Students Survey 20241

Visualize Data

Open dataset

```
# Load necessary libraries
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

Load Dataset

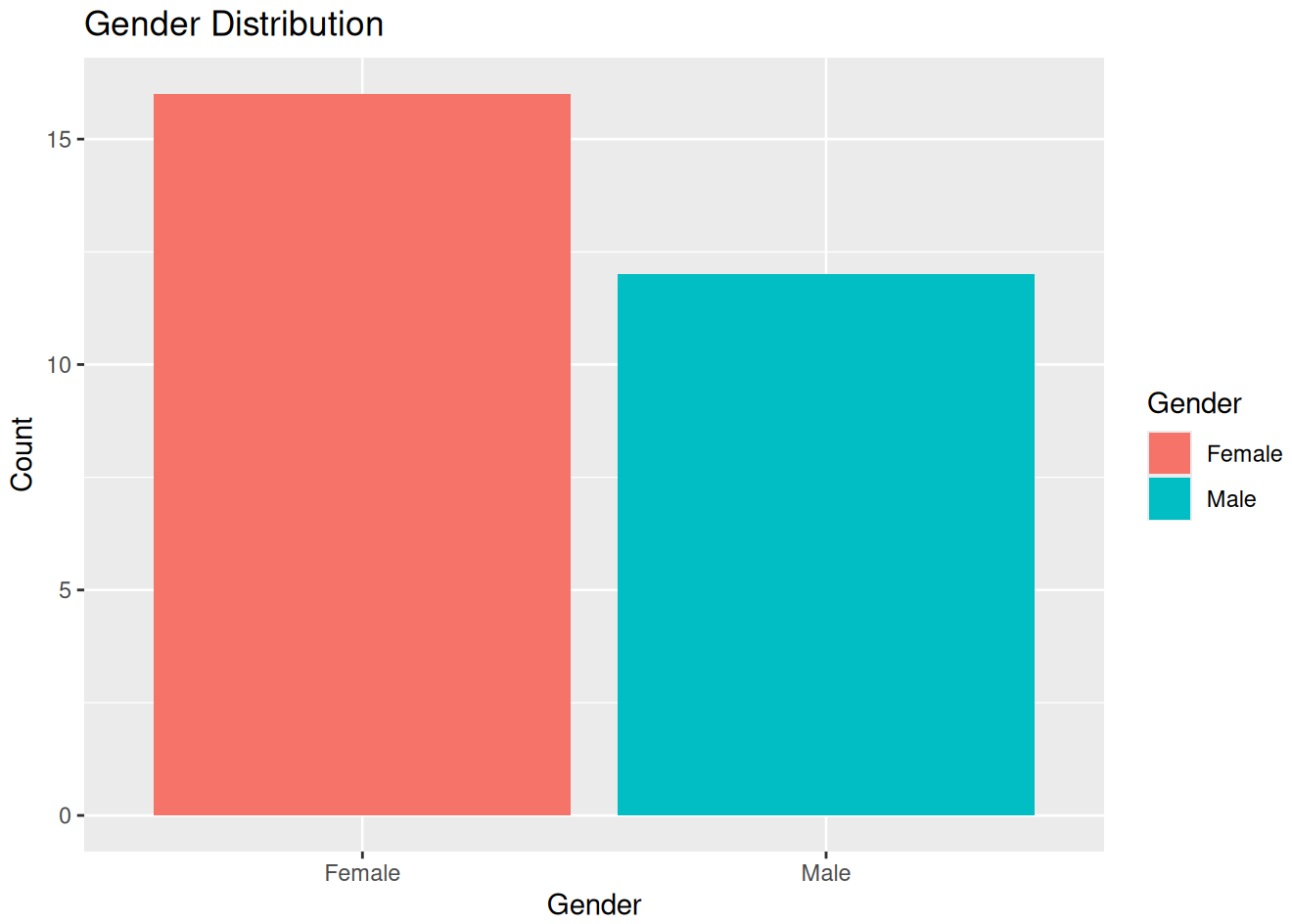
```
# Load the dataset
survey_data <- read.csv("students_sruvey_307307_20241.csv")
```

Demographics Overview

```
# Gender Distribution
gender_dist <- survey_data %>%
  group_by(Gender) %>%
  summarise(Count = n())
print(gender_dist)
```

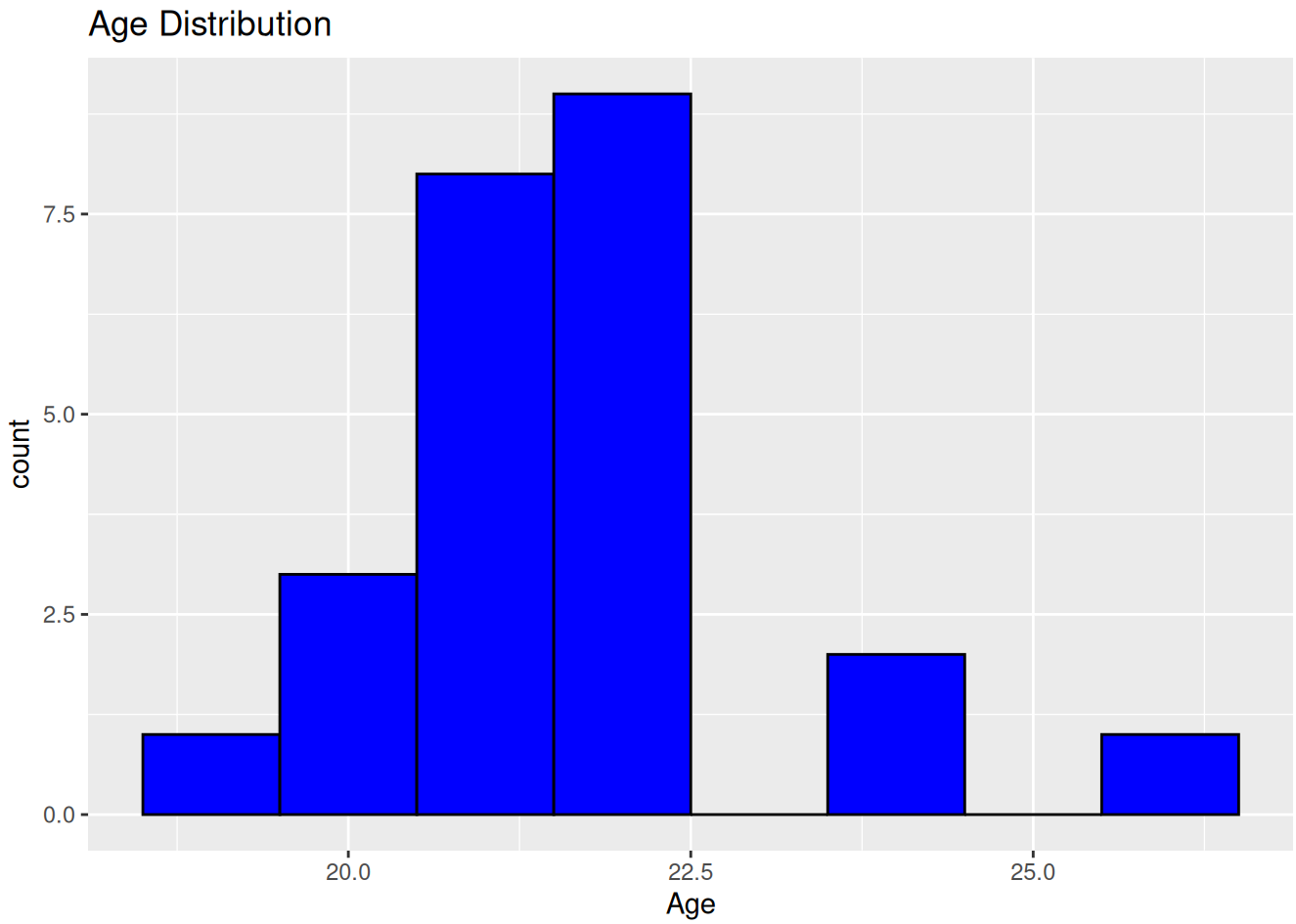
```
# A tibble: 2 × 2
  Gender Count
  <chr>  <int>
1 Female    16
2 Male      12
```

```
ggplot(gender_dist, aes(x = Gender, y = Count, fill = Gender)) +
  geom_bar(stat = "identity") +
  ggtitle("Gender Distribution")
```




```
# Age Distribution
ggplot(survey_data, aes(x = Age)) +
  geom_histogram(binwidth = 1, fill = "blue", color = "black") +
  ggtitle("Age Distribution")
```

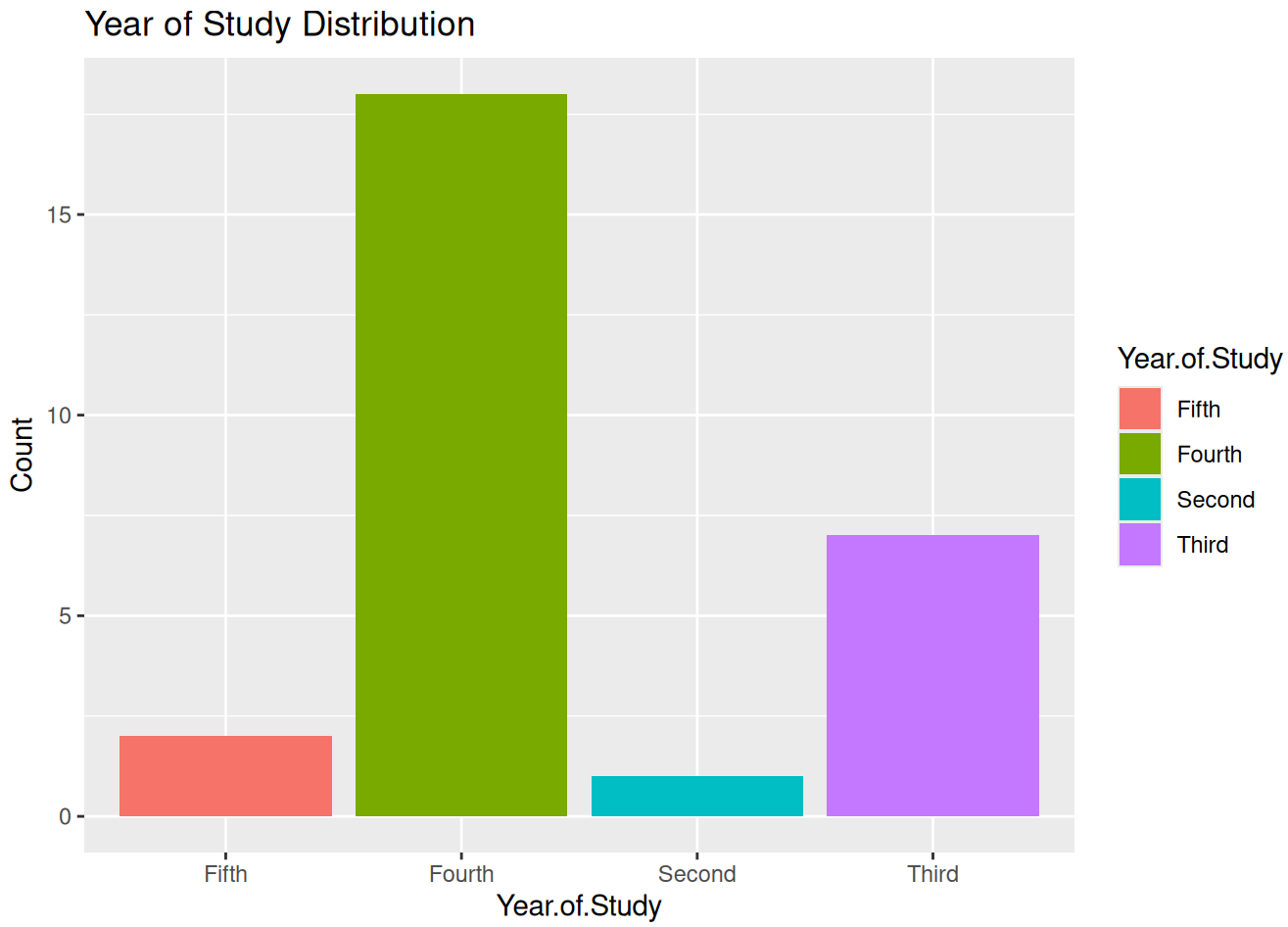
```
Warning: Removed 4 rows containing non-finite outside the scale range
(`stat_bin()`).
```




```
# Year of Study Distribution
year_dist <- survey_data %>%
  group_by(Year.of.Study) %>%
  summarise(Count = n())
print(year_dist)
```

```
# A tibble: 4 × 2
  Year.of.Study Count
  <chr>         <int>
1 Fifth             2
2 Fourth           18
3 Second            1
4 Third             7
```

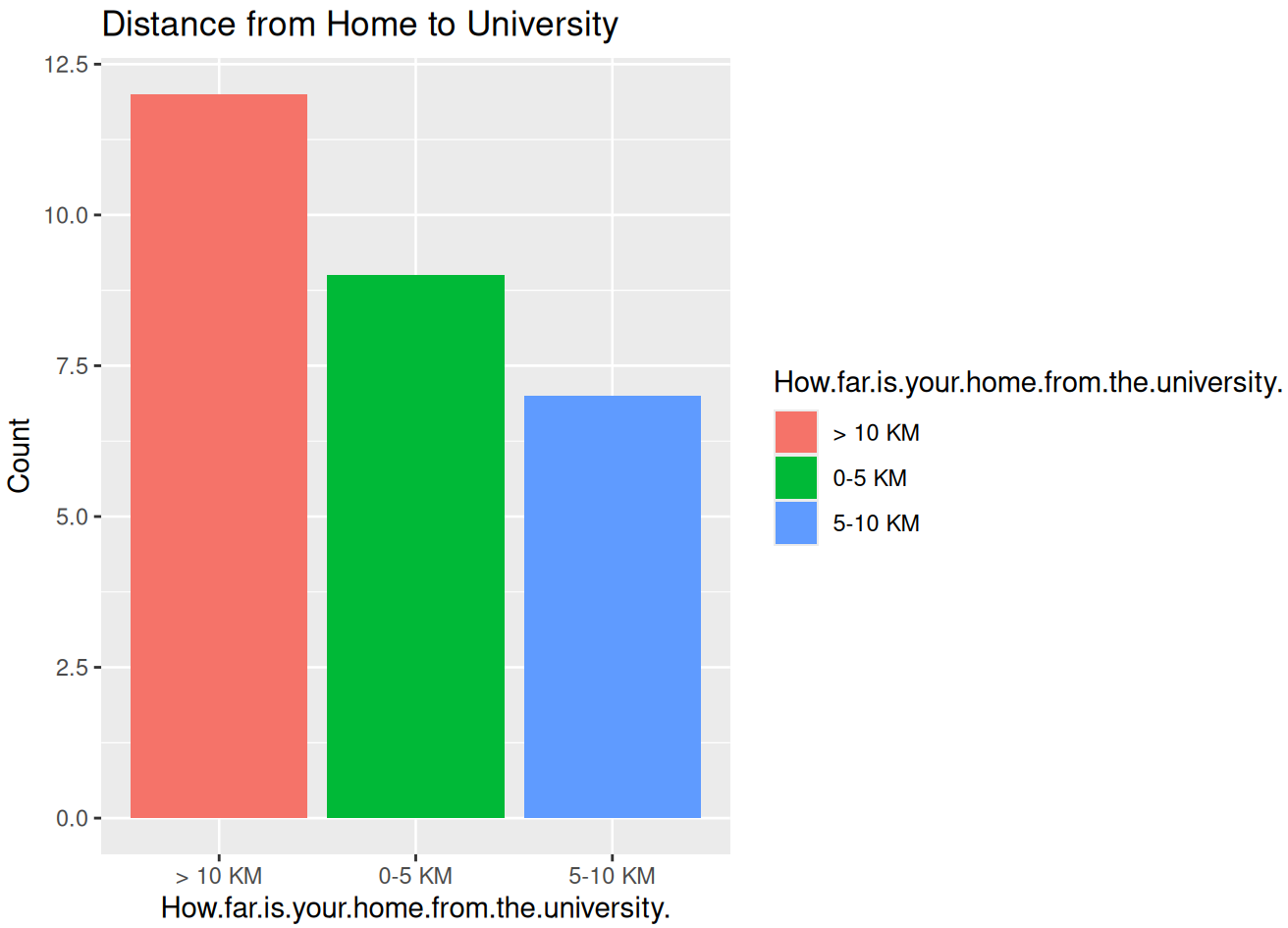
```
ggplot(year_dist, aes(x = Year.of.Study, y = Count, fill = Year.of.Study)) +
  geom_bar(stat = "identity") +
  ggtitle("Year of Study Distribution")
```



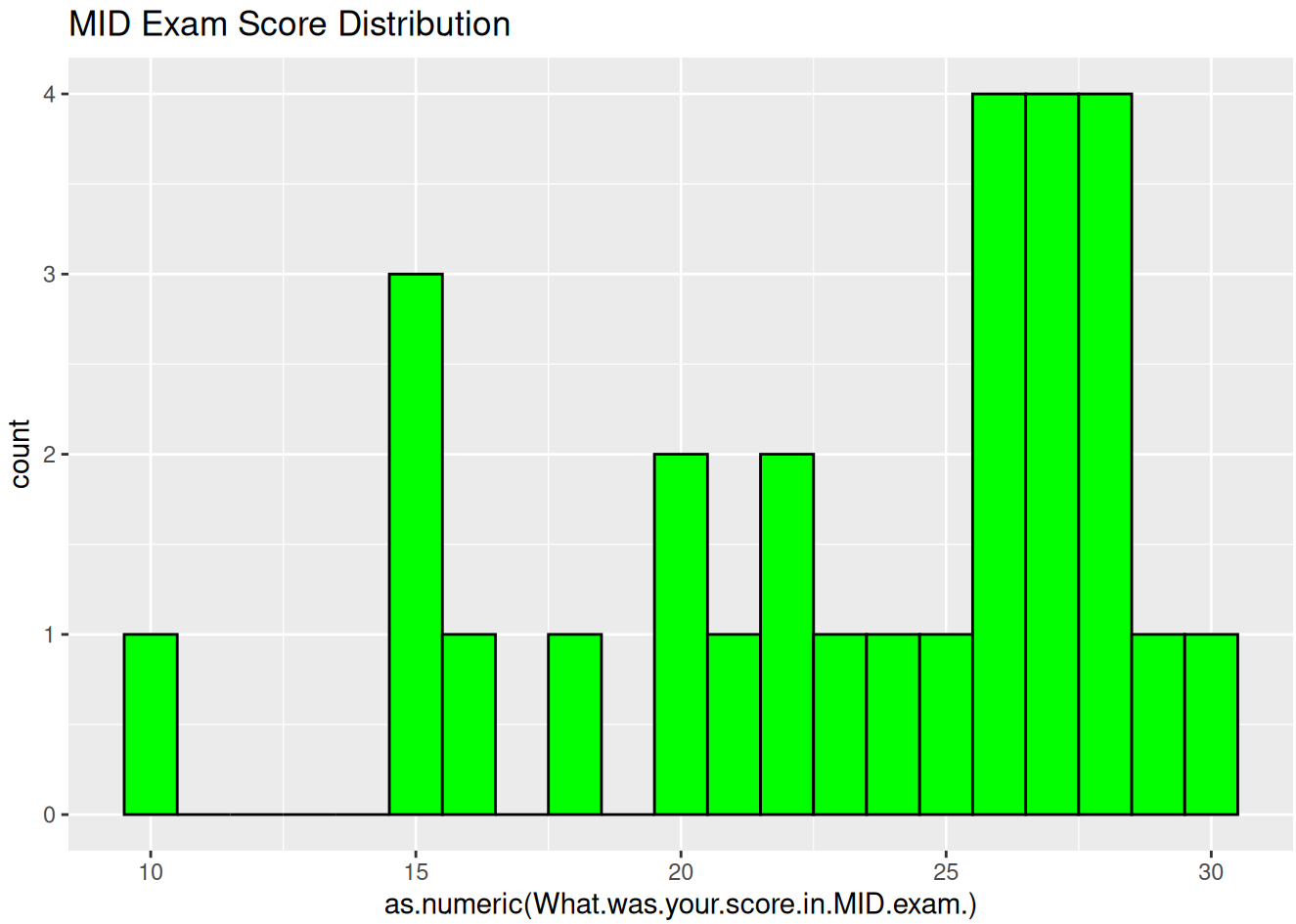

```
# Geographical Insights
home_distance <- survey_data %>%
  group_by(How.far.is.your.home.from.the.university.) %>%
  summarise(Count = n())
print(home_distance)
```

```
# A tibble: 3 × 2
  How.far.is.your.home.from.the.university. Count
  <chr>                                     <int>
1 0-5 KM                                        9
2 5-10 KM                                       7
3 > 10 KM                                      12
```

```
ggplot(home_distance, aes(x = How.far.is.your.home.from.the.university., y = Count, fill =
  geom_bar(stat = "identity") +
  ggtitle("Distance from Home to University")
```



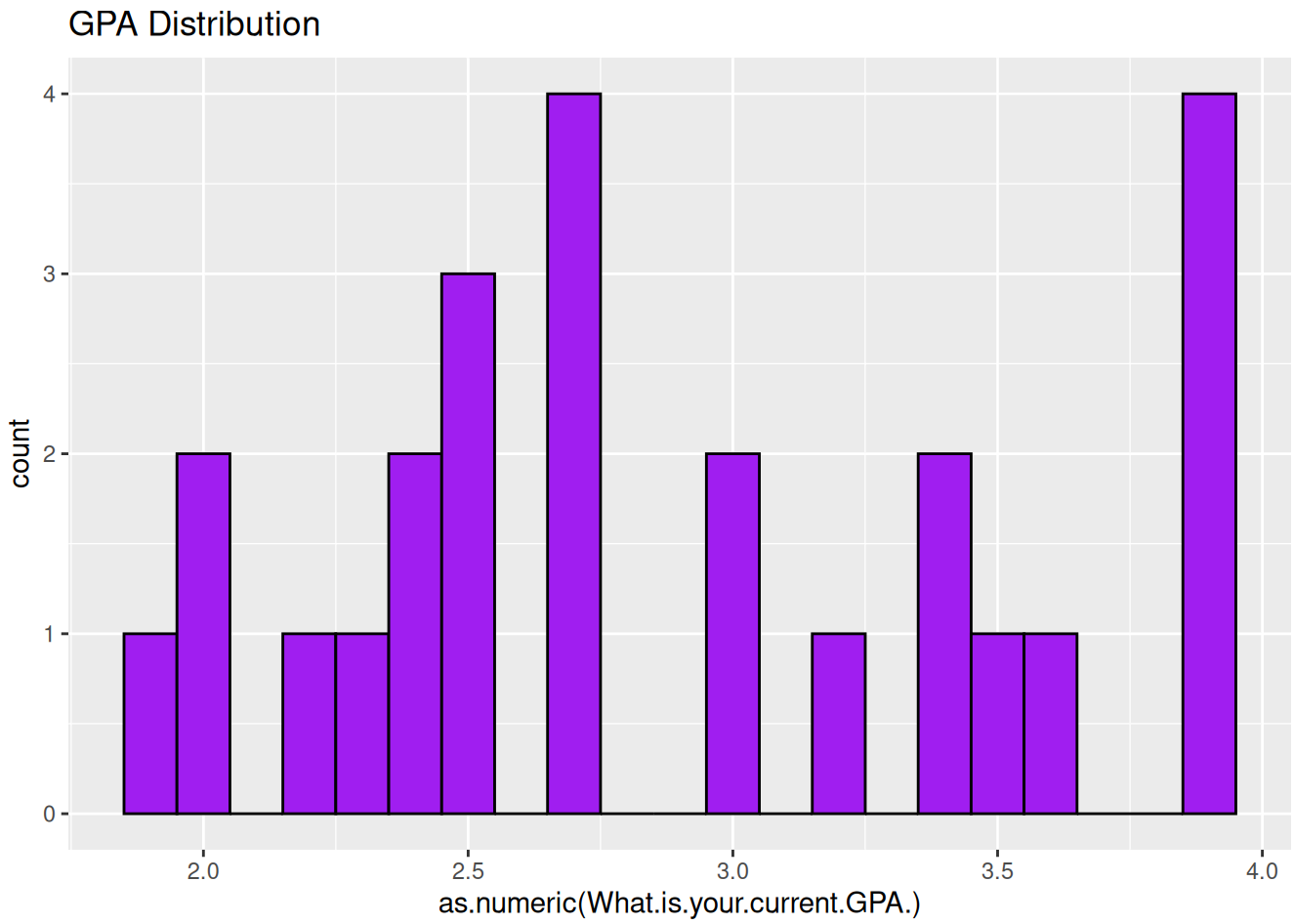

```
# Academic Performance
# MID Exam Scores
ggplot(survey_data, aes(x = as.numeric(What.was.your.score.in.MID.exam.))) +
  geom_histogram(binwidth = 1, fill = "green", color = "black") +
  ggtitle("MID Exam Score Distribution")
```




```
# Current GPA Distribution
ggplot(survey_data, aes(x = as.numeric(What.is.your.current.GPA.))) +
  geom_histogram(binwidth = 0.1, fill = "purple", color = "black") +
  ggtitle("GPA Distribution")
```

```
Warning in FUN(X[[i]], ...): NAs introduced by coercion

Warning: Removed 3 rows containing non-finite outside the scale range
(`stat_bin()`).
```




```
# Work-Study Relationship
work_status_gpa <- survey_data %>%
  group_by(Do.you.work.while.studying.) %>%
  summarise(Average_GPA = mean(as.numeric(What.is.your.current.GPA.), na.rm = TRUE))
```

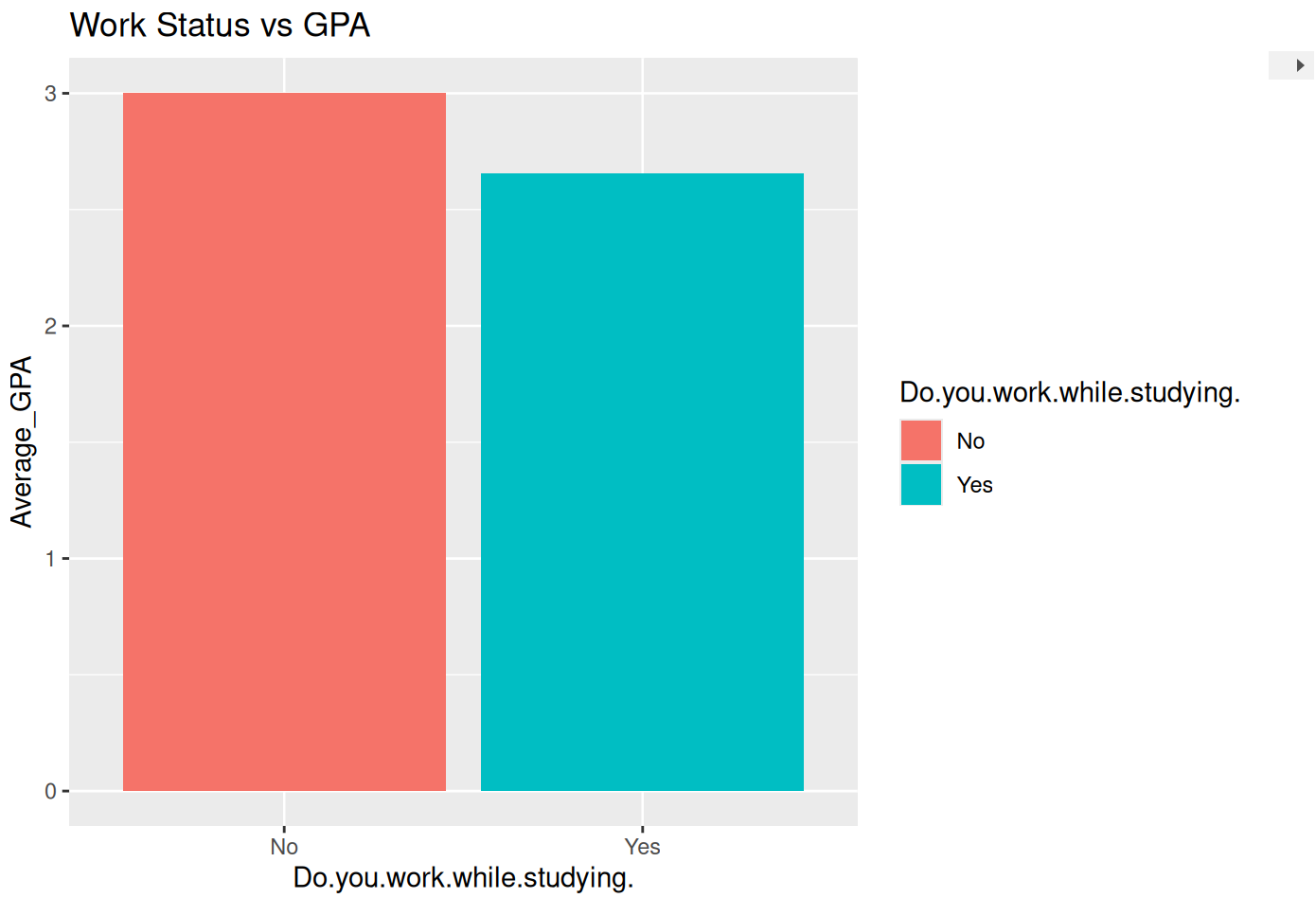
```
Warning: There was 1 warning in `summarise()`.
ℹ In argument: `Average_GPA = mean(as.numeric(What.is.your.current.GPA.), na.rm
  = TRUE)`.
ℹ In group 2: `Do.you.work.while.studying. = "Yes"`.
Caused by warning in `mean()`:
! NAs introduced by coercion
```

```
print(work_status_gpa)
```

```
# A tibble: 2 × 2
  Do.you.work.while.studying. Average_GPA
  <chr>                             <dbl>
1 No                                 3.00
2 Yes                                2.66
```

```
ggplot(work_status_gpa, aes(x = Do.you.work.while.studying., y = Average_GPA, fill = Do.y
  geom_bar(stat = "identity") +
```

```
  ggtitle("Work Status vs GPA")
```




```
# Satisfaction Ratings
satisfaction_university <- survey_data %>%
  summarise(Average_Satisfaction = mean(as.numeric(How.satisfied.are.you.about.your.exper
print(satisfaction_university)
```

```
  Average_Satisfaction
1                 3.75
```

```
satisfaction_major <- survey_data %>%
  summarise(Average_Satisfaction = mean(as.numeric(How.satisfied.are.you.about.the.Busine
print(satisfaction_major)
```

```
  Average_Satisfaction
1                 4.25
```

```
# Likelihood to Recommend
likelihood_recommend <- survey_data %>%
  summarise(
```

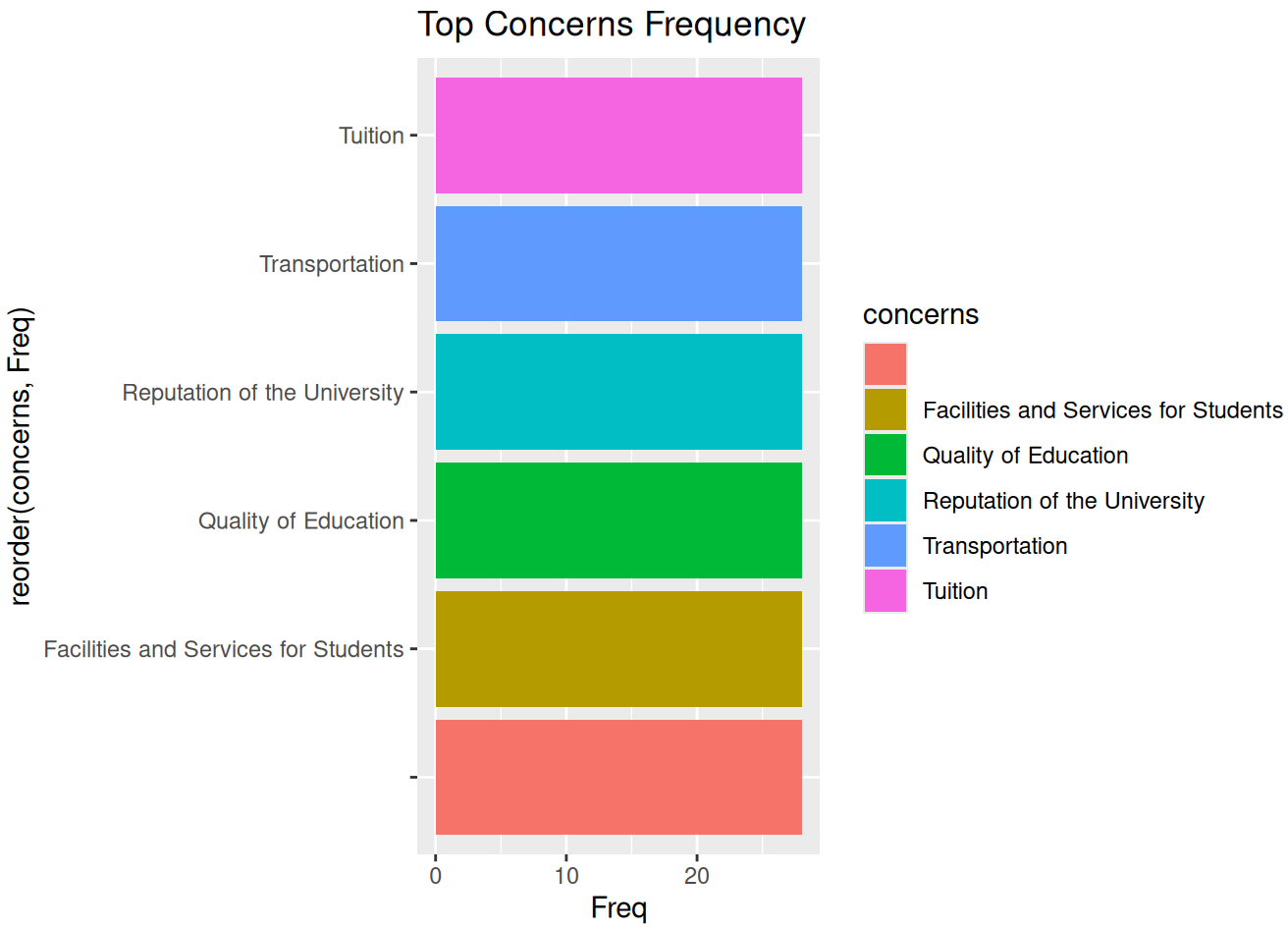
```
    Recommend_University = mean(as.numeric(How.likely.are.you.to.recommend.the.University
    Recommend_Major = mean(as.numeric(How.likely.are.you.to.recommend.studying.Business.I
)
print(likelihood_recommend)
```

```
  Recommend_University Recommend_Major
1             7.481481        7.714286
```

```
# Concerns Analysis
library(stringr)
concerns <- unlist(str_split(survey_data$Rank.your.top.concerns.about.your.experience.in.
concerns_freq <- as.data.frame(table(concerns))
concerns_freq <- concerns_freq %>% arrange(desc(Freq))
print(concerns_freq)
```

```
                            concerns Freq
1                                      28
2 Facilities and Services for Students   28
3                 Quality of Education   28
4         Reputation of the University   28
5                       Transportation   28
6                              Tuition   28
```

```
ggplot(concerns_freq, aes(x = reorder(concerns, Freq), y = Freq, fill = concerns)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  ggtitle("Top Concerns Frequency")
```



Here are some hypothesis tests you can perform on your survey data, including **t-tests**, **chi-square tests**, and **ANOVA**:

---

1. T-Tests

Hypothesis:

- Students who work have different GPAs compared to those who do not.

Null Hypothesis ((H_0)):

There is no difference in the mean GPA between students who work and those who do not.

Alternative Hypothesis ((H_a)):

There is a difference in the mean GPA between students who work and those who do not.

```
# Perform a t-test
t_test_result <- t.test(
  as.numeric(What.is.your.current.GPA.) ~ Do.you.work.while.studying.,
  data = survey_data,
  na.rm = TRUE
)
print(t_test_result)
```

2. Chi-Square Test

Hypothesis:

- Satisfaction with the university is independent of gender.

Null Hypothesis ((H_0)):

Satisfaction with the university is independent of gender.

Alternative Hypothesis ((H_a)):

Satisfaction with the university is not independent of gender.

```
# Create a contingency table
satisfaction_gender_table <- table(
  survey_data$Gender,
  as.numeric(survey_data$How.satisfied.are.you.about.your.experience.at.the.University.of
)

# Perform a chi-square test
chi_square_result <- chisq.test(satisfaction_gender_table)
print(chi_square_result)
```

---

3. ANOVA (Analysis of Variance)

Hypothesis:

- Students from different years of study have different levels of satisfaction with the university.

Null Hypothesis ((H_0)):

There is no difference in mean satisfaction with the university across years of study.

Alternative Hypothesis ((H_a)):

There is a difference in mean satisfaction with the university across years of study.

```
# Perform ANOVA
anova_result <- aov(
  as.numeric(How.satisfied.are.you.about.your.experience.at.the.University.of.Petra.) ~ Y
  data = survey_data
)
summary(anova_result)
```

Additional Ideas

1. **T-Test for Recommendation**:

- Compare the likelihood to recommend the university between male and female students.

2. **Chi-Square for High School Type**:
   - Test if high school type (National vs. International) is associated with working while studying.
3. **ANOVA for Distance**:
   - Analyze if the distance from home to university influences satisfaction.