

Top Spotify Songs by Year

Bruno Salgado

I pledge my honor that I have abided by the Stevens Honor System

The topic I chose for this experiment in data visualization is the top songs on Spotify, by year, from 2010 to 2019. This topic speaks volumes (literally) to me specifically because, although I am a Software Engineer by training and the field is a huge passion of mine, I am also a self-taught music producer and I generally have a vast wealth of knowledge surrounding songs and albums.

To create the graphs seen in this report, I first had to start by finding a viable source of data. While Professor Moriarty had suggested a few websites to find this data, including one related to NYC specific data, I took it upon myself to search elsewhere and I stumbled across Kaggle. I recognized the name as it is popular amongst Machine Learning enthusiasts and that is a field I'm interested in learning more about.

I searched for music data and came across this CSV, or comma separated values, file which can be seen in the GitHub repo folder where this project is contained (top10s_spotify_2010-2019.csv). This data a bit unrefined because, for example, the genre titles for each song are broken down into oddly specific subgenres that don't allow a totally clear picture to be painted. For instance, "Love The Way You Lie" by Eminem is considered "detroit hip hop" instead of just "hip hop." Further refinement could have been done to parse out key words such as "hip" and then lumped into its own unified category, but this feature was deemed outside of my skill set for the time being and is something I plan on revisiting in the future.

Finally, the actual relationships I chose to model were:

- Genre frequency, across all years
- Beats per minute (BPM) vs year
- "Danceability" vs loudness (dB)

I chose these specific variables to compare against one another specifically because they paint interesting and insightful pictures into the change of the popular music landscape over time as well as what kind of music really dominated the 2010s which is still something many music analysts are still trying to figure out themselves.

Overall, while this was my first experience in R, it was a rewarding one and I can absolutely see myself using R in my professional life and in general for integration into web applications as well.

Link to original data download:

<https://www.kaggle.com/leonardopena/top-spotify-songs-from-20102019-by-year>

Link to GitHub repository:

<https://github.com/bsalgado98/fe-550-data-visualization/tree/master/ggplot2-assignment>