The Effects of Music on Mental Health: A Data-driven Analysis

2024-04-30

load packages

```
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(tidyr)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.3
```

```
## corrplot 0.92 loaded
```

```
library(readr)
library(pheatmap)
```

```
## Warning: package 'pheatmap' was built under R version 4.3.3
```

```
library(cluster)
```

```
## Warning: package 'cluster' was built under R version 4.3.3
```

```
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.3.3
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

loading data

```
data <- read.csv("C:/Users/cgonz/Dropbox/My PC (LAPTOP-CUJK30BH)/Downloads/mxmh_survey_results.csv")

head(data)
```

```
##            Timestamp Age Primary.streaming.service Hours.per.day While.working
## 1 8/27/2022 19:29:02  18                   Spotify           3.0           Yes
## 2 8/27/2022 19:57:31  63                   Pandora           1.5           Yes
## 3 8/27/2022 21:28:18  18                   Spotify           4.0            No
## 4 8/27/2022 21:40:40  61             YouTube Music           2.5           Yes
## 5 8/27/2022 21:54:47  18                   Spotify           4.0           Yes
## 6 8/27/2022 21:56:50  18                   Spotify           5.0           Yes
##   Instrumentalist Composer        Fav.genre Exploratory Foreign.languages BPM
## 1             Yes      Yes            Latin         Yes               Yes 156
## 2              No       No             Rock         Yes                No 119
## 3              No       No Video game music          No               Yes 132
## 4              No      Yes             Jazz         Yes               Yes  84
## 5              No       No              R&B         Yes                No 107
## 6             Yes      Yes             Jazz         Yes               Yes  86
##   Frequency..Classical. Frequency..Country. Frequency..EDM. Frequency..Folk.
## 1               Rarely              Never          Rarely            Never
## 2             Sometimes              Never           Never           Rarely
## 3                 Never              Never Very frequently            Never
## 4             Sometimes              Never           Never           Rarely
## 5                 Never              Never          Rarely            Never
## 6               Rarely          Sometimes           Never            Never
##   Frequency..Gospel. Frequency..Hip.hop. Frequency..Jazz. Frequency..K.pop.
## 1              Never           Sometimes            Never   Very frequently
## 2          Sometimes              Rarely  Very frequently            Rarely
## 3              Never              Rarely           Rarely   Very frequently
## 4          Sometimes               Never  Very frequently         Sometimes
## 5             Rarely     Very frequently            Never   Very frequently
## 6              Never           Sometimes  Very frequently   Very frequently
##   Frequency..Latin. Frequency..Lofi. Frequency..Metal. Frequency..Pop.
## 1   Very frequently           Rarely             Never Very frequently
## 2         Sometimes           Rarely             Never       Sometimes
## 3             Never        Sometimes         Sometimes          Rarely
## 4   Very frequently        Sometimes             Never       Sometimes
## 5         Sometimes        Sometimes             Never       Sometimes
## 6            Rarely  Very frequently            Rarely Very frequently
##   Frequency..R.B. Frequency..Rap. Frequency..Rock. Frequency..Video.game.music.
## 1       Sometimes Very frequently            Never                     Sometimes
## 2       Sometimes          Rarely  Very frequently                        Rarely
## 3           Never          Rarely           Rarely              Very frequently
## 4       Sometimes           Never            Never                         Never
## 5 Very frequently Very frequently            Never                        Rarely
## 6 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.
```

Data summary

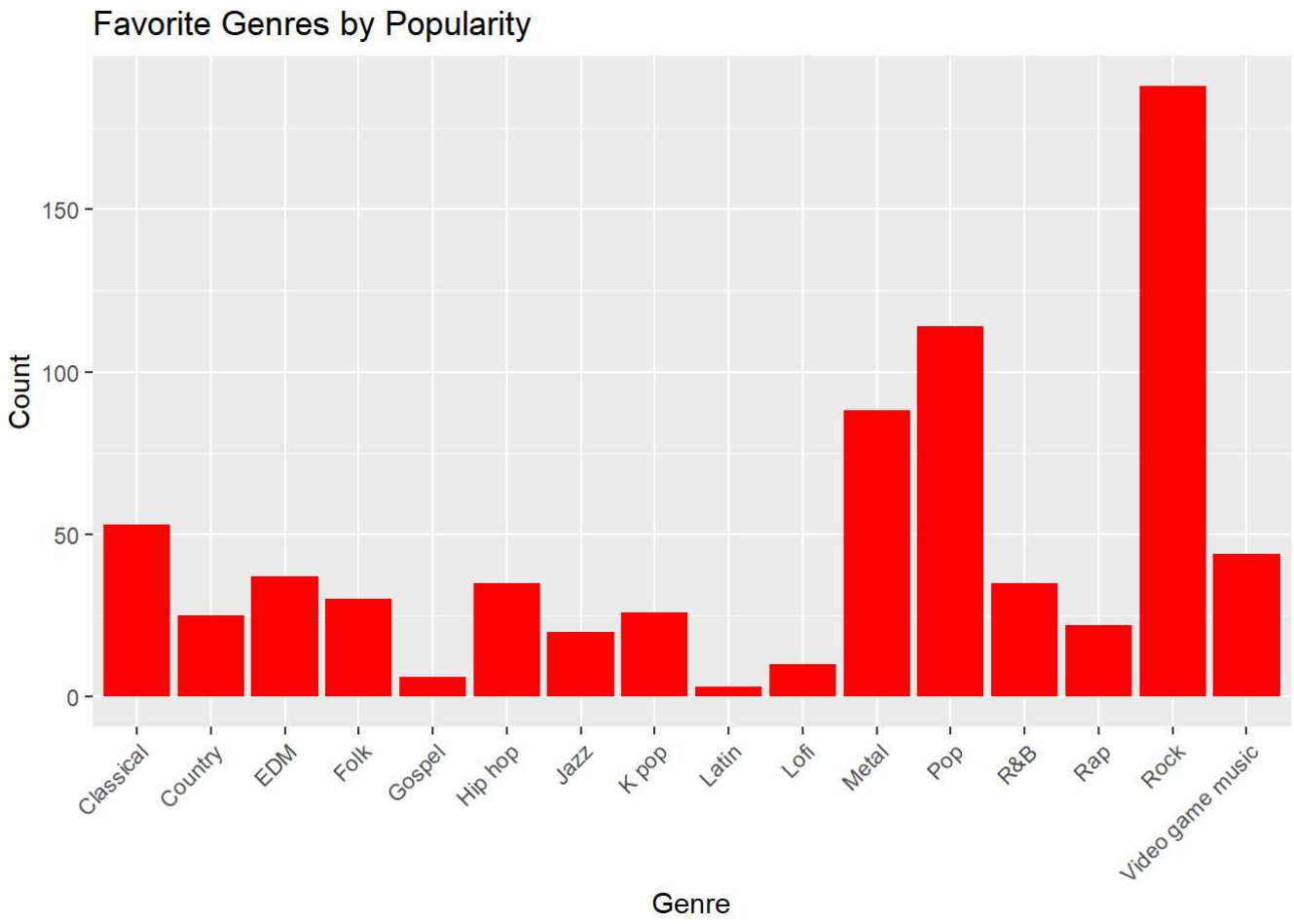
```
summary(data)
```

```
##   Timestamp              Age        Primary.streaming.service Hours.per.day
##  Length:736         Min.   :10.00   Length:736                Min.   : 0.000
##  Class :character   1st Qu.:18.00   Class :character          1st Qu.: 2.000
##  Mode  :character   Median :21.00   Mode  :character          Median : 3.000
##                     Mean   :25.21                             Mean   : 3.573
##                     3rd Qu.:28.00                             3rd Qu.: 5.000
##                     Max.   :89.00                             Max.   :24.000
##                     NA's   :1
##  While.working      Instrumentalist      Composer          Fav.genre
##  Length:736         Length:736         Length:736         Length:736
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Exploratory        Foreign.languages       BPM           Frequency..Classical.
##  Length:736         Length:736         Min.   :0.00e+00   Length:736
##  Class :character   Class :character   1st Qu.:1.00e+02   Class :character
##  Mode  :character   Mode  :character   Median :1.20e+02   Mode  :character
##                                        Mean   :1.59e+06
##                                        3rd Qu.:1.44e+02
##                                        Max.   :1.00e+09
##                                        NA's   :107
##  Frequency..Country. Frequency..EDM.    Frequency..Folk.   Frequency..Gospel.
##  Length:736          Length:736         Length:736         Length:736
##  Class :character    Class :character   Class :character   Class :character
##  Mode  :character    Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Frequency..Hip.hop. Frequency..Jazz.   Frequency..K.pop.  Frequency..Latin.
##  Length:736          Length:736         Length:736         Length:736
##  Class :character    Class :character   Class :character   Class :character
##  Mode  :character    Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Frequency..Lofi.   Frequency..Metal.  Frequency..Pop.    Frequency..R.B.
##  Length:736         Length:736         Length:736         Length:736
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  Frequency..Rap.    Frequency..Rock.   Frequency..Video.game.music.
##  Length:736         Length:736         Length:736
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
```

```
##
##
##
##
##     Anxiety         Depression        Insomnia           OCD
##  Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
##  1st Qu.: 4.000   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 0.000
##  Median : 6.000   Median : 5.000   Median : 3.000   Median : 2.000
##  Mean   : 5.838   Mean   : 4.796   Mean   : 3.738   Mean   : 2.637
##  3rd Qu.: 8.000   3rd Qu.: 7.000   3rd Qu.: 6.000   3rd Qu.: 5.000
##  Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000
##
##  Music.effects      Permissions
##  Length:736         Length:736
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

1. What correlations exist between different music genres and their self-reported mental health conditions?

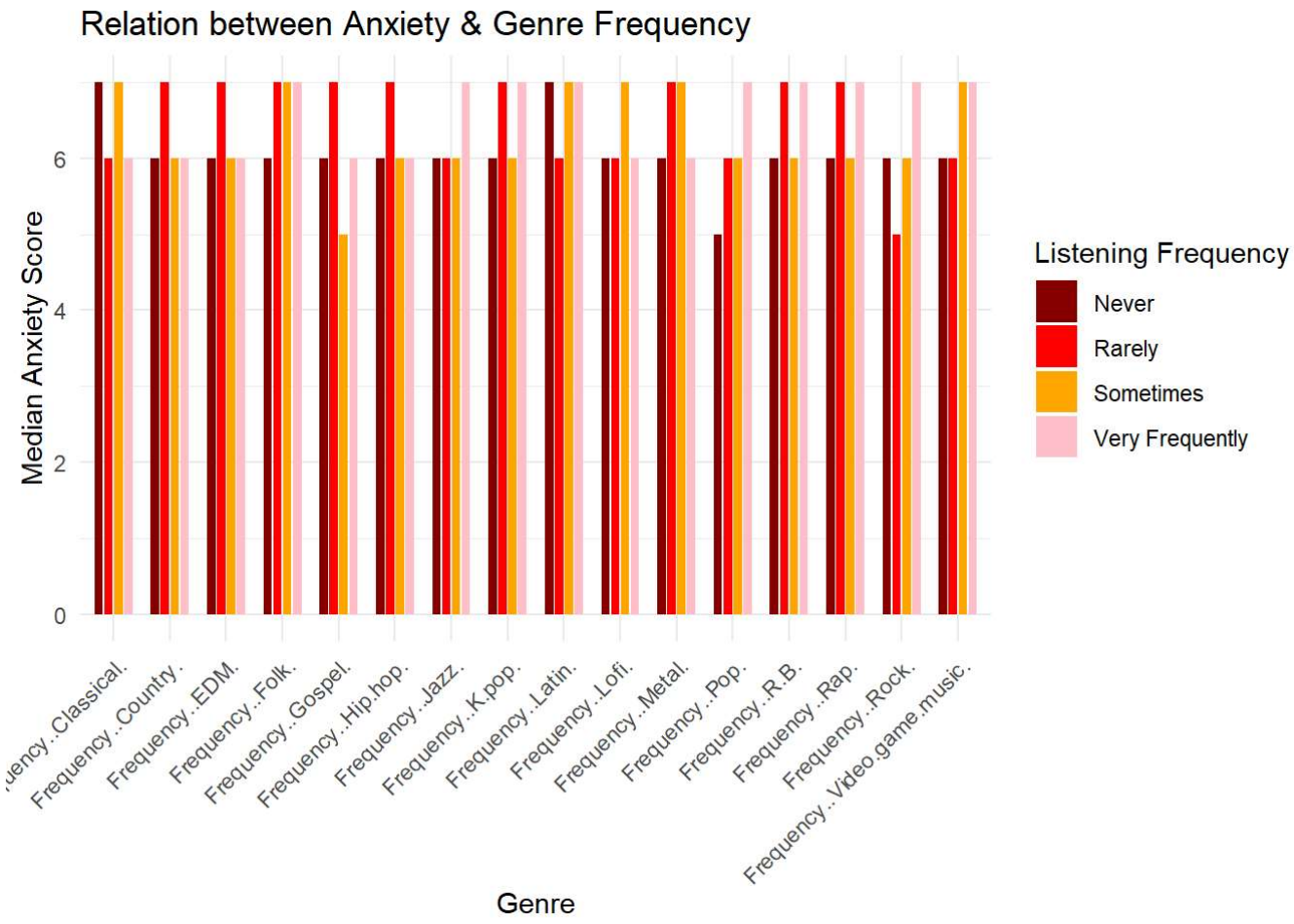
```
favgenre_data <- data %>%
  count(Fav.genre) %>%
  arrange(desc(n))  # arranges frequency in descending order

ggplot(favgenre_data, aes(x=Fav.genre, y=n)) +
  geom_bar(stat="identity", fill="red") +
  labs(title="Favorite Genres by Popularity", x="Genre", y="Count") +
  theme(axis.text.x = element_text(angle=45, hjust=1))
```



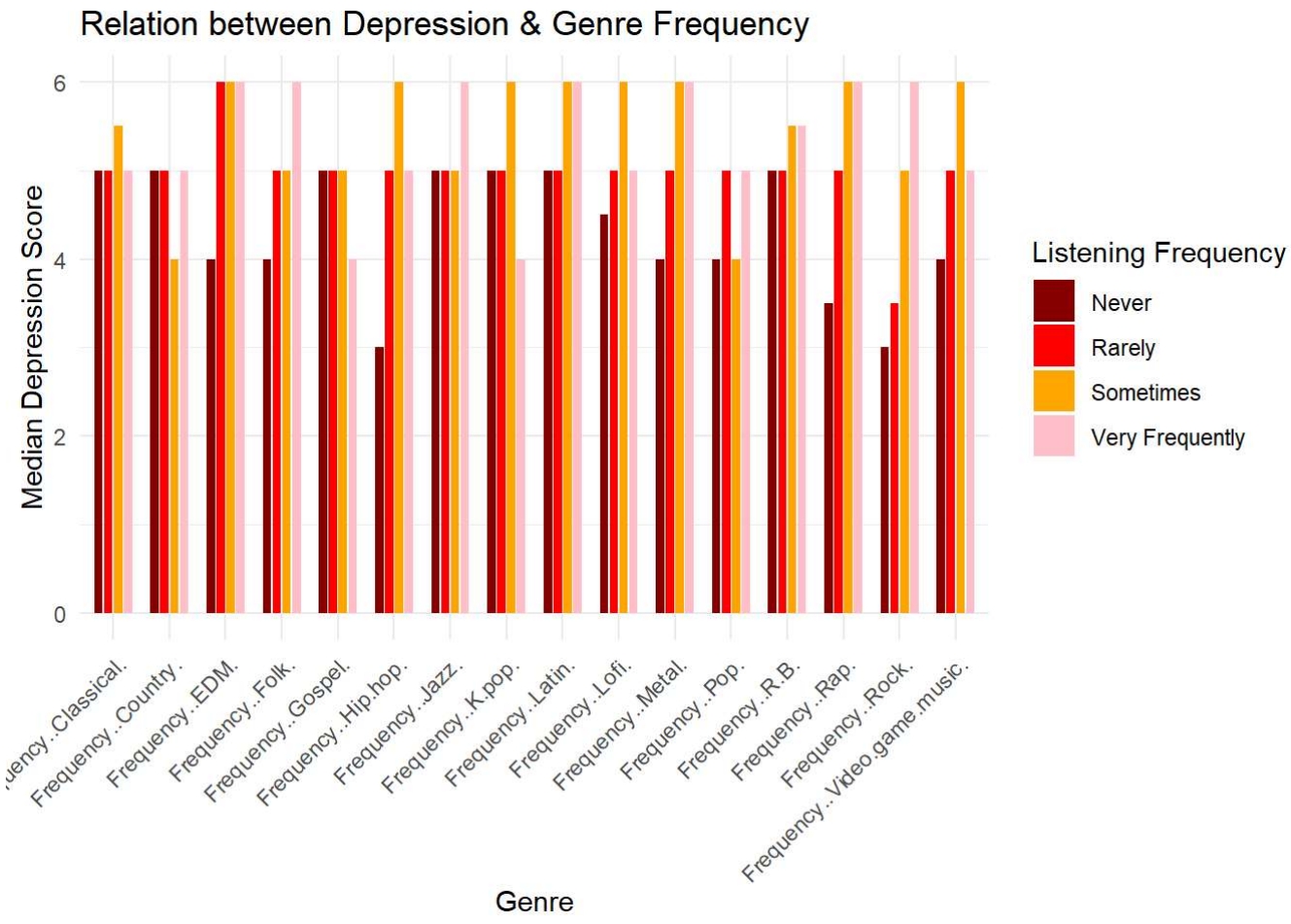

```
this_data <- data %>%
  mutate(across(starts_with("Frequency"), ~case_when(
    . == "Never" ~ 0,
    . == "Rarely" ~ 1,
    . == "Sometimes" ~ 2,
    . == "Very frequently" ~ 3,
    TRUE ~ NA_real_
  )))

genre_columns <- names(data)[grepl("^Frequency", names(data))]

data_long <- data %>%
  pivot_longer(cols = genre_columns, names_to = "Genre", values_to = "Frequency") %>%
  group_by(Genre, Frequency) %>%
  summarise(
    Anxiety_Median = median(Anxiety, na.rm = TRUE),
    Depression_Median = median(Depression, na.rm = TRUE),
    Insomnia_Median = median(Insomnia, na.rm = TRUE),
    OCD_Median = median(OCD, na.rm = TRUE),
    .groups = 'drop'
  )
```

```
## Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.
## ℹ Please use `all_of()` or `any_of()` instead.
##   # Was:
##   data %>% select(genre_columns)
##
##   # Now:
##   data %>% select(all_of(genre_columns))
##
## See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

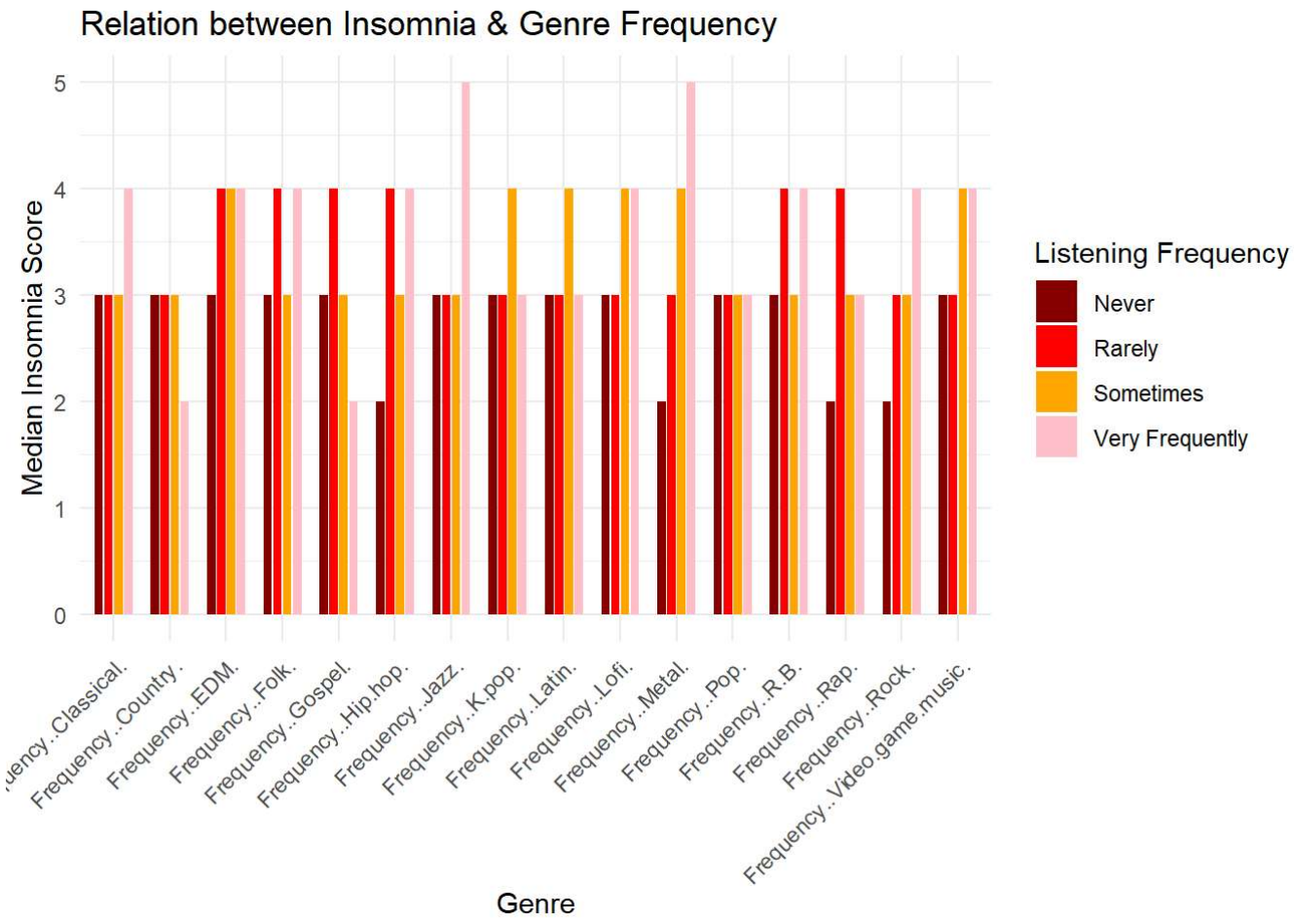
```
# function to create bar plots for each condition
plot_condition <- function(data, condition, title, ylab) {
  ggplot(data, aes(x = Genre, y = get(condition), fill = factor(Frequency))) +
    geom_bar(stat = "identity", position = position_dodge(width = 0.7), width = 0.6) +
    scale_fill_manual(values = c("darkred", "red", "orange", "pink"),
                      labels = c("Never", "Rarely", "Sometimes", "Very Frequently")) +
    labs(title = title, x = "Genre", y = ylab, fill = "Listening Frequency") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
}

# Plotting each condition
plot_condition(data_long, "Anxiety_Median", "Relation between Anxiety & Genre Frequency", "Media
n Anxiety Score")
```



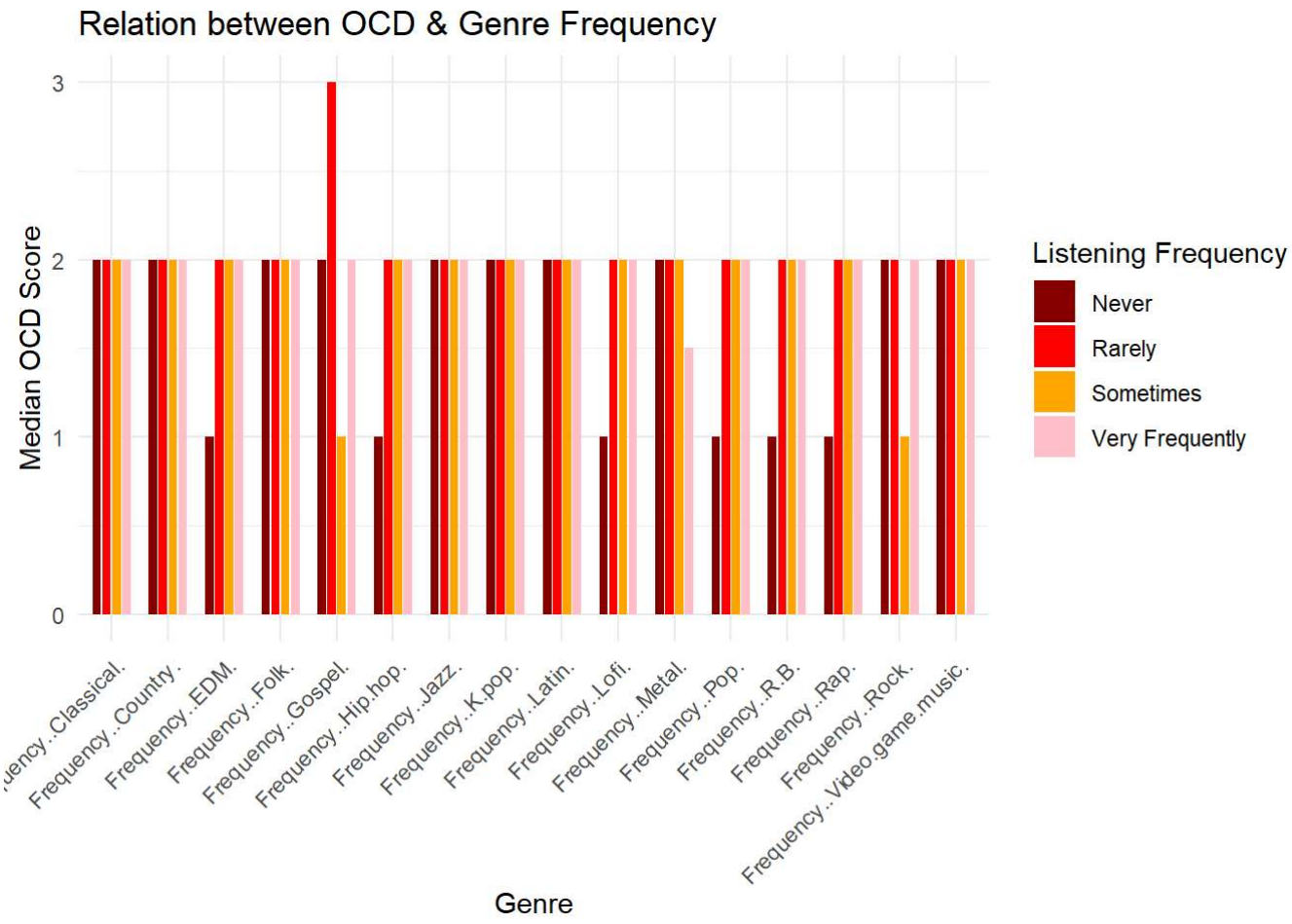

```
plot_condition(data_long, "Depression_Median", "Relation between Depression & Genre Frequency",
"Median Depression Score")
```




```
plot_condition(data_long, "Insomnia_Median", "Relation between Insomnia & Genre Frequency", "Med
ian Insomnia Score")
```

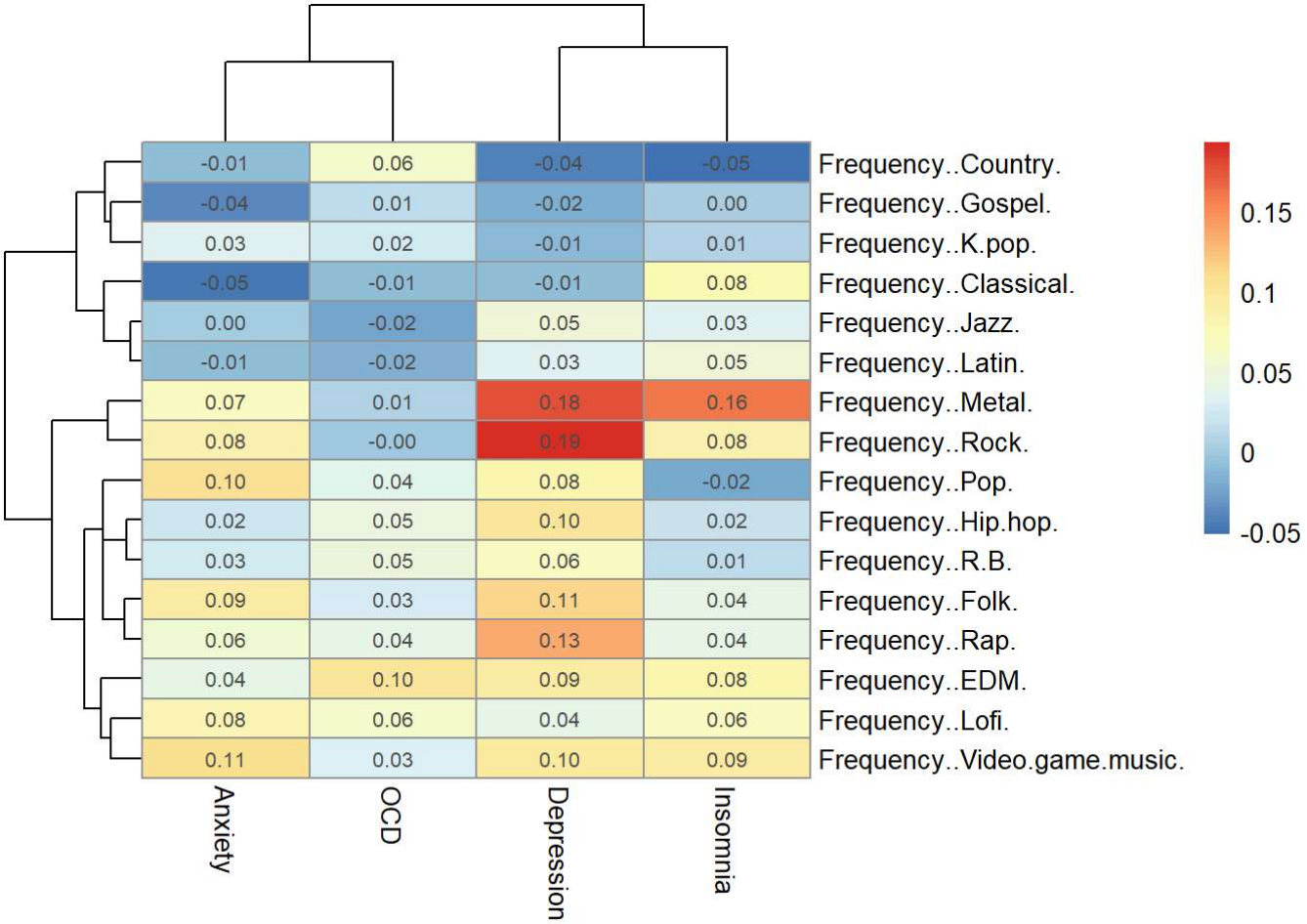



```
plot_condition(data_long, "OCD_Median", "Relation between OCD & Genre Frequency", "Median OCD Sc
ore")
```



Correlation matrix

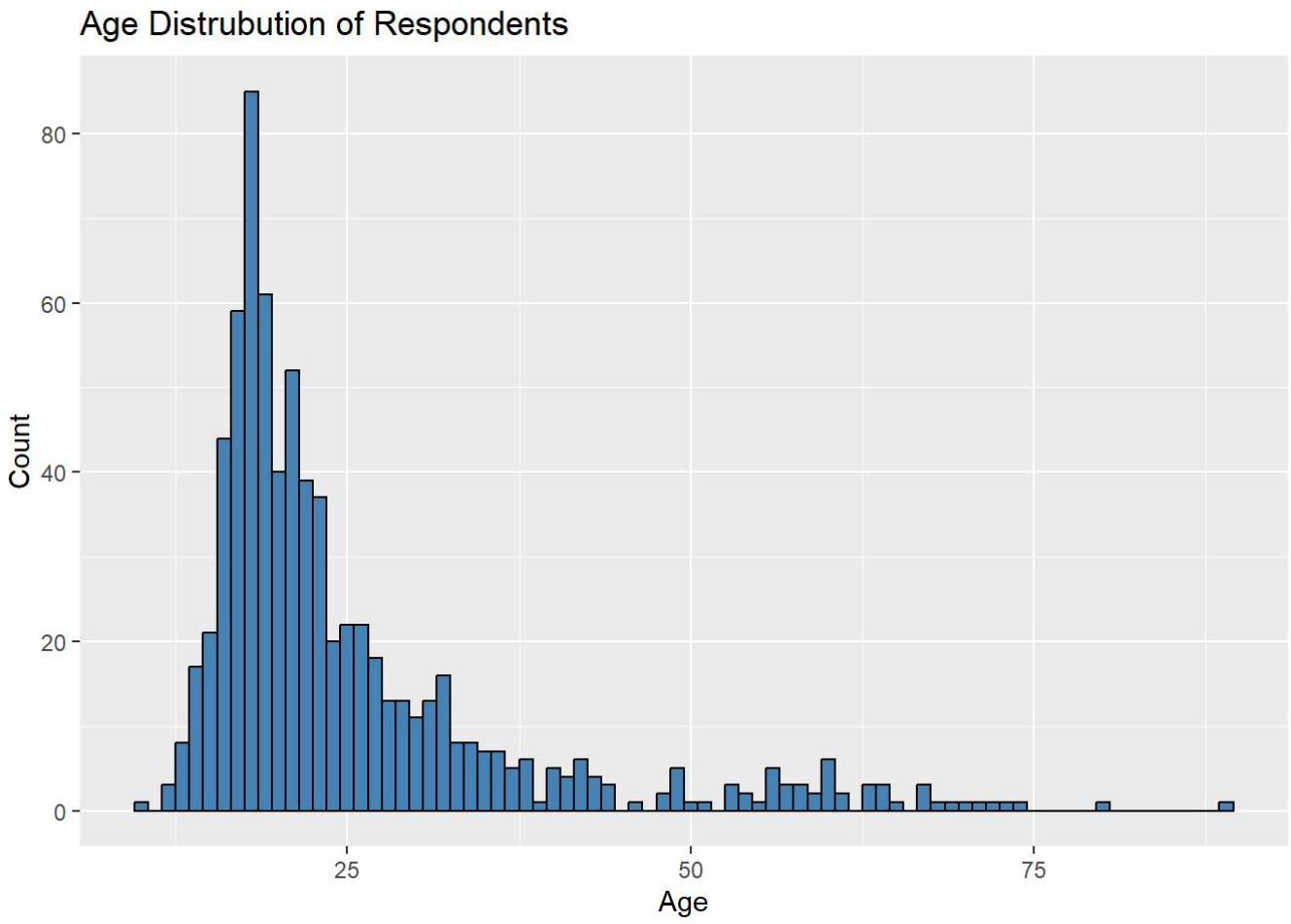
```
mental_health_vars <- c("Anxiety", "Depression", "Insomnia", "OCD")
genre_cols <- grep("Frequency", names(this_data), value = TRUE)


cor_matrix <- cor(this_data[genre_cols], this_data[mental_health_vars], use="complete.obs")

pheatmap(cor_matrix, display_numbers = TRUE, title = "Correlation Heatmap")
```



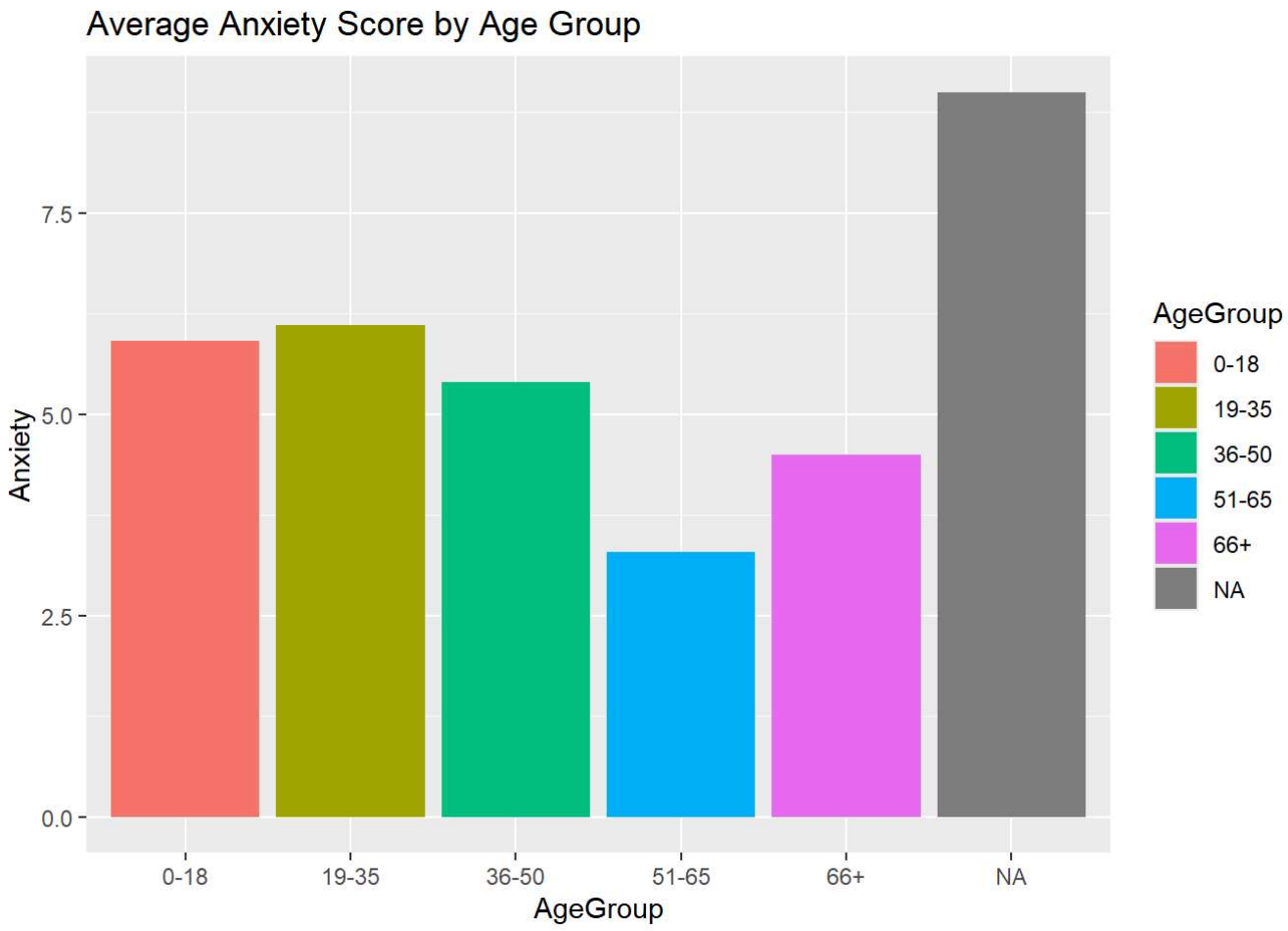
“Which age group experiences the most mental health issues?”

```
ggplot(data, aes(x=Age)) +
  geom_histogram(binwidth = 1, fill="steelblue", color="black") +
  ggtitle("Age Distrubution of Respondents") +
  xlab("Age") +
  ylab("Count")
```

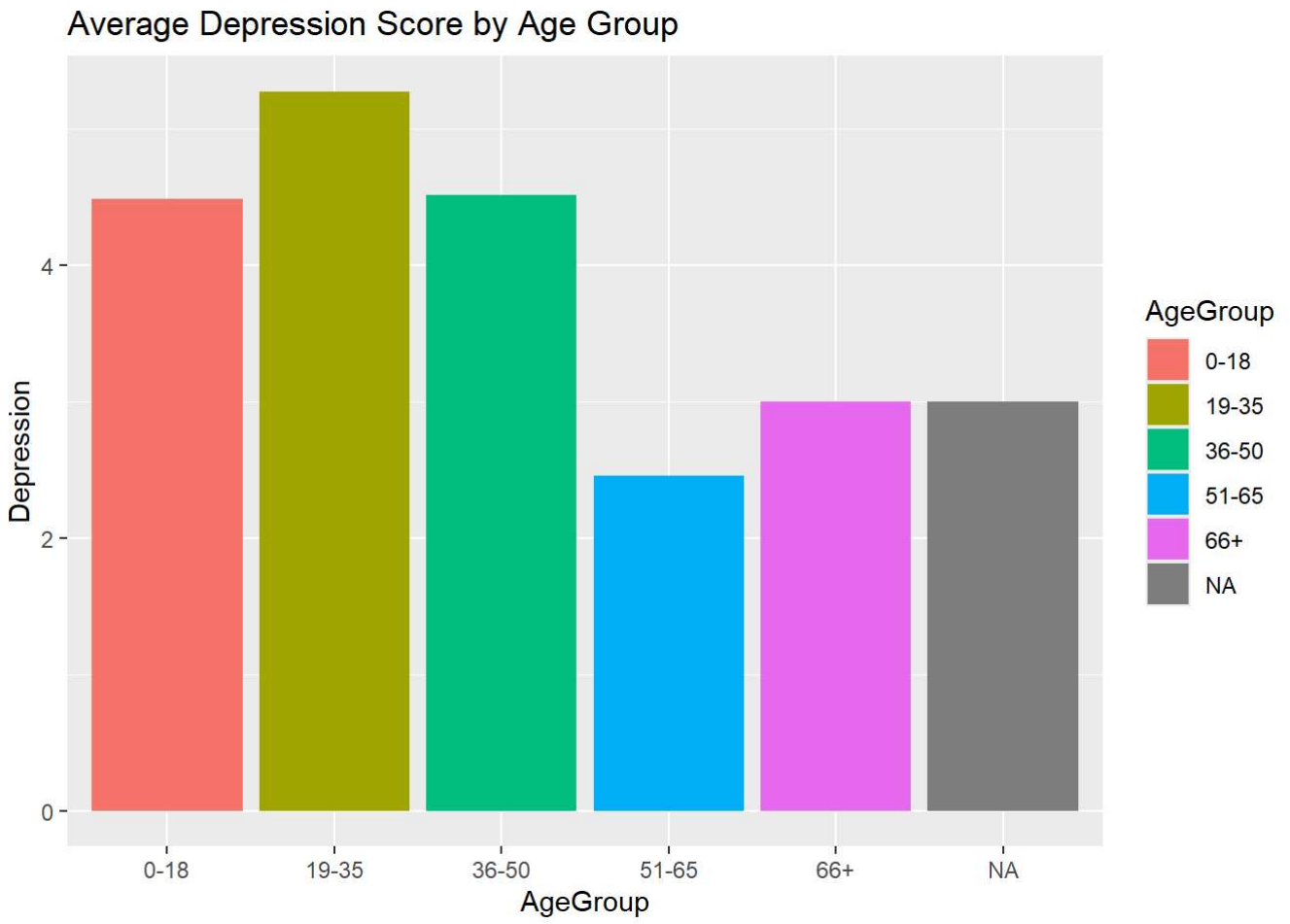
```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_bin()`).
```



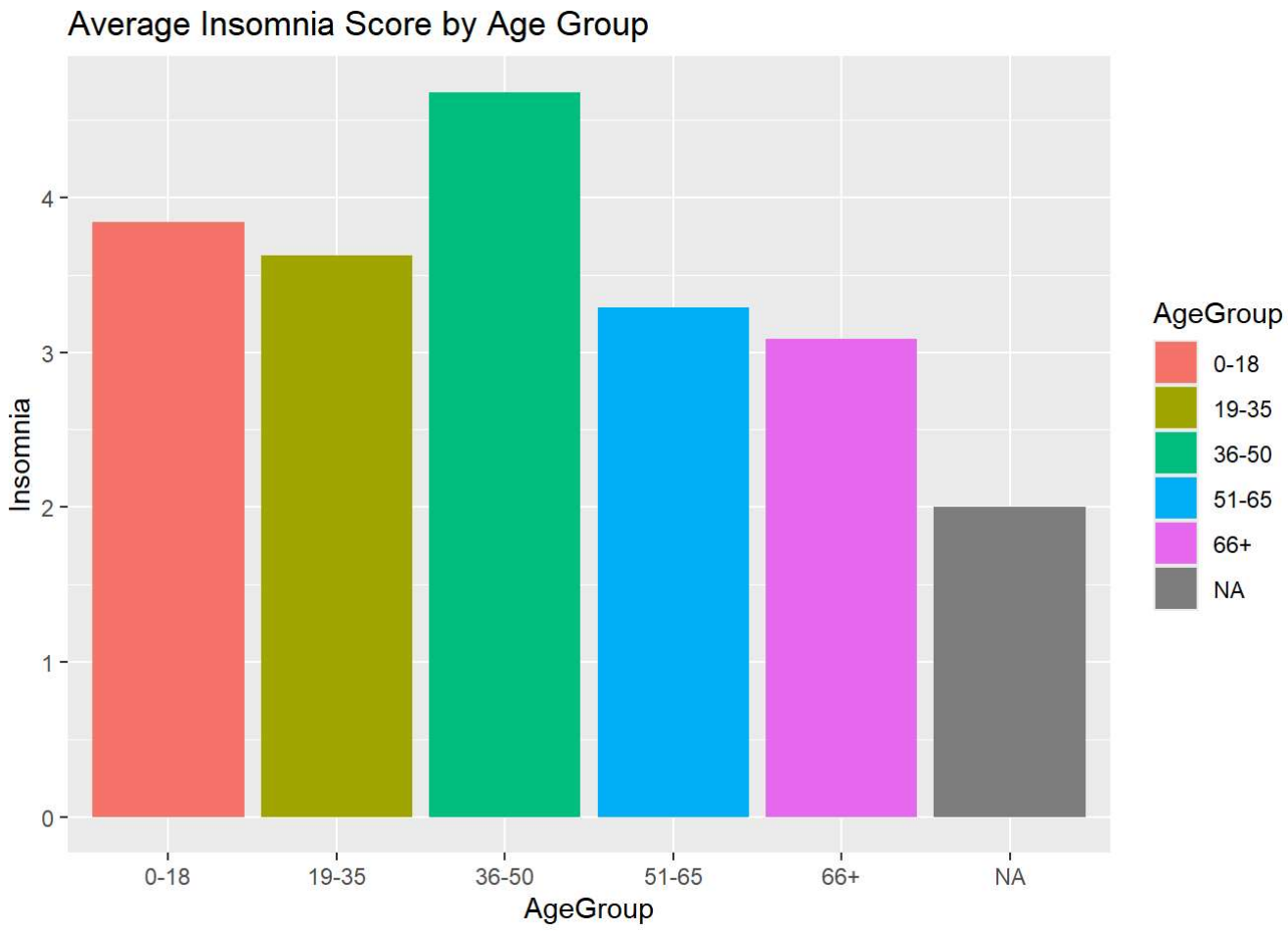

```
this_data$AgeGroup <- cut(this_data$Age, breaks=c(0, 18, 35, 50, 65, Inf), labels=c("0-18", "19-35", "36-50", "51-65", "66+"))

ggplot(this_data, aes(x=AgeGroup, y=Anxiety, fill=AgeGroup)) +
  geom_bar(stat="summary", fun="mean") +
  labs(title="Average Anxiety Score by Age Group")
```



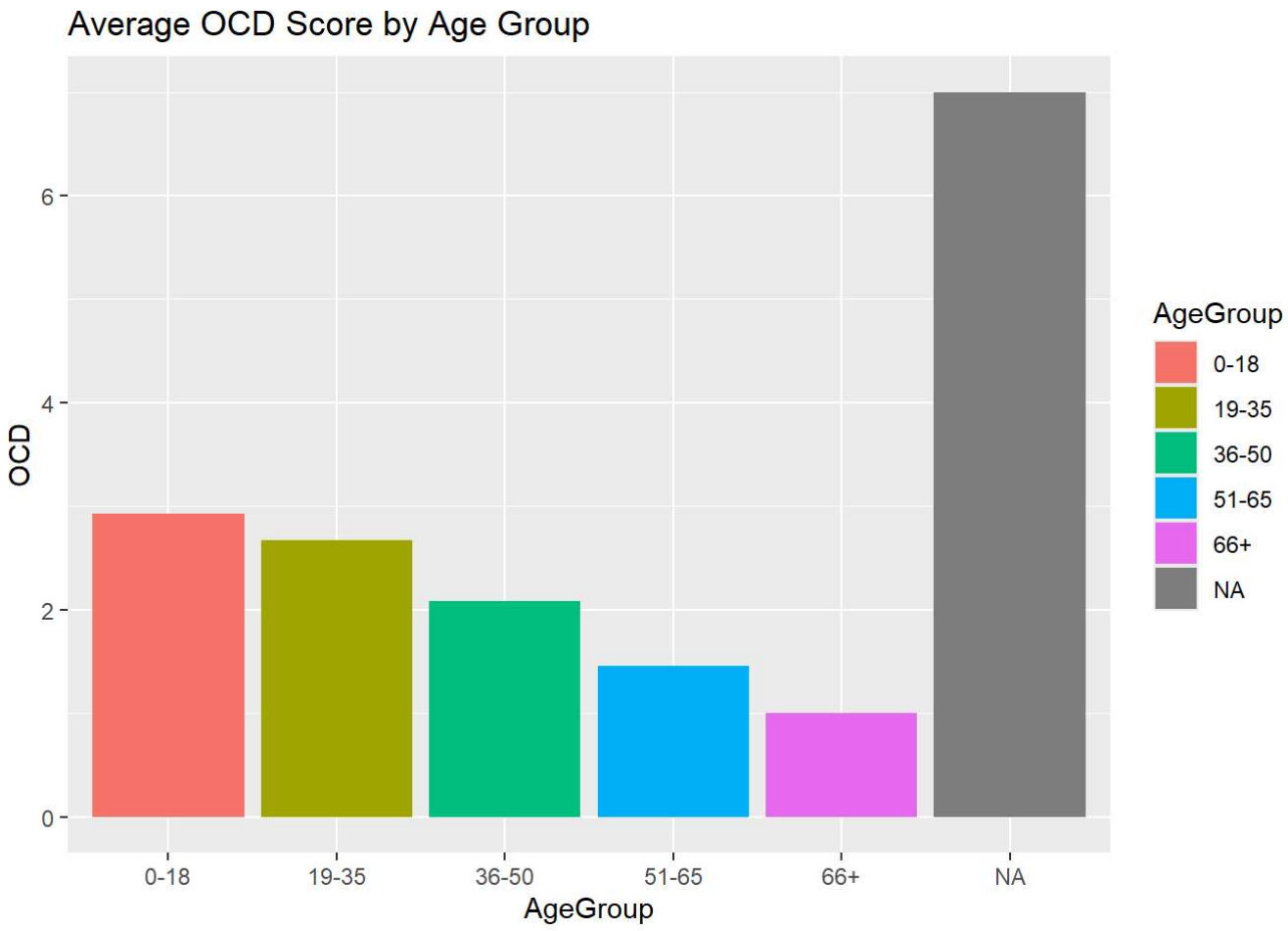

```
ggplot(this_data, aes(x=AgeGroup, y=Depression, fill=AgeGroup)) +
  geom_bar(stat="summary", fun="mean") +
  labs(title="Average Depression Score by Age Group")
```




```
ggplot(this_data, aes(x=AgeGroup, y=Insomnia, fill=AgeGroup)) +
  geom_bar(stat="summary", fun="mean") +
  labs(title="Average Insomnia Score by Age Group")
```




```
ggplot(this_data, aes(x=AgeGroup, y=OCD, fill=AgeGroup)) +
  geom_bar(stat="summary", fun="mean") +
  labs(title="Average OCD Score by Age Group")
```



ANOVA

```
# Performing ANOVA for each mental health variable
anova_anxiety <- aov(Anxiety ~ AgeGroup, data=this_data)
summary(anova_anxiety)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## AgeGroup      4    288   72.10   9.684 1.24e-07 ***
## Residuals   730   5435    7.45
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```

```
cat("\nDEPRESSION\n")
```

```
##
## DEPRESSION
```

```
anova_depression <- aov(Depression ~ AgeGroup, data=this_data)
summary(anova_depression)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## AgeGroup      4    349   87.34   9.978 7.27e-08 ***
## Residuals   730   6390    8.75
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```

```
cat("\nINSOMNIA\n")
```

```
##
## INSOMNIA
```

```
anova_insomnia <- aov(Insomnia ~ AgeGroup, data=this_data)
summary(anova_insomnia)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## AgeGroup      4     65  16.149   1.698  0.149
## Residuals   730   6944   9.513
## 1 observation deleted due to missingness
```

```
cat("\nOCD\n")
```

```
##
## OCD
```

```
anova_OCD <- aov(OCD ~ AgeGroup, data=this_data)
summary(anova_OCD)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## AgeGroup      4    117  29.295   3.687 0.00556 **
## Residuals   730   5800   7.946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```

Correlation tests

We need to see if Hours per Day is normally distributed:

```
shapiro.test(this_data$`Hours.per.day`)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  this_data$Hours.per.day
## W = 0.77592, p-value < 2.2e-16
```

This implies our data is normally distributed (low p value). Since this subset of our data is normally distributed, we're going to be performing a Pearson correlation test.

```
cat("ANXIETY\n")
```

```
## ANXIETY
```

```
cor.test(this_data$Hours.per.day, this_data$Anxiety, method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  this_data$Hours.per.day and this_data$Anxiety
## t = 1.3378, df = 734, p-value = 0.1814
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.0230299  0.1211538
## sample estimates:
##       cor
## 0.0493189
```

```
cat("\nDEPRESSION\n")
```

```
##
## DEPRESSION
```

```
cor.test(this_data$Hours.per.day, this_data$Depression, method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  this_data$Hours.per.day and this_data$Depression
## t = 3.0129, df = 734, p-value = 0.002676
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.03856881 0.18134569
## sample estimates:
##       cor
## 0.1105275
```

```
cat("\nINSOMNIA\n")
```

```
##
## INSOMNIA
```

```
cor.test(this_data$Hours.per.day, this_data$Insomnia, method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  this_data$Hours.per.day and this_data$Insomnia
## t = 3.8815, df = 734, p-value = 0.0001132
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.07027401 0.21191533
## sample estimates:
##       cor
## 0.1418205
```

```
cat("\nOCD\n")
```

```
##
## OCD
```

```
cor.test(this_data$Hours.per.day, this_data$OCD, method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  this_data$Hours.per.day and this_data$OCD
## t = 3.2396, df = 734, p-value = 0.001251
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.04686438 0.18937089
## sample estimates:
##      cor
## 0.118729
```

Is there a correlation between self-reported mental health and identifying with being a musician

(i.e. instrumentalist or composer)?

```
data$Is_Musician <- ifelse(data$Instrumentalist == "Yes" | data$Composer == "Yes", "Yes", "No")

summary_stats <- data %>%
  group_by(Is_Musician) %>%
  summarise(across(c(Anxiety, Depression, Insomnia, OCD),
                   list(mean = ~mean(.x, na.rm = TRUE),
                        sd = ~sd(.x, na.rm = TRUE))))
```

T-test for: “Is there a correlation btwn being a musician and mental health issues?”

```
# T-test for Anxiety
t_test_anxiety <- t.test(Anxiety ~ Is_Musician, data = data)
print(t_test_anxiety)
```

```
##
##  Welch Two Sample t-test
##
## data:  Anxiety by Is_Musician
## t = -0.32132, df = 567.2, p-value = 0.7481
## alternative hypothesis: true difference in means between group No and group Yes is not equal
to 0
## 95 percent confidence interval:
##  -0.4865675  0.3497534
## sample estimates:
##  mean in group No mean in group Yes
##          5.812634          5.881041
```

```
# T-test for Depression
t_test_depression <- t.test(Depression ~ Is_Musician, data = data)
print(t_test_depression)
```

```
##
##  Welch Two Sample t-test
##
## data:  Depression by Is_Musician
## t = -0.29038, df = 585.29, p-value = 0.7716
## alternative hypothesis: true difference in means between group No and group Yes is not equal
to 0
## 95 percent confidence interval:
##  -0.5150575  0.3823748
## sample estimates:
##  mean in group No mean in group Yes
##          4.771949          4.838290
```

```r
# T-test for Insomnia
t_test_insomnia <- t.test(Insomnia ~ Is_Musician, data = data)
print(t_test_insomnia)
```

```
##
##  Welch Two Sample t-test
##
## data:  Insomnia by Is_Musician
## t = -1.2315, df = 552.85, p-value = 0.2187
## alternative hypothesis: true difference in means between group No and group Yes is not equal
to 0
## 95 percent confidence interval:
##  -0.7580043  0.1738040
## sample estimates:
##  mean in group No mean in group Yes
##          3.631692          3.923792
```

```r
# T-test for OCD
t_test_ocd <- t.test(OCD ~ Is_Musician, data = data)
print(t_test_ocd)
```

```
##
##  Welch Two Sample t-test
##
## data:  OCD by Is_Musician
## t = -0.11105, df = 575.93, p-value = 0.9116
## alternative hypothesis: true difference in means between group No and group Yes is not equal
to 0
## 95 percent confidence interval:
##  -0.4472870  0.3994136
## sample estimates:
##  mean in group No mean in group Yes
##          2.628480          2.652416
```

MODELS

Supervised: Multiple Regressions & Diagnostic Plots

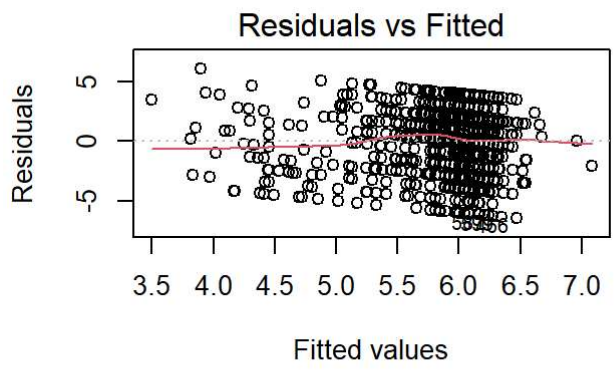
```r
cat("ANXIETY\n")
```

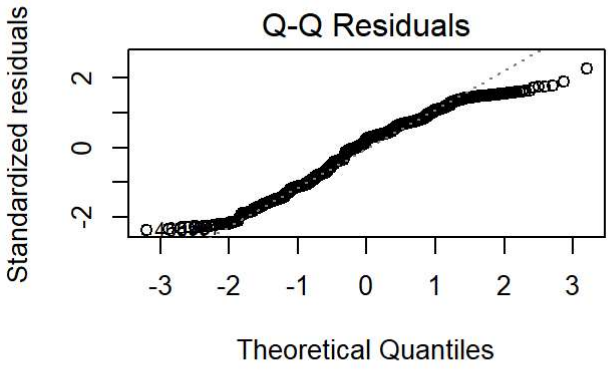
```
## ANXIETY
```

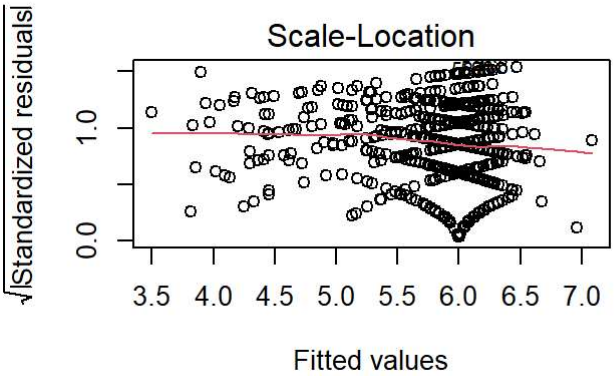
```r
# predicting anxiety from Age, Hours per Day, and Classical Music Frequency
lm_model <- lm(Anxiety ~ Age + Hours.per.day + Frequency..Classical., data=this_data)
summary(lm_model)
```

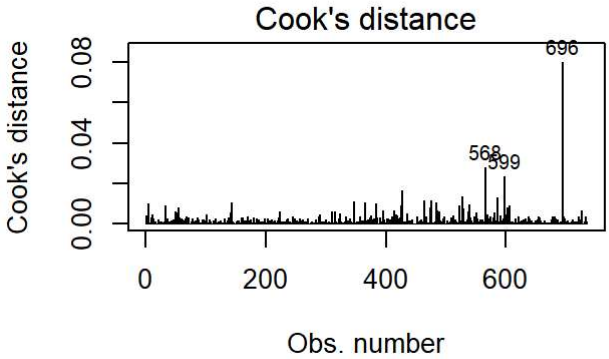
```
##
## Call:
## lm(formula = Anxiety ~ Age + Hours.per.day + Frequency..Classical.,
##     data = this_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4709 -2.1170  0.5882  1.9984  6.1026
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           6.809090   0.293937  23.165  < 2e-16 ***
## Age                  -0.039839   0.008472  -4.702 3.07e-06 ***
## Hours.per.day         0.037944   0.033548   1.131    0.258
## Frequency..Classical. -0.080186   0.103200  -0.777    0.437
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.751 on 731 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.03376,    Adjusted R-squared:  0.02979
## F-statistic: 8.512 on 3 and 731 DF,  p-value: 1.458e-05
```

```
par(mfrow = c(2, 2))  # Set up the plotting area for multiple plots
plot(lm_model, which = 1:4)  # Plots for residuals
```









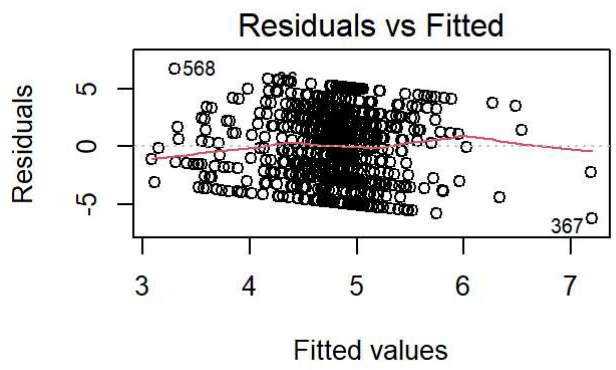

```
cat("\nDEPRESSION\n")
```

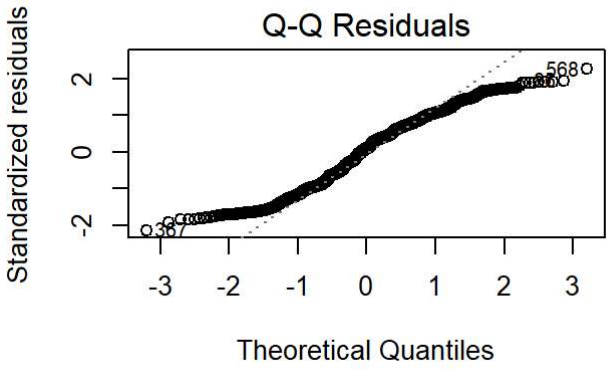
```
##
## DEPRESSION
```

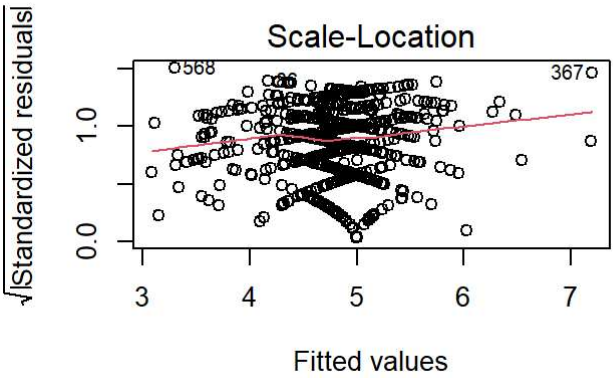
```
# predicting depression from Age, Hours per Day, and Classical Music Frequency
lm_model <- lm(Depression ~ Age + Hours.per.day + Frequency..Classical., data=this_data)
summary(lm_model)
```

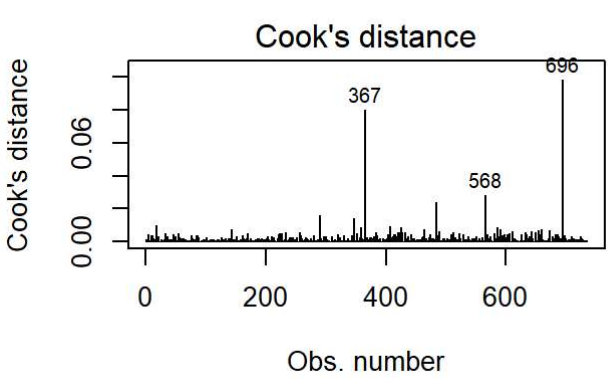
```
##
## Call:
## lm(formula = Depression ~ Age + Hours.per.day + Frequency..Classical.,
##     data = this_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.1967 -2.6769  0.2934  2.3771  6.7045
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           5.141173   0.320257  16.053  < 2e-16 ***
## Age                  -0.029531   0.009231  -3.199  0.00144 **
## Hours.per.day         0.105336   0.036552   2.882  0.00407 **
## Frequency..Classical.  0.019035   0.112441   0.169  0.86562
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.997 on 731 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.02586,    Adjusted R-squared:  0.02186
## F-statistic: 6.468 on 3 and 731 DF,  p-value: 0.0002529
```

```
par(mfrow = c(2, 2))  # Set up the plotting area for multiple plots
plot(lm_model, which = 1:4)  # Plots for residuals
```


Residuals vs Fitted
Residuals
Fitted values
568
367

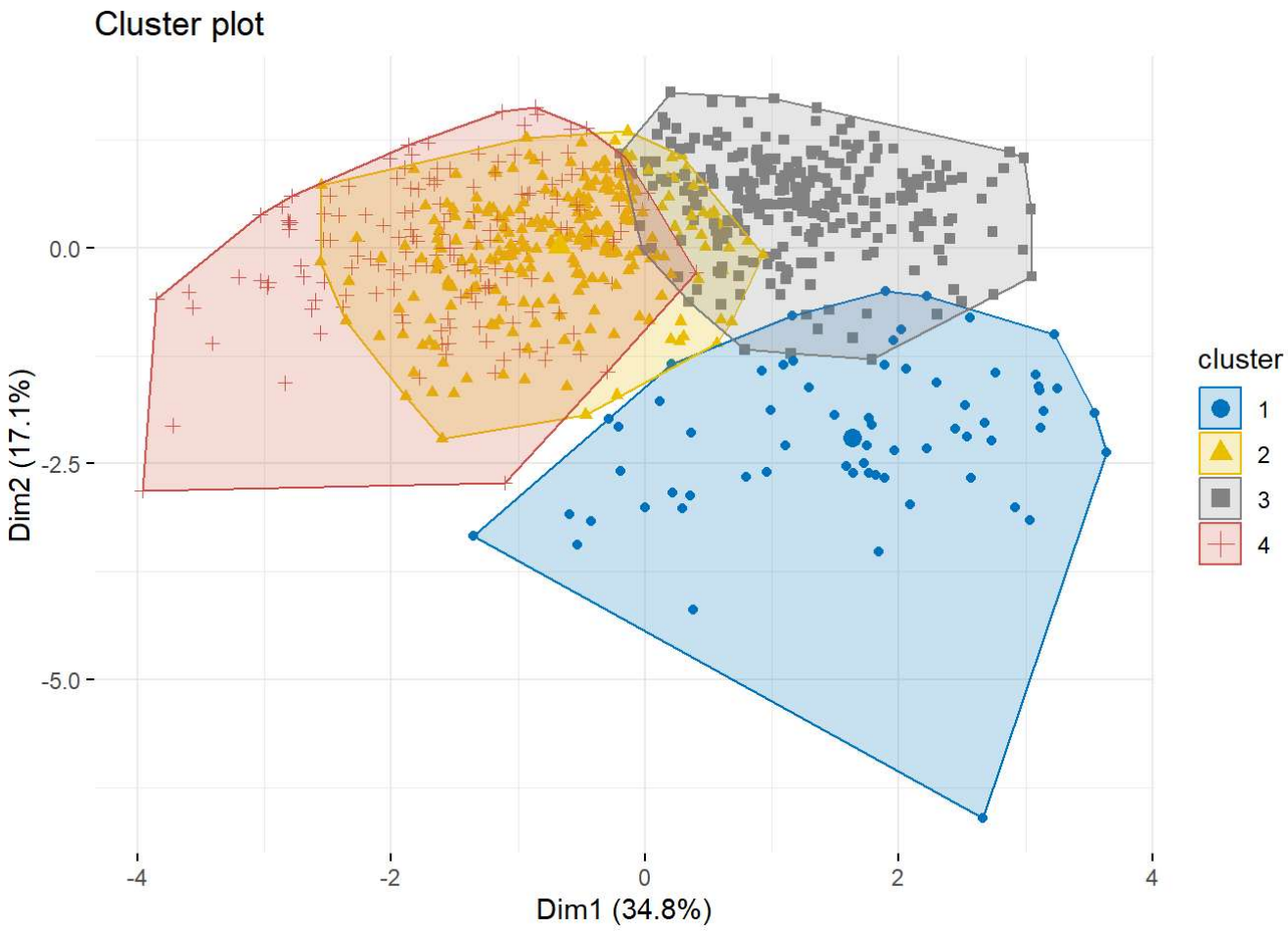

Q-Q Residuals
Standardized residuals
Theoretical Quantiles
568
367


Scale-Location
√|Standardized residuals|
Fitted values
568
367


Cook's distance
Cook's distance
Obs. number
696
367
568

Unsupervised: K-means Clustering

```
data_selected <- data %>%
  select(Age, Hours.per.day, Anxiety, Depression, Insomnia, OCD) %>%
  na.omit()

# save row indices of data used for clustering
row_indices <- row.names(data_selected)

# standardize data
data_scaled <- scale(data_selected)

# perform clustering (k = 4)
set.seed(123)  # Ensure reproducibility
kmeans_result <- kmeans(data_scaled, centers = 4, nstart = 25)

# data frame has a 'Cluster' column initialized with NA
data$Cluster <- NA

valid_indices <- which(complete.cases(data[, c("Age", "Hours.per.day", "Anxiety", "Depression",
"Insomnia", "OCD")]))

# assign the cluster results back to the original data using the valid indices
data$Cluster[valid_indices] <- kmeans_result$cluster

fviz_cluster(kmeans_result, data = data_scaled, geom = "point", ellipse.type = "convex",
             palette = "jco", ggtheme = theme_minimal())
```


Cluster plot
0.0
-2.5
-5.0
Dim2 (17.1%)
-4
-2
0
2
4
Dim1 (34.8%)
cluster
1
2
3
4