Now Playing

Presented by:

Bree PiperWilliams, Daniel Rodriguez, Grace Yoo, Kiana Navarre, Pooja Rajasekharan

Research Goal

- Analyze the top tracks over a period of 20 years to identify various trends, find key indicators that make songs top tracks, and showcase the results.

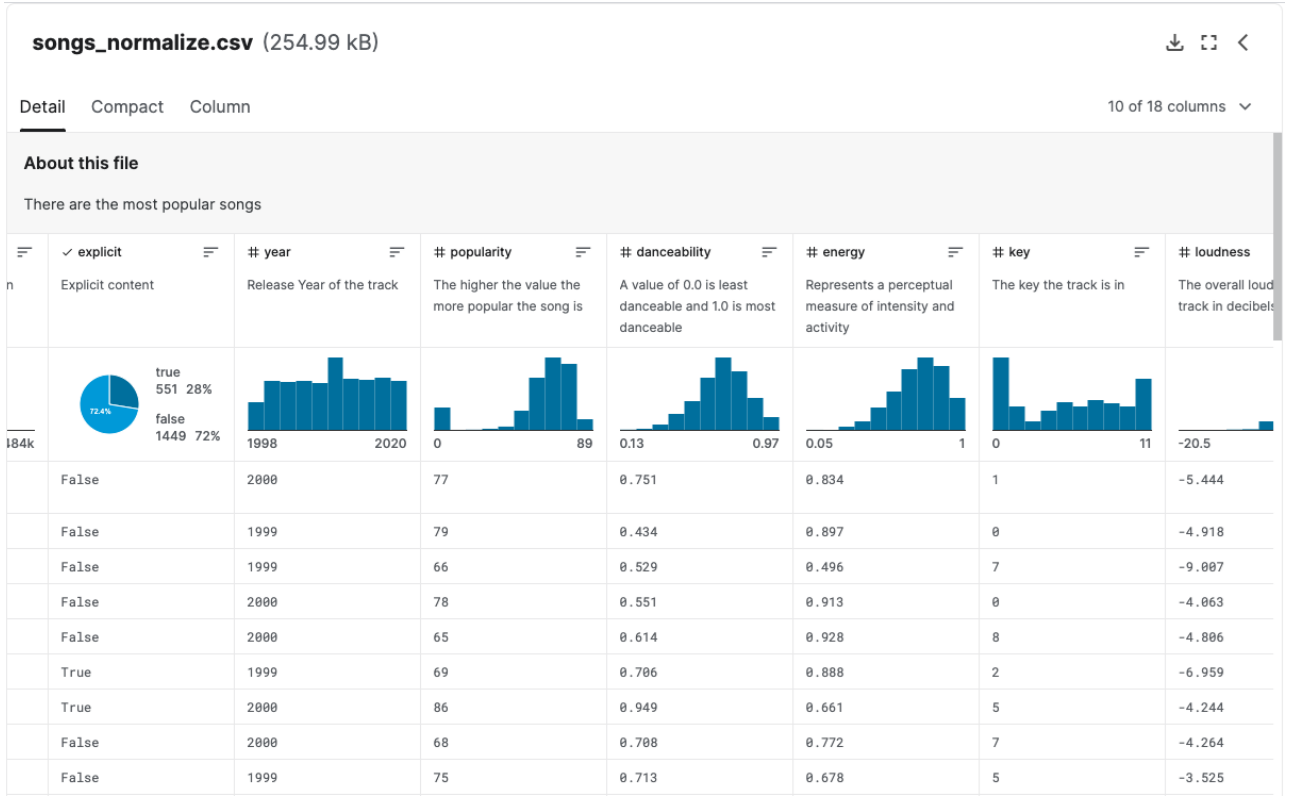
Analysis Performed:
- by Popularity
- by Year
- by Mode
- by Key

Potential Target Audience:
- upcoming artists
- music labels
- music composers

Spotify

The Dataset

- Source: Kaggle
- Top 2,000 songs
- 2000 to 2019



Metrics Analyzed

- Artists
- Genre
- Duration
- Loudness
- Popularity
- Beats per Minute (BPM)
- Valence
- Key Signature
- Mode

Spotify

Data Cleaning

- Used “UTF-8-Sig” encoding to read special characters
- Removed songs outside of date range in Excel
- Removed duplicate values
- Removed songs with unclear genres
- Converted milliseconds to seconds

Spotify

Assumptions About Data

- Data is complete and accurate.
- Metrics were calculated using level of scientific rigor. (e.g. valence)

Analyzing by Popularity

- Home
- Search
- Playlist

- Topic
  - Popular Genres
  - Song Length and Popularity
  - Loudness and Popularity

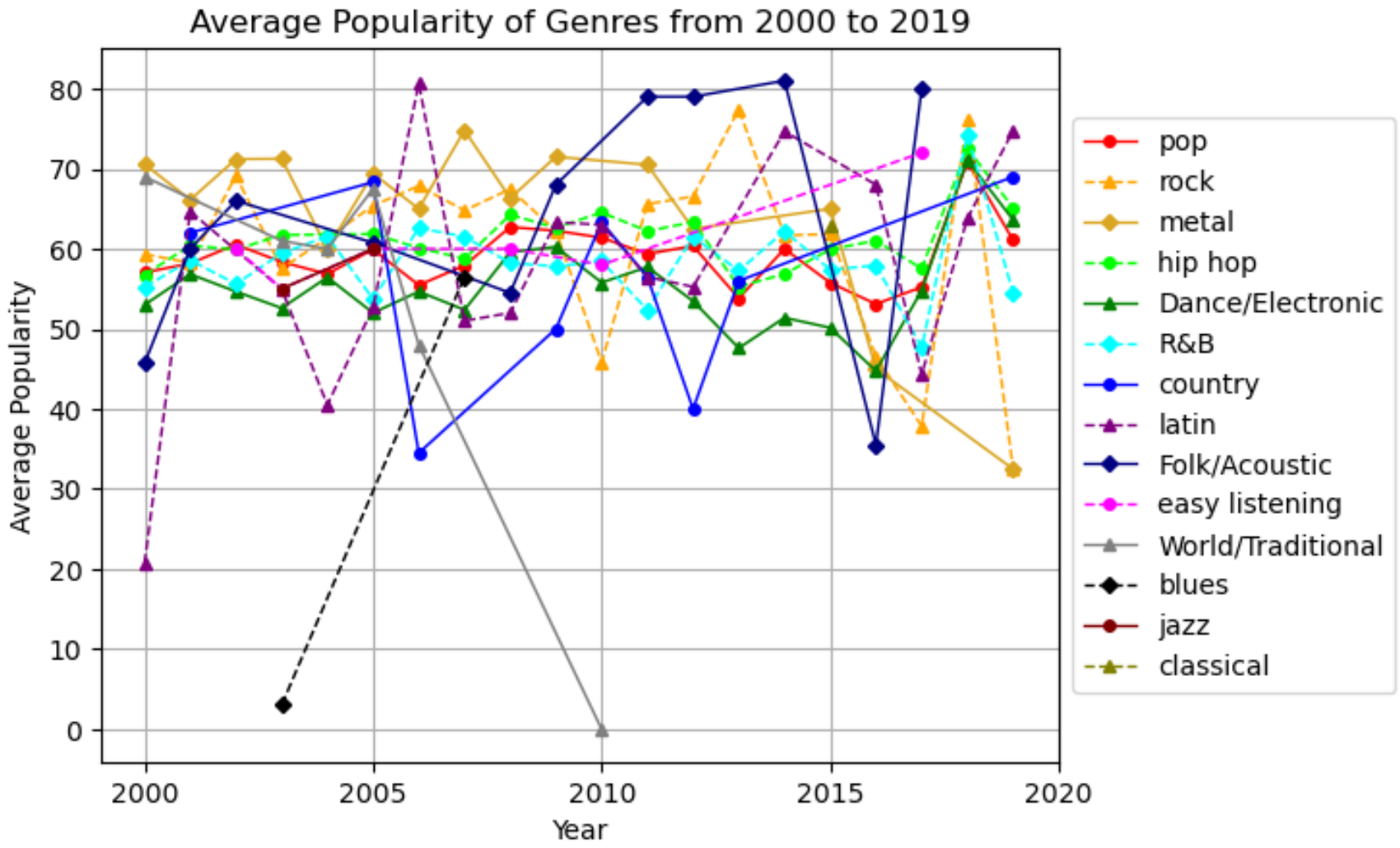
Popular Genre

Song Length vs Popularity

Loudness vs Popularity

Analyzing By Popularity

Year vs Popularity Per Genre



Spotify

Analyzing By Popularity

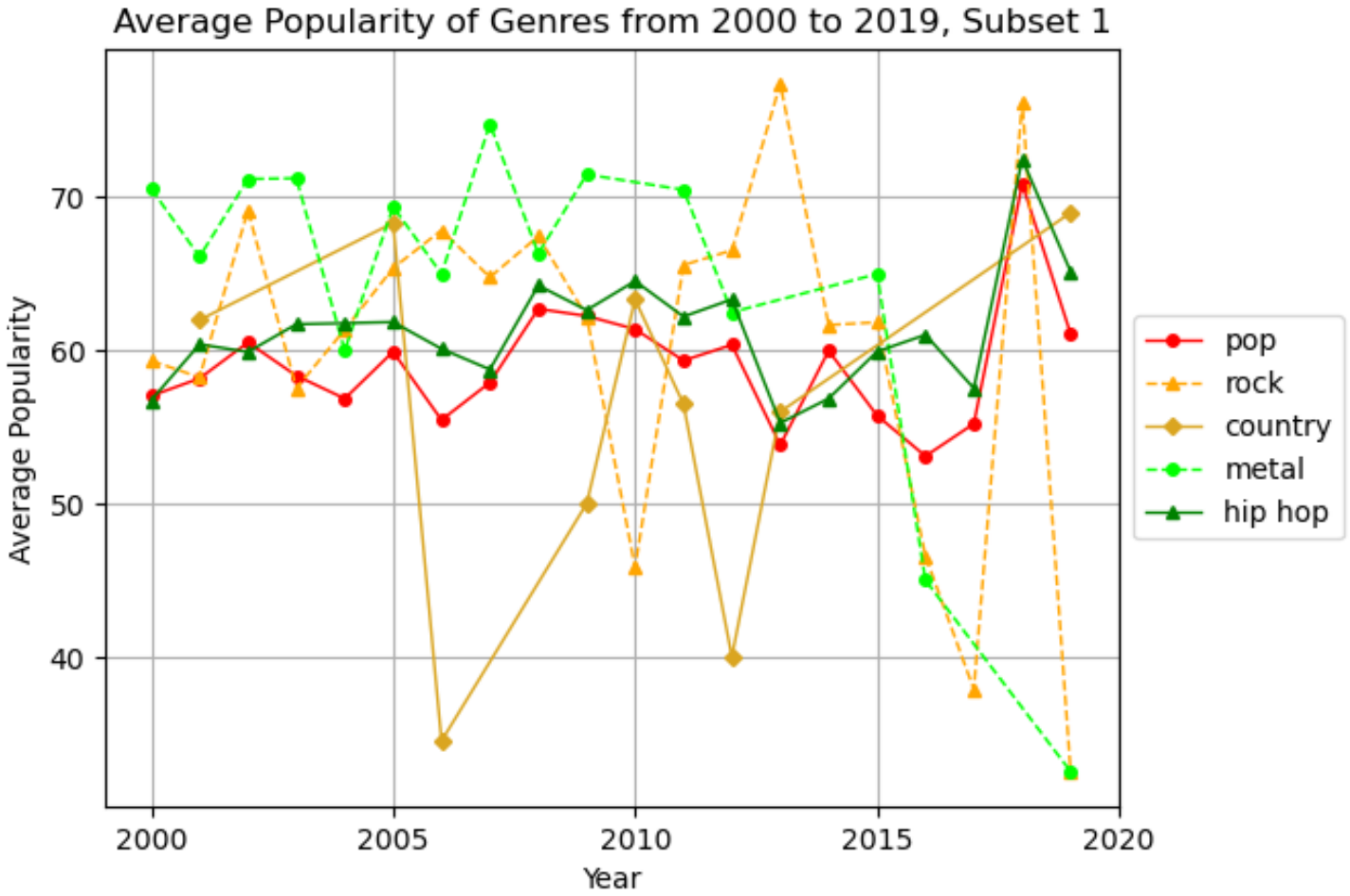
Pop, Rock, Country, Metal and Hip Hop

Hypothesis: Pop is a consistently popular genre in the years of 2000 to 2019.

Null Hypothesis: Pop will not remain above 50% popularity from 2000 to 2019.

Alternative Hypothesis: Pop will remain above 50% popularity from 2000 to 2019.

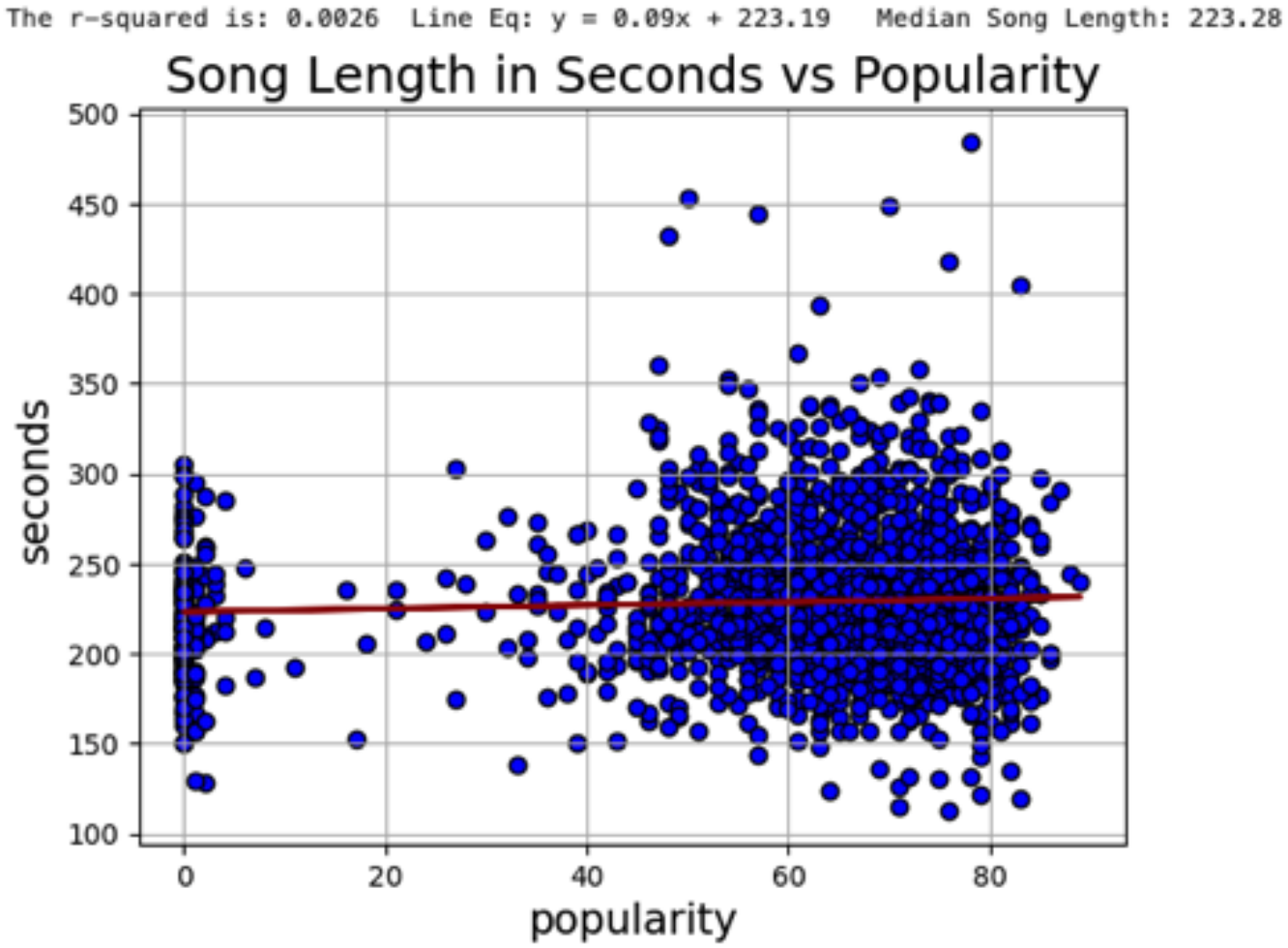
Conclusion: As all values for each year are above 50%, the null hypothesis is rejected.



Spotify

Analyzing By Popularity

Song Length vs Popularity



Hypothesis: If song length can be used as an indicator to determine its popularity.

Null Hypothesis: There is no significant correlation between song length and its popularity.

Alternative Hypothesis: Song length is an indicator of song popularity.

Observation: Accept the Null Hypothesis. There is no significant correlation between the song's length and popularity.

The median song length of 223.28 seconds cannot be used as an indicator of song popularity.

R- value: .0026

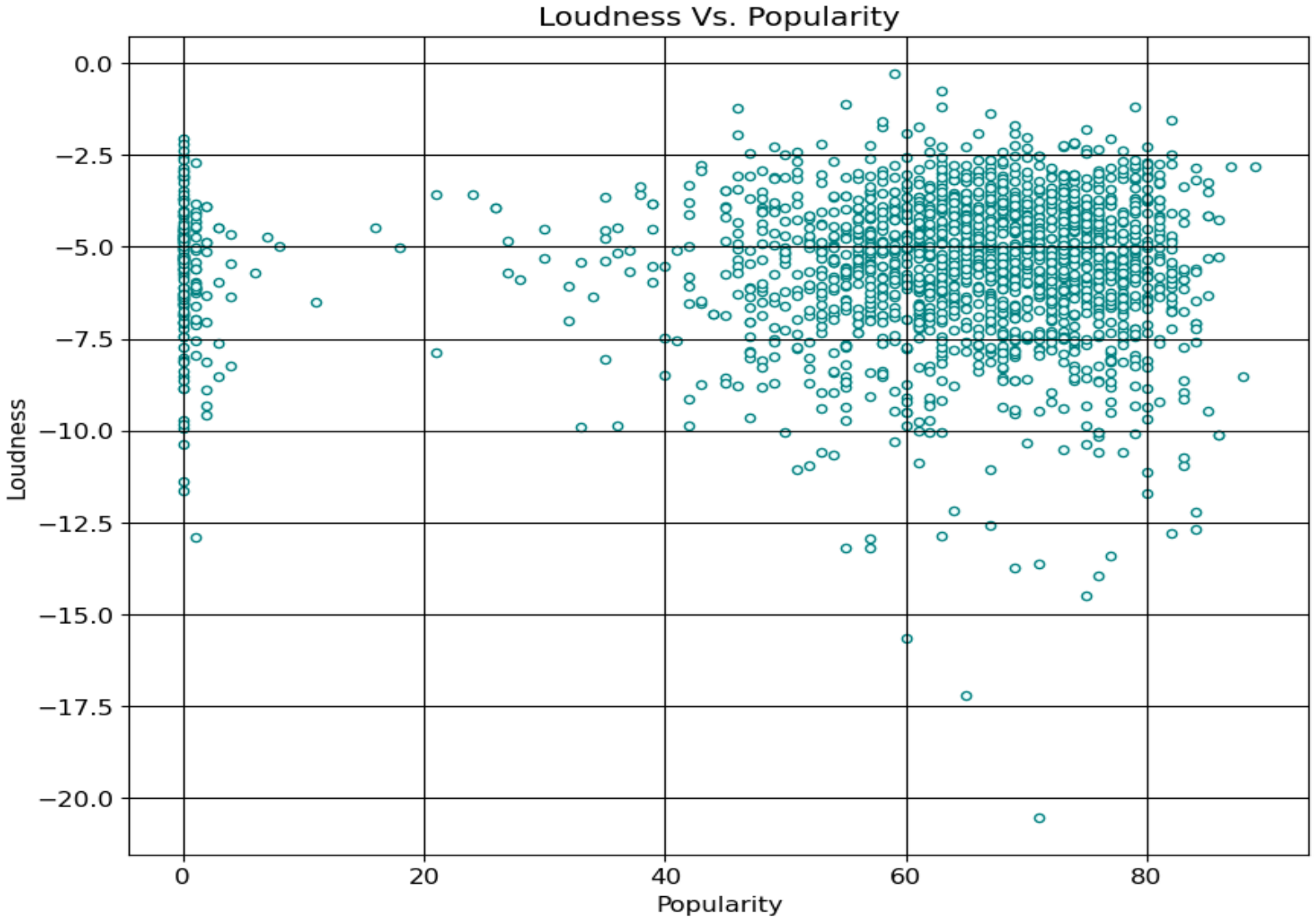
Linear Regression: Positive Correlation

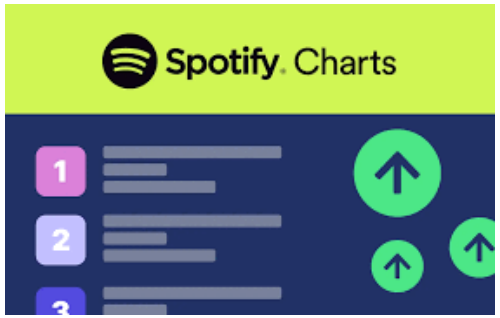
y = .09 + 223.19

Spotify

Analyzing By Popularity

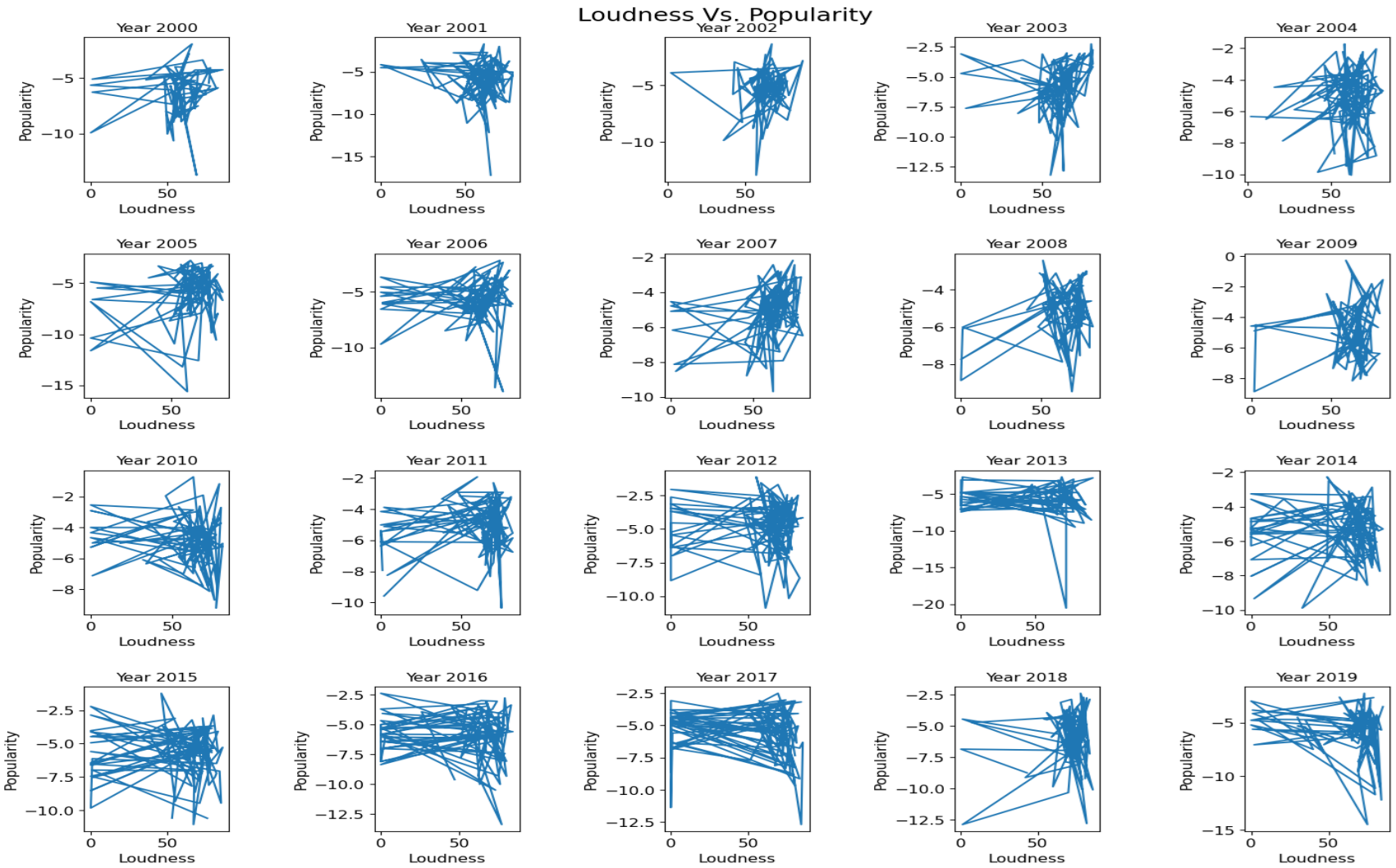
Loudness vs Popularity





Spotify

Analyzing By Popularity
Loudness vs Popularity by The Year



Hypothesis: The loudness of a song determines the popularity of the song

Alternative Hypothesis: The loudest songs are the most popular.

Null Hypothesis: There is no correlation between loudness and popularity.

Conclusion: Data supports the Null Hypothesis. There is no significant correlation between the loudness of a song and its popularity. Separating the data by year confirms there is no relationship.

Spotify

Analyzing by Year

Home

Search

Playlist

Topic

- Loudness vs Year
- BPM vs Year
- Valence vs Year
- Duration vs Year

Loudness vs Year

BPM vs Year

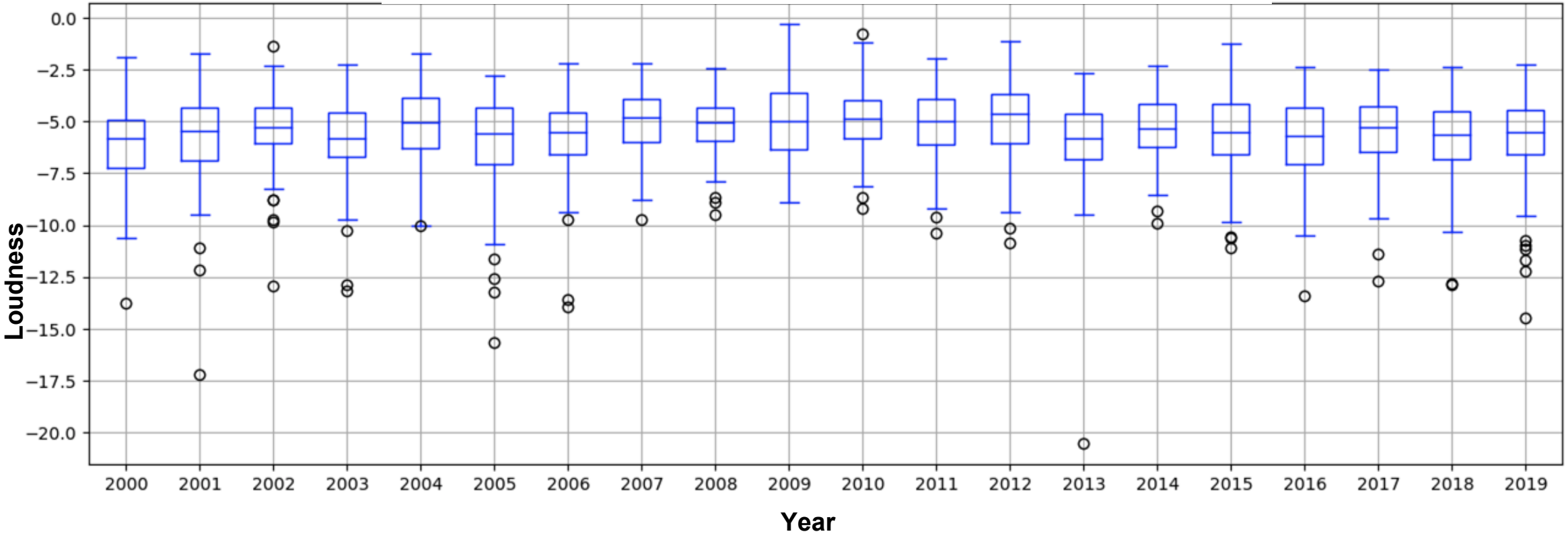
Valence vs Year

Duration vs Year

Analyzing By Year

Loudness vs Year

Boxplot Chart for Loudness vs Year

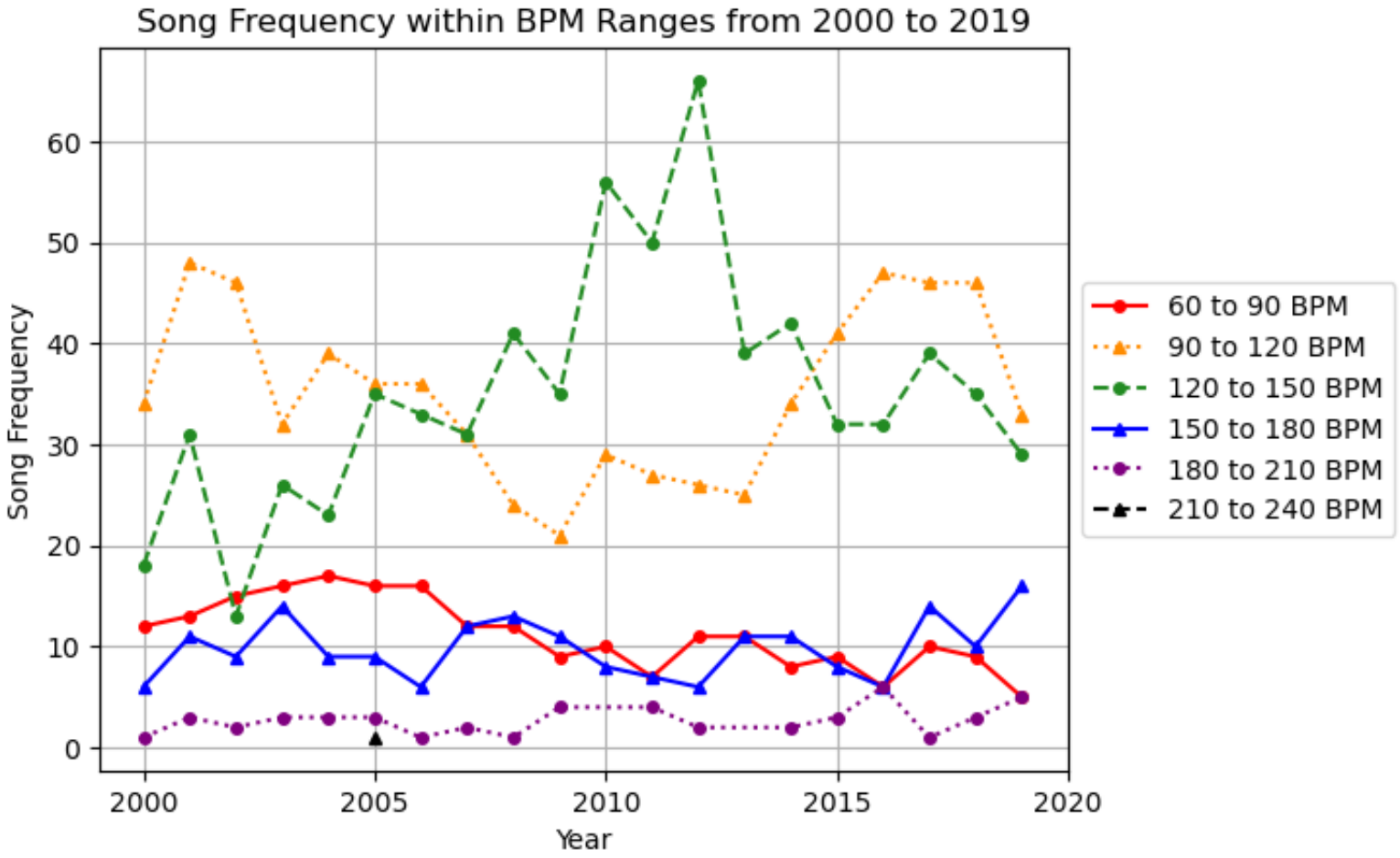


Mean: -5.27 **Median**: -5.497 **P.Stdev**: 1.928

Hypothesis: If the year a song became popular on Spotify have any influence on the loudness of the song
Null Hypothesis: The year the song became popular does not influence the loudness of the song.
Alternative Hypothesis: The year the song became popular does influence the loudness of the song.
Observation: No evidence shows the year the song became popular influences the loudness of the song. The song's loudness stayed consistent with P.Stdev of 1.928 and a Median of -5.497.

Analyzing By Year

BPM vs Year



Dance/Electronic music was popular in the 2010s. (e.g. Turn Down for What by Lil Jon)

Assumption is that the typical BPM for Dance/Electronic music average around 125 BPM.

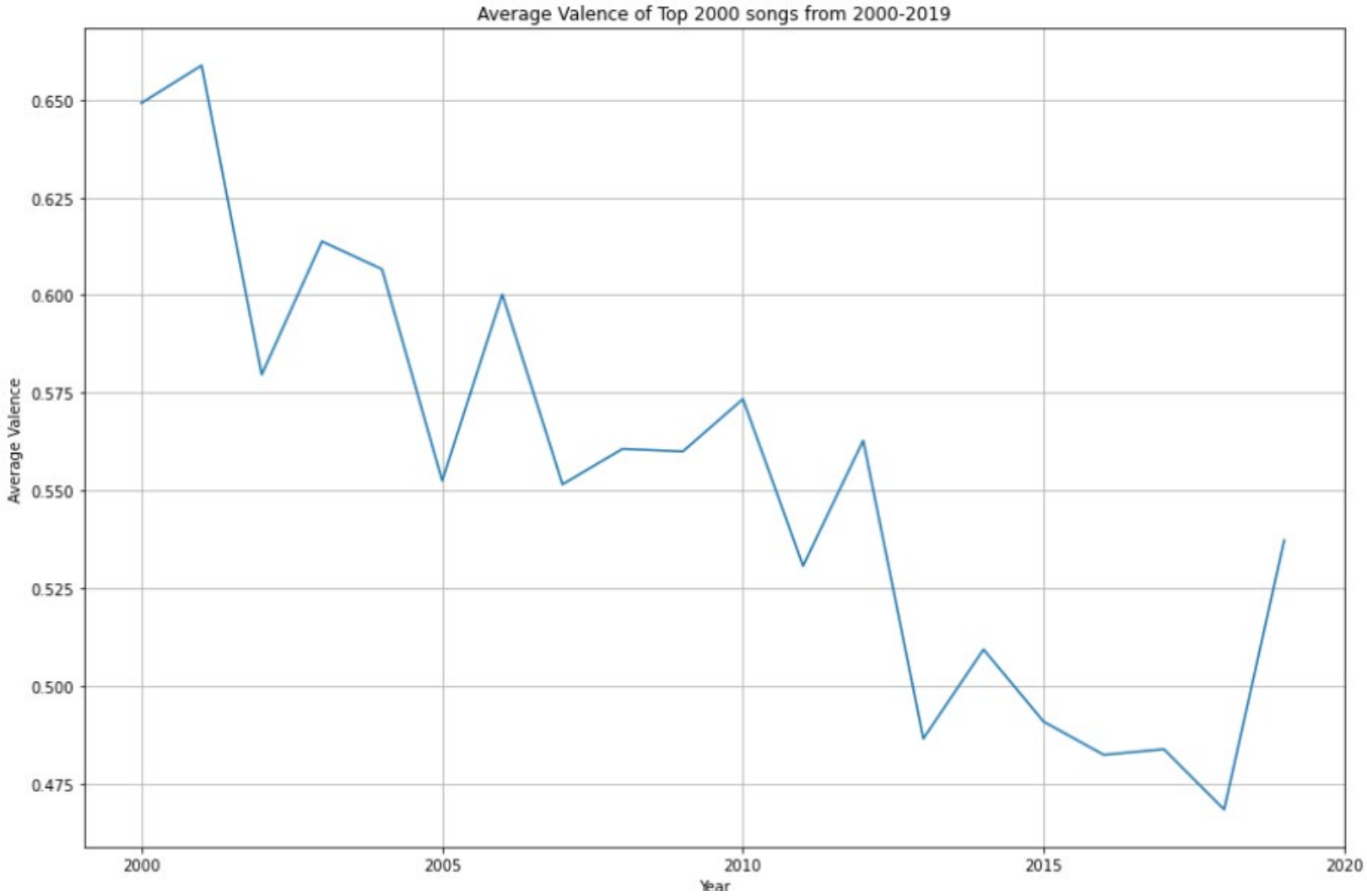
Null Hypothesis: Songs with 120-150 BPM should not occur more than 20 times every year in 2000-2019.

Alternative Hypothesis: Songs with 120-150 BPM should occur more than 20 times every year in 2000-2019.

Spotify

Analyzing By Year

Valence vs Year



Null Hypothesis: Average valence of top songs stayed consistent between the years 2000 - 2019.

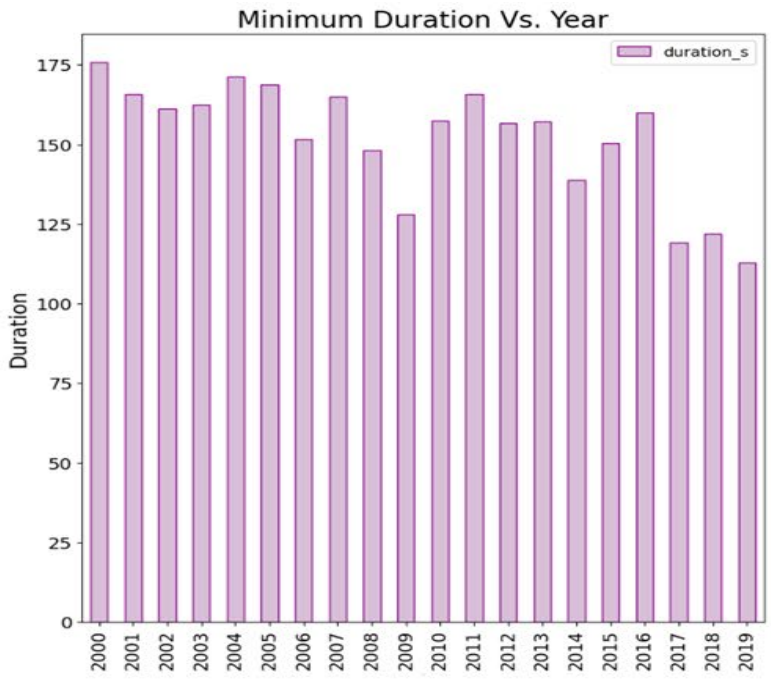
Alternative Hypothesis: The average valence of top songs changed between the years 2000 - 2019.

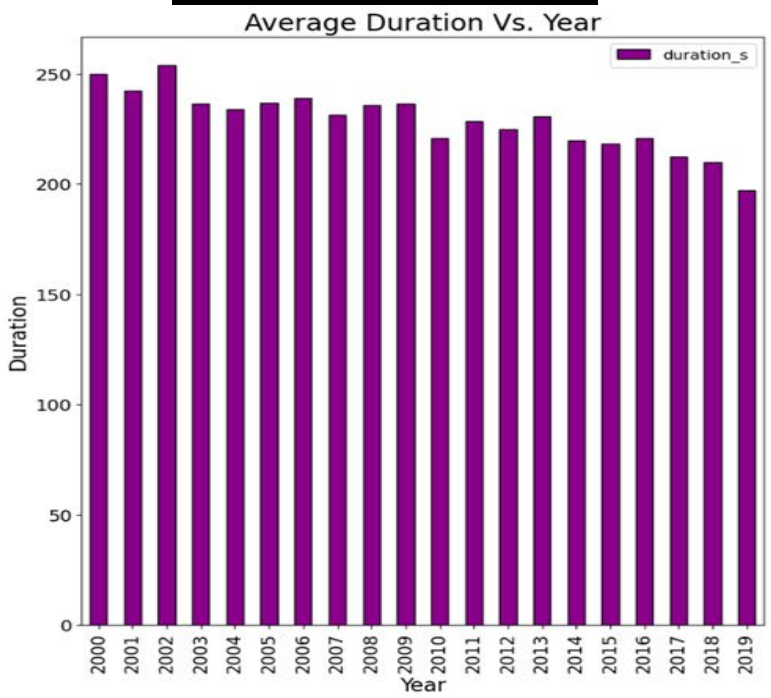
Conclusion: Our data supports the rejection of the null hypothesis. The trendline suggests that the average valence over the last 20 years has been steadily decreasing.

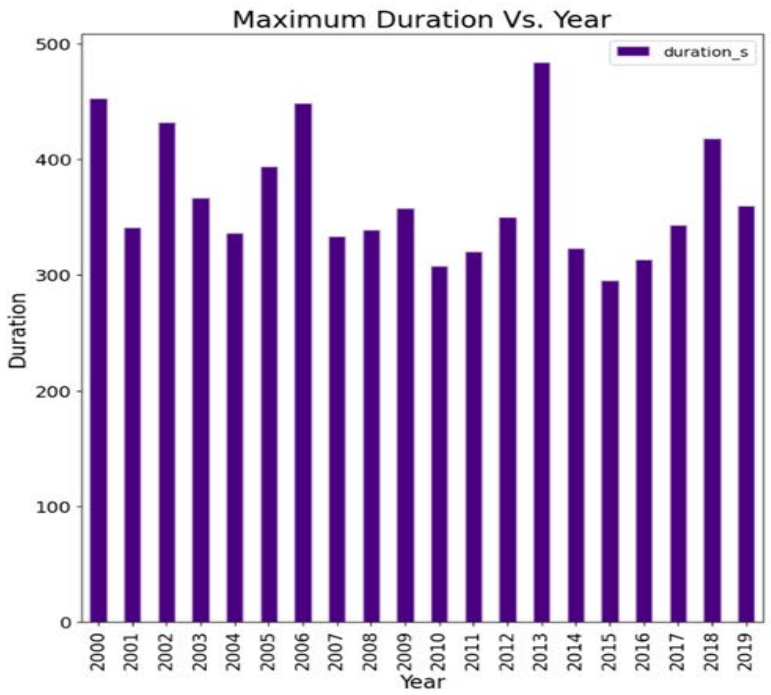
Spotify

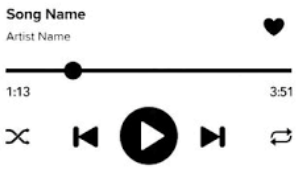
Analyzing By Year

Duration vs Year









Hypothesis: The year determines the duration of the song.

Alternative Hypothesis: The song duration will decrease over time.

Null Hypothesis: There is no correlation between the year and the duration of songs.

Conclusion: The data confirms that the hypothesis is correct however it is not a strong correlation. Over time the duration of songs have decreased with 2019 having the lowest duration minimum and lowest average duration.

Spotify

Home

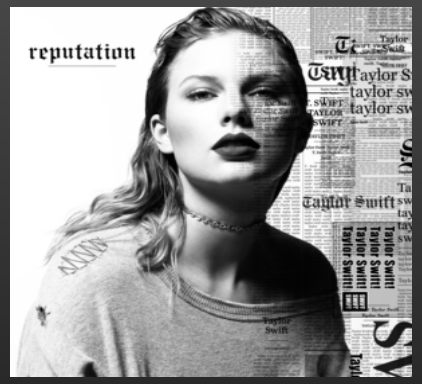
Search

Playlist

Topic

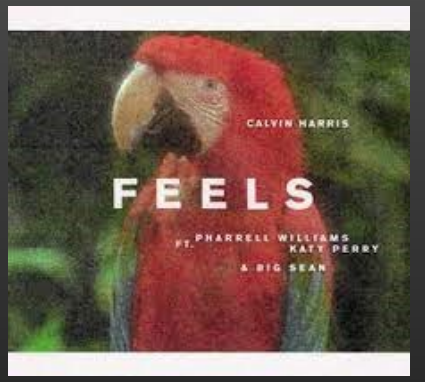
- Key/Mode vs Popularity
- Mode vs Year
- Key vs Artist
- Key vs Year

Analyzing by Key/Mode



Key/Mode vs Popularity

Mode vs Year

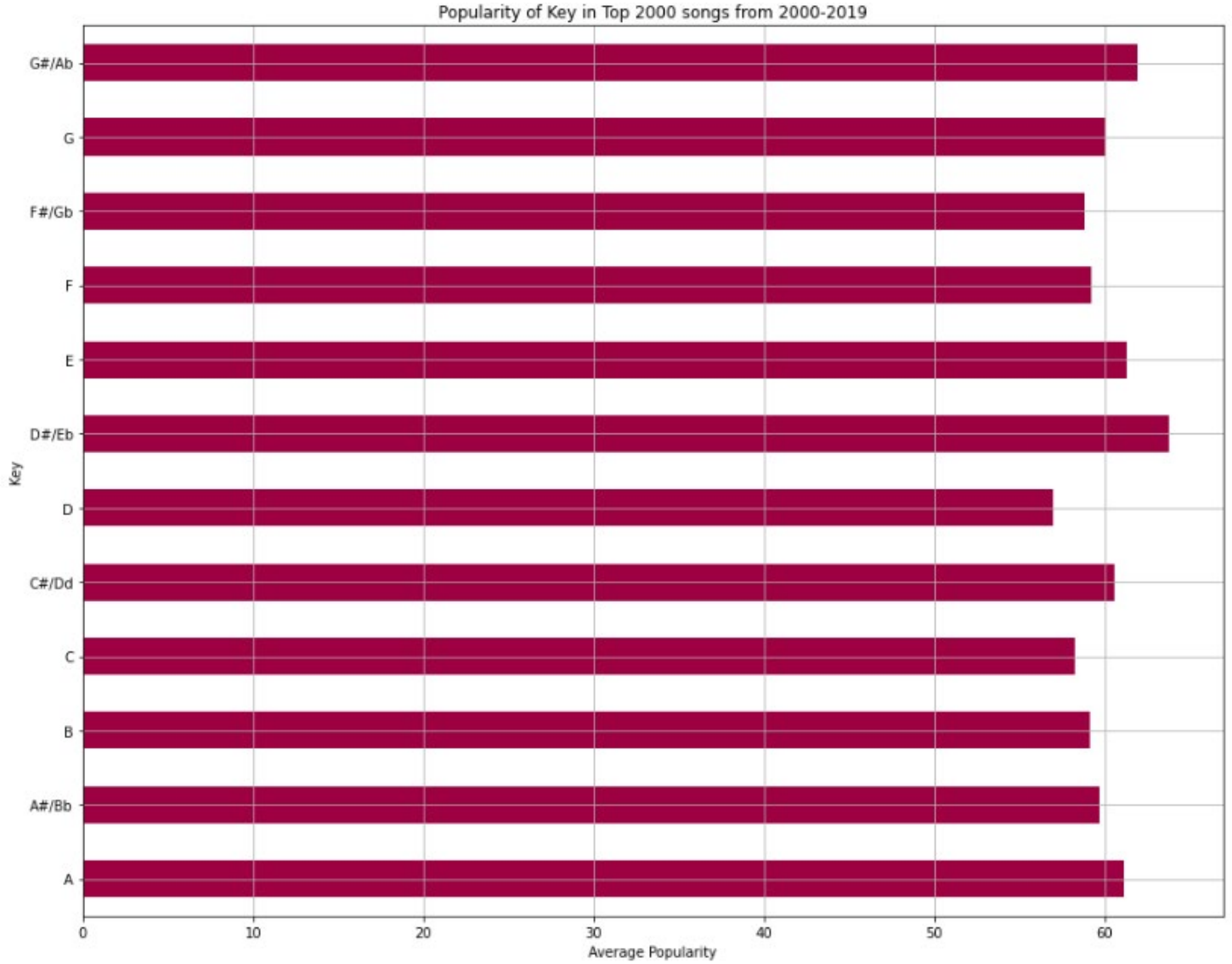


Key vs Artist

Key vs Year

Analyzing By Key

Key vs Popularity



Null Hypothesis: There is no significant difference in the distribution of average popularity across musical key.

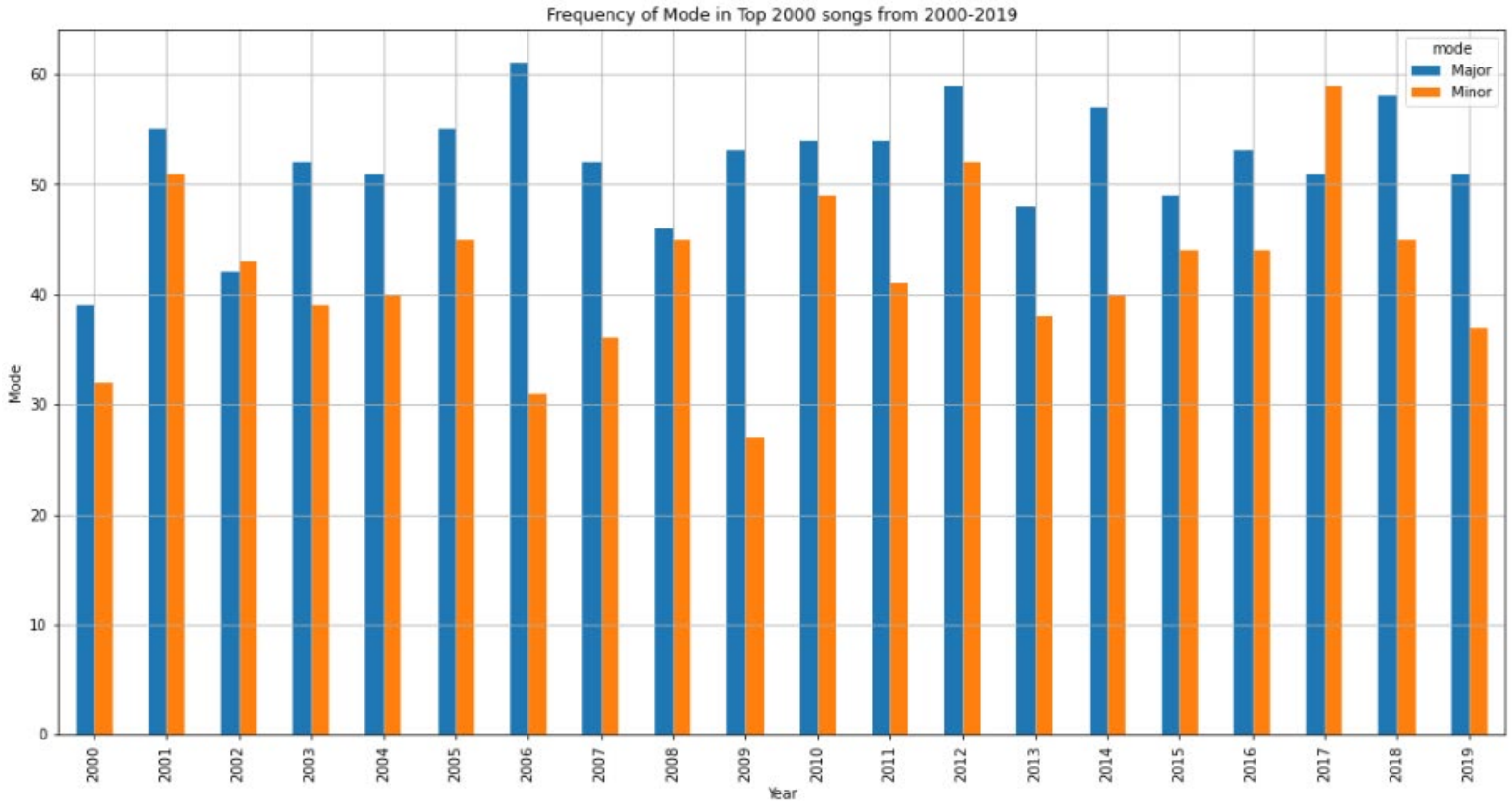
Alternative Hypothesis: There is a significant difference in the distribution of average popularity across musical key.

Conclusion: We cannot reject the null hypothesis; P-Value = 0.956465621

Spotify

Analyzing by Mode

Mode vs Year



Null Hypothesis: Top 2000 songs between 2000 and 2019 have an even number of records in both the major and minor modalities.

Alternative Hypothesis:

Top 2000 songs between 2000 and 2019 do not have an even number of records in both the major and minor modalities.

Conclusion: The data provides evidence to reject our null hypothesis. There appear to be consistently more top songs in the “Major” modality from 2000 to 2019.

Spotify

Analyzing By Key: Key vs. Year

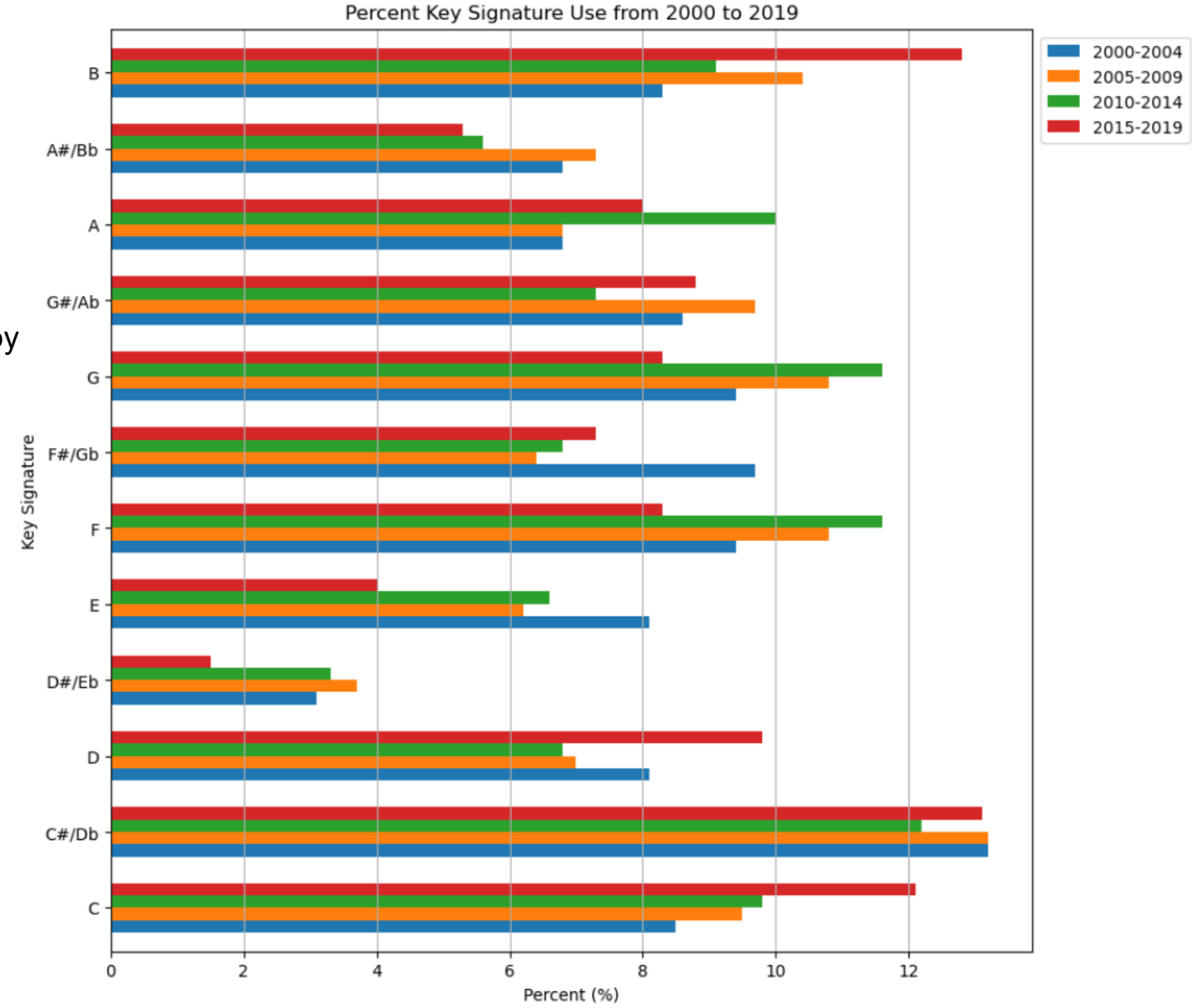
Hypothesis: Key signature usage in the top spotify tracks remains consistent throughout the past 20 years.

Null Hypothesis: There is no relationship between year and key signature, therefore key signature usage is not dependent or influenced by year.

Alternative Hypothesis: Key signature usage varies significantly by year.

Observations:

- D#/Eb key is used less than all other keys with an average usage percentage of 3.4% for 2000-2014 and a usage percentage of 1.5% for 2015-2019.
- C key has been steadily increasing in usage from 2000 to 2019. Starting at a usage of 8.5% for the year range 2000-2004 and increasing to a usage of 12.1% for 2015-2019.
- The key of A#/Bb is decreasing in usage with a usage of 7.3% for the year range 2005-2009 and decreasing to a usage of 5.3% for 2015-2019.
- The key of F#/Gb has decreased in use from a usage of 9.7% for 2000 to 2004 to a average usage of 6.8% for the remaining 15 years.
- The key of C#/Db has consistently been used key throughout 2000 to 2019 with an average usage of 12.9% for all 20 years.



Conclusion: A Chi-Squared analysis with alpha = 0.05, results in p-value = 0.199.
p-value > alpha, so the null hypothesis holds true and the alternative hypothesis is rejected.

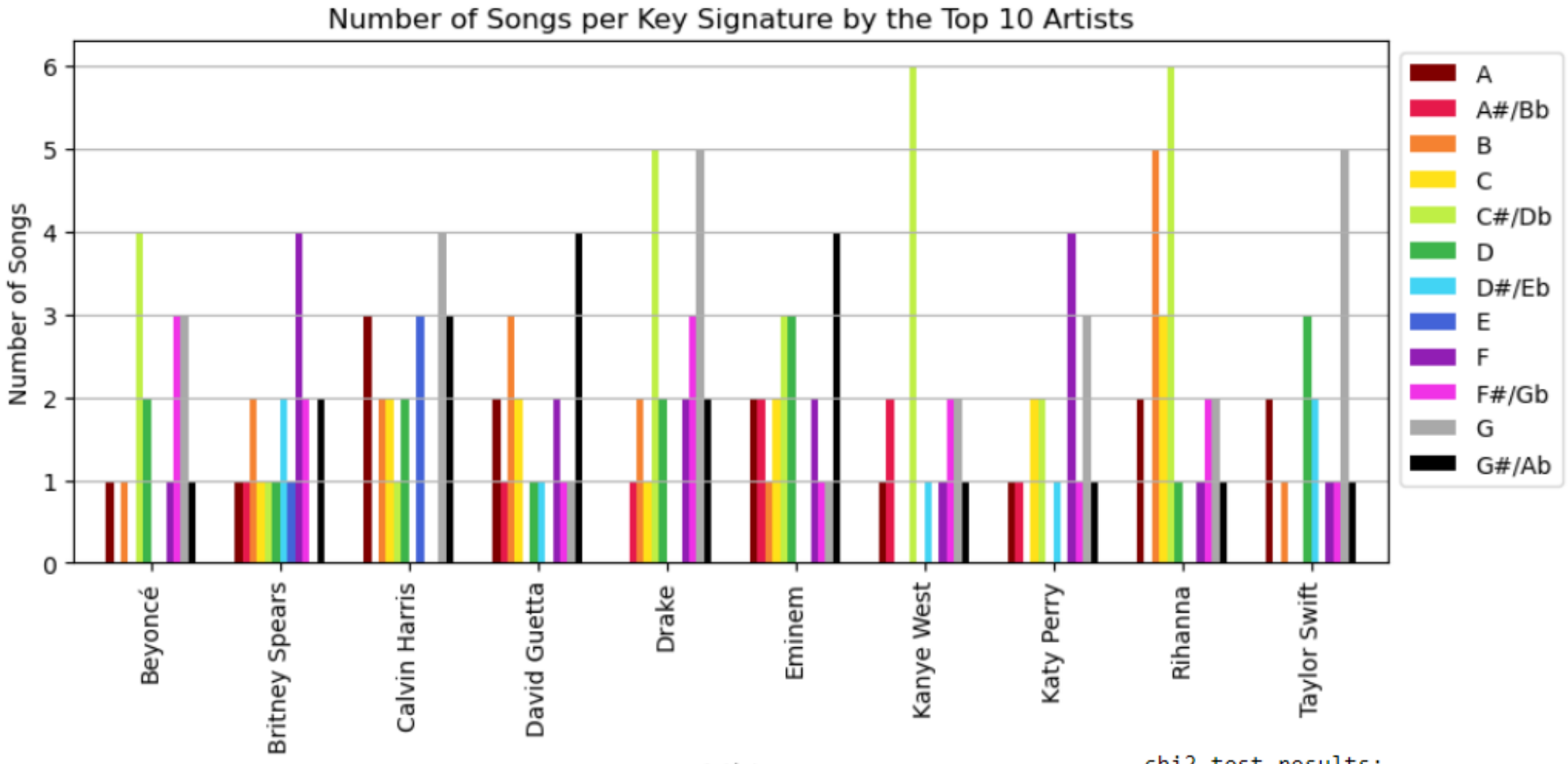
```
chi2 test results:
p value is 0.19929633122313042
Null hypothesis holds true
```

Analyzing By Key:
Key vs Artist

Hypothesis: Artists will tend to favor certain keys over others, and will therefore use said key for the majority of their songs.

Null Hypothesis: There is no relation between artist and key signature.

Alternative Hypothesis: There is a significant relationship between artist and key signature.




```
chi2 test results:
p value is 0.16108563855645297
Null hypothesis holds true
```

Observations:

- C#/Db key appears to be favored for Beyonce, Kanye West, and Rihanna.
- G key appears to be favored by Calvin Harris and Taylor Swift.
- Key E is only used by Calvin Harris (in this list of artists).
- Britney Spears has songs in all keys except G.

Conclusions:

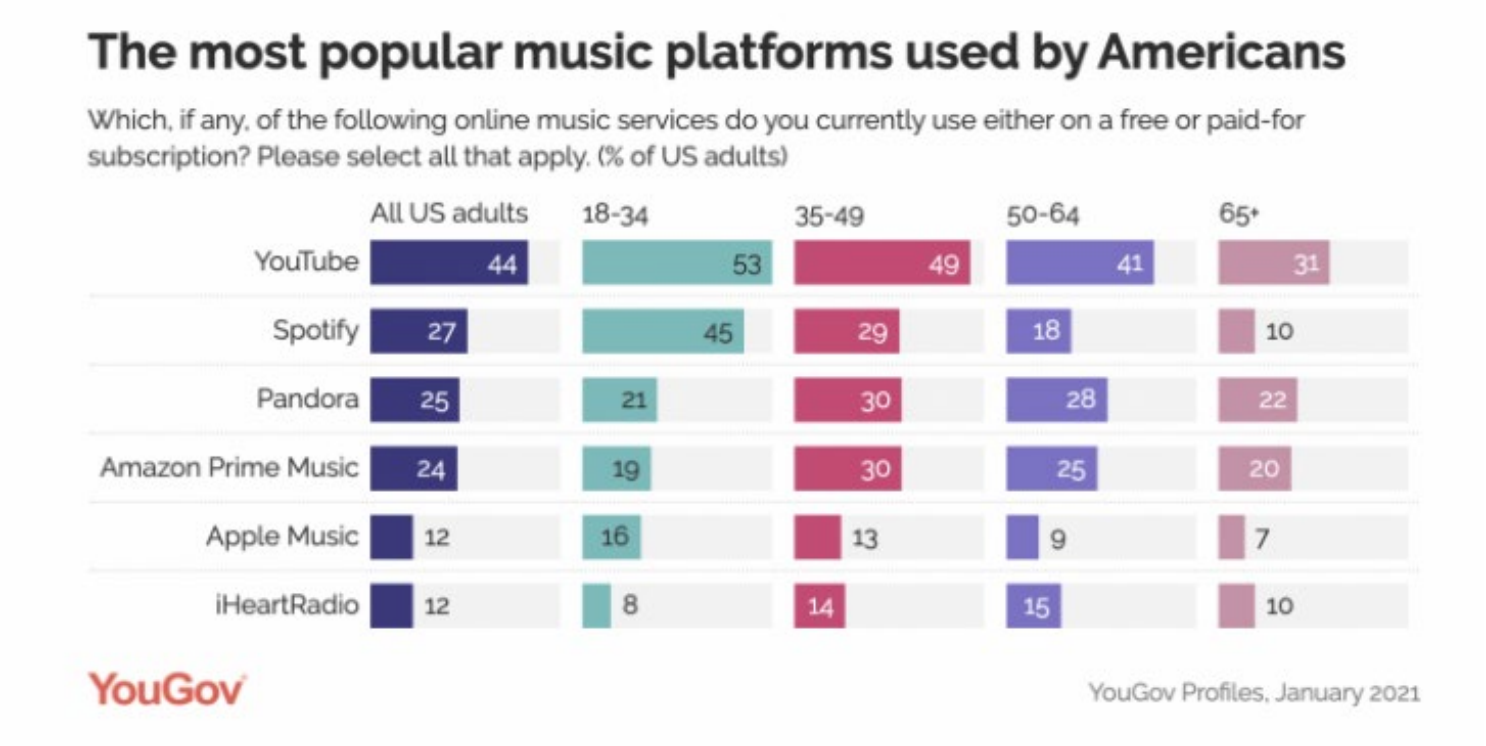
A Chi-Squared analysis with alpha = 0.05, results in p-value = 0.161.

p-value > alpha, so the null hypothesis holds true and the alternative hypothesis is rejected.

Spotify

Challenges

- Simplifying data
- Making graphs readable
- Cleaning data
- Github
- Genre mis-match
- Assumptions about data



Next Steps

- Comparing data to other streaming services
- Include current data
- Include age demographics

Spotify

Summary

- Popularity
  - Certain genres were more popular than others.
- Year
  - Analyzing by year resulted in noticeable trends.
- Key/Mode
  - The key of D#/E flat was less commonly used.
  - Popular songs were more commonly in a major key rather than a minor key.

Spotify®

Thank you!