Visual Analysis

Anahit Baghdasaryan

2023-12-04

Data Preprocessing

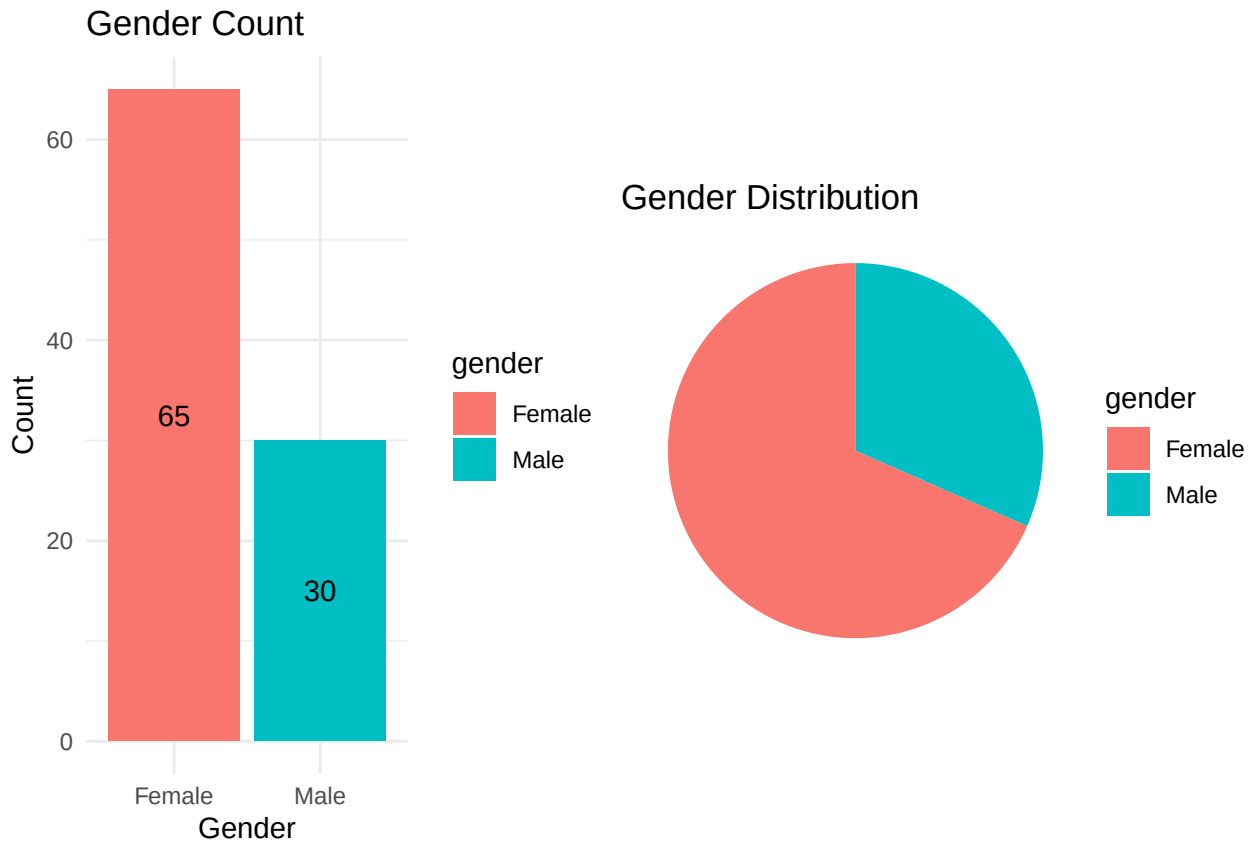
```r
data <- read.csv("data_tidy.csv")
#Uncomment to see
#head(data)
```

```r
# List of columns to exclude from conversion
columns_to_exclude <- c("id", "timestamp", "field_of_study", "university", "gpa")

# List of columns to convert to categorical
columns_to_convert <- setdiff(names(data), columns_to_exclude)

# Convert selected columns to categorical
data[columns_to_convert] <- lapply(data[columns_to_convert], as.factor)

# Verify the changes
#Uncomment to see
#str(data)
```

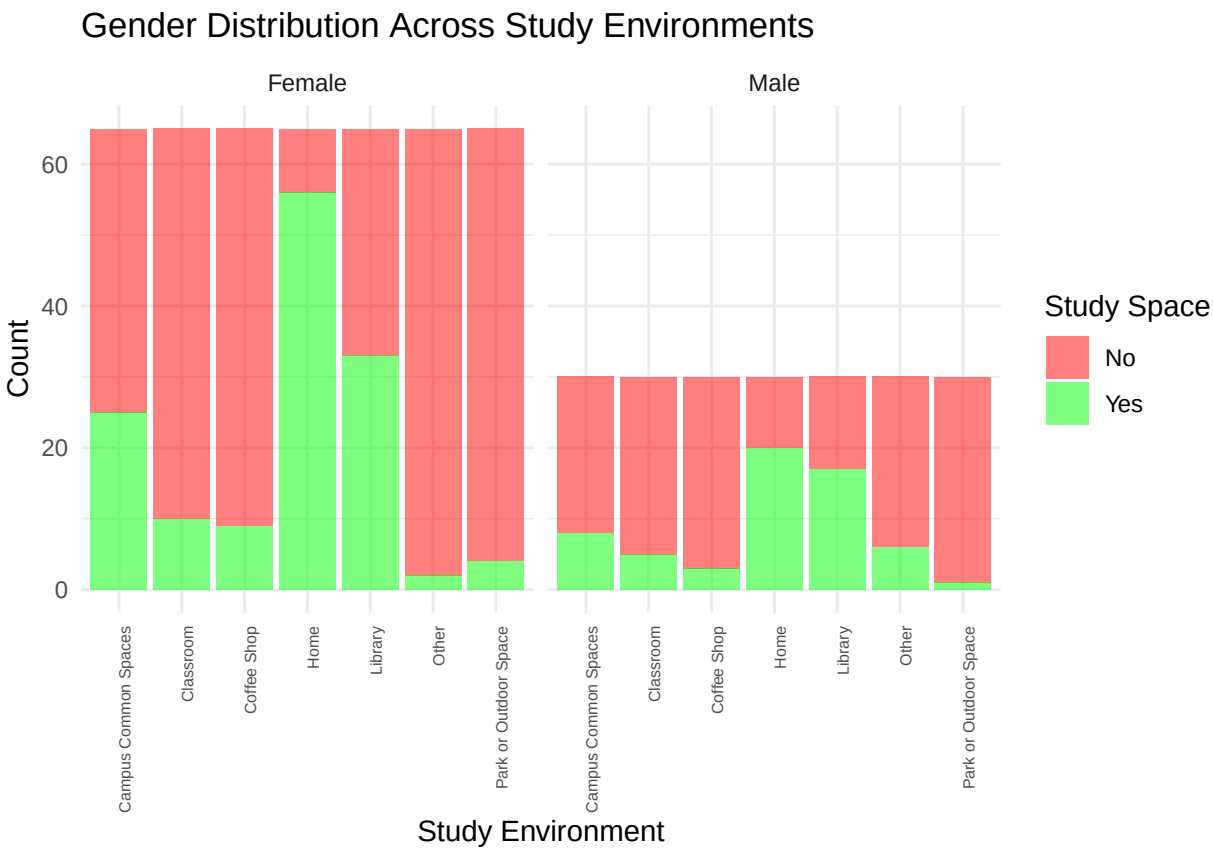
```r
# Summary statistics
#Uncomment to see
#summary(data)
```

Analysis on Gender

```
# Create a countplot of the 'gender' column
countplot <- ggplot(data, aes(x = gender, fill = gender)) +
  geom_bar() +
  geom_text(stat = 'count', aes(label = after_stat(count)),
            position = position_stack(vjust = 0.5)) +
  labs(title = 'Gender Count', x = 'Gender', y = 'Count') +
  theme_minimal()

# Create a pie chart of the 'gender' column
pie_chart <- ggplot(data, aes(x = 1, fill = gender)) +
  geom_bar() +
  coord_polar(theta = 'y') +
  labs(title = 'Gender Distribution') +
  theme_void()

# Display the plots side by side
grid.arrange(countplot, pie_chart, ncol = 2)
```

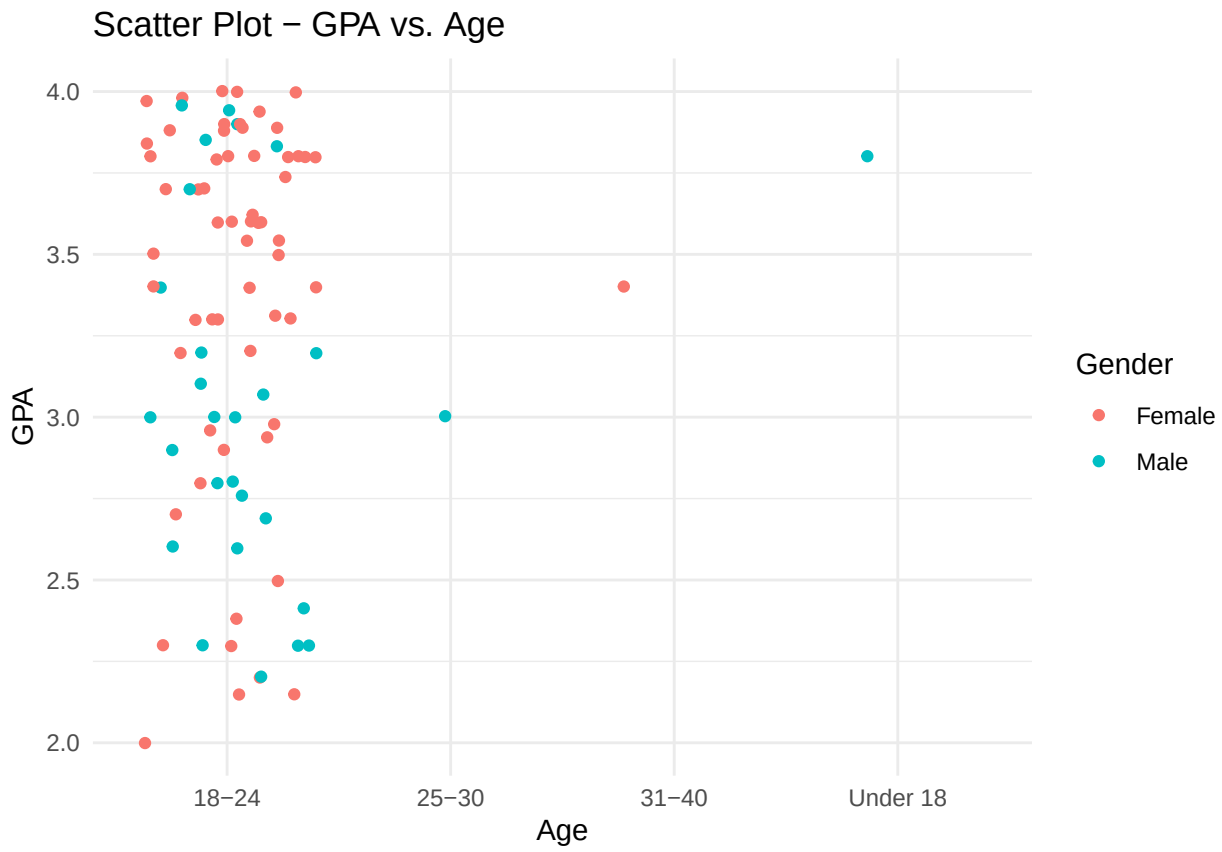


```r
# Selecting relevant columns
study_env_gender <- data[, c("study_env_Campus.Common.Spaces",
                             "study_env_Classroom", "study_env_Coffee.shop",
                             "study_env_Home", "study_env_Library", "study_env_Other",
                             "study_env_Park.or.outdoor.space", "gender")]

# Reshaping data for plotting
study_env_gender_long <- tidyr::gather(study_env_gender,
                                       key = "Study_Environment",
                                       value = "Value", -gender)

# Define labels for Study_Environment
labels <- c(
  "study_env_Campus.Common.Spaces" = "Campus Common Spaces",
  "study_env_Classroom" = "Classroom",
  "study_env_Coffee.shop" = "Coffee Shop",
  "study_env_Home" = "Home",
  "study_env_Library" = "Library",
  "study_env_Other" = "Other",
  "study_env_Park.or.outdoor.space" = "Park or Outdoor Space"
)

# Stacked bar plot for gender distribution across study environments
stacked_bar_plot <- ggplot(study_env_gender_long,
                           aes(x = Study_Environment, fill = Value)) +
  geom_bar(alpha = 0.5) +
  facet_wrap(~ gender) +
  labs(title = "Gender Distribution Across Study Environments",
       x = "Study Environment",
       y = "Count",
       fill = "Study Space") +
  scale_fill_manual(values = c("0" = "red", "1" = "green"),
                    labels = c("0" = "No", "1" = "Yes")) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 6)) +
  scale_x_discrete(labels = labels)  # Set custom labels for Study_Environment

print(stacked_bar_plot)
```

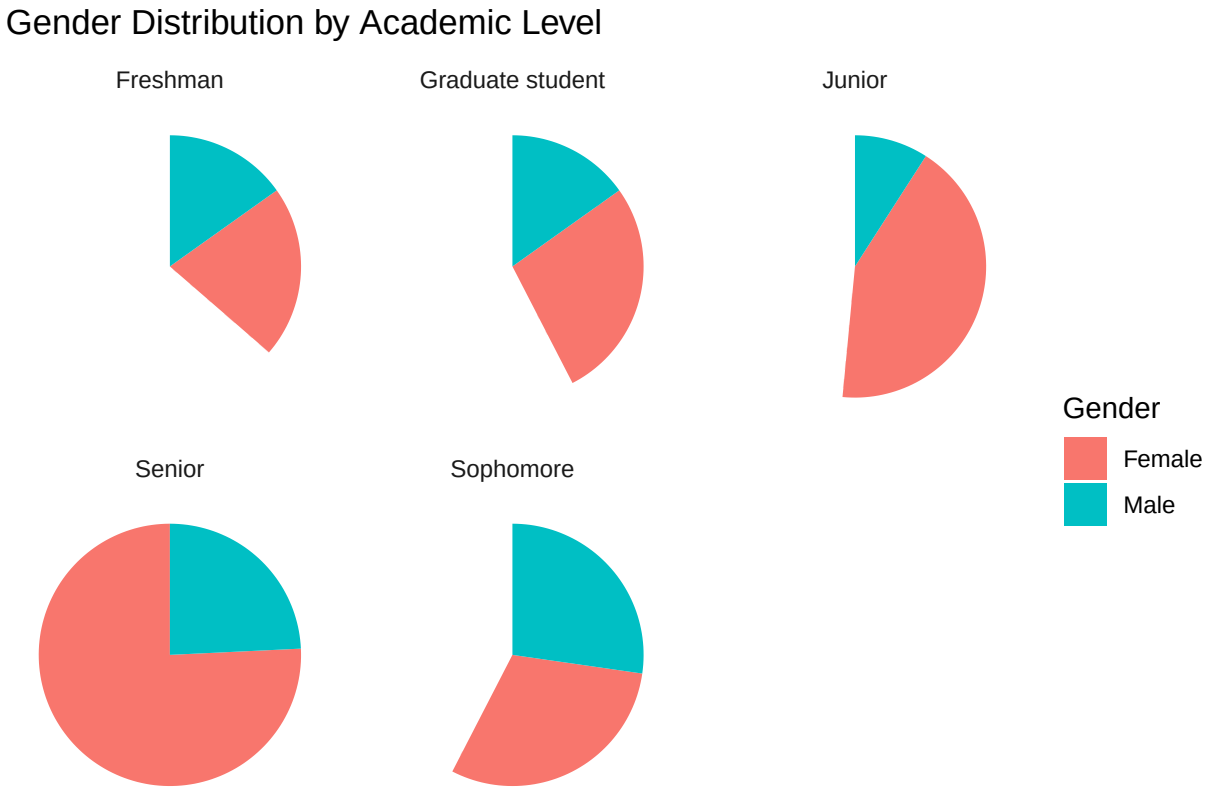

Gender Distribution Across Study Environments
Female
Male
Count
60
40
20
0
Campus Common Spaces
Classroom
Coffee Shop
Home
Library
Other
Park or Outdoor Space
Study Environment
Study Space
No
Yes

```r
# Scatter plot for GPA vs. Age colored by gender with jittering
scatter_plot_gpa_age <- ggplot(data, aes(x = age, y = gpa, color = gender)) +
  geom_jitter() +
  labs(title = "Scatter Plot - GPA vs. Age",
       x = "Age",
       y = "GPA",
       color = "Gender") +
  theme_minimal()

print(scatter_plot_gpa_age)
```

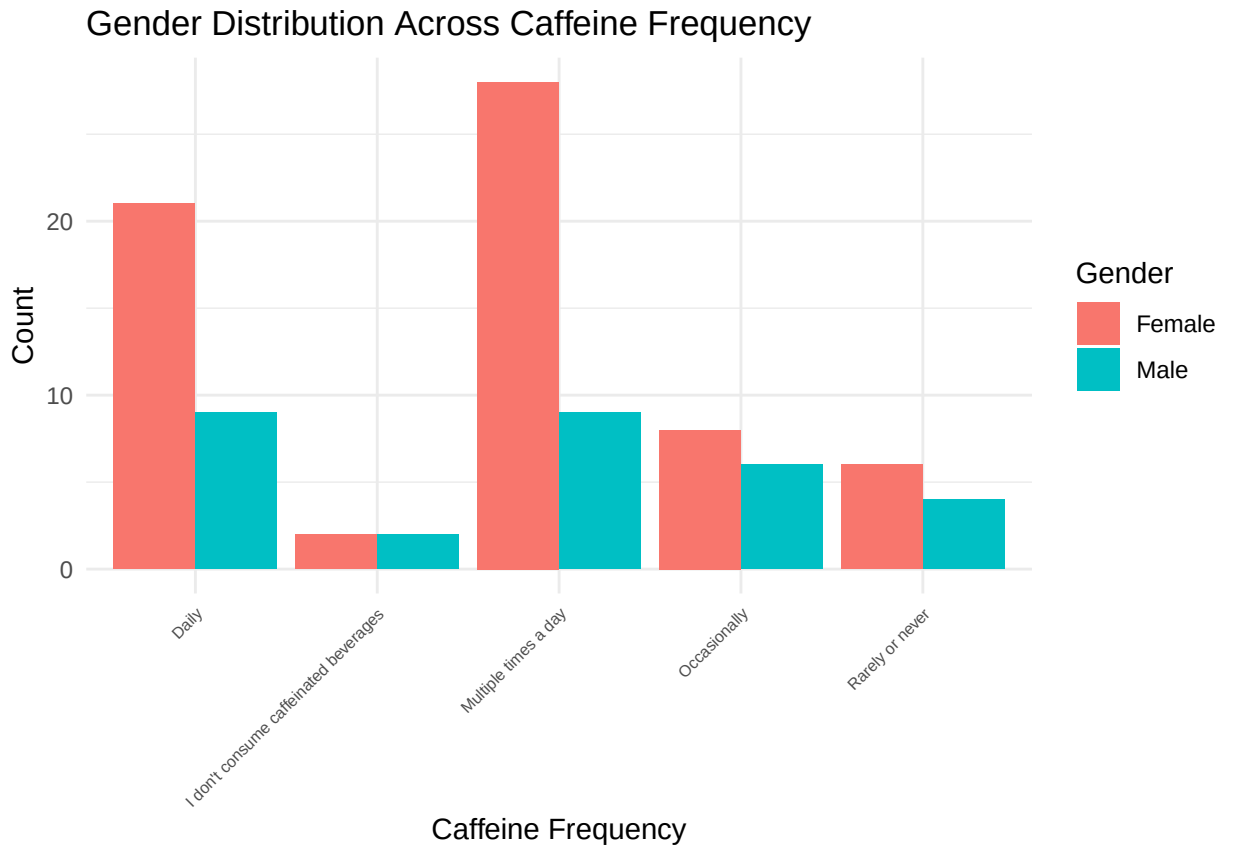
```
## Warning: Removed 7 rows containing missing values (`geom_point()`).
```



```r
# Donut plot for gender distribution by academic level
donut_plot_academic_gender <- ggplot(data, aes(x = "", fill = gender)) +
  geom_bar(stat = "count", width = 1) +
  coord_polar("y") +
  facet_wrap(~ academic_level) +
  labs(title = "Gender Distribution by Academic Level",
       fill = "Gender") +
  theme_minimal() +
  theme(axis.text = element_blank(),
        axis.title = element_blank(),
        panel.grid = element_blank())

print(donut_plot_academic_gender)
```



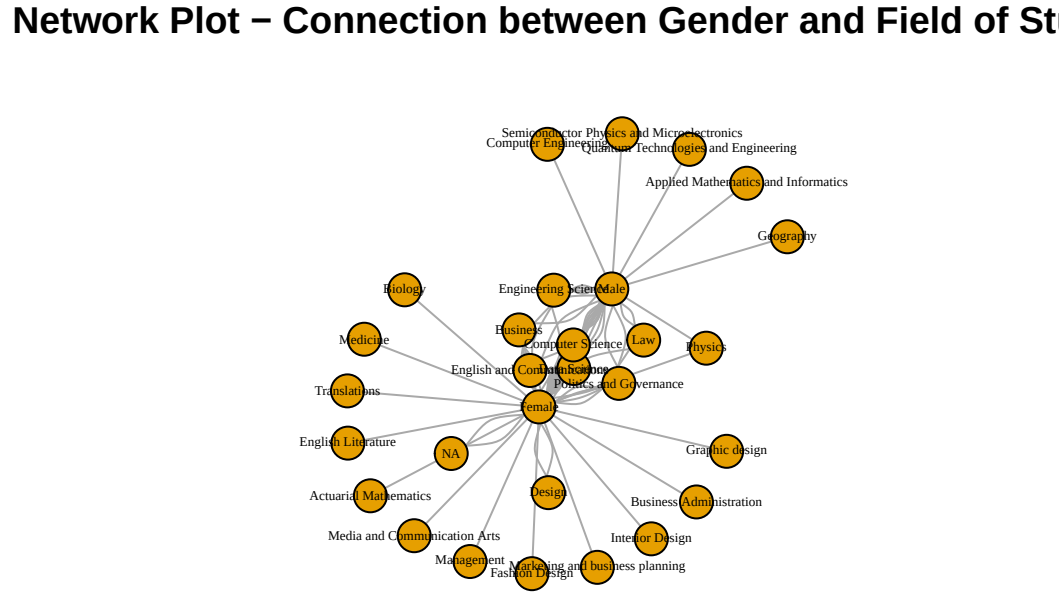
```r
# Grouped bar plot for gender distribution across caffeine frequency
grouped_bar_plot_caffeine_gender <-
  ggplot(data, aes(x = caffeine_freq, fill = gender)) +
  geom_bar(position = "dodge") +
  labs(title = "Gender Distribution Across Caffeine Frequency",
       x = "Caffeine Frequency",
       y = "Count",
       fill = "Gender") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 6))

print(grouped_bar_plot_caffeine_gender)
```



```r
# Create a graph based on the connections between gender and field of study
edges <- data[, c("gender", "field_of_study")]
graph <- graph_from_data_frame(edges, directed = FALSE)
```

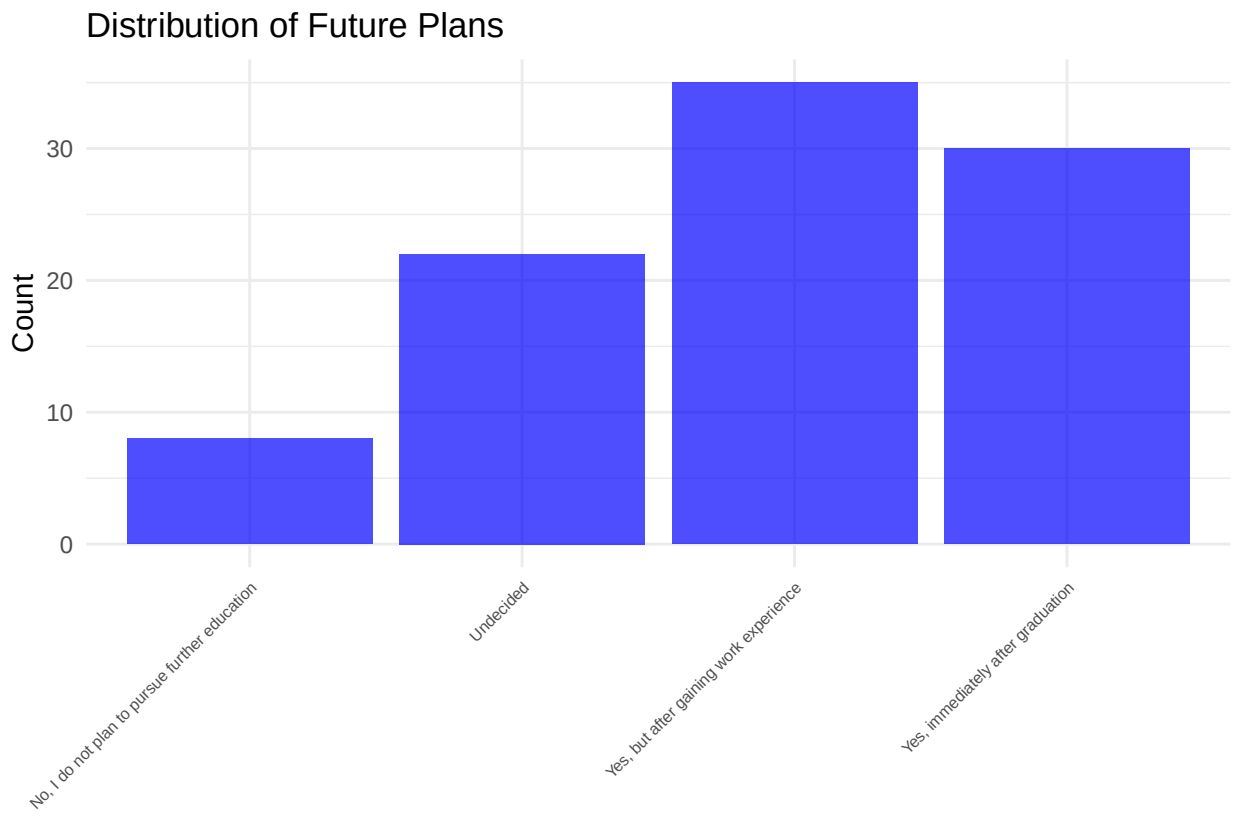
```
## Warning in graph_from_data_frame(edges, directed = FALSE): In 'd' 'NA' elements
## were replaced with string "NA"
```

```r
# Plot the network
set.seed(123)  # Set seed for reproducibility
network_plot <-
  plot(graph, layout = layout_with_fr, edge.arrow.size = 10,
       vertex.label.cex = 0.4, vertex.label.color = "black",
       main = "Network Plot - Connection between Gender and Field of Study")
```

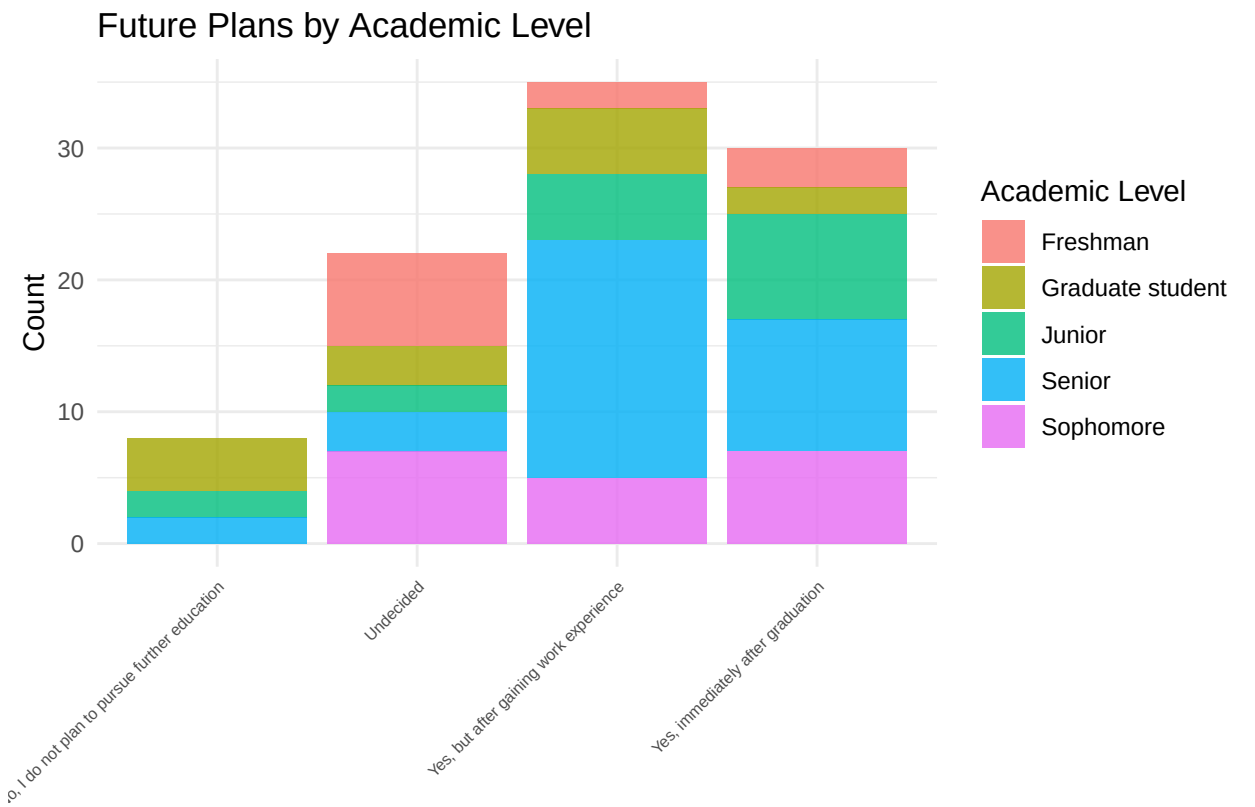


Analysis on Future Plans

```
# Plot the distribution of future plans
ggplot(data, aes(x = future_plans)) +
  geom_bar(fill = "blue", alpha = 0.7) +
  labs(title = "Distribution of Future Plans",
       x = "Do you plan to pursue further education after completing your current academic program?",
       y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 6))
```



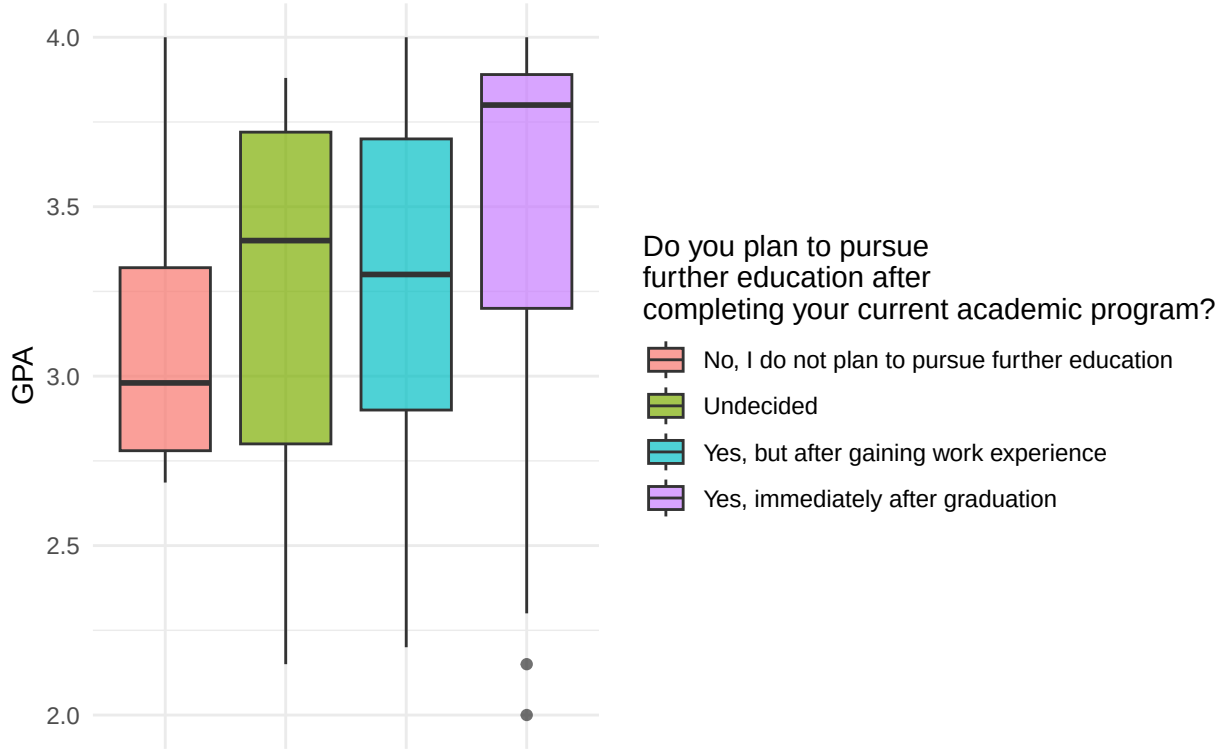
```r
# Create a stacked bar plot of future plans by academic level
ggplot(data, aes(x = future_plans, fill = academic_level)) +
  geom_bar(position = "stack", alpha = 0.8) +
  labs(title = "Future Plans by Academic Level",
       x = "Do you plan to pursue further education after completing your current academic program?",
       y = "Count",
       fill = "Academic Level")+
       theme_minimal() +
       theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 6))
```



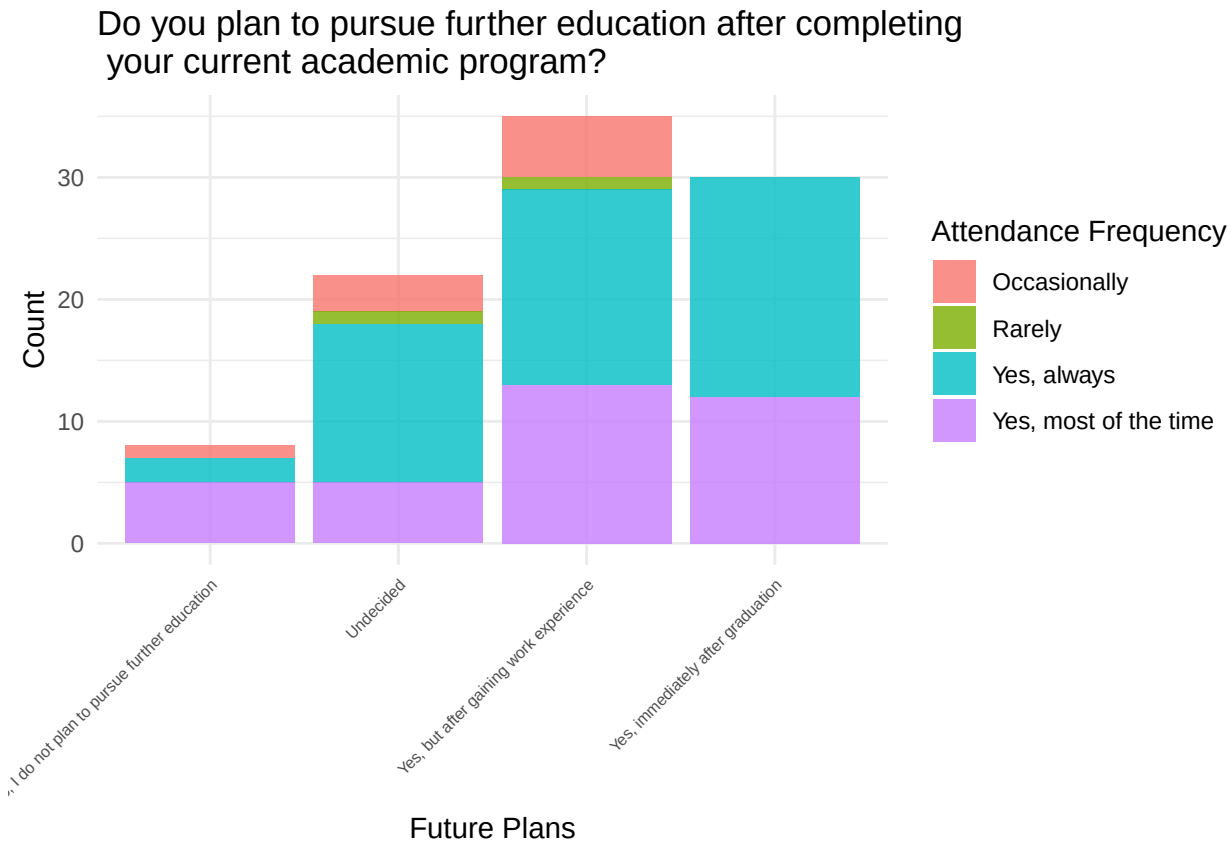
u plan to pursue further education after completing your current academic program?

```r
# Create a boxplot of GPA by future plans
ggplot(data, aes(x = future_plans, y = gpa, fill = future_plans)) +
  geom_boxplot(alpha = 0.7) +
  labs(title = "",
       x = " ",
       y = "GPA",
       fill = "Future Plans")+
  theme_minimal() +
  theme(axis.text.x = element_blank())+
  scale_fill_discrete(name =
                        "Do you plan to pursue \nfurther education after \ncompleting your current acad
```

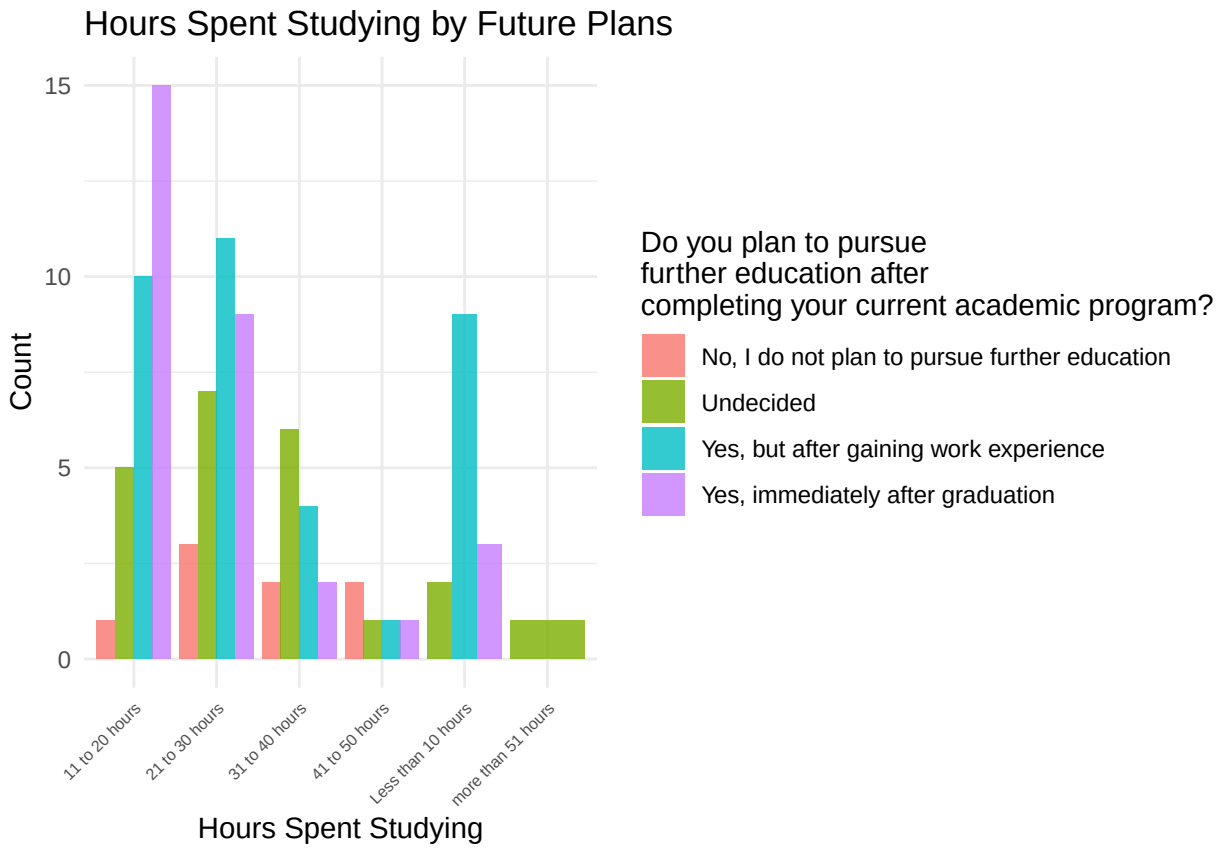
```
## Warning: Removed 7 rows containing non-finite values (‘stat_boxplot()‘).
```



```r
# Create a bar plot of attendance frequency by future plans
ggplot(data, aes(x = future_plans, fill = attendance)) +
  geom_bar(position = "stack", alpha = 0.8) +
  labs(title = "Do you plan to pursue further education after completing\n your current academic progra
       x = "Future Plans",
       y = "Count",
       fill = "Attendance Frequency")+
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 6))
```

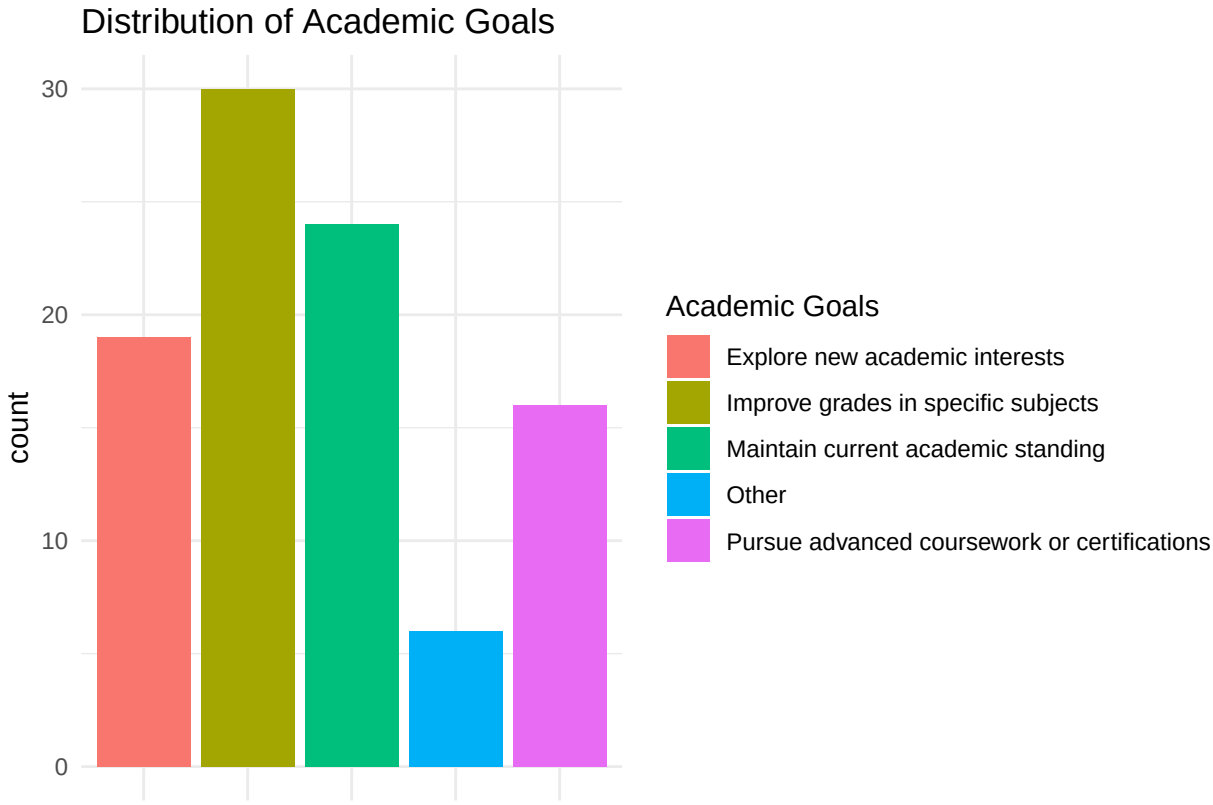


```r
# Create a histogram of hours spent studying by future plans
ggplot(data, aes(x = hours_spend_studying, fill = future_plans)) +
  geom_bar(position = "dodge", alpha = 0.8) +
  labs(title = "Hours Spent Studying by Future Plans",
       x = "Hours Spent Studying",
       y = "Count",
       fill = "Future Plans") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 6))+
  scale_fill_discrete(name =
                        "Do you plan to pursue \nfurther education after \ncompleting your current acad
```

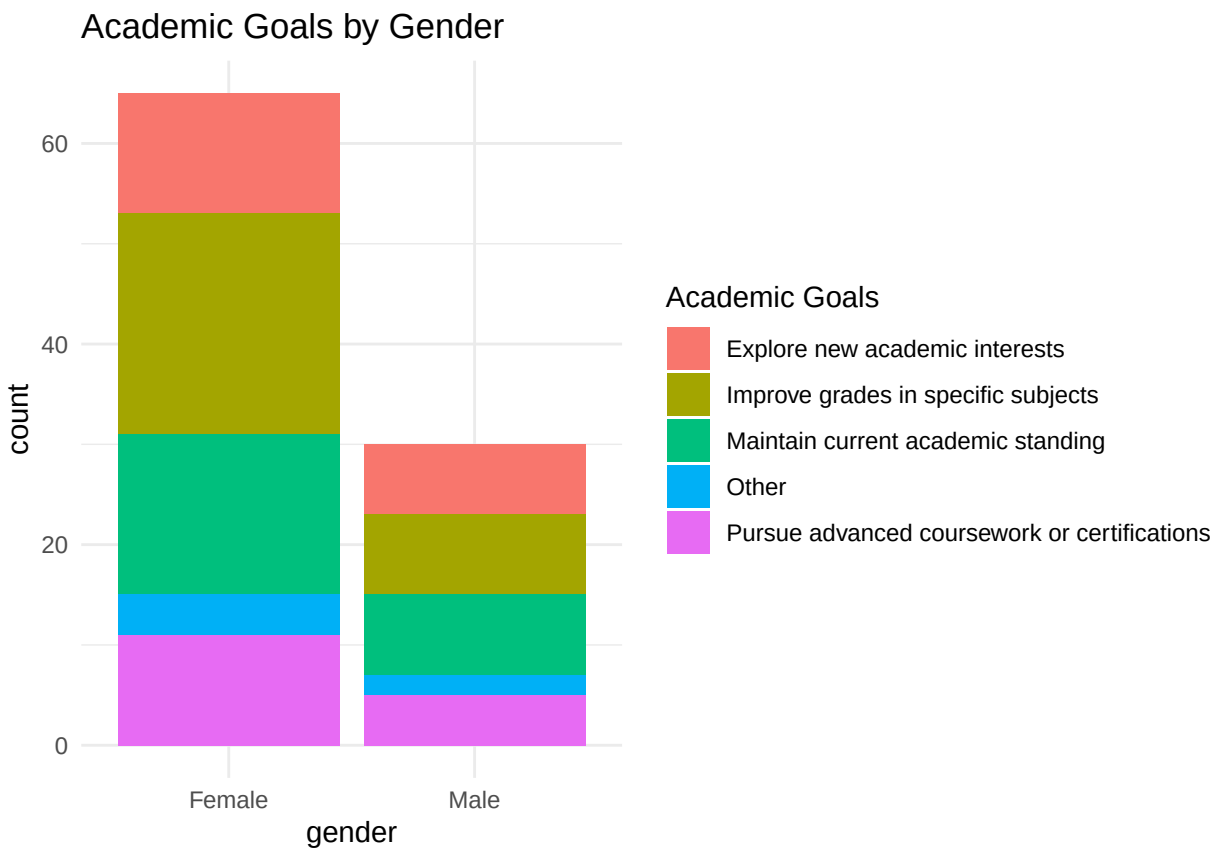


Academic Goals Analysis

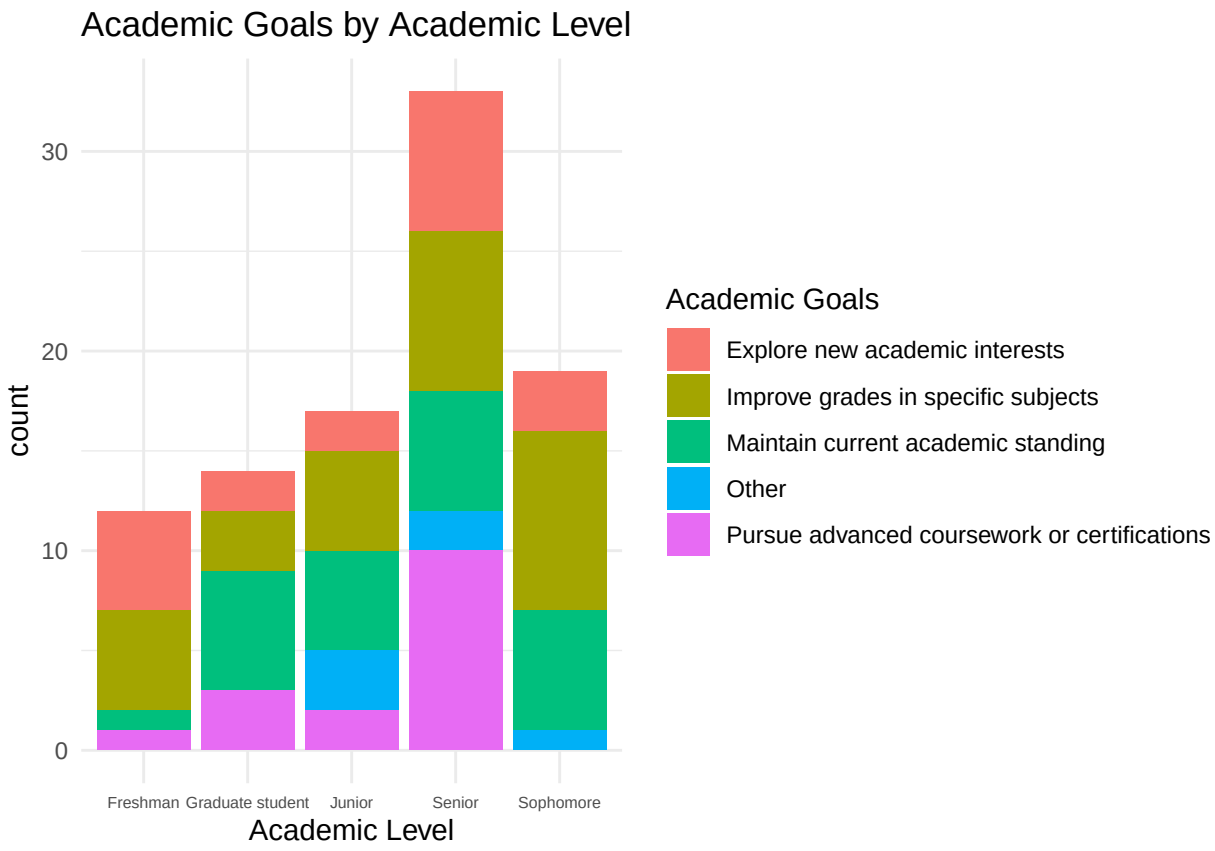
```r
# Bar plot for academic goals distribution
ggplot(data, aes(x = academic_goals, fill = academic_goals)) +
  geom_bar() +
  labs(title = "Distribution of Academic Goals", x = "") +
  theme_minimal()+
  theme(axis.text.x = element_blank())+
  scale_fill_discrete(name = "Academic Goals")
```



```r
# Stacked bar plot for academic goals by gender
ggplot(data, aes(x = gender, fill = academic_goals)) +
  geom_bar(position = "stack") +
  labs(title = "Academic Goals by Gender") +
  theme_minimal()+
  scale_fill_discrete(name = "Academic Goals")
```

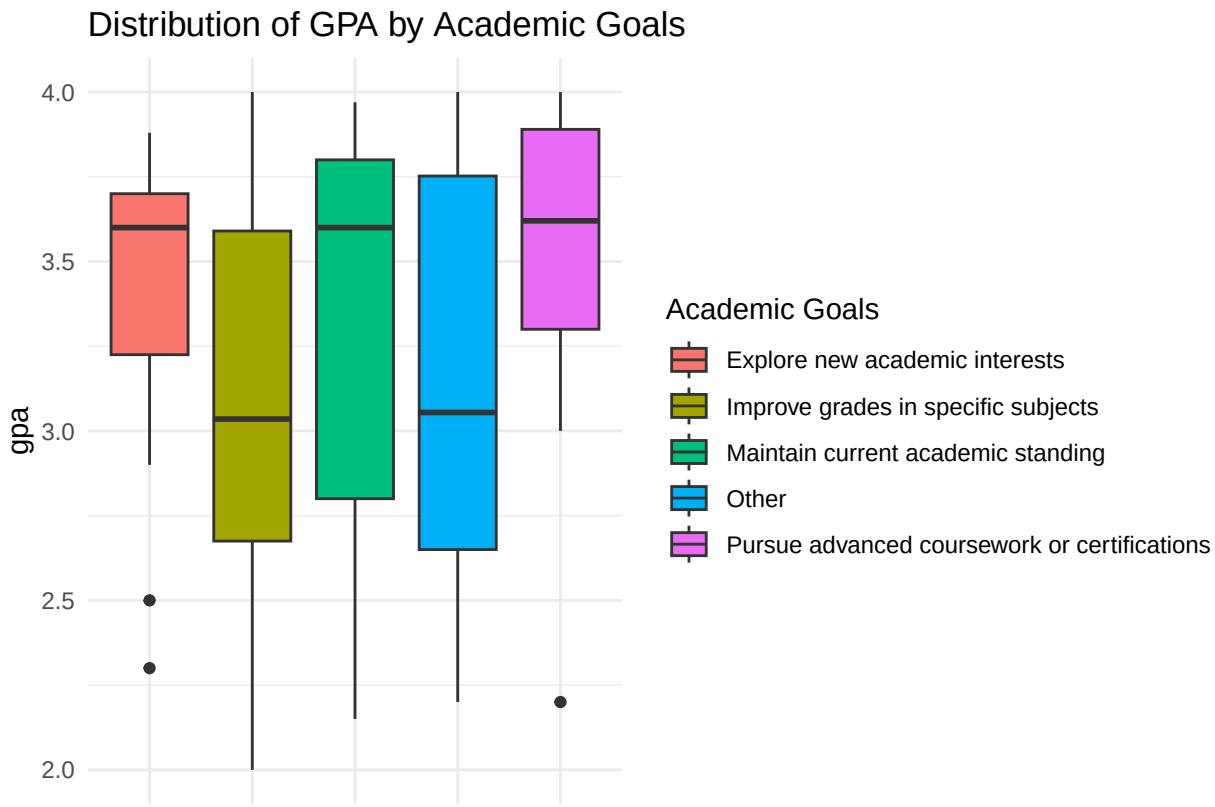


```r
# Stacked bar plot for academic goals by academic level
ggplot(data, aes(x = academic_level, fill = academic_goals)) +
  geom_bar(position = "stack") +
  labs(title = "Academic Goals by Academic Level", x = "Academic Level") +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 6))+
  scale_fill_discrete(name = "Academic Goals")
```



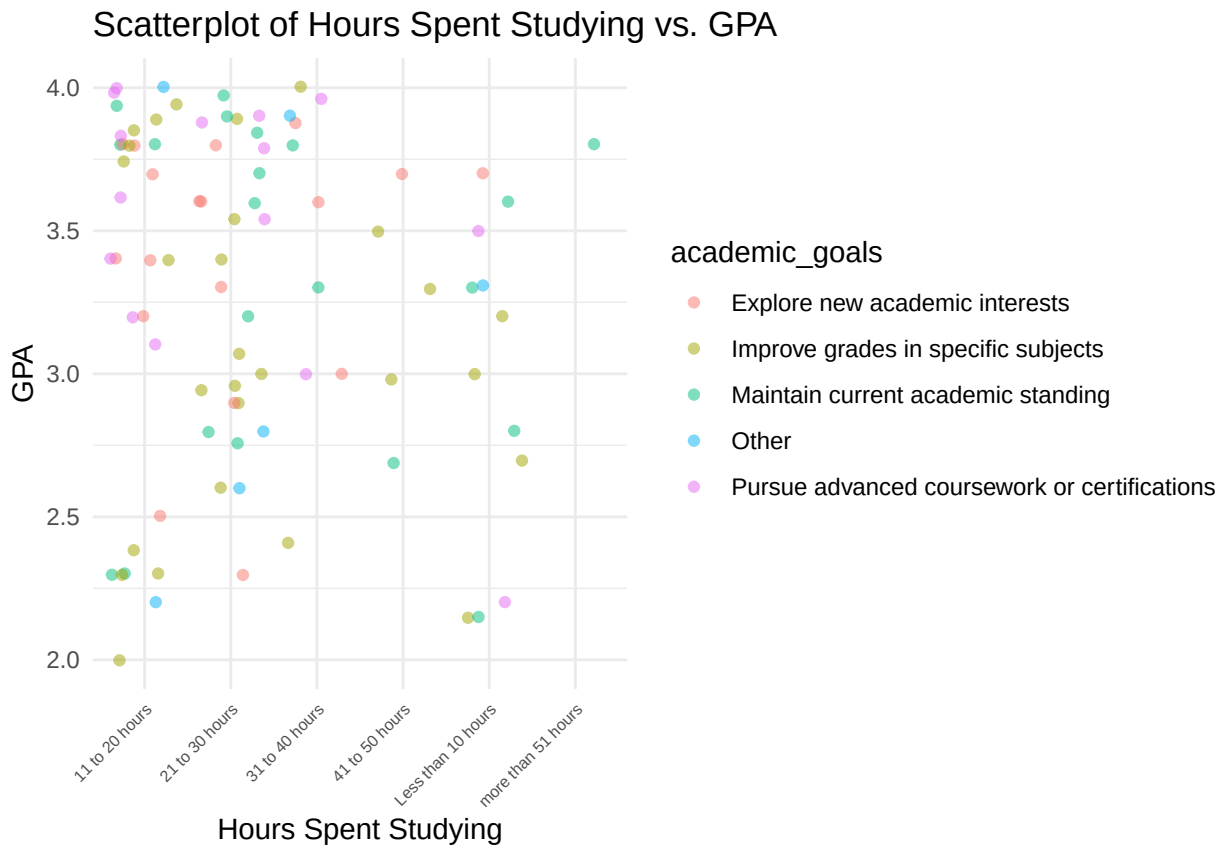
```r
# Boxplot of GPA by Academic Goals
ggplot(data, aes(x = academic_goals, y = gpa, fill = academic_goals)) +
  geom_boxplot() +
  labs(title = "Distribution of GPA by Academic Goals", x = "") +
  theme_minimal() +
  theme(axis.text.x = element_blank())+
  scale_fill_discrete(name = "Academic Goals")
```

```
## Warning: Removed 7 rows containing non-finite values (‘stat_boxplot()‘).
```



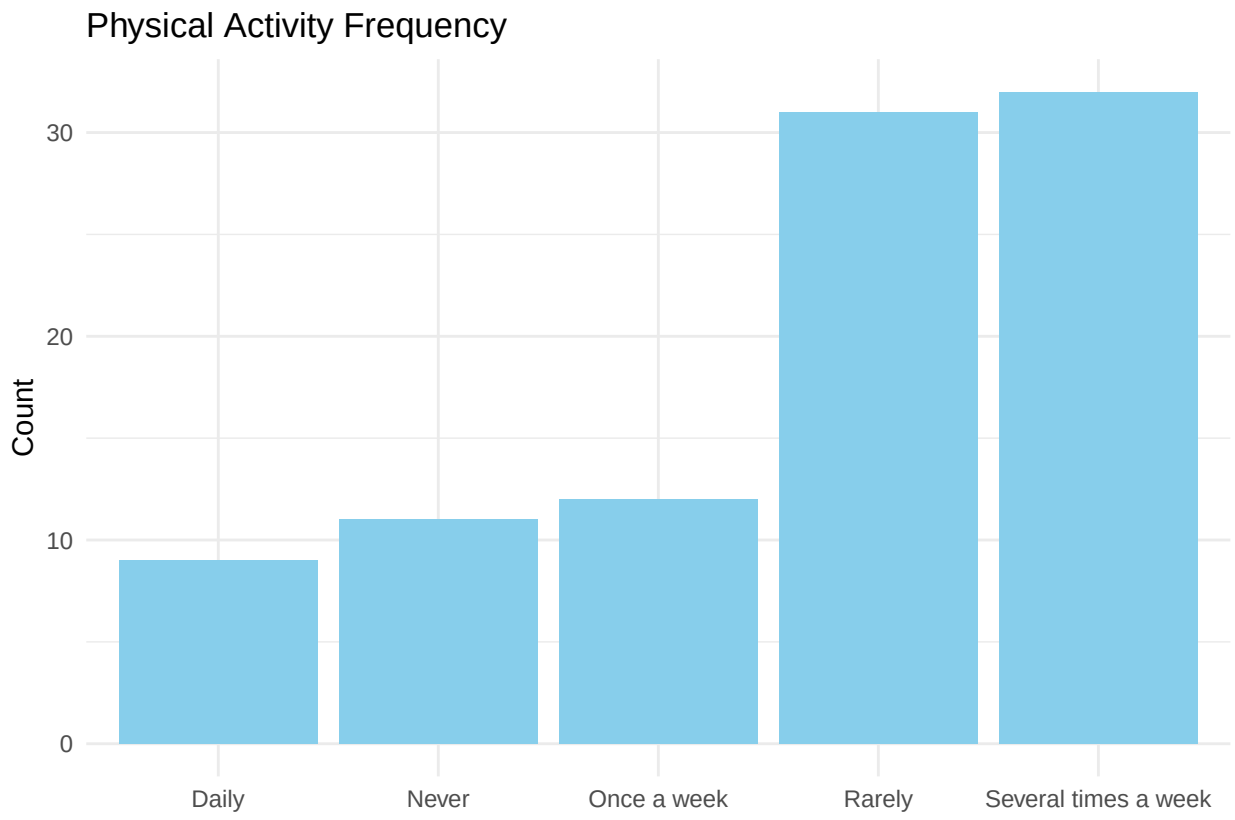
```r
# Scatterplot of Hours Spent Studying vs. GPA
ggplot(data, aes(x = hours_spend_studying, y = gpa, color = academic_goals)) +
  geom_jitter(alpha =0.5) +
  labs(title = "Scatterplot of Hours Spent Studying vs. GPA",
       x = "Hours Spent Studying", y = "GPA") +
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 6))+
  scale_fill_discrete(name = "Academic Goals")
```

```
## Warning: Removed 7 rows containing missing values (`geom_point()`).
```

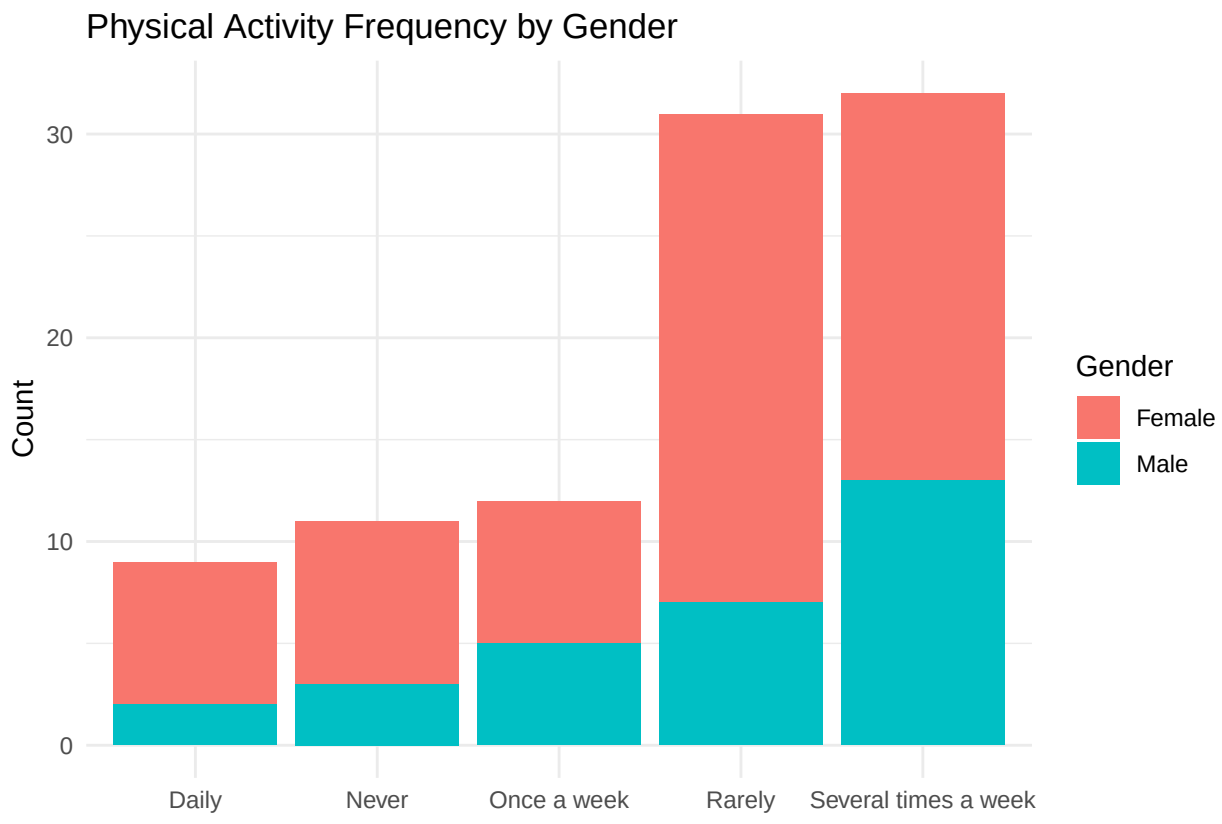


Physical Activity Analysis

```r
# Bar plot for physical_activity_freq
ggplot(data, aes(x = physical_activity_freq)) +
  geom_bar(fill = "skyblue") +
  labs(title = "Physical Activity Frequency", x = " ", y = "Count")+
  theme_minimal()
```

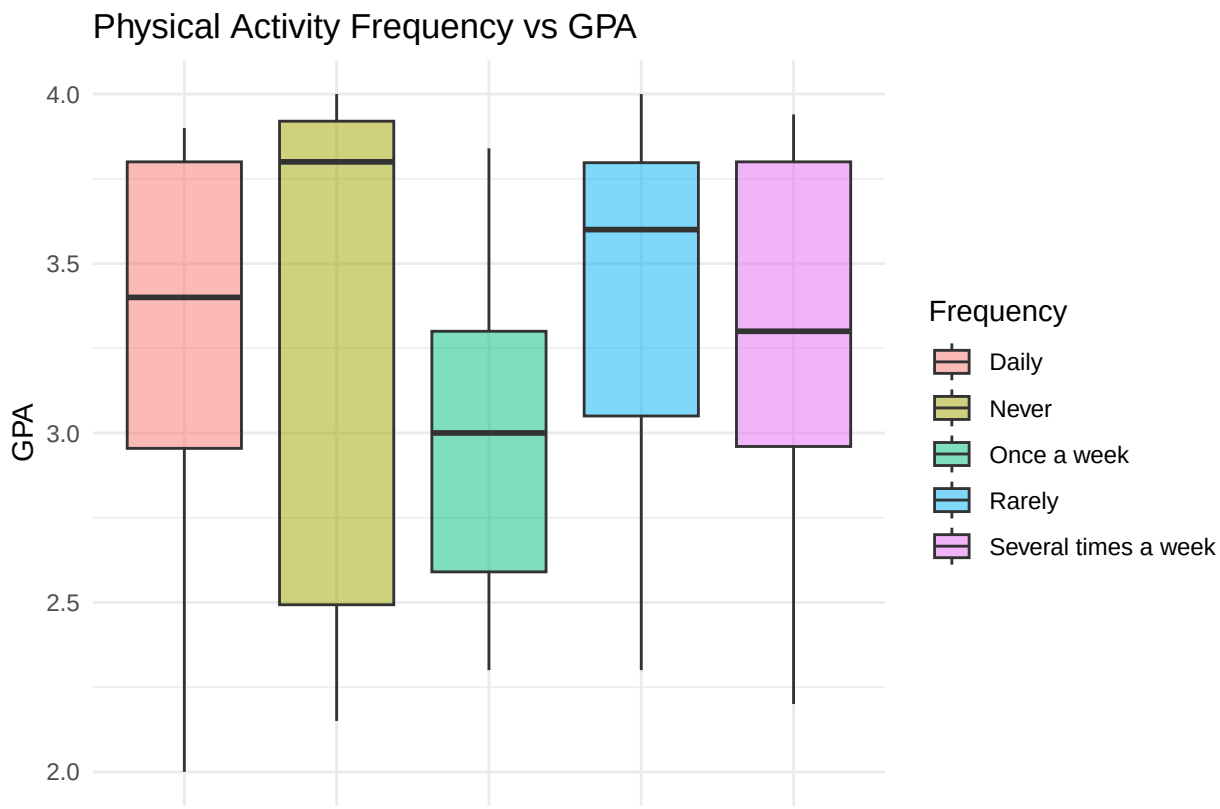


```r
# Stacked bar plot for physical_activity_freq by gender
ggplot(data, aes(x = physical_activity_freq, fill = gender)) +
  geom_bar(position = "stack") +
  labs(title = "Physical Activity Frequency by Gender", x = "", y = "Count",
       fill = "Gender")+
  theme_minimal()
```



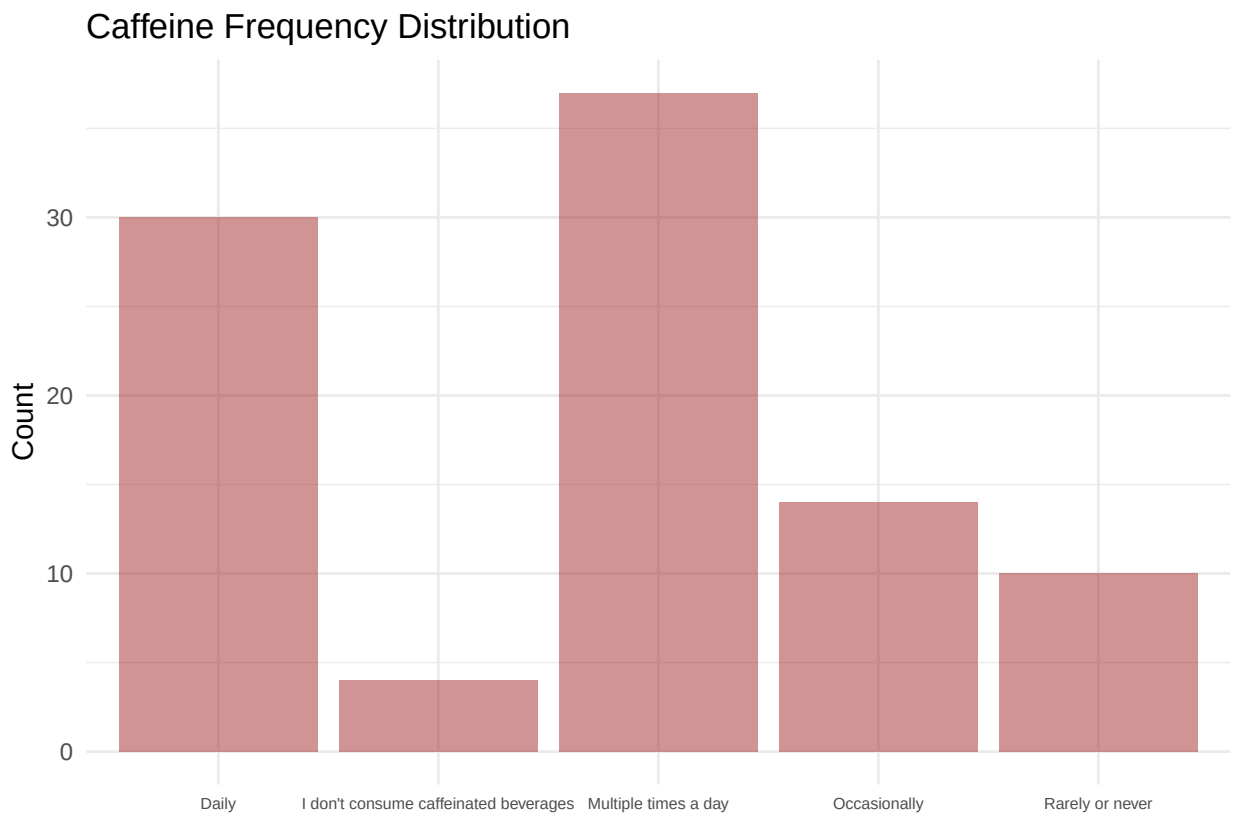
```r
# Box plot for physical_activity_freq
ggplot(data, aes(x = physical_activity_freq, y = gpa, fill = physical_activity_freq)) +
  geom_boxplot(alpha = 0.5) +
  labs(title = "Physical Activity Frequency vs GPA", x = "", y = "GPA",
       fill = "Frequency") +
  theme_minimal()+
  theme(axis.text.x = element_blank())
```

```
## Warning: Removed 7 rows containing non-finite values (‘stat_boxplot()‘).
```



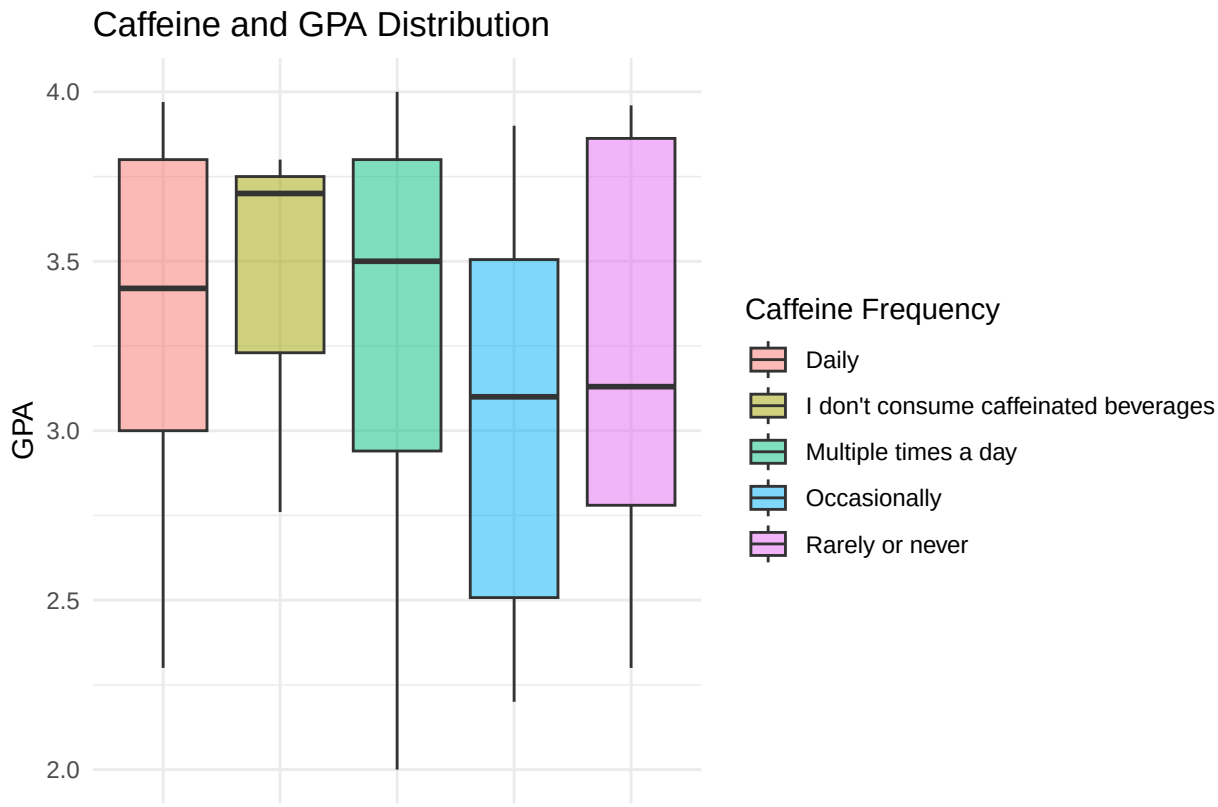
Caffeine Frequency Analysis

```
# Bar plot for caffeine frequency
ggplot(data, aes(x = caffeine_freq)) +
  geom_bar(fill = "brown", alpha = 0.5) +
  labs(title = "Caffeine Frequency Distribution",
       x = "",
       y = "Count")+
  theme_minimal()+
  theme(axis.text.x = element_text(size = 6))
```

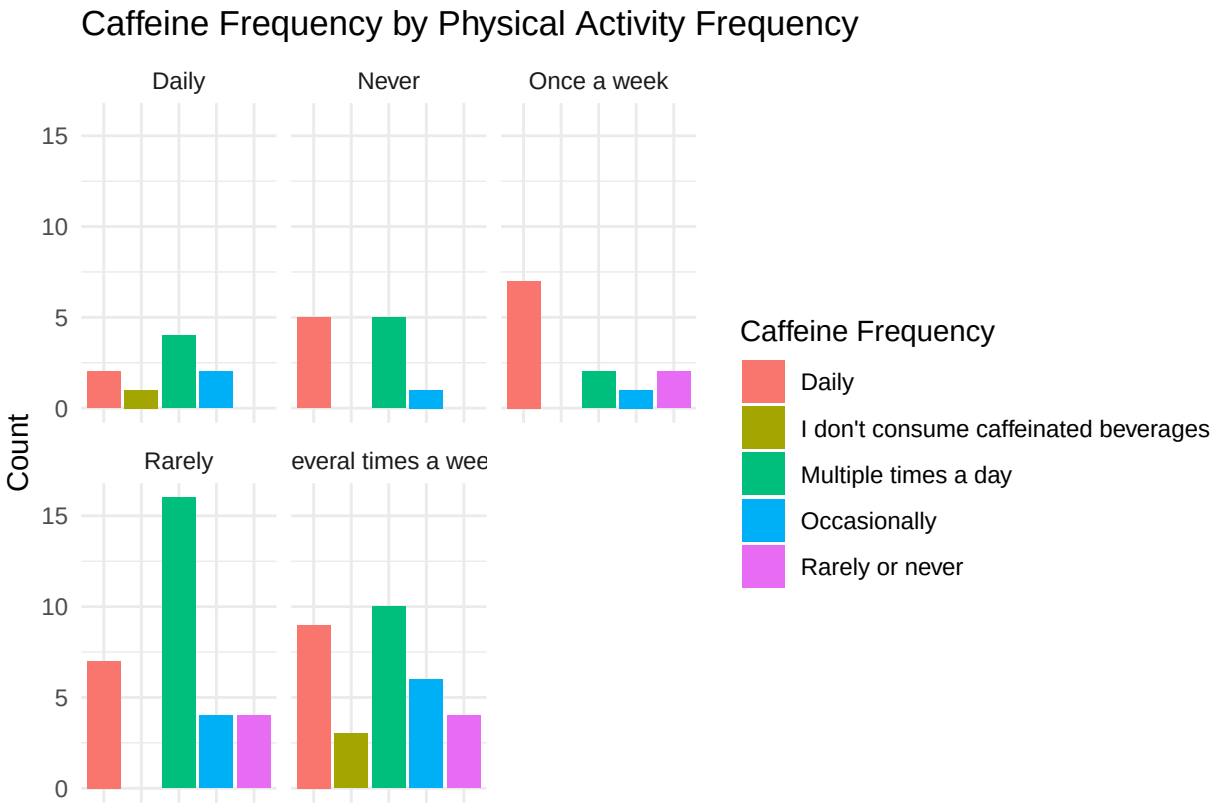


```r
# Box plot for caffeine and GPA
ggplot(data, aes(x = caffeine_freq, y = gpa, fill = caffeine_freq)) +
  geom_boxplot(alpha = 0.5) +
  labs(title = "Caffeine and GPA Distribution",
       x = "",
       y = "GPA",
       fill = "Caffeine Frequency") +
  theme_minimal() +
  theme(axis.text.x = element_blank())
```

```
## Warning: Removed 7 rows containing non-finite values (`stat_boxplot()`).
```

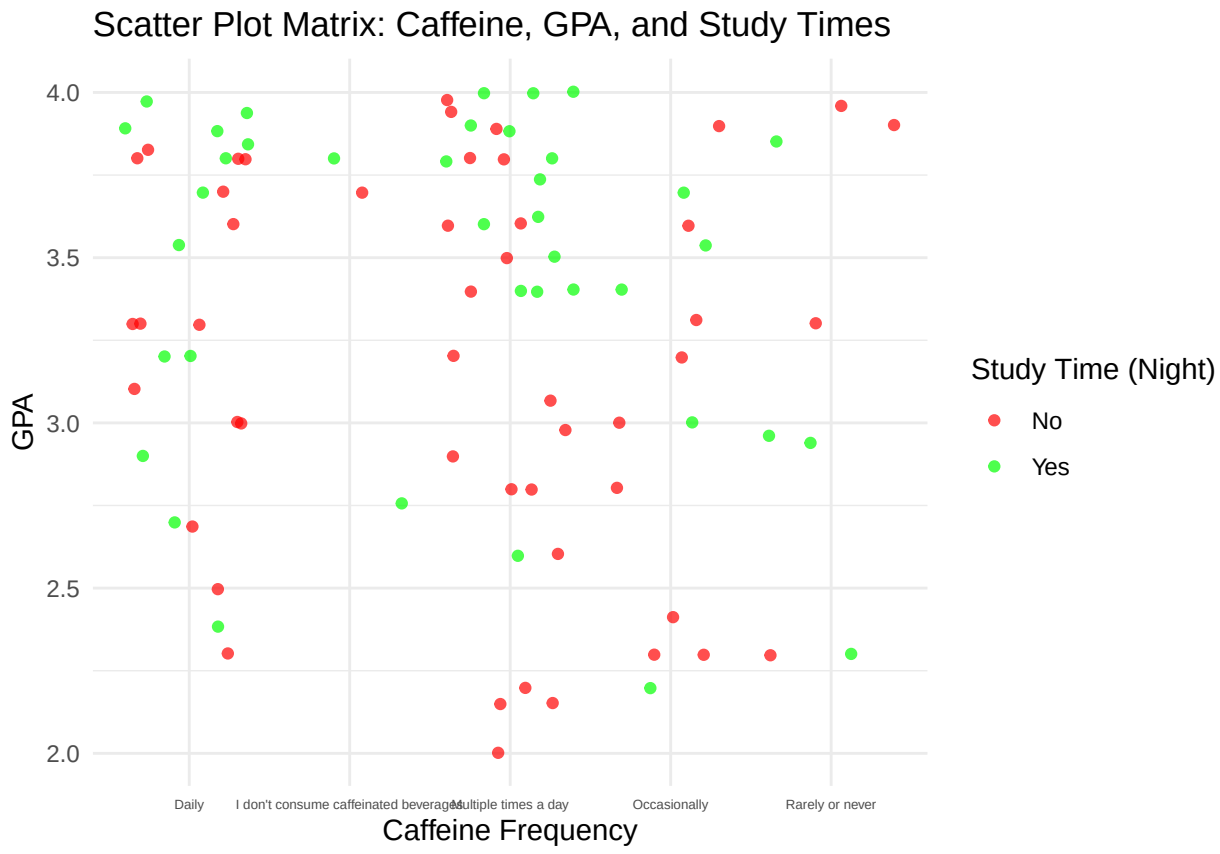


```r
# Faceted histogram for caffeine frequency by physical activity frequency
ggplot(data, aes(x = caffeine_freq, fill = caffeine_freq)) +
  geom_bar() +
  facet_wrap(~physical_activity_freq) +
  labs(title = "Caffeine Frequency by Physical Activity Frequency",
       x = "",
       y = "Count",
       fill = "Caffeine Frequency")+
  theme_minimal()+
  theme(axis.text.x = element_blank())
```

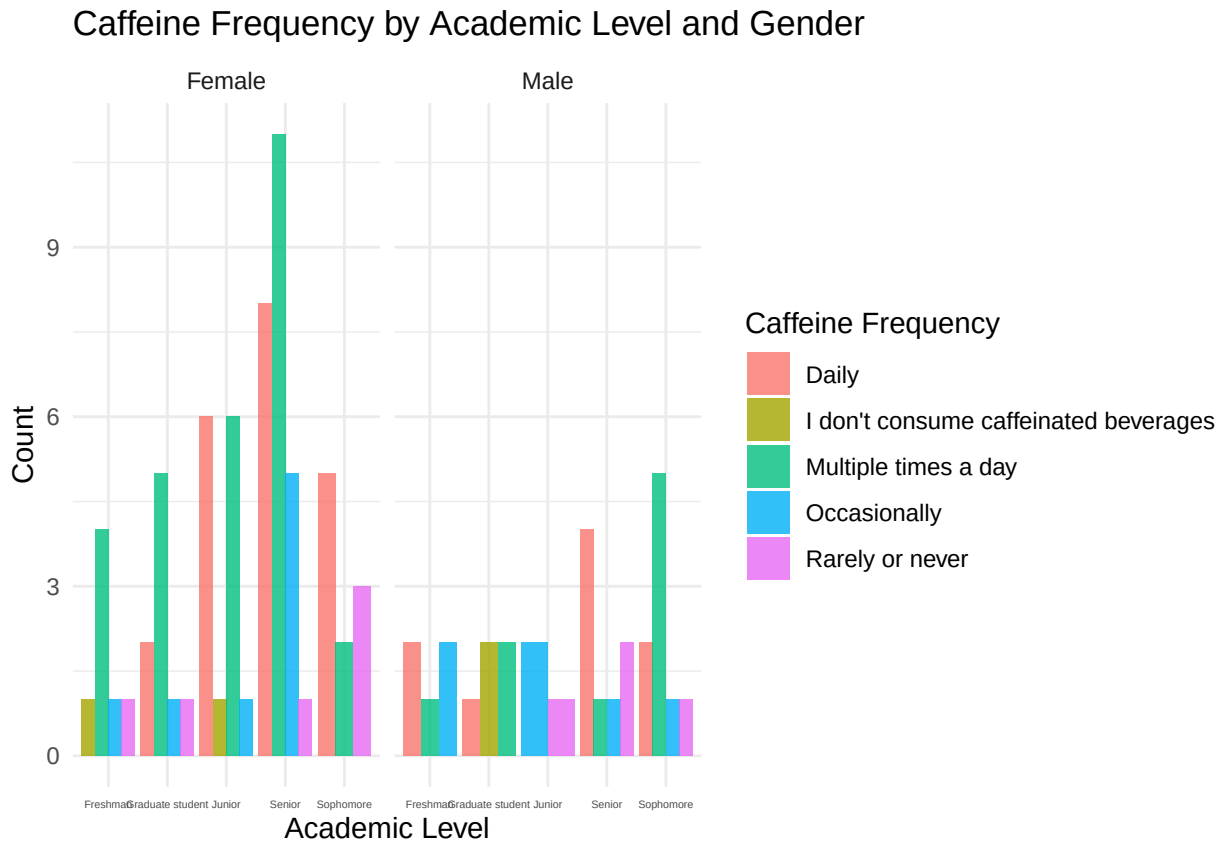


```r
# Scatter plot matrix for caffeine, GPA, and study times
ggplot(data, aes(x = caffeine_freq, y = gpa, color = study_time_Night)) +
  geom_jitter(alpha = 0.7) +
  labs(title = "Scatter Plot Matrix: Caffeine, GPA, and Study Times",
       x = "Caffeine Frequency",
       y = "GPA",
       color = "Study Time (Night)") +
  scale_color_manual(values = c("0" = "red", "1" = "green"),
                     labels = c("0" = "No", "1" = "Yes")) +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 5))
```

```
## Warning: Removed 7 rows containing missing values (`geom_point()`).
```

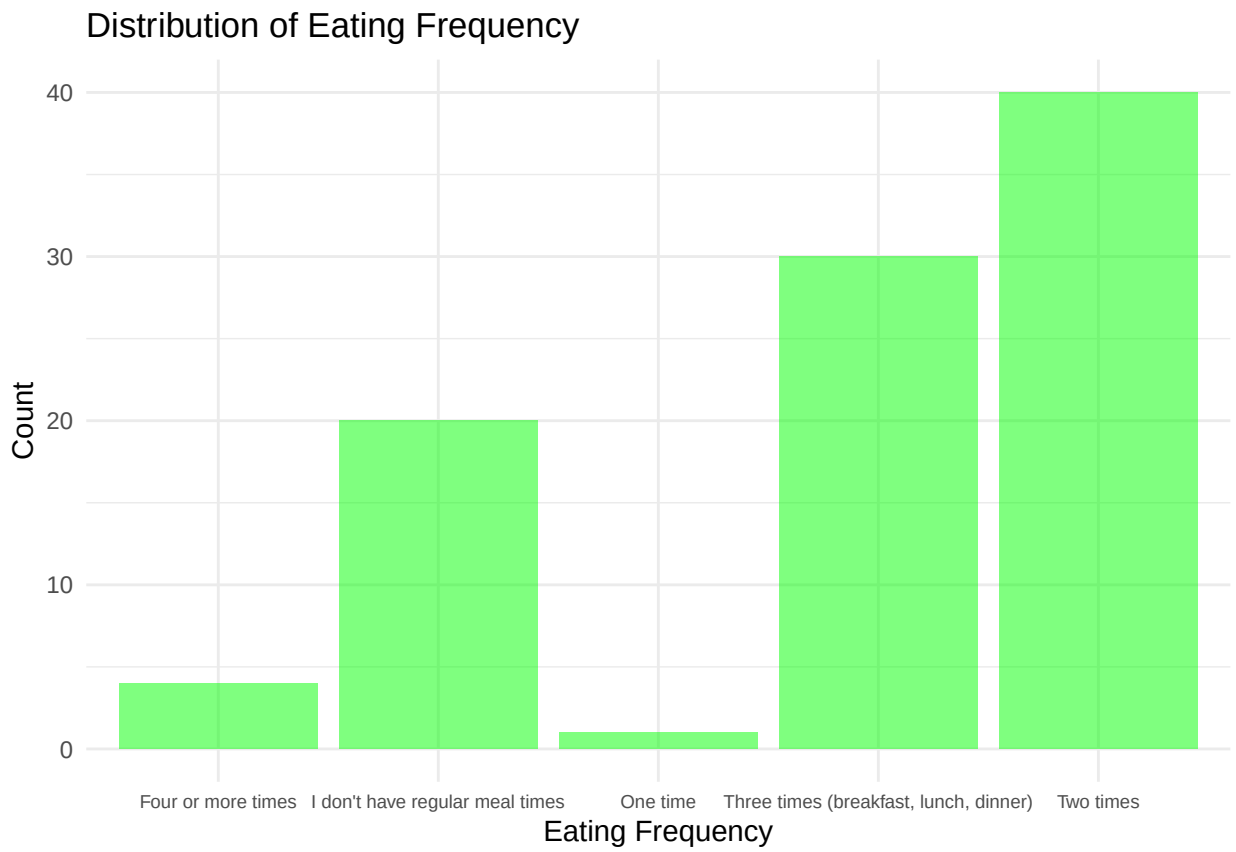


```r
# Grouped bar plot for caffeine frequency by academic level and gender
ggplot(data, aes(x = academic_level, fill = caffeine_freq)) +
  geom_bar(position = "dodge", alpha = 0.8) +
  facet_wrap(~gender) +
  labs(title = "Caffeine Frequency by Academic Level and Gender",
       x = "Academic Level",
       y = "Count",
       fill = "Caffeine Frequency") +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 4.3))
```

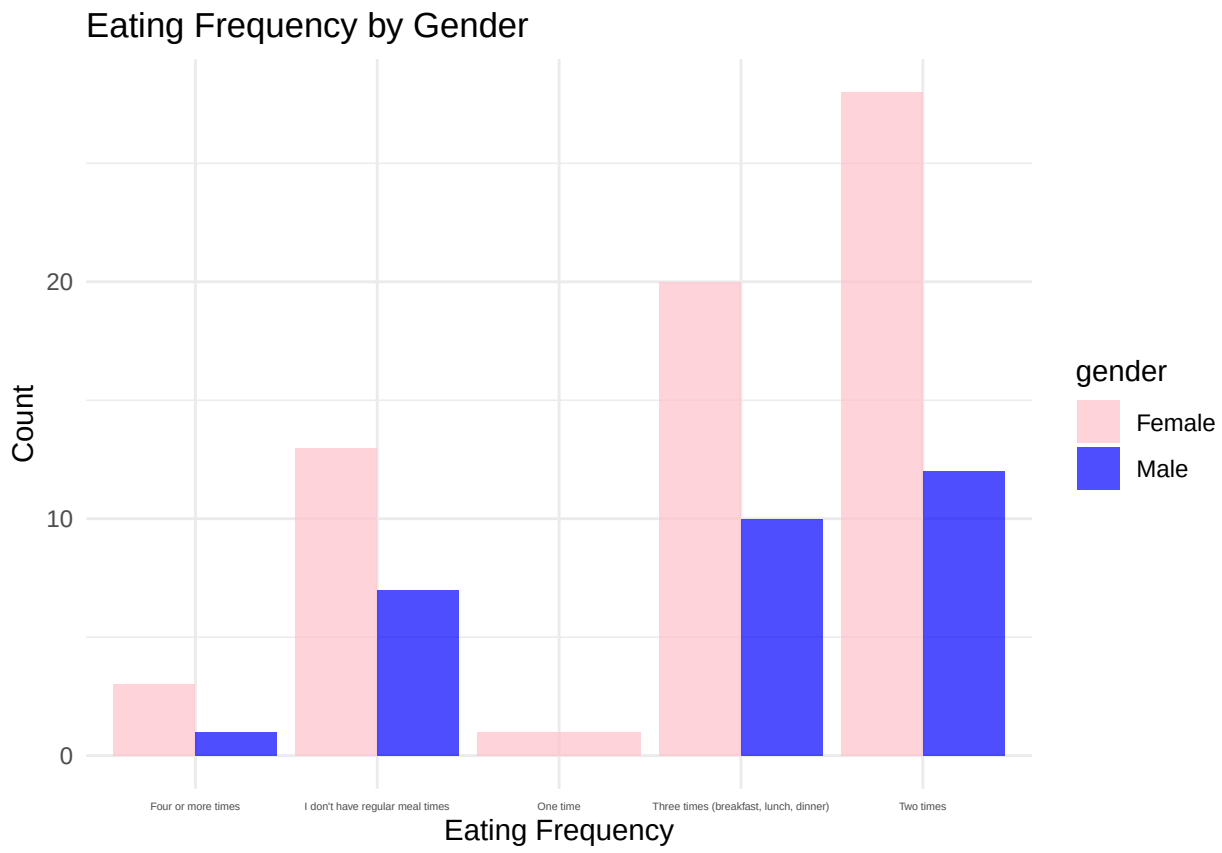


Eating Frequency Analysis

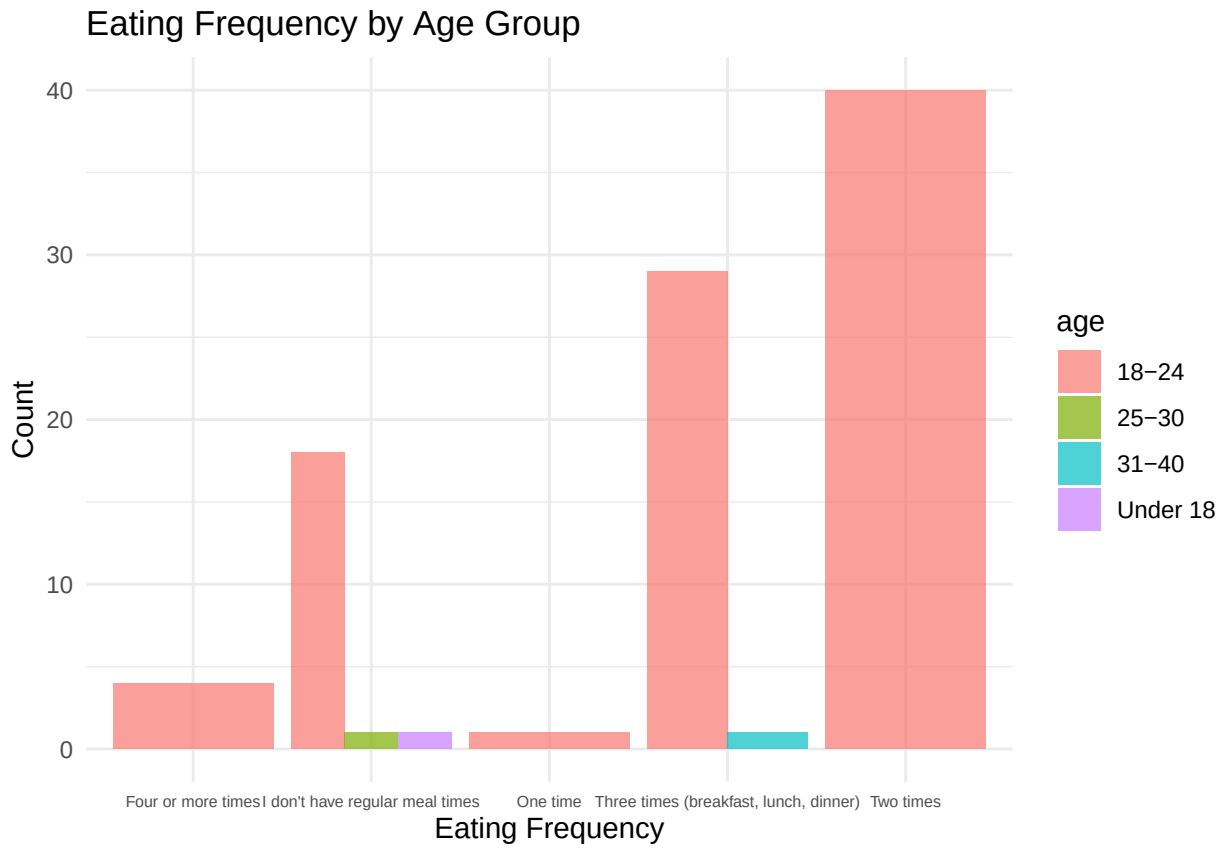
```
ggplot(data, aes(x = eat_freq)) +
  geom_bar(fill = "green", alpha = 0.5) +
  labs(title = "Distribution of Eating Frequency", x = "Eating Frequency",
       y = "Count")+
  theme_minimal() +
  theme(axis.text.x = element_text(size = 7))
```



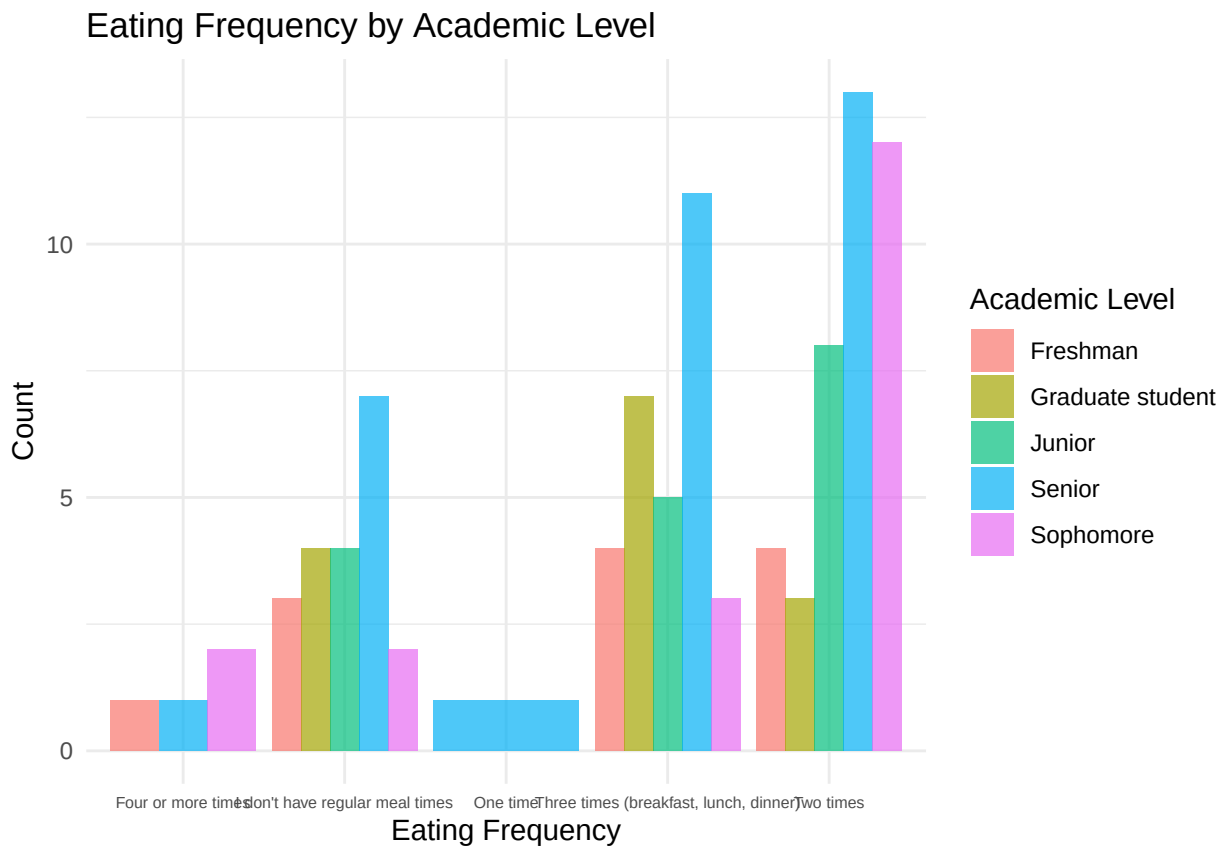
```r
ggplot(data, aes(x = eat_freq, fill = gender)) +
  geom_bar(position = "dodge", alpha = 0.7) +
  labs(title = "Eating Frequency by Gender", x = "Eating Frequency",
       y = "Count") +
  scale_fill_manual(values = c("Female" = "pink", "Male" = "blue"))+
  theme_minimal()+
  theme(axis.text.x = element_text(size = 4))
```



```r
ggplot(data, aes(x = eat_freq, fill = age)) +
  geom_bar(position = "dodge", alpha = 0.7) +
  labs(title = "Eating Frequency by Age Group", x = "Eating Frequency",
       y = "Count")+
  theme_minimal()+
  theme(axis.text.x = element_text(size = 6))
```

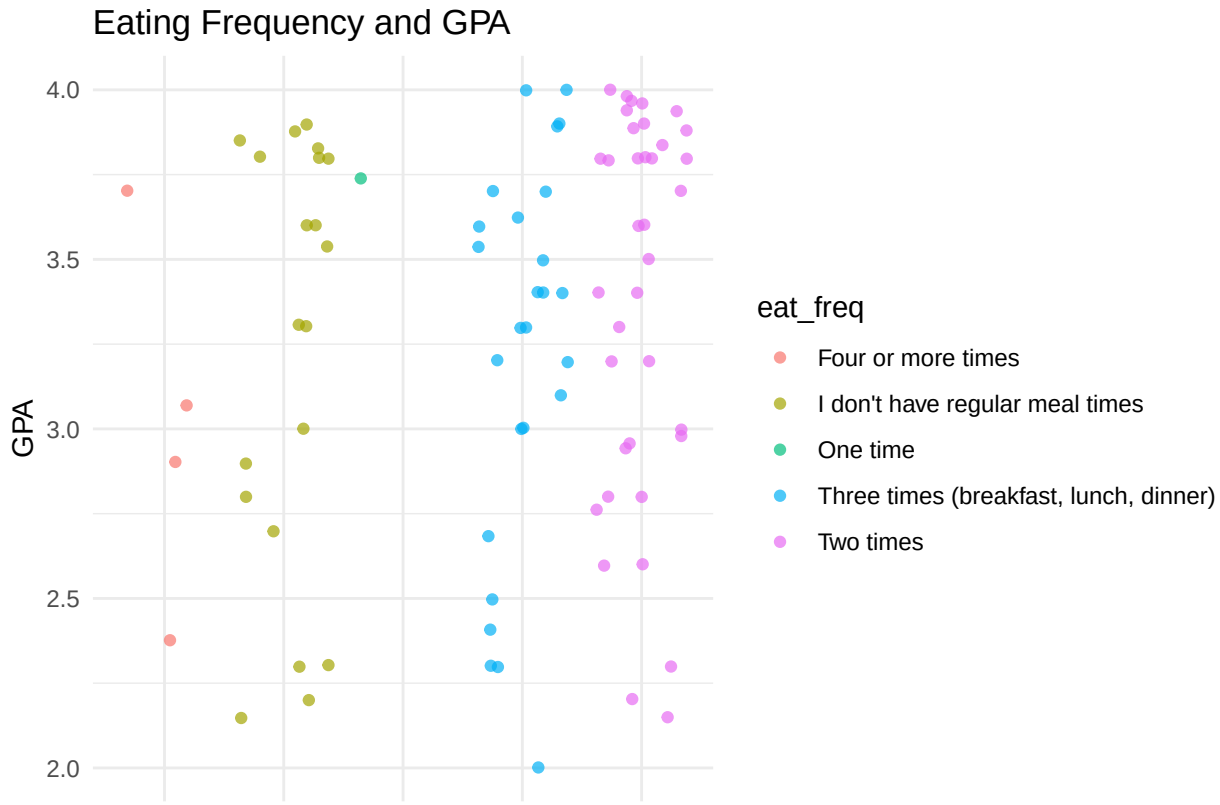


```r
ggplot(data, aes(x = eat_freq, fill = academic_level)) +
  geom_bar(position = "dodge", alpha = 0.7) +
  labs(title = "Eating Frequency by Academic Level", x = "Eating Frequency",
       y = "Count",
       fill = "Academic Level")+
  theme_minimal()+
  theme(axis.text.x = element_text(size = 6))
```



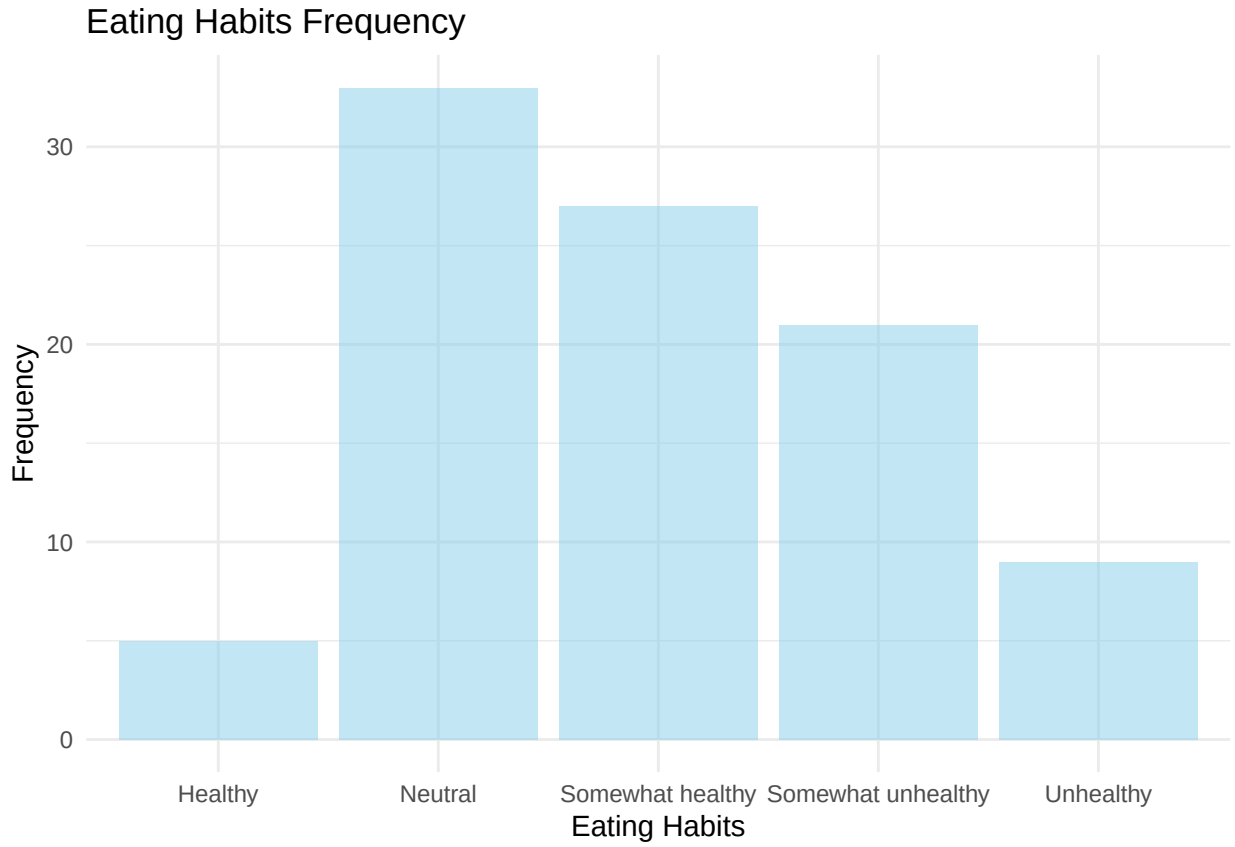
```r
ggplot(data, aes(x = eat_freq, y = gpa, color = eat_freq)) +
  geom_jitter(alpha = 0.7) +
  labs(title = "Eating Frequency and GPA", x = "", y = "GPA",
       fill = "Eating Frequency")+
  theme_minimal()+
  theme(axis.text.x = element_blank())
```

```
## Warning: Removed 7 rows containing missing values (`geom_point()`).
```

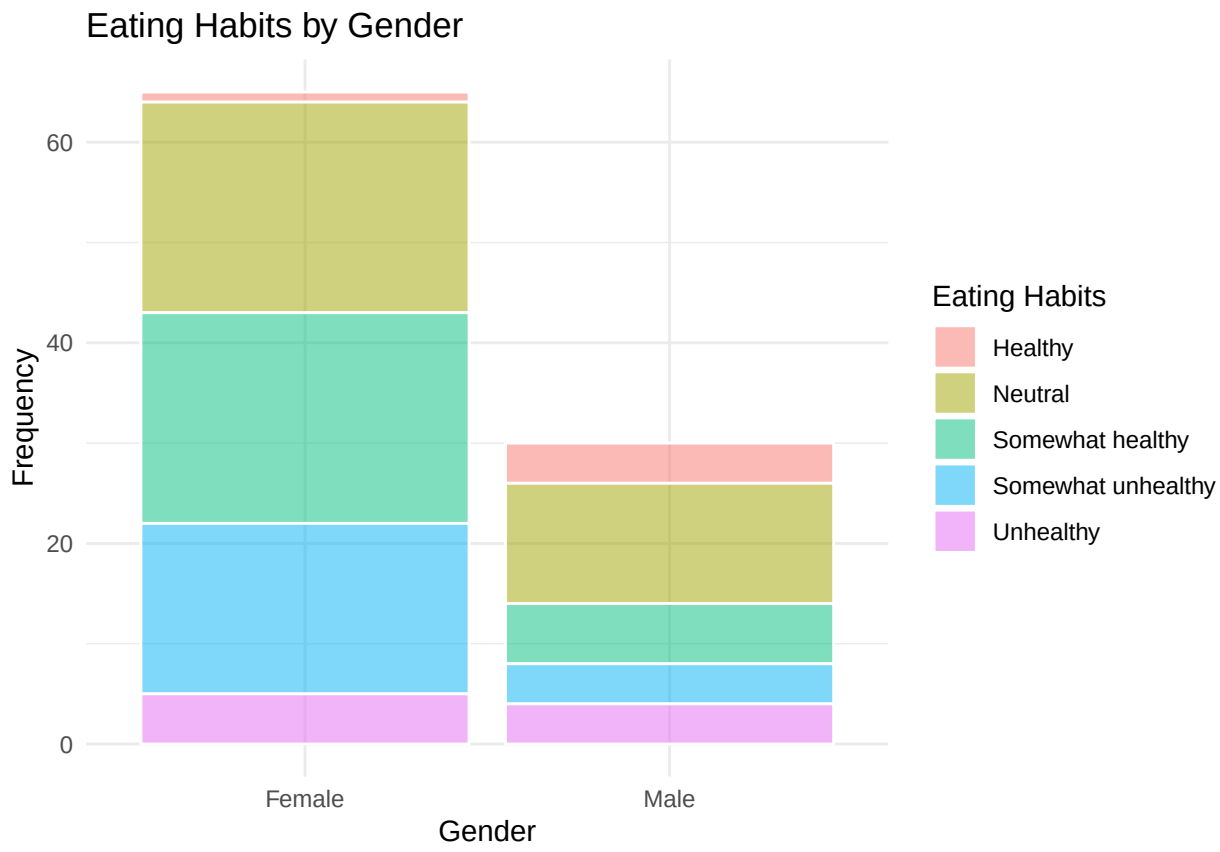


Eating Habits Analysis

```r
# Bar Plot for Eating Habits Frequency
ggplot(data, aes(x = eating_habits)) +
  geom_bar(fill = "skyblue", alpha = 0.5) +
  labs(title = "Eating Habits Frequency",
       x = "Eating Habits",
       y = "Frequency") +
  theme_minimal()
```



```r
# Stacked Bar Plot for Eating Habits by Gender
ggplot(data, aes(x = gender, fill = eating_habits)) +
  geom_bar(position = "stack", color = "white", alpha = 0.5) +
  labs(title = "Eating Habits by Gender",
       x = "Gender",
       y = "Frequency",
       fill = "Eating Habits") +
  theme_minimal()
```

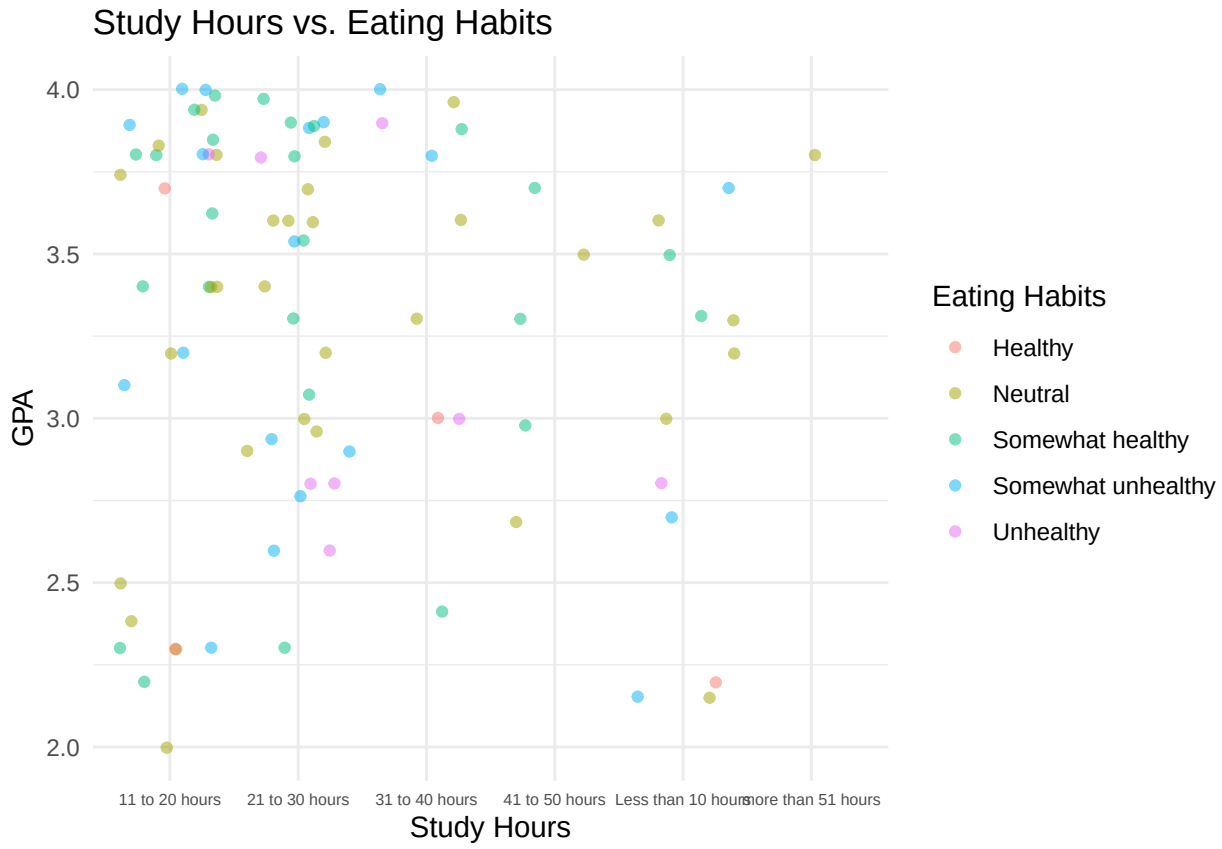


```r
# Boxplot for GPA by Eating Habits
ggplot(data, aes(x = eating_habits, y = gpa, fill = eating_habits)) +
  geom_boxplot(alpha = 0.7) +
  labs(title = "GPA by Eating Habits",
       x = "",
       y = "GPA",
       fill = "Eating Habits") +
  theme_minimal()+
  theme(axis.text.x = element_blank())
```

```
## Warning: Removed 7 rows containing non-finite values (`stat_boxplot()`).
```

```r
# Scatter Plot for Study Hours vs. Eating Habits
ggplot(data, aes(x = hours_spend_studying, y = gpa, color = eating_habits)) +
  geom_jitter(alpha=0.5) +
  labs(title = "Study Hours vs. Eating Habits",
       x = "Study Hours",
       y = "GPA",
       color = "Eating Habits") +
  theme_minimal()+
  theme(axis.text.x = element_text(size = 6))
```

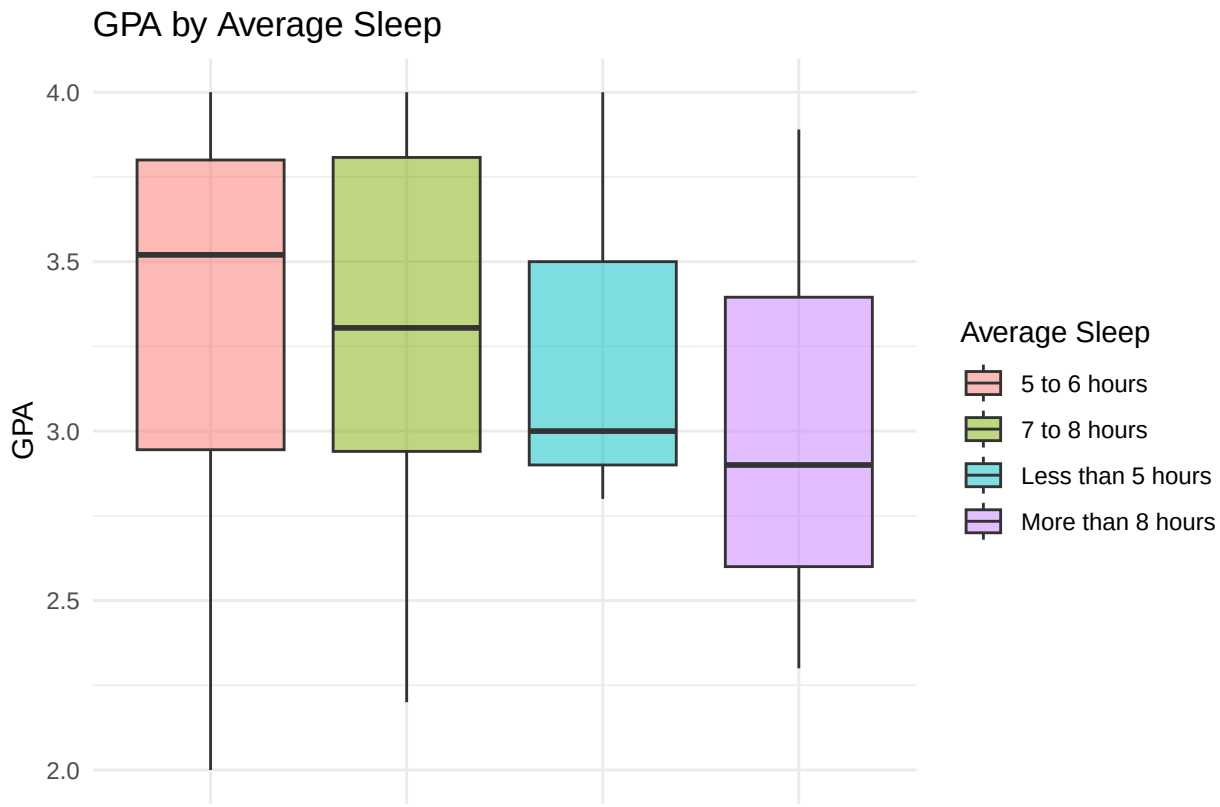
```
## Warning: Removed 7 rows containing missing values (`geom_point()`).
```



Average Sleep Analysis

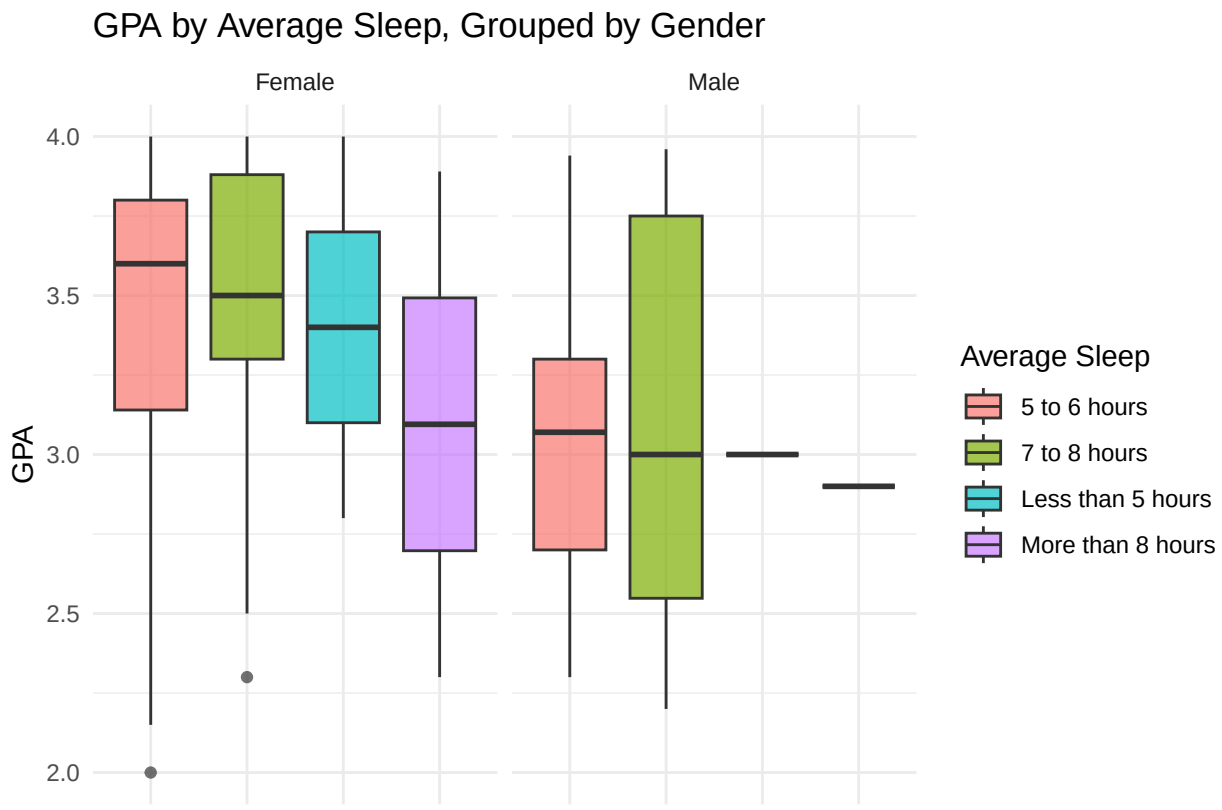
```
# Boxplot for GPA by Average Sleep
ggplot(data, aes(x = avg_sleep, y = gpa, fill = avg_sleep)) +
  geom_boxplot(alpha = 0.5) +
  labs(title = "GPA by Average Sleep",
       x = "",
       y = "GPA",
       fill = "Average Sleep") +
  theme_minimal()+
  theme(axis.text.x = element_blank())
```

```
## Warning: Removed 7 rows containing non-finite values (‘stat_boxplot()‘).
```



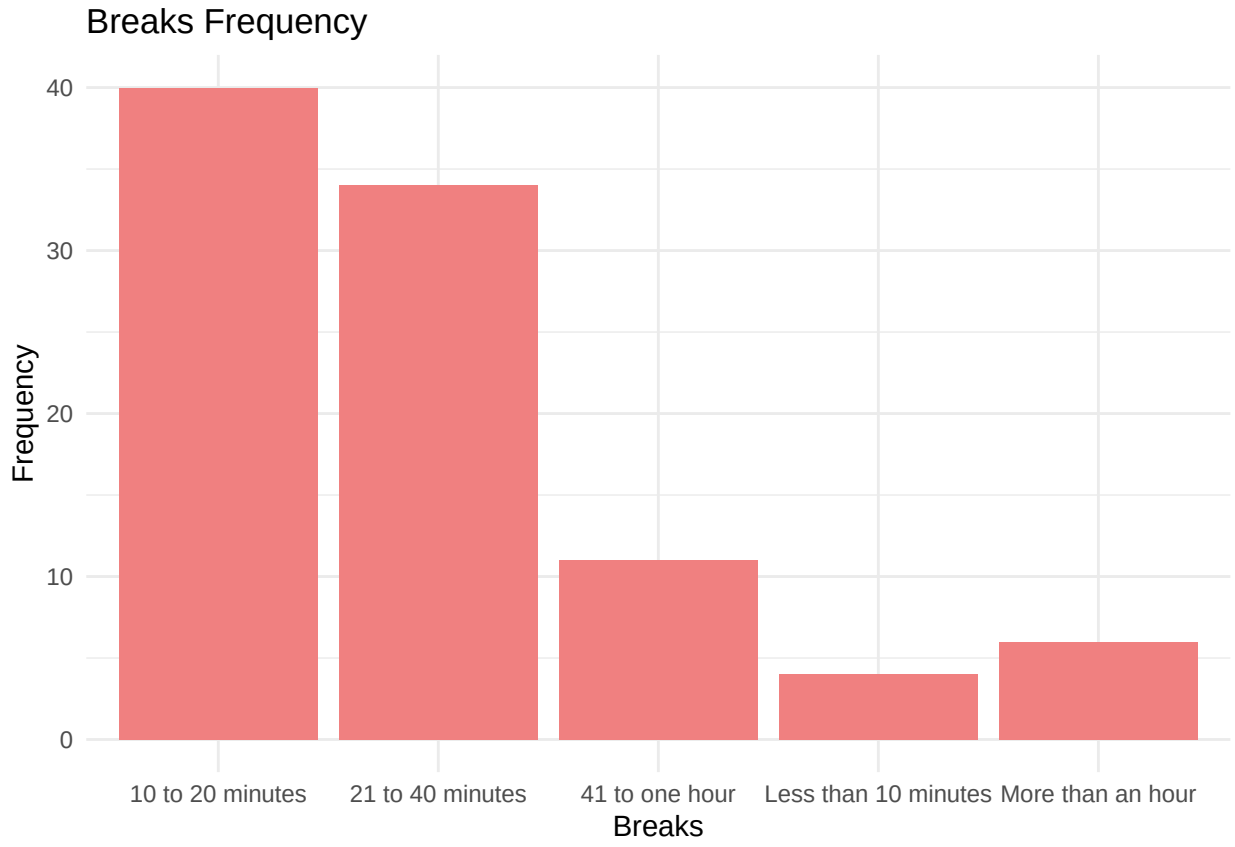
```r
# Boxplot for GPA by Average Sleep, Grouped by Gender
ggplot(data, aes(x = avg_sleep, y = gpa, fill = avg_sleep)) +
  geom_boxplot(alpha = 0.7) +
  facet_grid(~gender) +
  labs(title = "GPA by Average Sleep, Grouped by Gender",
       x = "",
       y = "GPA",
       fill = "Average Sleep") +
  theme_minimal()+
  theme(axis.text.x = element_blank())
```

```
## Warning: Removed 7 rows containing non-finite values (`stat_boxplot()`).
```



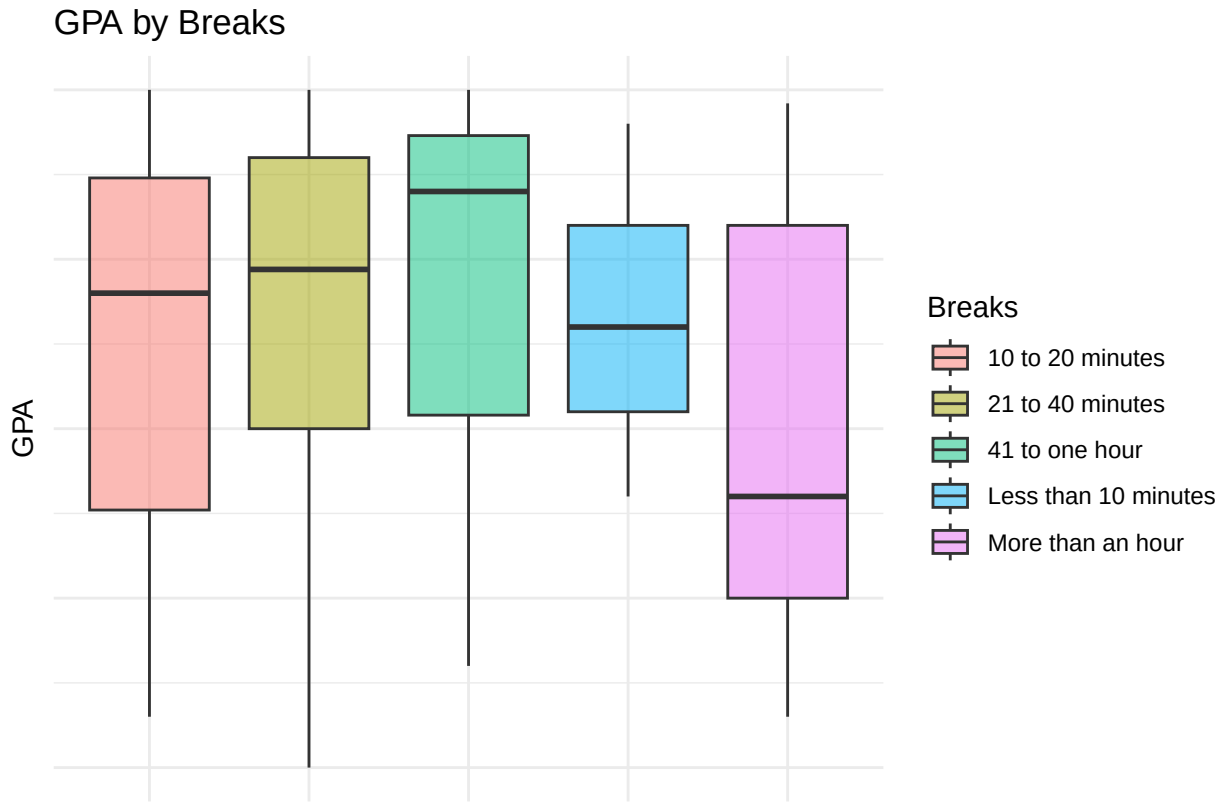
Breaks Analysis

```r
# Bar Plot for Breaks Frequency
ggplot(data, aes(x = breaks)) +
  geom_bar(fill = "lightcoral") +
  labs(title = "Breaks Frequency",
       x = "Breaks",
       y = "Frequency") +
  theme_minimal()
```

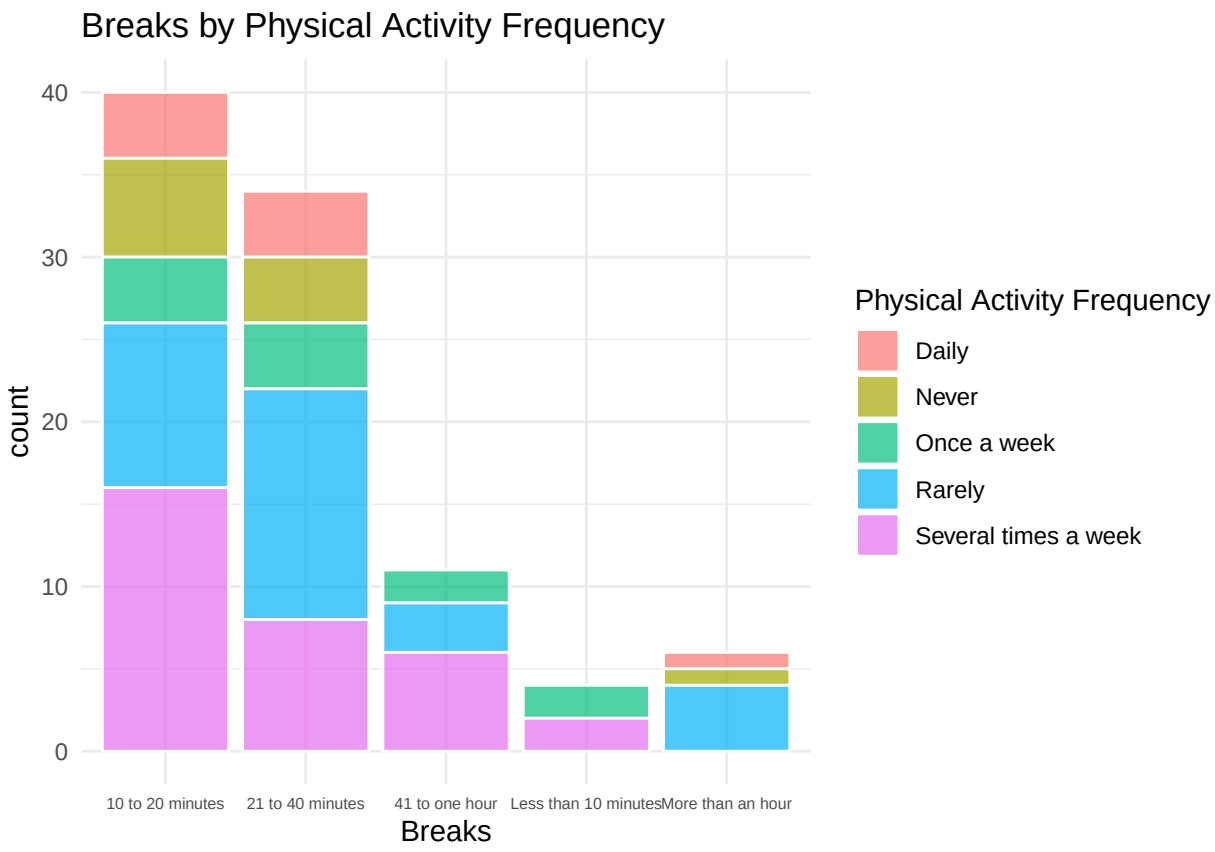


```r
# Boxplot for GPA by Breaks
ggplot(data, aes(x = breaks, y = gpa, fill = breaks)) +
  geom_boxplot(alpha = 0.5) +
  labs(title = "GPA by Breaks",
       x = "",
       y = "GPA",
       fill = "Breaks") +
  theme_minimal()+
  theme(axis.text = element_blank())
```

```
## Warning: Removed 7 rows containing non-finite values (‘stat_boxplot()‘).
```



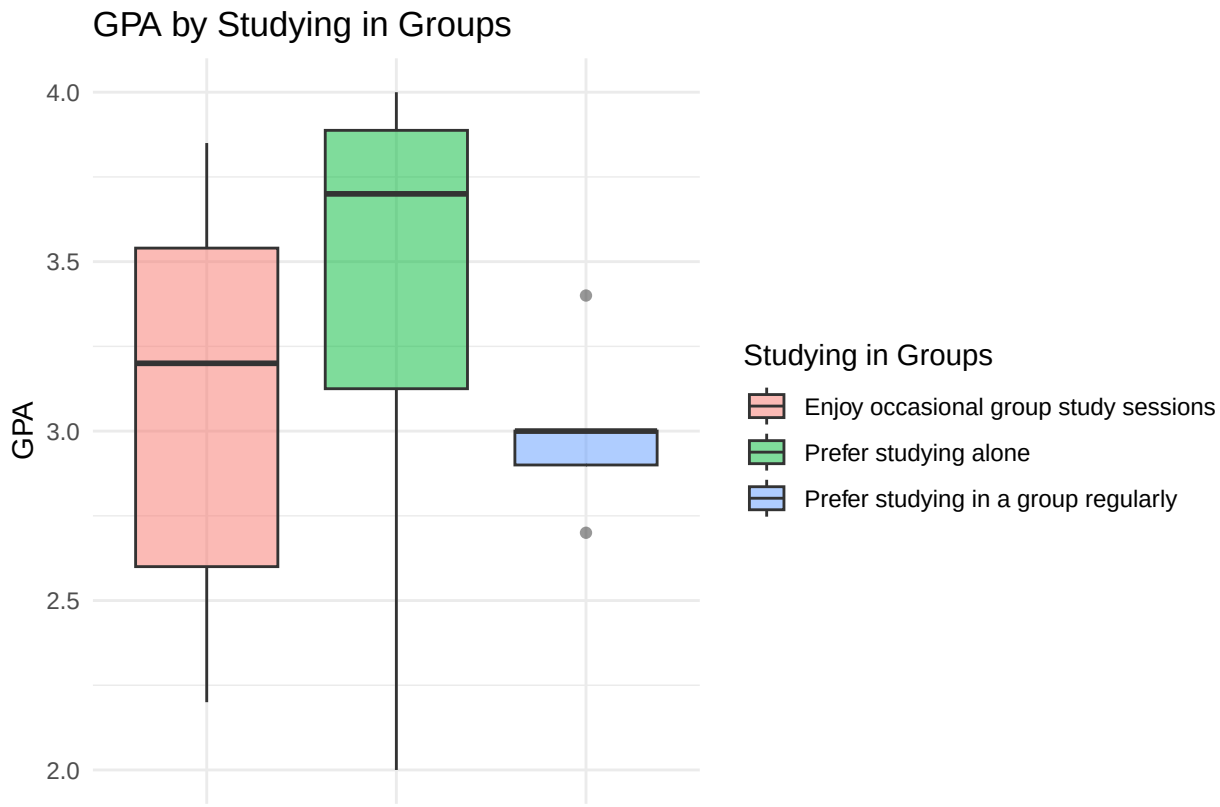
```r
# Boxplot for Breaks by Study Hours
ggplot(data, aes(x = breaks, fill = physical_activity_freq)) +
  geom_bar(position = "stack", color = "white", alpha = 0.7) +
  labs(title = "Breaks by Physical Activity Frequency",
       x = "Breaks",
       fill = "Physical Activity Frequency") +
  theme_minimal()+
  theme(axis.text.x = element_text(size = 6))
```



Studying in Groups Analysis

```
# Boxplot for GPA by Studying in Groups
ggplot(data, aes(x = studying_in_groups, y = gpa, fill = studying_in_groups)) +
  geom_boxplot(alpha = 0.5) +
  labs(title = "GPA by Studying in Groups",
       x = "",
       y = "GPA",
       fill = "Studying in Groups") +
  theme_minimal()+
  theme(axis.text.x = element_blank())
```

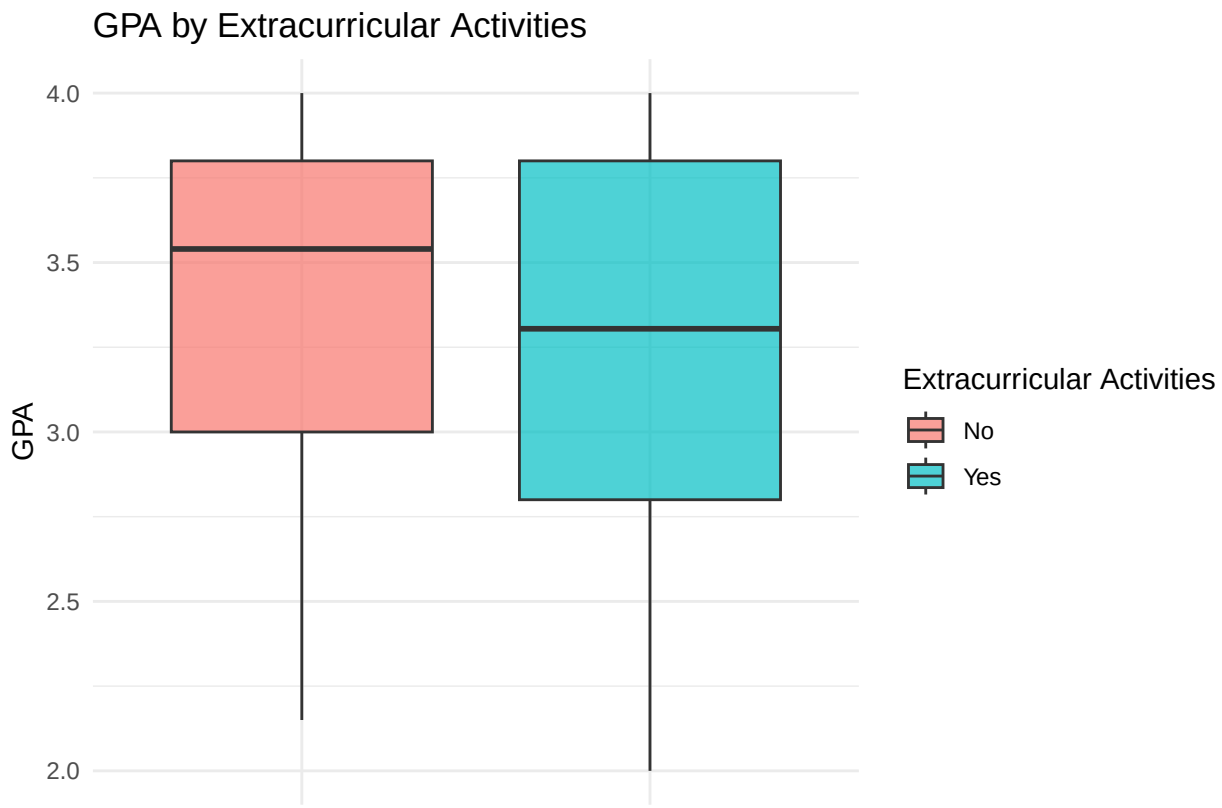
```
## Warning: Removed 7 rows containing non-finite values (‘stat_boxplot()‘).
```



Extracurricular Activities Analysis

```
# Boxplot for GPA by Extracurricular Activities
ggplot(data, aes(x = extracurricular_activities, y = gpa,
                 fill = extracurricular_activities)) +
  geom_boxplot(alpha = 0.7) +
  labs(title = "GPA by Extracurricular Activities",
       x = "",
       y = "GPA",
       fill = "Extracurricular Activities") +
  theme_minimal()+
  theme(axis.text.x = element_blank())
```

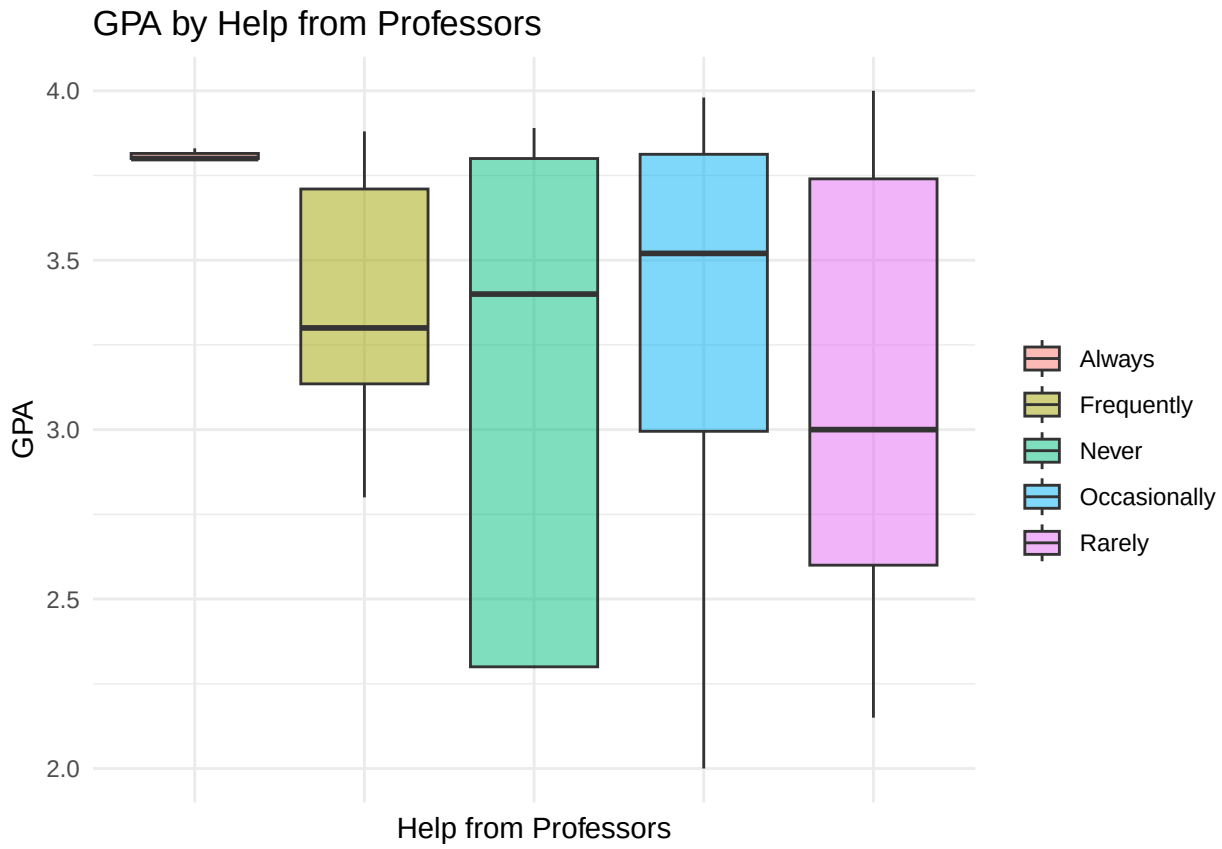
```
## Warning: Removed 7 rows containing non-finite values (`stat_boxplot()`).
```



Help from Professors Analysis

```
ggplot(data, aes(x = help_from_professors, y = gpa,
                 fill = help_from_professors)) +
  geom_boxplot(alpha = 0.5) +
  labs(title = "GPA by Help from Professors",
       x = "Help from Professors",
       y = "GPA",
       fill = "") +
  theme_minimal()+
  theme(axis.text.x = element_blank())
```

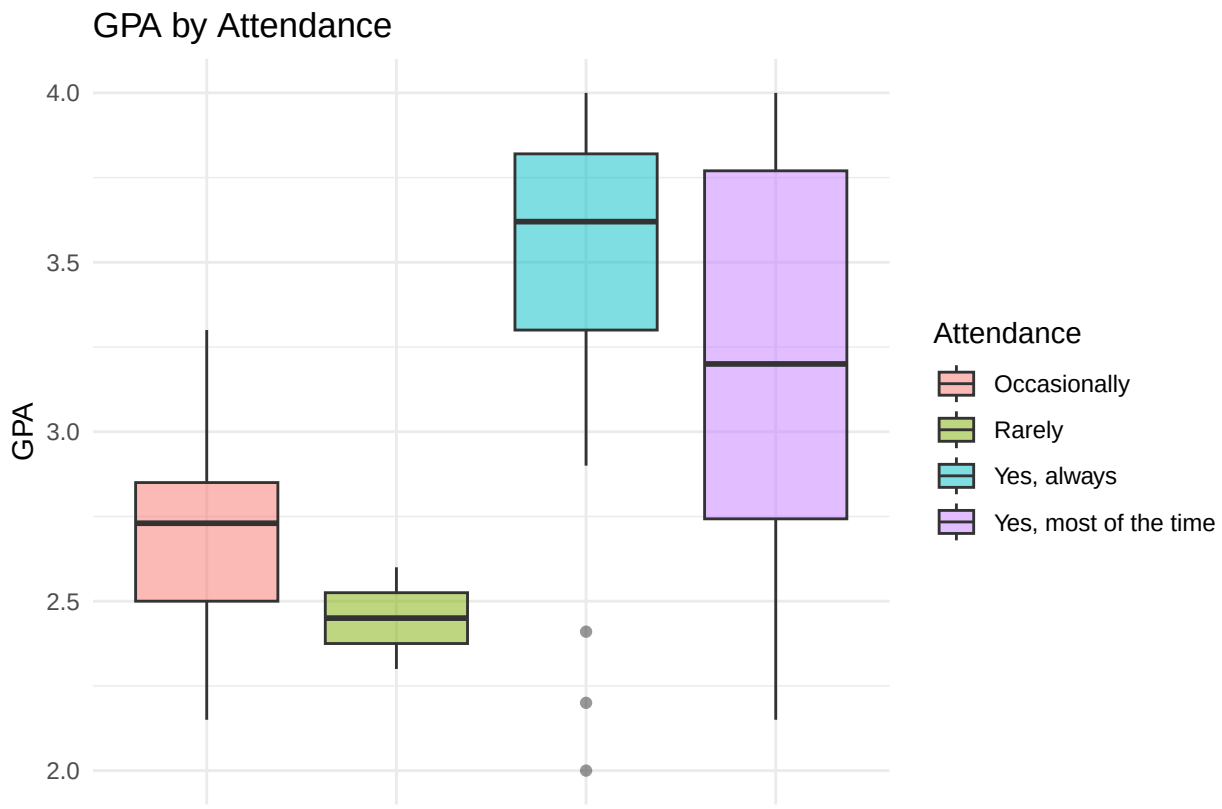
```
## Warning: Removed 7 rows containing non-finite values (`stat_boxplot()`).
```



Attendance Analysis

```
ggplot(data, aes(x = attendance, y = gpa, fill = attendance)) +
  geom_boxplot(alpha=0.5) +
  labs(title = "GPA by Attendance",
       x = "",
       y = "GPA",
       fill = "Attendance") +
  theme_minimal()+
  theme(axis.text.x = element_blank())
```

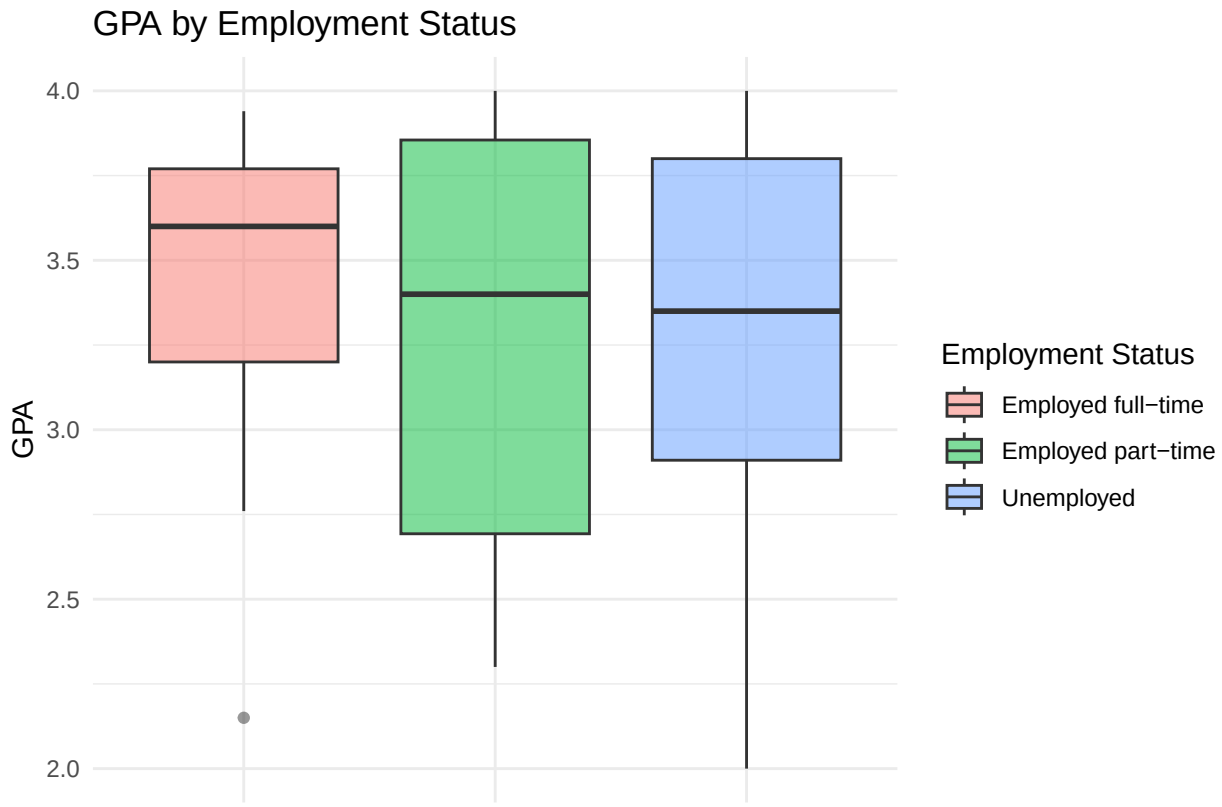
```
## Warning: Removed 7 rows containing non-finite values (‘stat_boxplot()‘).
```



Employment Status Analysis

```
ggplot(data, aes(x = employment_status, y = gpa, fill = employment_status)) +
 geom_boxplot(alpha = 0.5) +
 labs(title = "GPA by Employment Status",
      x = "",
      y = "GPA",
      fill = "Employment Status") +
 theme_minimal()+
 theme(axis.text.x = element_blank())
```

```
## Warning: Removed 7 rows containing non-finite values (‘stat_boxplot()‘).
```



Age Analysis

```
ggplot(data, aes(x = age, y = gpa, fill = age)) +
  geom_boxplot(alpha = 0.5) +
  labs(title = "GPA by Age",
       x = "",
       y = "GPA",
       fill = "Age") +
  theme_minimal()+
  theme(axis.text.x = element_blank())
```

```
## Warning: Removed 7 rows containing non-finite values (‘stat_boxplot()‘).
```