INFO 3010

Spring 2024

The Effects of Music on Mental Health: An Analysis in R

1. Abstract (10 pts)

This project delves into the relationship between music listening habits and self-reported mental health conditions, exploring how different genres of music influence symptoms of anxiety, depression, insomnia, and OCD. Utilizing the MxMH dataset, which contains data from various demographics, this analysis employs both exploratory data analysis and statistical testing to investigate correlations between music genre preferences, musicianship, and mental health across different age groups. Through data visualization, correlation analysis, linear regression models, and K-means clustering, the study reveals diverse patterns in how different music genres potentially affect mental health symptoms. Notably, genres like Metal and Rock are associated with higher depression and insomnia scores, whereas Classical music shows a negative correlation with anxiety. These findings also challenge the "tortured artist" stereotype by showing no significant differences in mental health conditions between musicians and non-musicians. This research thus contributes valuable insights into the field of music therapy, suggesting that tailored music listening, considering personal mental health needs and preferences, could enhance therapeutic outcomes.

2. Introduction (20 pts)

Music therapy has been recognized for its ability to improve mental health by leveraging music's ability to stimulate "happy" hormones such as oxytocin. Despite its benefits, the application of music therapy varies considering the diverse genres being listened to across a diverse population. The MxMH dataset raises the following questions:

1. What correlations exist between different music genres and their self-reported mental health conditions?
2. Is there a correlation between self-reported mental health and identifying with being a musician (i.e. instrumentalist or composer)?
3. Which age group most struggles with mental health? Does it correlate with listening hours?

This report aims to determine if specific genres of music are more effective in alleviating symptoms of mental health issues like anxiety, depression, insomnia, and OCD. By analyzing these correlations, this project seeks to improve the application of music therapy by aligning music listening preferences with personal mental health needs, potentially leading to more effective therapeutic outcomes.

This report will include an overview of the dataset, followed by an exploratory data analysis to visualize and summarize the data. Later sections will delve into more rigorous statistical testing and data modeling to uncover patterns and relationships. The findings from these analyses will be later discussed in the context of their implications for music therapy practices and general mental health awareness. Ultimately, this investigation intends to contribute to a broader understanding of how music can influence mental health, providing valuable insights for music enthusiasts, psychologists, therapists, and other members of the public.

INFO 3010

Spring 2024

3. Dataset (30 pts)

The MxMH dataset contains quantitative and categorical data collected from respondents on a Google Form survey. This survey was posted on online platforms like Reddit, Discord, and other social media as well as advertisements at various public locations. It encompasses a domain focused on the intersection between music preferences and mental health— a domain that not only relates to the arts and entertainment but also to health and psychology. This integration of music with mental well-being aligns with the therapeutic practice of Music Therapy (MT), which leverages musical interaction to alleviate stress and enhance mood through the stimulation of hormones like oxytocin.

The dataset was sourced from Kaggle.com (Rasgastis[1]). The timestamps of the survey results span from August to November 2022. Most of the data was collected upon its release in August 2022, with a decreasing average rate of responses[2] over the subsequent months (see time distribution).

The dataset includes the following attributes:

1. *Timestamp* (date-time)
2. *Age* (years)
3. *Primary streaming service*
4. *Hours per day*
5. *Demographics and listening habits*: musical (instrumentalist/composer) and personal habits (while working, exploratory, foreign language)
6. *Favorite genre*
7. *Beats per minute (BPM)*
8. *Frequency of Listening to Various Music Genres*: how often they listen to 16 different genres, choosing from 'Never,' 'Rarely,' 'Sometimes,' or 'Very frequently.'
9. *Mental Health Metrics*: self-reported levels of Anxiety, Depression, Insomnia, and OCD, ranked on an ordinal scale from 0 to 10.
10. *Permissions*

4. Exploratory Data Analysis (EDA) (60 pts)

The summary statistics provided a detailed overview of the participants' music listening habits, demographic information, and reported mental health conditions. The ages of participants range from 10 to 89, with a median age of 21 and a mean of 25.21, indicating a young demographic skewed towards younger individuals (since the mean is greater than the median. Further evidenced by the 1st quartile at 18 years, and the 3rd quartile at 28 years). This is evident in *Figure 6: Age Distribution of Respondents*, as a right-skewed distribution is clearly shown. Participants report listening to music between 0 and 24 hours per day (which is obvious since there are only 24 hours in a day), with a median of 3 and a mean of approximately 3.57 hours (Q1 = 2 and Q3 = 5— another right skew). When looking at BPM, the data shows a range of 0 – 999999999 BPM which indicates that some of the responses might be erroneous.

---

[1] https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results?resource=download

[2] Note: average rate of responses refers to the use of a trend line, rather than the derivative of the function (i.e. instantaneous rate)

The metrics for Anxiety, Depression, Insomnia, and OCD are on a normalized scale from 0 to 10 with interquartile ranges for these metrics decreasing in the respective aforementioned order. Lastly, the attributes for timestamp, music listening habits, instrumentalist/composer, genre frequency, and permissions contained character data and thus contain little information in the summary statistics.

The bar chart in Figure 1 demonstrates that the median anxiety scores across various genres are relatively high, showing less variation in response to listening frequency compared to other mental health conditions. Anxiety levels appear consistently high across most genres and frequencies, suggesting a broad impact of music listening habits on self-reported anxiety, irrespective of genre. Figure 2, on the other hand, reveals a less consistent pattern in median depression scores across genres compared to anxiety. However, there is a noticeable increase in depression scores associated with frequent listening to genres such as Metal, Jazz, and Rock. This suggests that these genres might have a more pronounced impact on depression or that individuals with depression may gravitate towards these music types. Insomnia scores in Figure 3 show moderate consistency but with certain genres like Jazz and Metal showing higher scores particularly at higher listening frequencies. This might indicate a potential stimulating effect of certain music genres on sleep disturbances, or conversely, that individuals suffering from insomnia might use certain types of music as a coping mechanism during sleepless nights. OCD scores in Figure 4 exhibit a generally consistent pattern across different genres and listening frequencies, like insomnia but with generally lower median scores. This uniformity suggests a lesser impact of music listening preferences on OCD symptoms compared to anxiety or depression.

The correlation heatmap in Figure 5 provides a more specific view of the relationships between music genres and mental health conditions. Notably, there is a positive correlation between listening to Metal and higher scores of Insomnia and Depression, and Rock with Depression, highlighting these genres' significant associations with negative mental health outcomes. Conversely, a negative correlation is observed between Anxiety and Classical music, and Depression and Country music, suggesting these genres might be inversely related to these mental health issues.

The histogram in Figure 6 shows a significant portion of respondents in the younger age groups, particularly around the early 20s, with the count decreasing for older age groups. This skewed distribution indicates a predominantly younger demographic within the dataset, which could influence mental health reporting given younger individuals are overrepresented. Furthermore, the skew likely results from selection bias in the data collection method as the dataset was gathered through online platforms and public advertisements in places like universities, where both are primarily frequented by a younger demographic. When looking at Figure 7, depression scores peak in the mid-life age groups (19-35 and 36-50) and decrease among older adults (66+). The highest average anxiety scores are observed in the 'NA' category followed closely by the 19-35 and 0-18 age groups. Anxiety levels are notably high among younger adults and adolescents, possibly reflecting the unique stressors and challenges faced by these age groups, such as academic pressures, career challenges, social dynamics, etc. In Figure 8, the 19-35 age group shows the highest level of depression, with the 0-18 and 36-50 age groups displaying approximately equal levels, which are slightly lower. Young adults (19-35) experience the highest levels of depression, which may be linked to the transitional nature of this life period, such as academics, careers, relationships, children, or existential dread. In Figure 9, the 36-50 age group reports the highest insomnia scores, followed by the 0-18 and 19-35 groups. This trend could be attributed to the various stressors prevalent in this life stage, including career pressures, physiological change, health complications, and family responsibilities. Younger groups also report significant insomnia, potentially exacerbated by lifestyle choices and digital consumption habits. In Figure 10, the 'NA' category shows high OCD scores, significantly exceeding those in categorized age groups, where scores are generally less than 3. The extreme scores in the 'NA' group could indicate data entry errors or a specific subset of respondents with severe OCD conditions who neglected to specify their age for some reason. For the rest, OCD symptoms appear relatively moderate and less varied across age groups compared to other mental health metrics.

The ANOVA results provided insights into the impact of age on various mental health conditions among the respondents. Each test investigates the differences in mean scores for Anxiety, Depression, Insomnia, and OCD across defined age groups. The ANOVA for Anxiety yields a highly significant result with an F-value of 9.684 and a p-value of 1.24e-07, indicating strong evidence of differences in anxiety levels across the age groups. This significant result suggests that age significantly influences anxiety symptoms among the respondents, with certain age groups potentially experiencing higher or lower levels of anxiety. Similarly, the ANOVA for Depression also shows a highly significant effect of age on depression scores, with an F-value of 9.978 and a p-value of 7.27e-08. This reinforces the idea that age is a critical factor in the variance of depression levels among different demographics. The significance underscores the varying challenges or life conditions associated with different life stages that may influence depression. On the other hand, the ANOVA of Insomnia does not reveal a significant difference in insomnia scores across age groups, evidenced by an F-value of 1.698 and a p-value of 0.149. This lack of statistical significance implies that age may not be as influential on insomnia symptoms as it is on anxiety or depression, suggesting that factors other than age might play a more pivotal role in affecting sleep patterns among the surveyed population. For OCD, the ANOVA results indicate a significant difference, with an F-value of 3.687 and a p-value of 0.00556. Although the statistical significance is lower compared to Anxiety and Depression, it still shows that age groups differ in their OCD symptoms. This statistic highlights that age might influence OCD conditions, albeit less powerful than in Anxiety and Depression.

The Pearson correlation tests provide insights into how listening hours correlate with different mental health conditions. For anxiety, there is a very weak positive correlation (cor = 0.049) with a non-significant p-value (p = 0.1814). This suggests that increased listening hours do not significantly affect anxiety levels. In contrast, moderate positive correlation (cor = 0.110) exists between listening hours and depression, significant at p = 0.002676. This might imply that those who listen to music more frequently could be using it as a coping mechanism for existing depression symptoms or that extensive listening is somewhat associated with higher depression. Furthermore, there is a significant positive correlation (cor = 0.142) between listening hours and insomnia (p = 0.0001132), indicating that those who listen to more music report higher insomnia levels. This could reflect usage patterns where individuals engage with music to combat sleeplessness, which might instead contribute to or exacerbate their symptoms. A moderate positive correlation (*cor* = 0.119) is observed between listening hours and OCD scores, with statistical significance (p = 0.001251). This may suggest a behavioral pattern where higher engagement with music correlates with OCD tendencies, potentially as a part of repetitive behaviors or rituals associated with OCD.

Across all four mental health conditions, the t-tests demonstrate no significant differences between musicians and non-musicians. This suggests that identifying as a musician — whether as an instrumentalist or a composer — does not correlate significantly with higher or lower levels of these specific mental health conditions in this dataset. These findings can be interpreted to suggest that any stereotypes or assumptions about musicians (like the "tortured artist" stereotype) facing different mental health challenges compared to non-musicians are not supported by this data. However, it is essential to consider other factors such as lifestyle, type of music, and the professional environment that might influence mental health but were not controlled for in these analyses.

5. Models (75 pts)

In exploring the relationship between anxiety/depression, age, hours per day spent listening to music, and frequency of listening to classical music, linear regression models were employed to quantify these relationships. The results from these models were further analyzed using diagnostic plots to assess the adequacy and assumptions underlying the linear regression analysis.

The Anxiety model indicates a significant negative association between age and anxiety levels, suggesting that anxiety decreases with age. However, the hours per day spent listening to music and the frequency of listening to classical music did not show significant effects on anxiety levels. The model explains a small portion of the variance in anxiety scores (Adjusted R-squared: 0.02979), highlighting that while age is a significant predictor, many other unaccounted factors likely influence anxiety levels. Figure 11 showed some potential issues such as non-linearity and non-constant variance of residuals, suggesting that the model assumptions may not fully hold. Depression was analyzed similarly through a linear regression model, considering the same set of predictors. Unlike anxiety, both age and hours per day spent listening to music were significant predictors of depression. Age showed a negative association, indicating that depression levels might decrease with age. Interestingly, more hours spent listening to music were associated with higher levels of reported depression. The model's fit was slightly lower than that for anxiety (Adjusted R-squared: 0.02186), but it still captured some of the variability in depression scores. Figure 12 also indicated potential issues with the regression assumptions, like the anxiety model. These regression analyses and the accompanying diagnostic evaluations suggest that while certain relationships between the predictors and mental health conditions were statistically significant, the models' assumptions were not fully satisfied.

In the exploration of data related to various mental health measures—Anxiety, Depression, Insomnia, OCD—and other relevant factors like Age and Hours per day spent listening to music, a k-means clustering model was utilized. This unsupervised model helped to identify inherent groupings or patterns in the data without any pre-assigned labels, offering insights into how respondents might be segmented based on these characteristics. The cluster plot shows the distribution of the dataset into four distinct groups, each represented by different colors and plotted against the first two principal components, which capture the most significant variance in the dataset. The red and yellow clusters show significant overlap, indicating a potential similarity between these two groups in terms of their age, hours spent listening to music, and mental health metrics. This overlap suggests that while these clusters are distinct, they share some common characteristics that might pertain to moderate levels of mental health issues and music listening habits.

Interestingly, there is a small region in the plot where clusters 2 (yellow), 3 (red), and 4 (grey) intersect, but cluster 1 (blue) does not participate in this overlap. This area might represent a unique blend of characteristics from the overlapping clusters, potentially indicative of individuals who exhibit mixed features of these clusters, such as varying degrees of mental health concerns combined with diverse music listening patterns. Cluster 1, which is distinctly separate for the most part, encompasses the largest area compared to the other clusters and includes a very clear outlier. This suggests that cluster 1 is the most heterogeneous group, containing individuals with possibly the lowest average mental health issues and varying ages and music listening habits. The presence of an outlier within this cluster further underscores its variability, highlighting an individual who significantly deviates from the rest in terms of the analyzed variables.

6. Summary of learning (5 pts)

Throughout this project, I delved into the relationships between mental health and attributes such as music listening habits, age, and other demographic information. I faced challenges, particularly in finding meaningful ways to present the data and writing the code required for effective analysis. This required outside research and self-study, improving my analytical and technical skills. One of the most interesting findings from this project was the investigation of the "tortured artist" trope, which suggests that artists (musicians, in this case) are more susceptible to mental health issues. Contrary to popular belief and existing cultural narratives, the statistical analysis did not support this stereotype. This revelation highlights the importance of empirical insights in challenging and potentially dispelling widespread myths. Overall, this project was a comprehensive learning experience that enhanced my ability to independently navigate complex datasets and apply a variety of data analysis techniques to derive meaningful conclusions.

INFO 3010

Spring 2024

Tables and Figures

Table 1: Summary Statistics

```
  Timestamp              Age         Primary.streaming.service Hours.per.day     While.working       Instrumentalist       Composer
 Length:736         Min.   :10.00   Length:736                Min.   : 0.000   Length:736         Length:736         Length:736
 Class :character   1st Qu.:18.00   Class :character          1st Qu.: 2.000   Class :character   Class :character   Class :character
 Mode  :character   Median :21.00   Mode  :character          Median : 3.000   Mode  :character   Mode  :character   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
 Frequency..EDM.    Frequency..Folk.   Frequency..Gospel. Frequency..Hip.hop. Frequency..Jazz.   Frequency..K.pop.  Frequency..Latin.
 Length:736         Length:736         Length:736         Length:736          Length:736         Length:736         Length:736
 Class :character   Class :character   Class :character   Class :character    Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character    Mode  :character   Mode  :character   Mode  :character


 Frequency..Video.game.music.    Anxiety         Depression        Insomnia           OCD         Music.effects       Permissions
 Length:736                   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Length:736         Length:736
 Class :character             1st Qu.: 4.000   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 0.000   Class :character   Class :character
 Mode  :character             Median : 6.000   Median : 5.000   Median : 3.000   Median : 2.000   Mode  :character   Mode  :character
                              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

  Fav.genre          Exploratory        Foreign.languages       BPM          Frequency..Classical. Frequency..Country.
 Length:736         Length:736         Length:736         Min.   :0.00e+00   Length:736            Length:736
 Class :character   Class :character   Class :character   1st Qu.:1.00e+02   Class :character      Class :character
 Mode  :character   Mode  :character   Mode  :character   Median :1.20e+02   Mode  :character      Mode  :character
                                                          Mean   :1.59e+06
                                                          3rd Qu.:1.44e+02
                                                          Max.   :1.00e+09
                                                          NA's   :107
 Frequency..Lofi.   Frequency..Metal.  Frequency..Pop.    Frequency..R.B.    Frequency..Rap.    Frequency..Rock.
 Length:736         Length:736         Length:736         Length:736         Length:736         Length:736
 Class :character   Class :character   Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

Table 2: ANOVA Tests

```
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
----------------------------------------------------------------

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
----------------------------------------------------------------

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
----------------------------------------------------------------

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
----------------------------------------------------------------
```

INFO 3010

Spring 2024

Table 3: Correlation Tests

```
ANXIETY

        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
```

```
DEPRESSION

        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
```

```
INSOMNIA

        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
```

```
OCD

        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
```

INFO 3010

Spring 2024

Table 4: T-tests

```
        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

----------------------------------------------------------------

        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

----------------------------------------------------------------
        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

----------------------------------------------------------------

        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
```

INFO 3010

Spring 2024

Figure 1: Anxiety & Genre Frequency

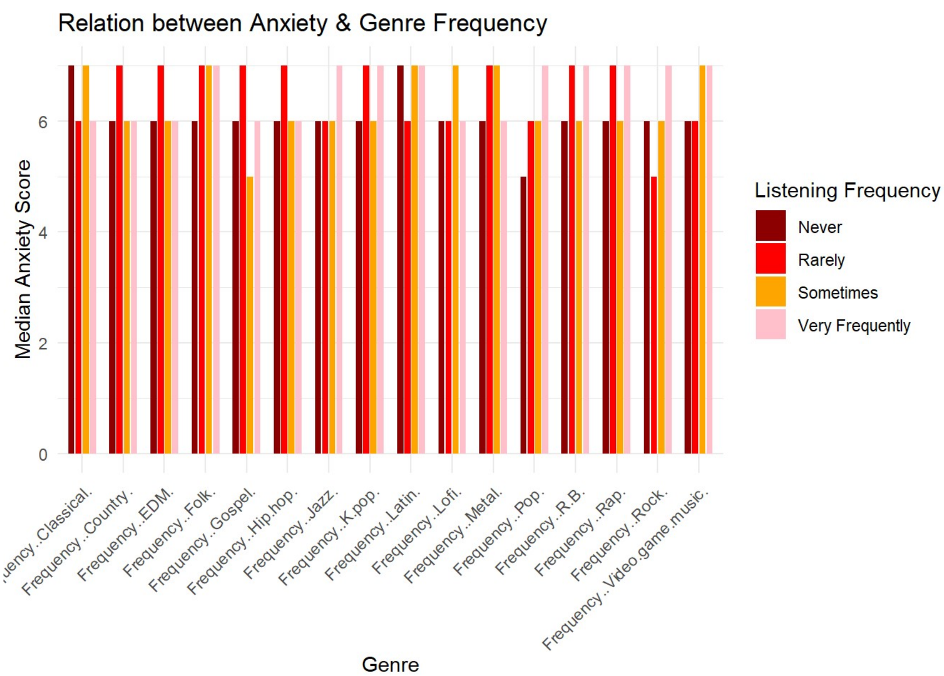
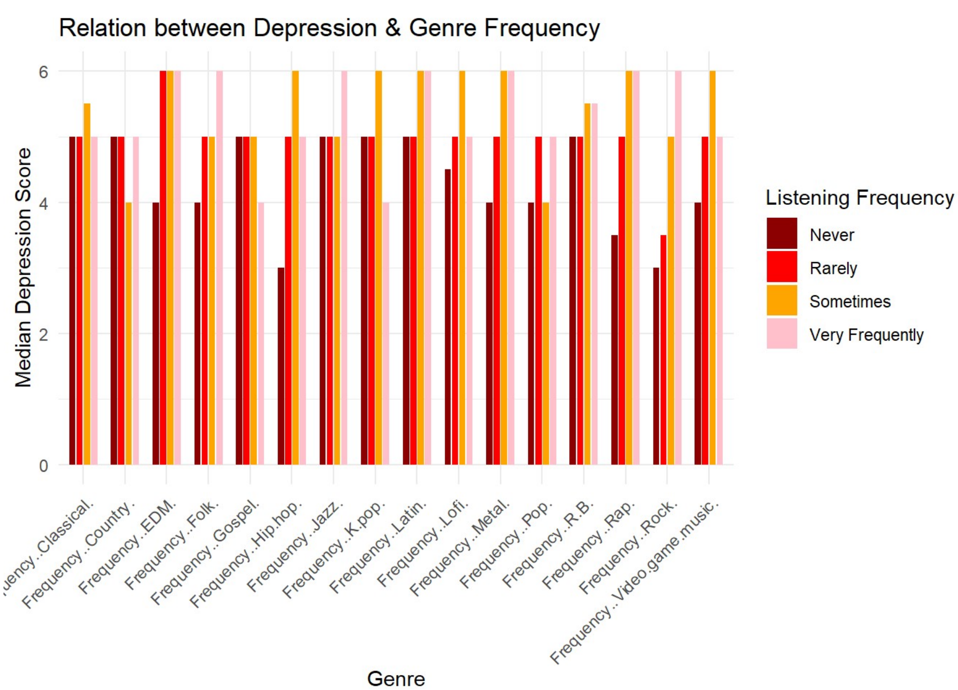


Figure 2: Depression & Genre Frequency



INFO 3010

Spring 2024

Figure 3: Insomnia & Genre Frequency

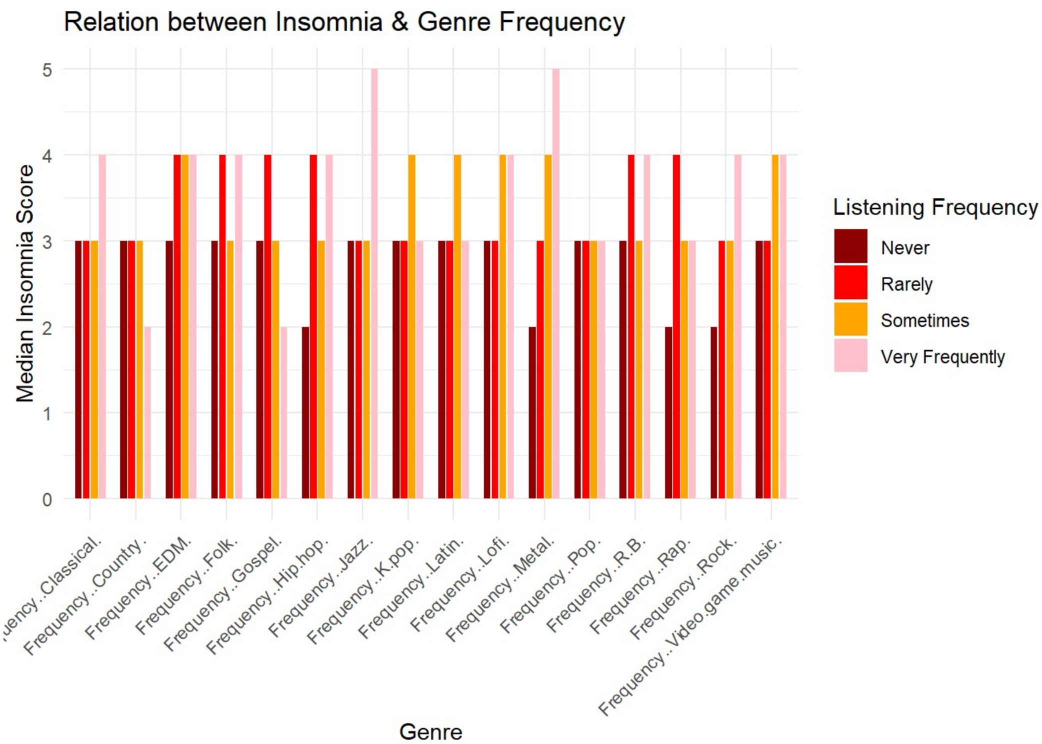


Figure 4: OCD & Genre Frequency

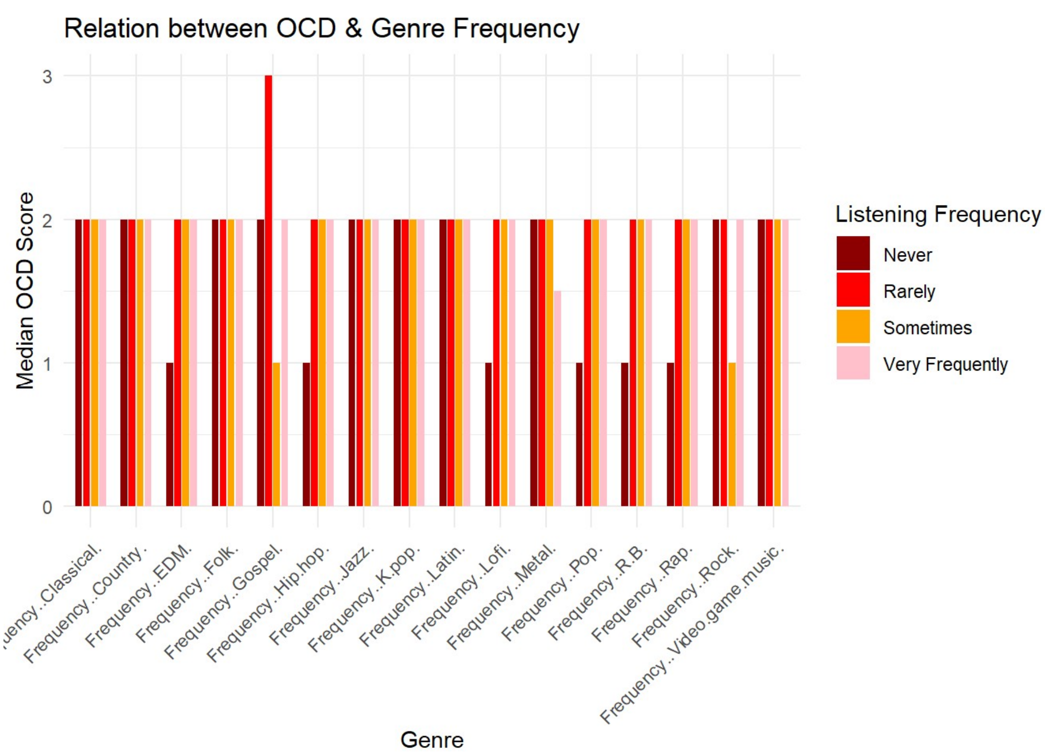


Figure 5: Correlation Table of Genre Frequency and Self-reported Mental Health

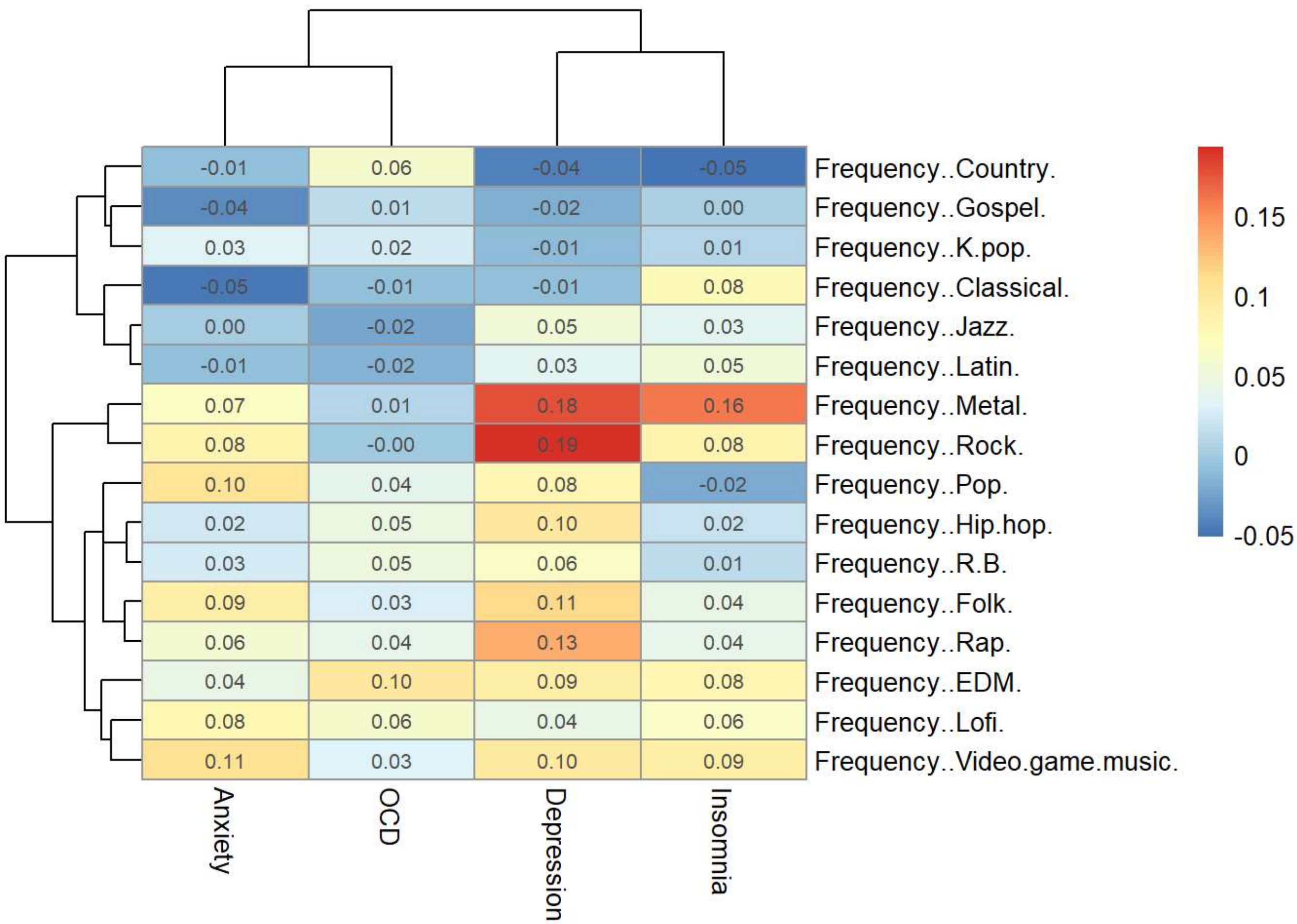
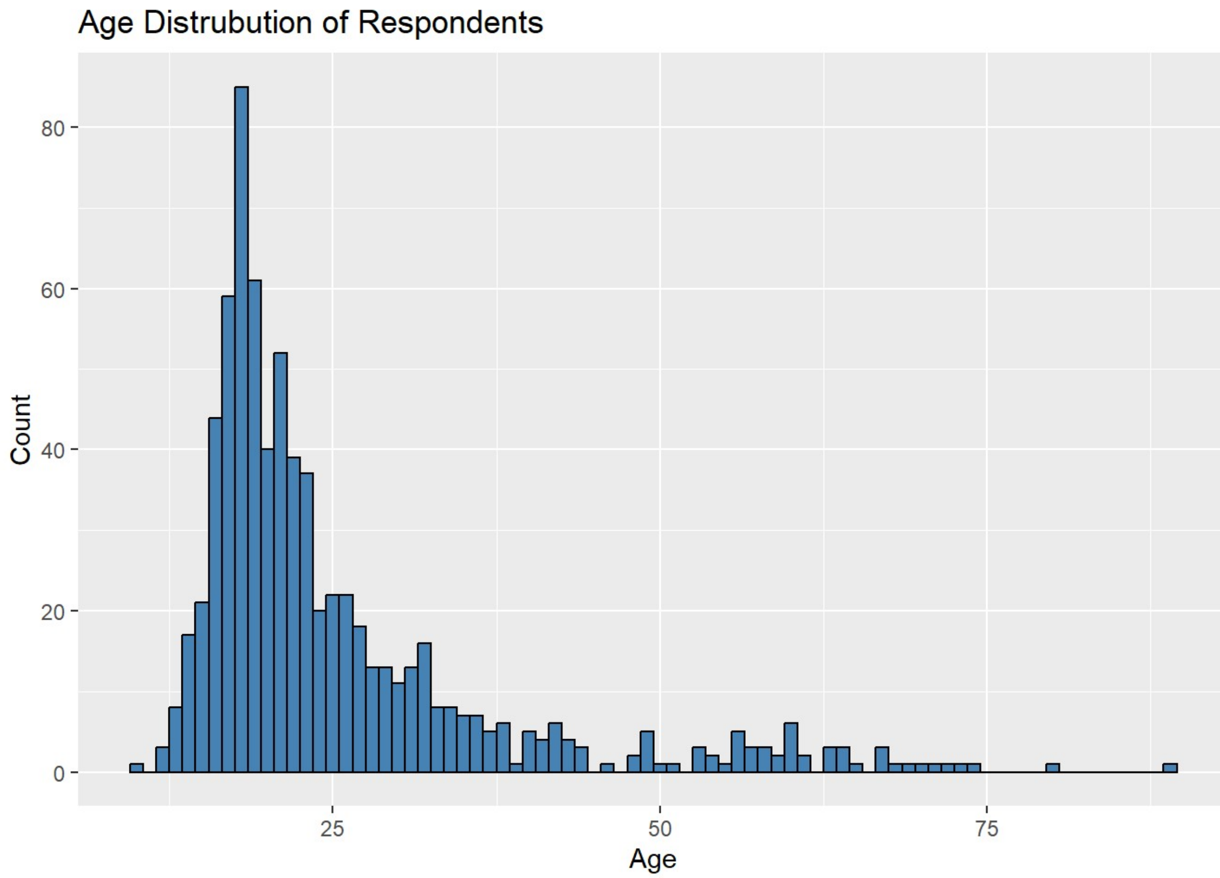


Figure 6: Age Distribution of Respondents



INFO 3010

Spring 2024

Figure 7: Anxiety by Age Group

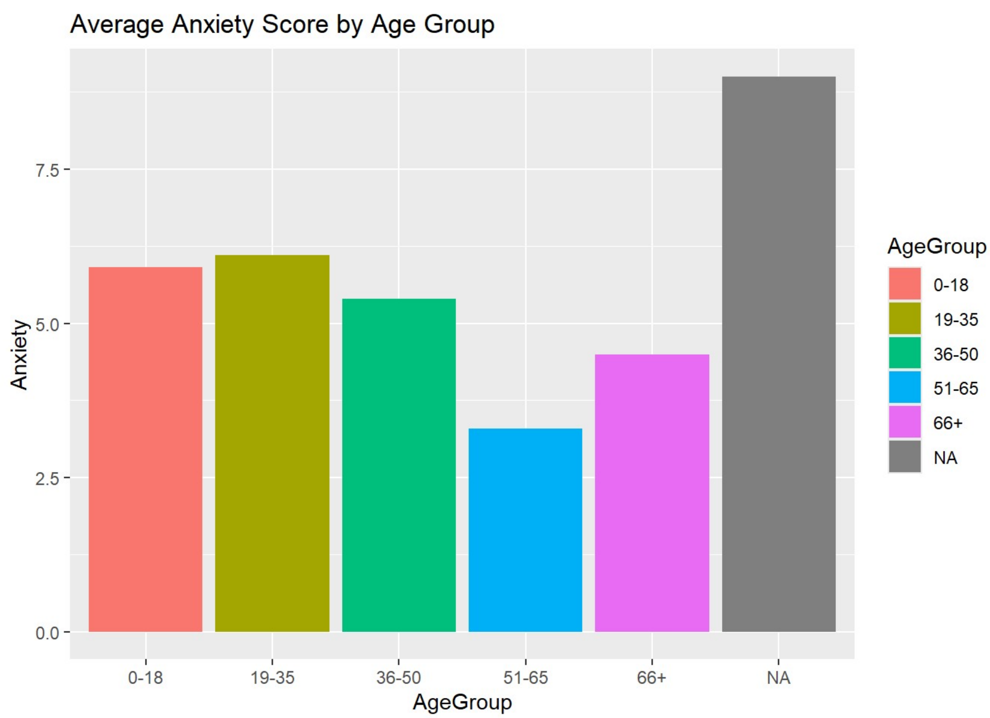
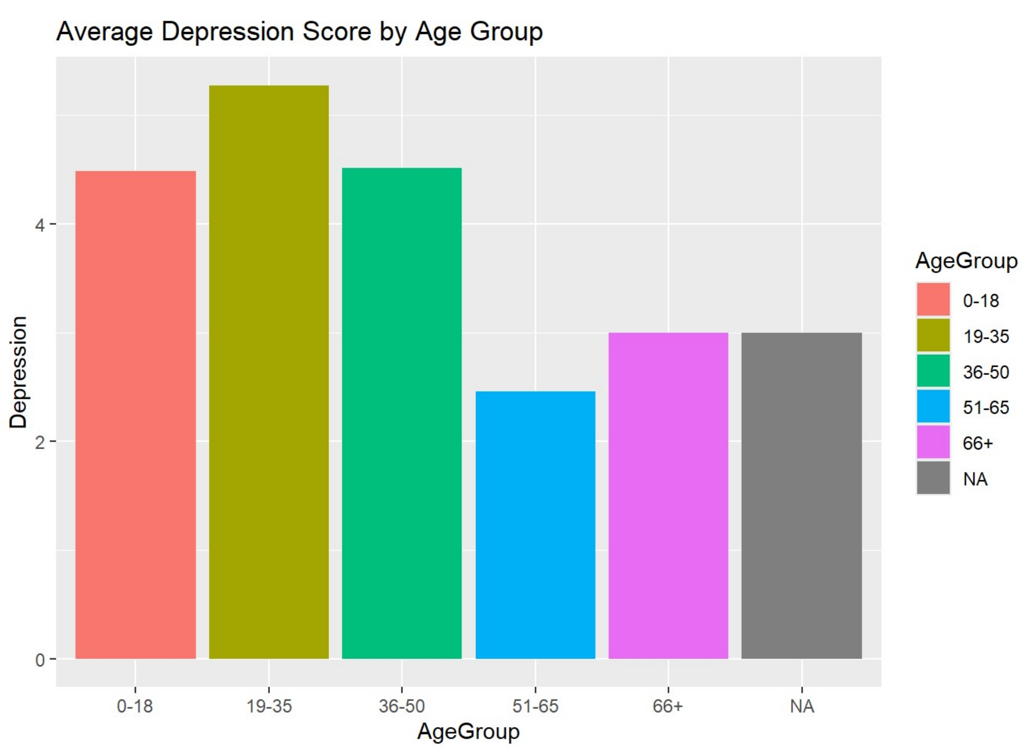


Figure 8: Depression by Age Group



INFO 3010

Spring 2024

Figure 9: Insomnia by Age Group

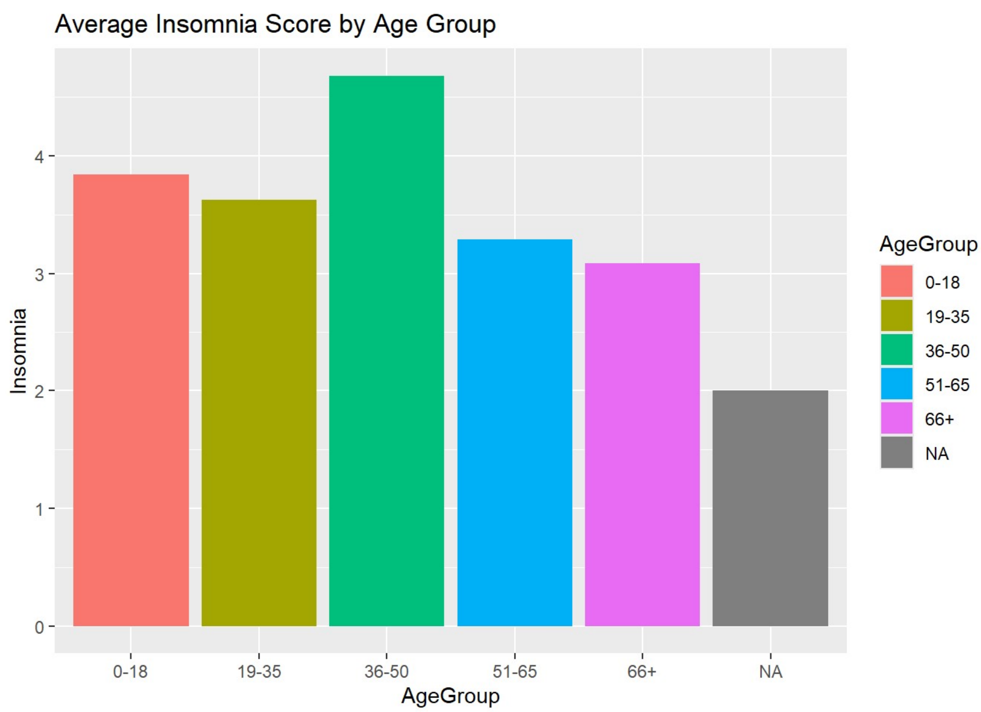
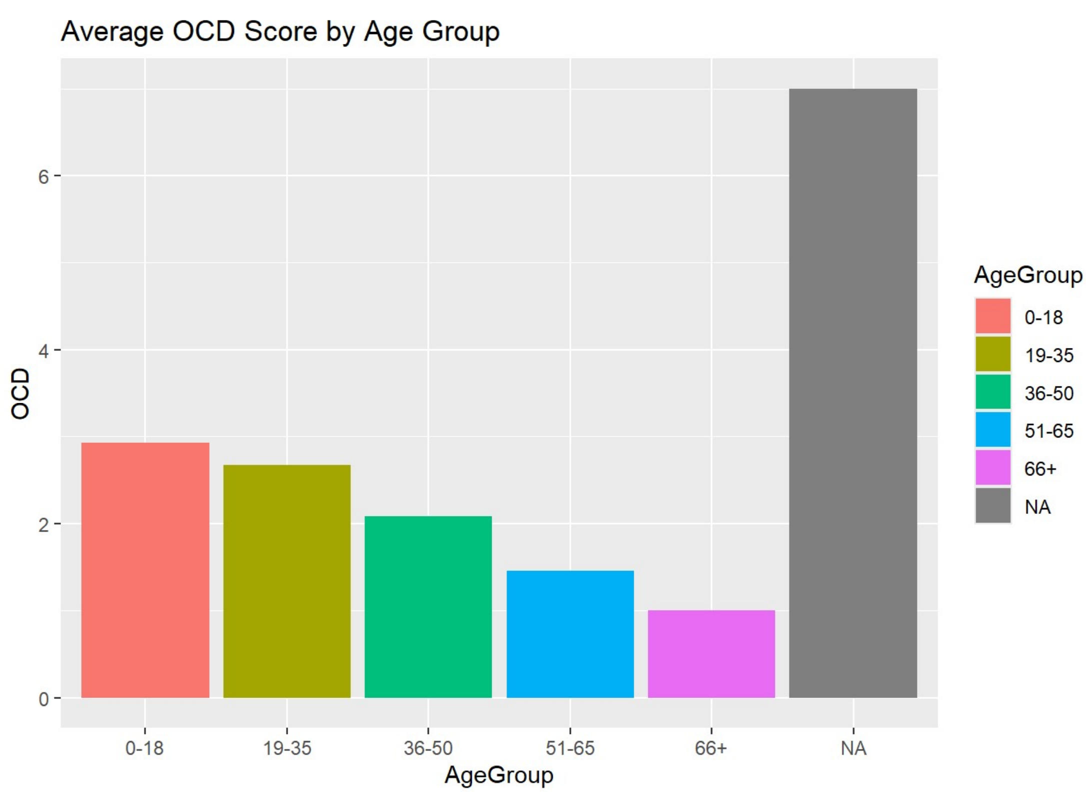


Figure 10: OCD by Age Group



INFO 3010

Spring 2024

Figure 11: Diagnostic Plot of a Multiple Regression (Classical music, Listening Hours, and Anxiety)

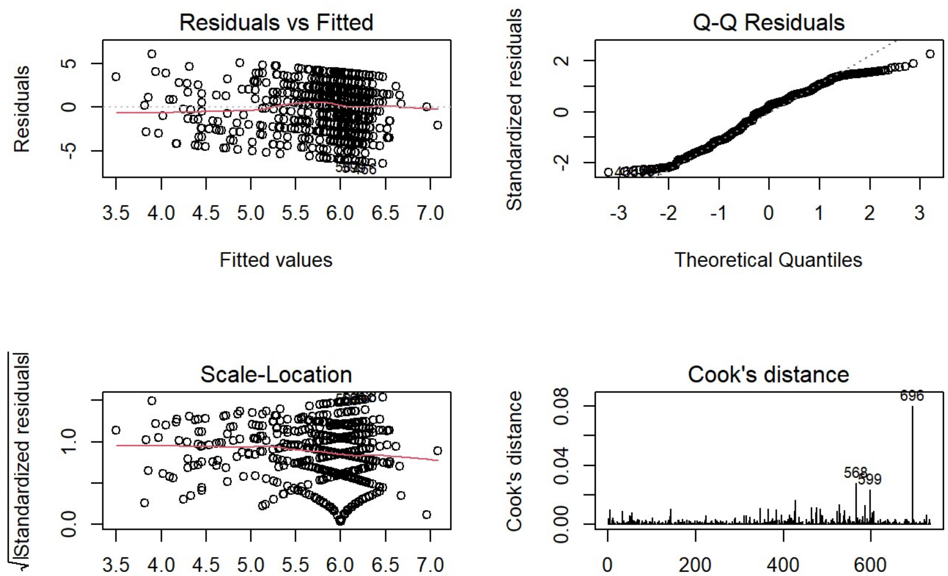
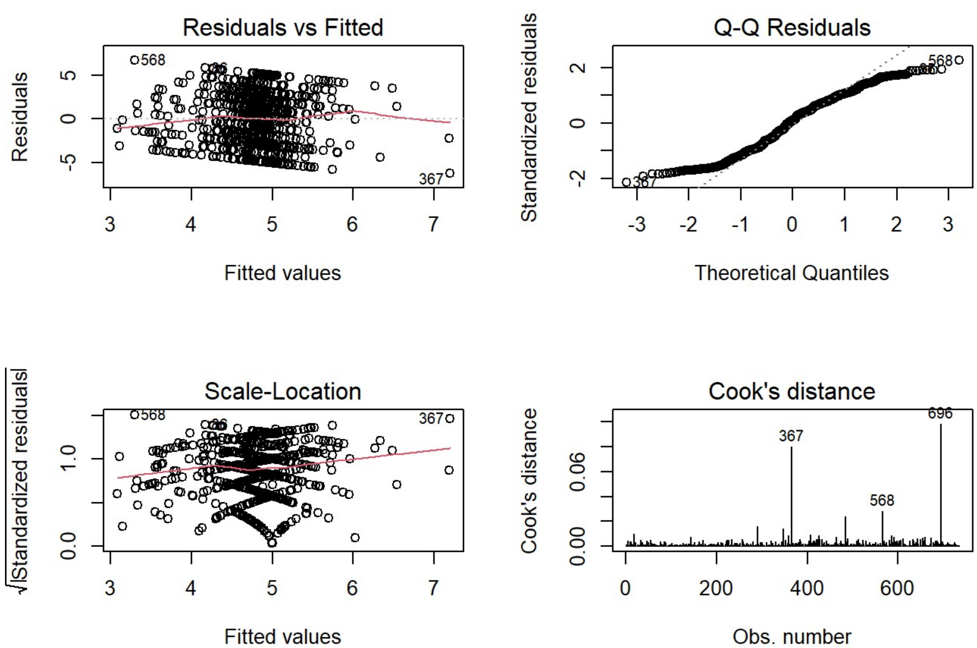


Figure 12: Diagnostic Plot of a Multiple Regression (Classical music, Listening Hours, and Depression)



INFO 3010

Spring 2024

Figure 13: K-means Clustering on Age, Listening Hours, and Self-reported Mental Health