestamp', 'name','good_and_gather_score', 'chameleon_score', 'age', 'gender
# reorder column names for treatment
treatment <- treatment[, c('timestamp', 'good_and_gather_score', 'chameleon_score', 'age', 'gender','hou

#--- Control ----

control$treatment <- 0
control$age <- as.integer(control$age)
```

```
## Warning: NAs introduced by coercion
```

```r
# re-labeling gender
control <- control %>%
  mutate(gender = case_when(
    gender == "F" ~ "Female",
    gender == "M" ~ "Male",
    TRUE ~ "Unknown"
  ))

# removing rows where age is null
control<- control %>%
  filter(!is.na(age))

#--- Treatment ----

treatment$treatment <- 1
treatment$age <- as.integer(treatment$age)
```

```r
# re-labeling gender
treatment <- treatment %>%
  mutate(gender = case_when(
    gender == "F" ~ "Female",
    gender == "M" ~ "Male",
    TRUE ~ "Unknown"
  ))

# removing rows where age is null
treatment<- treatment %>%
  filter(!is.na(age))
```

Balancing Control and Treatment

There are more participants in Control than in Treatment groups. To help create balance between the two groups, will perform random sampling to match the size of the treatment group.

```r
print("Before Random Sampling:")
```

```
## [1] "Before Random Sampling:"
```

```r
cat("Control size:", nrow(control))
```

```
## Control size: 51
```

```r
cat("\nTreatment size:", nrow(treatment))
```

```
##
## Treatment size: 39
```

```r
# selecting smaller group size
n_control <- nrow(control)
min_size <- min(n_control, nrow(treatment))

# random sampling the control group
control <- control[sample(1:n_control, min_size), ]

# combined data
d <- rbind(control, treatment)
table(d$treatment)
```

```
##
##  0  1
## 39 39
```

Organizing the rest of the data from dataset “d”

```r
# creating age groups
d$age_group <- cut(d$age,
                   breaks = c(0, 20, 30, 40, 50, Inf),
                   labels = c("Under 20", "20-30", "31-40", "41-50", "Over 50"),
                   right = FALSE)


# Convert how_often_drink_coffee to integer by factoring
d$how_often_drink_coffee <- factor(d$how_often_drink_coffee,
                                   levels = c("Never",
                                              "Occasionally (up to 1 time a week)",
                                              "Sometimes (a few times a week)",
                                              "Often (almost every day)",
                                              "Every day"))

# yes/no flag for if the participant is aware of the coffee brand at all
d$chameleon_awareness_flag <- ifelse(d$chameleon_awareness == "No", 0, 1)
d$good_and_gather_awareness_flag <- ifelse(d$good_and_gather_awareness == "No", 0, 1)

cat("\nNumber of Rows after cleaning:",nrow(d),"\n")
```

```
##
## Number of Rows after cleaning: 78
```

```r
str(d)
```

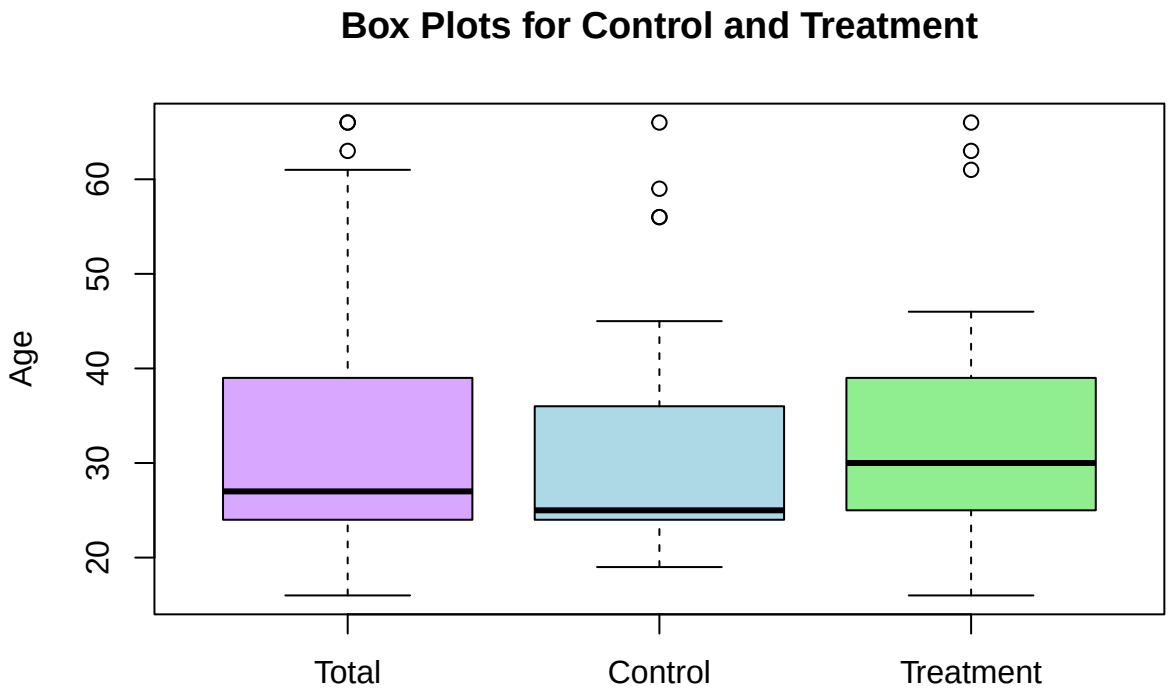
```
## 'data.frame':    78 obs. of  16 variables:
##  $ timestamp                    : chr  "11/23/2024 18:32:49" "11/11/2024 10:21:52" "11/11/2024 10:2
##  $ good_and_gather_score        : int  1 4 3 1 5 2 3 5 3 4 ...
##  $ chameleon_score              : int  5 2 5 5 4 4 6 3 2 5 ...
##  $ age                          : int  24 24 25 27 31 25 27 38 20 45 ...
##  $ gender                       : chr  "Female" "Male" "Male" "Female" ...
##  $ how_often_drink_coffee       : Factor w/ 5 levels "Never","Occasionally (up to 1 time a week)",.
##  $ hot_or_cold                  : chr  "Cold Coffee" "Cold Coffee" "Cold Coffee" "Cold Coffee" ...
##  $ sweet_or_not_sweet           : chr  "Sweet" "Not Sweet" "Sweet" "Not Sweet" ...
##  $ good_and_gather_awareness    : chr  "Yes, Neutral" "Yes, Neutral" "Yes, Neutral" "Yes, Positive"
##  $ chameleon_awareness          : chr  "No" "No" "No" "Yes, Positive" ...
##  $ medical_condition            : chr  "No" "No" "No" "No" ...
##  $ name                         : chr  "Shivani Bangalore" "Chris L" "Stephen Hei" "Liz Ren" ...
##  $ treatment                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ age_group                    : Factor w/ 5 levels "Under 20","20-30",..: 2 2 2 2 3 2 2 3 2 4 ...
##  $ chameleon_awareness_flag     : num  0 0 0 1 0 0 1 0 0 0 ...
##  $ good_and_gather_awareness_flag: num  1 1 1 1 1 0 0 0 0 0 ...
```

Exploratory Data Analysis

```r
# box plot for age by treatment and control

boxplot(d$age, control$age, treatment$age,
        names = c("Total", "Control", "Treatment"),
```

```
        main = "Box Plots for Control and Treatment",
        ylab = "Age",
        col = c("#D8A7FF", "lightblue", "lightgreen"),
        border = "black")
```



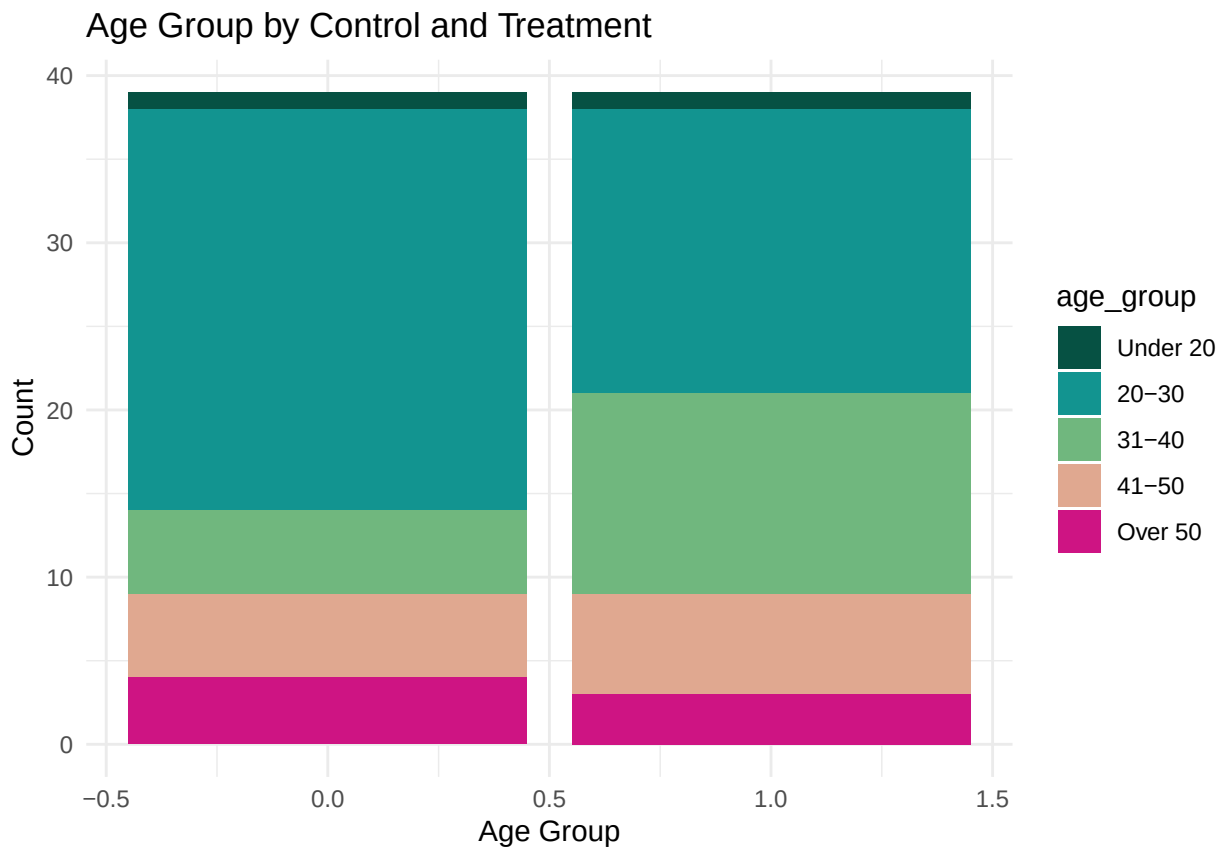

```
cat("Number of Rows for Treatment Group:", sum(d$treatment == 1))
```

```
## Number of Rows for Treatment Group: 39
```

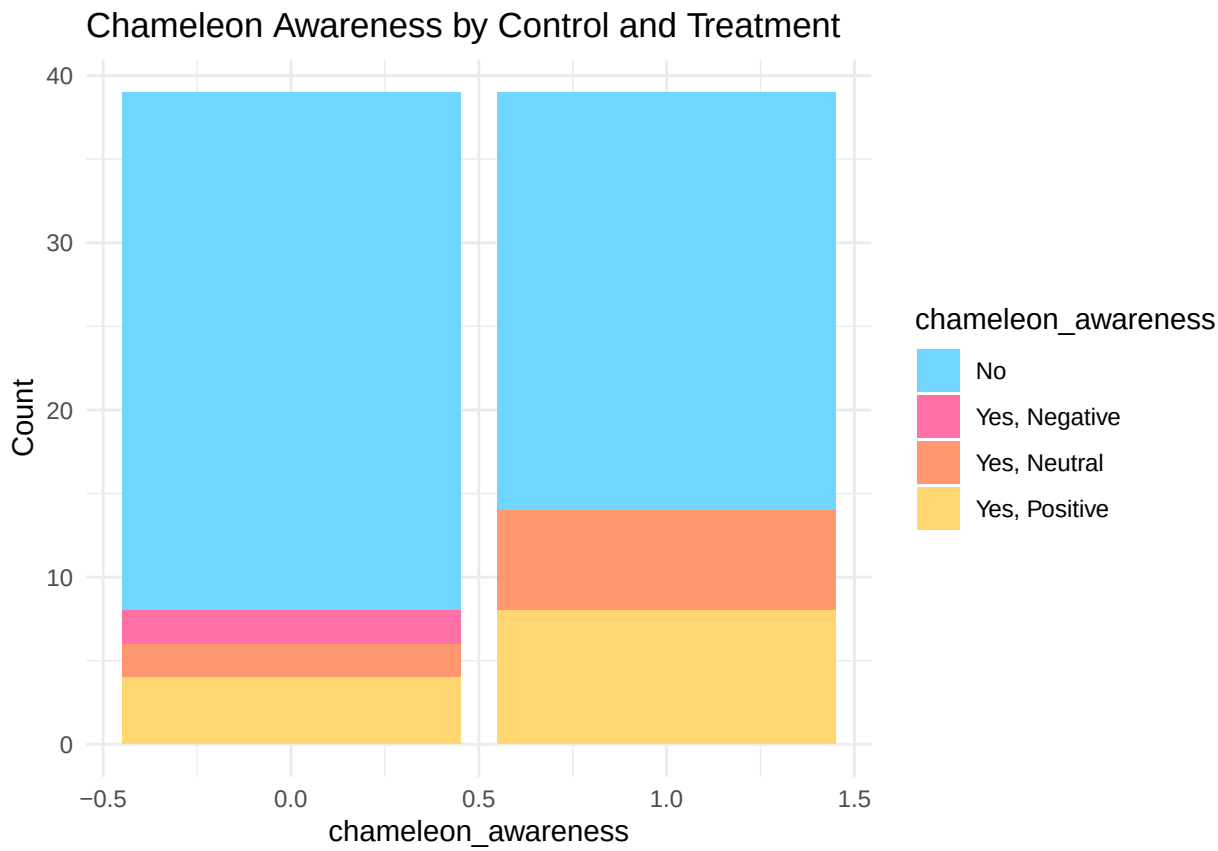
```
cat("\nNumber of Rows for Control Group:", sum(d$treatment == 0))
```

```
##
## Number of Rows for Control Group: 39
```

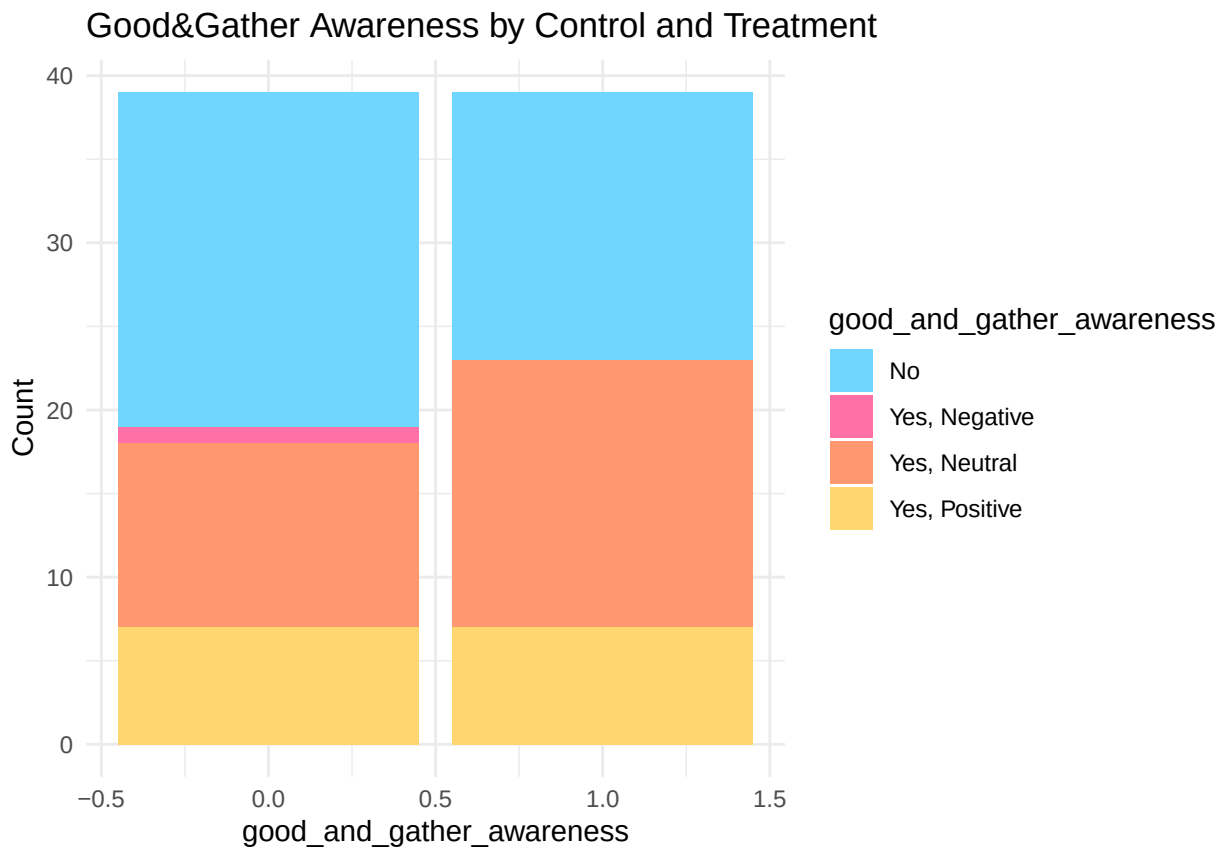
```
ggplot(d, aes(x = treatment, fill = age_group)) +
  geom_bar(position = "stack") +
  labs(title = "Age Group by Control and Treatment", x = "Age Group", y = "Count") +
  scale_fill_manual(values = c("#065143", "#129490", "#70B77E", "#E0A890", "#CE1483"))  +
  theme_minimal()
```




```
ggplot(d, aes(x = treatment, fill = chameleon_awareness)) +
  geom_bar(position = "stack") +
  labs(title = "Chameleon Awareness by Control and Treatment", x = "chameleon_awareness", y = "Count") +
  scale_fill_manual(values = c("#70D6FF", "#FF70A6", "#FF9770", "#FFD670")) +
  theme_minimal()
```




```
ggplot(d, aes(x = treatment, fill = good_and_gather_awareness)) +
  geom_bar(position = "stack") +
  labs(title = "Good&Gather Awareness by Control and Treatment", x = "good_and_gather_awareness", y = "
  scale_fill_manual(values = c("#70D6FF", "#FF70A6", "#FF9770", "#FFD670")) +
  theme_minimal()
```



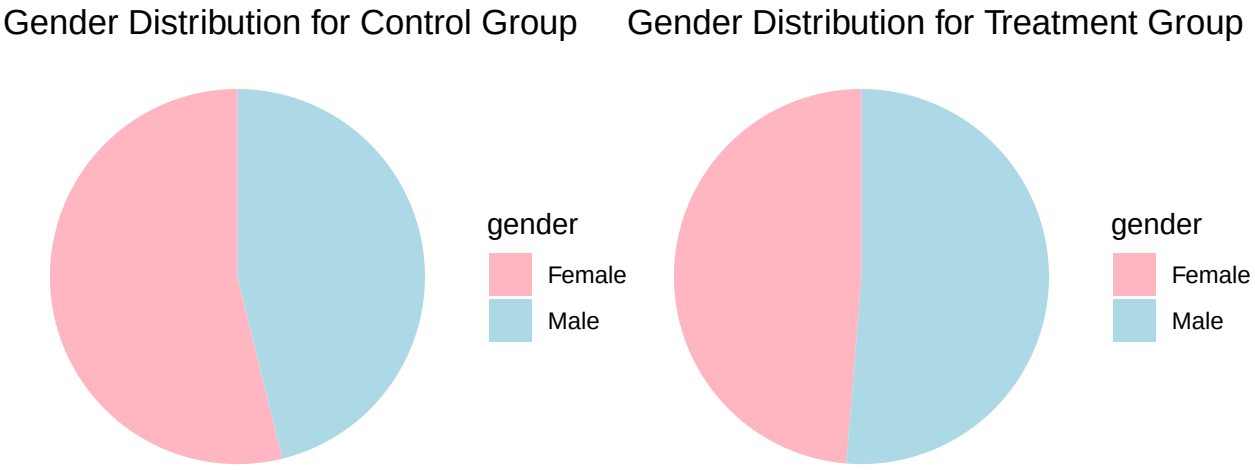

```
### Control Group Gender ###
control_gender_counts <- control %>%
  group_by(gender) %>%
  tally()

control_pie <- ggplot(control_gender_counts, aes(x = "", y = n, fill = gender)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  labs(title = "Gender Distribution for Control Group") +
  scale_fill_manual(values = c("lightpink", "lightblue", "purple")) +
  theme_void()

### Treatment Group Gender ###
treatment_gender_counts <- treatment %>%
  group_by(gender) %>%
  tally()

treatment_pie <- ggplot(treatment_gender_counts, aes(x = "", y = n, fill = gender)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  labs(title = "Gender Distribution for Treatment Group") +
  scale_fill_manual(values = c("lightpink", "lightblue", "purple")) +
  theme_void()

grid.arrange(control_pie, treatment_pie, ncol = 2)
```



Simple Average Treatment Effect

```
ate_good_and_gather <- mean(d$good_and_gather_score[d$treatment == 1], na.rm = TRUE) -
                       mean(d$good_and_gather_score[d$treatment == 0], na.rm = TRUE)
cat("ATE Good & Gather:", ate_good_and_gather)
```

```
## ATE Good & Gather: -0.4358974
```

```
ate_chameleon <- mean(d$chameleon_score[d$treatment == 1], na.rm = TRUE) -
                       mean(d$chameleon_score[d$treatment == 0], na.rm = TRUE)
cat("\nATE Chameleon:", ate_chameleon)
```

```
##
## ATE Chameleon: 0.4102564
```

Average Treatment Effect using Linear Regression

```
# Basic Linear regression to estimate ATE
model_gg <- lm(good_and_gather_score ~ treatment, data=d)
ate_regression <- coef(model_gg)["treatment"]
print(ate_regression)
```

```
##  treatment
## -0.4358974
```

```
summary(model_gg)
```

```
##
## Call:
## lm(formula = good_and_gather_score ~ treatment, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.05128 -1.05128 -0.05128  1.27564  2.94872
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.0513     0.2317   17.48   <2e-16 ***
## treatment    -0.4359     0.3277   -1.33    0.187
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.447 on 76 degrees of freedom
## Multiple R-squared:  0.02275,    Adjusted R-squared:  0.009896
## F-statistic:  1.77 on 1 and 76 DF,  p-value: 0.1874
```

```
# Basic Linear regression to estimate ATE
model_c <- lm(chameleon_score ~ treatment, data=d)
ate_regression <- coef(model_c)["treatment"]
print(ate_regression)
```

```
## treatment
## 0.4102564
```

```
summary(model_c)
```

```
##
## Call:
## lm(formula = chameleon_score ~ treatment, data = d)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.051 -1.641  0.359  1.359  2.359
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.6410     0.2457  14.821   <2e-16 ***
## treatment     0.4103     0.3474   1.181    0.241
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.534 on 76 degrees of freedom
## Multiple R-squared:  0.01802,    Adjusted R-squared:  0.005097
## F-statistic: 1.394 on 1 and 76 DF,  p-value: 0.2413
```

ATE Adjusted for Covariates

```
model_gg_covariates <- lm(good_and_gather_score ~ treatment + log(age) + gender + chameleon_awareness ,

ate_with_covariates <- coef(model_gg_covariates)["treatment"]
print(ate_with_covariates)
```

```
##  treatment
## -0.6688623
```

```
summary(model_gg_covariates)
```

```
##
## Call:
## lm(formula = good_and_gather_score ~ treatment + log(age) + gender +
##     chameleon_awareness, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3406 -0.8404 -0.0484  0.8124  2.5604
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      -0.9548     1.6826  -0.567  0.57218
## treatment                        -0.6689     0.3104  -2.155  0.03456 *
## log(age)                          1.4542     0.4853   2.997  0.00376 **
## genderMale                        0.4417     0.2985   1.480  0.14338
## chameleon_awarenessYes, Negative -1.9733     0.9541  -2.068  0.04227 *
## chameleon_awarenessYes, Neutral   0.9138     0.5118   1.786  0.07845 .
## chameleon_awarenessYes, Positive -0.6917     0.4220  -1.639  0.10565
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.307 on 71 degrees of freedom
## Multiple R-squared:  0.2553, Adjusted R-squared:  0.1924
## F-statistic: 4.057 on 6 and 71 DF,  p-value: 0.001492
```

```
model_gg_covariates_v2 <- lm(good_and_gather_score ~ treatment + gender + log(age) + chameleon_awareness

anova(model_gg_covariates , model_gg_covariates_v2)
```

```
## Analysis of Variance Table
##
## Model 1: good_and_gather_score ~ treatment + log(age) + gender + chameleon_awareness
## Model 2: good_and_gather_score ~ treatment + gender + log(age) + chameleon_awareness +
##     good_and_gather_awareness
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     71 121.26
## 2     68 116.18  3    5.0837 0.9918  0.402
```

Interpretation We tested multiple covariates to see if we can improve the regression model for Good&Gather Score. The main covariates we see has a positive impact is how a participate views the Chameleon brand and age group.

When it comes to age, participants in the age group 30 - 39 and 40 - 49 are likely to rate Good & Gather higher after the brand is revealed. Because these two variables have some significants, age group does play a part in how a participant rates the coffee after treatment is provided.

When it comes to the Chameleon, even though the participants has a negative view of Chameleon coffee as a brand, they are still likely to score Good & Gather -1.8285 after treatment is provided. The p-value for Chameleon awareness is 0.0561, which means this variable is marginally significant.

We also wanted to test if adding Good&Gather brand awareness as a variable to model has an significant effect to the model. From the ANOVA test we can see that the p-value is 0.5600 which is greater than 0.05. This indicated Good&Gather brand awareness has no statistically significant impact on scoring the coffee.

```
stargazer(model_gg_covariates, model_gg_covariates_v2,
          type = "text",    # Use "html" for HTML output or "latex" for LaTeX
          title = "Regression Results for Good and Gather Score",
          covariate.labels = c("Treatment", "log(Age)", "Gender", "Chameleon Awareness", "Good and Gathe
          star.cutoffs = c(0.10, 0.05, 0.01, 0.001),    # Significance stars
          out = "regression_table.txt")  # Optional: Save output to a text file
```

```
##
## Regression Results for Good and Gather Score
## ================================================================================
##                                                  Dependent variable:
##                                              ---------------------------------------
##                                                    good_and_gather_score
##                                                  (1)                    (2)
## --------------------------------------------------------------------------------
## Treatment                                      -0.669**               -0.631**
##                                                 (0.310)                (0.314)
##
## log(Age)                                       1.454***                1.273**
##                                                 (0.485)                (0.498)
##
## Gender                                          0.442                  0.284
##                                                 (0.298)                (0.312)
##
## Chameleon Awareness                            -1.973**               -2.238**
##                                                 (0.954)                (0.969)
##
## Good and Gather Awareness                       0.914*                 0.980*
##                                                 (0.512)                (0.525)
##
## chameleon_awarenessYes, Positive                -0.692                 -0.563
##                                                 (0.422)                (0.440)
##
## good_and_gather_awarenessYes, Negative                                 -0.967
##                                                                        (1.359)
##
## good_and_gather_awarenessYes, Neutral                                  -0.606
##                                                                        (0.366)
##
```

```
## good_and_gather_awarenessYes, Positive                                -0.279
##                                                                       (0.444)
##
## Constant                                          -0.955              -0.024
##                                                   (1.683)             (1.779)
##
## ------------------------------------------------------------------------------------
## Observations                                        78                  78
## R2                                                 0.255               0.287
## Adjusted R2                                        0.192               0.192
## Residual Std. Error                          1.307 (df = 71)     1.307 (df = 68)
## F Statistic                                4.057*** (df = 6; 71) 3.034*** (df = 9; 68)
## ====================================================================================
## Note:                                                    *p<0.1; **p<0.05; ***p<0.01
```

```
model_c_covariates <- lm(chameleon_score ~ treatment + log(age) + gender + chameleon_awareness_flag + g

ate_with_covariates <- coef(model_c_covariates)["treatment"]
print(ate_with_covariates)
```

```
##  treatment
## 0.05415126
```

```
summary(model_c_covariates)
```

```
##
## Call:
## lm(formula = chameleon_score ~ treatment + log(age) + gender +
##     chameleon_awareness_flag + good_and_gather_awareness_flag,
##     data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1784 -0.9022  0.0574  0.9978  2.4911
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -1.71481    1.81840  -0.943 0.348819
## treatment                       0.05415    0.31841   0.170 0.865434
## log(age)                        1.40090    0.51195   2.736 0.007818 **
## genderMale                      0.22952    0.32169   0.713 0.477854
## chameleon_awareness_flag        1.40292    0.36032   3.894 0.000219 ***
## good_and_gather_awareness_flag  0.44192    0.33600   1.315 0.192598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.366 on 72 degrees of freedom
## Multiple R-squared:  0.2623, Adjusted R-squared:  0.2111
## F-statistic:  5.12 on 5 and 72 DF,  p-value: 0.0004505
```

```
model_c_covariates_v2 <- lm(chameleon_score ~ treatment + log(age) + gender + chameleon_awareness_flag,

anova(model_c_covariates , model_c_covariates_v2)
```

```
## Analysis of Variance Table
##
## Model 1: chameleon_score ~ treatment + log(age) + gender + chameleon_awareness_flag +
##     good_and_gather_awareness_flag
## Model 2: chameleon_score ~ treatment + log(age) + gender + chameleon_awareness_flag
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     72 134.38
## 2     73 137.61 -1   -3.2286 1.7299 0.1926
```

```
stargazer(model_gg_covariates, model_gg_covariates_v2,
          type = "text",   # Use "html" for HTML output or "latex" for LaTeX
          title = "Regression Results for Good and Gather Score",
          covariate.labels = c("Treatment", "log(Age)", "Gender", "Chameleon Awareness", "Good and Gathe
          star.cutoffs = c(0.10, 0.05, 0.01, 0.001),   # Significance stars
          out = "regression_table.txt")  # Optional: Save output to a text file
```

```
##
## Regression Results for Good and Gather Score
## ================================================================================
##                                                  Dependent variable:
##                                          ---------------------------------------
##                                                  good_and_gather_score
##                                                (1)                   (2)
## --------------------------------------------------------------------------------
## Treatment                                   -0.669**              -0.631**
##                                              (0.310)               (0.314)
##
## log(Age)                                    1.454***               1.273**
##                                              (0.485)               (0.498)
##
## Gender                                       0.442                 0.284
##                                              (0.298)               (0.312)
##
## Chameleon Awareness                         -1.973**              -2.238**
##                                              (0.954)               (0.969)
##
## Good and Gather Awareness                    0.914*                0.980*
##                                              (0.512)               (0.525)
##
## chameleon_awarenessYes, Positive             -0.692                -0.563
##                                              (0.422)               (0.440)
##
## good_and_gather_awarenessYes, Negative                             -0.967
##                                                                    (1.359)
##
## good_and_gather_awarenessYes, Neutral                              -0.606
##                                                                    (0.366)
##
## good_and_gather_awarenessYes, Positive                             -0.279
##                                                                    (0.444)
##
## Constant                                     -0.955                -0.024
##                                              (1.683)               (1.779)
##
```

```
## ------------------------------------------------------------------------------
## Observations                                78                    78
## R2                                         0.255                 0.287
## Adjusted R2                                0.192                 0.192
## Residual Std. Error                  1.307 (df = 71)       1.307 (df = 68)
## F Statistic                     4.057*** (df = 6; 71) 3.034*** (df = 9; 68)
## ==============================================================================
## Note:                                                *p<0.1; **p<0.05; ***p<0.01
```