Billboard 200 Data Analysis

*Taylor Feagin and Sarah Gaines*

# Introduction

The intent of this project is to study the attributes of popular music albums, to see if there is a pattern in what makes an album a hit. Based on this pattern, we hope to propose what musical features can make an album more successful.

Additionally we looked for patterns over time, to see if music preferences changed overtime. (just an idea to try, don’t have to keep this if it doesn’t work)

For the purpose of this project, popularity is defined by an album’s ranking on the Billboard 200. “Billboard 200 is a record chart ranking the 200 most popular music albums… in the United States” and is published weekly (reference 1).

Table of Contents

[Introduction 1](#_Toc25919500)

[Data 2](#_Toc25919501)

[Data Visualization and Results 4](#_Toc25919502)

[Conclusions 4](#_Toc25919503)

[References 4](#_Toc25919504)

List of Figures

Figure 1:

List of Tables

Table 1: “albums” column description

Table 2: “acoustic\_features” column description

# Data

The data used in this project was obtained from Kaggle (reference 2) under the name “The Billboard 200 acoustic data”. This data contains info on the Billboard 200 albums from 1963-2019. It consists of two tables within an SQLite database:

* “albums” lists all Billboard 200 albums for every week between 1963-2019. It has a unique row for each instance an album appeared on the charts. There are 574,000 rows total. The columns are as follows:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| id | int64 | Where available from Spotify and not null in the table |
| date | object | Week of the chart |
| artist | object | Artist name |
| album | object | Album name |
| rank | object | Album’s place in the charts |
| length | float64 | Number of tracks |
| track\_length | float64 | Length of the album in milliseconds |

\*descriptions from (reference 2)

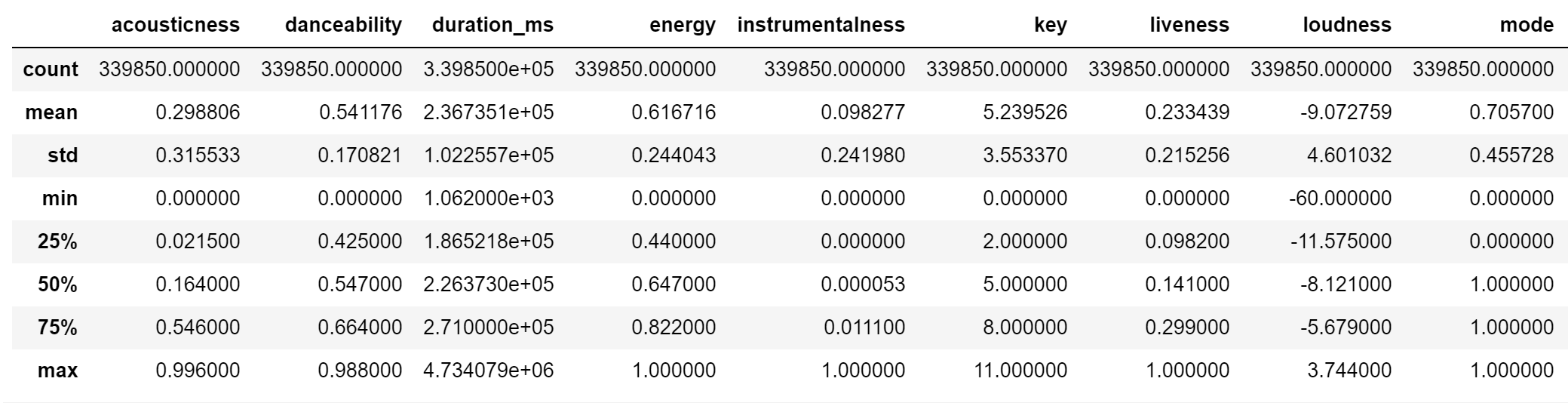
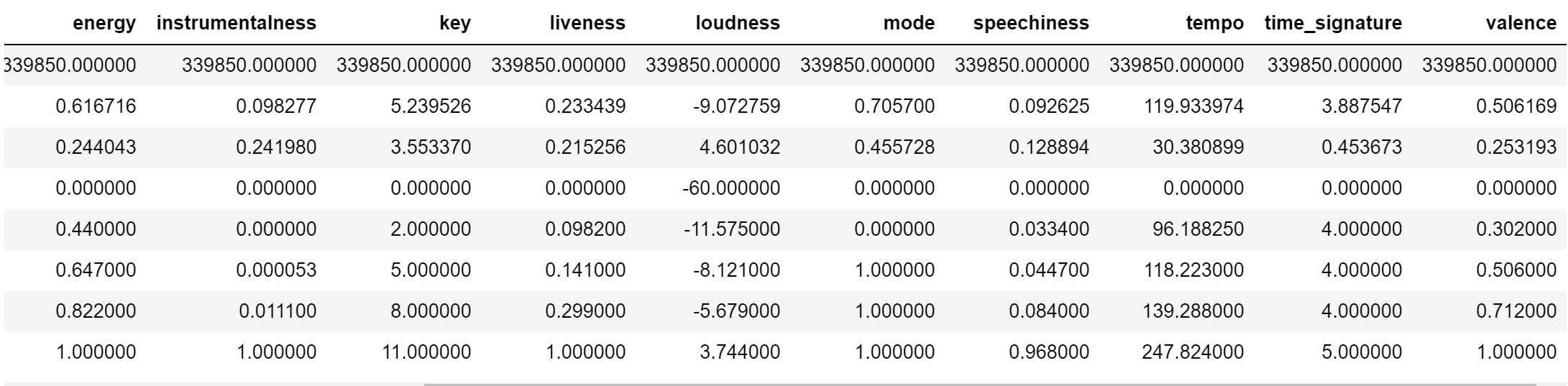
* “acoustic\_features” provides the Spotify EchoNest acoustic data of every track within all the albums on the Billboard 200 from 1963-2019. It has a unique row per song in every albums. There are 340,000 rows total and it contains no null values. The description of each column can be found at (reference 3).

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| id | object | Track ID on Spotify |
| song | object | Track name |
| album | object | Album name |
| artist | object | Artist name |
| acousticness | float64 | A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic. |
| danceability | float64 | how suitable a track is for dancing based on a combination of tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable. |
| duration\_ms | float64 | Duration in milliseconds |
| energy | float64 | A measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy. |
| instrumentalness | float64 | Predicts whether a track contains no vocals. The closer the value is to 1.0, the greater likelihood the track contains no vocal content. |
| key | float64 | The key the track is in. Integers map to pitches using standard Pitch Class notation. |
| liveness | float64 | Detects the presence of an audience in the recording. Higher values represent an increased probability that the track was performed live. |
| loudness | float64 | The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typical range between -60 and 0 db. |
| mode | float64 | Indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0. |
| speechiness | float64 | Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording, the closer to 1.0 the attribute value. |
| tempo | float64 | The overall estimated tempo of a track in beats per minute. |
| time\_signature | float64 | An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure). |
| valence | float64 | A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive, while tracks with low valence sound more negative. |
| album\_id | object | Album ID on Spotify |
| date | object | Release date of the album |

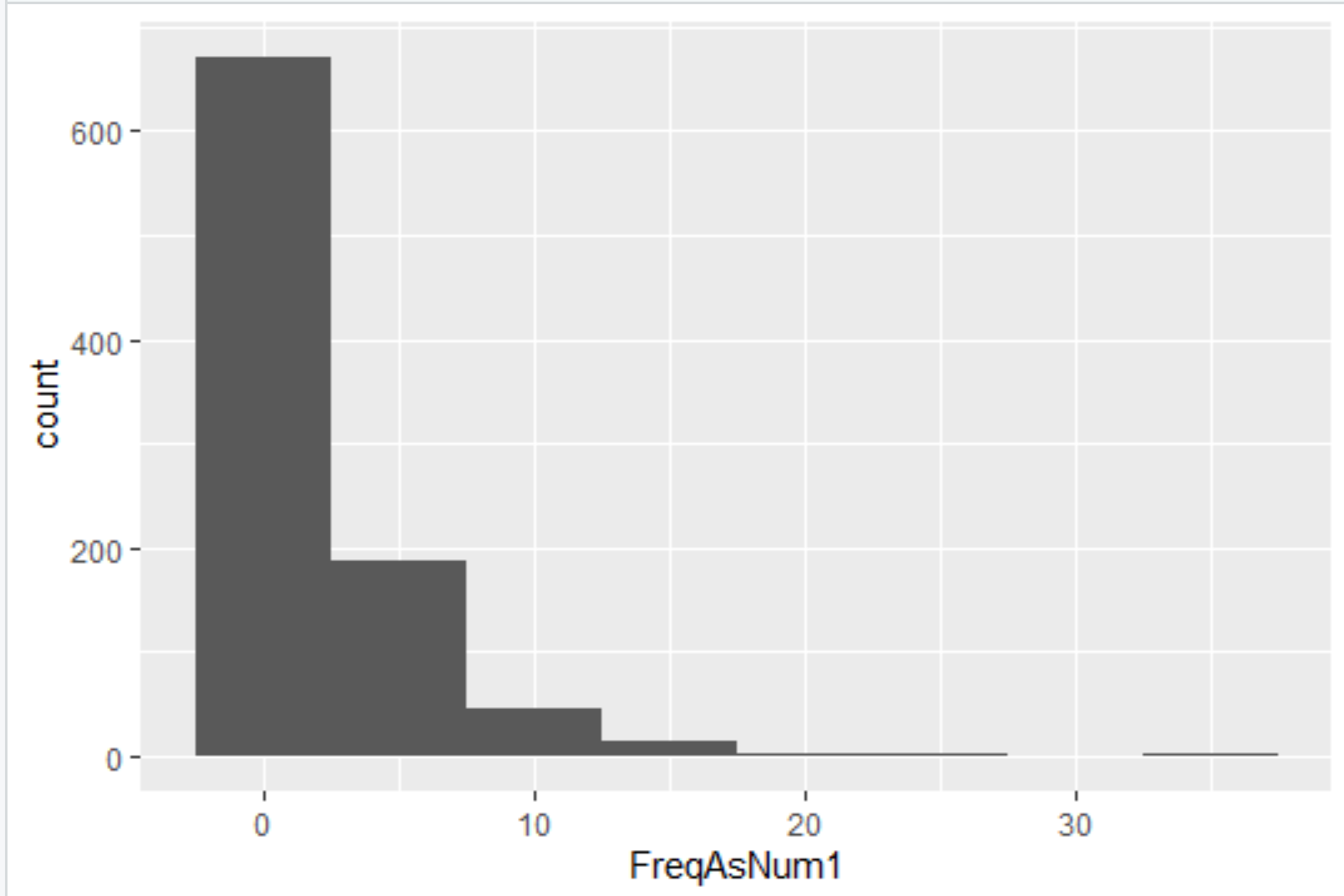
\*description from ref 2 and ref 3

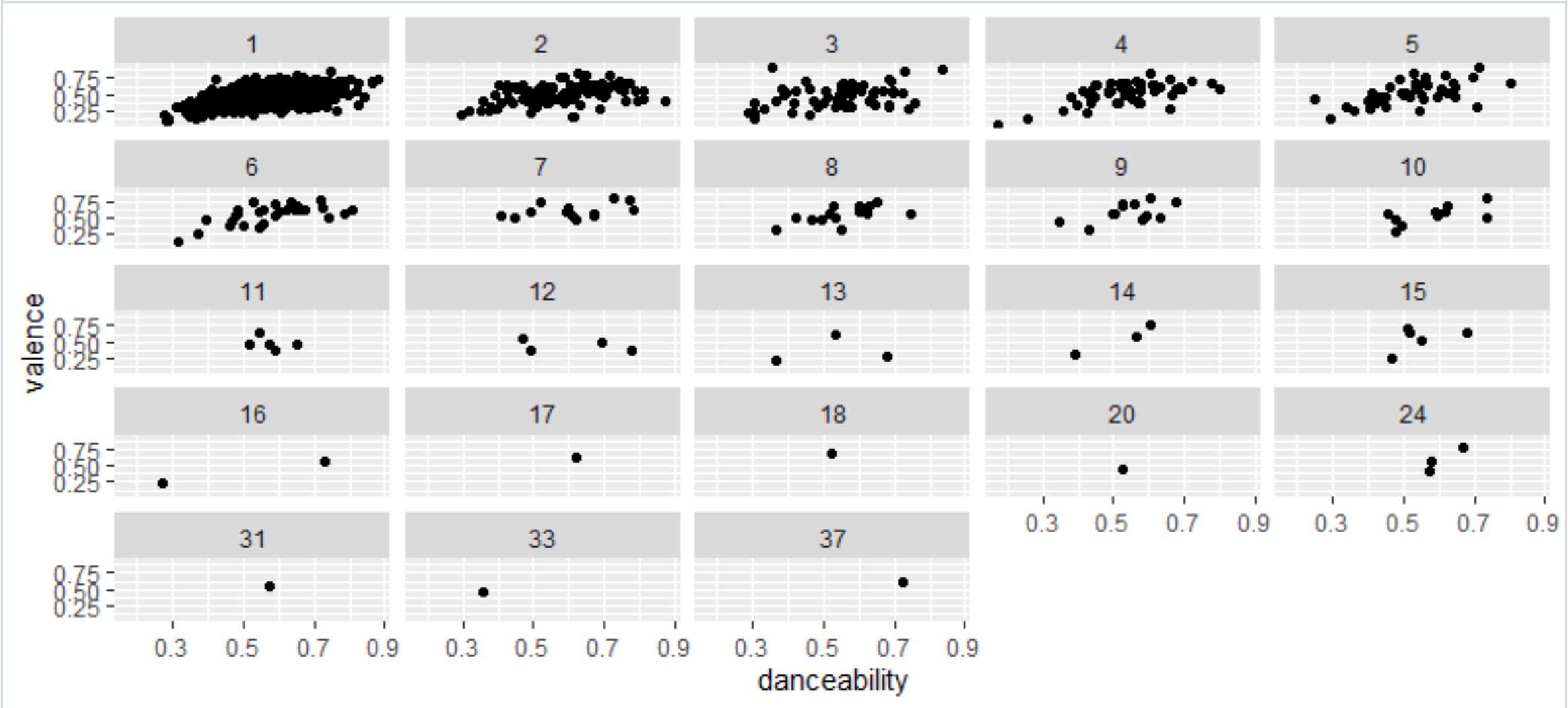
The data was already “clean” but we did need to look at the initial behavior of the data to isolate any outliers. In addition to removing outliers, we had to isolate the data of interest since the data set was so large.

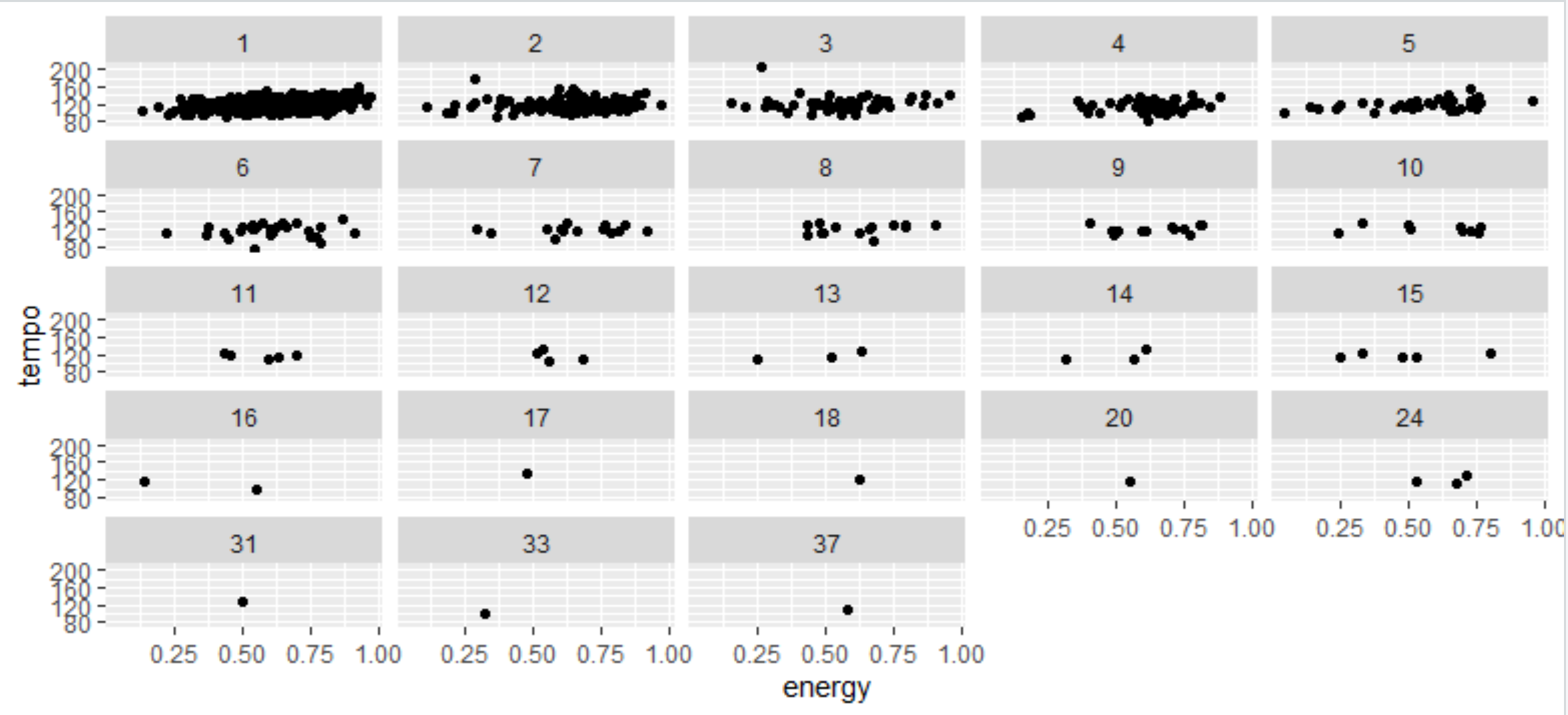
The info we need from the “albums” table is the rank, and the frequency an album appeared on the Billboard 200.

# Data Visualization and Results







# Conclusions

# References

1. <https://en.wikipedia.org/wiki/Billboard_200>
2. <https://www.kaggle.com/snapcrack/the-billboard-200-acoustic-data/>
3. <https://www.kaggle.com/nadintamer/top-tracks-of-2017#featuresdf.csv>