

Do Mental Disorders Influence the Type of Music Listened To?

The question of whether mental disorders influence music preferences is indeed broad and complex.

Mental disorders such as PTSD, depression, borderline personality disorder, and others have always captured my attention. Initially perceived as burdens and deep flaws in one's personality, mental disorders are more common than we might think and are often inherently part of what makes human beings so paradoxically beautiful. Many of these disorders stem from heightened sensitivity or exposure to difficult life events and traumas, leading to a unique sensibility and perspective. Music, being a reflection of our innermost feelings and tastes, serves as an excellent medium to explore these differences.

Mental Disorders and Music Popularity: First Exploratory Data Analysis

Kaggle Dataset

In this first part, the aim is to explore broad trends extracted from a useful dataset found on Kaggle, titled "Twitter Mental Disorder Tweets and Musics Dataset," which attempts to link mental disorders with music preferences.

Using the Kaggle API, I extracted two datasets: the "control" and "disorder" datasets. These datasets link each user and their disorder (either PTSD, depression, anxiety, panic disorder, borderline personality disorder, bipolar disorder, or no disorder) to the titles and artists of the music they mentioned.

An important question arose from the outset: When the dataset's author extracted the music information, what did the "music tweet" say exactly? The dataset's construction and its implications are somewhat unclear. We don't know the context in which the music was mentioned. For this analysis, I assumed that a person would naturally tweet more about the music they like since they listen to it more frequently. Thus, I inferred that users liked the music they mentioned, allowing me to build models of "popularity" based on these mentions.

The Question of Popularity and Disorder Proportions

Without direct popularity information linked to each music piece, the only way to measure broad trends was to count the number of times each artist was mentioned for different disorders.

However, a basic bar plot (Figure 1) raises an important question: Is the proportion of users from each disorder equal in the overall dataset? This question is crucial since we want to know the proportion of users affected by a specific mental disorder for each artist. A similar but more accurate plot (Figure 2) shows that the proportion of unique users per disorder is indeed unequal, which could lead to misinterpretations.

The plot in Figure 4, which presents a podium for each disorder independently, doesn't face this issue because it doesn't compare disorders directly. One consideration is that multiple mentions of the same artist by the same user were counted equally, which might be questionable. Should a user's obsession with an artist count significantly more than a general liking?

In Figure 3, I chose not to modify the dataset, aiming to show the most significant differences between the top popular artists among the general ("control") population and a population

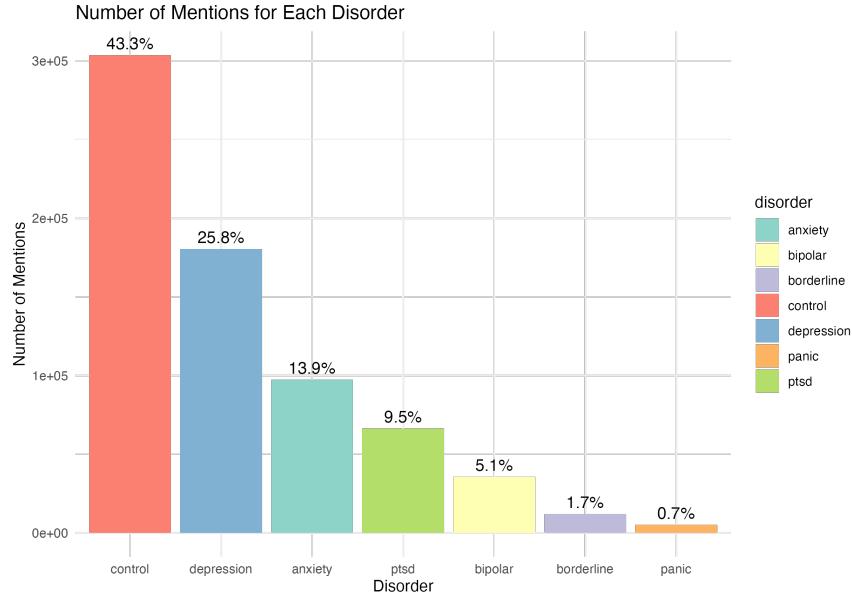


Figure 1

comprising people with mental disorders. This helps illustrate that even when considering all disorders together (and their natural proportions within this group), their music preferences differ from the control population.

Artist Popularity Showing Disorder Correlation

This quick analysis inevitably leads to considering the similarity between disorders. For the plots in this part, I emphasized re-scaling the dataset counts of disorders for each artist. I weighted the counts by dividing by the mention percentages of each disorder, leading to the correlation plot in Figure 5.

This plot reveals that some disorders tend to be more correlated than others, and overall, there is a correlation with the control population. A positive correlation indicates that higher counts of one disorder are associated with higher counts of another, while a negative correlation suggests that higher counts of one disorder are associated with lower counts of another.

We can observe that anxiety is highly correlated with borderline, control, and depression, and to a lesser extent with panic and PTSD. The control group shows lower correlation with bipolar disorder and PTSD compared to other disorders. Overall, bipolar disorder is the least correlated with the other disorders.

Figure 6 also followed the weighting process and shows how the three most popular artists in the global dataset influence each disorder population differently.

Mental Disorders and Music Features: Second Exploratory Data Analysis

After exploring genre preferences and the distribution of the most popular artists among different disorders, I was curious whether mental disorders influence music taste in terms of features such as energy, danceability, key, tempo, or speechiness.

I built the dataset by fetching features of each song using the Spotify API. I began by plotting the correlations between features over the entire dataset to get an initial idea of how features might have similar patterns, which could help identify those specific to one disorder (see Figure 7).

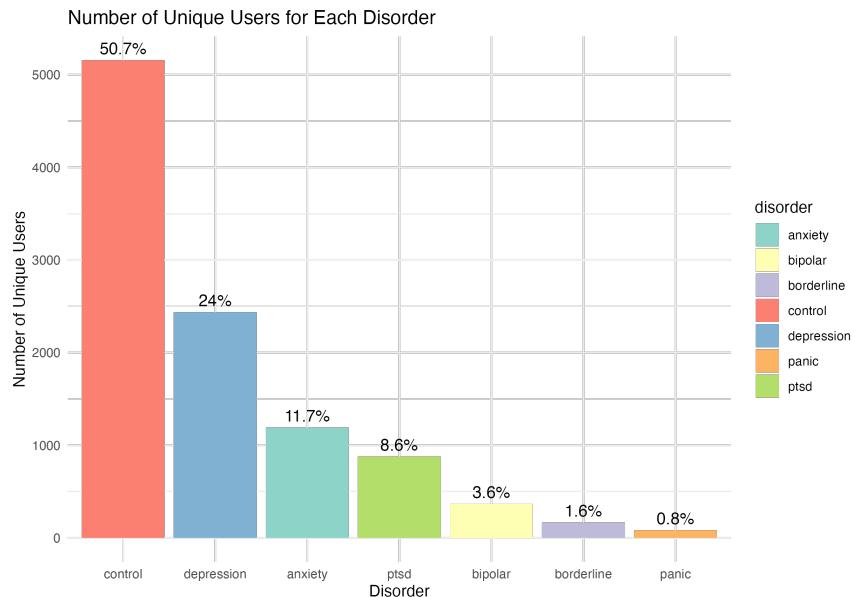


Figure 2

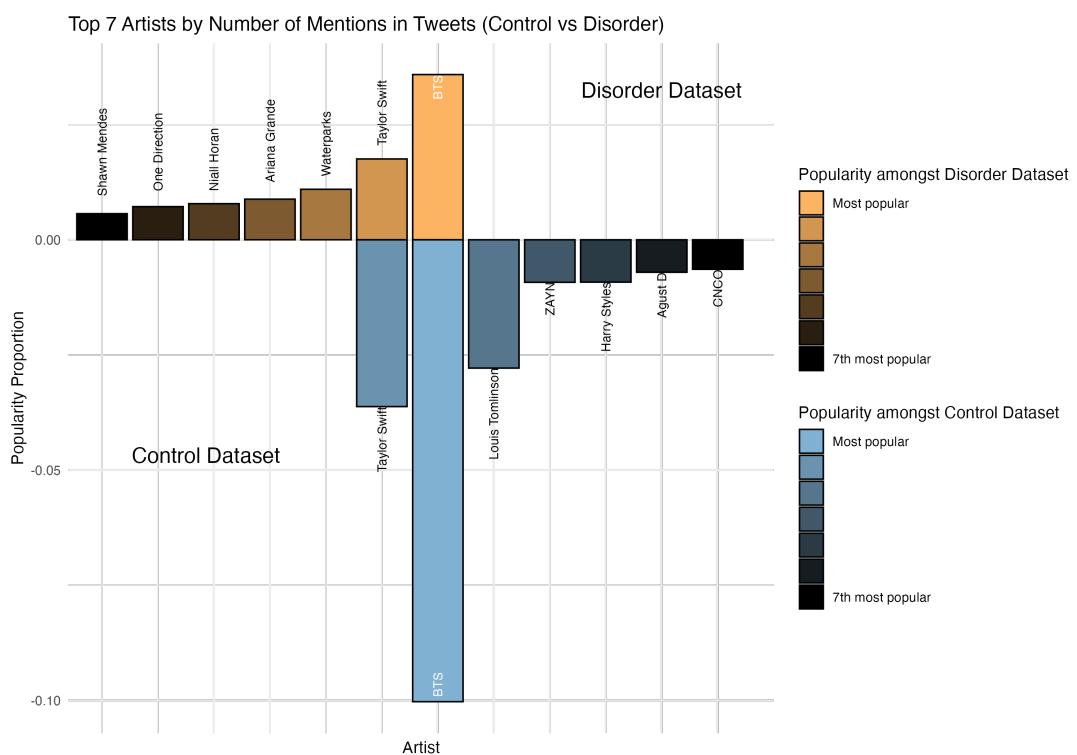


Figure 3

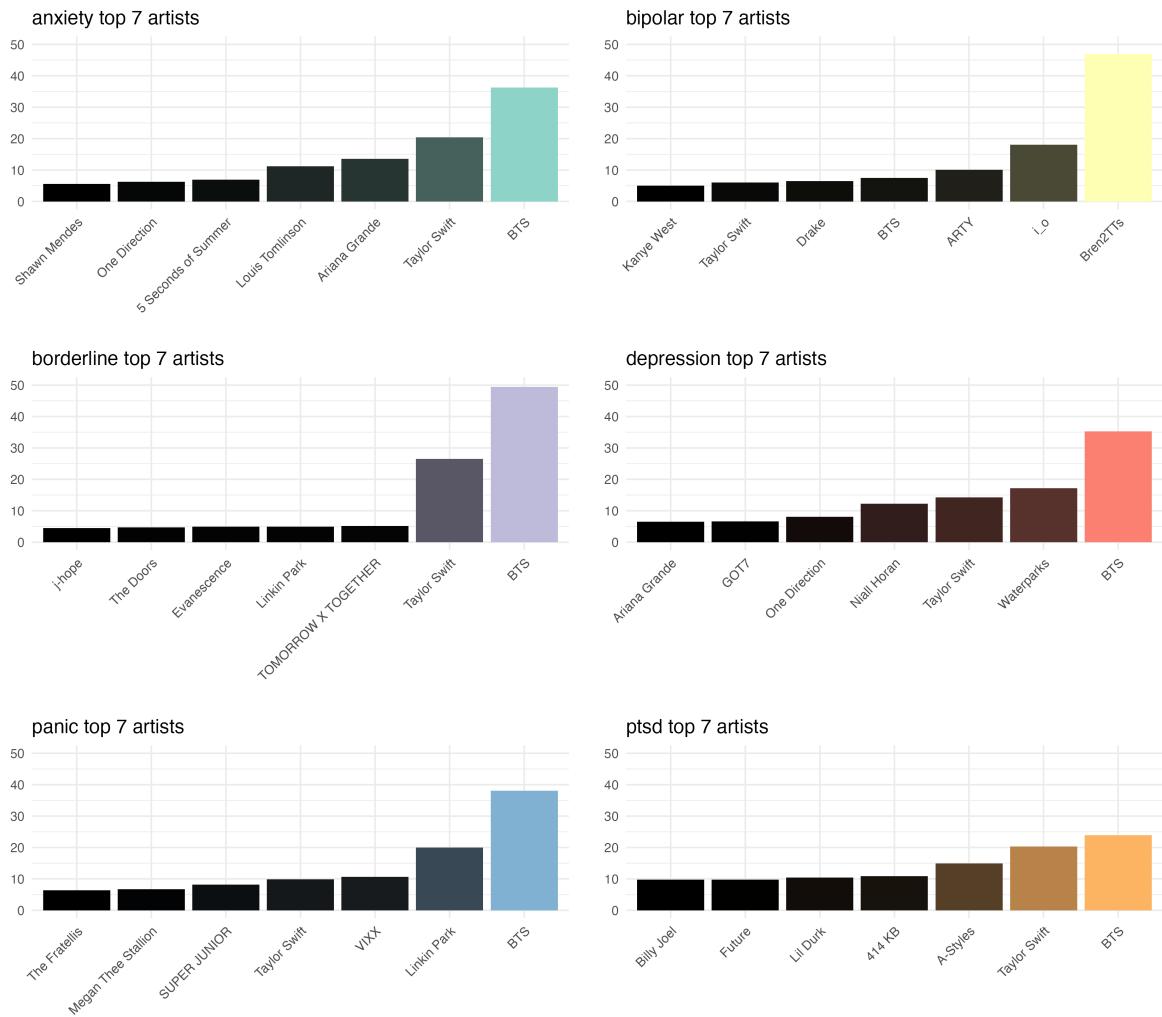


Figure 4

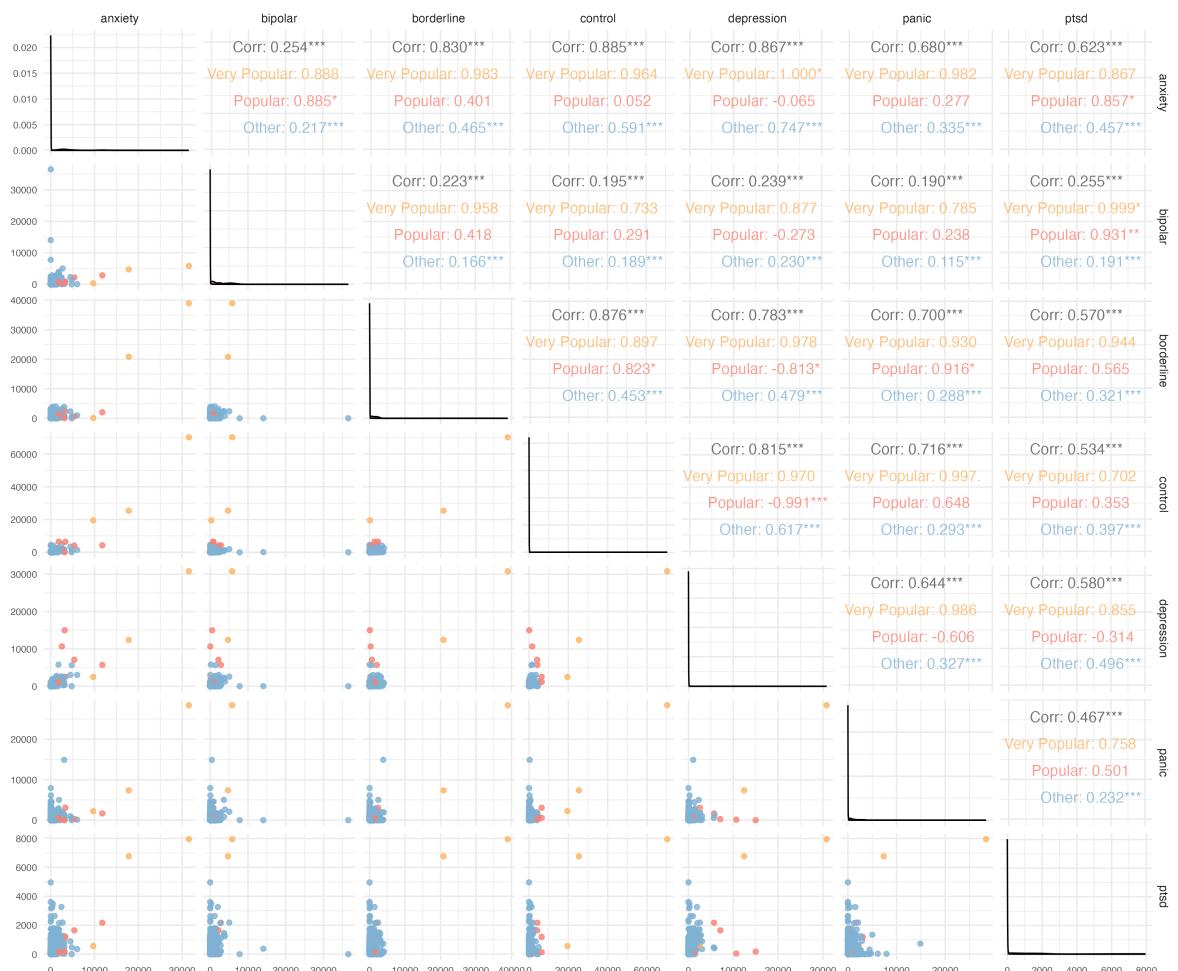


Figure 5

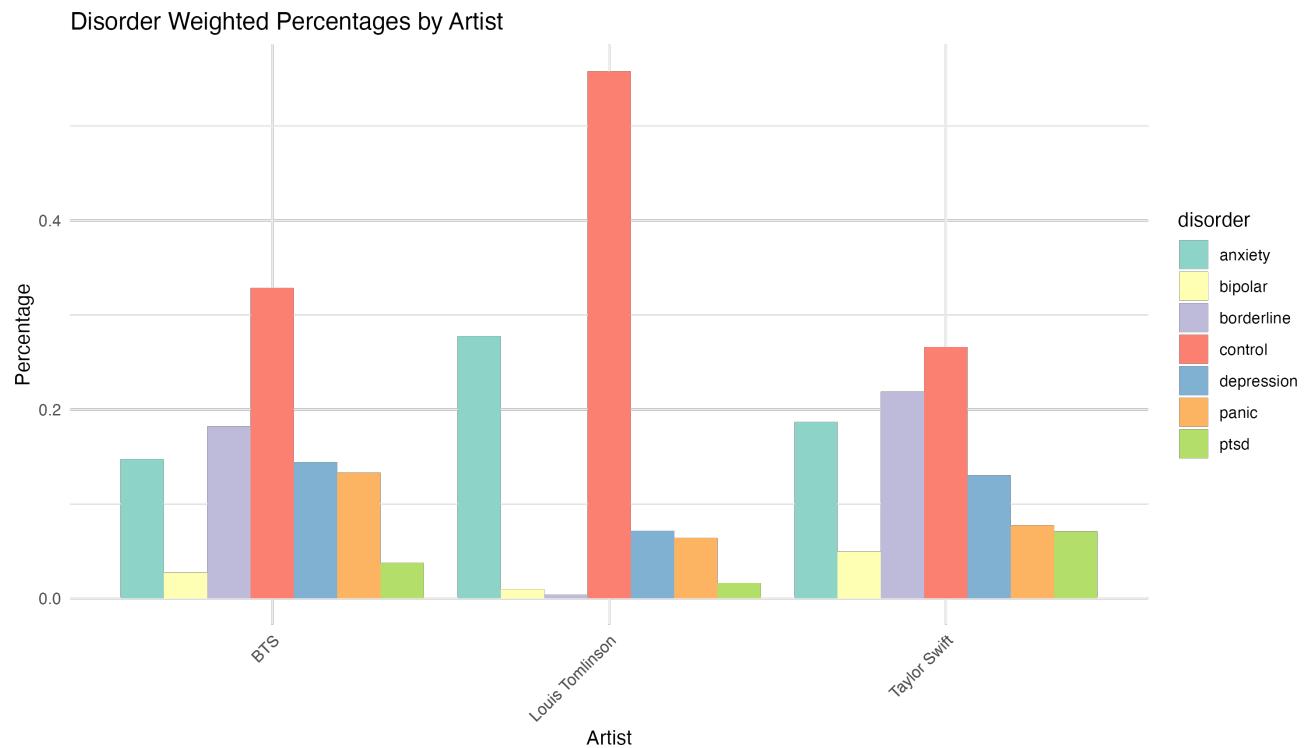


Figure 6

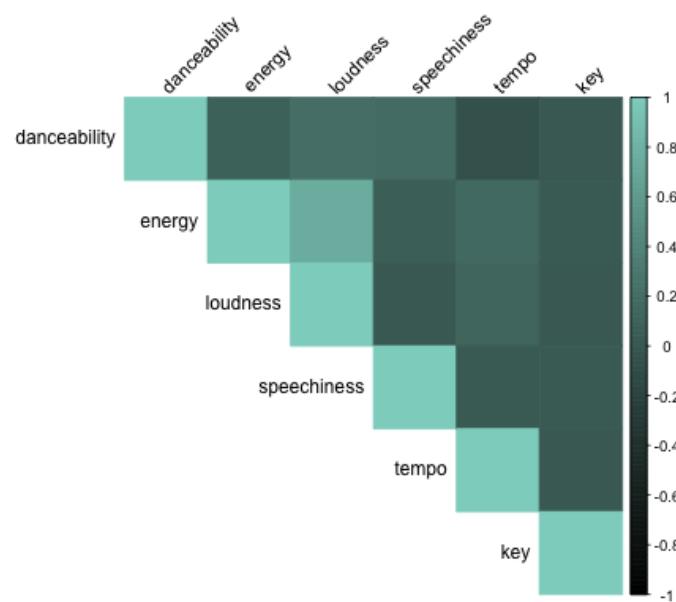


Figure 7

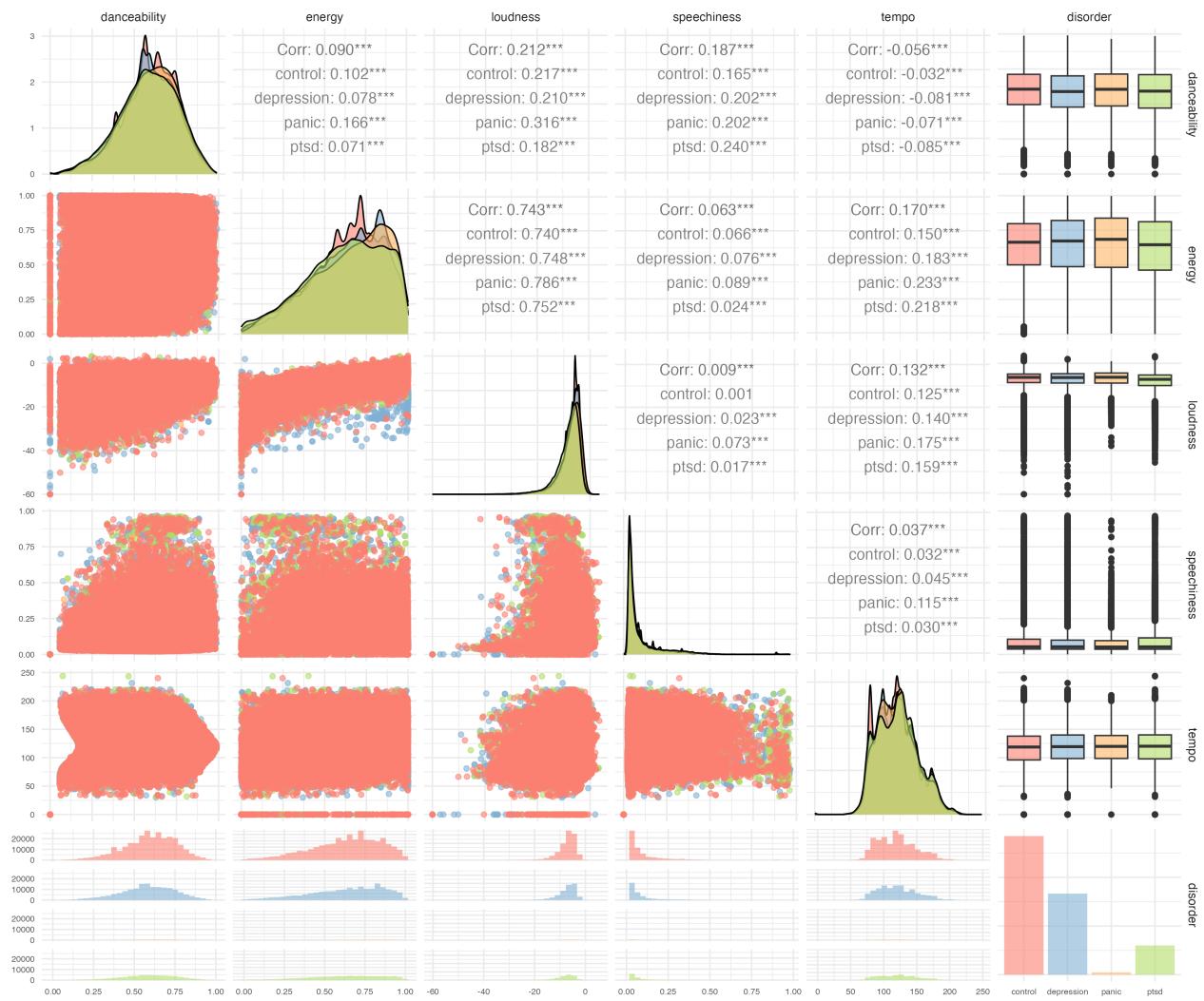


Figure 8

This correlation plot shows that loudness and energy are highly correlated, as well as loudness and danceability, speechiness and danceability, tempo and energy, and tempo and loudness. These correlations are generally what we would expect.

Overall Patterns

To visualize the main patterns, pair plots are useful. I plotted two of them (for the first three disorders and the next three, both including control points for reference) - see Figures 8 and 9.

Although the overall resemblance between the control points and the others might be disappointing, some particularities caught my attention. I provided some boxplots, density plots, and violin plots of specific features to see where the disorders might have an effect:

Energy Patterns

The energy plots (Figures 10 and 11) revealed an interesting pattern: the depression disorder shows a high density region around 0.8, which the control dataset lacks. Panic and anxiety disorders also show this tendency to some extent. While I am not deeply versed in the subject, one symptom of depression is feeling deeply anxious or worried. Energy in music measures intensity and activity, and it makes sense that high energy would correlate with anxiety and

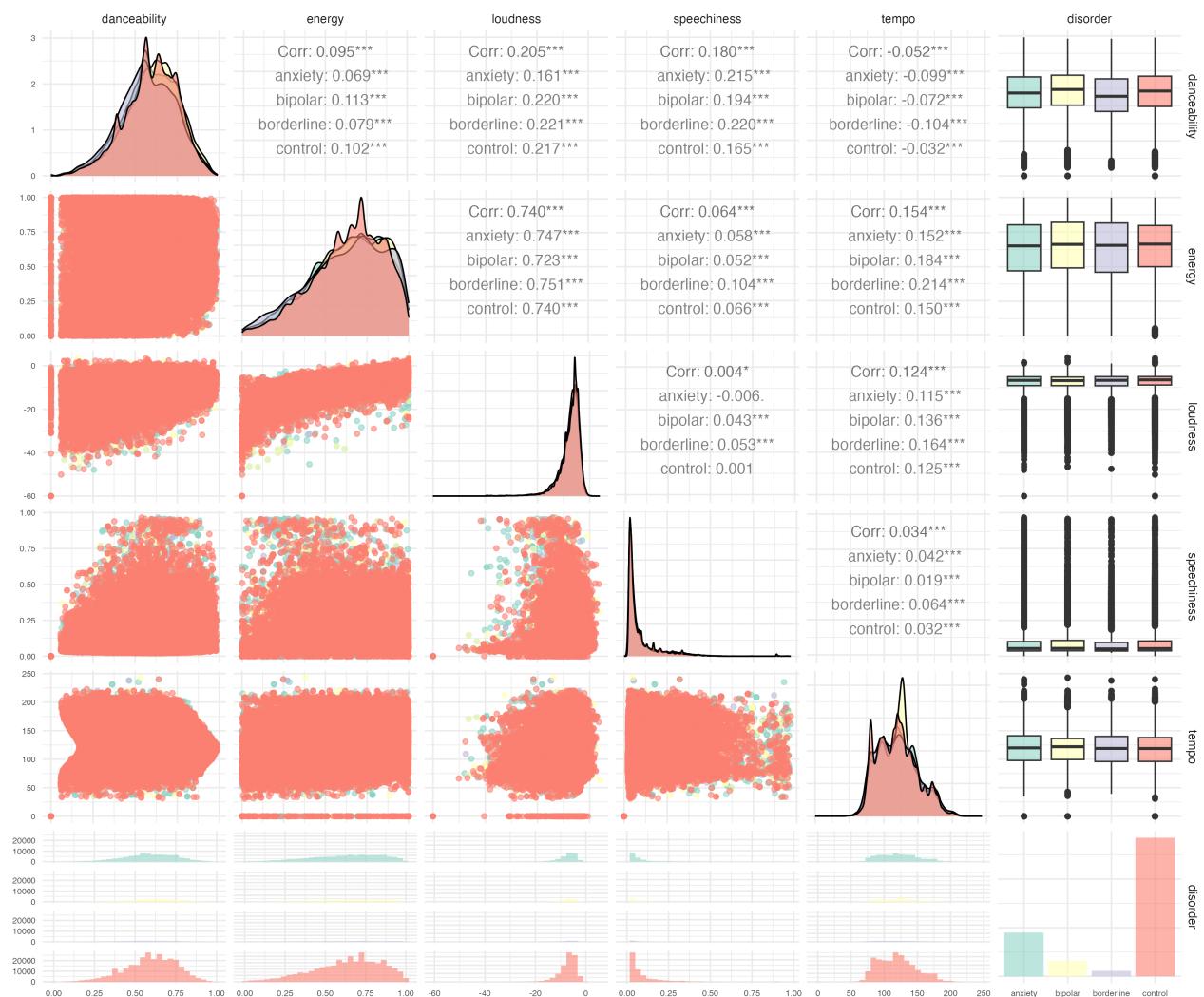


Figure 9

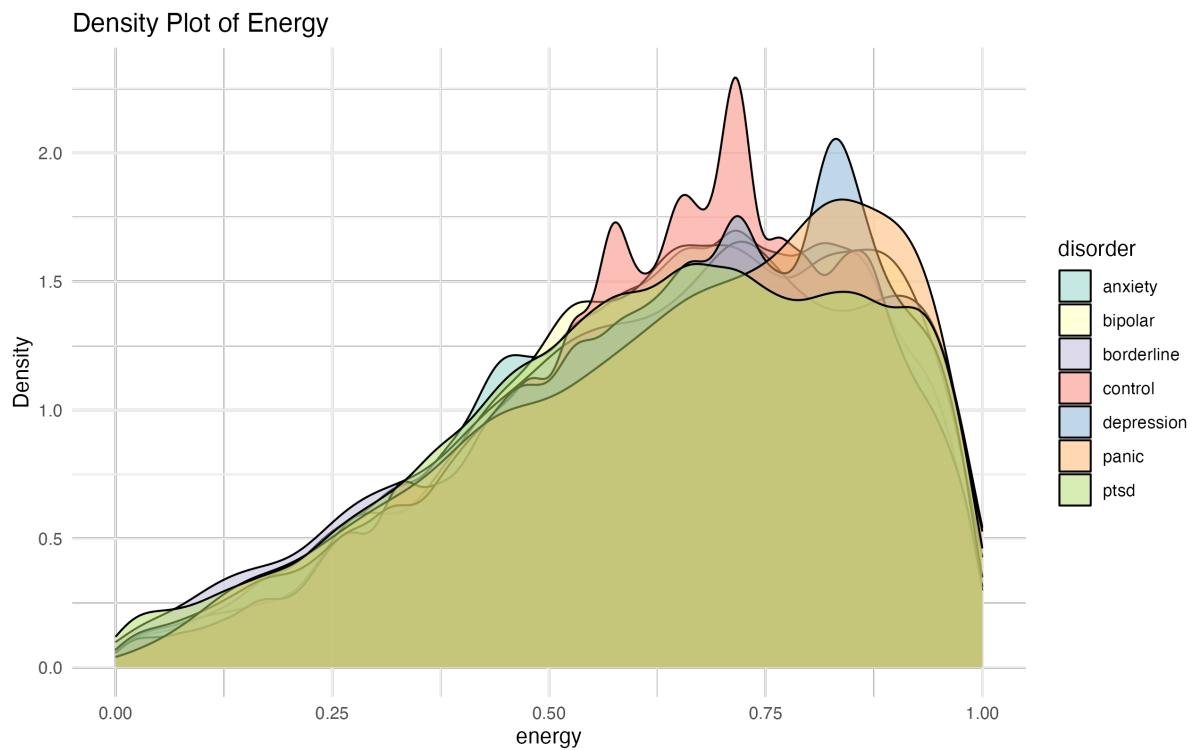


Figure 10

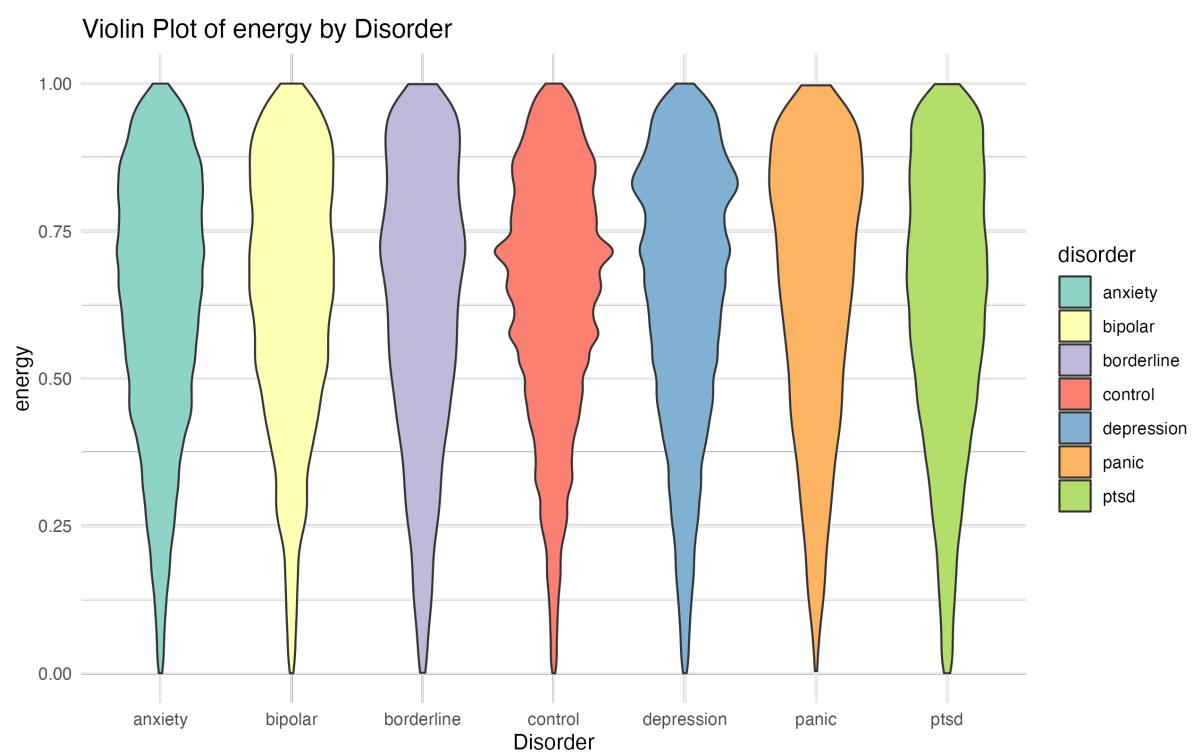


Figure 11

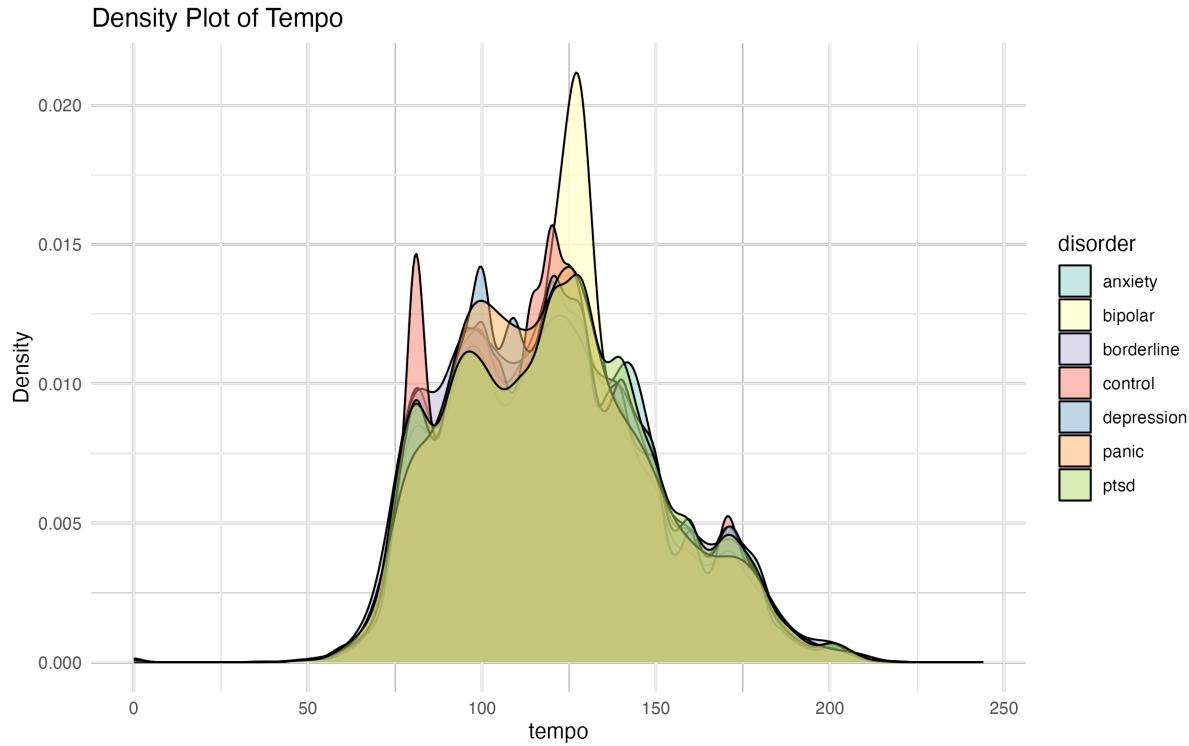


Figure 12

panic. However, the specific peak at 0.8 warrants further investigation.

Danceability and Tempo Trends

Looking at the tempo feature, Figure 12 shows that bipolar disorder stands out at a specific tempo (around 130 BPM). Additionally, the panic disorder population is less present at lower tempos (around 70 BPM) compared to the control population. This suggests a link between high energy, high tempos, and panic/anxiety disorders.

The danceability violin plot for different disorders (Figure 13) reveals that borderline personality disorder has a low density for low danceability. To explore this further, I plotted danceability densities for comparison (Figure 14). It appears that borderline individuals listen to fewer low danceability songs but also fewer high danceability songs compared to others. The bipolar pattern also shows high danceability around 0.60 - 0.70, where borderline individuals are less represented.

Keys

The boxplot of keys by disorder (Figure 15) shows that bipolar disorder stands out. This raises the question: Is the bipolar dataset representative of the bipolar population, or is there a real pattern? Considering the correlation plot (Figure 7), the key feature isn't particularly correlated with danceability or tempo.

Questionable Interpretations

Interpreting these results requires caution, especially considering the differing representation counts of each disorder. While depression is one of the most represented disorders, bipolar, borderline, and panic disorders are relatively rare in this dataset. Therefore, results highlighting these disorders need to be carefully contextualized.

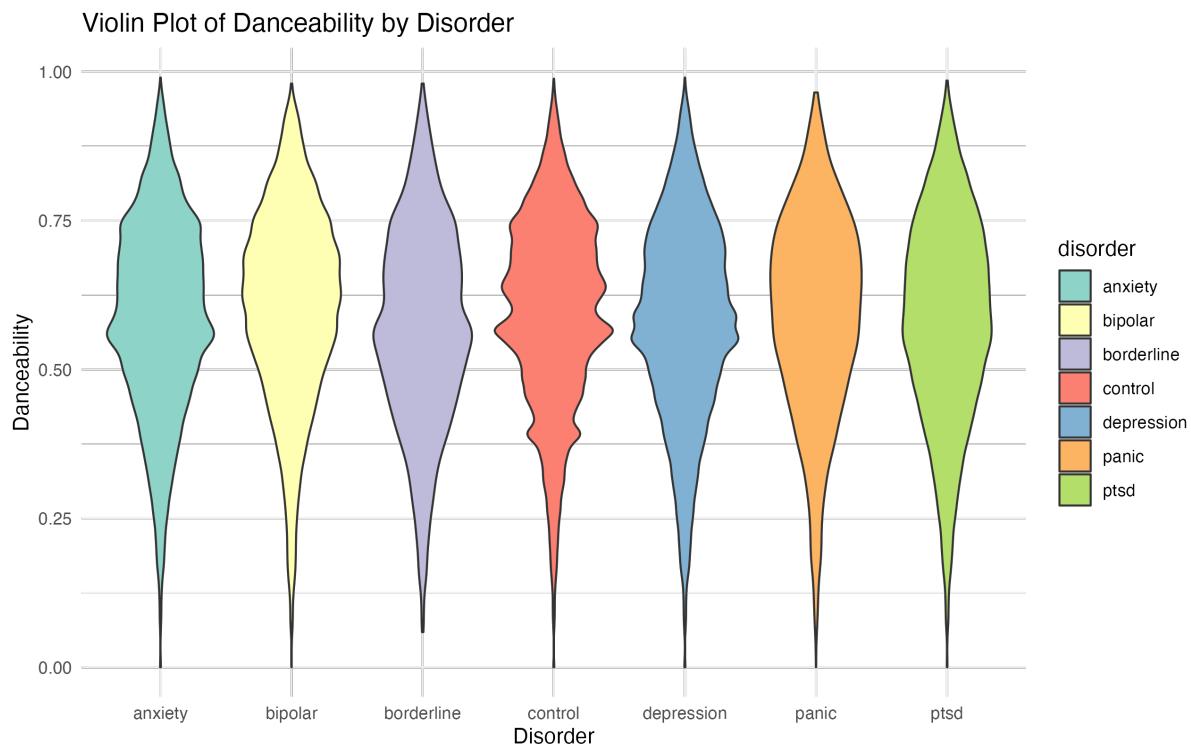


Figure 13

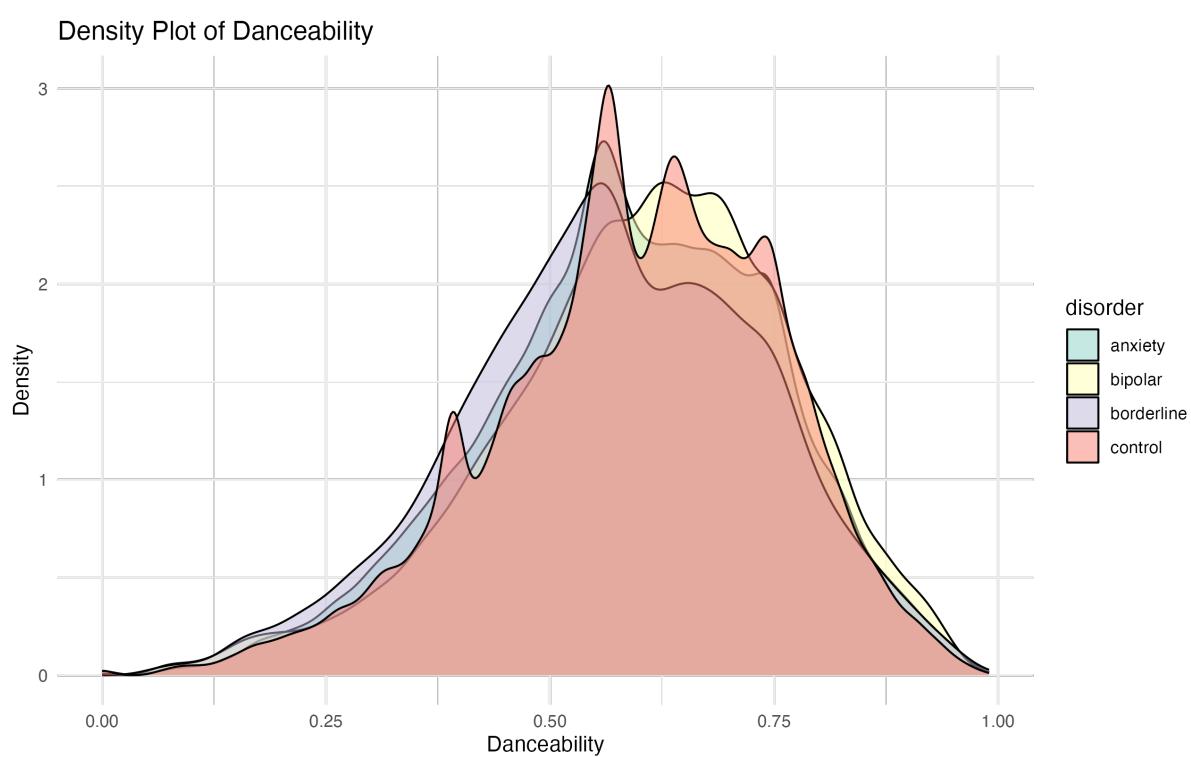


Figure 14

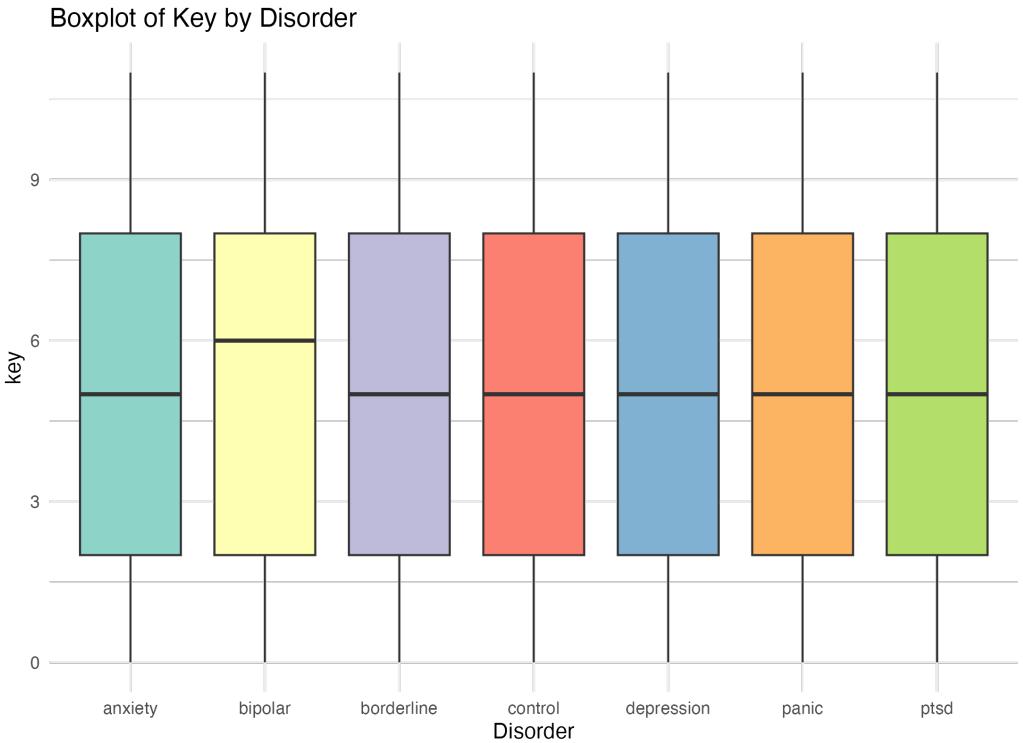


Figure 15

Linking these patterns with music genres could further refine our interpretations.

Mental Disorders and Music Genres: Third Exploratory Data Analysis

This exploratory data analysis focuses on artists and their weighted counts (as mentioned in the first section) for each disorder. I intended to add the genres of each artist using the Spotify API. Unfortunately, due to time constraints and the time-consuming nature of data fetching, the rest of this analysis remains to be discovered.