Spotify Warped

By: Brian and Luke

Tuning the Research Question

Our main question is, which traits make a song danceable? In other words, if we want to make the most danceable song ever which elements do we need to include?

We initially chose the spotify data set because of its vastness and how through it is. It includes some trivial columns like track name or artist, but it also includes some strange columns like danceability, liveliness, and valiance. These standout measurements made us want to investigate the data set.

Spotify wrapped dropped on the 29th so it's just been front and center in our minds.

Variables

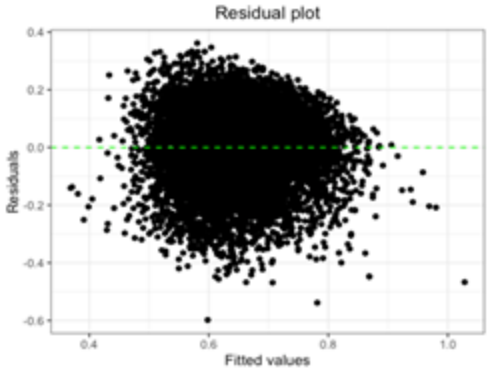
We will be using the following 11 predictors and performing a stepwise selection to see which are the best.

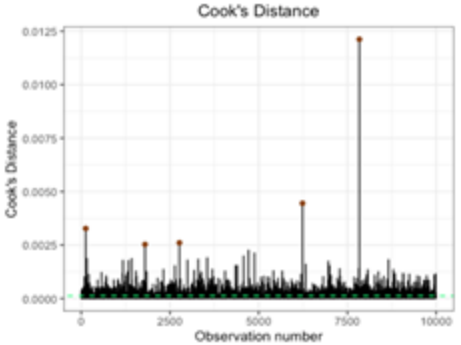
Energy: represents perpetual measure of activity.
Key: Estimated overall key.
Loudness: Overall loudness in decibels.
Mode: Indicates if the track in in a major or minor scale.
Speechiness: Presence of spoken words.
Acousticness: Indicated if the track is acoustic or not.

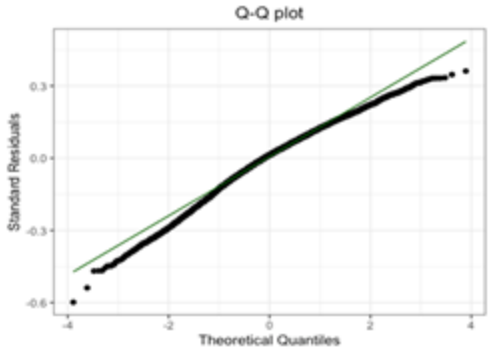
Instrumentalness: Indicates if the track is fully instrumental.
Liveness: Indicates if the track is recorded in front of an audience.
Valence: Describes the musical positiveness.
Tempo: Overall tempo in BPM
duration_ms: Duration of the song im MS

High Notes and Low Notes

- With a stepwise selection of 0.1 significance,only *key* is taken out of the model. Which isn't so exciting because we still have all of our predictors
- We have to assume independence because it is going to be unavoidable. Musicians inspire other musicians, there is no way to know for certain that an artist did/didn't make an impression on another.







Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.041581	Pr > D	<0.0100
Cramer-von Mises	W-Sq	18.66469	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	110.3544	Pr > A-Sq	<0.0050

Now Playing: The Model

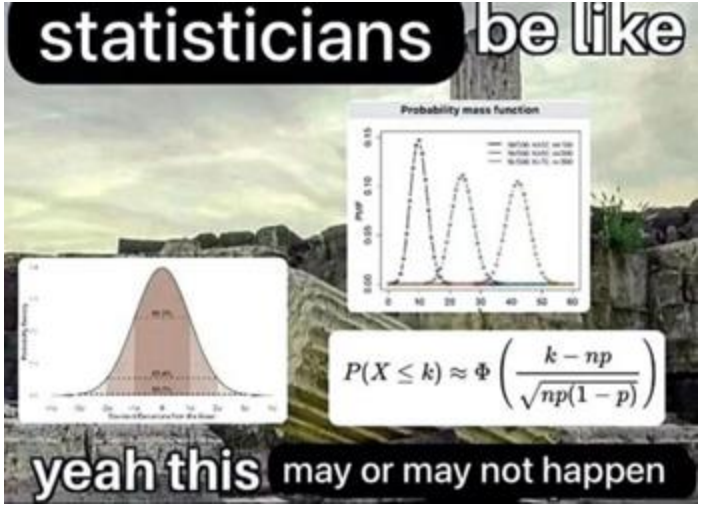
Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	0.9176745072	B	0.00709045	129.42	<.0001
energy	-.2477408806		0.00615541	-40.25	<.0001
loudness	0.0092940347		0.00033382	27.84	<.0001
mode 0	0.0122006912	B	0.00141450	8.63	<.0001
mode 1	0.0000000000	B	.	.	.
speechiness	0.2391707588		0.00701588	34.09	<.0001
acousticness	-.0997403906		0.00381299	-26.16	<.0001
instrumentalness	0.0760176851		0.00329135	23.10	<.0001
liveness	-.0920572867		0.00461429	-19.95	<.0001
valence	0.2277063896		0.00313261	72.69	<.0001
tempo	-.0009218516		0.00002639	-34.94	<.0001
duration_ms	-.0000001527		0.00000001	-12.77	<.0001

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	164.1774552	16.4177455	1022.65	<.0001
Error	32822	526.9279437	0.0160541		
Corrected Total	32832	691.1053989			

R-Square	Coeff Var	Root MSE	danceability Mean
0.237558	19.34869	0.126705	0.654850

Conclusion

Everything is significant but we can explain 24% of what our model is saying, so for now the scientifically created most danceable song remains out of reach. However we would like to keep trying with different sampling methods, song filters, or transformations on the predictors.



However, we are aware of the possibility that this may not be the best move. Perhaps people consider shorter songs to be more danceable. Or something else like so. The mentality of a culture is much more difficult to predict, and a more in-depth analysis would be required to understand this complex effect.