Quantitative Sentiment Analysis of Lyrics in Popular Music

ABSTRACT Popular music has been changing significantly over the years, revealing clear, audible differences when compared with songs written in other eras. A pop music composition is normally made of two parts—the tune and the lyrics. Here we use a digital humanities and data science approach to examine how lyrics changed between the 1950’s and the more recent years, and apply quantitative analysis to measure these changes. To identify possible differences, we analyzed the sentiments expressed in the songs of the Billboard Hot 100, which reflects the preferences of popular music listeners and fans in each year. Automatic sentiment analysis of 6,150 Billboard 100 songs covering all the years from 1951 through 2016 shows clear and statistically significant changes in sentiments expressed through the lyrics of popular music, generally towards a more negative tone. The results show that anger, disgust, fear, sadness, and conscientiousness have increased significantly, while joy, confidence, and openness expressed in pop song lyrics have declined.

KEYWORDS: Popular music, lyrics, data science, sentiment analysis, Billboard, tone analyzer

Popular music has changed significantly since its early days in the 1950’s, as shown in clear musical differences between songs written then and those produced in later periods. These differences are reflected in the numerous genres that developed since the early days of pop music, each with its own unique musical characteristics. A pop music fan can normally discriminate easily between the audio of a song recorded in the 1960’s and one made in the 2000’s.

In most cases, pop songs have lyrics in addition to a tune, and the lyrics provide another layer to the songs’ emotion, attitude, and narrative. Although love and romance have always been the most dominant topic of popular music, lyrics and lyrical styles have changed significantly across the decades, reflecting social and political changes.

Several studies focused on the changes in the topics discussed in popular music lyrics, and their psychological, social, and economic implications. For instance, the expression of political ideas through pop music lyrics has changed over time, and is also influenced

by events outside of the world of music, linking musical preferences to political orientation. Popular music lyrics have also changed the way sexual content is expressed. Another example is the explicit mentioning of drugs during the mid-1990s, and in general, drugs and other substances are mentioned more frequently in pop music during more recent years. LGBTQ identity and gay rights have also appeared more frequently in pop music lyrics in tandem with social changes. Flynn et al. looked for gender differences within cross-genre lyrics from the Billboard Top 20 songs between 2009 and 2013. Pettijohn and Sacco Jr assessed the content of Billboard Hot 100 number one songs between 1995 and 2003, and found that during difficult economic or social instances in history, there was a greater presence of comforting, romantic, and meaningful lyrics. Yoo et al. applied statistical analysis and text mining to identify changes in the patterns of words used in Korean pop music, and revealed, for example, that English words are more common in more recent Korean pop music lyrics. Ballard et al. identified differences between lyrics in four different genres: heavy metal, rap, country, and pop. Their study shows that lyrics of different genres have a different impact on the listener’s behavior. Namely, the findings show that rap and heavy metal lyrics are less likely to inspire prosocial behavior.

While little work has been focused on large-scale sentiment analysis and text mining in pop music, several previous studies isolated different defined emotions and measured these emotions from each song to profile songs or listener’s preference patterns. Yeh et al. classified popular songs using Thayer’s model of four emotions: angry, happy, relaxed, and depressed. Their automatic classifier used the emotions of extracted choruses to categorize the emotion of the entire song, and achieved an average prediction rate of 92% in the prediction of songs’ emotions. Batcho used survey participants to rate songs for anger, happiness, sadness, relevance, liking, and nostalgia. The study showed that participants who favored personal nostalgia also favored happy lyrics, while sad lyrics were more associated with historical nostalgia. Knobloch and Zillmann tested mood-management theory by having subjects in a bad or good moods choose to listen to songs from the top 10 chart. These songs were pre-categorized according to their joyfulness and energy. The results

showed that subjects in bad mood were more likely than subjects in a good mood to listen to joyful, energetic songs for a greater length of time.

Here we use automatic sentiment analysis to measure the sentiments of the songs in the Billboard Hot 100 songs between 1951 and 2016. The different sentiments are quantified to measure the changing sentiments expressed through pop music lyrics across the years, and their correlation with the year is used to identify trends in pop music lyrics and music fan preferences.

**AUTOMATIC SENTIMENT ANALYSIS**

Automatic sentiment analysis, a subset of text mining, is the examination of sentiments (opinions, emotions, attitudes, and feelings) found within text. Computational methods are implemented to locate, isolate, and categorize these sentiments. Sentiment analysis is often paired with social media mining, as a means of discerning the sentiments found in text extracted from on-line product reviews, discussion forums, blogs, and other social networking services. The implementation of sentiment analysis is a beneficial method of comprehending the opinion of a certain person or a certain population, and is often used in the context of a specific subject, such as politics, current news, entertainment, etc. The insights it provides from both product reviews and customer service reviews are vital information for businesses, enabling them to discover customer needs, satisfaction, and concerns, as well as helping them to provide quality products and services and achieve the intended brand image.

Sentiment analysis can be done at three levels: the document-level, sentence level, or aspect-level. Document-level analysis determines the average sentiment for an entire document. Sentence-level sentiment analysis views each sentence as a separate document, hence determining the individual average sentiment for each sentence or line of a given document. Aspect/feature-level analysis locates specific features of an entity (the main subject of the document), and uses the writer’s opinion of those features to classify the sentiment regarding the entity in question. An example of an entity-feature pair would be: “I had a nice stay at the Cupertino Hotel. The daily breakfast was delicious.” In this document, the entity is Cupertino Hotel and the feature is breakfast. When studying the sentiment found in text, it is important to note that sentiment can be explicit or implicit. An explicit sentiment would be: “This is my new favorite restaurant.” Whereas, an implicit sentiment would be: “The restaurant’s apple pie had the right balance between sweetness and tartness.”

Classifying the sentiment of a document or sentence can be separated into several common approaches. Text can be classified according to its polarity/valence (positive, negative, or neutral). Cho et al. used multiple sentiment dictionaries to classify Amazon.com reviews on books, smartphones, and movies as either positive or negative. Taking polarity one step further, text can be classified according to the strength of the sentiment (very good, very bad, etc). Villarroel Ordonez et al. used the pre-existing five-star scale ratings of customer reviews from BN.com, Amazon.com, and TripAdvisor.com to classify the strength of both explicit and implicit sentiments. An alternative approach to polarity classification is appraisal theory, where emotions brought by situations found within the text are used to classify the sentiment of the text. The number of emotion categories varies, depending on the data and purpose of the analysis. Twelve emotion classes were suggested by Storm and Storm: anxiety, fear, anger, hostility, disgust, shame, sadness, contentment, liking, happiness, pride, and love. Shaver et al. suggested the use of just five emotions: anger, sadness, fear, happiness, and affection.

METHODS

Data: The Billboard Hot 100
Pop music is a very prevalent form of art, and a large number of artists record and release a very large number of songs each year. Songs can become popular and have impact through the traditional channels of communication such as radio stations, but also through peer-to-peer communication and specific groups of interest. In the post-information era songs can also be communicated through social media or the web. Measuring all of these songs, however, can lead to biased results because the vast majority of these tunes do not become popular among pop music listeners, and therefore do not have a significant impact or indication on the preferences of pop music consumers. Therefore, the history of pop music can be profiled by the most popular songs in each year that had the highest impact and were consumed by the highest number of pop music listeners.

The annual Billboard Hot 100 is a common tool to characterize the most popular songs in a certain year, often has been used in pop music studies to identify trends and typical preferences of pop music artists and fans.

Since 1958, Billboard, an entertainment media magazine, has been a prevalent source in relaying the most popular music hits with its introduction of the Billboard Hot 100 chart. This chart dictates the 100 most favored songs for a given year through the combined use of three measurements. Originally, these measurements were the following:

1. The number of sales per single.
2. The number of times the single was played on a jukebox.
3. The number of times the single was played on the radio. (The Top 40 airtime chart was used as a reference.)

The grouping of these measurements was intended to capture the initial popularity of a single and test the longevity of its popularity over time. Over the years, the music industry has evolved, making the original Billboard Hot 100 measurements irrelevant. Jukebox sales have declined sharply and practically disappeared, and with the introduction of the web, Billboard magazine has made additions to its core measurements to properly reflect changes in technology. In addition to in-store and concert sales, Billboard magazine now tracks both physical and digital sales made on-line. In 2011, streaming ratings were introduced as a measurement for the Billboard Hot 100. This includes the streaming services: Spotify, Rhapsody, Muve Music, etc. In 2013, video plays were incorporated from YouTube and Vevo. Billboard also takes into account information collected from social media.

Before the Billboard Hot 100, Billboard generated three separate charts, each based on one of the original measurements. Of these charts, the ranking used by this analysis for songs earlier than 1958 is the Best Sellers in Stores (Best Sellers) chart as a reflection of music fan song preferences.

Certain Billboard songs were sold as part of a pair. The second song was either a B-side, an A-side, or double-A side single that was sold together with one of the charting songs. The Billboard Hot 100 has changed its rules over time regarding two-song charting, but remained consistent in considering the first song listed in the pair considered to have a stronger performance than the second one. Since it is not clear if the second song was equally as popular as the first or simply tolerated because of the popularity of the first, this analysis did not include the lyrics of any second songs from two-song chartings.

The source for the lyrics used in the analysis was www.azlyrics.com and www.oldlyrics.com, and the lyrics were collected and downloaded automatically. For songs

not present in AZLyrics or OldieLyrics, secondary sources include genius.com, www.lyricsfreak.com, www.metrolyrics.com, www.lyricsmode.com, www.lyricsondemand.com, lyrics.com, www.songlyrics.com, and www.flashlyrics.com. Before, the lyrics were saved, any presence of labeling brackets was removed. Examples of these include: [Chorus], [Verse], [x2], [instrumental], etc. This minimizes the risk of non-lyrical, but descriptive, text distorting the tone analysis results.

Sentiment analysis: Tone Analyzer

The tool used for the sentiment analysis is Tone Analyzer, which is a mature computational linguistic tool provided as part of the IBM Watson developer cloud. The IBM Watson Tone Analyzer Application Programming Interface (API) is a tone analysis service that processes text input for tone and outputs a JavaScript Object Notation (JSON) file containing document level and/or sentence-level scores of 13 types of tones. Scores range from 0 (no presence of given tone) to 1.0, which means 100% presence of tone. A score of at least 0.5 indicates the presence of a given tone, whereas any score greater than 0.75 indicates a strong presence of the given tone. Tone Analyzer separates these 13 tones into three categories:

1. Emotional Tone: anger, disgust, fear, joy, sadness.
2. Language Tone: analytical, confident, tentative.
3. Social Tone: openness, conscientiousness, extraversion, agreeableness, emotional range.

Tone analyzer uses a combination of psycholinguistics and machine learning to determine the type of tone found in text. Psycholinguistics is the assessment of the association between how our minds work and how we learn, use, and comprehend language. Tone Analyzer links words with different tones associated with them. These words are balanced by the observation that negative emotions are normally expressed in a more intensive fashion. The combination of different words and tones are handled by Tone Analyzer by using a Support Vector Machine (SVM), with a one-vs-rest method that expands SVM to more than two classes. The choice of words provides substantial information used by the computer to determine the tone and sentiments expressed in the text, and the personality of the writer.

For example, the first sentence of the song “Total eclipse of the heart” (Bonnie Tyler) is “Turnaround, every now and then I get a little bit lonely and you’re never coming round”. The most dominant sentiment in that sentence according to Tone Analyzer is sadness, with value of 0.786. Sadness is also a dominant tone for that entire song, with a score of 0.52. Joy, on the other hand, is scored low for that song, with a score of 0.09. Fear has a value of 0.53, conscientiousness 0.08, extraversion is 0.02, and openness has a score of 0.48 for the song.

The first line of the Village People’s “Y.M.C.A.” is “Young man, there’s no need to feel down”, and the dominant tone in that line as analyzed by Tone Analyzer is tentative, with a score of 0.61. Extraversion score for that entire song is 0.55, and the joy score is 0.65. Anger, disgust, and fear are scored much lower, with 0.11, 0.07, and 0.09, respectively.

The first lines of Queen’s “We will rock you” is “Buddy you’re a boy make a big noise, playin’ in the street gonna be a big man some day”, analyzed by Tone Analyzer as agreeableness (0.64), extraversion (0.85), and fear (0.39). Tones with lower scores are disgust and sadness, with a score of ~0.07).

The Bee Gees song “Too much heaven” opens with “Nobody gets too much heaven no more, it’s much harder to come by, I’m waiting in line”. For that part of the song, Sadness is the most dominant tone as deduced by Tone Analyzer, with a score of 0.64. For that entire song Tone Analyzer computed low scores for anger (0.01), disgust (0.003), and fear (0.01), while joy (0.62), agreeableness (0.95), and extraversion (0.77) are scored high.

RESULTS

A total of 6,150 Billboard Hot 100 songs were collected, representing 66 years of popular music, from 1951 through 2016. Of these 6,150 songs, 65 songs were instrumental-only pieces, and due to the absence of lyrics these songs were omitted from the tone analysis in order to avoid skewing the data with zero values. The analysis therefore represents only the 6,085 songs that had lyrics. Figure 1 shows the number of songs used in each year, excluding instrumental pieces that do not have lyrics.

The songs were grouped together by year, and their tone scores were averaged for each year. For example, the average anger score from songs in 1951 was 0.0751. Averaging the data produced 66 data points (the total number of years) with 13 unique averaged tones. The standard deviation and standard error were calculated for each averaged tone of each year. Considering the calculations from all 13 tones, the range in standard error was 0.61% to 7.11%. Ninety percent of the standard error values were less than 3.51%, and just 10% of the standard error ranged from 3.51% to 7.11%. Focusing on the standard error (SE) of each individual tone category, 90% of emotional tone SE was less than 2.84%, 90% of language tone SE was less than 3.74%, and 90% of social tone SE was less than 3.64%.

Pearson Correlation and Linear Regression

From the averaged tone scores, two tests were conducted to identify the presence of a linear correlation between the year and a specific tone from lyrics of that year. The first test was a Pearson correlation coefficient test: using Equation 1, where \( x \) is the year, \( \bar{x} \) is the sample mean of the years, \( y \) is the tone score, and \( \bar{y} \) is the sample mean of the tone scores.
The magnitude of the resulting coefficient indicates the polarity and strength of the relationship. The second test used is a simple linear regression. It examines the relationship between two continuous variables, where the value of one variable \( x \) is capable of predicting the value of the second variable \( y \). A linear regression uses the equation \( y = mx + b \) to test the linearity of the dependence between \( x \) and \( y \). The regression line can be used in trend analysis as the best-fit line of the relationship between \( x \) and \( y \). In this study the regression line shows the existence of a trend of a consistent change in sentiments over the observed period of time. To generate this equation, linear regression uses the Least Squares Method as shown in Equation 2.

\[
S = \sum (y - \bar{y})^2
\]

The Least Squares Method is the minimized sum of the squared vertical differences between each actual data point and the best fit line. If the resulting slope of Equation 2 is statistically different than zero, it indicates the existence of a linear dependency between the year \( x \) and the tone \( y \) of its songs. The polarity of the slope indicates the existence of a positive or negative trend.

Based on the results of the Pearson correlation coefficients, anger, disgust, fear, and conscientiousness are sentiments that have strong positive correlations with the year. Their respective coefficients were 0.8897, 0.7817, 0.7790, and 0.7629, and the correlations are all statistically significant \((P < 0.0001 \text{ in all cases})\). Figures 2, 3, 4, and 6 show the average measured anger, disgust, fear, and conscientiousness, respectively.
Figure 2. Average tone of anger measured for every year using Billboard Hot 100 song lyrics from 1951-2016. The error bars are the standard error of the mean.

Figure 3. Average tone of disgust measured for every year using Billboard Hot 100 song lyrics from 1951-2016.

Figure 4. Average tone of fear measured for every year using Billboard Hot 100 song lyrics from 1951-2016.
The Pearson correlation shows that the emotional tones of anger, disgust, and fear, as well as the social tone of conscientiousness, have become more prevalent in the Billboard Hot 100 and Best Sellers lyrics. Anger started to increase more rapidly during the second half of the 1980s, and continued to increase through 2016. Disgust was lower in the early 1950s and early 1980s, but increased starting the 1980s and until 2016. Fear remained relatively steady until the early 1980s, where it started to increase until 2016.

Joy has a negative correlation with time, with a Pearson correlation coefficient of $-0.7293$ ($P < 0.0001$). That demonstrates that the Billboard Hot 100 and Best Sellers lyrics had been becoming less joyful over the years. Figure 5 shows the change in joy expressed in pop song lyrics. The graph show that anger joy was generally higher during the 1950’s through the 1970’s.

Unlike these four sentiments, not all sentiments that were measured showed a significant long-term trend over the years. Figures 7 and 8 show the change in measured extroversion and agreeableness, respectively. The graphs show that these sentiments have not changed significantly. The Pearson correlation of the extroversion is $\sim 0.0687$, and is not
Extroversion

![Figure 7](image1.png)

**FIGURE 7.** Average tone of extroversion in Billboard Popular Lyrics from 1951-2016.

Agreeableness

![Figure 8](image2.png)

**FIGURE 8.** Average tone of agreeableness in Billboard Popular Lyrics from 1951-2016.

statistically significant \((P' < 0.584)\). The Pearson correlation of the change in agreeableness is \(\sim 0.126\), and is also not statistically significant \((P' < 0.313)\). The graph also shows that extroversion decreased during the 1980’s.

Looking at the simple linear regression trend line, the emotional tone of anger had the strongest linear relationship out of all the tones analyzed, with a \(R^2\) value of 0.7915.

The tones of disgust, fear, joy, and conscientiousness were only mildly reflected by a linear regression line. Their \(R^2\) values ranged between 0.5319 and 0.6111. Figures 2-6 show visualizations of anger, disgust, fear, joy, and conscientiousness, respectively within Billboard Hot 100 and Best Sellers lyrics over the decades. All graphs show the mean of the sentiment scores of all songs in each year, as computed by Tone Analyzer.

In addition to the seven sentiments mentioned above, other measurements that were made are sadness, analytical, confidence, tentativeness, and openness. Sadness and tentativeness had positive Pearson correlations of \(\sim 0.655\) and \(\sim 0.58\), respectively \((P < 0.0001)\), showing that popular music lyrics became more tentative and sadder in time.
Figure 9 shows how sadness changed over time. The graph shows that sadness started to increase during the late 1980s, and peaked during the first decade of the 21st century compared to the other decades.

On the other hand, analytics, confidence, and openness have significant negative Pearson correlations of $-0.661$, $-0.612$, and $-0.523$, respectively ($P < 0.0001$), showing that the expression of these sentiments in the Billboard Hot 100 pop music lyrics decreased in time.

We also analyzed the different tones expressed in lyrics in different genres. Three common musical genres were tested, which are country, rap, and pop. The analysis was based on 148 country songs, 170 rap songs, and 281 pop songs. All songs were taken from the Billboard Hot 100 dataset used in the previous experiments. Figure 10 shows the mean and standard error of each of the 13 tones.

As the analysis shows, the tone in rap song lyrics is substantially different from the tone expressed in the lyrics of pop and country songs. Namely, anger and disgust are tones that are much more common in rap songs compared to pop and country, while joy is expressed less often in rap lyrics compared to pop and country. These observations are aligned with the work of Ballard et al., which showed that rap music lyrics are less likely to inspire prosocial behavior. The graph also shows that rap lyrics tend to be less analytical and less confident, while being more extroversive compared to pop and country.

**Conclusion**

Analytical tools and automatic text mining are useful methods that allow quantitative analysis of large datasets of text files, providing a new perspective on the analysis of art and human creations. These tools have been becoming more common through the emerging field of the digital humanities, which is a relatively new academic discipline that combines computer and information technologies with fields within the humanities such as
literature, poetry, music, and visual art. That combination allows to archive, catalog, distribute, and process larger amounts of information, that were not possible in the pre-information era. The digitalization of the data also allows to measure that information quantitatively to analyze and profile human creations in a quantitative manner. Here we performed automatic analysis of the tone of the Billboard Hot 100 songs between 1951 through 2016 to identify trends in the sentiments expressed through the pop music lyrics.

This analysis shows that the tone in popular music lyrics has shifted significantly over the years. Anger, disgust, fear, sadness, tentativeness, and conscientiousness have increased over time, while joy, analytics, confidence, and openness have declined. Extroversion and agreeableness did not show a clear long-term trend, although extroversion showed a decline during the 1980s. In general, the results show a clear trend toward a more negative tone in pop music lyrics, with a more significant change around the early 1990s. That trend can also be explained by changes in social values, reflected through changes in mainstream popular music. Using the Billboard Hot 100 songs aims to ensure that the analysis is based on the songs that were the most popular in each year, as a reflection of music fan preferences during that time.

The study is limited by the tones that can be measured by Tone Analyzer. Clearly, lyrics can be subjective, and can be interpreted differently by different listeners, making it more difficult to make a precise measurement of the tone. The study also depends on the ability of Tone Analyzer to capture the sentiments, and the ability of the Billboard Hot 100 chart to reflect the true preferences of music fans in different years, as the measuring scheme of Billboard Hot 100 changed over time to adjust to the technological and social changes of popular music consumption. However, such quantitative approaches to popular music studies provide new insights from the historical popular music data, and can provide new information that can be difficult
to quantify and profile without using text analysis algorithms. It is expected that with the continuous growth of the digital humanities, the application of text mining methods to popular music will become more prevalent in popular music studies, and will enable new discoveries and observations into the history of popular music.

WOR Lloyd}


Yarkoni, Tal, “Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers,” *Journal of Research in Personality* 44, no. 3 (2010): 363–373.

