Spotify Tracks Dataset Report

109511097 江廷威

In the Spotify tracks dataset, each track has several attributes: popularity, duration(ms), explicit, danceability, energy, and etc. From the visualization below (all track genre included), the lower-left of the grid shows the correlation coefficient of each pair of the attributes, the upper-right of the grid show the scatter plot of each pair of the attributes.

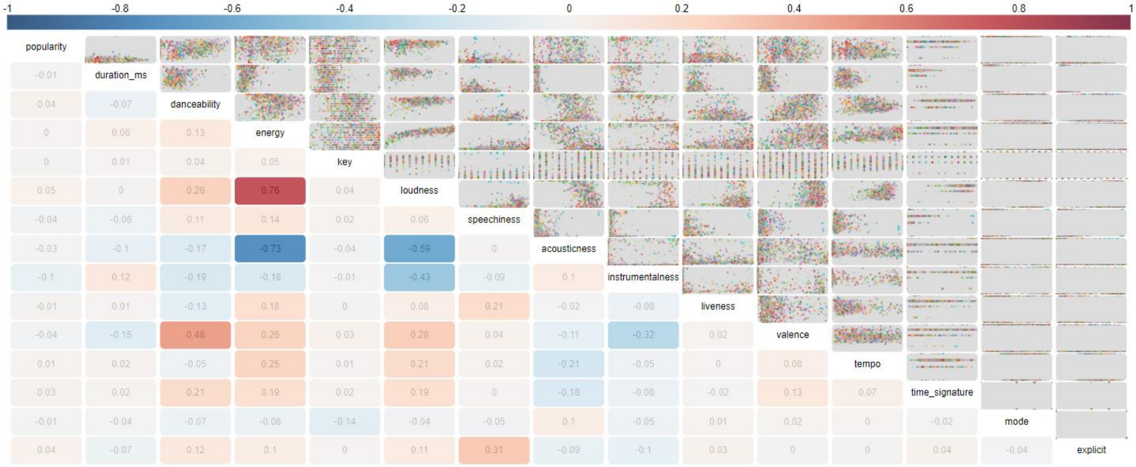


Figure 1

First of all, from the correlation coefficient we can clearly see that most of the attributes have weak correlation with each other. The pairs that have strong correlation are (energy, loudness), (energy, acousticness), (loudness, accousticness), and (danceability, valence). Moreover, we can also see that their corresponding scatter plots have a relatively strong trend going up or down.

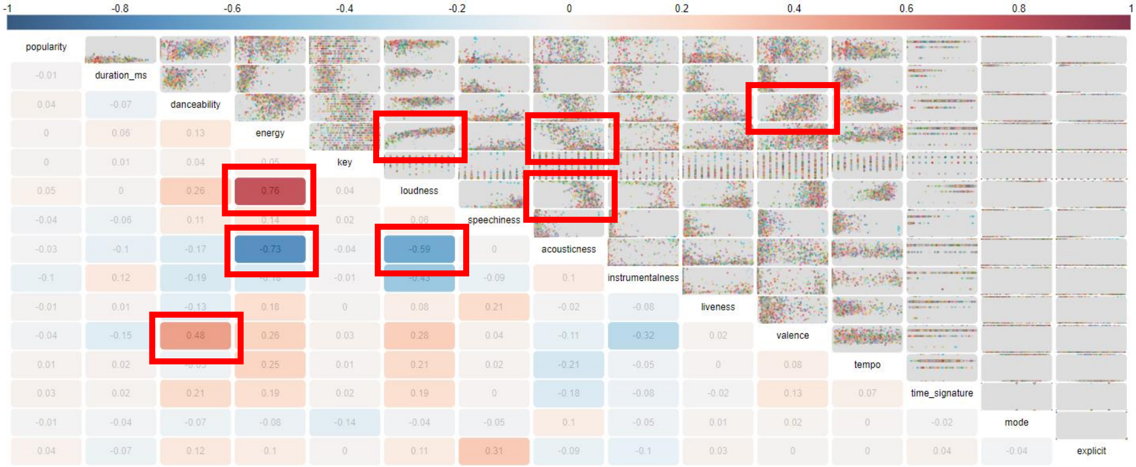


Figure 2

Let’s dive deeper into the visualization of the (energy, loudness) pair. According to the 2D density plot, we can see that most of the samples have loudness larger that -20. While the energy of the samples spans from 0.0 to 1.0, we can still see that most of the sample lies in the range between 0.3 and 1.0. Since there are over 100,000 samples in this dataset, if we visualize them using the ordinary scatter plot, there will be many occlusions (see Figure 4). However, by combining density plot with scatter plot, we can see the true distribution of the samples, and which of them are the outliers.

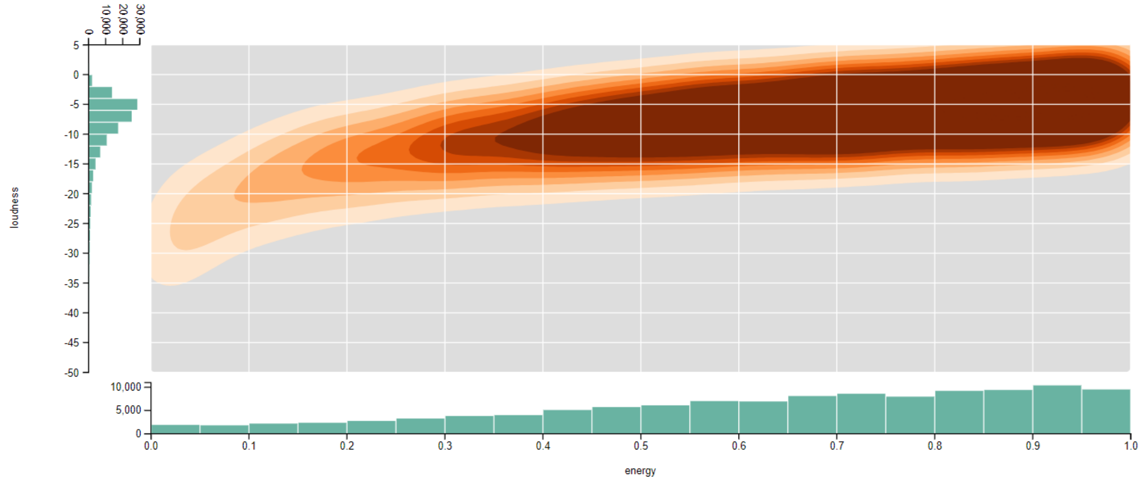


Figure 3

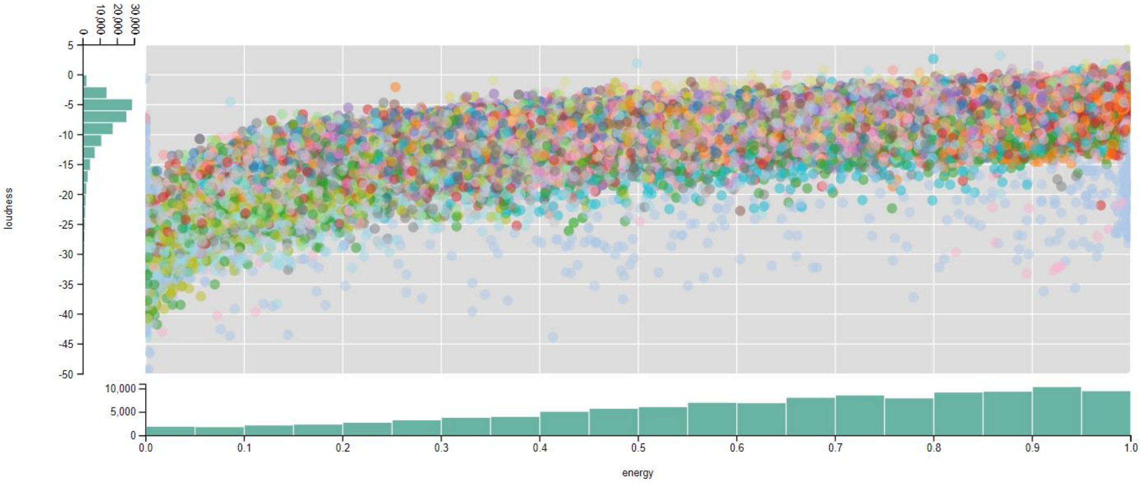


Figure 4

The color of the dots corresponds to the track genre of the sample. However, in the visualization above, there are too many samples that we cannot see the distribution of each class clearly. In this case, we can use the checkboxes on the left side to select some genre that we are interested in. For example, in the plot below we select 4 kinds of track genre: classical, comedy, death metal, and sleep. First, we can see that classical, comedy and death metal have large difference in terms of their energy, where classical has the lowest, and death metal have the highest. Second, we can see that the loudness of sleep genre is relatively smaller than the other 3 genres, and its energy can be either very low or very high.

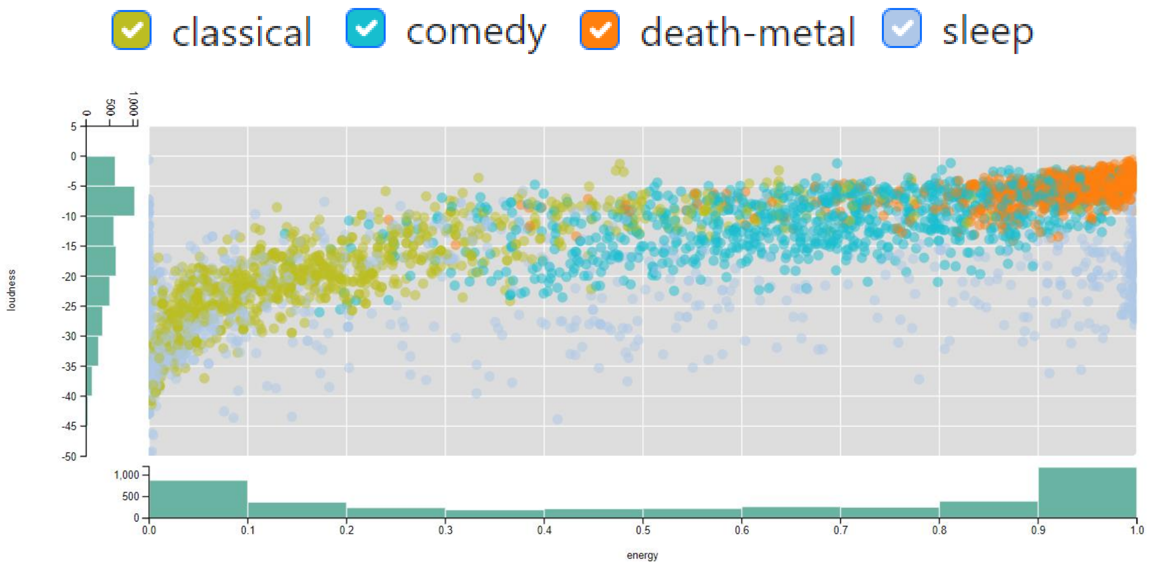
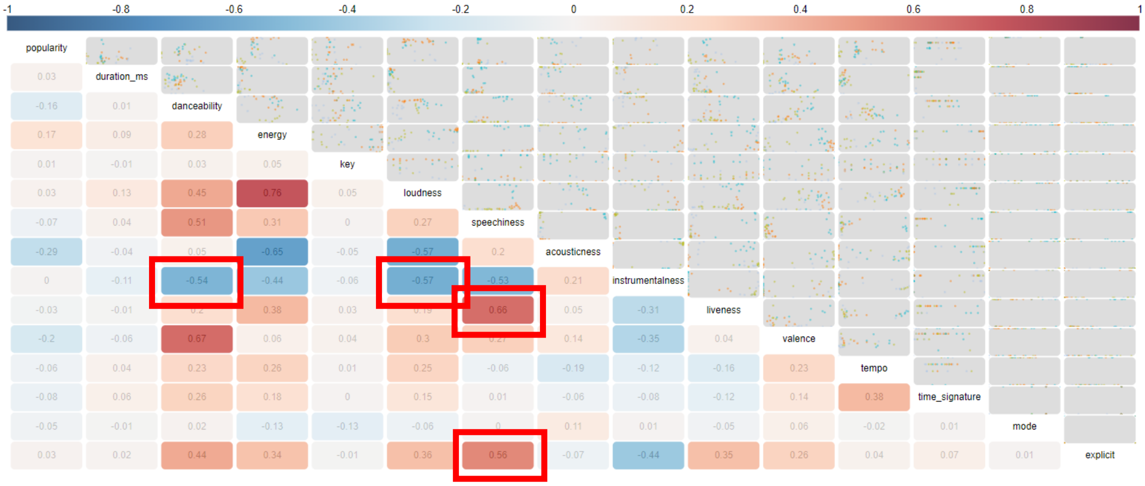


Figure 5

When we take a look at the gird again after we selected the 4 genres that we are interested in, we can see that correlation of some of the attribute pair get stronger: (danceability, valence), (loudness, instrumentalness), (speechiness, liveness), and (speechiness, explicit). From this we can say that the attribute pairs from different track genre might not have similar correlation.



popularity															
0.03	duration_ms														
-0.16	0.01	danceability													
0.17	0.09	0.28	energy												
0.01	-0.01	0.03	0.05	key											
0.03	0.13	0.45	0.76	0.05	loudness										
-0.07	0.04	0.51	0.31	0	0.27	speechiness									
-0.29	-0.04	0.05	-0.65	-0.05	-0.57	0.2	acousticness								
0	-0.11	-0.54	-0.44	-0.06	-0.57	-0.53	0.21	instrumentalness							
-0.03	-0.01	0.2	0.38	0.03	0.19	0.66	0.05	-0.31	liveness						
-0.2	-0.06	0.67	0.06	0.04	0.3	0.27	0.14	-0.35	0.04	valence					
-0.06	0.04	0.23	0.26	0.01	0.25	-0.06	-0.19	-0.12	-0.16	0.23	tempo				
-0.08	0.06	0.26	0.18	0	0.15	0.01	-0.06	-0.08	-0.12	0.14	0.38	time_signature			
-0.05	-0.01	0.02	-0.13	-0.13	-0.06	0	0.11	0.01	-0.05	0.06	-0.02	0.01	mode		
0.03	0.02	0.44	0.34	-0.01	0.36	0.56	-0.07	-0.44	0.35	0.26	0.04	0.07	0.01	explicit	

Figure 6

Figure 7 is the scatter plot with respect to speechiness and liveness. We can see that comedy tracks has high liveness and speechiness, which is quite different from the 3 other genres.

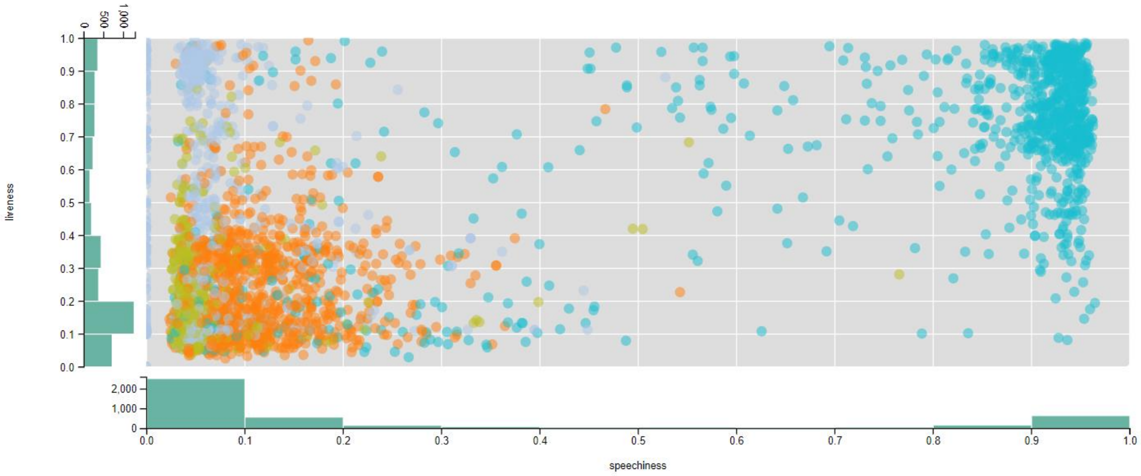


Figure 7

We can also brush the scatter plot to zoom in to the occluded area (lower-left corner). Most of the classical music has low liveness and speechiness.

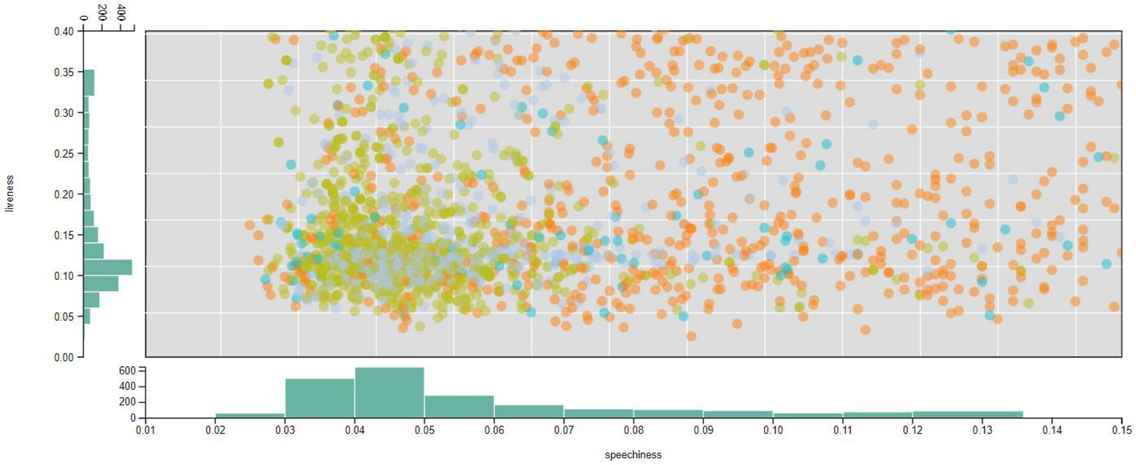


Figure 8

The original data might be a bit overwhelming, but using some visualization and interaction technique we can focus on the specific part of the data we are interested in and help us gain insight from them.