Report : Does Mental Disorders have an influence on the type of music listened to ?

This is a really broad question, I have to admit.

Mental disorders (like PTSD, Depression, Borderline etc...) never ceased to get most of my attention. At first perceived like burdens and deep flaws of the personality, mental disorders are actually more common than we think and are usually inherently part of what makes human beings so paradoxally beautiful. Mostly, most of those disorders come from a place of high sensitivity or exposure to difficult life events and traumas... which often leeds to a particular sensibility and vision. What most adequate to see the difference in sensitivity and taste than through music ?

Mental disorders and music popularity : First exploratory data analysis

Kaggle dataset

The aim is, in this first part, to explore the first broad tendances easily extracted from a useful dataset found on Kaggle named "Twitter Mental Disorder Tweets and Musics Dataset", that tries to link mental disorders with music listened to.

Using the Kaggle API, I extracted only the two datasets that really made sense to what I wanted to show which was the "control" and "disorder" datasets linking each users, their disorder (either PTSD, Depression, Anxiety, Panic, Borderline Personality, Bipolar or Control) to the title and artist of the music he mentioned.

However, an important question arised from the very beginning : when the author of this dataset extracted the music's informations, what did the "music tweet" said exactly ? The construction of the dataset and what it means is a bit blurry. We do not know what the music really represents in the end for the user. Maybe the tweet was really negative about the specific music. For the sake of the analysis and because we can think that a person will naturally tend to tweet more about the music it likes because he listens to them much more than the others, I took for granted the fact that each users liked the music he was mentionning and I could therefore build models of "popularity" based on it.

The question of popularity and disorders proportion

Without popularity information linked to each music, the only way of measuring broad tendences with this dataset was to count for each artist the number of times he was mentionned for the different disorders.

But before creating such dataset, visualizing a basic first bar plot (Figure 1) can lead to an important question : Is the proportion of users from each disorder equal in the overall dataset ? This question is of importance since we want to know the proportion of users affected by a specific mental disorder for each artist. By plotting a similar plot but most adequate (see Figure 2), we can see that indeed, the proportion of unique users per disorder is not equal between disorders and could therefore lead to false interpretations.

The Plot that actually doesn't need to take care of this question typically is presented Figure 4. Because we do not compare disorders directly, but we realize a podium for each of

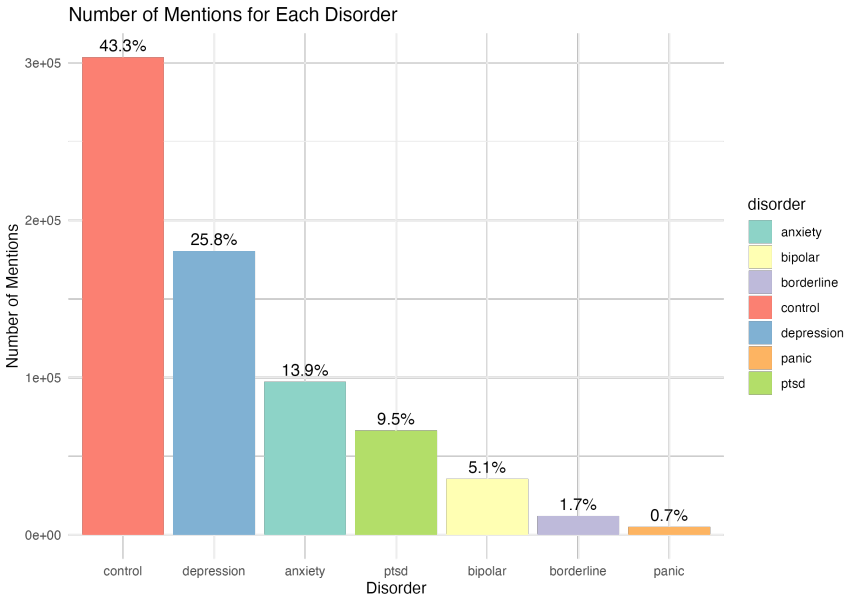

Figure 1

them independently, the problem is not of actuality. One element to take into account though is that we allowed multiple mentions of the same artists by the same user, and counted each mention as an important one, which could be questionnable. Is obsession for example of one's user on an artist should really counts 10 times more than a normal attraction to an artist's music ?

In Figure 3, we volontarily didn't modify the dataset since I wanted to show the most important difference for the top popular artists in between the general ("control") population and a population exclusively comprising people with mental disorders, with the "natural" proportions of each disorder found on twitter. This helps see that even by considering all disorders together (and the natural proportion of them within this "combined mental disorders" group), their sensitivity tends to show other tendencies than with the control population.

artists popularity showing disorders correlation

This quick analysis of the dataset leads enavitably to consider similarity between disorders. To create the plots of this part, I put a lot of importance in re-scaling the dataset counts of disorders for each artists. I therefore weighted the counts by dividing by the mentions pourcentages of each disorder and it leads to the correlation plot in Figure 5. We can see thanks to this plot that some disorders tends to be more correlated than others, and there overall correlation between the control population. A positive correlation means that the points tend to rise together, indicating that higher counts of one disorder tend to be associated with higher counts of another disorder and on the opposite, if the points tend to fall together, ther is negative correlation, indicating that higher counts of one disorder tend to be associated with lower counts of another disorder.

Figure 6 also followed the weighting process and shows how the 3 most popular artists over the global dataset have a different influence for each disorder population.

Mental disorders and music features : Second exploratory data analysis

After exploring genres preferences between disorders and the repartition of the most popular artists amongst them, I was curious to see if some mental disorders tend to influence the music taste of the population in terms of features like energy, danceability, key, tempo or speechiness.

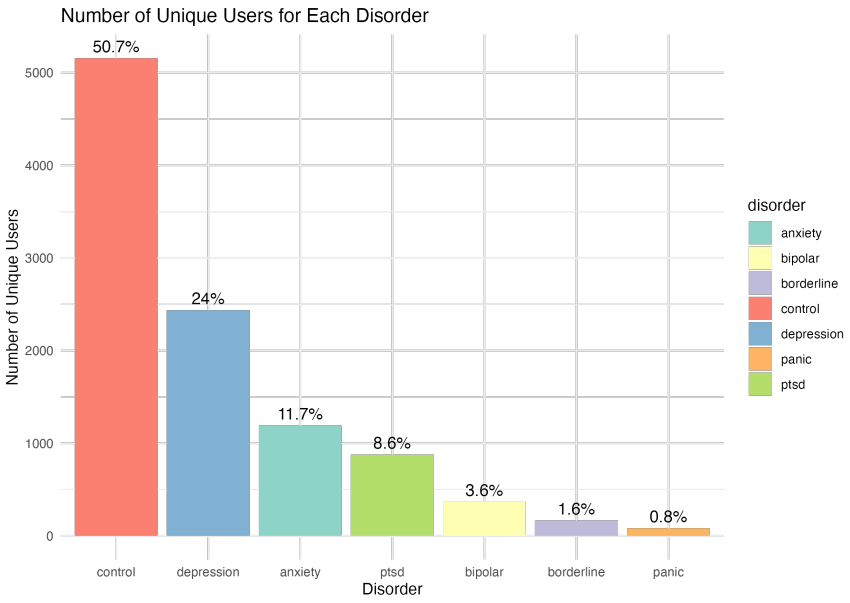

Figure 2

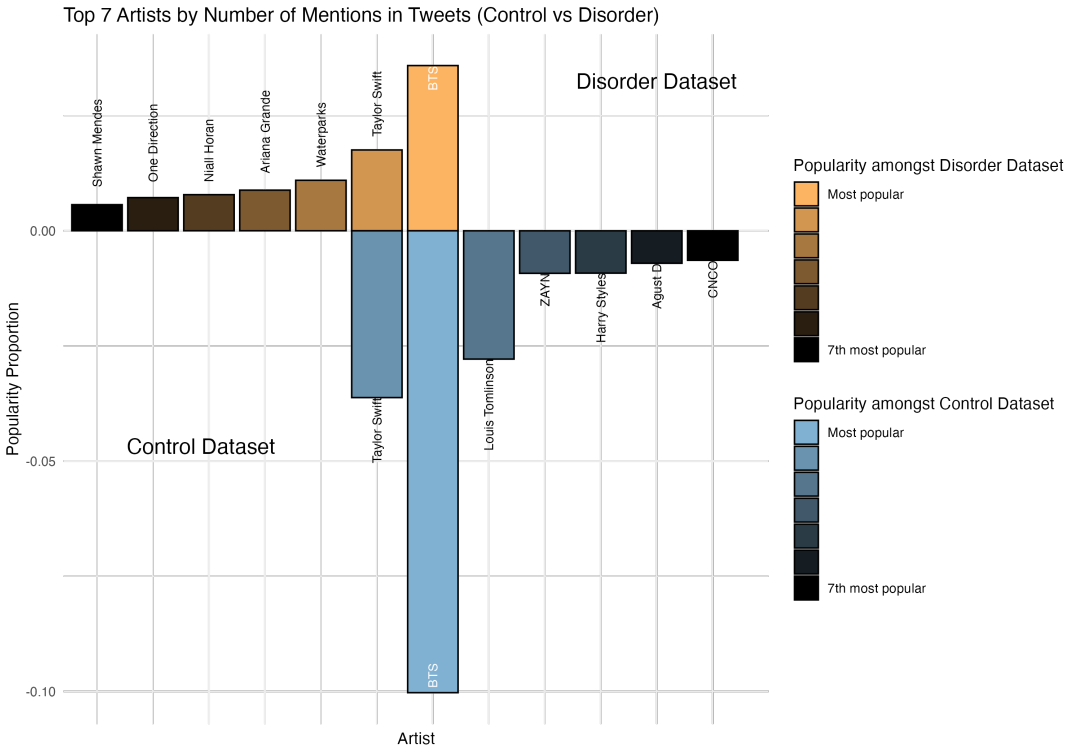

Figure 3

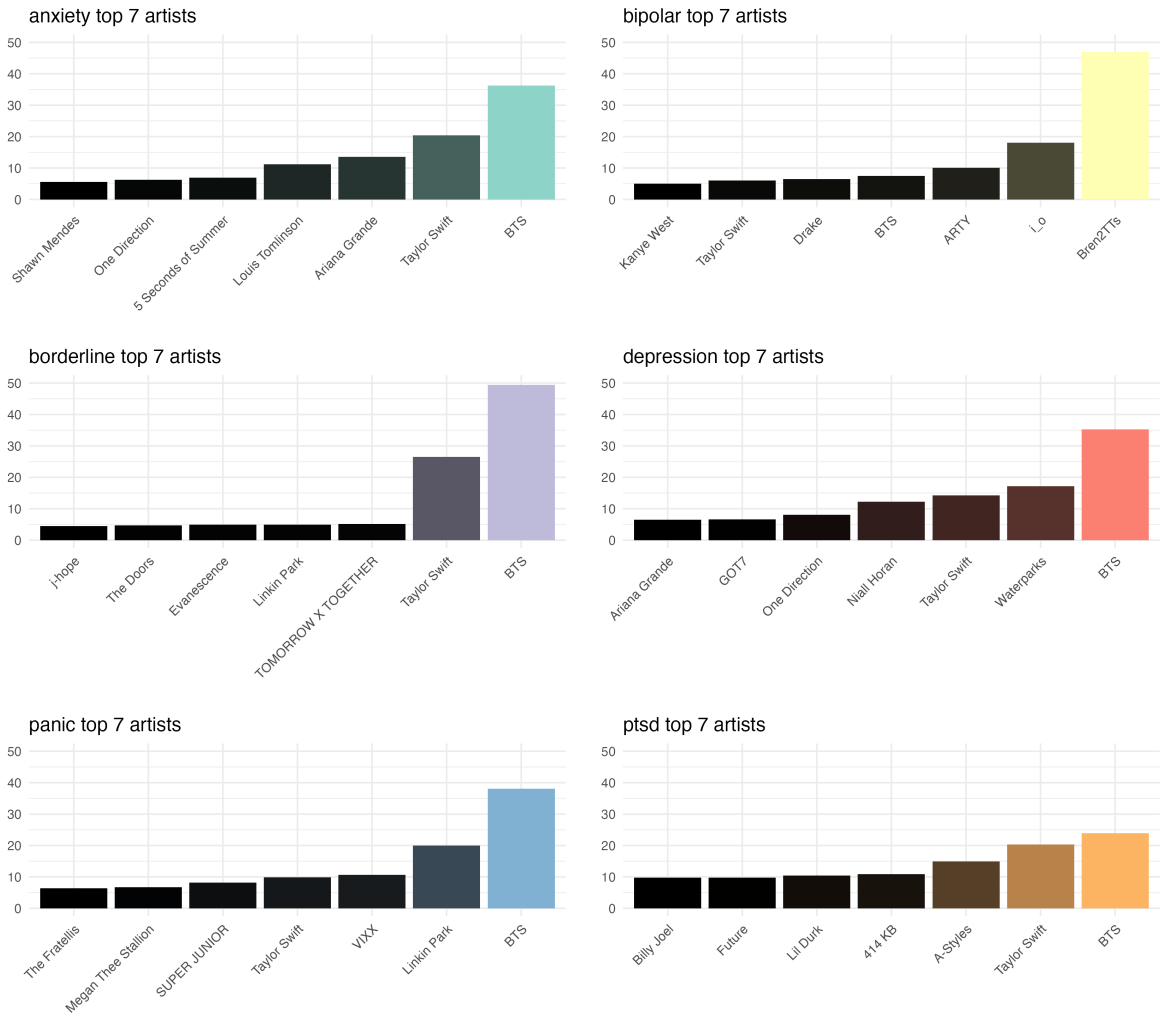
anxiety top 7 artists
50
40
30
20
10
0
Shawn Mendes
One Direction
5 Seconds of Summer
Louis Tomlinson
Ariana Grande
Taylor Swift
BTS
bipolar top 7 artists
50
40
30
20
10
0
Kanye West
Taylor Swift
Drake
BTS
ARTY
i_o
Bren2TTs
borderline top 7 artists
50
40
30
20
10
0
j-hope
The Doors
Evanescence
Linkin Park
TOMORROW X TOGETHER
Taylor Swift
BTS
depression top 7 artists
50
40
30
20
10
0
Ariana Grande
GOT7
One Direction
Niall Horan
Taylor Swift
Waterparks
BTS
panic top 7 artists
50
40
30
20
10
0
The Fratellis
Megan Thee Stallion
SUPER JUNIOR
Taylor Swift
VIXX
Linkin Park
BTS
ptsd top 7 artists
50
40
30
20
10
0
Billy Joel
Future
Lil Durk
414 KB
A-Styles
Taylor Swift
BTS

Figure 4

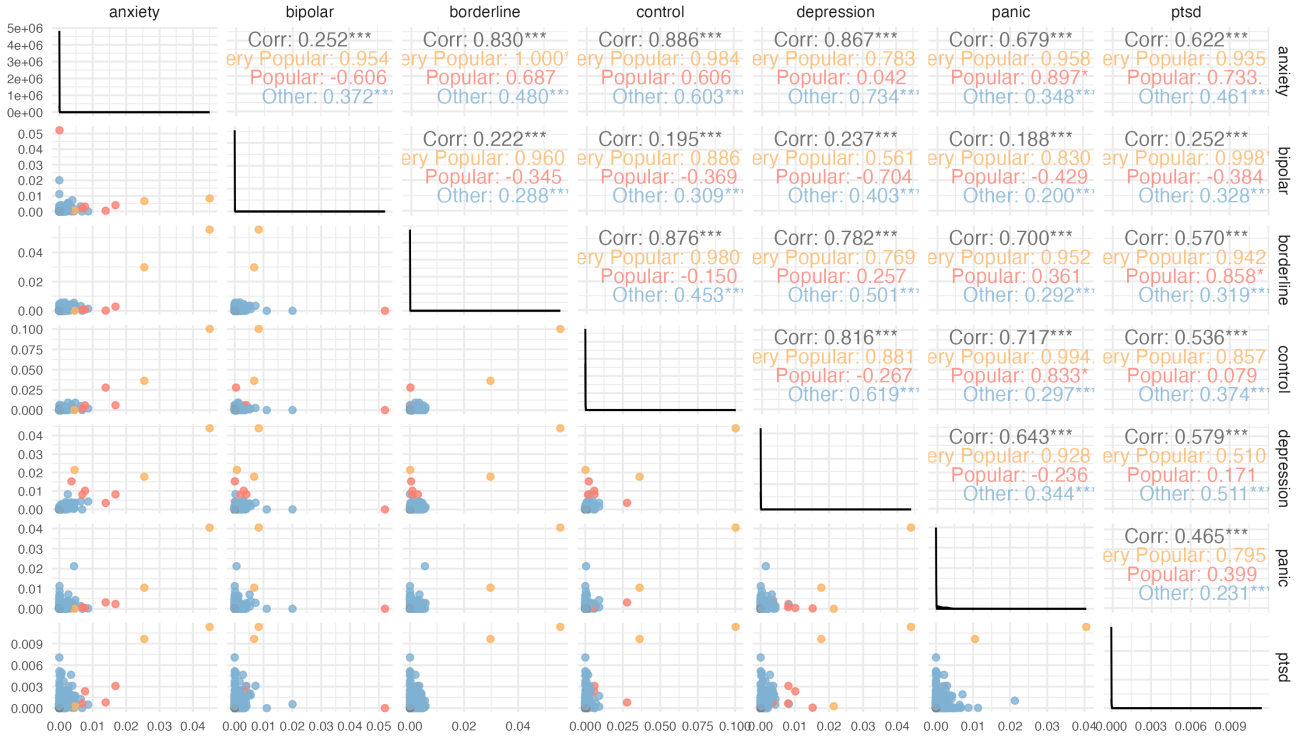

Figure 5

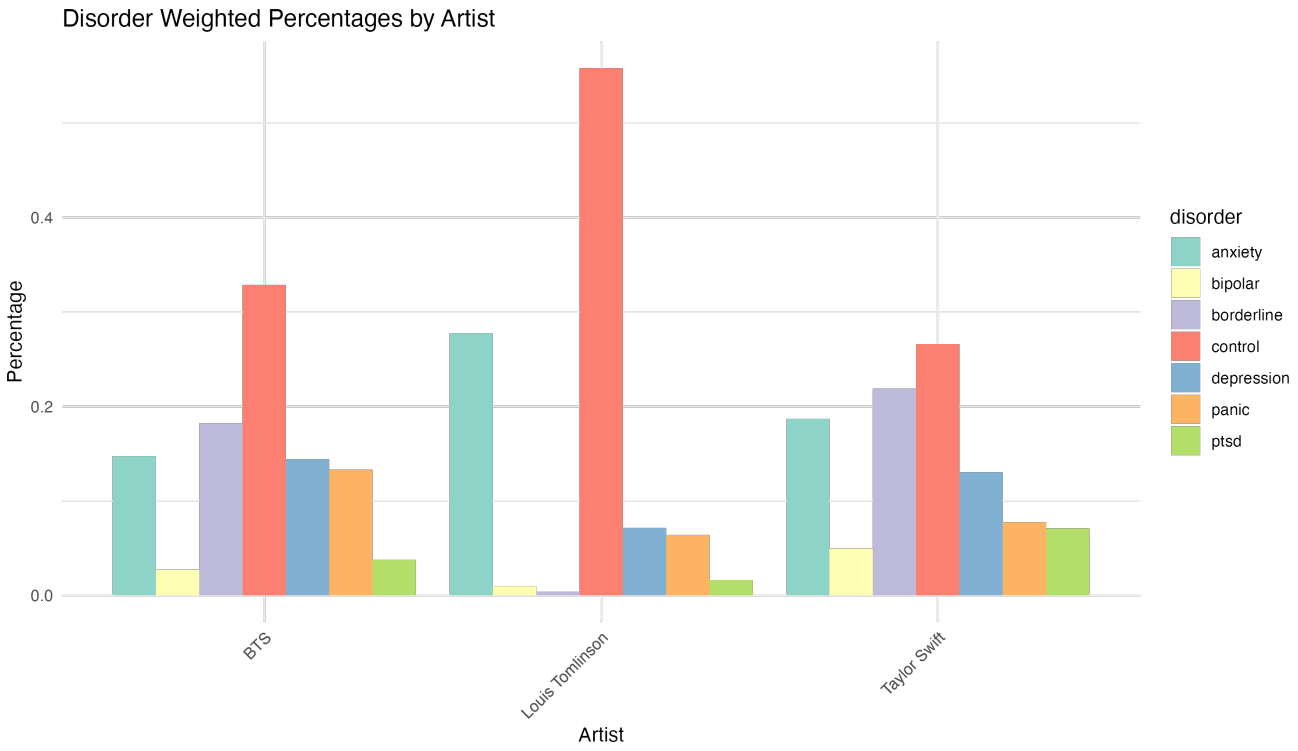

Figure 6

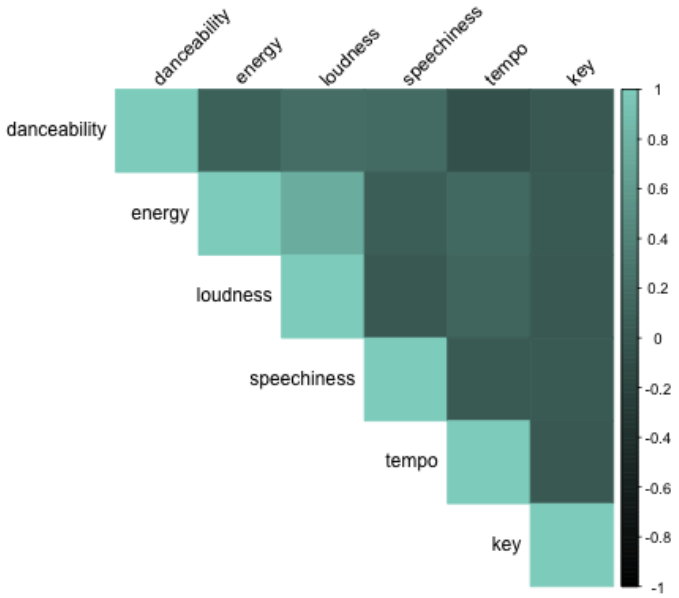

Figure 7

I built the dataset by fetching features of each song thanks to Spotify API and began by plotting the correlations between features over the entire dataset to have a first idea of how features might have similar patterns and help us identify those who are specific to one disorder (see 7).

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. Correaltions that are overall what we would expect.

Overall patterns

To visualize the main patterns, pairs-plots are really useful. I plotted two of them (for the first 3 disorders, and the next 3 others, both accompanied by the control points to have a common reference) - see Figures 8, and 9

If we might be a little disappointed by the overall resemblance between the control points and the others, some particularities caught my attention and I provided some boxplots, density plots and violin plots of specific features to see where the disorders might have an effect:

Energy patterns

With the energy plots (10, 11), one specific pattern caught my attention : the depression disorder has a very particular high density region around 0.8, where the control dataset doesn't. The only other two disorders that might have a little bit of this tendence are the panic and anxiety ones. I will not go into too much interpretation here as I am not well enough documented on this subject but one of the symptomes of depression is to feel deeply anxious or worried. Energy in music is a perceptible measure of intensity and activity. It makes a bit of sense to me that high energy would correspond to anxiety and panic, but why specifically 0.8 and not hust above as seen in the control dataset ? There is so many interpretations we could give to


danceability
energy
loudness
speechiness
tempo
disorder
Corr: 0.090***
control: 0.102***
depression: 0.078***
panic: 0.166***
ptsd: 0.071***
Corr: 0.212***
control: 0.217***
depression: 0.210***
panic: 0.316***
ptsd: 0.182***
Corr: 0.187***
control: 0.165***
depression: 0.202***
panic: 0.202***
ptsd: 0.240***
Corr: -0.056***
control: -0.032***
depression: -0.081***
panic: -0.071***
ptsd: -0.085***
Corr: 0.743***
control: 0.740***
depression: 0.748***
panic: 0.786***
ptsd: 0.752***
Corr: 0.063***
control: 0.066***
depression: 0.076***
panic: 0.089***
ptsd: 0.024***
Corr: 0.170***
control: 0.150***
depression: 0.183***
panic: 0.233***
ptsd: 0.218***
Corr: 0.009***
control: 0.001
depression: 0.023***
panic: 0.073***
ptsd: 0.017***
Corr: 0.132***
control: 0.125***
depression: 0.140***
panic: 0.175***
ptsd: 0.159***
Corr: 0.037***
control: 0.032***
depression: 0.045***
panic: 0.115***
ptsd: 0.030***
control
depression
panic
ptsd

Figure 8

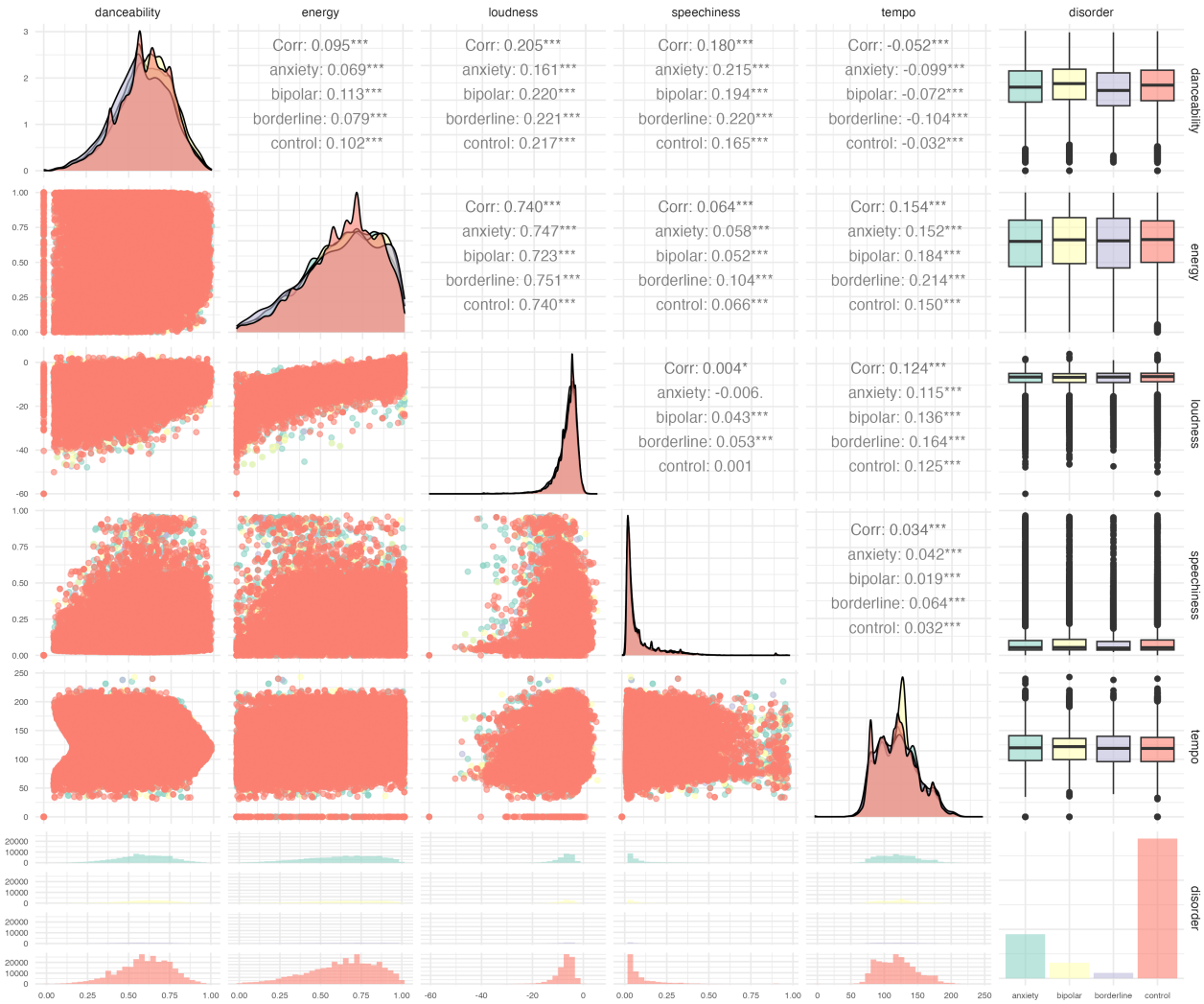
danceability
energy
loudness
speechiness
tempo
disorder
Corr: 0.095***
anxiety: 0.069***
bipolar: 0.113***
borderline: 0.079***
control: 0.102***
Corr: 0.205***
anxiety: 0.161***
bipolar: 0.220***
borderline: 0.221***
control: 0.217***
Corr: 0.180***
anxiety: 0.215***
bipolar: 0.194***
borderline: 0.220***
control: 0.165***
Corr: -0.052***
anxiety: -0.099***
bipolar: -0.072***
borderline: -0.104***
control: -0.032***
Corr: 0.740***
anxiety: 0.747***
bipolar: 0.723***
borderline: 0.751***
control: 0.740***
Corr: 0.064***
anxiety: 0.058***
bipolar: 0.052***
borderline: 0.104***
control: 0.066***
Corr: 0.154***
anxiety: 0.152***
bipolar: 0.184***
borderline: 0.214***
control: 0.150***
Corr: 0.004*
anxiety: -0.006.
bipolar: 0.043***
borderline: 0.053***
control: 0.001
Corr: 0.124***
anxiety: 0.115***
bipolar: 0.136***
borderline: 0.164***
control: 0.125***
Corr: 0.034***
anxiety: 0.042***
bipolar: 0.019***
borderline: 0.064***
control: 0.032***
anxiety
bipolar
borderline
control

Figure 9

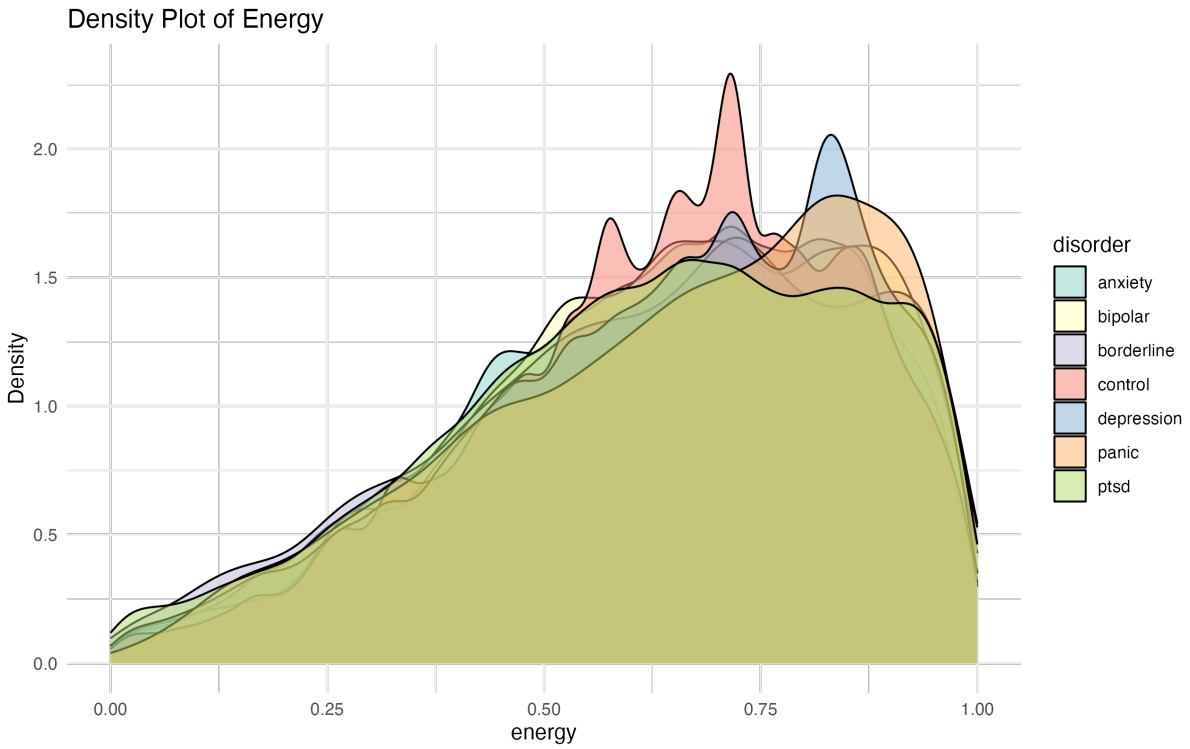

Figure 10

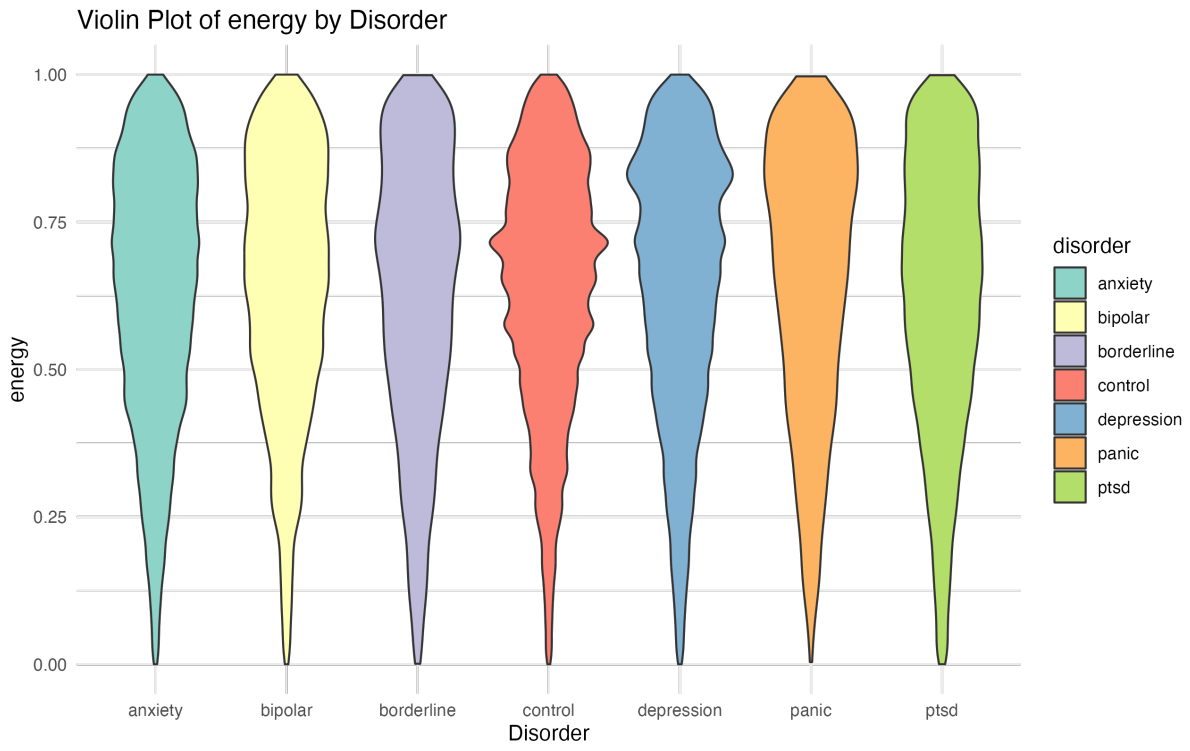

Figure 11

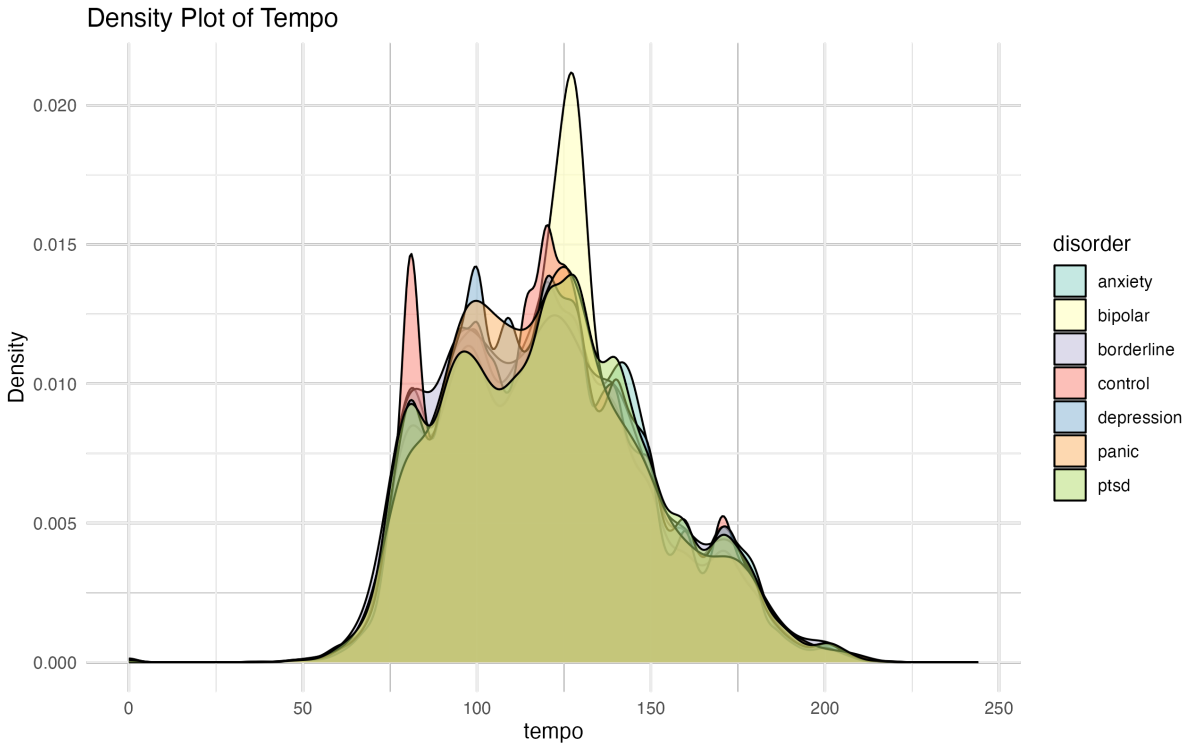

Figure 12

this pattern but it is hard to find the right one without having insights on what energy really represents.

danceability and tempo tendences

Looking at the Tempo feature now, we can see on Figure 12 that bipolar disorder clearly stands out at one specific tempo (around 130 BPM). Also, it seems like the tempo of panic disorder population is a lot less present in the lower tempos (around 70 BPM) than the control population is. Is panic and and anxiety that well related to high energy and high tempos?

I also show a violin plot of danceability amongst disorders because it seems to me that borderline danceability violin density is very low for low danceability. to explore more of the borderline pattern, I plotted the densities of danceability to compare some disorders with the borderline one and the control one (see Figure 14). But I found out that a pattern more visbile is that the borderline danceability density isn't as dense as for the others around 0.70. So borderline would listen to less "very low" danceability songs but also a bit less of higher danceability rate. We can spot also the resemblance in the bipolar pattern of high danceability around this specific area where borderlines are less sensitive (0.60 - 0.70).

We can ask again what does danceability represent exactly, but this would leed to a lot of questions.

Keys

We can finally see in the boxplot of key by disorders (see Figure 15) that bipolar once more stands out. Would this mean that the bipolar dataset is not really represnetable of bipolar population ? Or is there a real pattern around it ? This question arises by taking into account the correlation plot Figure 7, the key feature is not particularly correlated to danceability, nor to tempo.

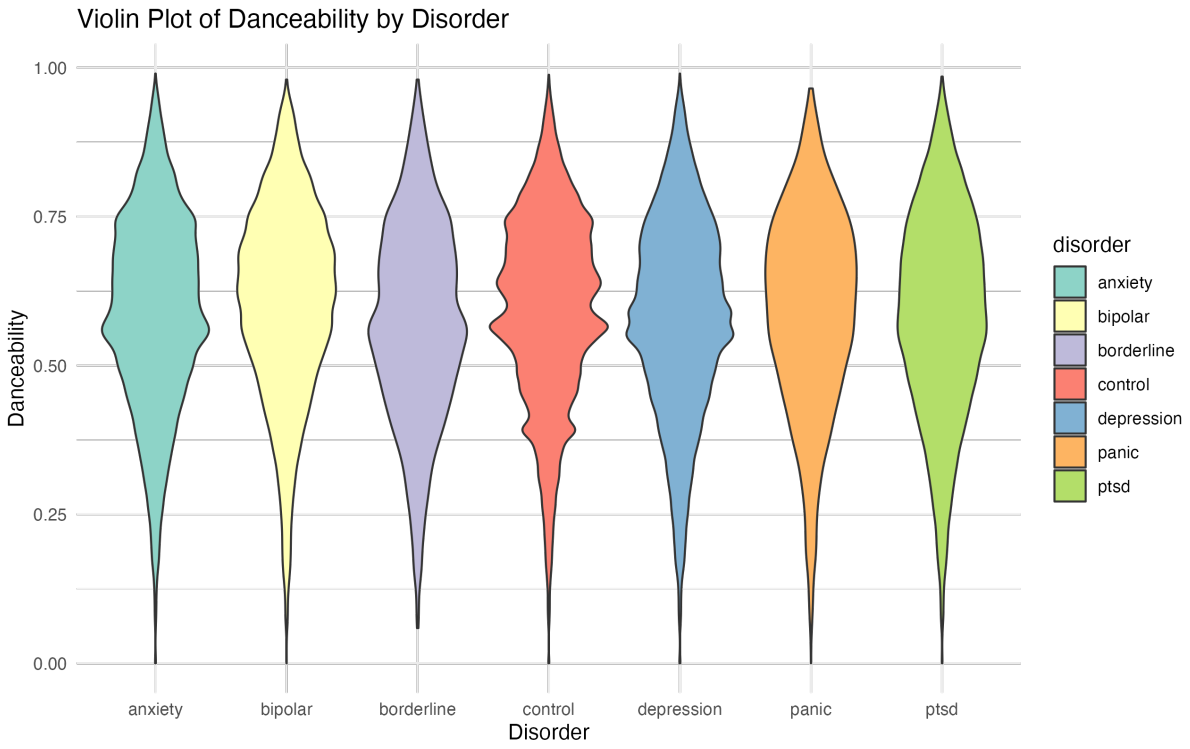

Figure 13

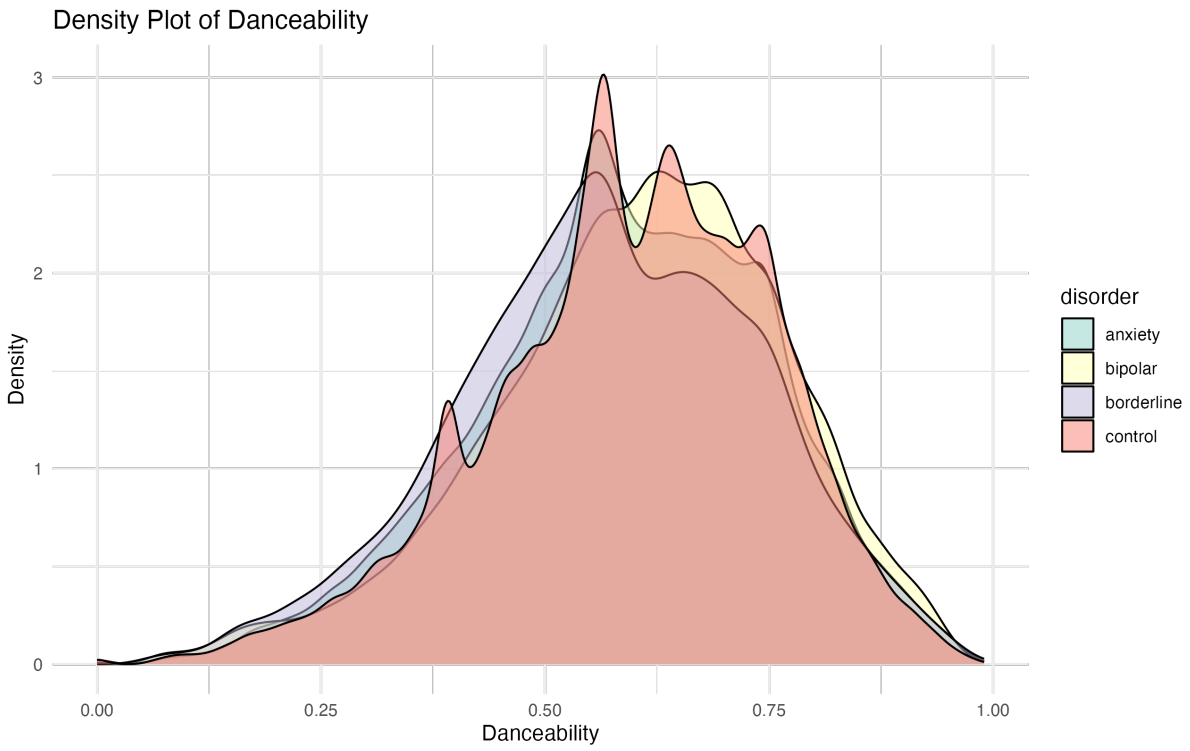

Figure 14

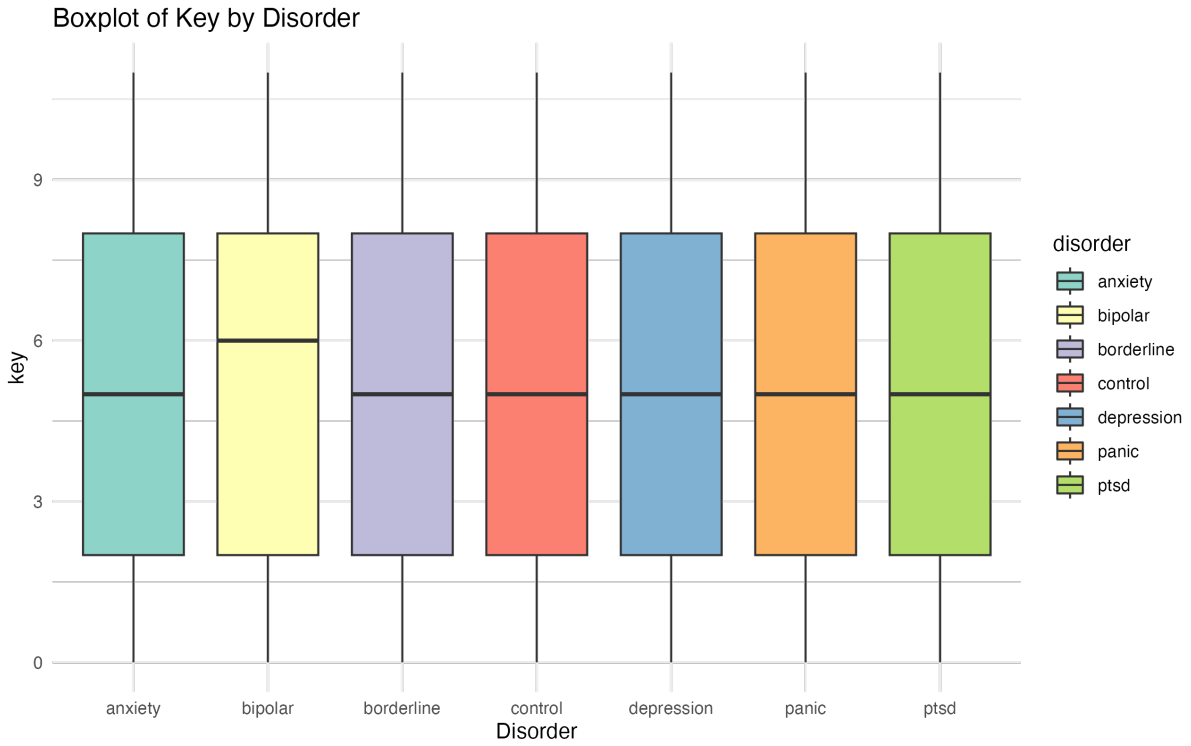

Figure 15

Questionnable interpretations

There is of course a lot to take into account to interpret those results, especially considering the difference of representation count of each disorder. If the depression disorder is one of the most represented disorder, the bipolar, borderline and panic ones are quite rare in this dataset. Therefore results that put them in the picture needs to be carefully put in perspective.

Also, by linking those patterns with the genre of music for example, we could thighten our interpretations.

Mental disorders and music genres : Third exploratory data analysis

This EDA is focusing only on the artists and their weighted counts (as mentioned in the first section) for each disorder. I wanted to add to this dataset the genres of each artists thanks to Spotify API.

Unfortunately, by lack of time and because fetching this dataset is too time - consuming, the rest of this analysis is yet to be discovered...