

DSCI304_FinalProj

Eleazar Martin

2024-12-03

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
```

```
## 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(RColorBrewer)
```

```
data <- read.csv("/Users/eleazarmartin/Desktop/mxmh_survey_results.csv")
```

```
## Most healthy pie chart
```

```
data_healthy <- data %>%
```

```
  filter(Depression == 0) %>%
```

```
  filter(Anxiety == 0) %>%
```

```
  filter(Insomnia == 0) %>%
```

```
  filter(OCD == 0)
```

```
data_summary <- data_healthy %>%
```

```
  count(Fav.genre) %>%
```

```
  mutate(percentage = n / sum(n) * 100)
```

```
data_summary <- data_summary %>%
```

```
  arrange(desc(percentage)) %>%
```

```
  mutate(Fav.genre = factor(Fav.genre, levels = Fav.genre))
```

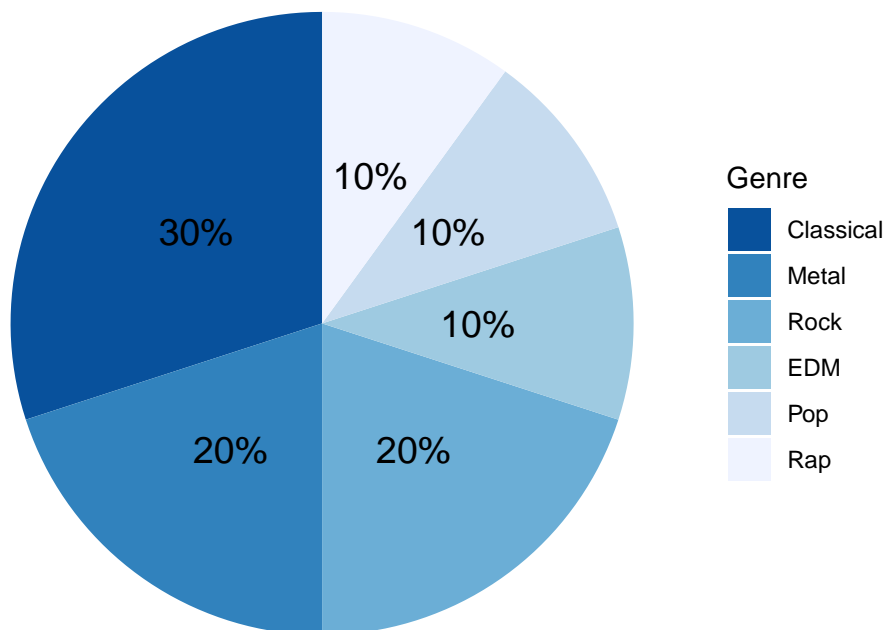
```

blues_palette <- rev(brewer.pal(n = nrow(data_summary), name = "Blues"))

ggplot(data_summary, aes(x = "", y = percentage, fill = Fav.genre)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  geom_text(aes(label = paste0(round(percentage, 1), "%"),
    position = position_stack(vjust = 0.5),
    color = "black", size = 5) +
  labs(title = "Classical music is the most common favorite genre among most \nmentally-healthy respondents")
  theme_void() +
  scale_fill_manual(values = blues_palette)

```

Classical music is the most common favorite genre among most mentally-healthy respondents



```

## Most unhealthy pie chart

data_unhealthy <- data %>%
  filter(Depression >= 8) %>%
  filter(Anxiety >= 8) %>%
  filter(Insomnia >= 8) %>%
  filter(OCD >= 8)

data_summary2 <- data_unhealthy %>%
  count(Fav.genre) %>%
  mutate(percentage = n / sum(n) * 100)

data_summary2 <- data_summary2 %>%

```

```

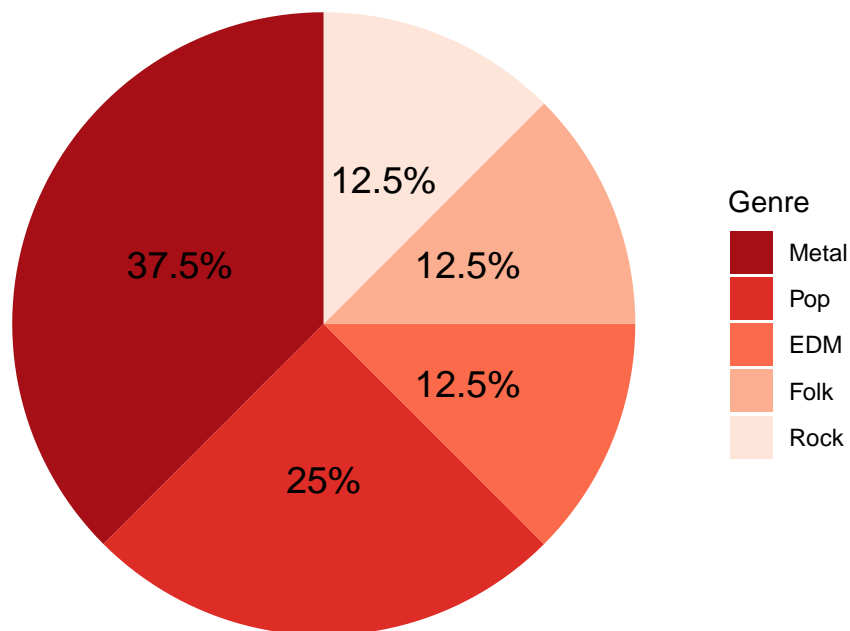
arrange(desc(percentage)) %>%
mutate(Fav.genre = factor(Fav.genre, levels = Fav.genre))

reds_palette <- rev(brewer.pal(n = nrow(data_summary2), name = "Reds"))

ggplot(data_summary2, aes(x = "", y = percentage, fill = Fav.genre)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  geom_text(aes(label = paste0(round(percentage, 1), "%")),
            position = position_stack(vjust = 0.5),
            color = "black", size = 5) +
  labs(title = "Metal music is the most common favorite genre among most \nmentally-unhealthy respondents",
        theme_void() +
  scale_fill_manual(values = reds_palette)

```

Metal music is the most common favorite genre among most mentally-unhealthy respondents



```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```

# Categorizing severity
categorize_severity <- function(value) {
  if (value <= 3) {
    return("Mild")
  } else if (value >= 4 & value <= 6) {

```

```

    return("Moderate")
  } else {
    return("Severe")
  }
}

data <- data %>%
  filter(!is.na(Depression) & !is.na(Anxiety) & !is.na(Insomnia) & !is.na(OCD))

data <- data %>%
  mutate(
    Anxiety_Severity = sapply(Anxiety, categorize_severity),
    Depression_Severity = sapply(Depression, categorize_severity),
    Insomnia_Severity = sapply(Insomnia, categorize_severity),
    OCD_Severity = sapply(OCD, categorize_severity)
  )

data_long <- data %>%
  pivot_longer(
    cols = c(Anxiety_Severity, Depression_Severity, Insomnia_Severity, OCD_Severity),
    names_to = "Condition",
    values_to = "Severity"
  ) %>%
  mutate(Condition = sub("_Severity", "", Condition))

grouped_data <- data_long %>%
  group_by(Condition, Severity) %>%
  summarize(Avg_Hours = mean(Hours.per.day))

```

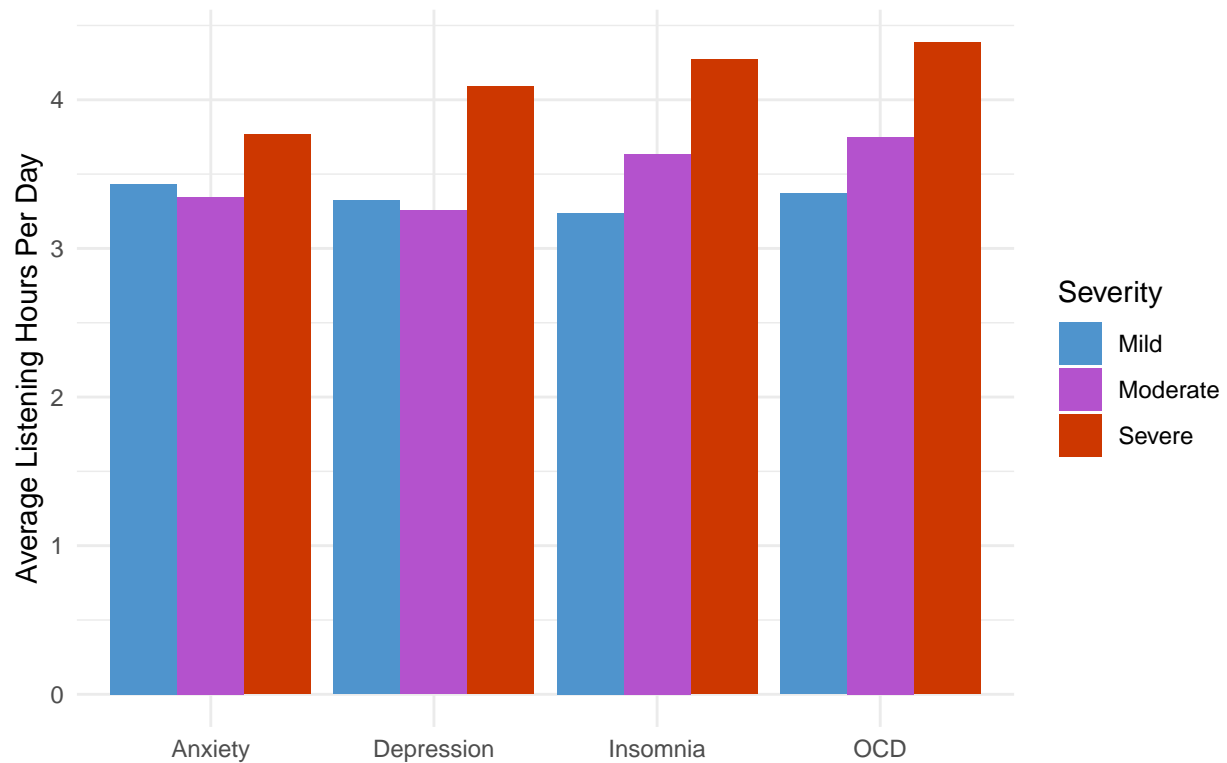
'summarise()' has grouped output by 'Condition'. You can override using the
'.groups' argument.

```

ggplot(grouped_data, aes(x = Condition, y = Avg_Hours, fill = Severity)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "People in the Severe Category Listen to the Most Music Daily",
    x = "",
    y = "Average Listening Hours Per Day",
    fill = "Severity"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 0, hjust = 0.5)) +
  scale_fill_manual(values = c("Mild" = "steelblue3", "Moderate" = "mediumorchid3", "Severe" = "oranger

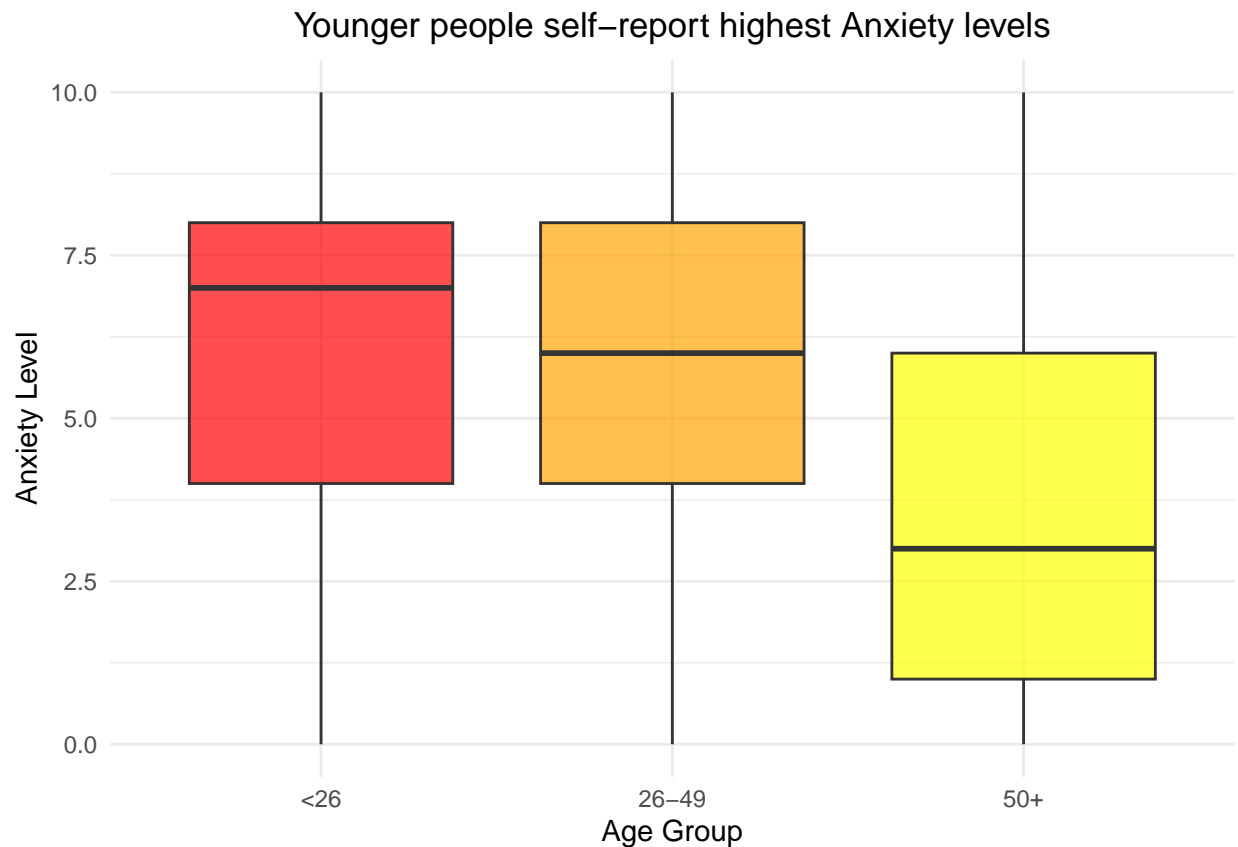
```

People in the Severe Category Listen to the Most Music Daily



```
data <- data %>%
  mutate(Age_Group = case_when(
    Age <= 25 ~ "<26",
    Age >= 50 ~ "50+",
    TRUE ~ "26-49"
  ))

ggplot(data, aes(x = Age_Group, y = Anxiety, fill = Age_Group)) +
  geom_boxplot(alpha = 0.7, outlier.color = "red", outlier.shape = 16) +
  scale_fill_manual(values = c("<26" = "red", "26-49" = "orange", "50+" = "yellow")) +
  labs(
    title = "Younger people self-report highest Anxiety levels",
    x = "Age Group",
    y = "Anxiety Level",
    fill = "Age Group"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 0, hjust = 0.5),
    legend.position = "none"
  )
```



```
library(ggplot2)
library(dplyr)
library(stargazer)
```

```
##
```

```
## Please cite as:
```

```
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
data <- data %>% filter(!is.na(Age))
```

```
data <- data %>%
  mutate(Fav.genre_encoded = ifelse(Fav.genre == "Classical", 0, 1))
```

```
data <- data %>%
  mutate(While.working_encoded = ifelse(While.working == "Yes", 1, 0))
```

```
model <- lm(Anxiety ~ Age + Fav.genre_encoded + Hours.per.day, data = data)
summary(model)
```

```
##
```

```
## Call:
```

```
## lm(formula = Anxiety ~ Age + Fav.genre_encoded + Hours.per.day,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.5128 -2.1951  0.6475  1.9604  5.8988
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.841107   0.448812  13.015 < 2e-16 ***
## Age           -0.040066   0.008402  -4.769 2.24e-06 ***
## Fav.genre_encoded  0.951128   0.391699   2.428  0.0154 *
## Hours.per.day    0.033474   0.033484   1.000  0.3178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.741 on 731 degrees of freedom
## Multiple R-squared:  0.0407, Adjusted R-squared:  0.03676
## F-statistic: 10.34 on 3 and 731 DF, p-value: 1.142e-06
```

```
stargazer(
  model,
  type = "text",
  title = "Coefficients from the Linear Regression Model",
  align = TRUE,
  no.space = TRUE,
  digits = 3,
  single.row = TRUE
)
```

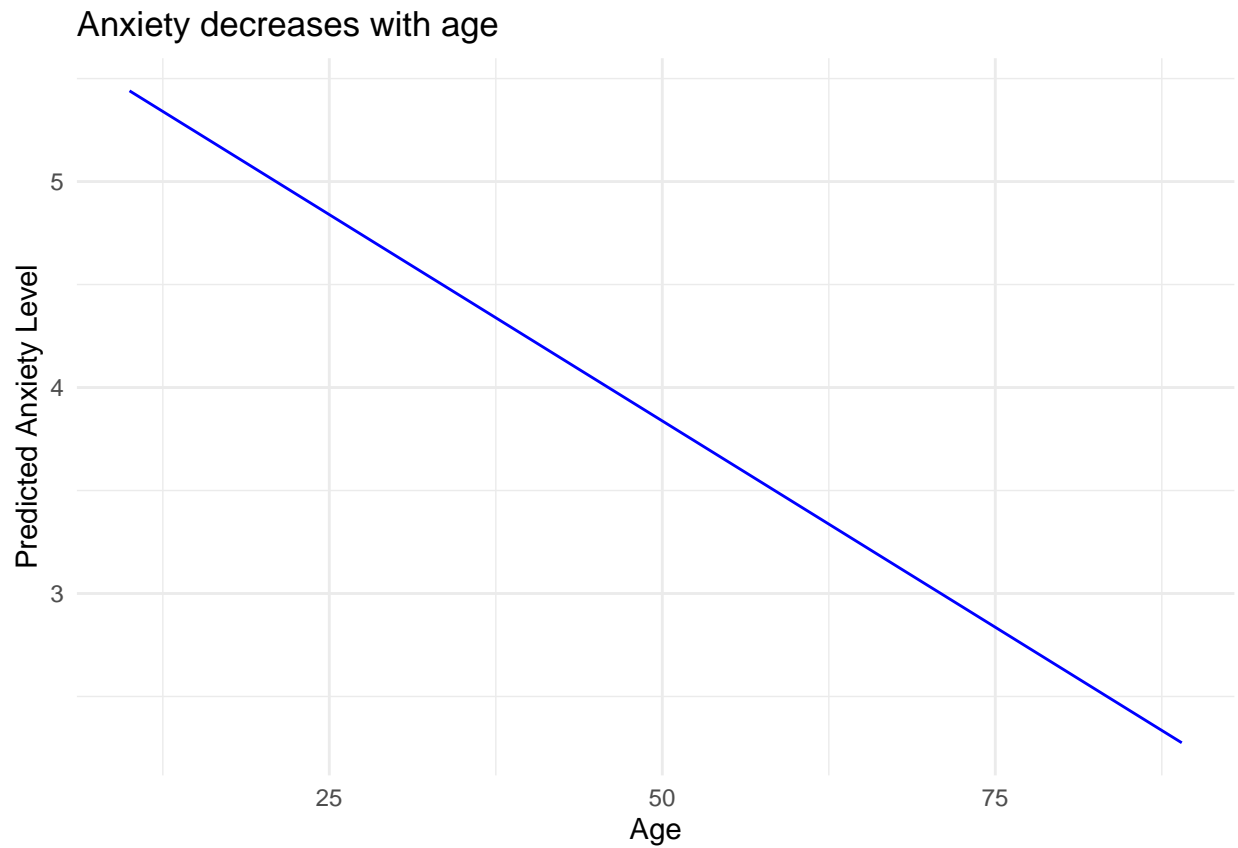
```
##
## Coefficients from the Linear Regression Model
## =====
##                               Dependent variable:
##                               -----
##                               Anxiety
## -----
## Age                          -0.040*** (0.008)
## Fav.genre_encoded            0.951** (0.392)
## Hours.per.day                0.033 (0.033)
## Constant                     5.841*** (0.449)
## -----
## Observations                  735
## R2                            0.041
## Adjusted R2                   0.037
## Residual Std. Error          2.741 (df = 731)
## F Statistic                   10.337*** (df = 3; 731)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

```
age_coeff <- coef(model)["Age"]
age_seq <- seq(min(data$Age), max(data$Age), length.out = 100)
age_effect <- data.frame(
```

```

Age = age_seq,
Predicted_Anxiety = coef(model)["(Intercept)"] + age_coeff * age_seq
)
ggplot(age_effect, aes(x = Age, y = Predicted_Anxiety)) +
  geom_line(color = "blue") +
  labs(
    title = "Anxiety decreases with age",
    x = "Age",
    y = "Predicted Anxiety Level"
  ) +
  theme_minimal()

```



```

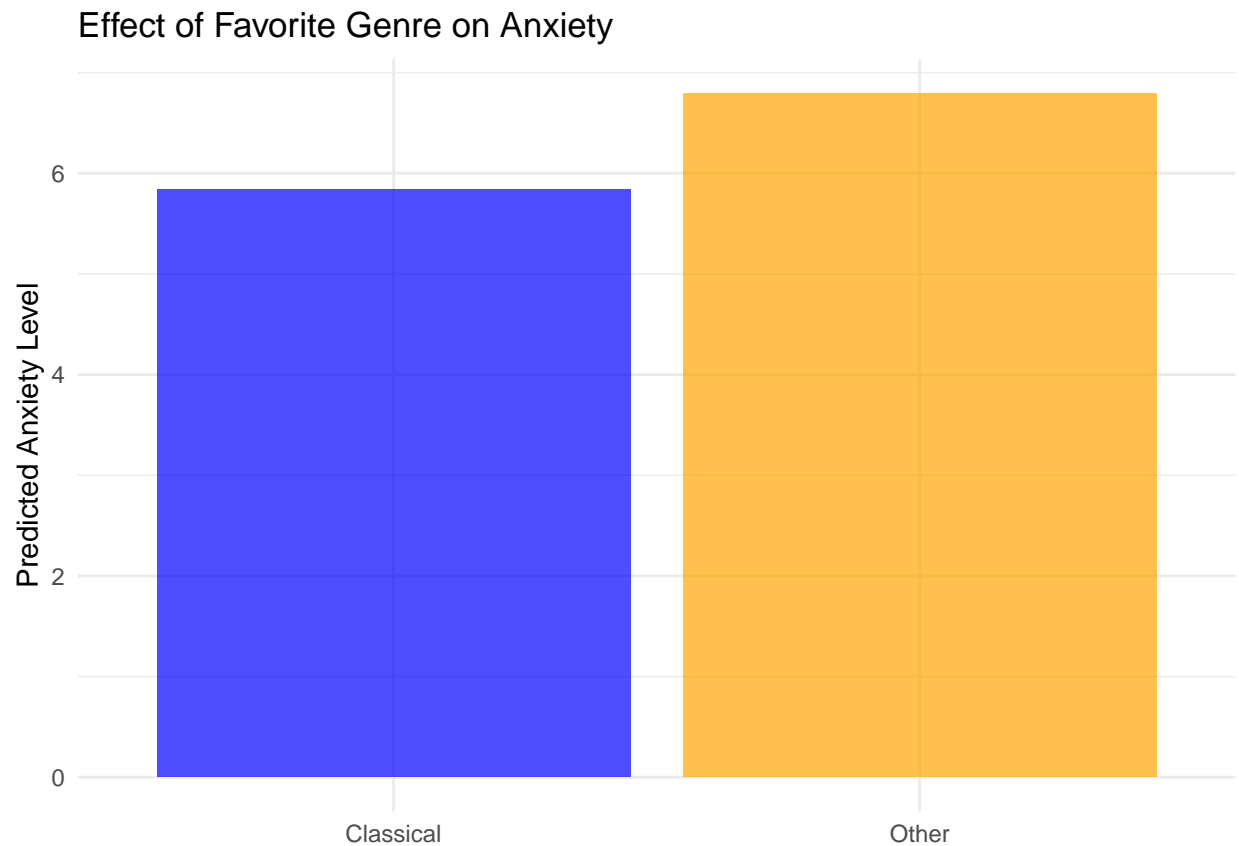
genre_coeff <- coef(model)["Fav.genre_encoded"]
genre_effect <- data.frame(
  Fav.genre_encoded = c("Classical", "Other"),
  Predicted_Anxiety = coef(model)["(Intercept)"] + genre_coeff * c(0, 1)
)

ggplot(genre_effect, aes(x = Fav.genre_encoded, y = Predicted_Anxiety)) +
  geom_bar(stat = "identity", fill = c("blue", "orange"), alpha = 0.7) +
  labs(
    title = "Effect of Favorite Genre on Anxiety",
    y = "Predicted Anxiety Level"
  ) +
  theme_minimal() +

```



```
theme(
  axis.title.x = element_blank()
)
```



```
data <- data %>%
  filter(!is.na(Age), !is.na(Hours.per.day)) %>%
  mutate(
    Fav.genre_encoded = ifelse(Fav.genre == "Classical", 0, 1),
    While.working_encoded = ifelse(While.working == "Yes", 1, 0)
  )

model_interaction <- lm(Anxiety ~ Age + Fav.genre_encoded * Hours.per.day, data = data)
summary(model_interaction)
```

```
##
## Call:
## lm(formula = Anxiety ~ Age + Fav.genre_encoded * Hours.per.day,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.4731 -2.2043  0.6481  1.9644  5.8998
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)          5.347475    0.717071    7.457 2.51e-13 ***
## Age                 -0.040399    0.008412   -4.803 1.90e-06 ***
## Fav.genre_encoded    1.471527    0.707816    2.079  0.038 *
## Hours.per.day        0.207784    0.200281    1.037  0.300
## Fav.genre_encoded:Hours.per.day -0.179380  0.203204   -0.883  0.378
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.741 on 730 degrees of freedom
## Multiple R-squared:  0.04172,    Adjusted R-squared:  0.03647
## F-statistic: 7.945 on 4 and 730 DF,  p-value: 2.853e-06
```

```
interaction_effect <- expand.grid(
  Fav.genre_encoded = c(0, 1),
  Hours.per.day = seq(min(data$Hours.per.day, na.rm = TRUE),
    max(data$Hours.per.day, na.rm = TRUE),
    length.out = 100),
  Age = mean(data$Age, na.rm = TRUE)
)

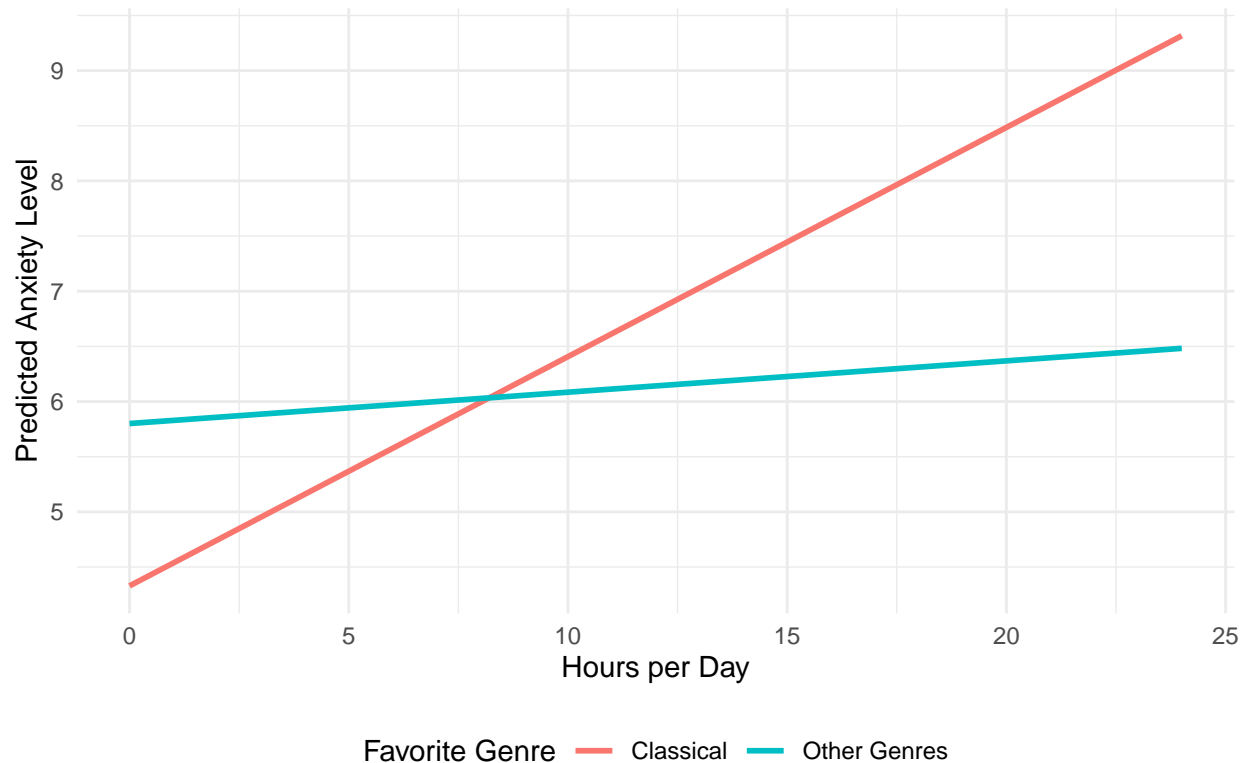
interaction_effect$Predicted_Anxiety <- predict(model_interaction, newdata = interaction_effect)

interaction_effect$Fav.genre_label <- ifelse(interaction_effect$Fav.genre_encoded == 0, "Classical", "0")

ggplot(interaction_effect, aes(x = Hours.per.day, y = Predicted_Anxiety, color = Fav.genre_label)) +
  geom_line(size = 1) +
  labs(
    title = "More than 7 hours a day for Classical lovers increases anxiety",
    x = "Hours per Day",
    y = "Predicted Anxiety Level",
    color = "Favorite Genre"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    legend.position = "bottom"
  )
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

More than 7 hours a day for Classical lovers increases anxiety



```
library(dagitty)
```

```
## Warning: package 'dagitty' was built under R version 4.2.3
```

```
library(ggdag)
```

```
##
```

```
## Attaching package: 'ggdag'
```

```
## The following object is masked from 'package:stats':
```

```
##
```

```
## filter
```

```
dag <- dagitty("  
  dag {  
    Age -> Favorite_Genre  
    Favorite_Genre -> Listening_While_Working  
    Listening_While_Working -> Listening_Hours  
    Listening_Hours -> Anxiety  
    Age -> Anxiety  
  }  
")
```

```
label_map <- c(  
  Favorite_Genre = "Favorite Genre"  
  Listening_While_Working = "Listening While Working"  
  Listening_Hours = "Listening Hours"  
  Anxiety = "Anxiety Level"  
)
```

```

"Age" = "Age",
"Favorite_Genre" = "Favorite\nGenre",
"Listening_While_Working" = "Listening\nWhile\nWorking",
"Listening_Hours" = "Listening\nHours",
"Anxiety" = "Anxiety"
)

ggdag(dag, layout = "circle") +
  geom_dag_node(size = 20) +
  geom_dag_text(aes(label = label_map[name]), size = 2) +
  geom_dag_edges_link(edge_width = 0.8, arrow_directed = grid::arrow(length = grid::unit(4, "mm"))) +
  coord_fixed() +
  theme_void() +
  labs(title = "Hypothesized Pathway: Age to Anxiety via Listening Habits") +
  theme(plot.title = element_text(hjust = 0.5, size = 16, face = "bold"))

## Warning in geom_dag_edges_link(edge_width = 0.8, arrow_directed =
## grid::arrow(length = grid::unit(4, : Ignoring unknown parameters:
## 'arrow_directed'

```

Hypothesized Pathway: Age to Anxiety via Listening Habits

