18CSC403 **/** Practical Techniques for Big Data Processing

**Basics of R and Exploratory Data Analysis (EDA)**

**ROLL NUM:** CB.SC.I5DAS21049

**NAME:** Pratiksha Cauvery K P

**Analysis Questions about the dataset:**

1. What is the distribution of different age groups in the dataset?
2. How does the frequency of listening to music vary across different age groups?
3. What is the most common genre of music preferred by people in the dataset?
4. Is there a correlation between the time spent listening to music and self-reported mental health status?
5. How does the impact of music on mental health differ between genres?
6. What is the average time spent listening to music per day?
7. How does the preference for listening to music while working vary?
8. How do stress levels vary with different genres of music?
9. What is the distribution of music preference based on mental health status?
10. Does music actually have a significantly positive impact on a person’s life?

**Descriptive Analysis and Exploratory Data Analysis (EDA):**

IMPORTING LIBRARIES AND DATASET TO CURRENT WORKING DIRECTORY:

setwd("/Users/pratiksha/Desktop/Sem7/Practical Techniques for Big Data Processing/Assignment-1/")

> install.packages("readr")

> library(readr)

> music\_mental\_health <- read.csv("mxmh.csv")

STRUCTURE AND SUMMARY OF THE DATASET:

> head(music\_mental\_health)

Timestamp Age Primary.streaming.service Hours.per.day While.working Instrumentalist

1 8/27/2022 19:29:02 18 Spotify 3.0 Yes Yes

2 8/27/2022 19:57:31 63 Pandora 1.5 Yes No

3 8/27/2022 21:28:18 18 Spotify 4.0 No No

4 8/27/2022 21:40:40 61 YouTube Music 2.5 Yes No

5 8/27/2022 21:54:47 18 Spotify 4.0 Yes No

6 8/27/2022 21:56:50 18 Spotify 5.0 Yes Yes

Composer Fav.genre Exploratory Foreign.languages BPM Frequency..Classical.

1 Yes Latin Yes Yes 156 Rarely

2 No Rock Yes No 119 Sometimes

3 No Video game music No Yes 132 Never

4 Yes Jazz Yes Yes 84 Sometimes

5 No R&B Yes No 107 Never

6 Yes Jazz Yes Yes 86 Rarely

Frequency..Country. Frequency..EDM. Frequency..Folk. Frequency..Gospel. Frequency..Hip.hop.

1 Never Rarely Never Never Sometimes

2 Never Never Rarely Sometimes Rarely

3 Never Very frequently Never Never Rarely

4 Never Never Rarely Sometimes Never

5 Never Rarely Never Rarely Very frequently

6 Sometimes Never Never Never Sometimes

Frequency..Jazz. Frequency..K.pop. Frequency..Latin. Frequency..Lofi. Frequency..Metal.

1 Never Very frequently Very frequently Rarely Never

2 Very frequently Rarely Sometimes Rarely Never

3 Rarely Very frequently Never Sometimes Sometimes

4 Very frequently Sometimes Very frequently Sometimes Never

5 Never Very frequently Sometimes Sometimes Never

6 Very frequently Very frequently Rarely Very frequently Rarely

Frequency..Pop. Frequency..R.B. Frequency..Rap. Frequency..Rock. Frequency..Video.game.music.

1 Very frequently Sometimes Very frequently Never Sometimes

2 Sometimes Sometimes Rarely Very frequently Rarely

3 Rarely Never Rarely Rarely Very frequently

4 Sometimes Sometimes Never Never Never

5 Sometimes Very frequently Very frequently Never Rarely

6 Very frequently Very frequently Very frequently Very frequently Never

Anxiety Depression Insomnia OCD Music.effects Permissions

1 3 0 1 0 I understand.

2 7 2 2 1 I understand.

3 7 7 10 2 No effect I understand.

4 9 7 3 3 Improve I understand.

5 7 2 5 9 Improve I understand.

6 8 8 7 7 Improve I understand.

str(music\_mental\_health)

'data.frame': 736 obs. of 33 variables:

$ Timestamp : chr "8/27/2022 19:29:02" "8/27/2022 19:57:31" "8/27/2022 21:28:18" "8/27/2022 21:40:40" ...

$ Age : int 18 63 18 61 18 18 18 21 19 18 ...

$ Primary.streaming.service : chr "Spotify" "Pandora" "Spotify" "YouTube Music" ...

$ Hours.per.day : num 3 1.5 4 2.5 4 5 3 1 6 1 ...

$ While.working : chr "Yes" "Yes" "No" "Yes" ...

$ Instrumentalist : chr "Yes" "No" "No" "No" ...

$ Composer : chr "Yes" "No" "No" "Yes" ...

$ Fav.genre : chr "Latin" "Rock" "Video game music" "Jazz" ...

$ Exploratory : chr "Yes" "Yes" "No" "Yes" ...

$ Foreign.languages : chr "Yes" "No" "Yes" "Yes" ...

$ BPM : int 156 119 132 84 107 86 66 95 94 155 ...

$ Frequency..Classical. : chr "Rarely" "Sometimes" "Never" "Sometimes" ...

$ Frequency..Country. : chr "Never" "Never" "Never" "Never" ...

$ Frequency..EDM. : chr "Rarely" "Never" "Very frequently" "Never" ...

$ Frequency..Folk. : chr "Never" "Rarely" "Never" "Rarely" ...

$ Frequency..Gospel. : chr "Never" "Sometimes" "Never" "Sometimes" ...

$ Frequency..Hip.hop. : chr "Sometimes" "Rarely" "Rarely" "Never" ...

$ Frequency..Jazz. : chr "Never" "Very frequently" "Rarely" "Very frequently" ...

$ Frequency..K.pop. : chr "Very frequently" "Rarely" "Very frequently" "Sometimes" ...

$ Frequency..Latin. : chr "Very frequently" "Sometimes" "Never" "Very frequently" ...

$ Frequency..Lofi. : chr "Rarely" "Rarely" "Sometimes" "Sometimes" ...

$ Frequency..Metal. : chr "Never" "Never" "Sometimes" "Never" ...

$ Frequency..Pop. : chr "Very frequently" "Sometimes" "Rarely" "Sometimes" ...

$ Frequency..R.B. : chr "Sometimes" "Sometimes" "Never" "Sometimes" ...

$ Frequency..Rap. : chr "Very frequently" "Rarely" "Rarely" "Never" ...

$ Frequency..Rock. : chr "Never" "Very frequently" "Rarely" "Never" ...

$ Frequency..Video.game.music.: chr "Sometimes" "Rarely" "Very frequently" "Never" ...

$ Anxiety : num 3 7 7 9 7 8 4 5 2 2 ...

$ Depression : num 0 2 7 7 2 8 8 3 0 2 ...

$ Insomnia : num 1 2 10 3 5 7 6 5 0 5 ...

$ OCD : num 0 1 2 3 9 7 0 3 0 1 ...

$ Music.effects : chr "" "" "No effect" "Improve" ...

$ Permissions : chr "I understand." "I understand." "I understand." "I understand." ...

>summary(music\_mental\_health)

Timestamp Age Primary.streaming.service Hours.per.day While.working

Length:736 Min. :10.00 Length:736 Min. : 0.000 Length:736

Class :character 1st Qu.:18.00 Class :character 1st Qu.: 2.000 Class :character

Mode :character Median :21.00 Mode :character Median : 3.000 Mode :character

Mean :25.21 Mean : 3.573

3rd Qu.:28.00 3rd Qu.: 5.000

Max. :89.00 Max. :24.000

NA's :1

Instrumentalist Composer Fav.genre Exploratory Foreign.languages

Length:736 Length:736 Length:736 Length:736 Length:736

Class :character Class :character Class :character Class :character Class :character

Mode :character Mode :character Mode :character Mode :character Mode :character

BPM Frequency..Classical. Frequency..Country. Frequency..EDM. Frequency..Folk.

Min. :0.00e+00 Length:736 Length:736 Length:736 Length:736

1st Qu.:1.00e+02 Class :character Class :character Class :character Class :character

Median :1.20e+02 Mode :character Mode :character Mode :character Mode :character

Mean :1.59e+06

3rd Qu.:1.44e+02

Max. :1.00e+09

NA's :107

Frequency..Gospel. Frequency..Hip.hop. Frequency..Jazz. Frequency..K.pop. Frequency..Latin.

Length:736 Length:736 Length:736 Length:736 Length:736

Class :character Class :character Class :character Class :character Class :character

Mode :character Mode :character Mode :character Mode :character Mode :character

Frequency..Lofi. Frequency..Metal. Frequency..Pop. Frequency..R.B. Frequency..Rap.

Length:736 Length:736 Length:736 Length:736 Length:736

Class :character Class :character Class :character Class :character Class :character

Mode :character Mode :character Mode :character Mode :character Mode :character

Frequency..Rock. Frequency..Video.game.music. Anxiety Depression Insomnia

Length:736 Length:736 Min. : 0.000 Min. : 0.000 Min. : 0.000

Class :character Class :character 1st Qu.: 4.000 1st Qu.: 2.000 1st Qu.: 1.000

Mode :character Mode :character Median : 6.000 Median : 5.000 Median : 3.000

Mean : 5.838 Mean : 4.796 Mean : 3.738

3rd Qu.: 8.000 3rd Qu.: 7.000 3rd Qu.: 6.000

Max. :10.000 Max. :10.000 Max. :10.000

OCD Music.effects Permissions

Min. : 0.000 Length:736 Length:736

1st Qu.: 0.000 Class :character Class :character

Median : 2.000 Mode :character Mode :character

Mean : 2.637

3rd Qu.: 5.000

Max. :10.000

EXPLORATORY DATA ANALYSIS (EDA)

1. Distribution of different age groups in the dataset

age\_distribution <- music\_mental\_health %>%

group\_by(Age) %>%

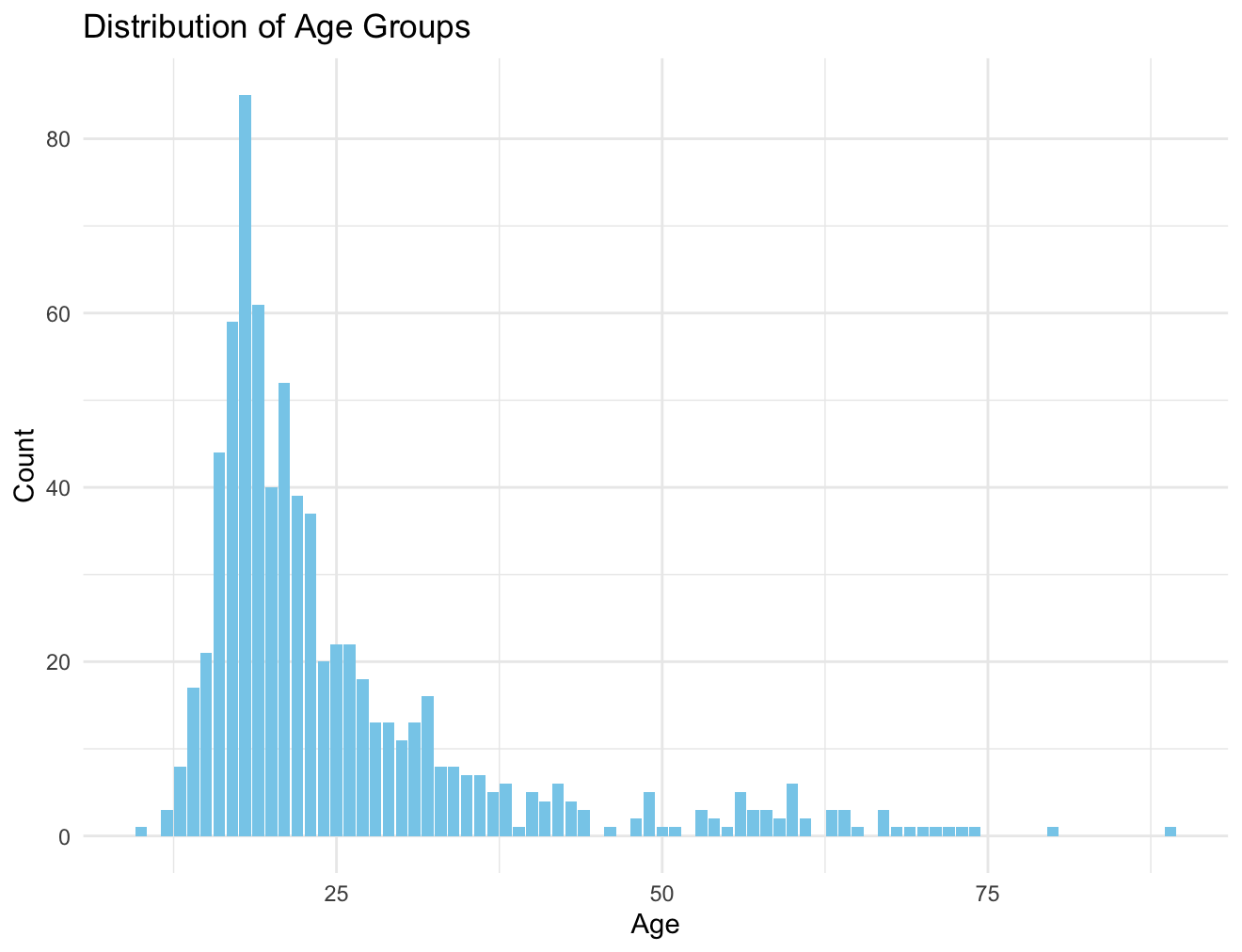
summarise(Count = n())

ggplot(age\_distribution, aes(x = Age, y = Count)) +

geom\_bar(stat = "identity", fill = "skyblue") +

theme\_minimal() +

labs(title = "Distribution of Age Groups", x = "Age", y = "Count")



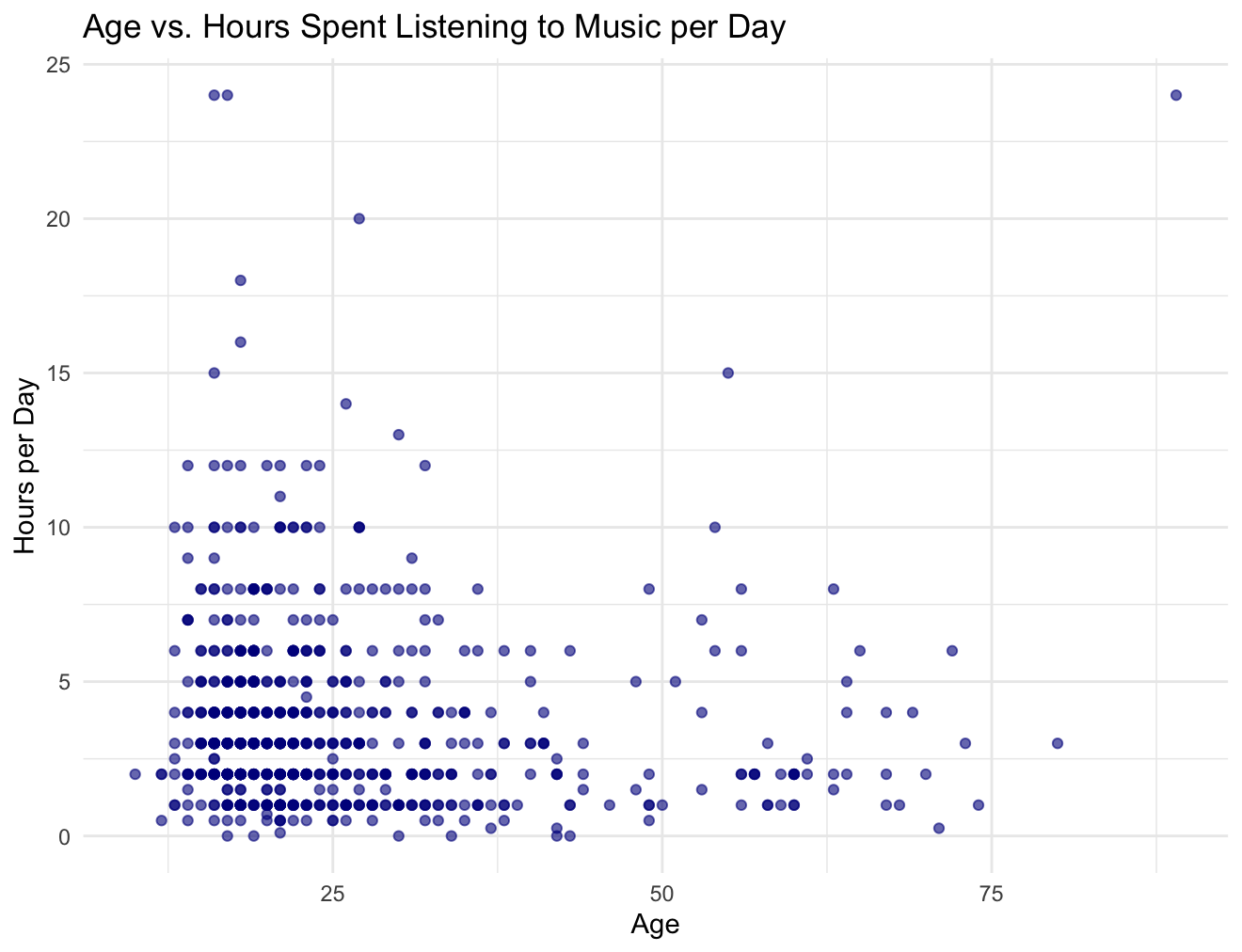
1. Age vs. Hours Spent Listening to Music per Day

ggplot(music\_mental\_health, aes(x = Age, y = Hours.per.day)) +

geom\_point(color = "darkblue", alpha = 0.6) +

theme\_minimal() +

labs(title = "Age vs. Hours Spent Listening to Music per Day", x = "Age", y = "Hours per Day")



1. Frequency of listening to different genres

classical\_frequency <- music\_mental\_health %>%

group\_by(Frequency..Classical.) %>%

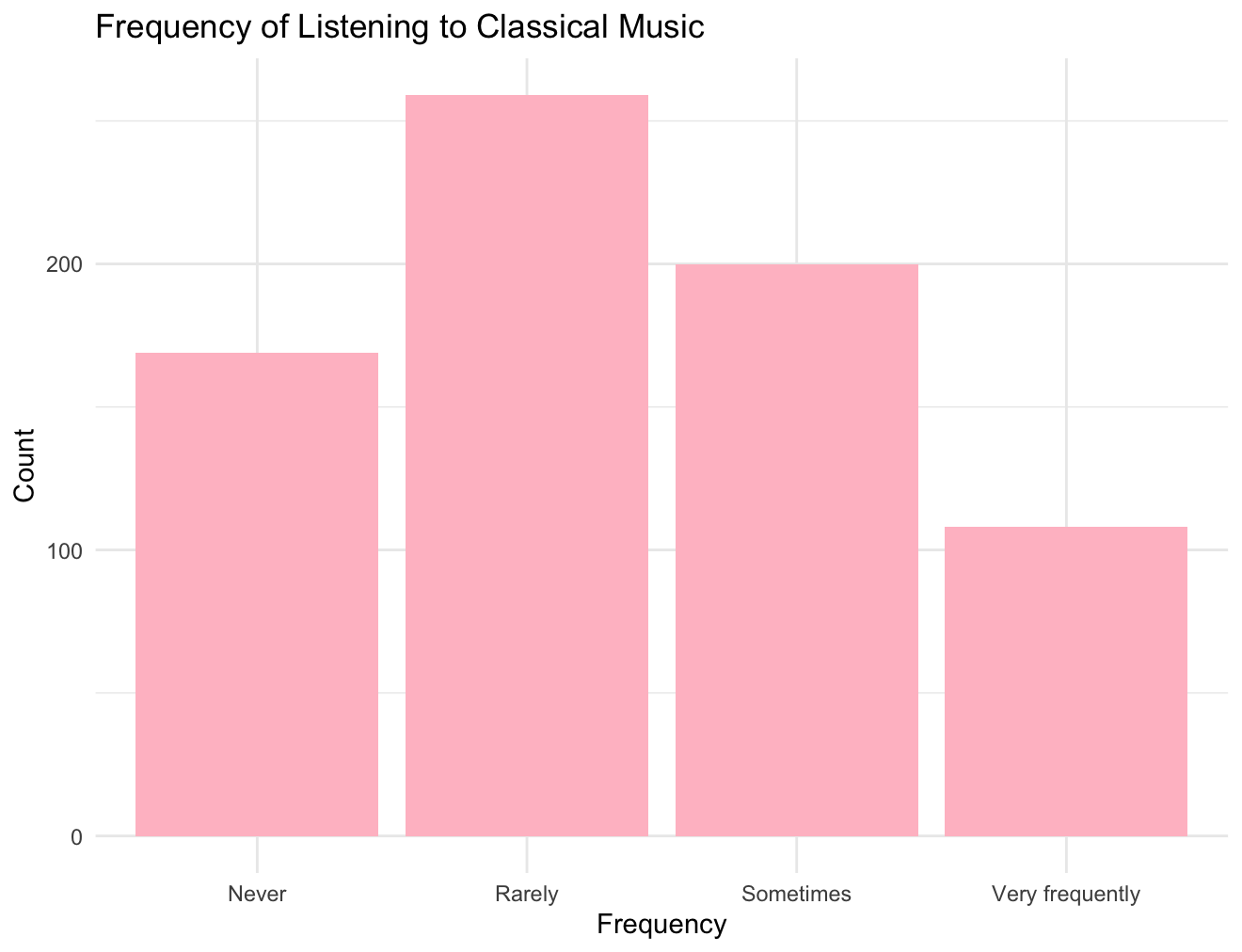
summarise(Count = n())

ggplot(classical\_frequency, aes(x = Frequency..Classical., y = Count)) +

geom\_bar(stat = "identity", fill = "pink") +

theme\_minimal() +

labs(title = "Frequency of Listening to Classical Music", x = "Frequency", y = "Count")



1. Most common genre of music preferred by individuals in the dataset

common\_genre <- music\_mental\_health %>%

group\_by(Fav.genre) %>%

summarise(Count = n()) %>%

arrange(desc(Count))

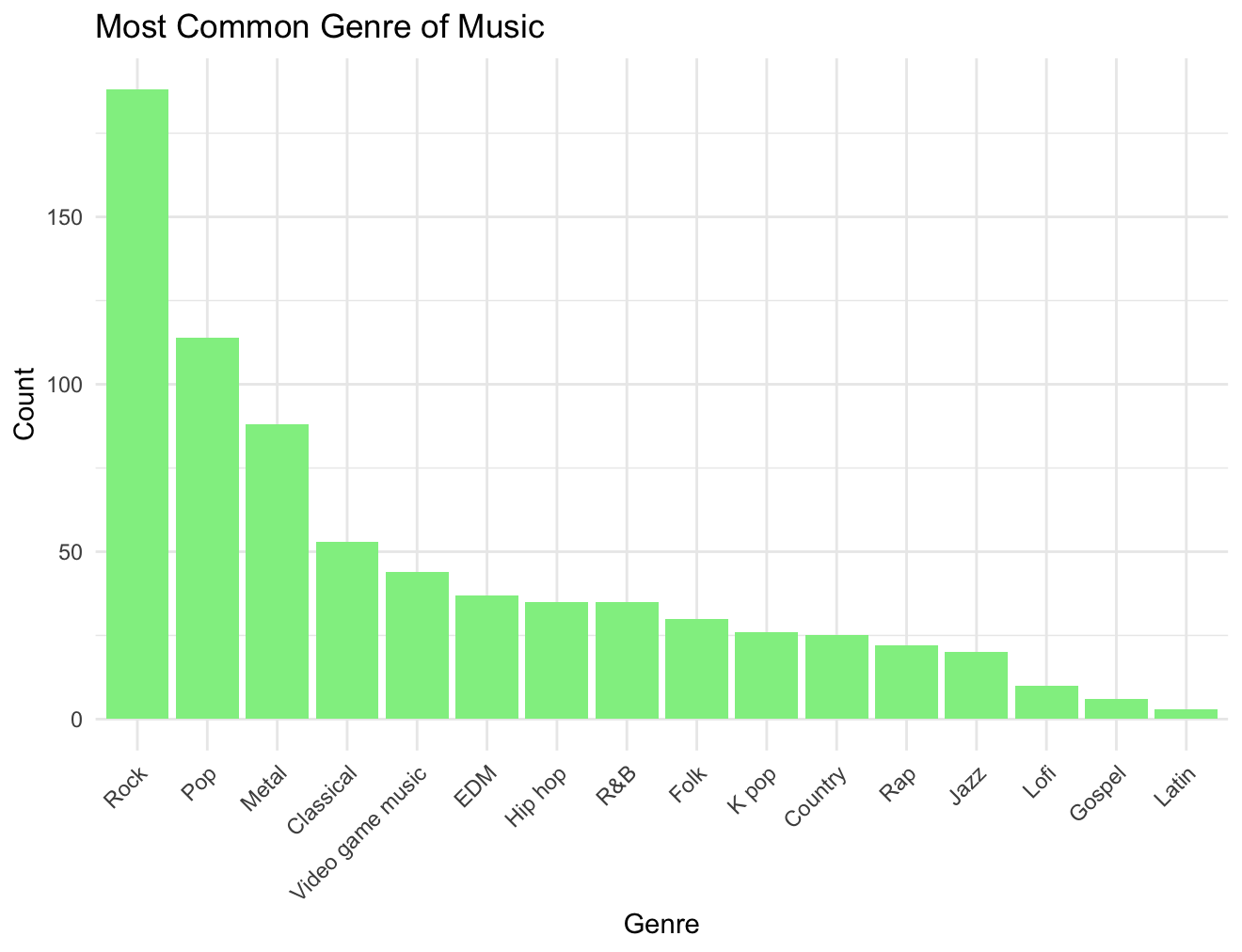
ggplot(common\_genre, aes(x = reorder(Fav.genre, -Count), y = Count)) +

geom\_bar(stat = "identity", fill = "lightgreen") +

theme\_minimal() +

labs(title = "Most Common Genre of Music", x = "Genre", y = "Count") +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))



1. Correlation between the time spent listening to music and mental health (Anxiety, Depression, etc.)

correlation\_analysis <- music\_mental\_health %>%

select(Hours.per.day, Anxiety, Depression, Insomnia, OCD) %>%

cor(use = "complete.obs")

> print(correlation\_analysis)

Hours.per.day Anxiety Depression Insomnia OCD

Hours.per.day 1.0000000 0.0493189 0.1105275 0.1418205 0.1187290

Anxiety 0.0493189 1.0000000 0.5199695 0.2926694 0.3483497

Depression 0.1105275 0.5199695 1.0000000 0.3789964 0.1969880

Insomnia 0.1418205 0.2926694 0.3789964 1.0000000 0.2263541

OCD 0.1187290 0.3483497 0.1969880 0.2263541 1.0000000

1. Impact of music on mental health differ between genres

impact\_genre <- music\_mental\_health %>%

group\_by(Fav.genre, Music.effects) %>%

summarise(Count = n())

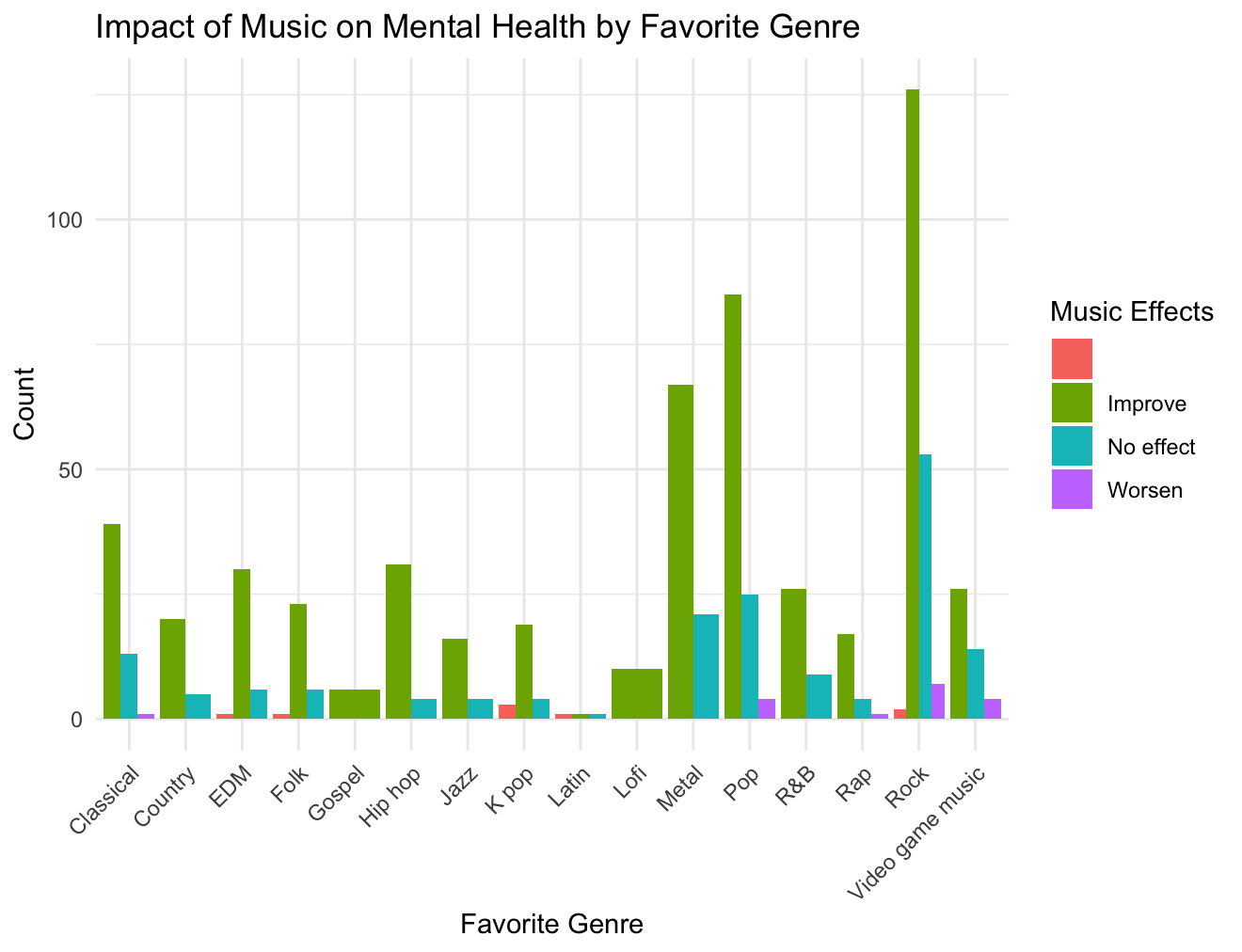
ggplot(impact\_genre, aes(x = Fav.genre, y = Count, fill = Music.effects)) +

geom\_bar(stat = "identity", position = "dodge") +

theme\_minimal() +

labs(title = "Impact of Music on Mental Health by Favorite Genre", x = "Favorite Genre", y = "Count", fill = "Music Effects") +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))



1. Average time spent listening to music per day

average\_time <- music\_mental\_health %>%

summarise(Average\_Hours = mean(Hours.per.day, na.rm = TRUE))

> print(average\_time)

Average\_Hours

1 3.572758

1. How does the preference for listening to music while working vary?

working\_music <- music\_mental\_health %>%

group\_by(While.working) %>%

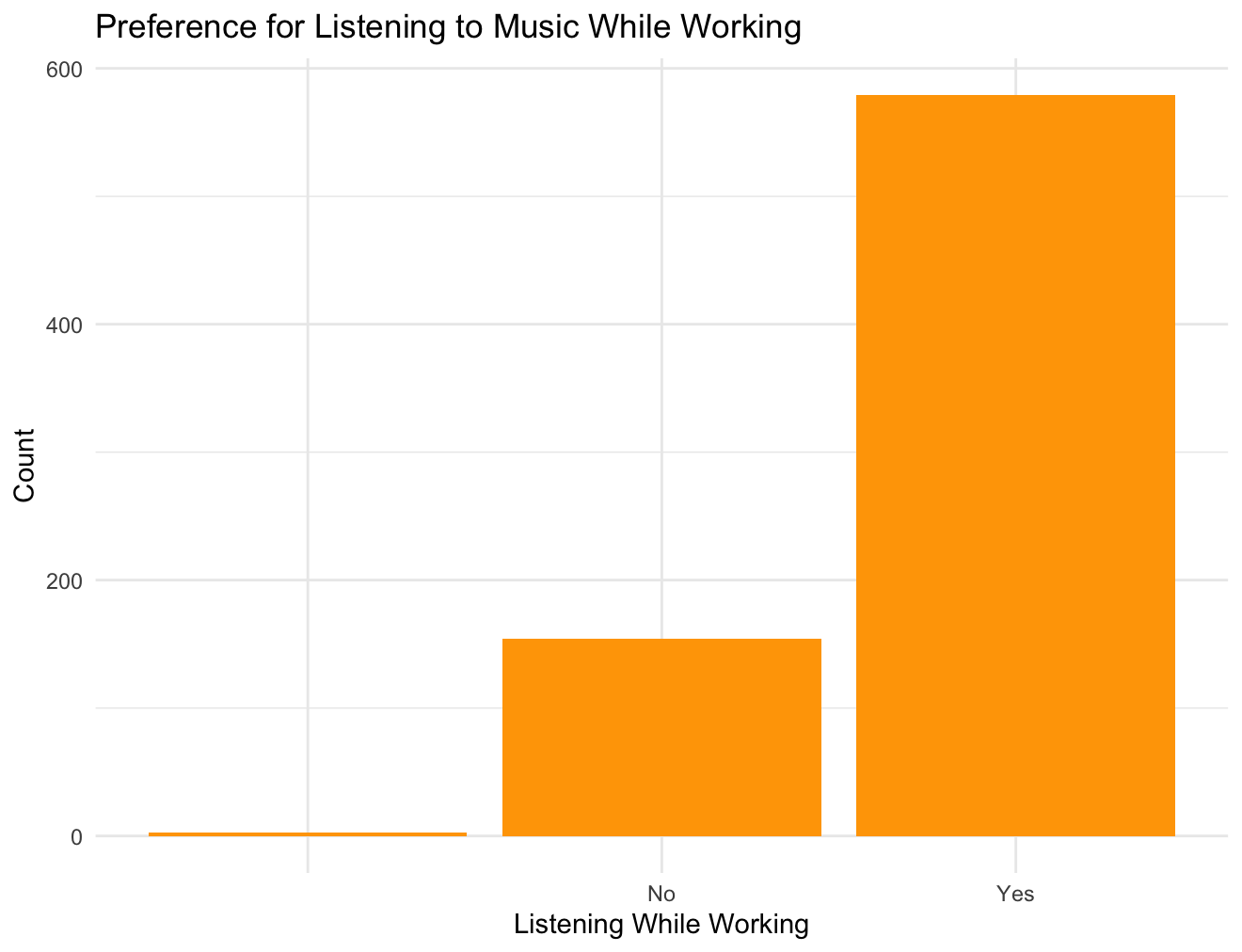
summarise(Count = n())

ggplot(working\_music, aes(x = While.working, y = Count)) +

geom\_bar(stat = "identity", fill = "orange") +

theme\_minimal() +

labs(title = "Preference for Listening to Music While Working", x = "Listening While Working", y = "Count")



1. Relationship Between Age and Music Effects on Mental Health

age\_music\_effects <- music\_mental\_health %>%

group\_by(Age, Music.effects) %>%

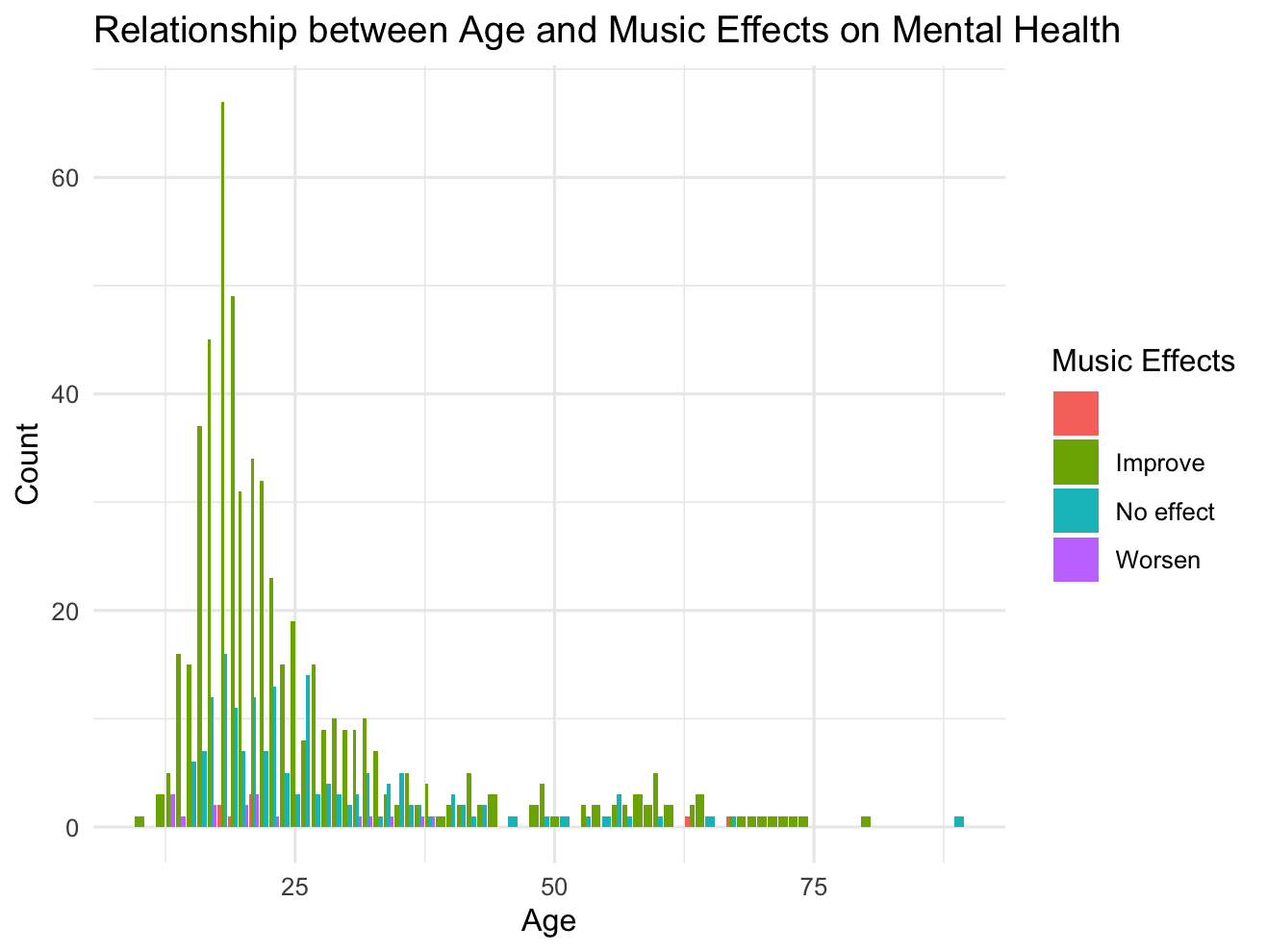
summarise(Count = n())

ggplot(age\_music\_effects, aes(x = Age, y = Count, fill = Music.effects)) +

geom\_bar(stat = "identity", position = "dodge") +

theme\_minimal() +

labs(title = "Relationship between Age and Music Effects on Mental Health", x = "Age", y = "Count", fill = "Music Effects")



10) Distribution of Favorite Music Genres

fav\_genre\_summary <- music\_mental\_health %>%

group\_by(Fav.genre) %>%

summarise(Count = n())

ggplot(fav\_genre\_summary, aes(x = "", y = Count, fill = Fav.genre)) +

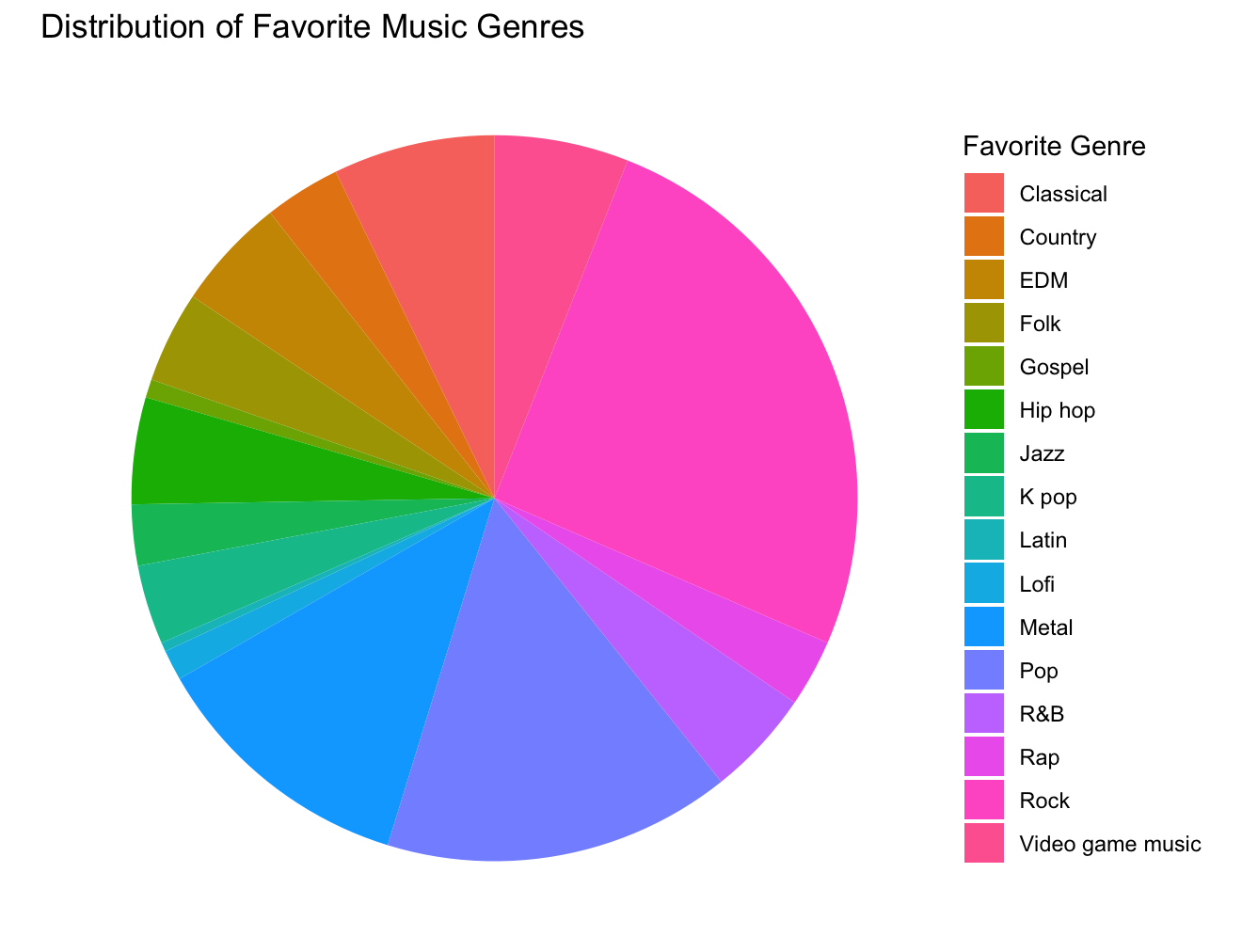
geom\_bar(width = 1, stat = "identity") +

coord\_polar("y", start = 0) +

theme\_void() +

labs(title = "Distribution of Favorite Music Genres", fill = "Favorite Genre") +

theme(legend.position = "right")



11) Hours Spent Listening to Music per Day by Favorite Genre

ggplot(music\_mental\_health, aes(x = Fav.genre, y = Hours.per.day)) + geom\_boxplot(fill = "skyblue", color = "black") + theme\_minimal() + labs(title = "Hours Spent Listening to Music per Day by Favorite Genre", x = "Favorite Genre", y = "Hours per Day") + theme(axis.text.x = element\_text(angle = 45, hjust = 1))

