

FinalProject in R

Keir G. Sumayo, Daven Parita, James Harry Prado

2025-12-14

```
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)

transport <- read.csv(
  "Transportation Uses for BSIT Students (Responses).csv",
  stringsAsFactors = FALSE
)

transport$Timestamp <- NULL
transport$X. <- NULL

transport <- transport %>%
  mutate(across(where(is.character), ~ gsub("^Bh$", "BH", .)))

data <- transport %>%
  rename(
    Q1 = X1..Where.do.you.live.,
    Q2 = X2..Please.indicate.how.many.rides.you.take.to.reach.the.school.,
    Q3 = X3..How.long.does.it.take.for.you.to.reach.school.from.home.,
    Q3A = a..How.much.is.your.one.way.fare.to.school.,
    Q3B = b..How.much.is.your.total.daily.fare..round.trip.,
    Feedback = What.are.your.thoughts.or.feedback.about.this.survey.
  )

str(transport)

## 'data.frame': 52 obs. of 10 variables:
```

```

## $ Name..Optional. : chr "Richele Pacot " "Mark Antl
## $ Age : int 19 19 20 19 19 20 69 18 19
## $ Sex : chr "Female" "Male" "Prefer no
## $ Year.Level : chr "2nd Year" "2nd Year" "2nd
## $ X1..Where.do.you.live. : chr "Family House" "Family Hou
## $ X2..Please.indicate.how.many.rides.you.take.to.reach.the.school.: chr "2 rides" "2 rides" "1 ride
## $ X3..How.long.does.it.take.for.you.to.reach.school.from.home. : chr "30-60 minutes" "30-60 min
## $ a..How.much.is.your.one.way.fare.to.school. : chr "P26 - P45" "P15 - P25" "P
## $ b..How.much.is.your.total.daily.fare..round.trip.. : chr "P76 - P100" "P50 - P75" "P
## $ What.are.your.thoughts.or.feedback.about.this.survey. : chr "Oks.lang" "None" "It may r

```

```
head(transport)
```

	Name..Optional.	Age	Sex	Year.Level
## 1	Richele Pacot	19	Female	2nd Year
## 2	Mark Anthony S. Bernarte	19	Male	2nd Year
## 3	Keir	20	Prefer not to say	2nd Year
## 4		19	Male	2nd Year
## 5		19	Female	2nd Year
## 6	Lean	20	Female	1st Year
## X1..Where.do.you.live.				
## 1	Family House			
## 2	Family House			
## 3	Family House			
## 4	BH			
## 5	Relative's House			
## 6	Family House			
## X2..Please.indicate.how.many.rides.you.take.to.reach.the.school.				
## 1			2 rides	
## 2			2 rides	
## 3			1 ride	
## 4			2 rides	
## 5			2 rides	
## 6			1 ride	
## X3..How.long.does.it.take.for.you.to.reach.school.from.home.				
## 1			30-60 minutes	
## 2			30-60 minutes	
## 3			15-30 minutes	
## 4			15-30 minutes	
## 5			15-30 minutes	
## 6			30-60 minutes	
## a..How.much.is.your.one.way.fare.to.school.				
## 1		P26 - P45		
## 2		P15 - P25		
## 3		P46 and above		
## 4		P26 - P45		
## 5		P26 - P45		
## 6		P46 and above		
## b..How.much.is.your.total.daily.fare..round.trip..				
## 1		P76 - P100		
## 2		P50 - P75		
## 3		P150 and above		
## 4		P76 - P100		
## 5		P76 - P100		

```

## 6                                     P50 - P75
##                                         What.are.your.thoughts.or.feedback.about.this.survey.
## 1                                         Oks.lang
## 2                                         None
## 3 It may not be perfect but it's nice, keep up the good work and achieve your goals
## 4                                         Good
## 5 The survey is clear and specific. It focuses on commuting habits of the students.
## 6                                         Good

```

```
colSums(is.na(transport))
```

```

##                                         Name..Optional.
##                                         0
##                                         Age
##                                         0
##                                         Sex
##                                         0
##                                         Year.Level
##                                         0
##                                         X1..Where.do.you.live.
##                                         0
##                                         X2..Please.indicate.how.many.rides.you.take.to.reach.the.school.
##                                         0
##                                         X3..How.long.does.it.take.for.you.to.reach.school.from.home.
##                                         0
##                                         a..How.much.is.your.one.way.fare.to.school.
##                                         0
##                                         b..How.much.is.your.total.daily.fare..round.trip..
##                                         0
##                                         What.are.your.thoughts.or.feedback.about.this.survey.
##                                         0

```

```

write.csv(
  data,
  "Transportation_Uses_Cleaned.csv",
  row.names = FALSE
)

```

```

data <- read.csv("Transportation_Uses_Cleaned.csv")
str(data)

```

```

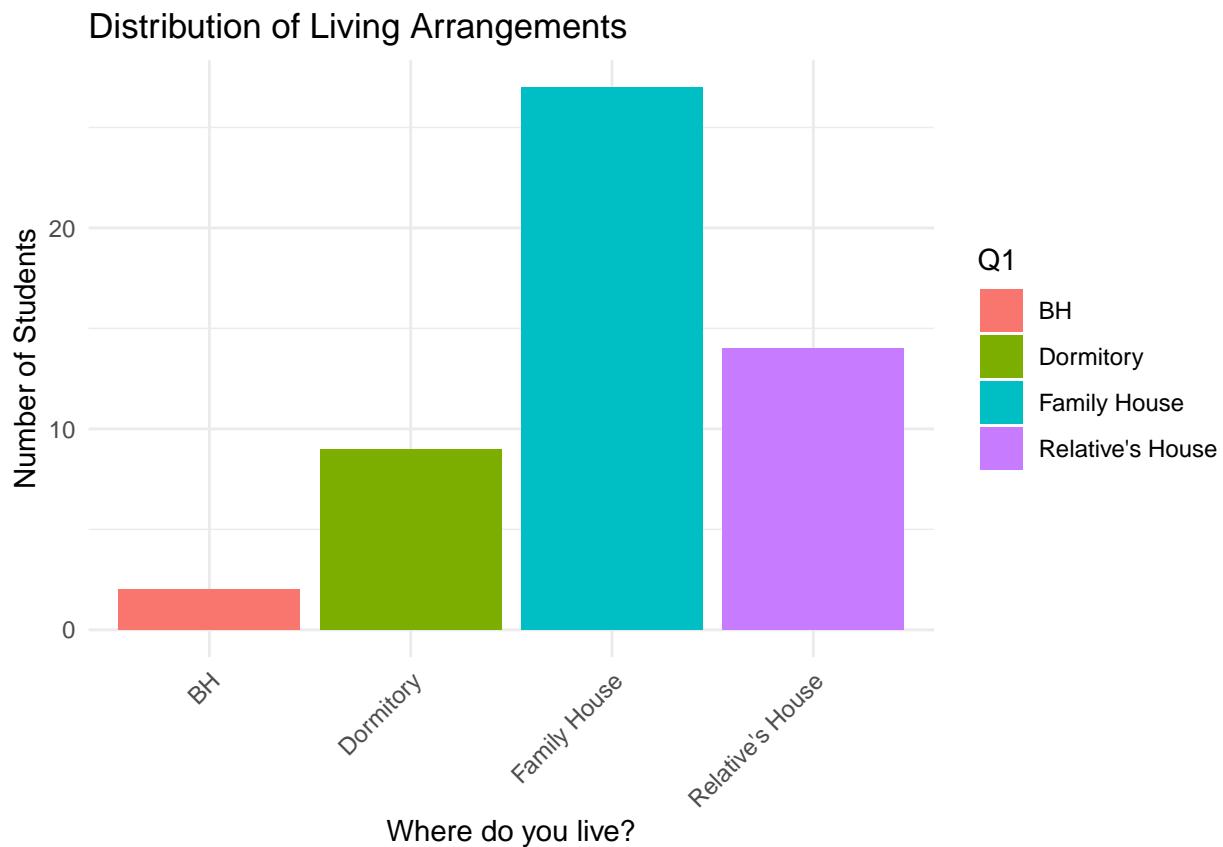
## 'data.frame':   52 obs. of  10 variables:
## $ Name..Optional.: chr  "Richele Pacot" "Mark Anthony S. Bernarte" "Keir" " "
## $ Age            : int  19 19 20 19 19 20 69 18 19 21 ...
## $ Sex            : chr  "Female" "Male" "Prefer not to say" "Male" ...
## $ Year.Level    : chr  "2nd Year" "2nd Year" "2nd Year" "2nd Year" ...
## $ Q1             : chr  "Family House" "Family House" "Family House" "BH" ...
## $ Q2             : chr  "2 rides" "2 rides" "1 ride" "2 rides" ...
## $ Q3             : chr  "30-60 minutes" "30-60 minutes" "15-30 minutes" "15-30 minutes" ...
## $ Q3A            : chr  "P26 - P45" "P15 - P25" "P46 and above" "P26 - P45" ...
## $ Q3B            : chr  "P76 - P100" "P50 - P75" "P150 and above" "P76 - P100" ...
## $ Feedback       : chr  "Oks.lang" "None" "It may not be perfect but it's nice, keep up the good wo

```

The bar chart titled “Distribution of Living Arrangements” shows where BSIT students from ISATU Lapaz live. The x-axis lists four categories: BH (Boarding House), Dormitory, Family House, and Relative’s House, while the y-axis shows the number of students in each category.

```
living_counts <- data %>%
  count(Q1)

ggplot(living_counts, aes(x = Q1, y = n, fill = Q1)) +
  geom_bar(stat = "identity") +
  labs(title = "Distribution of Living Arrangements",
       x = "Where do you live?",
       y = "Number of Students") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



This reflects common living patterns among college students in Iloilo City, where many commute from home, while others choose affordable boarding options near their school.

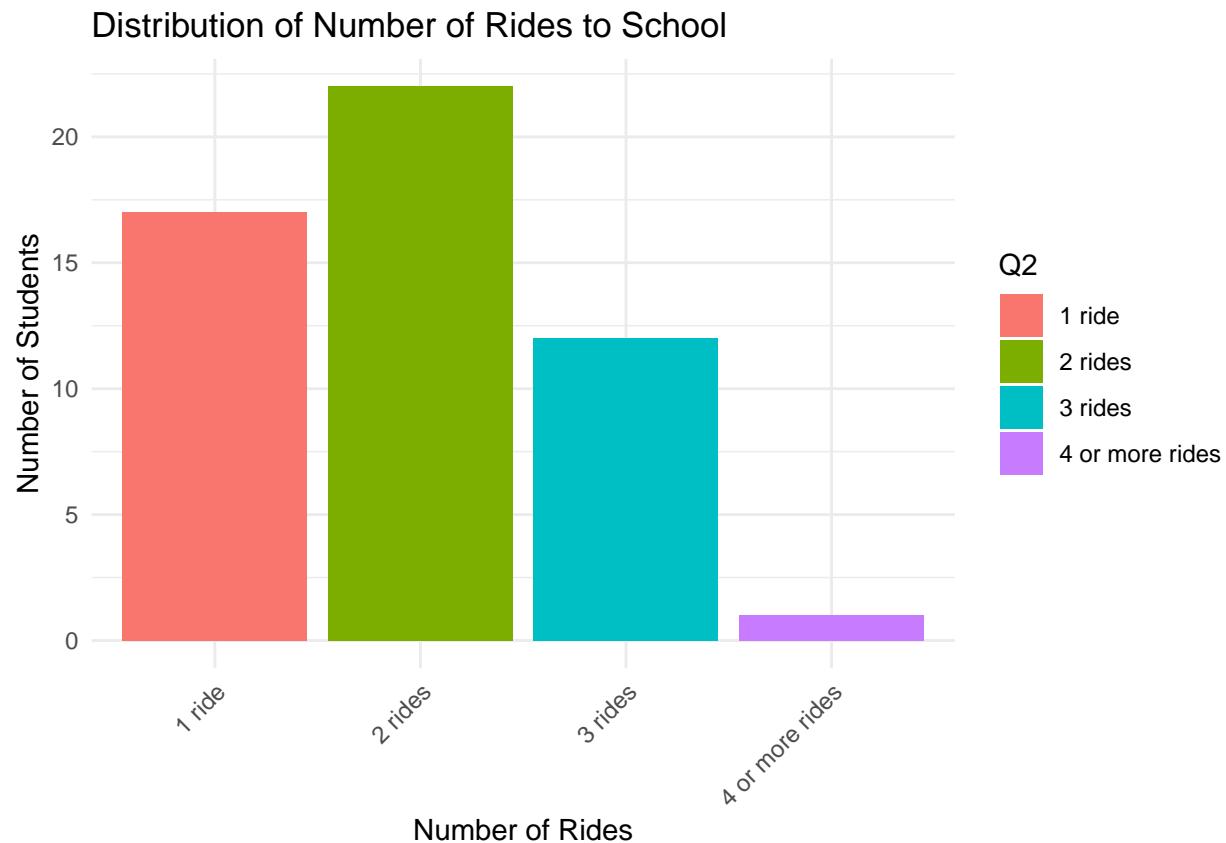
The bar chart titled “Distribution of Number of Rides to School” shows how many BSIT students from ISATU Lapaz take to reach campus. The x-axis represents the number of rides (1 ride, 2 rides, 3 rides, 4 or more rides), while the y-axis shows the number of students in each category.

```

rides_counts <- data %>%
  count(Q2)

ggplot(rides_counts, aes(x = Q2, y = n, fill = Q2)) +
  geom_bar(stat = "identity") +
  labs(title = "Distribution of Number of Rides to School",
       x = "Number of Rides",
       y = "Number of Students") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



Most students enjoy relatively simple commutes, reflecting ISATU Lapaz's accessibility within the city's transport network.

The bar chart titled “Distribution of One-Way Fare to School” shows how much BSIT students spend on a single trip to school. The x-axis displays fare ranges in Philippine pesos (₱), while the y-axis shows the number of students in each category.

```

time_counts <- data %>%
  count(Q3)

ggplot(time_counts, aes(x = Q3, y = n, fill = Q3)) +
  geom_bar(stat = "identity") +
  labs(title = "Distribution of Travel Time to School",
       x = "Travel Time Range (Q3)",
       y = "Number of Students")

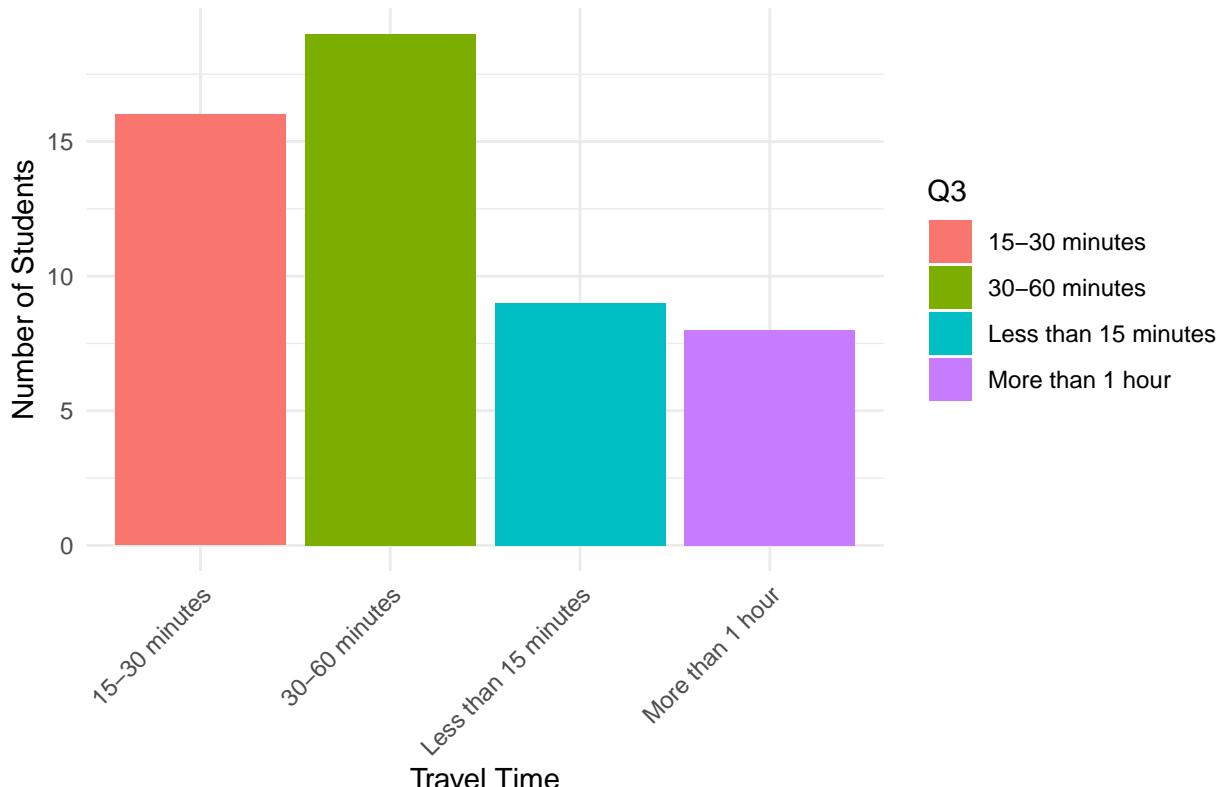
```

```

x = "Travel Time",
y = "Number of Students") +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

Distribution of Travel Time to School



The concentration of fares in this range reflects the affordability and accessibility of public transportation in Iloilo, supporting student mobility across the city. With safe, well-established routes and a student discount policy, most students can commute efficiently without excessive financial burden.

The bar chart titled “Distribution of One-Way Fare to School” shows how much BSIT students spend on a single trip to school. The x-axis displays fare ranges in Philippine pesos (₱), while the y-axis shows the number of students in each category.

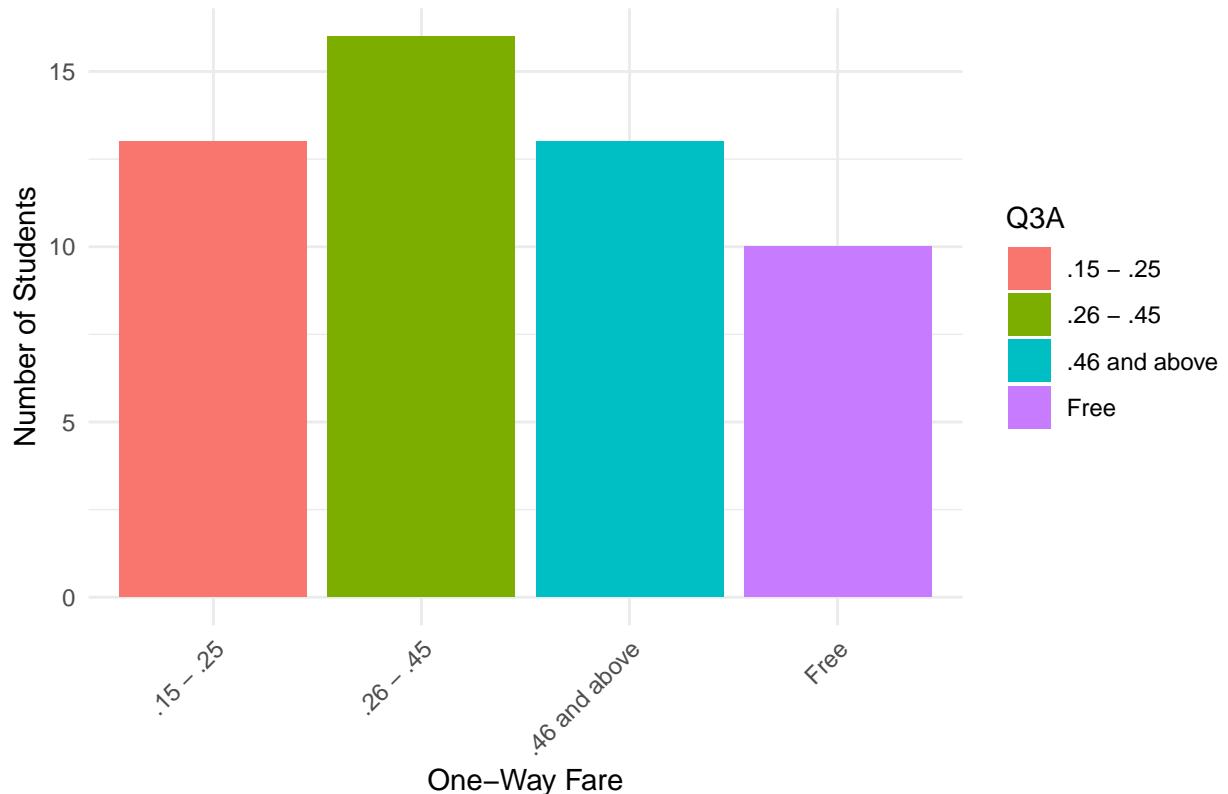
```

one_way_fare_counts <- data %>%
  count(Q3A)

ggplot(one_way_fare_counts, aes(x = Q3A, y = n, fill = Q3A)) +
  geom_bar(stat = "identity") +
  labs(title = "Distribution of One-Way Fare to School",
       x = "One-Way Fare",
       y = "Number of Students") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

Distribution of One-Way Fare to School



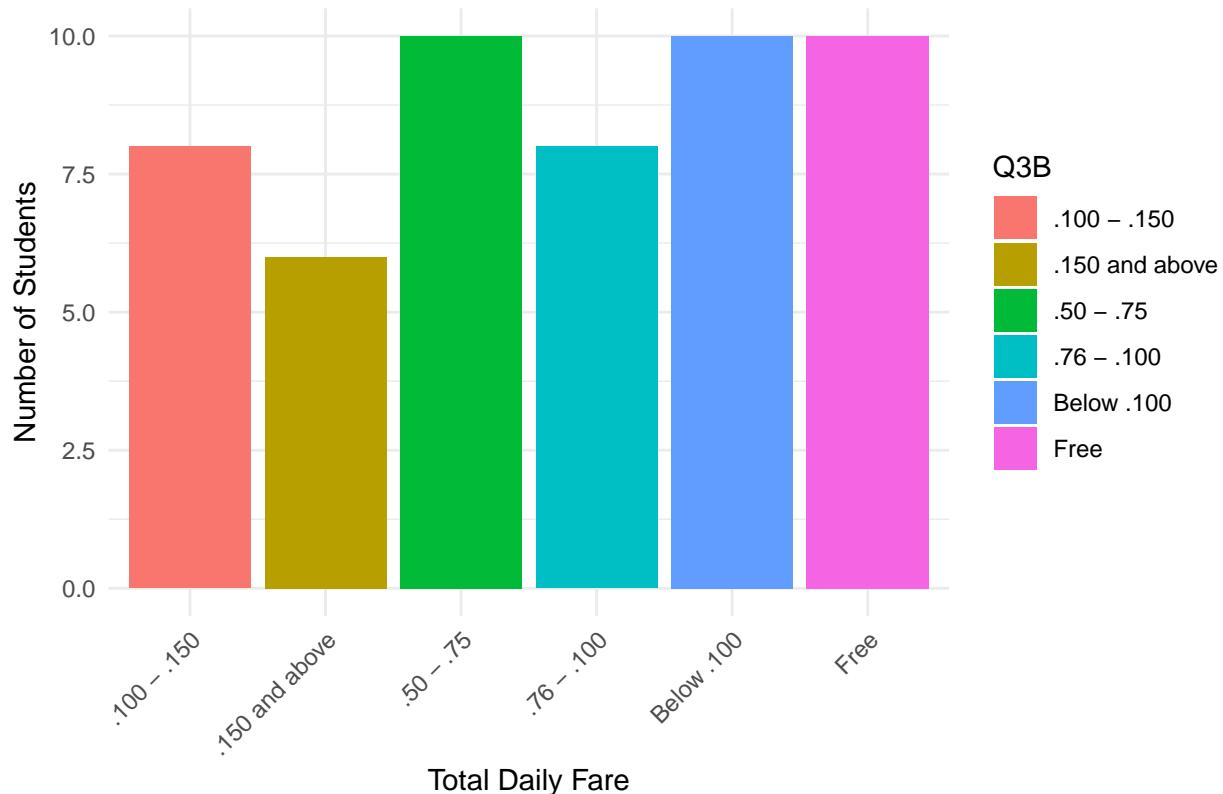
For students traveling from areas like Ungka, La Paz Plaza, or nearby barangays to ISATU, a one-way trip typically involves a short to moderate distance, resulting in fares between ₱13 and ₱45, depending on the starting point.

The bar chart titled “Distribution of Total Daily Fare (Round Trip)” shows how much BSIT students spend on their daily round-trip commute to school. The x-axis represents different fare ranges in Philippine pesos (₱), while the y-axis shows the number of students in each category.

```
total_fare_counts <- data %>%
  count(Q3B)

ggplot(total_fare_counts, aes(x = Q3B, y = n, fill = Q3B)) +
  geom_bar(stat = "identity") +
  labs(title = "Distribution of Total Daily Fare (Round Trip)",
       x = "Total Daily Fare",
       y = "Number of Students") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Distribution of Total Daily Fare (Round Trip)



This fare distribution reflects the affordability and accessibility of public transport in Iloilo, supporting student mobility across the city. With safe, well-established routes and a student-friendly discount system, most students can commute efficiently without excessive financial burden.

The bar chart titled “Sentiment Distribution” displays feedback classified into three categories: positive, neutral, and negative, with counts on the y-axis. Based on the visual, there are very few responses in each category, only a small number of feedback entries were provided by respondents.

```
df <- data.frame(
  Feedback = c(
    "It may not be perfect but it's nice, keep up the good work",
    "Good",
    "Oks lang",
    "None",
    "This is confusing and has a bug",
    "No option available",
    "Very helpful and clear"
  )
)

feedbacks <- df$Feedback[!is.na(df$Feedback)]

# Define keywords
positive_words <- c("good", "nice", "appreciate", "easy", "clear", "helpful", "perfect",
                     "keep up", "achieve", "relevant", "helps", "hope", "analyze", "not bad")
```

```

negative_words <- c("confusing", "wrong", "bugging", "problem", "no option", "bad")

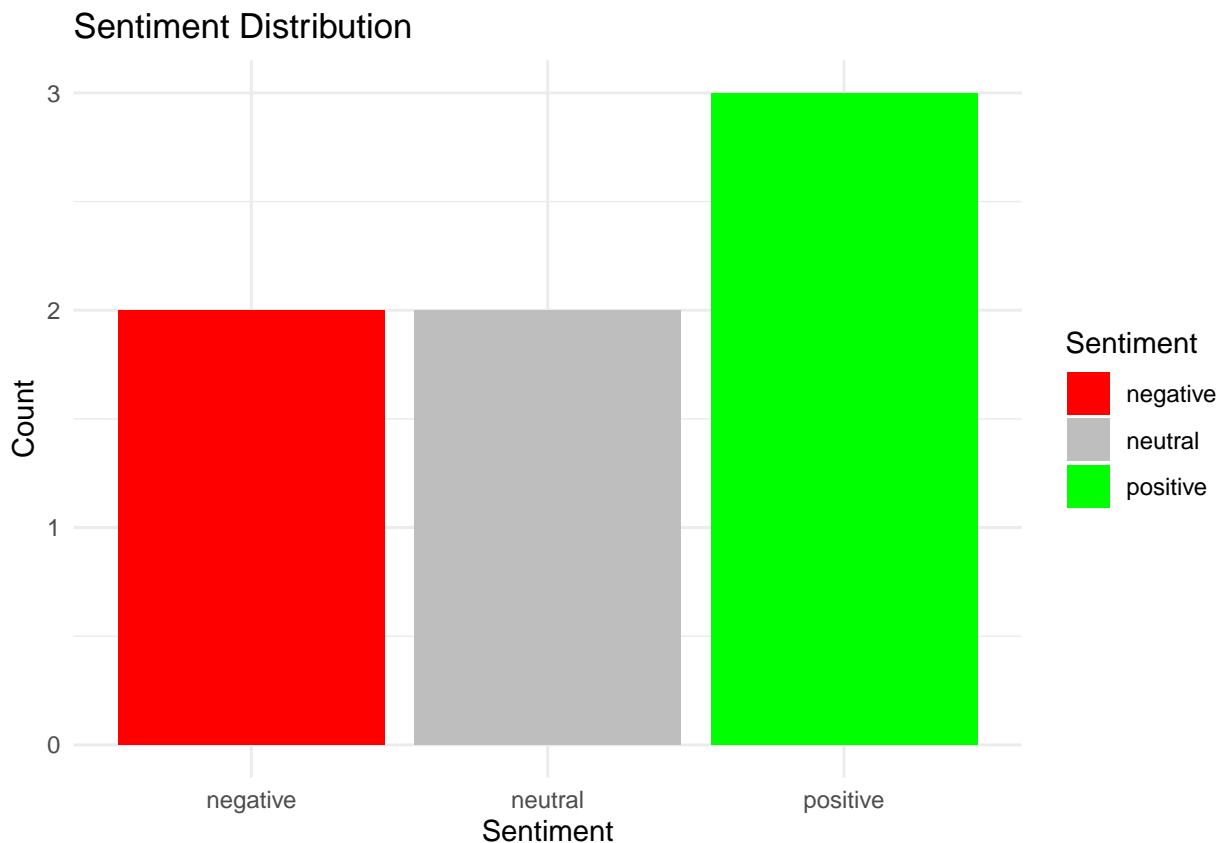
# Sentiment classification function
classify_sentiment <- function(fb) {
  fb_lower <- tolower(fb)
  if (any(sapply(negative_words, grepl, fb_lower, fixed = TRUE))) {
    return("negative")
  } else if (any(sapply(positive_words, grepl, fb_lower, fixed = TRUE))) {
    return("positive")
  } else {
    return("neutral")
  }
}

sentiments <- sapply(feedbacks, classify_sentiment)

sentiment_counts <- table(sentiments)
sentiment_df <- as.data.frame(sentiment_counts)
colnames(sentiment_df) <- c("Sentiment", "Count")

ggplot(sentiment_df, aes(x = Sentiment, y = Count, fill = Sentiment)) +
  geom_col() +
  labs(title = "Sentiment Distribution", x = "Sentiment", y = "Count") +
  theme_minimal() +
  scale_fill_manual(values = c("positive" = "green", "neutral" = "gray", "negative" = "red"))

```



key observation is that most respondents did not provide any feedback at all, leaving the field blank, marked

as “NA”, or writing minimal responses like “None”. This explains the low counts in the chart. As a result, the sample of readable, meaningful feedback is quite limited, which reduces the overall representativeness of the sentiment analysis.