Analysis Of Spotify Audio Features For Classification of Song Genre

Philip Franco
Multivariate Statistics
7/4/22

Introduction

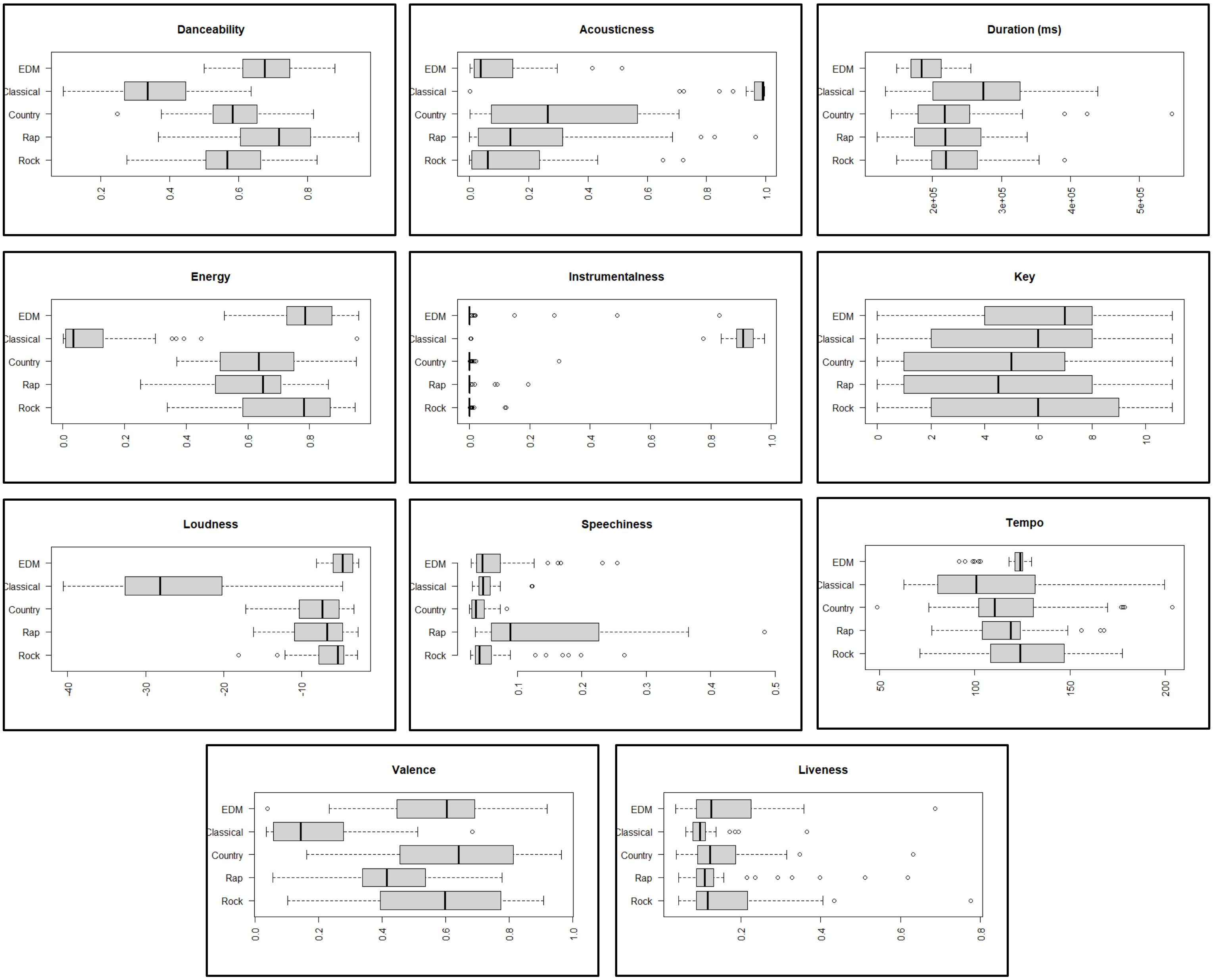
Music is becoming a popular platform do perform data analytics on and because of this, Spotify has an API that people can use for free. This API allows users to get valuable song data to look at different metrics for various machine learning tasks. This project will do it's best to use a data set created from scraping music from Spotify to classify songs into their proper genre.

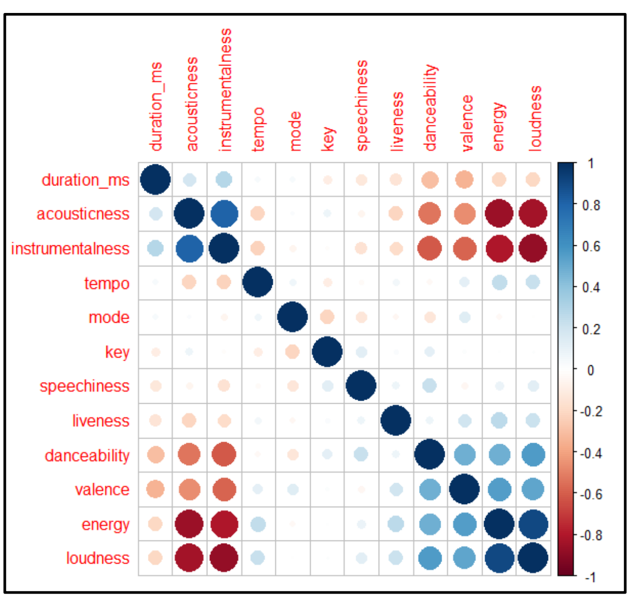
Dataset

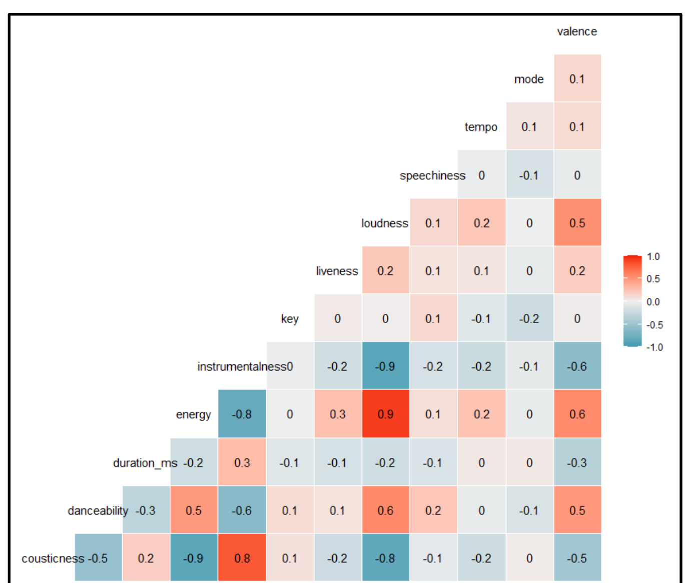
The data set was generated using SpotiPy, which is Spotify's Python API. The data sets consist of 250 of the most popular songs from the following genres: Rock, Country, Classical, Rap, and EDM. 50 songs from each genre was selected and scraped along with all the audio features provided. Spotify gives the following audio features that were later used to attempt to differentiate each song genre from each other.

1. Acousticness – Measures how acoustic a song is from a scale of 0-1.
2. Danceability – Takes tempo, rhythm stability, and beat strength into account to determine how suitable a track is for dancing. The scale is 0-1.
3. Duration_ms – The length of the song in milliseconds.
4. Energy – Songs that are high in energy have a fast tempo, are loud, and have good dynamic range. The scale is 0-1.
5. Instrumentalness – Predicts if there are no vocals in a song. Scale is 0-1, where values above 0.5 are intended to represent instrumental tracks.
6. Key – The key the track is in, and the scale maps to the pitch being used from -1-11. If no key was detected, -1 was used.
7. Liveness – Predicts if the track was played in front of an audience in the recording. Scale is 0-1 where 0.8 describes a high likelihood that there was an audience.
8. Loudness – This is measured in decibels.
9. Mode – Indicates modality of the track, or whether is in major, 1, or minor, 0.
10. Speechiness – Detects if there are spoken words in a song, where the higher values would be a speech only recording such as a podcast. Most songs with music and speech will fall between 0.3 and 0.6, and songs with no lyrics will fall below 0.3.
11. Tempo – This is an estimate of beats per minute of a song.
12. Valence - Predicts the sentiment of the song.

Below are box plots showing each variable separated by song genre. This gives us the full picture of the data set and helps us determine if classification can be possible using these metrics. Doing this points out that Classical music is inherently different than the other classifications by looking at Acousticness, Energy, and Loudness. Rap music is clearly differentiated by the Speechiness metric. EDM, Country, and Rock are more like each other than to Classical and Rap music.

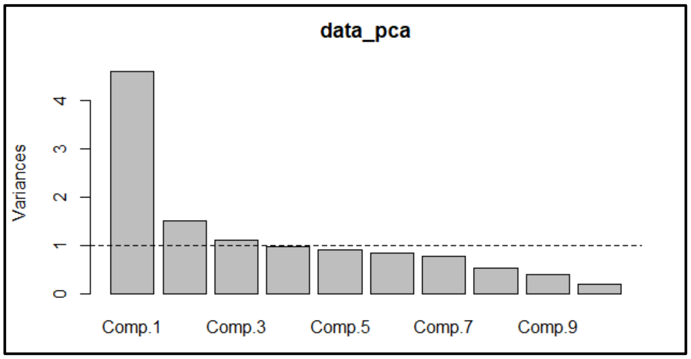

Danceability
Acousticness
Duration (ms)
Energy
Instrumentalness
Key
Loudness
Speechiness
Tempo
Valence
Liveness
EDM
Classical
Country
Rap
Rock

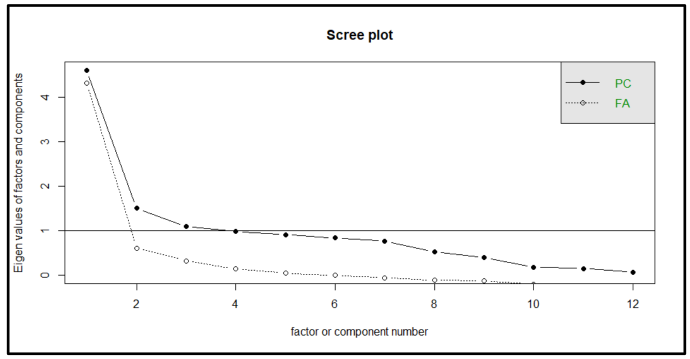

duration_ms
acousticness
instrumentalness
tempo
mode
key
speechiness
liveness
danceability
valence
energy
loudness
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


valence
mode 0.1
tempo 0.1 0.1
speechiness 0 -0.1 0
loudness 0.1 0.2 0 0.5
liveness 0.2 0.1 0.1 0 0.2
key 0 0 0.1 -0.1 -0.2 0
instrumentalness0 -0.2 -0.9 -0.2 -0.2 -0.1 -0.6
energy -0.8 0 0.3 0.9 0.1 0.2 0 0.6
duration_ms -0.2 0.3 -0.1 -0.1 -0.2 -0.1 0 0 -0.3
danceability -0.3 0.5 -0.6 0.1 0.1 0.6 0.2 0 -0.1 0.5
coustisness -0.5 0.2 -0.9 0.8 0.1 -0.2 -0.8 -0.1 -0.2 0 -0.5
1.0
0.5
0.0
-0.5
-1.0

The final two plots shown above are correlation matrix plots which show the correlation between each pair of metrics. Some metrics that have a high positive correlation are loudness and energy. Some metrics that have a high negative correlation are energy and acousticness. It is good to have these patterns in data because it means they have a relationship to each other. A strong relationship is good information to have when classifying data.

Principal Components Analysis



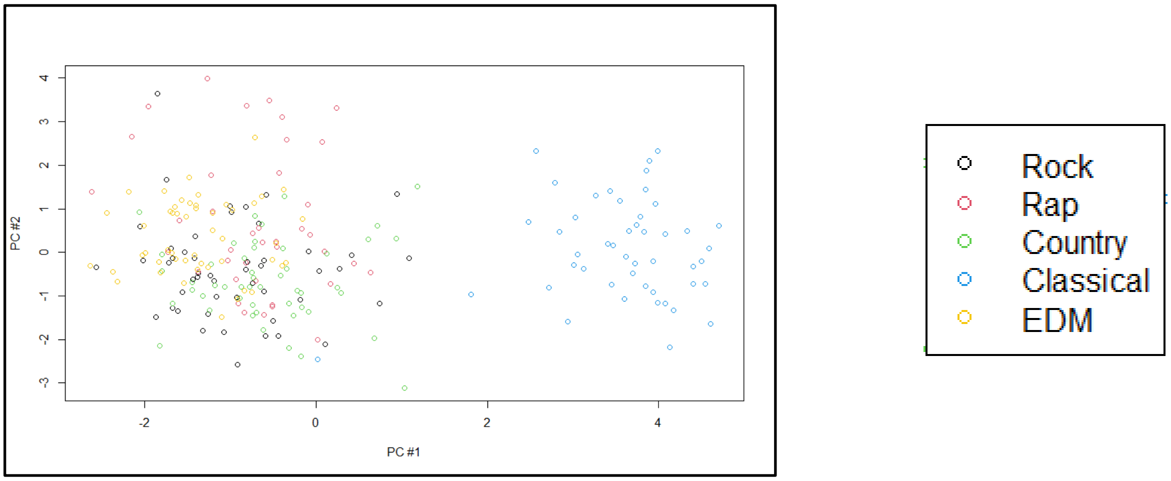


	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Standard Deviation	2.145	1.2247	1.0472	0.9908	0.9519	0.9157	0.8768
Proportion of Variance	0.3836	0.1250	0.0913	0.08180	0.07551	0.06988	0.06407
Cumulative Proportion	0.3836	0.5086	0.6000	0.6818	0.7573	0.8272	0.8913

Principal component analysis (PCA) is a dimension-reducing method that transforms a dataset with many variables into a dataset that has fewer variables, while still doing it's best to retain the information the larger dataset gives. When applying PCA with this dataset, the data shows that we can retain 82.72% of the information with only 6 variables, rather than using all 12.

	Acou	Danc	Dur	Ener	Instr	Key	Live	Loud	Spch	Tem	Mod	Val
CP1	0.41	-.235	0.17	-.426	0.43		-.139	-.432		-.114		-.323
CP2	0.11	0.27	-.284			0.51			0.43	-.363	-.487	
CP3	0.14	0.14	-.53	-.152				-.134	-.253	-.453	0.45	0.39

In the above table you can see the first three components that cover approximately 60% of the data. Component 1 appears to represent the Classical genre, because the Acousticness is higher, the Energy is lower, and the Valence is lower. Component 2 appears to be influenced by songs that have high Speechiness. Component 3 appears to look at songs that are shorter and have a lower tempo.



When plotting the first and second component, you can see how the Classical Genre is off towards the right of the plot. Rap music appears to be more focused towards the top left of the graph. EDM is focused on the middle left of the graph. The Rock and Country genres are blended towards the left/bottom of the graph.

Factor Analysis

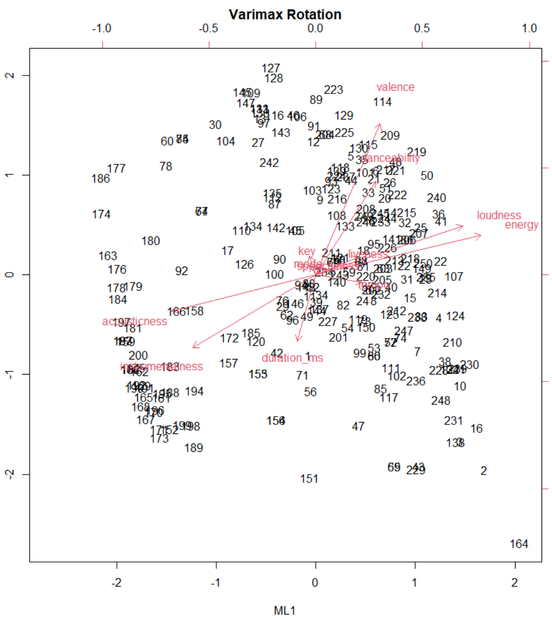
	K=4	K=5	K=6	K=7
P-Value	0.02042	0.30594	0.83735	NA
TLI	0.97	0.994	1.020	1.024
RMSE	0.052	0.024	0	0

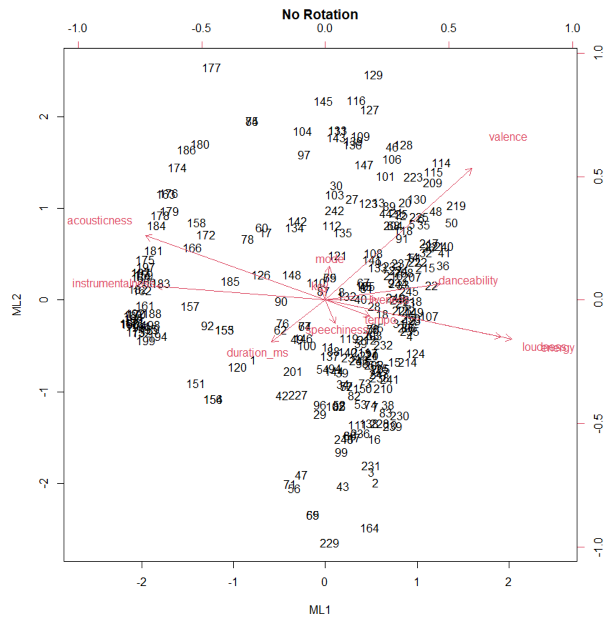
Factor Analysis is a model that looks at the measurements of the latent variables. When testing different values for K, we check the P-values. At K=5, the P-Value is first greater than 0.05 meaning the probability that the null hypothesis is true. The TLI value is greater than 0.95, and the RMSE value is lower than 0.05. K=5 is the best fit for this set of data, which makes a lot of sense when considering there are 5 categories for the dataset. Below is a table showing the 5 factors, with a data cut off of 0.25 to remove negligible values.

	Acou	Danc	Dur	Ener	Instr	Key	Live	Loud	Spch	Tem	Mod	Val
FA1	-.846	0.35	-.108	0.97	-.722		0.25	0.86		0.27		0.38
FA2	-.212	0.54	-.390	.23	-.425	0.11	0.10	0.29				0.88
FA3		0.3	-.128			0.33			0.33	-.262	-.491	-.194
FA4		0.3			-.44			0.28	0.27			-.16
FA5	0.47	-.26			0.13	0.11						

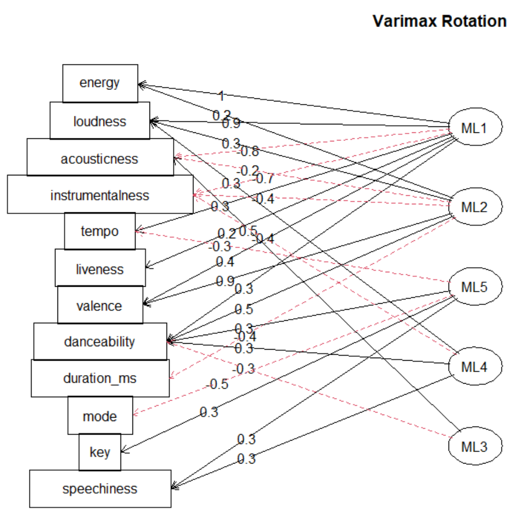
Factor 1 shows a very low acousticness, but a high amount of energy and loudness. This seems to be pulling out songs that are in the EDM genre. Factor 2 looks at songs that have high valence and danceability so songs that are positive and up-beat. I am not sure what Factors 3 and 4 represent. Factor 5 has high acousticness, so it could represent songs of the classical genre.

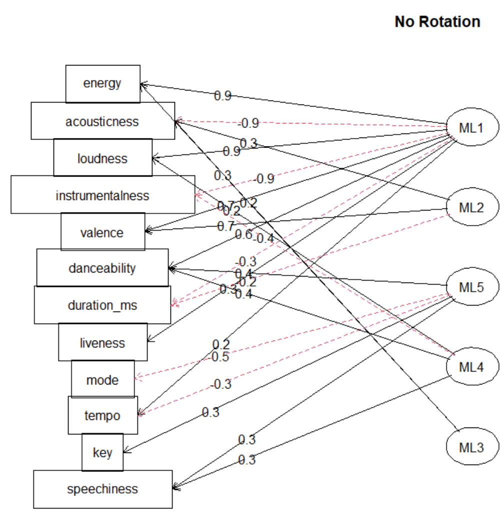
A varimax rotation was used for this data set, and by using this type of rotation the variance was increased in the data. Looking at the charts below, the loudness and energy are right on top of the same axis before the rotation. After the rotation, they are adjusted along with many of the other variables.

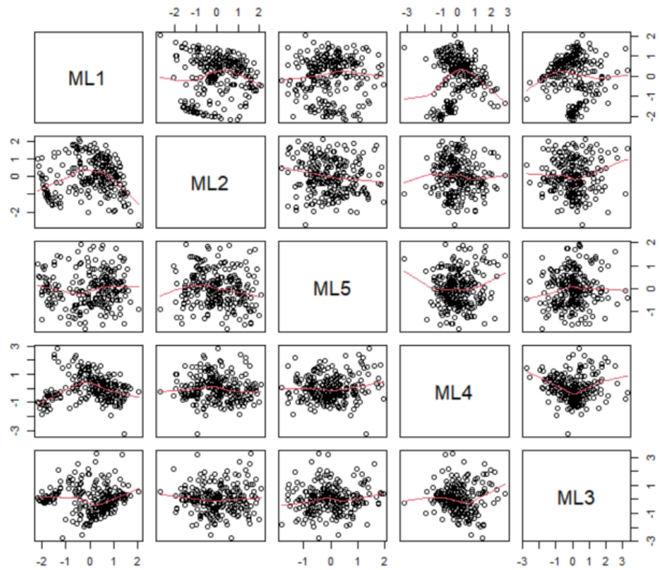




The graphs below show how the loadings changed form each variable when comparing the factor analysis using varimax rotation and not using varimax rotation. The rotation was a success by drastically increasing the variance in the dataset when providing factor analysis.





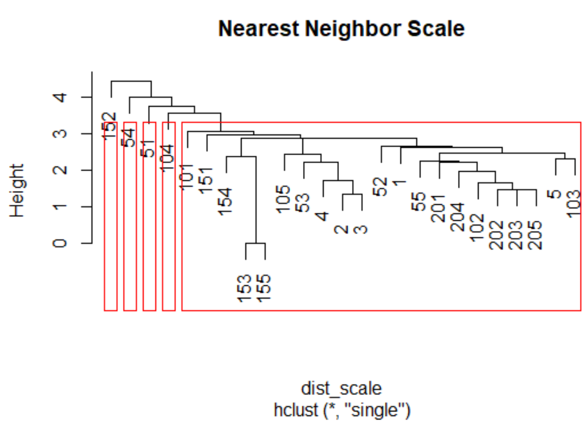


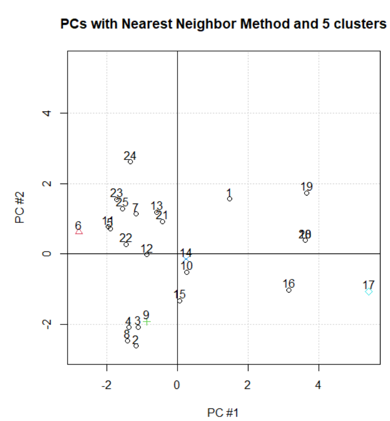
Cluster Analysis

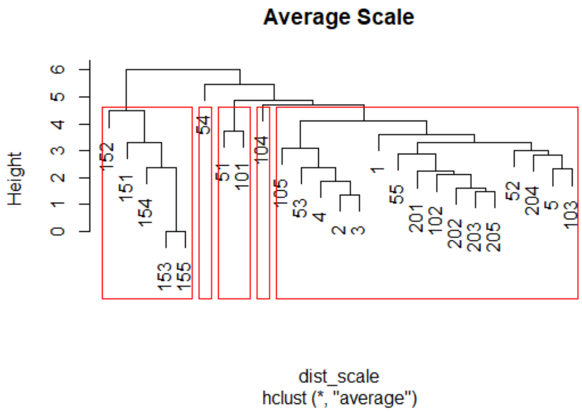
The next type of analysis completed was cluster analysis. Cluster analysis was performed on a smaller subset of the data so the plots would be more legible. Five songs were taken from each genre and used to create a dataset of 25 total songs. Below are the charts, where some types of cluster analysis clearly performed better than others. I did not include centroid scale because it performed the same as nearest neighbor, and both performed poorly.

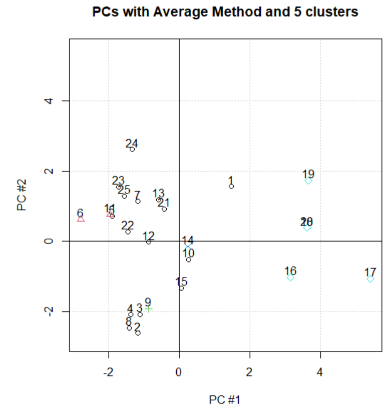
When looking at the charts below, the songs are grouped together by genre by number. In the cluster diagrams, 1-5 are rock, 51-55 are rap etc. As for the PCs charts numbers 1-5 are rock, numbers 6-10 are rap etc. Numbers 151-155 and 16-20 are for the classical genre which was able to be separated from the rest of the data using this analysis.

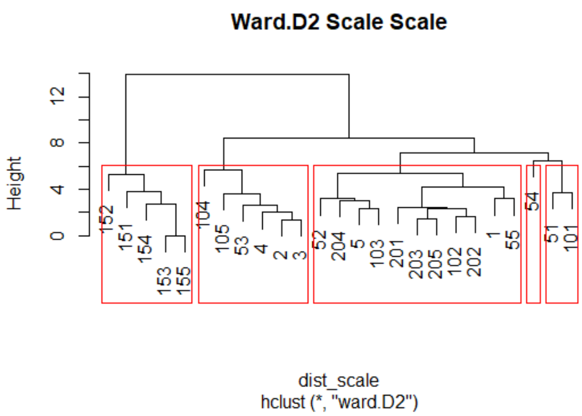
Complete scale and Ward.D2 scale performed the same when looking to see what gets clustered together. None of the other clusters only had one song genre in it. When using more samples, it did not help create more defined clusters for the 5 genres. Two clusters generally were created, which are classical music and not classical music.

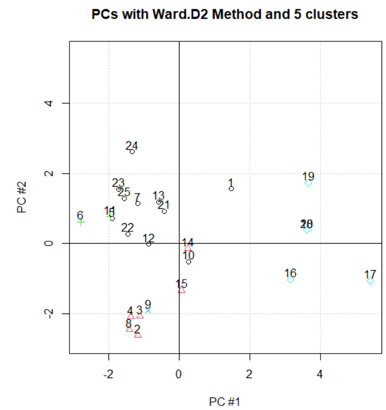


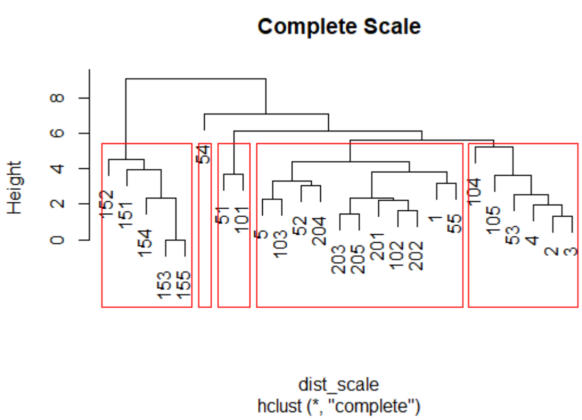


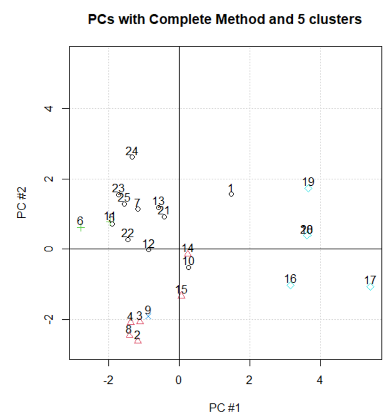












Discriminate Analysis

Performing discriminate analysis on the data yeilded interesting results. Using linear discriminate analysis, the algorithm was able to predict the genre correctly 66.4% of the time. Using quadratic discriminant analysis, the algorithm was able to predict the genre correctly 63.2% of the time. The following confusion matrixes were generated.

LDA	Classical	Country	EDM	Rap	Rock
Classical	48	1	0	0	1
Country	0	32	5	1	12
EDM	1	2	31	9	7
Rap	0	7	7	31	5
Rock	0	14	9	3	24

QDA	Classical	Country	EDM	Rap	Rock
Classical	48	1	0	0	1
Country	1	31	1	3	14
EDM	1	1	28	7	13
Rap	0	3	7	29	11
Rock	0	17	6	5	22

In these tests, LDA out performed QDA. Unexpectingly, it was easiest to categorize the classical genre, with both types of analysis properly classifying them 96% of the time. Country was the second most successful to classify, at 64% and 62% respectively. In both cases, Rock performed the worst often being confused for country soungs.