# Spotify Clarifies

## An Analysis of Music Popularity over Time

Britney Brown, Harrison DiStefano, Greg Eastman
Lisa Kaunitz, Tianyang Liu, Jeremy Weidner

06/04/2021

## Contents

**Abstract** **1**

**Introduction** **2**

**Data Description** **2**

**Methodology** **2**

**Analysis** **3**
- Top Artists . . . 3
- Song Titles . . . 4
- Music Theory Attributes . . . 5
  - Tempo . . . 5
  - Time Signature . . . 5
  - Key Signature . . . 6
- Danceability . . . 6
- Attributes of Hit Tracks by Decade . . . 8
  - Duration of Song . . . 8
  - Proportion of Major/Minor Modes . . . 8
  - Musical Qualities . . . 9

**Conclusions** **9**

**Bibliography** **10**

## Abstract

Write something

# Introduction

Everywhere we go, we are followed by melodies and catchy tunes that get stuck in our heads. But have you ever stopped to wonder why we love certain songs and have to turn the radio off on others? In this article, we will do just that by analyzing a Spotify database containing various musical attributes such as tempo, danceability, energy, and acousticness to get to the bottom of what truly makes a song popular. We will also examine if these features change over time, which would suggest a shift in musical preference as the entertainment industry, music, instruments, and we as a society evolve.

# Data Description

We conducted our analysis using a dataset compiling various music features for songs found on Spotify's Web API[1]. This data was obtained by sampling two different subsets of songs, hits and flops, from 1960-2019. Hits are defined as any song that made a top 100 'hit' list any week in a given decade. Flops are defined as the opposite, with additional requirements such as belonging to a genre considered non-mainstream or avant-garde. While the original purpose of this dataset was to predict whether or not a song will be successful, we will be utilizing the musical attributes to analyze the change in music over time. Since this dataset is stratified at the track level, each track acts as a unique unit of observation. With over 40,000 unique songs spread evenly across six decades, it provides plenty of information to analyze the changes in musical attributes across the last several years.

The majority of these variables are generated by Spotify's music analytic algorithm, which is unknown to the general public. Therefore, we have provided documentation to get a fairly robust understanding of the variables. The definitions for all our variables can be found in the attached JSON file as well as in the dataDictionary table when connected to the database instance.

# Methodology

This analysis will use a variety of investigative techniques to explore the data and understand music on a deeper level; answering intriguing questions about the songs we listen to but never question. We will begin by exploring the top producing artists of each decade to get a better understanding of the production of tracks over time. Just because the world's population is growing, will the number of tracks produced by artists also increase across each decade?

We will then shift into an anaylis of song titles and what terms appear most frequently. Nowadays, it is common practice to use certain pictures and words in social media and online videos to attract wider audiences. This same logic applies to song titles. However, instead of simply displaying the most common words, we will first "clean" the song titles using the tm[2] package and tidytext[3] to gain insight on the bigger picture. This cleaning process involves converting all text to lowercase then removing non-alphabetical characters, punctuation, and stop words that exist simply as determiners to mark nouns like "the", coordinating conjunctions like "but", or prepositions like "in". While these terms help to form a language, they do not hold any meaning for our song titles and were removed to decrease clutter in the text portion of the analysis. The final step of our cleaning process was to stem the text. This process alters words to return them to their root meaning. An example in our data was changing the words "remastered", "remaster", and "remastering" into their root: "remast". This allows us to interpret all these separate terms as the same term. With a clean text corpus, we then created a word cloud as a method to understand the frequent concepts in song titles.

We then continue with a univariate analysis involving a deeper dive into three main music theory attributes: tempo, time signature, and key signature. Any musician can look at sheet music, recognize these characteristics, and immediately know exactly how to play the song. But we will take it one step further and compare the various levels of these features to understand their effect on the popularity of a song.

Finally, we will use a visual analysis to explore how music has evolved in the last 50 years.

WHAT ABOUT SECTIONS????

# Analysis

QUICK DESCRIPTION OF ANALYSIS

## Top Artists

We begin our exploratory data analysis by creating a table of the top 5 artists or bands of each decade. For the purpose of this analysis, the top artists are defined in terms of their production output, or in other words, their total amount of songs created per decade. One interesting thing this table shows is the decreasing trend in the number of tracks from top artists over each decade. In the 1960s, the top artists/bands produced over 100 tracks per decade while artists in the 2010s produced 50 or less. This indicates a potential focus on quality rather than quantity as the music industry evolves with time.

Table 1: Top 5 Artist/Bands with the Most Songs per Decade

| Artist | Decade | Total Number of Songs |
|---|---|---:|
| Traditional | 60s | 145 |
| P. Susheela | 60s | 130 |
| Jerry Goldsmith | 60s | 118 |
| Harry Belafonte | 60s | 114 |
| Ennio Morricone | 60s | 96 |
| MPB4 | 70s | 59 |
| Buzzcocks | 70s | 50 |
| kalapana | 70s | 49 |
| John Coltrane | 70s | 44 |
| Vicente Fernández | 70s | 44 |
| The Cleaners From Venus | 80s | 47 |
| Malcolm Arnold | 80s | 45 |
| Nobuo Uematsu | 80s | 33 |
| Skinny Puppy | 80s | 33 |
| Running Wild | 80s | 32 |
| Luis Miguel | 90s | 28 |
| Madonna | 90s | 25 |
| Iggy Pop | 90s | 22 |
| Nobuo Uematsu | 90s | 19 |
| El Gran Combo De Puerto Rico | 90s | 17 |
| Toby Keith | 00s | 27 |
| Tim McGraw | 00s | 24 |
| Rascal Flatts | 00s | 24 |
| Iron Maiden | 00s | 23 |
| Kenny Chesney | 00s | 23 |
| Drake | 10s | 50 |
| Glee Cast | 10s | 41 |
| Taylor Swift | 10s | 35 |
| Luke Bryan | 10s | 25 |
| The Weeknd | 10s | 24 |

## Song Titles

Like music itself, our analysis is multifaceted, so we will transition our analysis from artists to song titles. As discussed in the Methodology section above, we extracted terms from the dataset and are now able to use the wordcloud2[4] package to create the following word cloud in the shape of a star, to resemble the talented artists we just explored.


There are a number of key takeaways from the above plot, but we will start by asking the very question Tina Turner asked in 1984, "What's love got to do with it?" Well it appears love has a lot to do with it, with "it" being song titles. This word appears over 2000 times, which is much more than any other common word found in song titles from 1960-2019. The abundance of this makes absolute sense, since love is one of the most powerful emotions humans have put into words. So powerful, that love songs are commonly seen as more than emotion, but a genre. In addition, we see that love songs are not the only emotional genre listed in our dataset. The blues, heavily prevalent in the 1960s, also appears in the word cloud, leading us to believe that more emotive phrases are popular in song titles.

Besides emotions and feelings, another topic arises as a common theme in song titles: people. More accurately, descriptions of who these people are to the artists of each song. This is reflected by the size the word "man", "girl", "baby", or "boy" have in the word cloud, indicating a higher frequency in our data's song titles. Often times, people seem to talk about themselves and the people in their lives, which seems to hold true for music as well. Therefore, using just the title of the songs, we have established two significant themes in music: emotions and relationships.

## Music Theory Attributes

Now that we've explored the artists and song titles, we will examine some of the musical attributes themselves. We will begin with three very important music theory concepts that every musician lives and breathes: tempo, time signature, and key signature.

### Tempo

Tempo describes the speed/pace of a song. In general, fast tempos evoke more positive emotions such as happiness and delight while slow tempos evoke negative emotions such as sadness and depression. To get an idea of the artists that tend to convey postive emotions compared to those that channel negative emotions, we will take a look at the top 10 artists with the fastest average tempos as well as the top 10 with the slowest.

Table 2: Top 10 Artist/Bands with the Fastest Average Tempos

| Artist | Average Song Tempo | Total Number of Songs |
|---|---|---|
| Crass | 172.8562 | 15 |
| Angerfist | 157.4122 | 13 |
| The Lurkers | 156.5629 | 19 |
| Allison | 156.0839 | 10 |
| The Vibrators | 150.0379 | 10 |
| The Dickies | 149.5186 | 19 |
| Avril Lavigne | 147.3464 | 15 |
| Foo Fighters | 147.1191 | 10 |
| Trifonic | 146.2458 | 11 |
| The Nashville Bluegrass Band | 145.9536 | 10 |

### Time Signature

Time signature indicates the rhythm of the song in terms of a beat's duration and the number of beats per measure. While it can get fairly technical, it is important to know that the most common time signature is 4 beats per measure (and is hence referred to as 'common time'). In fact, 88% of our dataset is made up of songs with common time. Compared to the other time signatures, common time has higher averages for danceability and energy but lower averages for acousticness. This time signature also has the highest success rate for hit songs with 43% of common time songs getting classified as a hit. As seen from the average rates in the table below, the success behind this time signature can be attributed to these songs being easier for people to dance to. Therefore, this table indicates that successful hit songs tend to higher levels of dancibility and energy.

Table 3: Summary of Time Signature Characteristics

| | | | | | |
|---|---|---|---|---|---|
| Time Signature | 4.000000e+00 | 3.0000000 | 5.0000000 | 1.0000000 | 0.0000000 |
| Average Danceability | 5.579703e-01 | 0.4049516 | 0.3993724 | 0.4014569 | 0.1334000 |
| Average Acousticness | 3.308935e-01 | 0.6177392 | 0.6153970 | 0.6211903 | 0.5906667 |
| Average Energy | 6.058251e-01 | 0.3784532 | 0.3824055 | 0.3927565 | 0.4109940 |
| Number of Tracks | 0.000000e+00 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 |
| Number of Hits | 1.569300e+04 | 973.0000000 | 117.0000000 | 61.0000000 | 1.0000000 |
| Percentage of Hit Tracks | 4.345000e-01 | 0.2555000 | 0.1993000 | 0.1644000 | 0.3333000 |

### Key Signature

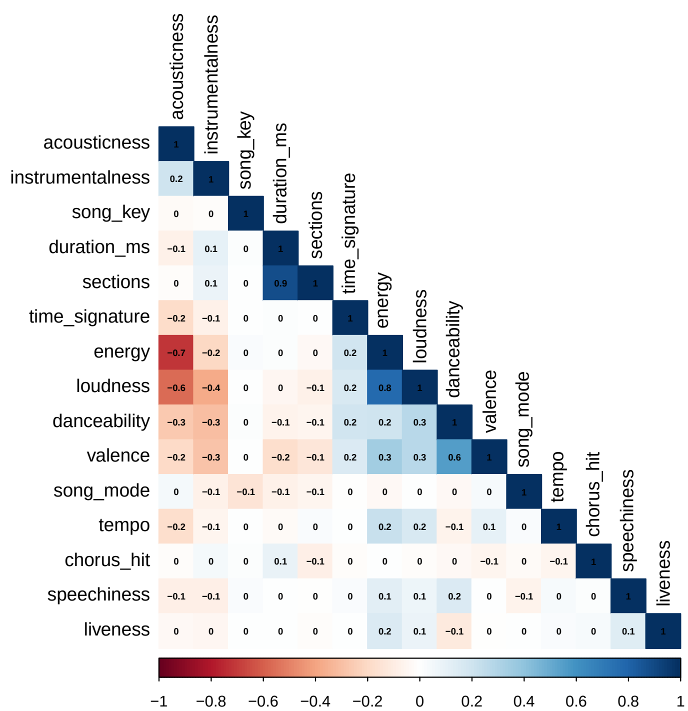
Key signature describes the combination of sharps and flats that determine the scale of a piece of music. In general, most songs shift from key to key to add variance and intrigue but this dataset simply estimates the overall key. As the table below indicates, the estimated key does not have a large effect on the success rate of songs since the 'Percentage of Hit Tracks' varies around 40% for each key.

Table 4: Summary of Key Signature Characteristics

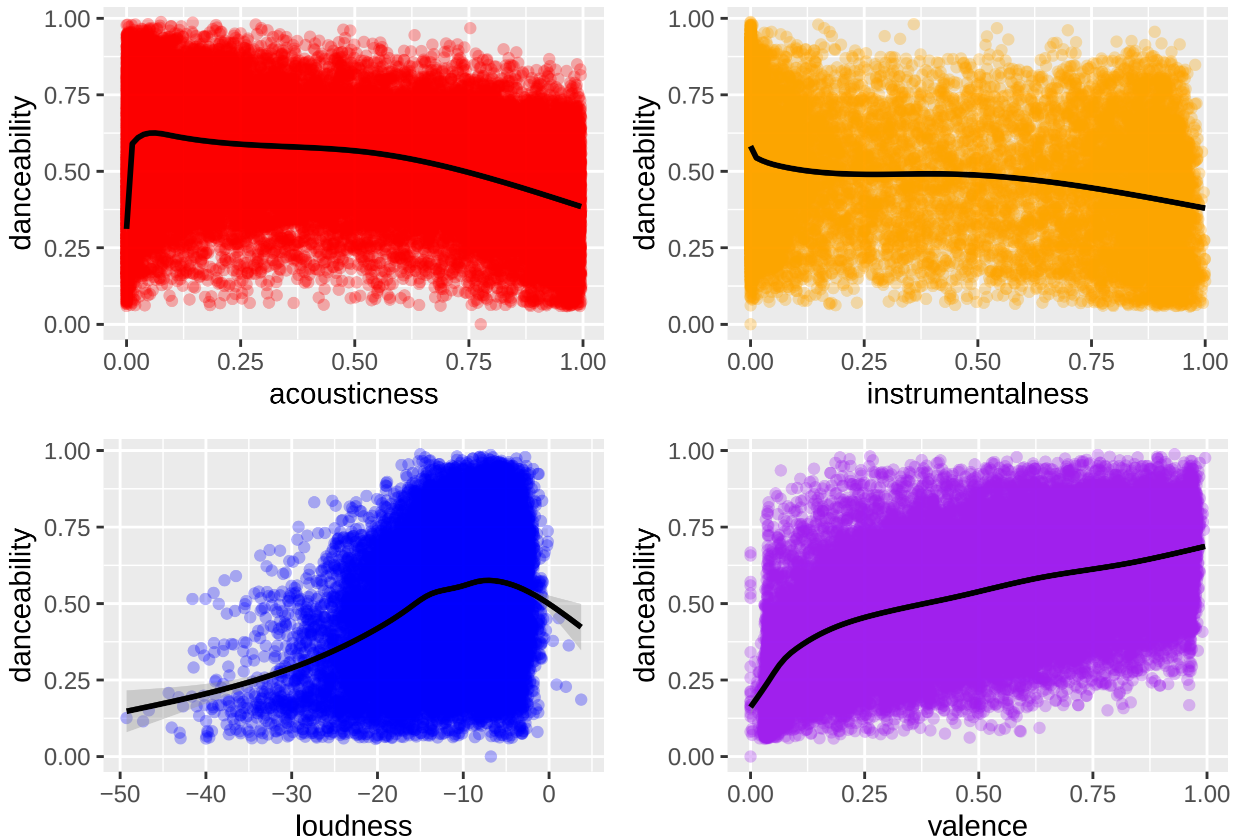
| Key Signature | Average Danceability | Average Acousticness | Average Energy | Percentage of Hit Tracks |
|---|---|---|---|---|
| C | 0.5391916 | 0.3912941 | 0.5613325 | 0.4078 |
| C#/Db | 0.5757106 | 0.2853243 | 0.6194482 | 0.4621 |
| D | 0.5225773 | 0.3628475 | 0.5817457 | 0.3876 |
| D#/Eb | 0.5051904 | 0.4968816 | 0.5100628 | 0.3822 |
| E | 0.5217508 | 0.3590309 | 0.5866685 | 0.3873 |
| F | 0.5329451 | 0.4307504 | 0.5385227 | 0.4084 |
| F#/Gb | 0.5568206 | 0.2918481 | 0.6267238 | 0.4417 |
| G | 0.5399274 | 0.3730568 | 0.5711438 | 0.4018 |
| G#/Ab | 0.5477378 | 0.3751159 | 0.5693247 | 0.4398 |
| A | 0.5261597 | 0.3525814 | 0.5924012 | 0.3834 |
| A#/Bb | 0.5593870 | 0.3920532 | 0.5595066 | 0.4569 |
| B | 0.5607593 | 0.2833212 | 0.6295437 | 0.4240 |

## Danceability

Through our analysis of the music theory features, we discovered the importance of danceability. This attribute is defined by Spotify as a way to describe how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value close to 0 indicates a song difficult to dance to while 1 represents dance-friendly songs. However, since we do not know the specifics of Spotify's algorithm, we will explore the correlation of this feature with other variables in our dataset.



TEXT TEXT TEXT

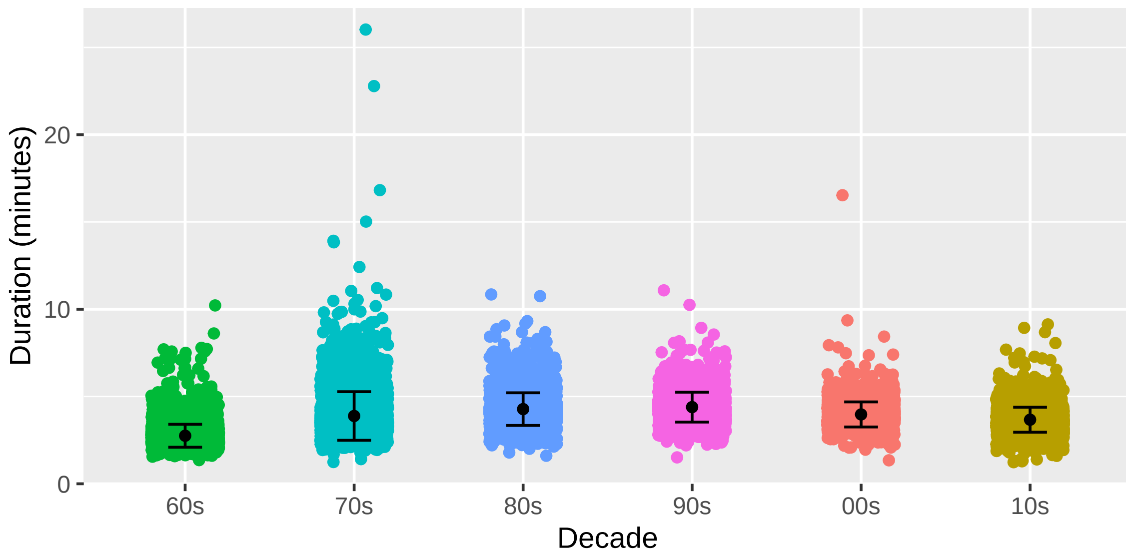


## Attributes of Hit Tracks by Decade

TEXT TEXT TEXT

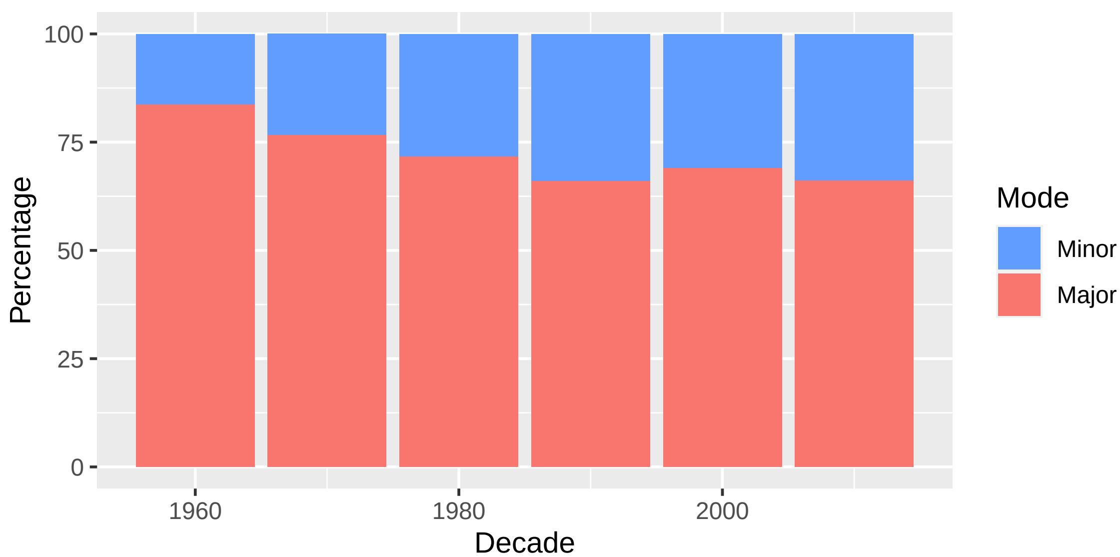
### Duration of Song

The duration of hit tracks is genrally under 10 minutes with an average closer to 4 minutes across the decades. One notable exception is in the 1970s, possibly due to the popularity of Progressive Rock during those years with the blending of Rock and Jazz Fusion into long, drawn-out concept albums and longer-winded tracks.



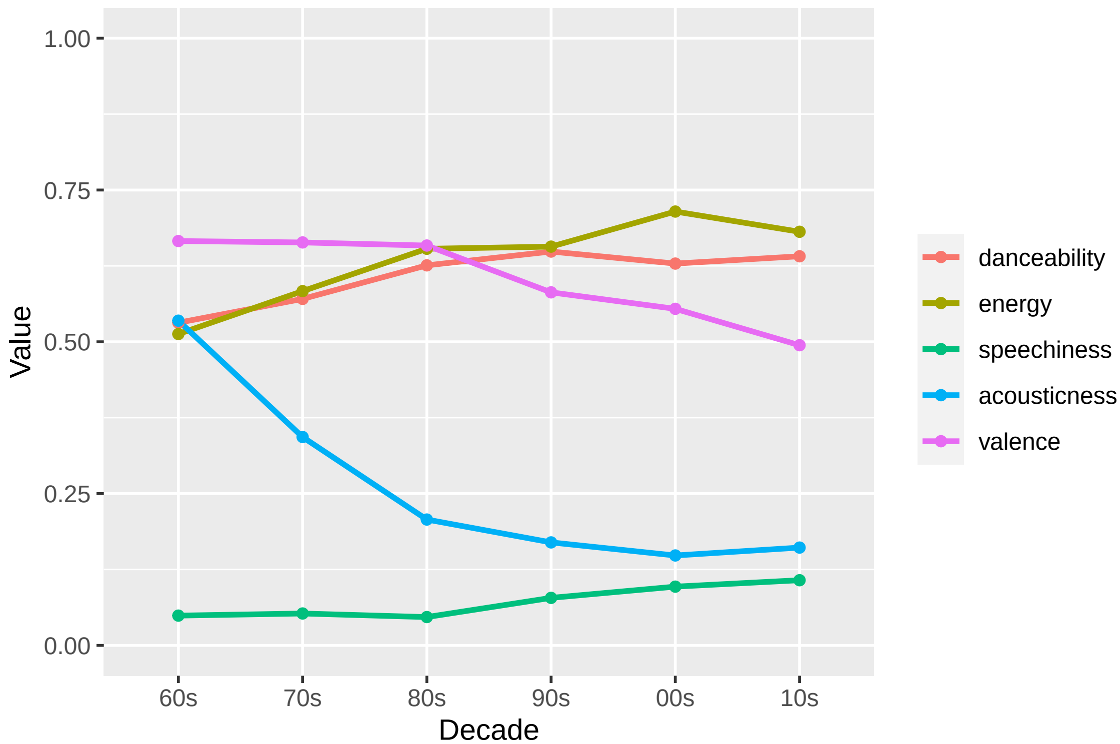

### Proportion of Major/Minor Modes

The proportion of hit songs in the minor mode has increased from about 15% in the 1960s to almost 30% in the 2010s. The minor mode is associated with sadness, while the major has a happier sound. This signifies a potential trend that people have developed a taste for more depressing, saddening popular music over the last couple of decades.



### Musical Qualities

In addition to the duration and mode of songs, there are also a few noticeable trends in the Spotify qualities of hit music across the decades. Danceability, energy, and speechiness (the use of vocals) all saw a general upward trend over time, leveling out or decreasing slightly in the 2010s. On the other hand, we see the use of acousticness, as opposed to electric instruments, in tracks decline significantly from the 60's onward. This downward trend is align with the historical development of more electronic instruments and their widespread use. Finally, valence, which describes the musical positiveness conveyed by a track, has also been trending downward since the 80's. This decrease agrees with our finding in the previous section that indicates people prefer sadder songs in recent years.




## Conclusions

# Bibliography

1. FortyTwo102. "Spotify and Billboard Top Hits Data." GitHub. Jan. & Feb., 2020. Accessed May 27, 2021. https://github.com/fortyTwo102/The-Spotify-Hit-Predictor-Dataset.

2. Ingo Feinerer, Kurt Hornik, and David Meyer (2008). Text Mining Infrastructure in R. Journal of Statistical Software 25(5): 1-54. URL: https://www.jstatsoft.org/v25/i05/.

3. Silge J, Robinson D (2016). "tidytext: Text Mining and Analysis Using Tidy Data Principles in R." *JOSS*, *1*(3). doi: 10.21105/joss.00037 (URL: https://doi.org/10.21105/joss.00037), <URL: http://dx.doi.org/10.21105/joss.00037>.

4. Dawei Lang and Guan-tin Chien (2018). wordcloud2: Create Word Cloud by 'htmlwidget'. R package version 0.2.1. https://CRAN.R-project.org/package=wordcloud2