Do Countries Have Similar Music Tastes?

Emelia Barron & Michelle Mak

Introduction

These days, music streaming is ubiquitous. People have the ability to listen to music from all corners of the world. It might be interesting to find out to what extent music preferences vary among countries.

TOP 50 SONGS

- Australia
- Germany
- Hong Kong
- Mexico
- New Zealand
- Norway
- Portugal
- Singapore
- United Kingdom
- United States

ATTRIBUTES

- Danceability
- Energy
- Loudness
- Key Mode
- Speechiness
- Acousticness
- Instrumentalness
- Liveness
- Valence
- Tempo

DATA QUESTIONS

Can songs be classified by country based on its audio features?

Who are the most popular artists by country and overall?

Are there any song attributes that are positively/negatively correlated?

What are the most common key modes?

What songs are the most joyful (valence)?

What is the average value for each song characteristic for each country?

Exploratory Analysis

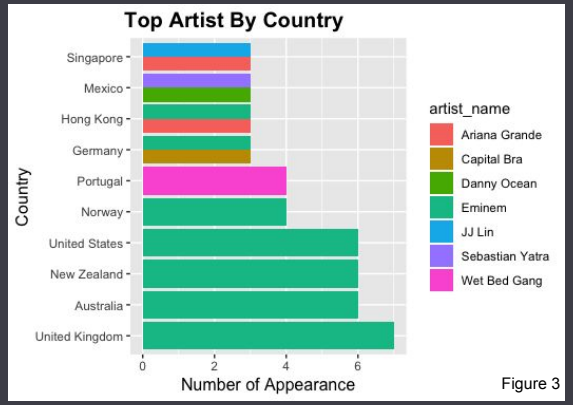

Figure 3

track_name	valence
Shape of You	0.931
Habibo	0.926
Dona Maria	0.924
Ahora Te Puedes Marchar	0.909
Mala Mía	0.901

Figure 4

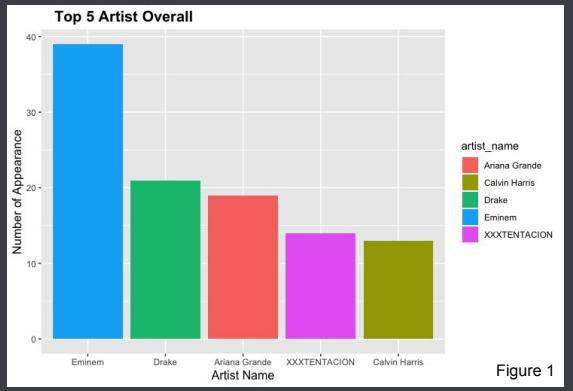

Figure 1

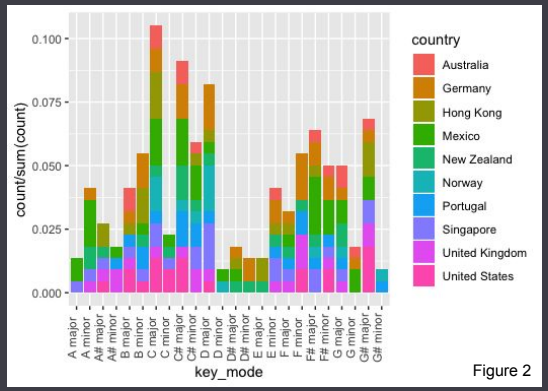

Figure 2

Eminem is the most popular artist

5 out of 10 countries had Eminem ranked as their top artist

2 out of the remaining 5 countries had Eminem tied as their top artist

Shape of You By Ed Sheeran was ranked the most joyful song

Most popular key modes were C Major, C# Major, and D Major

Figure 1: Top 5 Artists
Figure 2: Count of Key Mode By Country
Figure 3: Top Artist By Country
Figure 4: Top 5 Most Joyful Songs

Exploratory Analysis

Deeper Dive into Song Attributes

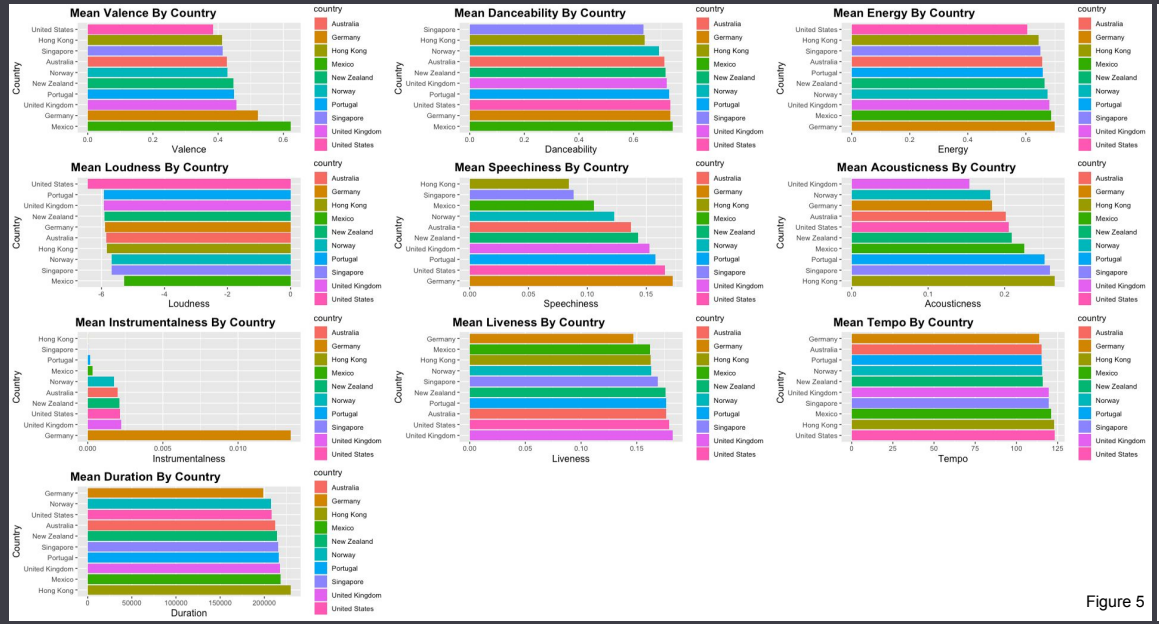

Figure 5

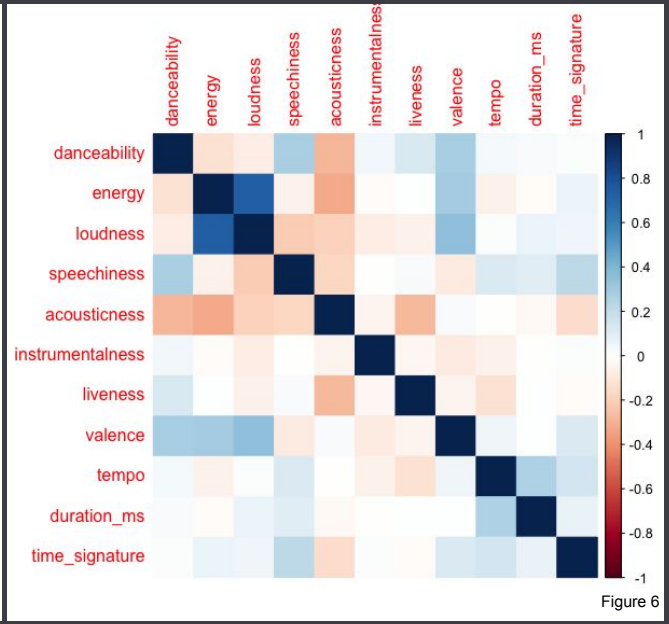

Figure 6

Valence, Speechiness, Acousticness, and Instrumentalness had higher variability among countries. Loudness and Energy were highly correlated, Acousticness and Energy were negatively correlated.

Figure 5: Mean Song Characteristics By Country
Figure 6: Song Characteristic Heat Map

Exploratory Analysis

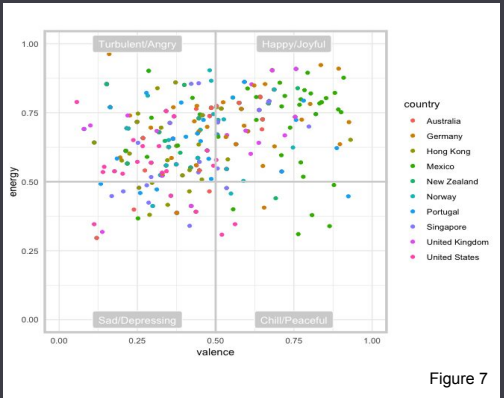

Figure 7

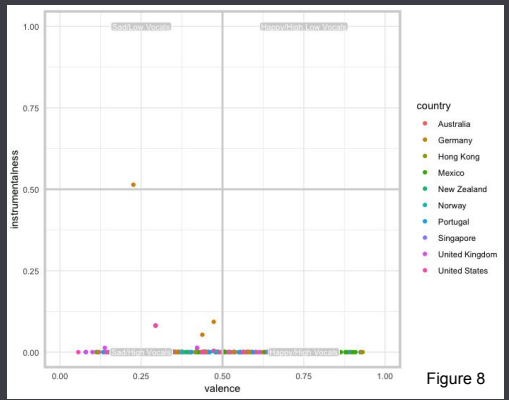

Figure 8

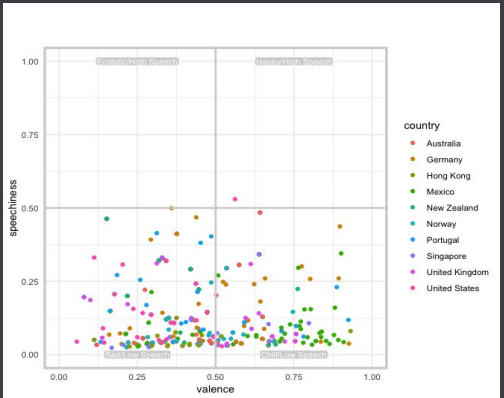

Figure 9

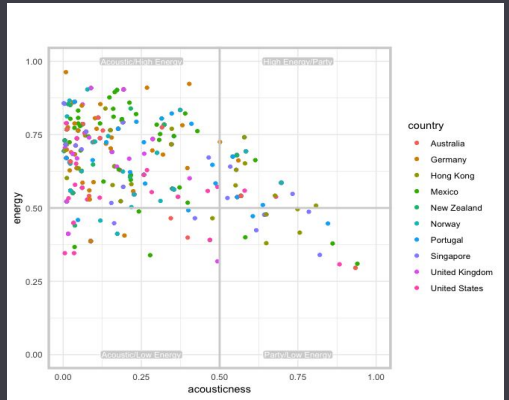

Figure 10

Figure 7: Energy vs. Valence
Figure 8: Instrumentalness vs. Valence
Figure 9: Speechiness vs. Valence
Figure 10: Energy vs. Acousticness

Clustering with K-Means

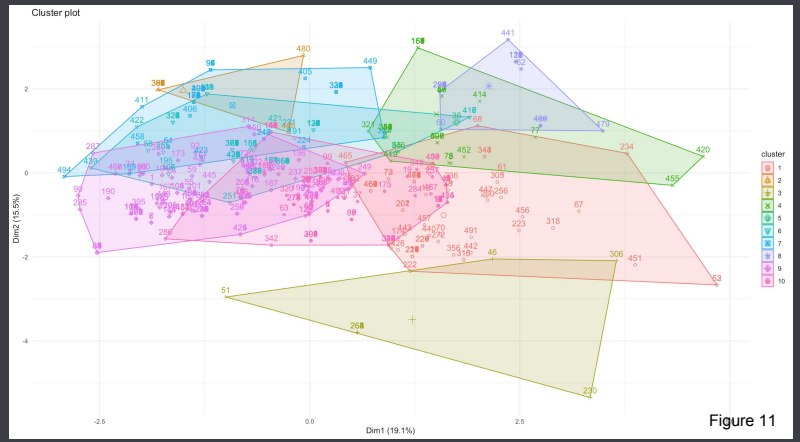

Figure 11

	1	2	3	4	5	6	7	8	9	10
Australia	12	1	1	3	7	8	0	6	4	8
Germany	20	2	1	3	3	3	1	6	4	7
Hong Kong	7	1	2	1	13	8	0	13	3	2
Mexico	23	0	0	1	4	8	0	4	5	5
New Zealand	12	1	1	3	6	6	0	8	5	8
Norway	14	1	2	2	4	7	0	9	5	6
Portugal	15	2	0	4	6	7	0	7	3	6
Singapore	11	1	2	2	14	8	0	9	1	2
United Kingdom	14	1	0	3	3	8	0	7	6	8
United States	4	1	0	3	9	7	0	10	5	11

Figure 12

Cluster 1 = Mexico

Cluster 2 = Germany or Portugal

Cluster 3 = Hong Kong, Norway, or Singapore

Cluster 4 = Portugal

Cluster 5 = Hong Kong or Singapore

Cluster 6 = Australia, Hong Kong, Mexico, Singapore, or UK

Cluster 7 = Germany

Cluster 8 = Hong Kong

Cluster 9 = United Kingdom

Cluster 10 = United States

Figure 11: Ten Cluster FVIZ Plot
Figure 12: Ten Cluster CLUS Plot

K-Nearest Neighbor

SpotifyTestSet$country	kNN_fit Australia	Germany	Hong Kong	Mexico	New Zealand	Norway	Portugal	Singapore	United Kingdom	United States	Total
Australia	0	1	2	1	1	2	0	0	0	2	9
	0.000	0.111	0.222	0.111	0.111	0.222	0.000	0.000	0.000	0.222	0.090
	0.000	0.062	0.133	0.091	0.200	0.250	0.000	0.000	0.000	0.118	
	0.00	0.01	0.02	0.01	0.01	0.02	0.00	0.00	0.00	0.02	
Germany	0	1	0	2	1	0	0	1	2	1	8
	0.000	0.125	0.000	0.250	0.125	0.000	0.000	0.125	0.250	0.125	0.080
	0.000	0.062	0.000	0.182	0.200	0.000	0.000	0.111	0.182	0.059	
	0.00	0.01	0.00	0.02	0.01	0.00	0.00	0.01	0.02	0.01	
Hong Kong	1	1	4	0	0	1	1	1	1	0	10
	0.100	0.100	0.400	0.000	0.000	0.100	0.100	0.100	0.100	0.000	0.100
	0.143	0.062	0.267	0.000	0.000	0.125	1.000	0.111	0.091	0.000	
	0.01	0.01	0.04	0.00	0.00	0.01	0.01	0.01	0.01	0.00	
Mexico	0	1	1	5	0	1	0	1	1	1	11
	0.000	0.091	0.091	0.455	0.000	0.091	0.000	0.091	0.091	0.091	0.110
	0.000	0.062	0.067	0.455	0.000	0.125	0.000	0.111	0.091	0.059	
	0.00	0.01	0.01	0.05	0.00	0.01	0.00	0.01	0.01	0.01	
New Zealand	2	3	0	0	0	2	0	0	1	6	14
	0.143	0.214	0.000	0.000	0.000	0.143	0.000	0.000	0.071	0.429	0.140
	0.286	0.188	0.000	0.000	0.000	0.250	0.000	0.000	0.091	0.353	
	0.02	0.03	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.06	
Norway	1	2	1	2	1	0	0	2	2	1	12
	0.083	0.167	0.083	0.167	0.083	0.000	0.000	0.167	0.167	0.083	0.120
	0.143	0.125	0.067	0.182	0.200	0.000	0.000	0.222	0.182	0.059	
	0.01	0.02	0.01	0.02	0.01	0.00	0.00	0.02	0.02	0.01	
Portugal	0	2	1	0	0	1	0	2	1	3	10
	0.000	0.200	0.100	0.000	0.000	0.100	0.000	0.200	0.100	0.300	0.100
	0.000	0.125	0.067	0.000	0.000	0.125	0.000	0.222	0.091	0.176	
	0.00	0.02	0.01	0.00	0.00	0.01	0.00	0.02	0.01	0.03	
Singapore	0	1	5	0	1	0	0	0	1	1	9
	0.000	0.111	0.556	0.000	0.111	0.000	0.000	0.000	0.111	0.111	0.090
	0.000	0.062	0.333	0.000	0.200	0.000	0.000	0.000	0.091	0.059	
	0.00	0.01	0.05	0.00	0.01	0.00	0.00	0.00	0.01	0.01	
United Kingdom	1	2	1	1	0	0	0	1	1	1	8
	0.125	0.250	0.125	0.125	0.000	0.000	0.000	0.125	0.125	0.125	0.080
	0.143	0.125	0.067	0.091	0.000	0.000	0.000	0.111	0.091	0.059	
	0.01	0.02	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01	
United States	2	2	0	0	1	1	0	1	1	1	9
	0.222	0.222	0.000	0.000	0.111	0.111	0.000	0.111	0.111	0.111	0.090
	0.286	0.125	0.000	0.000	0.200	0.125	0.000	0.111	0.091	0.059	
	0.02	0.02	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01	
Total	7	16	15	11	5	8	1	9	11	17	100
	0.07	0.16	0.15	0.11	0.05	0.08	0.01	0.09	0.11	0.17	

Figure 13

17% Accuracy

Figure 13: KNN CrossTable
Figure 14: FVIZ KNN

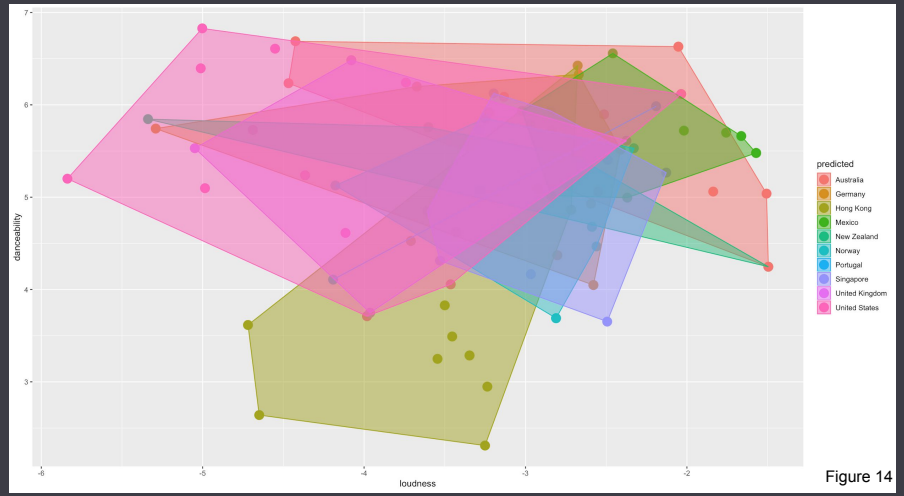

Figure 14

Decision Tree

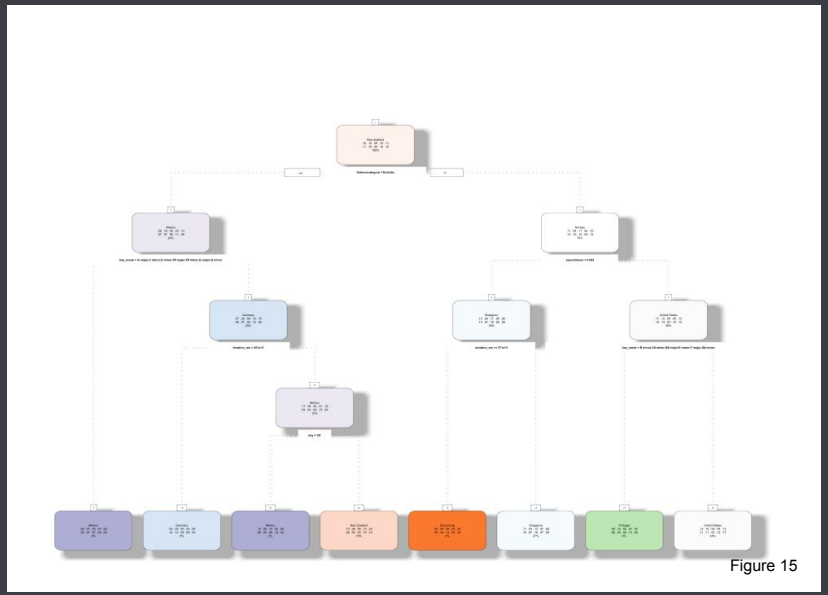

Figure 15

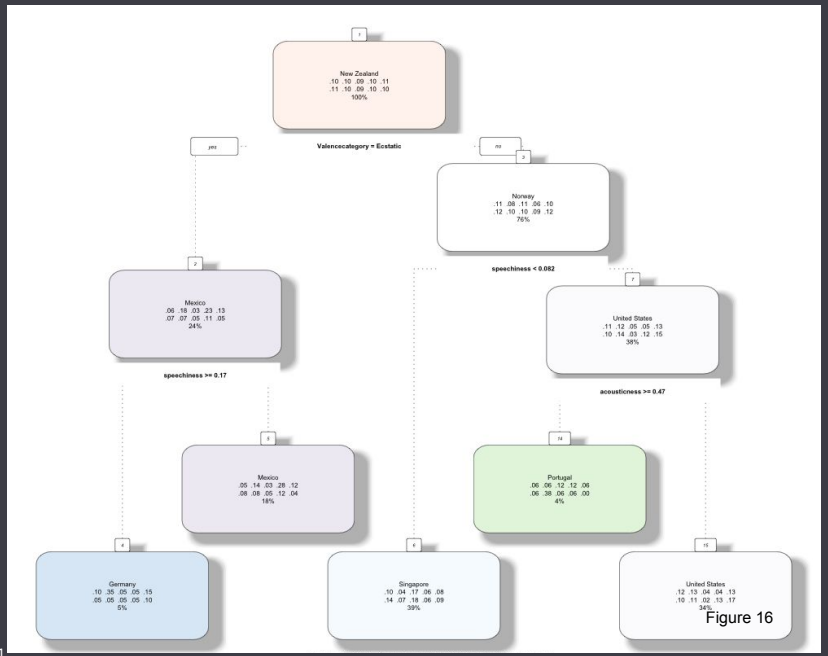

Figure 16

Figure 15: Decision Tree Model 1
Figure 16: Decision Tree Model 2

24.5% Accuracy

17.5% Accuracy

Pruned and targeted to valence, acousticness, speechiness, and instrumentalness

Random Forest

11% Accuracy

Figure 17: RF Histogram
Figure 18: RF Variable Importance

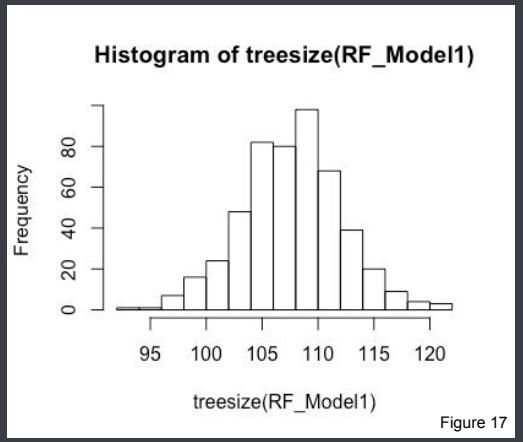

Figure 17

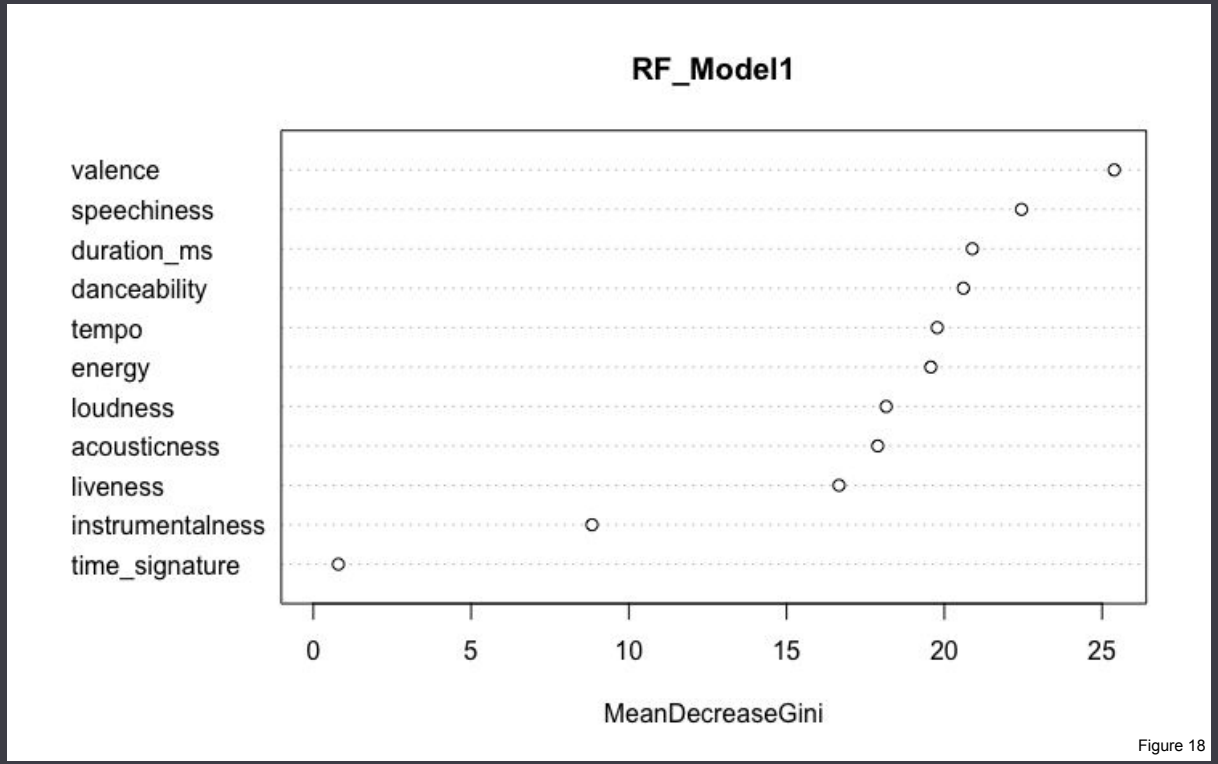

Figure 18

Support Vector Machine

United States	91%
Germany	92%
Mexico	89%
Hong Kong	90%
Singapore	91%
Australia	91%
New Zealand	86%
Norway	88%
United Kingdom	92%
Portugal	90%

Figure 19: SVM US
Figure 20: SVM Germany
Figure 21: SVM Mexico
Figure 22: SVM Hong Kong

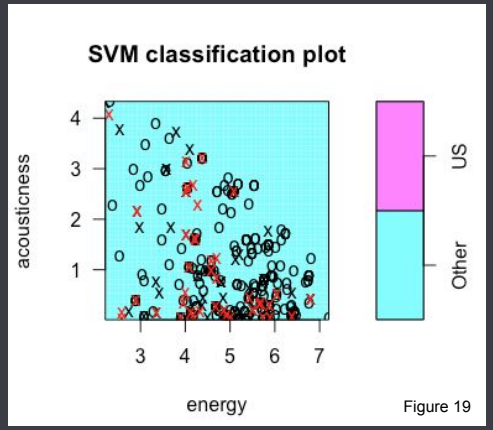

Figure 19

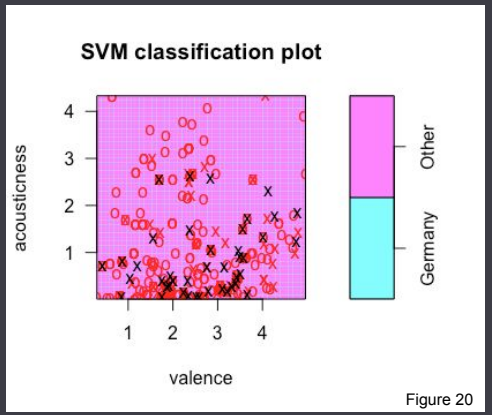

Figure 20

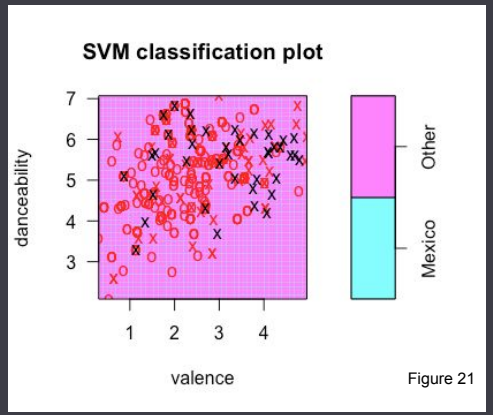

Figure 21

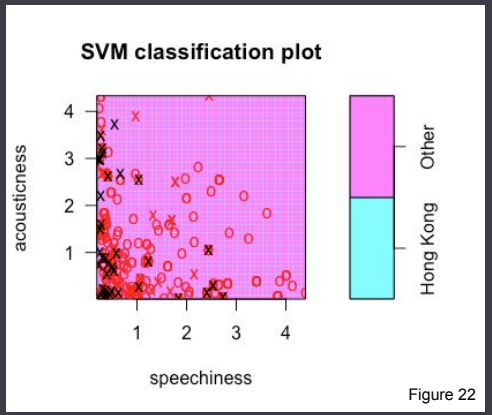

Figure 22

Conclusion

K-Means, KNN, Decision Tree, and Random Forest do not support support the hypothesis

However, after breaking out the countries for “one-against-all” models, SVM was able to most accurately predict classification of songs

Risiko by Bonez MC