EDA_141

2024-05-20

Loading libraries & datasets

```r
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.1
```

```
## Warning: package 'stringr' was built under R version 4.3.1
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```r
library(ggplot2)
library(dplyr)
library(broom)
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 4.3.1
```

```
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 4.3.1
```

```
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```r
library(lmerTest)
```

```
##
## Attaching package: 'lmerTest'
##
## The following object is masked from 'package:lme4':
##
##     lmer
##
```

```
## The following object is masked from 'package:stats':
##
##     step
```

```r
old_pre <- read.csv("cleaned_old_pre.csv")
old_post <- read.csv("cleaned_old_post.csv")
new_pre <- read.csv("cleaned_new_pre.csv")
new_post <- read.csv("cleaned_new_post.csv")

old_pre <- old_pre[,-c(29, 30)]
old_post <- old_post[,-c(29, 30)]
new_pre <- new_pre[,-c(29, 30)]
new_post <- new_post[,-c(29, 30)]
```

Q13 (column # 26) and Q2 (column # 15) are negative so their scores need to be reversed

```r
# Function to reverse the scoring for negative questions
reverse_score <- function(x) {
  return(ifelse(x == 1, 5, ifelse(x == 2, 4, ifelse(x == 3, 3, ifelse(x == 4, 2, ifelse(x == 5, 1, x))))
}

# Applying reverse scoring to columns 15 and 26
old_pre[, 15] <- reverse_score(old_pre[, 15])
old_pre[, 26] <- reverse_score(old_pre[, 26])

old_post[, 15] <- reverse_score(old_post[, 15])
old_post[, 26] <- reverse_score(old_post[, 26])

new_pre[, 15] <- reverse_score(new_pre[, 15])
new_pre[, 26] <- reverse_score(new_pre[, 26])

new_post[, 15] <- reverse_score(new_post[, 15])
new_post[, 26] <- reverse_score(new_post[, 26])
```

Correlation plots for the clusters

```r
create_corr_plot <- function(data, columns, title, filename) {
  cor_mat <- cor(data[, columns], use = "complete.obs")

  png(filename, width = 800, height = 800)
  par(mar = c(4, 4, 1, 1))
  corrplot(cor_mat, method = "circle", type = "upper", tl.pos = "d",
           tl.cex = 0.8, cl.cex = 0.8)
  mtext(side = 3, line = 0, title, cex = 1.2)
  dev.off()
}

# Cluster columns according to Prof. Esfandiari
knowledge_columns <- c(14, 17, 19, 21, 27)
attitude_columns <- c(15, 18, 20, 22, 24)
behavior_columns <- c(16, 23, 25, 26, 28)

# Dataset and title setup
cluster_info <- list(
```

```
  list(data = old_pre, columns = knowledge_columns, title = "Old Pre Knowledge Cluster"),
  list(data = old_post, columns = knowledge_columns, title = "Old Post Knowledge Cluster"),
  list(data = new_pre, columns = knowledge_columns, title = "New Pre Knowledge Cluster"),
  list(data = new_post, columns = knowledge_columns, title = "New Post Knowledge Cluster"),
  list(data = old_pre, columns = attitude_columns, title = "Old Pre Attitude Cluster"),
  list(data = old_post, columns = attitude_columns, title = "Old Post Attitude Cluster"),
  list(data = new_pre, columns = attitude_columns, title = "New Pre Attitude Cluster"),
  list(data = new_post, columns = attitude_columns, title = "New Post Attitude Cluster"),
  list(data = old_pre, columns = behavior_columns, title = "Old Pre Behavior Cluster"),
  list(data = old_post, columns = behavior_columns, title = "Old Post Behavior Cluster"),
  list(data = new_pre, columns = behavior_columns, title = "New Pre Behavior Cluster"),
  list(data = new_post, columns = behavior_columns, title = "New Post Behavior Cluster")
)

for (info in cluster_info) {
  create_corr_plot(info$data, info$columns, info$title, paste0(info$title, ".png"))
}
```

Creating composite scores for knowledge, attitude, and behavior clusters

```
# Function to create composite scores for a cluster
create_composite_score <- function(data, columns) {
  rowSums(data[, columns])
}

# Function to transform scores to a range from 0 to 100 (normalzing the scores)
transform_score <- function(score) {
  (score - 5) / (25 - 5) * 100
}

old_pre <- old_pre %>%
  mutate(Knowledge = transform_score(create_composite_score(old_pre, knowledge_columns)),
         Attitude = transform_score(create_composite_score(old_pre, attitude_columns)),
         Behavior = transform_score(create_composite_score(old_pre, behavior_columns)))

old_post <- old_post %>%
  mutate(Knowledge = transform_score(create_composite_score(old_post, knowledge_columns)),
         Attitude = transform_score(create_composite_score(old_post, attitude_columns)),
         Behavior = transform_score(create_composite_score(old_post, behavior_columns)))

new_pre <- new_pre %>%
  mutate(Knowledge = transform_score(create_composite_score(new_pre, knowledge_columns)),
         Attitude = transform_score(create_composite_score(new_pre, attitude_columns)),
         Behavior = transform_score(create_composite_score(new_pre, behavior_columns)))

new_post <- new_post %>%
  mutate(Knowledge = transform_score(create_composite_score(new_post, knowledge_columns)),
         Attitude = transform_score(create_composite_score(new_post, attitude_columns)),
         Behavior = transform_score(create_composite_score(new_post, behavior_columns)))
```

Combining datasets

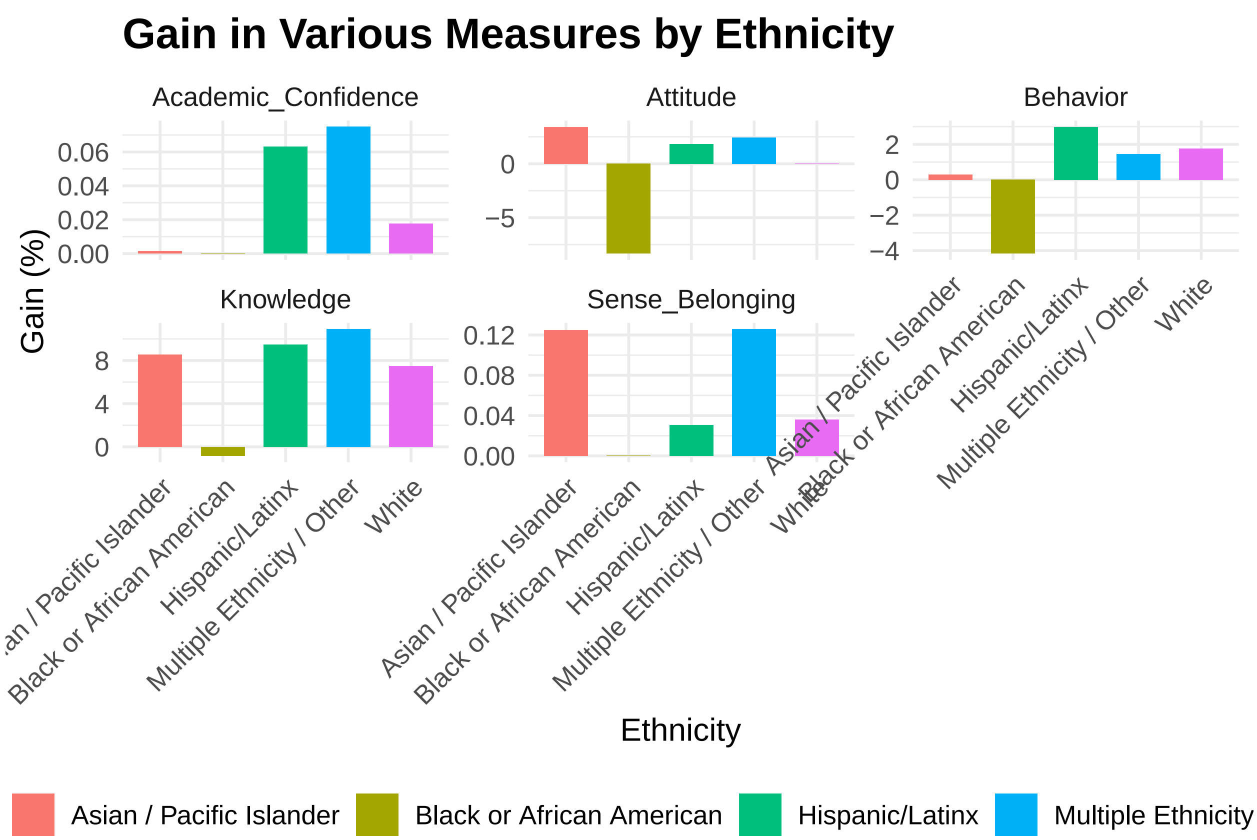
```
old_data <- bind_rows(old_pre %>% mutate(type = "pre", group = "old"),
                      old_post %>% mutate(type = "post", group = "old"))

new_data <- bind_rows(new_pre %>% mutate(type = "pre", group = "new"),
                      new_post %>% mutate(type = "post", group = "new"))

old_data <- na.omit(old_data)
new_data <- na.omit(new_data)

cols_to_factor <- c("Course", "Major", "Gender", "Ethnicity", "First_Gen", "Mother_Edu",
                    "Father_Edu", "Transfer", "GPA", "type", "group", "ID")

old_data <- old_data %>%
  mutate(across(all_of(cols_to_factor), as.factor))

new_data <- new_data %>%
  mutate(across(all_of(cols_to_factor), as.factor))

combined_data <- bind_rows(old_data, new_data)
```

Research Question 1: Gain in Knowledge, Attitude, Behavior, Sense of Belonging, and Academic Confidence by Ethnicity, Major, Transfer, First_Gen, and Gender Visuals

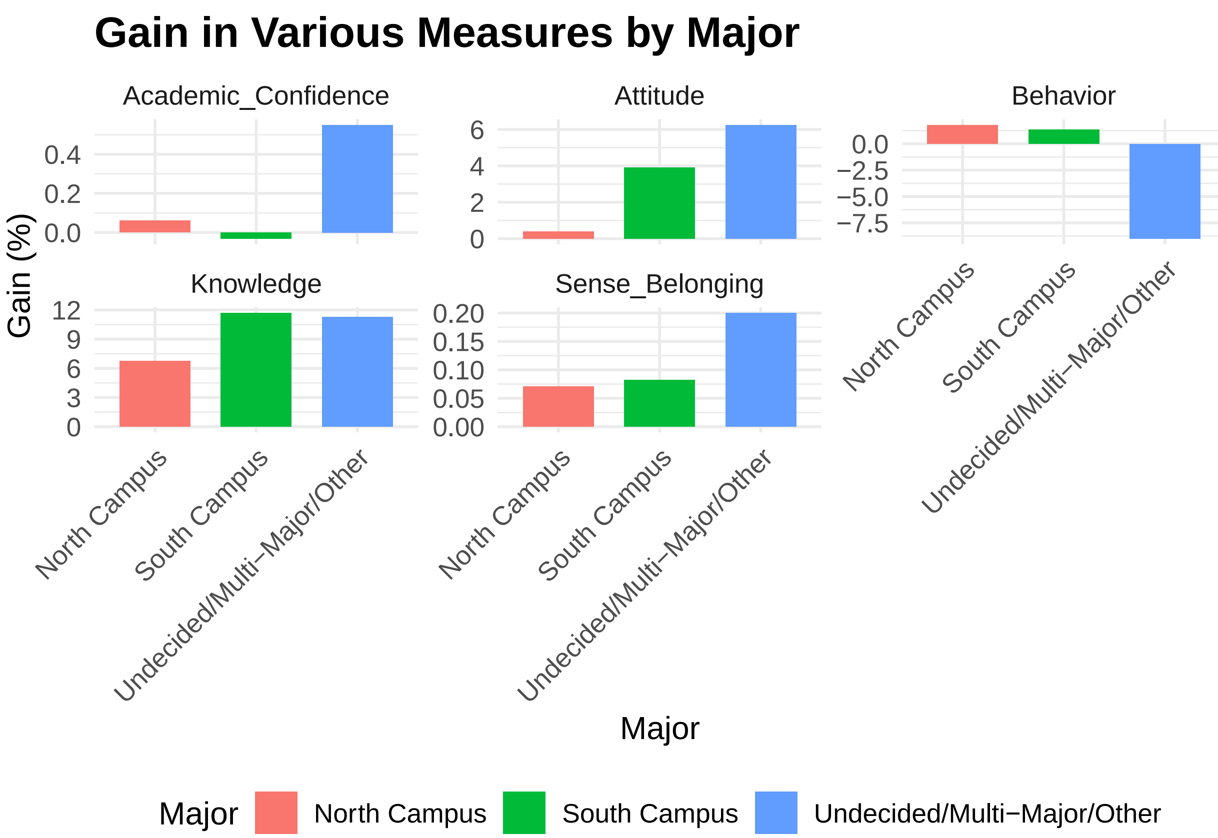
```
variables <- c("Knowledge", "Attitude", "Behavior", "Sense_Belonging", "Academic_Confidence")
group_vars <- c("Ethnicity", "Major", "Transfer", "First_Gen", "Gender")

prepare_gain_data <- function(data, variables, group_var) {
  data_long <- pivot_longer(data, cols = all_of(variables))

  gain_data <- data_long %>%
    group_by(!!sym(group_var), name, type) %>%
    summarise(mean_value = mean(value, na.rm = TRUE), .groups = 'drop') %>%
    pivot_wider(names_from = type, values_from = mean_value) %>%
    mutate(gain = post - pre)

  return(gain_data)
}

plot_gains <- function(gain_data, group_var) {
  p <- ggplot(gain_data, aes(x = !!sym(group_var), y = gain, fill = !!sym(group_var))) +
    geom_bar(stat = "identity", position = position_dodge(), width = 0.7) +
    facet_wrap(~ name, scales = "free_y") +
    labs(title = paste("Gain in Various Measures by", group_var), y = "Gain (%)", x = group_var) +
    theme_minimal() +
    theme(
      text = element_text(size = 12),
      axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1),
      strip_text = element_text(size = 13),
      plot.title = element_text(size = 16, face = "bold"),
      legend.position = "bottom"
    ) +
```

```
    guides(fill = guide_legend(title = group_var))
  return(p)
}

for (group_var in group_vars) {
  gain_data <- prepare_gain_data(combined_data, variables, group_var)
  plot <- plot_gains(gain_data, group_var)
  ggsave(plot, filename = paste0("gains_by_", group_var, ".png"), width = 14, height = 8)
  file_name <- paste0("plots/gains_by_", group_var, ".png")
  ggsave(file_name, plot, width = 14, height = 8, dpi = 300)
  print(plot)
}
```

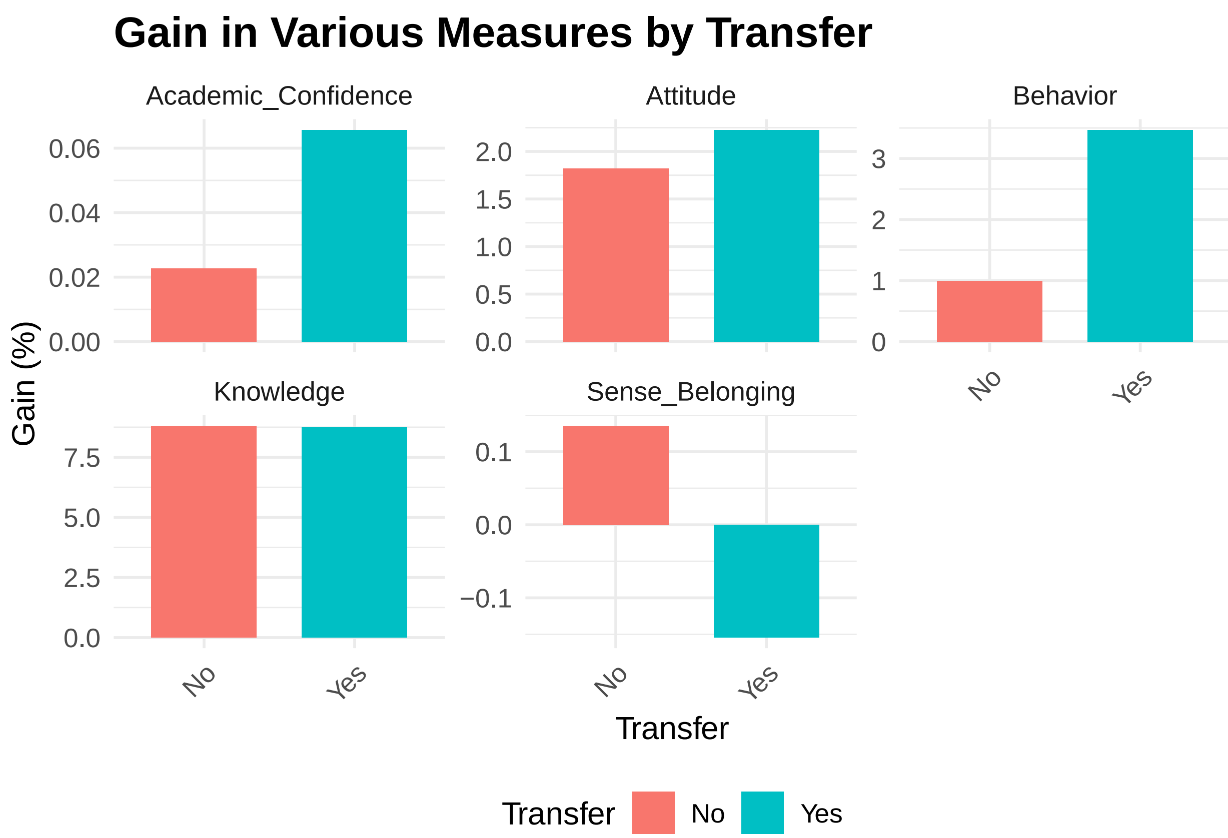
```
## Warning in plot_theme(plot): The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
```



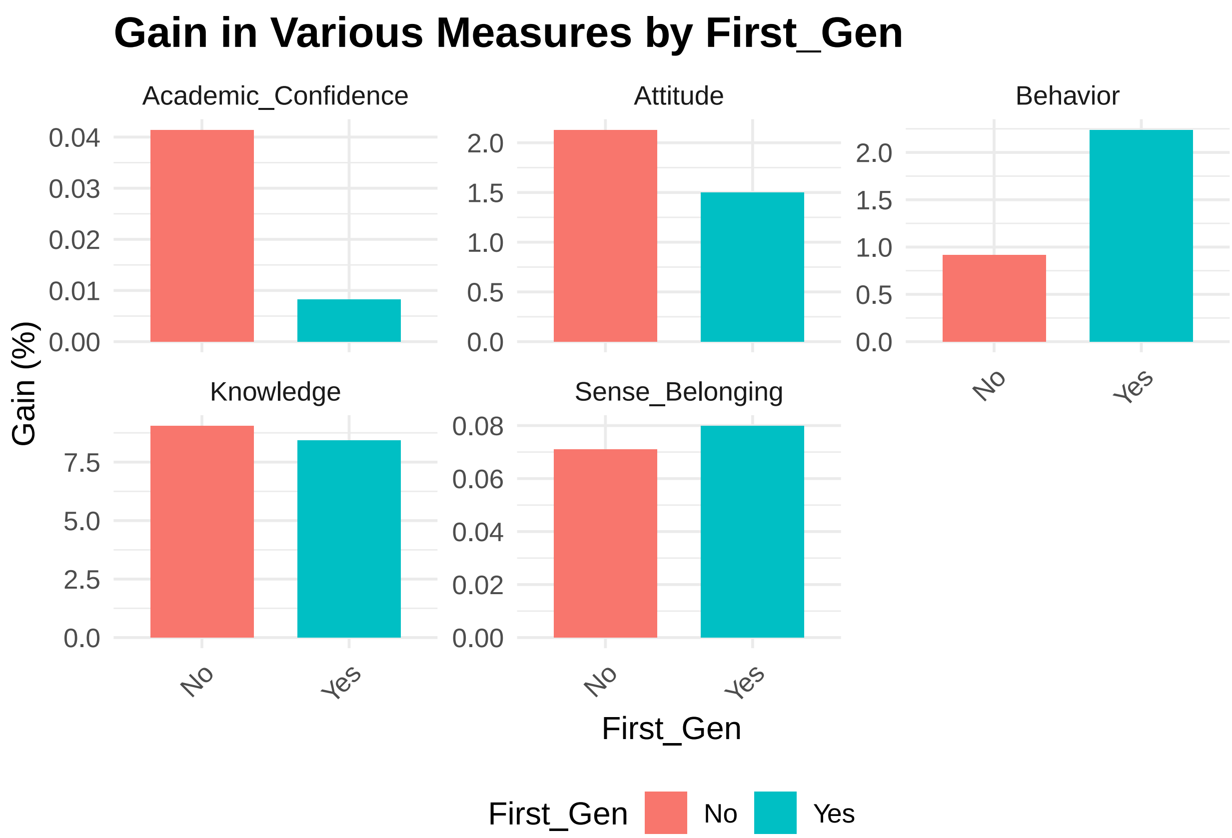

```
## Warning in plot_theme(plot): The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
```



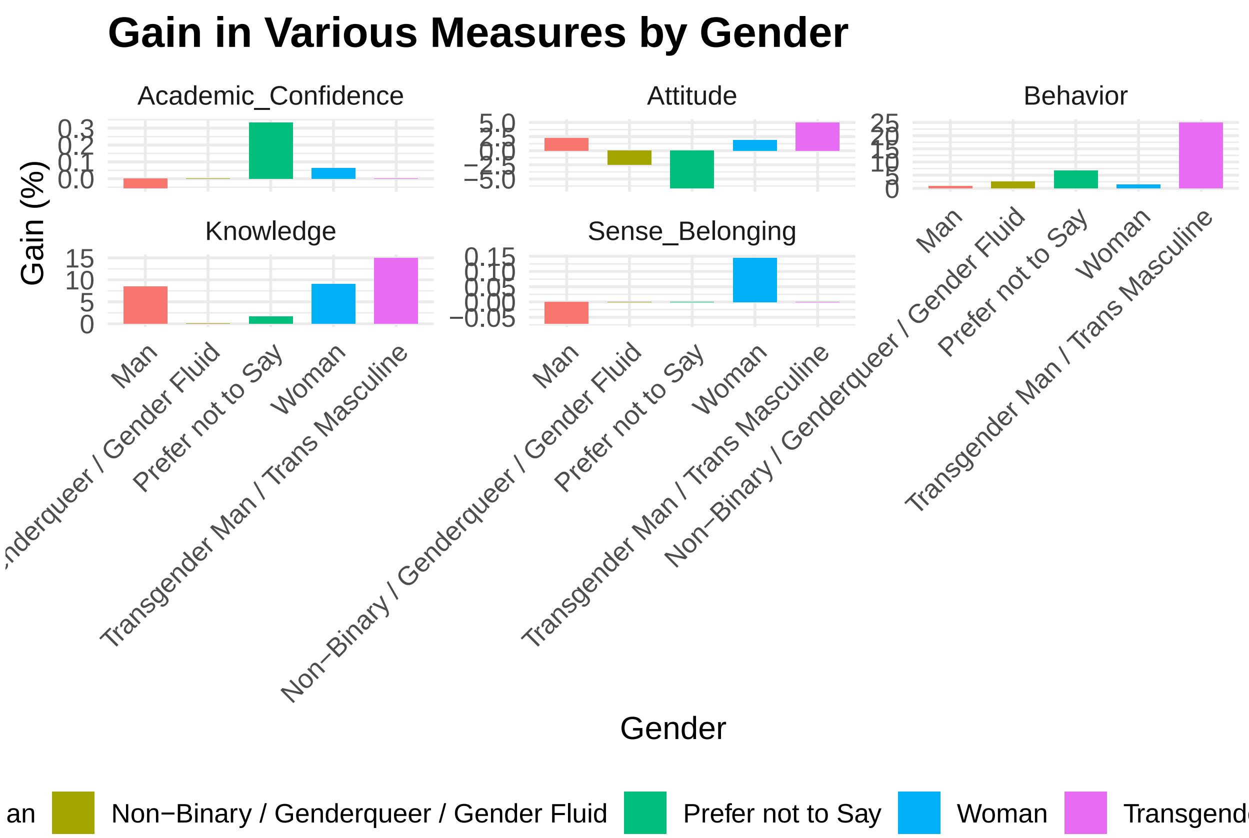

```
## Warning in plot_theme(plot): The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
```




```
## Warning in plot_theme(plot): The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
## The ‘strip_text‘ theme element is not defined in the element hierarchy.
```



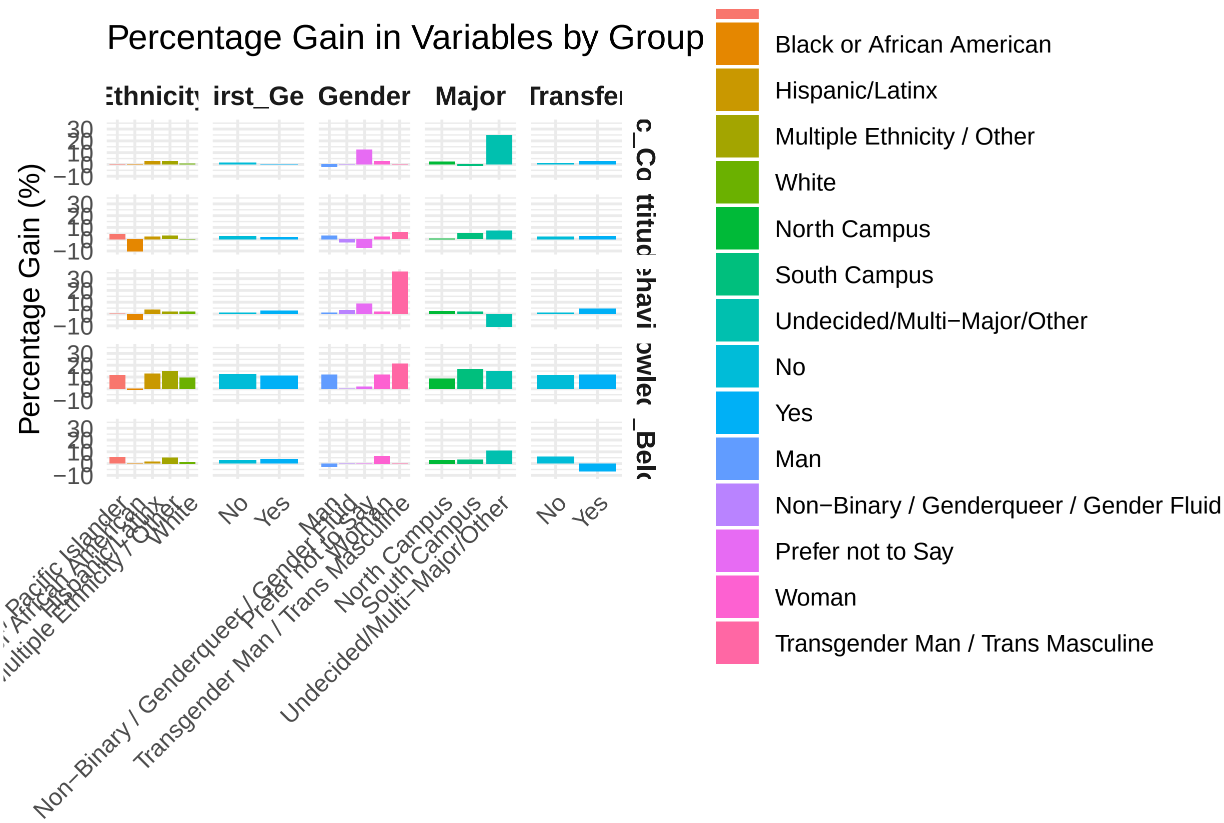

```
## Warning in plot_theme(plot): The ‘strip_text‘ theme element is not defined in
## the element hierarchy.
```



More Visuals

```
calculate_gain_plot <- function(data, variables, group_vars) {
  gains_data <- map_dfr(variables, function(variable) {
    map_dfr(group_vars, function(group_var) {
      data %>%
        group_by(!!sym(group_var), type) %>%
        summarise(mean_value = mean(!!sym(variable), na.rm = TRUE), .groups = 'drop') %>%
        spread(type, mean_value) %>%
        mutate(gain = (`post` - `pre`) / `pre` * 100,
               variable = variable,
               group_var = group_var,
               group_label = !!sym(group_var))
    })
  })

  gains_plot <- gains_data %>%
    ggplot(aes(x = group_label, y = gain, fill = group_label)) +
    geom_bar(stat = "identity", position = position_dodge()) +
    facet_grid(vars(variable), vars(group_var), scales = "free_x") +
    labs(title = "Percentage Gain in Variables by Group", y = "Percentage Gain (%)", x = "") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1),
          strip.background = element_blank(),
          strip.text.x = element_text(size = 10, face = "bold"),
          strip.text.y = element_text(size = 10, face = "bold"))
  return(gains_plot)
}
```

```r
variables <- c("Knowledge", "Attitude", "Behavior", "Sense_Belonging", "Academic_Confidence")
group_vars <- c("Ethnicity", "Major", "Transfer", "First_Gen", "Gender")

final_plot <- calculate_gain_plot(combined_data, variables, group_vars)
final_plot
```




```r
ggsave("final_gain_plot.png", plot = final_plot, width = 20, height = 15, dpi = 300)
```

```r
combined_data <- combined_data %>%
  select(ID, type, group, Ethnicity, Major, Transfer, First_Gen, Gender, Knowledge, Attitude, Behavior,
  mutate(type = factor(type, levels = c("pre", "post")))

library(ggplot2)

# Plot for Changes in Knowledge
knowledge_plot <- ggplot(combined_data, aes(x = type, y = Knowledge, fill = type)) +
  geom_boxplot() +
  facet_wrap(~group, scales = "free") +
  labs(title = "Changes in Knowledge", y = "Score", x = "") +
  theme_minimal()
ggsave("knowledge_plot.png", plot = knowledge_plot, width = 10, height = 6, dpi = 300)

# Plot for Changes in Attitude
attitude_plot <- ggplot(combined_data, aes(x = type, y = Attitude, fill = type)) +
  geom_boxplot() +
  facet_wrap(~group, scales = "free") +
  labs(title = "Changes in Attitude", y = "Score", x = "") +
```

```r
  theme_minimal()
ggsave("attitude_plot.png", plot = attitude_plot, width = 10, height = 6, dpi = 300)

# Plot for Changes in Behavior
behavior_plot <- ggplot(combined_data, aes(x = type, y = Behavior, fill = type)) +
  geom_boxplot() +
  facet_wrap(~group, scales = "free") +
  labs(title = "Changes in Behavior", y = "Score", x = "") +
  theme_minimal()
ggsave("behavior_plot.png", plot = behavior_plot, width = 10, height = 6, dpi = 300)

# Plot for Changes in Academic Confidence
academic_confidence_plot <- ggplot(combined_data, aes(x = type, y = Academic_Confidence, fill = type)) +
  geom_boxplot() +
  facet_wrap(~group, scales = "free") +
  labs(title = "Changes in Academic Confidence", y = "Score", x = "") +
  theme_minimal()
ggsave("academic_confidence_plot.png", plot = academic_confidence_plot, width = 10, height = 6, dpi = 3

# Plot for Changes in Sense of Belonging
sense_belonging_plot <- ggplot(combined_data, aes(x = type, y = Sense_Belonging, fill = type)) +
  geom_boxplot() +
  facet_wrap(~group, scales = "free") +
  labs(title = "Changes in Sense of Belonging", y = "Score", x = "") +
  theme_minimal()
ggsave("sense_belonging_plot.png", plot = sense_belonging_plot, width = 10, height = 6, dpi = 300)
```

Linear Mixed-Effects Models to assess changes

```r
# Model for Knowledge
knowledge_model <- lmer(Knowledge ~ type + (1 | ID) , data = combined_data)
summary(knowledge_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Knowledge ~ type + (1 | ID)
##    Data: combined_data
##
## REML criterion at convergence: 4361.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3083 -0.4940  0.0438  0.6282  2.5585
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 65.41    8.087
##  Residual             91.21    9.550
## Number of obs: 560, groups:  ID, 287
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  74.0332     0.7449 478.4960   99.39   <2e-16 ***
```

```
## typepost      8.7055     0.8139 283.6961   10.70   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## typepost -0.538
```

```
# Model for Attitude
attitude_model <- lmer(Attitude ~ type + (1 | ID), data = combined_data)
summary(attitude_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Attitude ~ type + (1 | ID)
##    Data: combined_data
##
## REML criterion at convergence: 4329.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9216 -0.4016  0.1140  0.4915  2.1277
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 95.60    9.777
##  Residual             69.07    8.311
## Number of obs: 560, groups:  ID, 287
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  81.2744     0.7628 425.0058 106.547   <2e-16 ***
## typepost      1.8387     0.7106 281.8020   2.588   0.0102 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## typepost -0.458
```

```
# Model for Behavior
behavior_model <- lmer(Behavior ~ type + (1 | ID), data = combined_data)
summary(behavior_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Behavior ~ type + (1 | ID)
##    Data: combined_data
##
## REML criterion at convergence: 4379
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.88690 -0.51517  0.04107  0.57908  2.41411
##
## Random effects:
```

```
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 86.32    9.291
##  Residual             83.86    9.157
## Number of obs: 560, groups:  ID, 287
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  80.8931     0.7760 449.9296 104.246   <2e-16 ***
## typepost      1.4569     0.7818 283.1348   1.863   0.0634 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## typepost -0.496
```

Knowledge: The p-value for this model is <2e-16. This value is much smaller than the standard 0.05 threshold, which means that participating in civic engagement courses has a very statistically significant effect on students' knowledge of approaches to community engagement (The most significant out of the 3 clusters). Attitude: The p-value for this model is 0.0102. This also is much smaller than the standard 0.05 threshold, indicating statistical significance, although it's closer to the threshold. This means that participating in civic engagement courses does have a statistically significant effect on students' attitude toward community engagement but less than their knowledge. Behavior: The p-value for this model is 0.0634. This is above the common threshold for significance, suggesting that this result may not be statistically significant. This means that participating in civic engagement courses does not have a statistically significant effect on students' behavior toward community engagement.

Linear Mixed-Effects Models to assess changes 2

```
# Model for Knowledge
knowledge_model_2 <- lmer(Knowledge ~ type + (1 | ID) + Major + Transfer + First_Gen, data = combined_da
summary(knowledge_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Knowledge ~ type + (1 | ID)
##    Data: combined_data
##
## REML criterion at convergence: 4361.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3083 -0.4940  0.0438  0.6282  2.5585
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 65.41    8.087
##  Residual             91.21    9.550
## Number of obs: 560, groups:  ID, 287
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  74.0332     0.7449 478.4960   99.39   <2e-16 ***
## typepost      8.7055     0.8139 283.6961   10.70   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## typepost -0.538
```

```r
# Model for Attitude
attitude_model_2 <- lmer(Attitude ~ type + (1 | ID) + Major + Transfer + First_Gen, data = combined_data
summary(attitude_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Attitude ~ type + (1 | ID)
##    Data: combined_data
##
## REML criterion at convergence: 4329.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9216 -0.4016  0.1140  0.4915  2.1277
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 95.60    9.777
##  Residual             69.07    8.311
## Number of obs: 560, groups:  ID, 287
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  81.2744     0.7628 425.0058 106.547   <2e-16 ***
## typepost      1.8387     0.7106 281.8020   2.588   0.0102 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## typepost -0.458
```

```r
# Model for Behavior
behavior_model_2 <- lmer(Behavior ~ type + (1 | ID) + Major + Transfer + First_Gen, data = combined_data
summary(behavior_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Behavior ~ type + (1 | ID)
##    Data: combined_data
##
## REML criterion at convergence: 4379
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.88690 -0.51517  0.04107  0.57908  2.41411
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 86.32    9.291
```

```
##  Residual             83.86    9.157
## Number of obs: 560, groups:  ID, 287
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  80.8931     0.7760 449.9296 104.246   <2e-16 ***
## typepost      1.4569     0.7818 283.1348   1.863   0.0634 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## typepost -0.496
```

```
# Checking frequency tables for Ethnicity, Gender, and Major
table(combined_data$Ethnicity)
```

```
##
##   Asian / Pacific Islander  Black or African American
##                        196                         12
##            Hispanic/Latinx Multiple Ethnicity / Other
##                        170                         80
##                      White
##                        102
```

```
table(combined_data$Gender)
```

```
##
##                                    Man Non-Binary / Genderqueer / Gender Fluid
##                                    157                                     4
##                      Prefer not to Say                                 Woman
##                                      5                                   392
##       Transgender Man / Trans Masculine
##                                      2
```

```
table(combined_data$Major)
```

```
##
##                North Campus                South Campus
##                         324                         227
## Undecided/Multi-Major/Other
##                           9
```

```
# Removing non-Man and non-Woman observations for Gender to ensure proper distribution
combined_data <- combined_data %>%
  filter(Gender %in% c("Man", "Woman"))

# Removing Undecided/Multi-Major/Other observations for Major to ensure proper distribution
combined_data <- combined_data %>%
  filter(Major != "Undecided/Multi-Major/Other")

# Combining African Americans with Multiple Ethnicity / Other for Ethnicity to ensure proper distributi
combined_data <- combined_data %>%
  mutate(Ethnicity = case_when(
    Ethnicity == "Black or African American" ~ "Multiple Ethnicity / Other",
    TRUE ~ Ethnicity
  ))
```

```r
combined_data <- combined_data %>%
  mutate(across(c(Gender, Major), droplevels))

# updated frequency tables
table(combined_data$Ethnicity)
```

```
##
##   Asian / Pacific Islander            Hispanic/Latinx
##                        192                        165
## Multiple Ethnicity / Other                      White
##                         86                         97
```

```r
table(combined_data$Gender)
```

```
##
##   Man Woman
##   153   387
```

```r
table(combined_data$Major)
```

```
##
## North Campus South Campus
##          315          225
```

```r
# Linear Mixed-Effects Models with additional predictors

# Model for Knowledge
knowledge_model_3 <- lmer(Knowledge ~ type + Ethnicity + Gender + Major + (1 | ID), data = combined_data
summary(knowledge_model_3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Knowledge ~ type + Ethnicity + Gender + Major + (1 | ID)
##    Data: combined_data
##
## REML criterion at convergence: 4172
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5283 -0.5007  0.0202  0.6359  2.4193
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 55.55    7.453
##  Residual             93.33    9.661
## Number of obs: 540, groups:  ID, 280
##
## Fixed effects:
##                                     Estimate Std. Error       df t value
## (Intercept)                          71.1081     1.5605 316.7912  45.569
## typepost                              8.7614     0.8403 270.9635  10.426
## EthnicityHispanic/Latinx              0.6182     1.5753 277.2070   0.392
## EthnicityMultiple Ethnicity / Other   2.0243     1.8155 306.2842   1.115
## EthnicityWhite                        4.3760     1.7807 274.2120   2.457
## GenderWoman                           4.0804     1.3828 269.3320   2.951
## MajorSouth Campus                    -3.3435     1.2606 299.5348  -2.652
```

```
##                                     Pr(>|t|)
## (Intercept)                          < 2e-16 ***
## typepost                             < 2e-16 ***
## EthnicityHispanic/Latinx             0.69506
## EthnicityMultiple Ethnicity / Other  0.26571
## EthnicityWhite                       0.01461 *
## GenderWoman                          0.00345 **
## MajorSouth Campus                    0.00842 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) typpst EthH/L EtME/O EthncW GndrWm
## typepost    -0.277
## EthnctyHs/L -0.434  0.011
## EthnctyME/O -0.377  0.031  0.401
## EthnictyWht -0.363 -0.012  0.406  0.344
## GenderWoman -0.550  0.007 -0.184 -0.053 -0.078
## MajrSthCmps -0.432 -0.002  0.239  0.068  0.072 -0.008
```

```
# Model for Attitude
attitude_model_3 <- lmer(Attitude ~ type + Ethnicity + Gender + Major + (1 | ID), data = combined_data)
summary(attitude_model_3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Attitude ~ type + Ethnicity + Gender + Major + (1 | ID)
##    Data: combined_data
##
## REML criterion at convergence: 4139.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9804 -0.3967  0.1083  0.4637  2.2533
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 82.78    9.098
##  Residual             71.30    8.444
## Number of obs: 540, groups:  ID, 280
##
## Fixed effects:
##                                     Estimate Std. Error       df t value
## (Intercept)                          75.2631     1.6483 313.5647  45.660
## typepost                              1.9728     0.7378 271.9260   2.674
## EthnicityHispanic/Latinx              1.6817     1.6826 287.9709   0.999
## EthnicityMultiple Ethnicity / Other   1.2422     1.9165 327.5815   0.648
## EthnicityWhite                        3.4990     1.9030 288.0070   1.839
## GenderWoman                           7.4132     1.4824 275.2990   5.001
## MajorSouth Campus                    -2.0980     1.3332 323.4717  -1.574
##                                     Pr(>|t|)
## (Intercept)                          < 2e-16 ***
## typepost                             0.00796 **
## EthnicityHispanic/Latinx             0.31840
## EthnicityMultiple Ethnicity / Other  0.51733
```

```
## EthnicityWhite                          0.06699 .
## GenderWoman                           1.02e-06 ***
## MajorSouth Campus                       0.11655
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) typpst EthH/L EtME/O EthncW GndrWm
## typepost    -0.231
## EthnctyHs/L -0.441  0.012
## EthnctyME/O -0.386  0.033  0.414
## EthnictyWht -0.367 -0.014  0.408  0.359
## GenderWoman -0.558  0.006 -0.182 -0.054 -0.080
## MajrSthCmps -0.433 -0.002  0.239  0.072  0.071 -0.008
```

```r
# Model for Behavior
behavior_model_3 <- lmer(Behavior ~ type + Ethnicity + Gender + Major + (1 | ID), data = combined_data)
summary(behavior_model_3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Behavior ~ type + Ethnicity + Gender + Major + (1 | ID)
##    Data: combined_data
##
## REML criterion at convergence: 4170.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.97336 -0.57608  0.09155  0.54683  2.45271
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 78.44    8.856
##  Residual             79.93    8.941
## Number of obs: 540, groups:  ID, 280
##
## Fixed effects:
##                                       Estimate Std. Error       df t value
## (Intercept)                            75.1144     1.6562 314.3881  45.354
## typepost                                1.4877     0.7803 272.0948   1.907
## EthnicityHispanic/Latinx                2.9672     1.6862 285.3888   1.760
## EthnicityMultiple Ethnicity / Other     4.8009     1.9275 321.7439   2.491
## EthnicityWhite                          8.8572     1.9070 284.5619   4.645
## GenderWoman                             4.4167     1.4839 274.2154   2.977
## MajorSouth Campus                      -1.3470     1.3403 316.7204  -1.005
##                                       Pr(>|t|)
## (Intercept)                            < 2e-16 ***
## typepost                               0.05764 .
## EthnicityHispanic/Latinx               0.07953 .
## EthnicityMultiple Ethnicity / Other    0.01325 *
## EthnicityWhite                        5.21e-06 ***
## GenderWoman                            0.00318 **
## MajorSouth Campus                      0.31565
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation of Fixed Effects:
##             (Intr) typpst EthH/L EtME/O EthncW GndrWm
## typepost    -0.243
## EthnctyHs/L -0.439  0.012
## EthnctyME/O -0.384  0.032  0.410
## EthnictyWht -0.366 -0.014  0.408  0.354
## GenderWoman -0.556  0.006 -0.182 -0.053 -0.080
## MajrSthCmps -0.433 -0.002  0.239  0.071  0.072 -0.008
```

```r
# Model for Sense of Belonging
sense_belonging_model <- lmer(Sense_Belonging ~ type + Ethnicity + Gender + Major + (1 | ID), data = com
summary(sense_belonging_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Sense_Belonging ~ type + Ethnicity + Gender + Major + (1 | ID)
##    Data: combined_data
##
## REML criterion at convergence: 964.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.94959 -0.33836 -0.04221  0.46709  2.19878
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept) 0.2349   0.4847
##  Residual             0.1751   0.4184
## Number of obs: 540, groups:  ID, 280
##
## Fixed effects:
##                                     Estimate Std. Error         df t value
## (Intercept)                         2.365096   0.085652 308.036823  27.613
## typepost                            0.062283   0.036600 266.565529   1.702
## EthnicityHispanic/Latinx           -0.196433   0.087612 285.185194  -2.242
## EthnicityMultiple Ethnicity / Other  0.003544   0.099452 328.230990   0.036
## EthnicityWhite                      0.176261   0.099084 286.029193   1.779
## GenderWoman                        -0.020024   0.077286 271.002105  -0.259
## MajorSouth Campus                   0.010584   0.069203 325.089274   0.153
##                                   Pr(>|t|)
## (Intercept)                         <2e-16 ***
## typepost                            0.0900 .
## EthnicityHispanic/Latinx            0.0257 *
## EthnicityMultiple Ethnicity / Other  0.9716
## EthnicityWhite                      0.0763 .
## GenderWoman                         0.7958
## MajorSouth Campus                   0.8785
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) typpst EthH/L EtME/O EthncW GndrWm
## typepost    -0.221
## EthnctyHs/L -0.442  0.013
```

```
## EthnctyME/O -0.388  0.033  0.418
## EthnictyWht -0.369 -0.015  0.409  0.364
## GenderWoman -0.560  0.006 -0.181 -0.054 -0.081
## MajrSthCmps -0.433 -0.003  0.239  0.073  0.071 -0.008
```

Interpretation of the Results

Knowledge Model

- **Pre/Post** The post-class scores are significantly higher by 8.76 points ($p < 2e - 16$).
- **Ethnicity:** Compared to the baseline (likely Asian / Pacific Islander):
  - Hispanic/Latinx: Not significantly different ($p = 0.695$).
  - Multiple Ethnicity / Other: Not significantly different ($p = 0.266$).
  - White: Significantly higher by 4.38 points ($p = 0.015$).
- **Gender:** Women have significantly higher scores by 4.08 points ($p = 0.003$).
- **North vs South Campus:** Students from the South Campus have significantly lower scores by 3.34 points ($p = 0.008$).

Attitude Model

- **(Intercept):** The baseline Attitude score is 75.26.
- **Pre/Post:** The post-class scores are significantly higher by 1.97 points ($p = 0.008$).
- **Ethnicity:** Compared to the baseline (likely Asian / Pacific Islander):
  - Hispanic/Latinx: Not significantly different ($p = 0.318$).
  - Multiple Ethnicity / Other: Not significantly different ($p = 0.517$).
  - White: Marginally higher by 3.50 points but still not significantly different ($p = 0.067$).
- **Gender:** Women have significantly higher scores by 7.41 points ($p = 1.02e - 06$).
- **North vs South Campus:** Not significantly different ($p = 0.117$).

Behavior Model

- **(Intercept):** The baseline Behavior score is 75.11.
- **Pre/Post:** The post-class scores are marginally higher by 1.49 points ($p = 0.058$).
- **Ethnicity:** Compared to the baseline (likely Asian / Pacific Islander):
  - Hispanic/Latinx: Marginally higher by 2.97 points ($p = 0.080$).
  - Multiple Ethnicity / Other: Significantly higher by 4.80 points ($p = 0.013$).
  - White: Significantly higher by 8.86 points ($p = 5.21e - 06$).
- **Gender:** Women have significantly higher scores by 4.42 points ($p = 0.003$).
- **North vs South Campus:** Not significantly different ($p = 0.316$).

Sense of Belonging Model

- **(Intercept):** The baseline Sense of Belonging score is 2.37.
- **Pre/Post:** The post-class scores are marginally higher by 0.06 points ($p = 0.090$).
- **Ethnicity:** Compared to the baseline (likely Asian / Pacific Islander):
  - Hispanic/Latinx: Significantly lower by 0.20 points ($p = 0.026$).
  - Multiple Ethnicity / Other: Not significantly different ($p = 0.972$).
  - White: Marginally higher by 0.18 points ($p = 0.076$).
- **Gender:** Not significantly different ($p = 0.796$).
- **North vs South Campus:** Not significantly different ($p = 0.879$).

Summary Conclusions: **Knowledge Model**: - Post-class scores are significantly higher. - White students and women have significantly higher scores. - Students from the South Campus have significantly lower scores than students from the North Campus.

Attitude Model:

- Post-class scores are significantly higher.
- Women have significantly higher scores.
- White students have marginally higher scores (still not meeting the threshold for statistical significance but close).

Behavior Model:

- Post-class scores are marginally higher.
- White students and women have significantly higher scores.
- Students identifying as Multiple Ethnicity / Other have significantly higher scores.

Sense of Belonging Model:

- Post-class scores are marginally higher.
- Hispanic/Latinx students have significantly lower scores.

Final Conclusions

- The civics study class has a relatively positive impact on Knowledge, Attitude, and Behavior scores, particularly for Knowledge.
- Gender differences are notable, with women showing higher improvements across Knowledge, Attitude, and Behavior.
- Ethnicity differences suggest that White students benefit more in Knowledge and Behavior, while Hispanic/Latinx students show a decrease in Sense of Belonging.