Exploratory Data Analysis Project –- Music and Mental Health

by Sarah Chauvin, Elizabeth Viramontes, and Gabriela Zarate

For our exploratory data analysis project, we wanted to analyze data related to mental health. We think this issue is important and impacts our daily lives, and we all enjoy listening to music. The dataset we used is based on a survey of music listening by genre and correlations with common mental health disorders. It was collected from a survey conducted mostly online, though there were some respondents who answered the questions in person. The creator of the dataset is Catherine Rasgaitis, a computer science student studying music therapy at the University of Washington in 2022. The data is made up of the responses of over 700 people, and it includes 16 music genres as well as four mental health disorders: anxiety, depression, insomnia, and obsessive-compulsive disorder (OCD).

Our initial study data included information we did not use for this particular analysis, including the user's primary streaming service, whether or not the user listened to music while working, whether or not they considered themselves instrumentalists or composers, and if they spoke more than one language. We figured we needed more information to be able to use any of these columns in an objective data analysis. We also could not compare changes in mental health rankings over time since this data only includes a timestamp of when each response was submitted to the survey, so we dropped that column as well.

We dropped rows with data we deemed unreliable, such as the respondent who said he listened to music with "999999999.0" beats per minute (hereafter noted as BPM) and 3 rows that claimed they listened to music 24 hours per day. We also dropped eight rows with null values for music effects and one row with no age listed so we could make sure that we were all using the same data for our various analytics.

The following charts demonstrate an overview of the data used for this project:

Figure 1

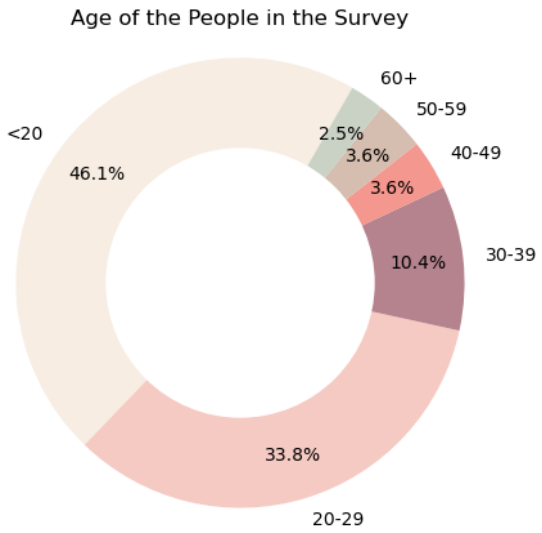
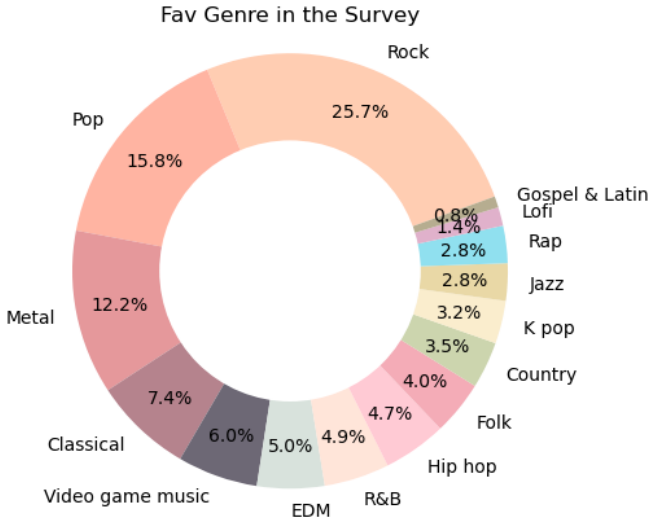


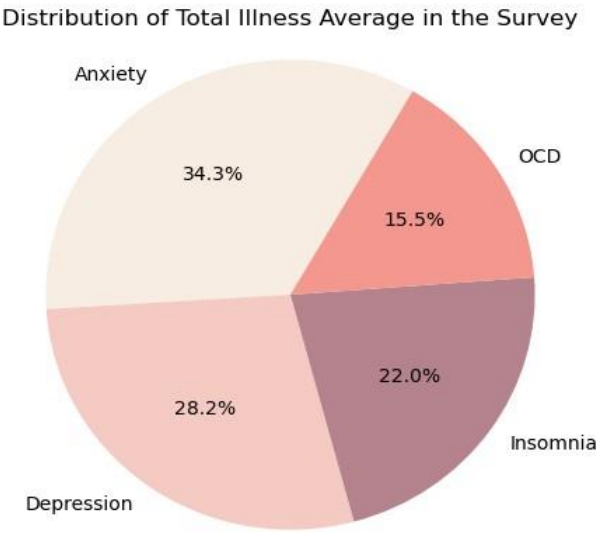
Figure 2



The first donut chart, Figure 1, shows the percentage distribution for different age groups of people who responded to the survey. The majority of respondents are under 30 years old. Note that we looked at groups of 10 years of age, but the survey did not have any respondents under 10 years old, so we categorized everything under 20 together. In Figure 2, we see the breakdown of users' favorite type of music according to their votes in the survey. The majority of the respondents listed rock, pop, and metal as their favorite genres.

To present the mental health disorders studied, Figure 3 is a pie chart showing the percentage of the disorders by average illness rank per person. This visualizes which mental problems were ranked highest or most often in this survey. This graph was not included in our presentation as it was used to give an overall comparison of the distribution of the disorders studied.

Figure 3



As shown in Figure 3, the people who took this survey have the highest percentage of anxiety and the lowest percentage of OCD. We reviewed these discrepancies in more detail as we looked for connections between music and mental health. All of this information enables an easier understanding of the context of the research questions raised during the implementation of this data analysis project.

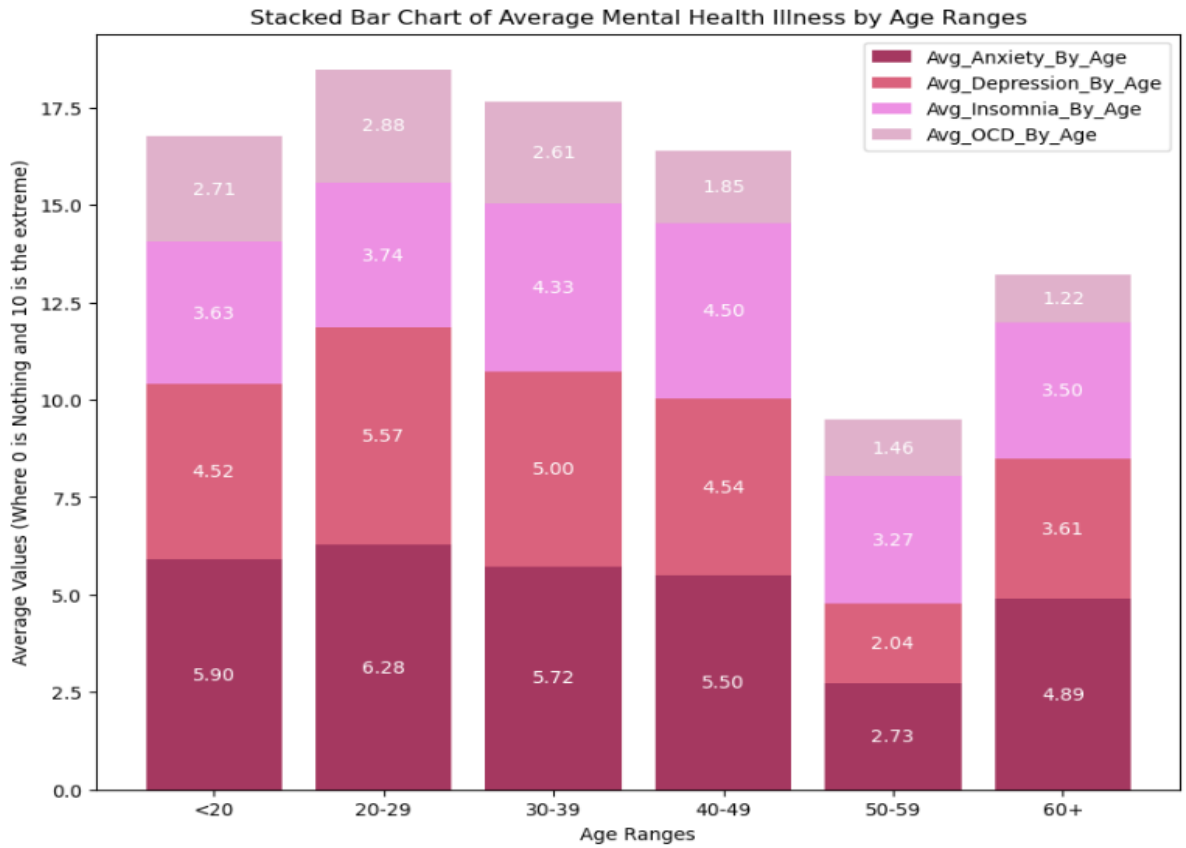
To initially understand the results of this research, we wanted to know if there were significant differences in mental health rankings based on age groups for those who frequently listen to specific genres. To understand this question and the following ones, we must first understand the distribution of diseases by age. To do that, we created the leaderboard in Figure 4 which shows the average disorder rank according to the grouped ages. For the disorder rankings, respondents chose their own numbers on a scale of 0 to 10, where 0 meant "I do not experience this (illness)" and 10 meant "I experience this regularly, constantly, or to an extreme" (Rasgaitis). There was no further information clarifying how these rankings were determined.

Figure 4

Age_Group	Avg_Anxiety_By_Age	Avg_Depression_By_Age	Avg_Insomnia_By_Age	Avg_OCD_By_Age
<20	5.90	4.52	3.63	2.71
20-29	6.28	5.57	3.74	2.88
30-39	5.72	5.00	4.33	2.61
40-49	5.50	4.54	4.50	1.85
50-59	2.73	2.04	3.27	1.46
60+	4.89	3.61	3.50	1.22

By viewing the ages of respondents as groups of 10, it became easier to visualize the information obtained in the survey regarding anxiety, depression, insomnia, and OCD. We observed some relationships between age groups and average levels of each mental disorder, as seen in the stacked bar chart in Figure 5.

Figure 5



Here we can observe average mental illness rank by age range, and we can easily compare the differences between the different mental problems and how their ranks change as users age. According to this data, the highest ranks of mental illness are indicated in users below the age of

40, especially in users in their 20s. Reasons for this could include many changes happening during that period of life, such as work, education, personal relationships, and economic issues. We can see that once users hit 30 years of age, these ranks start decreasing, and the biggest drop is after 50, which could be because people's lives tend to be more stable by that age (Goodwin et al).

Reviewing this information more specifically, we can see that there are sudden changes in each disorder when users reach their 50s. Disorder ranks remain relatively stable at other ages but decrease during this decade. One of the limitations of this research is that the majority of respondents to the survey were under 30 years old, so it is possible that the disorder ranks for those age groups are higher because of the overall number of values included in these averages. We had fewer respondents in their 50s, so their individual responses might be reflected more by the values in the 50s grouping. Interestingly, outside research suggests that people usually start having more issues with insomnia after the age of 50 (Brewster et al), and we do see in increase in insomnia rank from the 50s to the 60+ age group.

While this graph shows low ranks for OCD across the board, it is worth noting that OCD ranks significantly higher in users under 40. This disorder has been studied since 1980, and according to those results, it is mostly diagnosed in people between the ages of 14 and 35 (Zimmerman). Our graph does demonstrate higher OCD ranks for that age range. However, since OCD is not an overly common diagnosis, it is sometimes confused with other mental problems like anxiety and hyperactivity, and it may be seen as a symptom of these rather than a separate disorder (Zimmerman). Because of this ambiguity, we cannot be sure that the information obtained during this survey regarding mental health rankings is totally accurate.

Now that we have reviewed the mental health problems and how they present according to age, we can analyze the following scatterplots which reflect the relationship between ages and diseases with the favorite types of music listed in the survey. The color bar on the right side of the graphs visualizes the range of mental health rankings. The blue color represents a low rank of illness, and as the color becomes pinker, the disease becomes more prevalent. Each of these graphs represent a different mental disorder, as noted in their titles.

Figure 6a

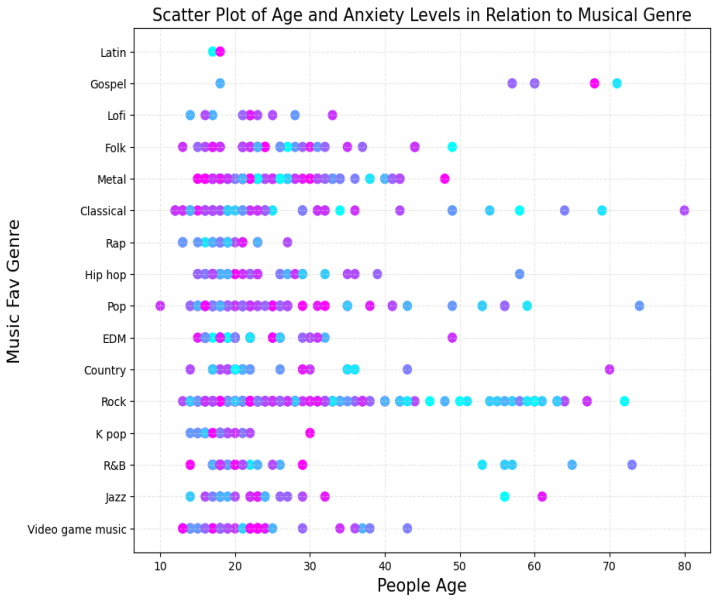


Figure 6b

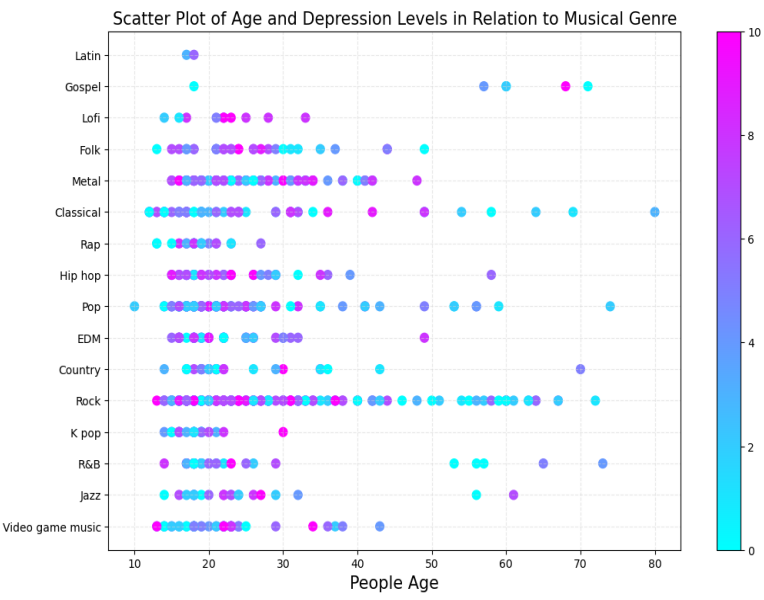


Figure 6c

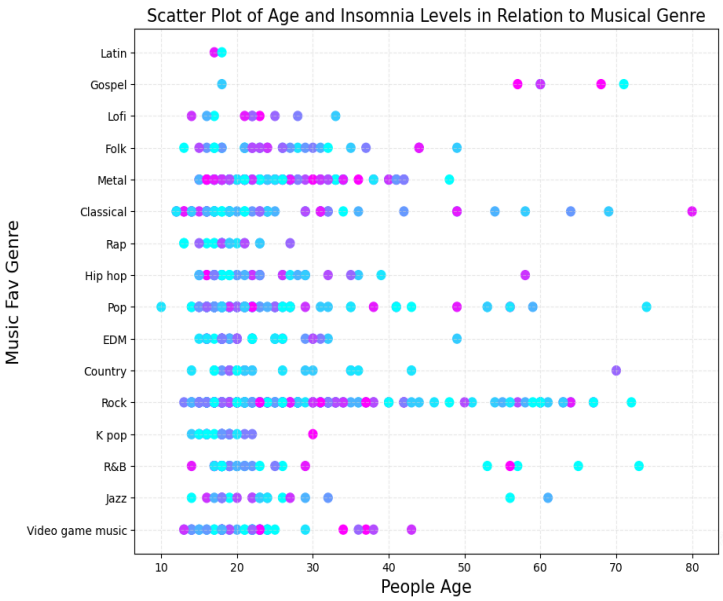
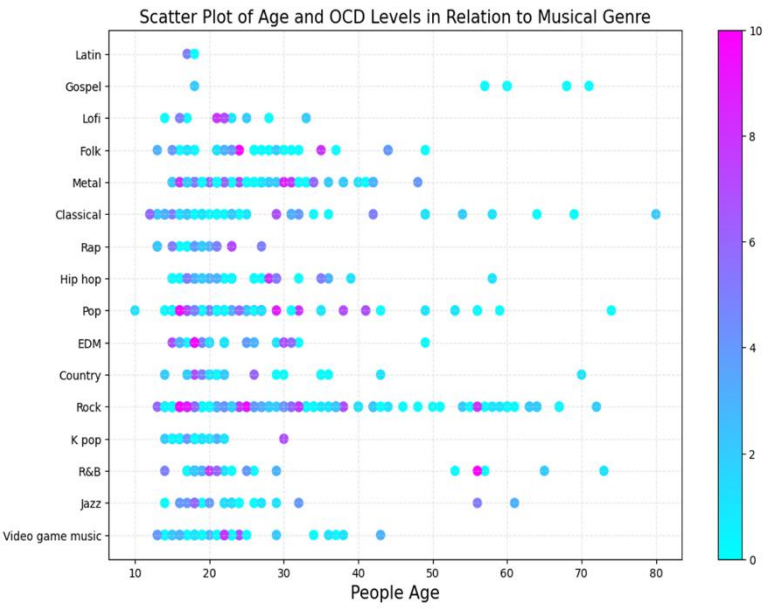


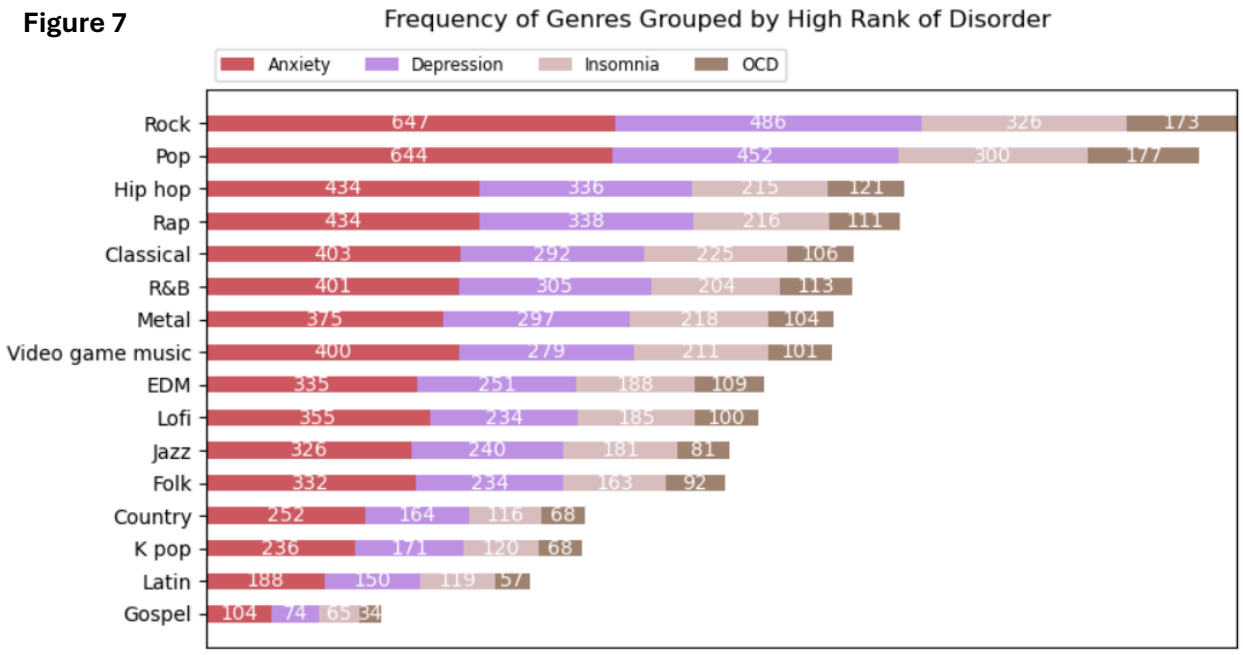
Figure 6d



With the Anxiety graph (Figure 6a), the pink colors are easy to observe in metal, pop, rock, and video game music, meaning that listeners of those genres score their disorder states higher compared to listeners of other types of music. The same observation can be made with depression and insomnia (Figures 6b and 6c). For these three diseases, disorder rank is higher in ages under 40, and the older respondents have lower disease ranks. This could be due to personal tastes in types of music, and it is possible that metal, pop, rock, and video game music are more attractive genres to those with poor mental health. But with the OCD graph (Figure 6d), most of the mental health rankings remain quite low, which could indicate either that music does not have much effect on this problem, or it has a positive effect, depending on the perspective of each respondent. We do observe some high disorder rank outliers in folk, metal, pop, electronic dance music (EDM), and rock, but we attribute these outliers to specific individuals as we found no connection between them.

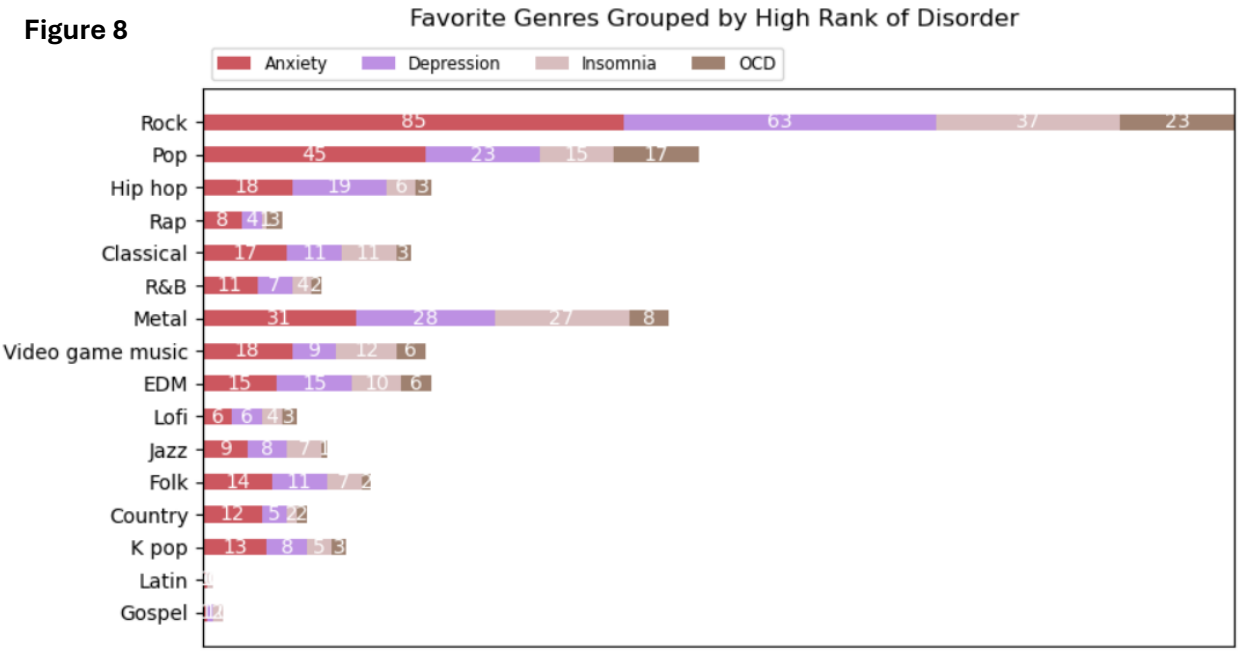
After breaking down our data by age group and individual disorder, we wanted to see if there was a correlation between users claiming a high rank of each disorder and the genres they listened to. In the initial dataset, the frequency columns contained string values to describe how often the respondent listened to each genre. We replaced those values with integers, with “Never” as 0, “Rarely” as 1, “Sometimes” as 2, and “Very frequently” as 3. For this research question, our main data frame was separated into four different data frames with one for each genre where the data was filtered by mental health rankings of 7/10 or higher. For each disorder data frame, we summed the integer values for each genre’s frequency in order to compare them objectively. Figure 7 shows that distribution:

Figure 7



In these same four dataframes, we also looked at the genres each respondent listed as their favorite. Interestingly, the results are somewhat similar, but actually vary quite a bit with those of the previous graph.

Figure 8



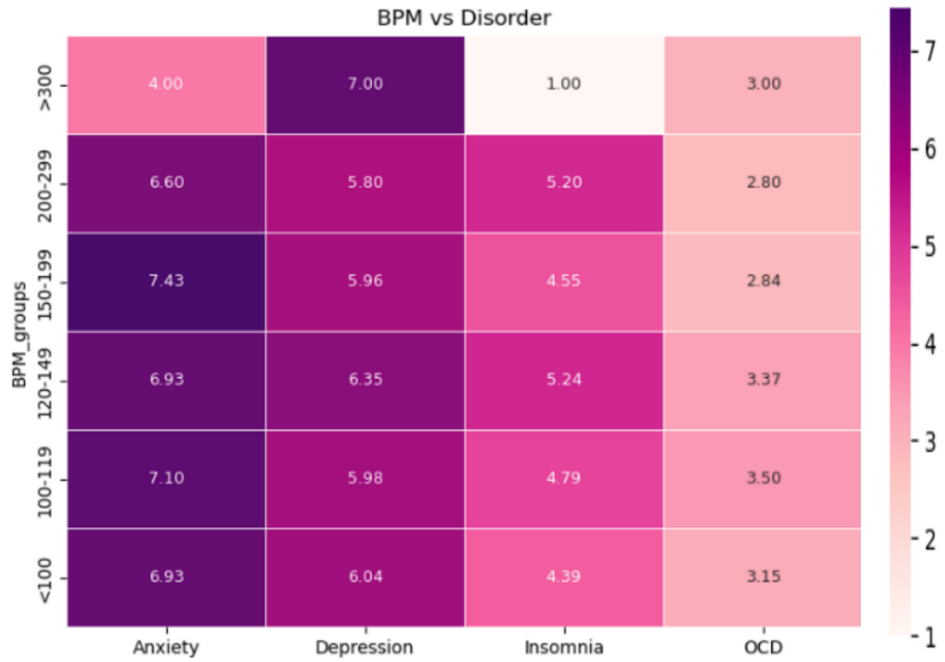
According to these values, the three genres most listed as favorites among users with high disorder rankings are rock, pop, and metal. But the frequency each of these is listened to does not exactly match, as metal is listened to 61% as much as rock, which aligns with favoritism, but metal is listened to 63% as much as pop, despite being just under pop in favoritism. This could indicate that listeners of metal are more likely to report having poor mental health, though we do not have

enough information here to determine if listening to metal music actually *causes* poor mental health.

It is worth noting that the numbers represented in each bar of the frequency graph do contain some repeat respondents, as many of them reported higher rankings for more than one disorder. Also, the numbers represent an aggregate of the data since each frequency value was changed to an integer and summed. There are not exactly 1632 different people in this data who listen to rock music frequently. Rather, 1632 is the total of aggregated responses to the frequency of listening to rock music among users with disorder rankings of 7 or higher. This specific data encompasses the responses of 417 users, so the frequency graph shows that many of these users listen to multiple types of music, and many of them also have high ranks for multiple mental health disorders.

Since the top favorite genres among our respondents with high mental disorder ranks are rock, pop, and metal, we thought there might be a relationship between high BPM and disorder rank. To evaluate this, we created groups of BPM counts for the initial data frame with null values for BPM dropped. We used the average disorder rank for each BPM group for each disorder to create the following heatmap:

Figure 9

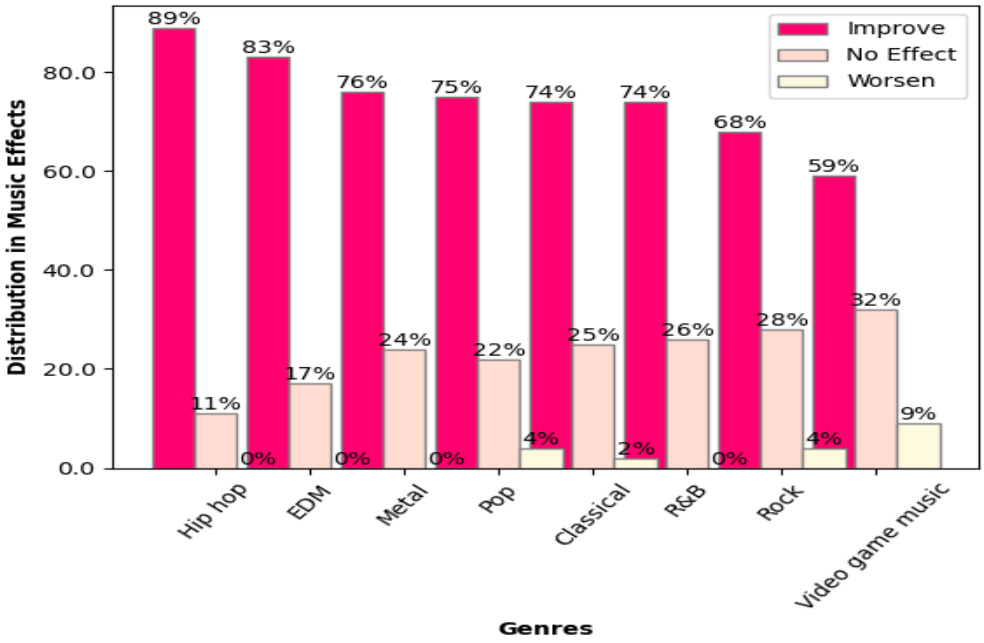


Disorder rank is indicated by the color bar. Though users could have ranked their illness levels as high as ten, there were no average values above 7.5 when all respondents who indicated a BPM value were included. This heatmap clearly shows that disorder rank is similar for almost all users within each individual disorder, despite their average BPM grouping. Since increasing the average BPM does not necessarily increase the disorder rank, we cannot say that BPM really makes much difference to disorder rank. We can glean more information from looking at the favorite

genres and frequencies of listening for each of those when looking for trends in mental health status.

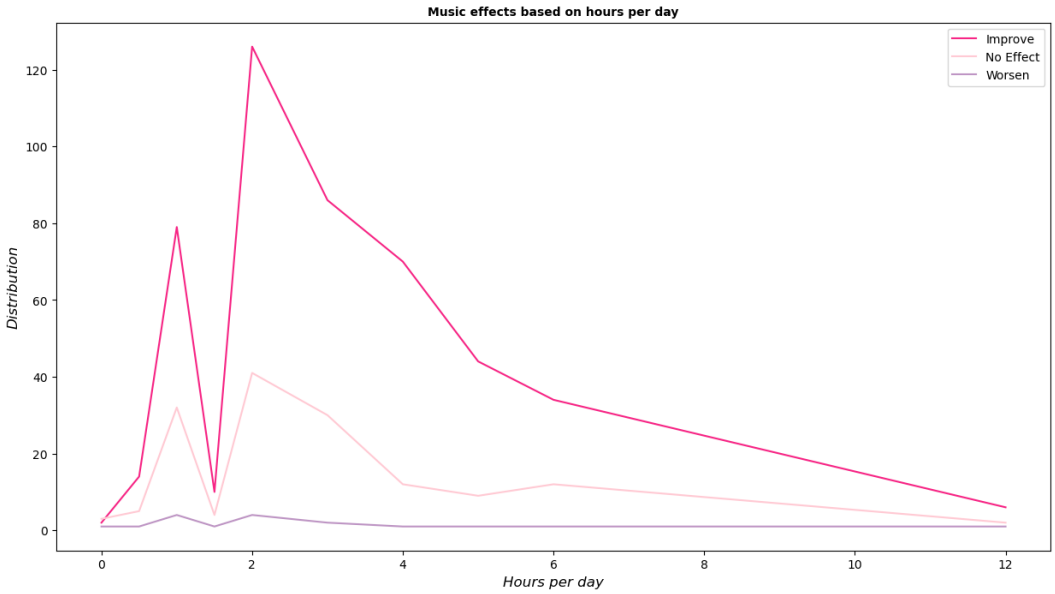
After analyzing the results for the listeners' favorite genres, we wanted to explore how they each perceived these genres' effects on their mental health. The survey asked each user to choose "improve," "worsen," or "no effect" for music effects overall. We created a new dataframe with hours per day, favorite genre, and music effects, grouped by favorite genres, and we looked at the value counts of how many votes each genre received for each effect. To observe the effects of the genres commonly listed as favorites, we filtered the data by genres that received at least 26 votes as favorites. The percentage of each effect per genre ensures a clearer view of the results in Figure 10. This process allowed us to focus on the genres that had a significant impact, like hip-hop, EDM, and pop, and drop Latin, gospel, and jazz, which received lower favoritism. By analyzing Figure 10, we can conclude that the genres with the most favorable results are hip-hop, EDM, and metal. While R&B, rock, and video game music received many positive votes for effect, they also received many no-effect rankings due to the large number of respondents listing these genres as favorites.

Figure 10



Continuing with our analysis, we wanted to determine whether there was any correlation between the time spent listening to music daily and its impact on mental health. Our initial hypothesis was that there is a positive correlation between the hours spent daily and positive effects of music. However, we discovered that the fewer hours spent listening to music, the higher the rate of improvement, as long as music was listened to for at least one hour per day. Above 6 hours per day, no evident correlation exists between the effect on mental health and the time listened. The highest rate of improvement was registered by users who spent 2-3 hours per day listening to music. These findings, as shown in Figure 11, are free from significant outliers, further reinforcing the validity of our research.

Figure 11



Finally, after analyzing all the information from the survey with the objective of making a correlation between two variables in the process, we realized that the correlations are weak with almost all of this data since the results are very subjective. We decided to select two different examples in order to present two different cases of correlation, as seen in Figure 12.

Figure 12a

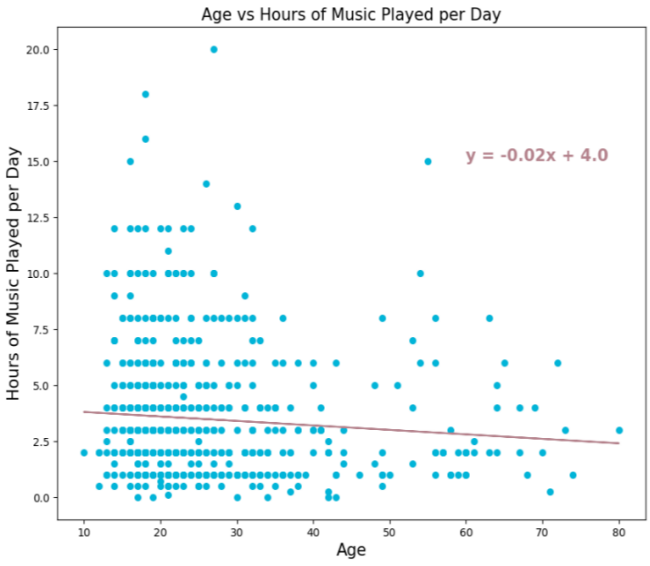
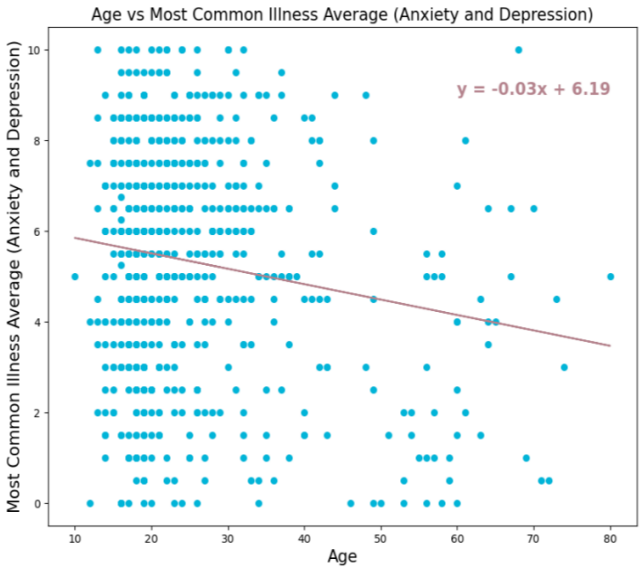


Figure 12b



The first regression is very weak, which means that there is little linear correlation between age and the number of hours per day users listened to music. We interpreted this to say that personal taste of each person matters more than age when observing how listening time affects mental health. Also, note that we do not know during which part of their day people are listening to music; they

may listen to more hours of music for pleasure during free time or while working due to the nature of their job.

In Figure 12b, we can see a slightly stronger negative correlation between illness rank and age. As we saw in Figure 5, the rates of anxiety and depression decrease as we age the most drastically of the disorders in our dataset, so these two disorders offer the clearest visualization for observing the relationship between mental health and age. As mentioned previously, many of our respondents do not fit this regression well, likely due to the nature of how the survey was conducted.

Based on the different analyses of our data and many possible correlations, we concluded that listening to music is likely to improve mental health, as seen in Figure 13, and we recommend listening for at least two to three hours per day. According to this dataset, the genres that offer the most improvement are hip hop, EDM, metal, pop, and classical. As the leaderboard in Figure 14 shows, the improvement results were most noticeable in listeners 20 years old and younger. However, it is important to mention that we cannot say that music was significantly less effective for listeners older than 20 due to how the information was collected. Since the majority of survey responses were collected via the internet, we do not have as many responses from older people, and more of those could alter the total impact music has on those age groups.

Figure 13

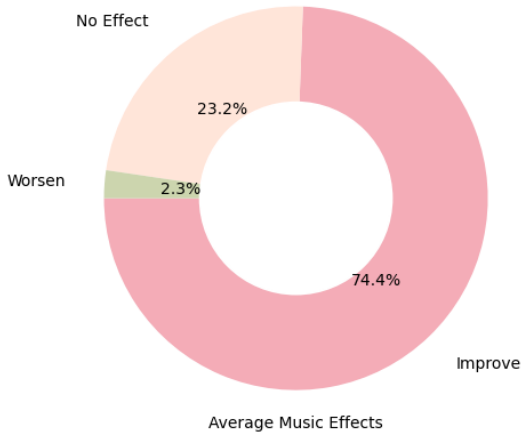


Figure 14

Age_Group	Improve	No Effect	Worsen
<20	443	125	12
20-39	45	26	5
40-49	19	7	0
50-59	18	8	0
60+	16	3	0

This dataset had quite a few limitations, and we developed some suggestions for improvements for future study. Foremost, we cannot judge mental health improvement with listening to music over time from this data. The same users reporting in over a set time period would offer more capability of establishing relationships with both frequency of genre usage and hours listened per day. Second, the mental health data was very subjective. Users were allowed to choose their rank for each disorder from 0 to 10, so it follows that each user might interpret their numbers differently. If these results were concentrated over a specific geographic location, for instance, the United States, we could compare their music statistics from the survey with actual published mental health statistics for more objective data.

The BPM section was not specific at all. It was optional, so about 100 users did not record a value, and we do not know if each user interpreted this number to be an average, the most common, or the highest BPM they listened to. Also, BPM is not something common music listeners are necessarily aware of, so many of these numbers could have been estimates. This is probably why we could not correlate anything with BPM. Additionally, the original data asked if users listened to music during work or not, but it did not ask for occupation or what other times during the day they listen. Clarifying this might make that part of the data more significant. We would also like to see gender of the respondent included, as we would have liked to have been able to compare how frequencies of listening to certain genres affect men and women differently.

Lastly, we have no basis for the results in the music effects section. Was the reported effect from listening to certain genres or listening for certain periods of time? Many of our respondents listed high ranks for multiple disorders, so does their reported effect apply to a certain disorder or multiples? It might have been useful to ask for ranks for each disorder over time, or after increasing or decreasing the amount of time the user listened to music each day. If we ran this survey in the future, we would clarify the questions to minimize these limitations and hopefully collect more objective data.

Resources and Works Cited

Brewster, Glenna S, et al. “Insomnia in the Older Adult.” *Sleep Medicine Clinics*, U.S. National Library of Medicine, Mar. 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC5847293/.

edX Boot Camps LLC. “Xpert Learning Assistant.” 2022, bootcampspot.com

Goodwin, Renee D, et al. “Trends in Anxiety among Adults in the United States, 2008-2018: Rapid Increases among Young Adults.” *Journal of Psychiatric Research*, U.S. National Library of Medicine, Nov. 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7441973/.

Hunter, John, et al. “Examples#.” *Examples - Matplotlib 3.9.0 Documentation*, Matplotlib Development Team, 2012, matplotlib.org/stable/gallery/index.html.

Rasgaitis, Catherine. “Music & Mental Health Eda.” Kaggle, Kaggle, 18 Dec. 2022, www.kaggle.com/code/catherinerasgaitis/music-mental-health-eda.

Zimmerman, Mark. “Obsessive-Compulsive Disorder (OCD) - Psychiatric Disorders.” *Merck Manual Professional Edition*, Sept. 2023, www.merckmanuals.com/professional/psychiatric-disorders/obsessive-compulsive-and related-disorders/obsessive-compulsive-disorder-ocd.

Special thanks to Professor Alexander Booth and teaching assistant Mike Wenner for helping us find our dataset and complete our coding.

Image Appendix