PSY6422 Data Analysis and Visualisation: Module Project

230170391

2024-04-16

Project Focus

The project concerns mental health data, based on its significance as a facet of health and wellbeing in society (https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response). I wanted to create a visualisation specifically looking at mood disorders (depression and anxiety scores), in relation to music listening and taste.

"Psychologists finds that when listening to music to relieve stress 'genre doesn't matter'."

https://medicalxpress.com/news/2024-01-psychologists-music-relieve-stress-genre.html

The 2024 article by Radboud University states that listening to music relieves stress, irrespective of genre. However, no quantitative data was reported illustrating the methodological approach to the findings. Therefore, for the module project I used a dataset reporting anxiety and depression scores at a continuous level, alongside reported listening hours per day and discrete genre affiliations.

The Data

Data was taken from 'kaggle', an online resource for gathering and reporting datasets. The specific webpage for the current data contained survey results for a large sample (n=736) of self-reported music taste and mental health data, as well as other variables such as listening frequencies.

A full reference of the data is: Rasgaitis, C. (2022). Music and Mental Health Survey Results. [Survey]. Kaggle.

The data can be accessed via the following link: https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results?select=mxmh_survey_results.csv

The data was downloaded as .csv to the working directory and then imported into r studio in this format:

```r
# Load the dataset in the csv file into R
dataset <- read.csv("music_mentalhealth_data.csv")
```

Data Preparation

The raw dataset included variables that were not relevant to this analysis, therefore several columns were removed for easier viewing. Filtering genres also provided a less crowded, more aesthetic approach to the visualisation that focused exclusively on the most popular ten music genres. Subsequent data preparation identified mean scores and reshaped the data, making it appropriate for analysis. The process of data cleaning and pre-processing in r is exemplified below, including the installation of necessary packages.

```r
# Set a CRAN mirror determining server of packages in r
chooseCRANmirror(ind = TRUE)

# Install packages and load libraries

install.packages("webshot2")
```

```
##
## The downloaded binary packages are in
##  /var/folders/8z/cwq9jwdn1xsc1rh4ty_s5ksr0000gn/T//RtmpBK5cRZ/downloaded_packages
```

```r
library(webshot2) # For representative screenshots of interactive plots in PDF

install.packages("here")
```

```
##
## The downloaded binary packages are in
##  /var/folders/8z/cwq9jwdn1xsc1rh4ty_s5ksr0000gn/T//RtmpBK5cRZ/downloaded_packages
```

```r
library(here) # To set working directory

install.packages("tidyverse")
```

```
##
## The downloaded binary packages are in
##  /var/folders/8z/cwq9jwdn1xsc1rh4ty_s5ksr0000gn/T//RtmpBK5cRZ/downloaded_packages
```

```r
library(tidyverse) # For data preparation

install.packages("dplyr")
```

```
##
## The downloaded binary packages are in
##  /var/folders/8z/cwq9jwdn1xsc1rh4ty_s5ksr0000gn/T//RtmpBK5cRZ/downloaded_packages
```

```r
library("dplyr") # For data manipulation

install.packages("ggplot2")
```

```
##
## The downloaded binary packages are in
##  /var/folders/8z/cwq9jwdn1xsc1rh4ty_s5ksr0000gn/T//RtmpBK5cRZ/downloaded_packages
```

```r
library(ggplot2) # To create plots

install.packages("plotly")
```

```
##
## The downloaded binary packages are in
##  /var/folders/8z/cwq9jwdn1xsc1rh4ty_s5ksr0000gn/T//RtmpBK5cRZ/downloaded_packages
```

```r
library("plotly") # To create interactive plots

# remove variables from dataset that are not relevant to the visualisation
filtered_data <- select(dataset, -c("Timestamp", "Age", "Primary.streaming.service", "While.working", "]

#label count of favourite genre on each bar
genre_counts <- table(filtered_data$Fav.genre)

#put favourite genres in decreasing order
sorted_genres <- sort(genre_counts, decreasing = TRUE)

#find top 10 favourite genres
top_10_genres <- names(sorted_genres)[1:10]

# Remove the remaining genres from the dataset
data_top_10 <- filtered_data[filtered_data$Fav.genre %in% top_10_genres, ]


# Create a new table with mean scores for anxiety and depression for hours of listening per day
mean_scores_by_hours <- data_top_10 %>%
  group_by(Hours.per.day) %>%
  summarise(
    depression = mean(Depression),
    anxiety = mean(Anxiety)
  )

# Create another table with the mean scores for anxiety and depression for each favourite genre_counts
mean_scores_by_genre <- data_top_10 %>%
  group_by(Fav.genre) %>%
  summarise(
    depression = mean(Depression),
    anxiety = mean(Anxiety)
  )




# Reshape both datasets to long format
mean_scores_long_genre <- pivot_longer(mean_scores_by_genre,
                              cols = c(depression, anxiety),
                              names_to = "Mental_Health",
                              values_to = "Mean_Score")

mean_scores_long_hours <- pivot_longer(mean_scores_by_hours,
                              cols = c(depression, anxiety),
                              names_to = "Mental_Health",
                              values_to = "Mean_Score")
```
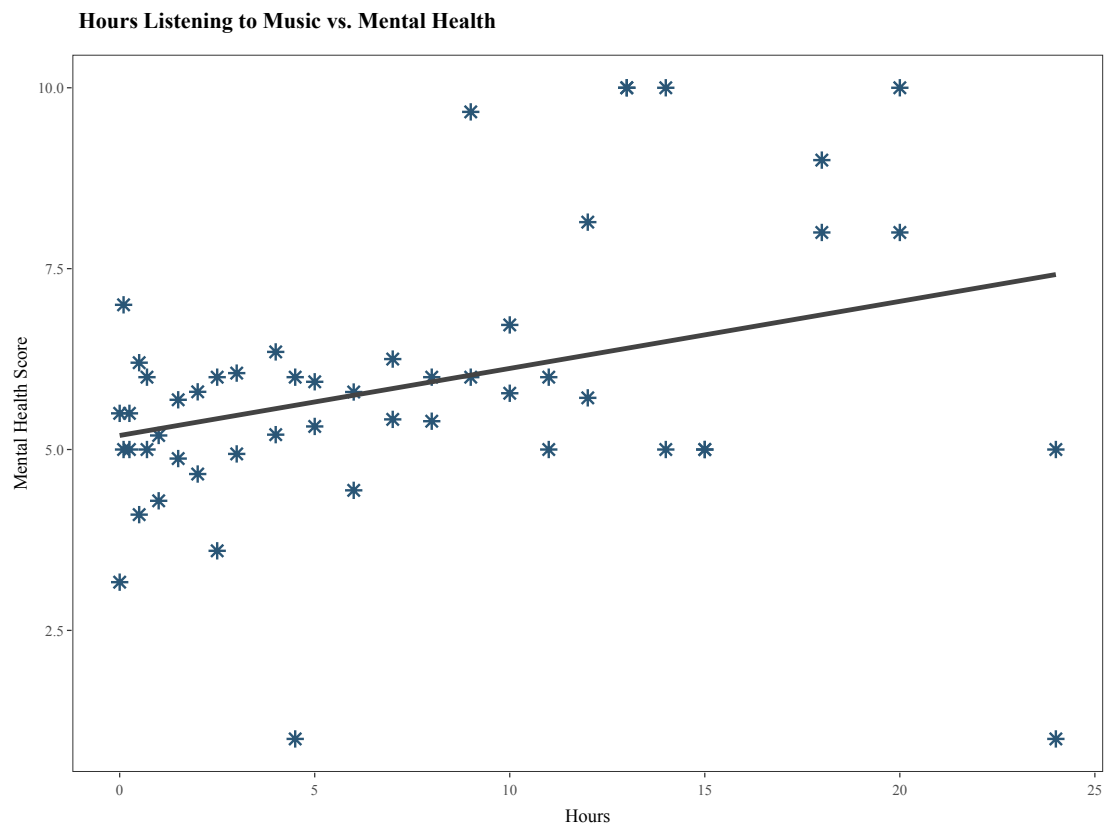
Visualisations

The colour palettes used in the visualisations were imported from the following websites: https://r-charts.com/color-palettes/ https://stackoverflow.com/questions/57153428/r-plot-color-combinations-that-are-colorblind-accessible

Scatterplot

The first visualisation is a scatterplot, created to inform the correlational relationship between hours spent
listening to music and mood disorder scores. The long data prepared for this analysis enabled identification
of a mean mood disorder score from distinct anxiety and depression scores. The colour scheme of dark blue
and grey aligns with mood disorder colour psychology (https://www.nbcnews.com/id/wbna35304133).

```r
# Create a scatterplot with hours of listening on the x axis and mean depression and anxiety scores on
# Add a title, subtitle and label axes
# Remove of background grid lines
# Make title bold
# Change font to Times New Roman
# Change colour, size and shape of points
p<- ggplot(mean_scores_long_hours, aes(x = Hours.per.day, y = Mean_Score)) + geom_point(colour = "#2A56
  geom_smooth(method = "lm", se = FALSE, colour = "#434343") +
  labs(x = "Hours", y = "Mental Health Score", subtitle = "The correlation between hours listening to mu
  ggtitle("Hours Listening to Music vs. Mental Health") +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        plot.title = element_text(face = "bold", family = "Times New Roman"),
        axis.text = element_text(family = "Times New Roman"),
        axis.title = element_text(family = "Times New Roman"),
        plot.subtitle = element_text(family = "Times New Roman"))

# Make the plot interactive by converting ggplot to plotly
plotly_p <- ggplotly(p)
plotly_p
```

Hours Listening to Music vs. Mental Health


```
# Save the scatterplot
ggsave("scatterplot.png", plot = p, width = 6, height = 4, dpi = 300)
```

Bar charts

Two identical visualisations were then created to show to relationship between favourite music genres and mood disorder scores. The first visualisation was created with custom colours whereas the second offers a colourblind friendly palette.

I visualised the mood disorder and favourite genre data using a bar chart, with individual bars for anxiety and depression to highlight their differences. This method complimented the categorical nature of the music genres in the dataset. I also added an interactive aspect to the bars, enabling visualisation of specific mean scores. The step-by-step process of the development of the visualisations are illustrated below.

Custom Bar Chart The first visualisation utilises dark grey and blue tones again, consistent with anxiety and depression affiliations and mood in colour psychology (https://www.nbcnews.com/id/wbna35304133).

```r
# Label two colours for the visualisation
custom_colours <- c(anxiety = "#434343", depression = "#2A5676")

# Create labels
custom_font <- "Times New Roman"
title_text <- "Mood Disorders by Favorite Music Genre"
subtitle_text <- "A visualisation using mean anxiety and depression scores"
x_label_text <- "Favorite Genre"
y_label_text <- "Score"
fill_label_text <- "Mood Disorders"
max_score <- 7

# Create a bar chart to show anxiety and depression scores by favourite music genre
# Add a title to the bar chart
# Make Title bold
# Add a subtitle to the bar chart
# Label the key
# Align the bars with the x axis
# Add integer lines on the y axis, increasing by 1
# Change colours to colourblind friendly colours
# Make the writing on the x axis slanted
# Delete grid lines
# Add bar lines
# Add ticks
# Centre the title and subtitle
# Change the font of labels and text to the custom font
# Outline the bars
# Turn the ggplot into an object and make the formatting the same as the original plot
# Make titles and axes labels bold
# Change font of labels and text to custom font
p1 <- ggplot(mean_scores_long_genre, aes(x = Fav.genre, y = Mean_Score, fill = Mental_Health, text = pa
  geom_bar(stat = "identity", position = "dodge", colour = "black") +
  labs(title = title_text,
       x = x_label_text,
       y = y_label_text,
       fill = fill_label_text) +
  scale_y_continuous(expand = c(0, 0), breaks = seq(0, max_score, by = 1)) +
  scale_fill_manual(values = custom_colours) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, family = custom_font),
```

```r
        axis.title = element_text(face = "bold", family = custom_font),
        legend.title = element_text(face = "bold", family = custom_font),
        legend.text = element_text(family = custom_font),
        legend.position = "right",
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(color = "black"),
        axis.ticks = element_line(color = "black"),
        plot.title = element_text(hjust = 0.5, family = custom_font),
        plot.subtitle = element_text(family = custom_font) )

# Add the subtitle into the plot again as it disappears
# Make the title bold and underlined
p1 <- p1 + ggtitle(paste("<b>", title_text, "</b>", "\n", subtitle_text))


# Make the plot interactive by converting ggplot to plotly
p1 <- ggplotly(p1, tooltip = "text")


# Adjust font sizes
p1 <- layout(p1,
         title = list(font = list(size = 15)),
         font = list(size = 6),
         xaxis = list(title = list(font = list(size = 13))),
         yaxis = list(title = list(font = list(size = 13))),
         legend = list(
           title = list(text = "Mood Disorders", font = list(size = 13, face = "bold"))
         ))

# Generate interactive plot
p1
```
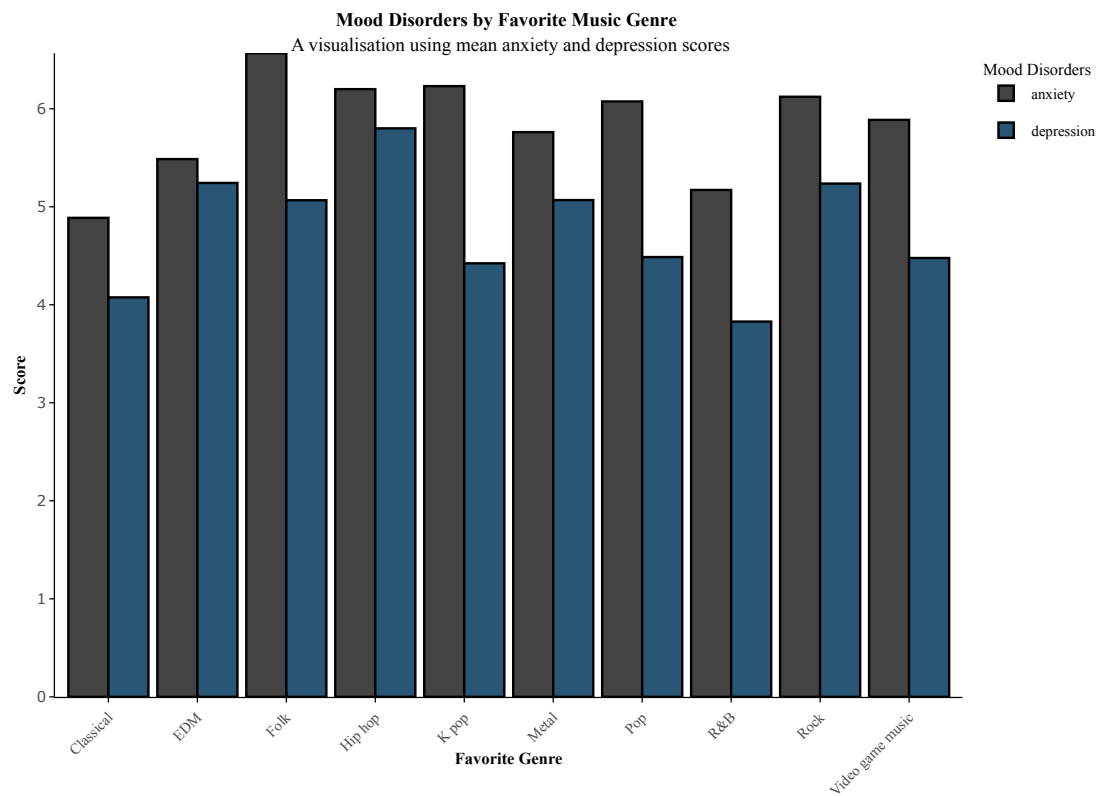
Mood Disorders by Favorite Music Genre

A visualisation using mean anxiety and depression scores


```
# Save the plot
htmlwidgets::saveWidget(p1, "plot1_customcolours.html")
```

Colourblind Bar Chart   The second visualisation offers colourblind friendly colours.

```r
# Label two colours from the colourblind palette
colourblind_colours <- c(anxiety = "#CC6677", depression = "#882255")

# Create a second, identical plot but with colourblind friendly colours
p2 <- ggplot(mean_scores_long_genre, aes(x = Fav.genre, y = Mean_Score, fill = Mental_Health, text = pas
  geom_bar(stat = "identity", position = "dodge", colour = "black") +
  labs(title = title_text,
       x = x_label_text,
       y = y_label_text,
       fill = fill_label_text) +
  scale_y_continuous(expand = c(0, 0), breaks = seq(0, max_score, by = 1)) +
  scale_fill_manual(values = colourblind_colours) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, family = custom_font),
        axis.title = element_text(face = "bold", family = custom_font),
        legend.title = element_text(face = "bold", family = custom_font),
        legend.text = element_text(family = custom_font),
        legend.position = "right",
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line = element_line(color = "black"),
        axis.ticks = element_line(color = "black"),
        plot.title = element_text(hjust = 0.5, family = custom_font),
        plot.subtitle = element_text(family = custom_font) )

# Add the subtitle into the plot again as it disappears
# Make the title bold and underlined
p2 <- p2 + ggtitle(paste("<b>", title_text, "</b>", "\n", subtitle_text))


# Make the plot interactive by converting ggplot to plotly
p2 <- ggplotly(p2, tooltip = "text")


# Adjust font sizes
p2 <- layout(p2,
             title = list(font = list(size = 15)),
             font = list(size = 6),
             xaxis = list(title = list(font = list(size = 13))),
             yaxis = list(title = list(font = list(size = 13))),
             legend = list(
               title = list(text = "Mood Disorders", font = list(size = 13, face = "bold"))
             ))

# Generate interactive plot
p2
```
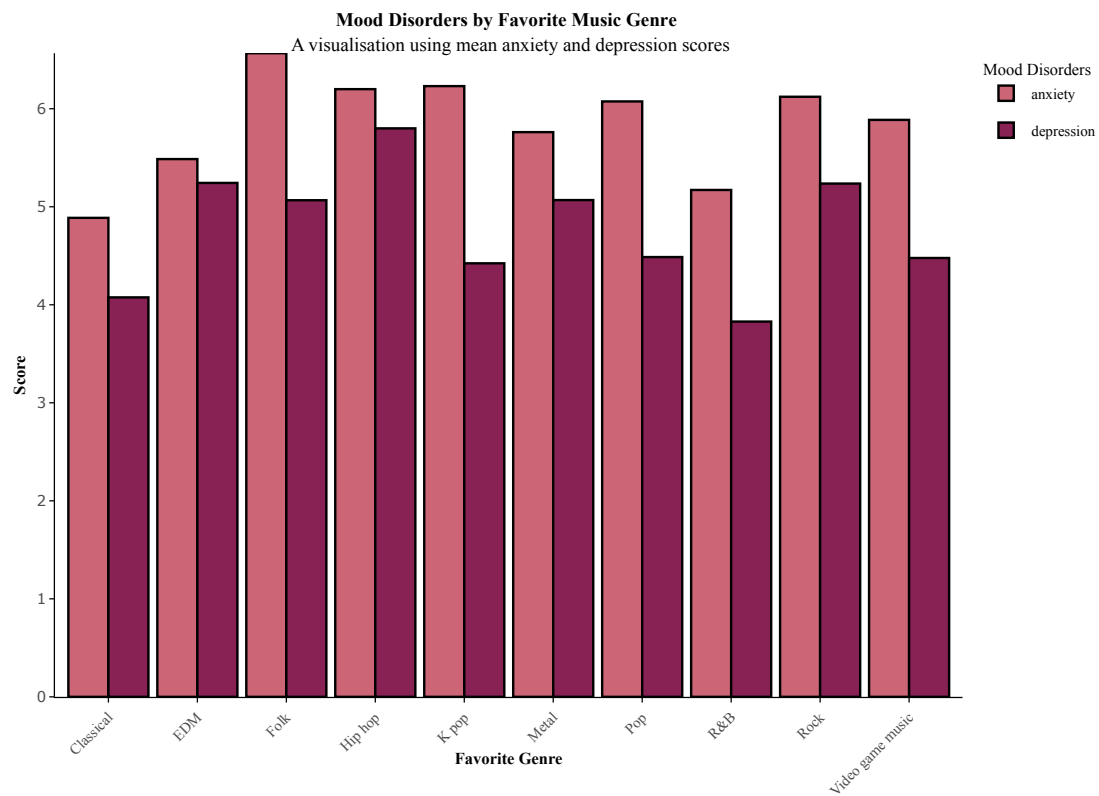
Mood Disorders by Favorite Music Genre
A visualisation using mean anxiety and depression scores


```
# Save the plot
htmlwidgets::saveWidget(p2, "plot2_colourblind.html")
```

Conclusions

Hours spent listening to music   The scatterplot suggests a small, positive relationship between mean mood disorder scores and hours spent listening to music. This appears to contradict the assumption that listening to music relieves stress. However, a possible explanation is that individuals suffering with their mental health are listening to more music per day with the intention of relieving their higher stress levels or because they are unable to engage in typical social and occupational activities to the same extent as others, therefore permitting more time for leisure.

Favourite genre   The bar-chart depicts that favouring hip-hop transcribes the highest mean score of depression, whereas the highest mean score in anxiety appears in connection with folk music. On the other hand, the lowest score in depression is visualised in relation to liking R&B, whereas the lowest score in anxiety correlates with an affiliation for classical music. Despite the general findings, the 2.0 range in depression scores and 1.7 range in anxiety scores suggests minimal difference in mental health dependent on favourite music genre.

Limitations   The correlational nature of the analysis limits any causal conclusions. Furthermore, based on the sample population of subclinical respondents, it is unclear as to whether clinically diagnosed mood disorders have any association with music taste. Future research could focus on a clinical population.