Erasmus School of Economics

It's all in the lyrics

Text analysis of song lyrics in popular music

Student	Joey Willemse
Student number	402511
Supervisor	F. Deutzmann, PhD
Study	Msc, Economics and Business, specialisation Marketing
Date	28th August 2014
Place	The Hague

Preface

This master thesis is the final part of my study. I have gotten the chance to work on a wide variety of topics that I find very interesting. One topic that I did not get the chance to work on was music, while this is a big passion of mine. This is why I choose to write my thesis about music and I have written my thesis with a lot of pleasure.

I had a lot of help from my supervisor Florian Deutzmann during the writing of this thesis. This helped me during difficult periods of the process and I would like thank him for his guidance.

I also want to thank my parents and my sister who always encouraged me during my study and always helped when needed.

The Hague, August 2014

Joey Willemse

Abstract

The goal of this study was to examine which themes in lyrics of rock and rap music have a positive effect on popularity. Content analysis with text mining showed that rock lyrics consist of seven themes, in order of frequency: heartbreak, empowerment, abandonment, violence, honesty hell and surviving. Rap lyrics consist of thirteen themes, in order of frequency: ghetto, sex, hitting rock bottom, bragging, wealth and status, women, being the best, punishment, drugs, prostitution, gangsta life, empowerment and shooting. The content analysis showed that there are big differences in content between rock and rap music. Comparisons with other research showed that the content of both genres has changed strongly over the years. Huber-White estimators showed that only empowerment has a significant, positive, effect on rock popularity. The analysis also showed that gangsta life is the only theme that has a significant, negative, effect on rap popularity.

Table of contents

Introduction

Revenues in the music industry have dropped drastically over the years because of file sharing. Music sales are estimated to have dropped as high as 20 percent tot 30 percent (Zentner, 2006; Hong, 2013). This is a very high decline for an industry that had revenues of $15 billion in 2013 (International Federation of the Phonographic Industry, 2014). One of the ways to keep the revenues at a high level is to know if a song has the potential to be a hit or not. Music consists of two main components and so these components affect the hit potential of a song: melody and lyrics (Serafine, Davidson, Crowder & Repp, 1986). This research focuses on the lyrics of music.

Lyrics are about themes and this research will look at the effect of different lyrical themes on the popularity of music. Rock and rap music were the first and second most popular music genres in 2008 in the United States of America respectively (Recording Industry Association of America, 2008). This means that these genres generate high revenues for the music industry and it is therefore important to know the effect of lyrical themes specifically for these genres.

There have been studies that showed the effects that lyrics have on listeners. For instance, Ali and Peynirciogly (2006) showed the effects of lyrics on how listeners convey positive and negative emotions. Another study showed that most teenagers focus on the music while a minority paid more attention to the lyrics (Prinsky & Rosenbaum, 1987; Arnett, 1991).

Former content analysis of rock lyrics showed that rock lyrics are mostly about defiance, rebellion, misogyny and expression of youthful angst (Bleich, Zillman & Weaver, 1991). Content analysis of rap lyrics showed that rap lyrics are mostly about misogyny, racial injustice / inequality and violence (Watkins, 2001; Mcfarland, 2003). This gives the impression that these are popular lyrical themes among listeners of both genres and that these

themes have a positive effect on popularity. However, the effect of lyrical themes on the popularity has not been studied yet.

This is why the following research question will be researched:

What is the difference of the effect of lyrical themes on the popularity of a song between rap and rock music?

Background and Theory

Music industry

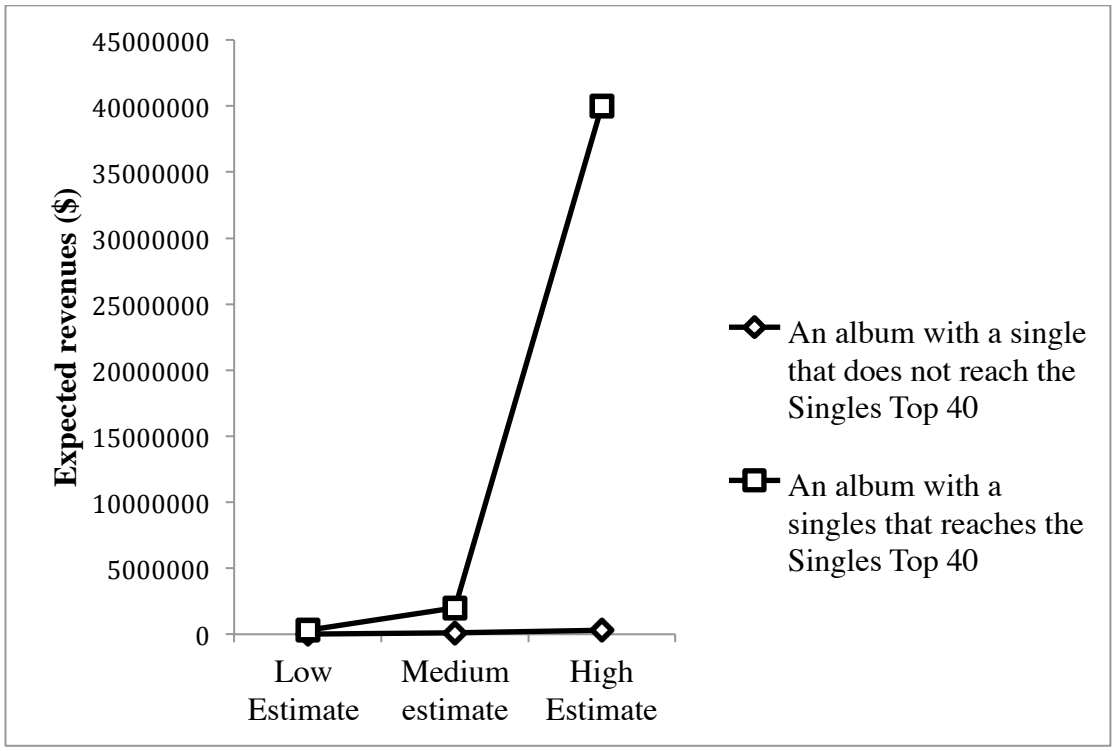
The record labels are the companies who sign musicians and make it able for them to reach a big audience (Shuker, 2001). The record companies are divided into two groups: the 'majors' who operate internationally and the independents. The majors are part of larger electronics or communication companies that provide a wide variety of products and services for international markets. The majors contain a number of labels. The independents are most of the time dependent on the majors for distribution and the majors also help to find and develop new musical talent (Negus, 1996; Shuker, 2001).

The record companies invest annually 4.5 billon dollar globally in musicians, repertoire and marketing (International Federation of the Phonographic Industry, 2014). This has resulted in revenues of 15.029 billion dollar in 2013. These revenues consisted of revenues from CD / DVD sales (7.730 billion dollar – 51 percent) streaming and subscription services (5.872 billion dollar – 39 percent) performance rights paid by internet radio services, venues and broadcasts (1.106 billion dollar – 8 percent) and synchronization deals, music used in films, advertisements and television programmes (322 million dollar – 2 percent) (International Federation of the Phonographic Industry, 2014). These revenues would have been higher if consumers would not use unlicensed services to download music. With the use

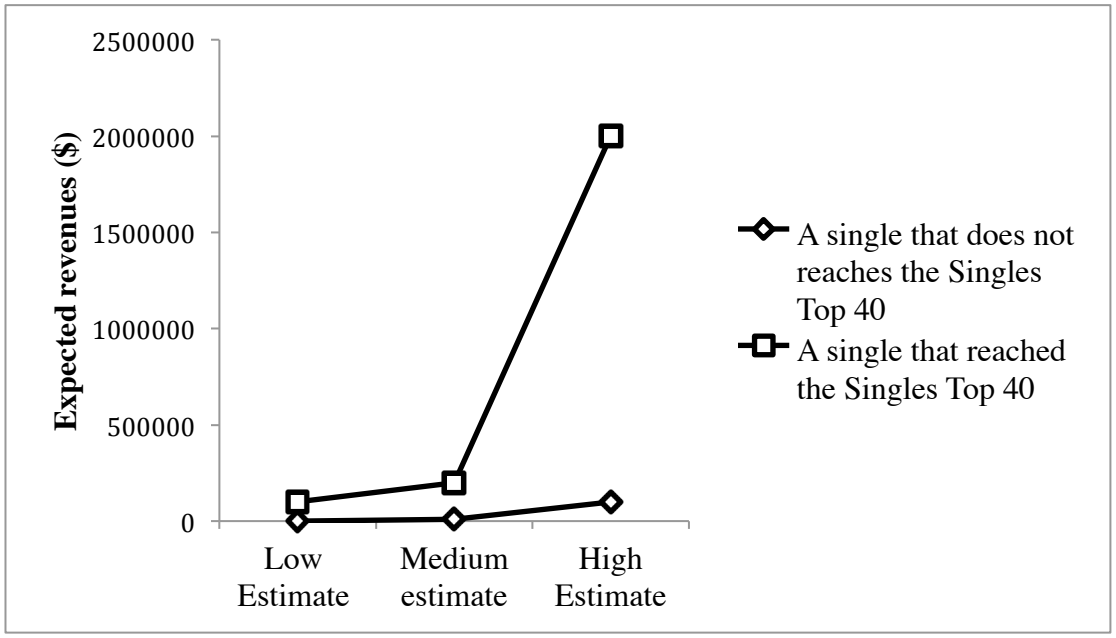
of ComScore/Nielsen data IFPI estimated that 26 percent of Internet users on a desktop regularly use unlicensed services to download music (International Federation of the Phonographic Industry, 2014).

More than 28 million consumers worldwide paid for a music subscription in 2013. The growing popularity of streaming music has resulted in growing revenues in most music markets (International Federation of the Phonographic Industry, 2014).

The company Polyphonic HMI estimated the revenues that a music single can generate in the United States (Elberse, Eliashberg & Villanueva, 2005). Graph 1 shows the expected revenues for albums with and without a single that is in the Top 40 chart position. Graph 2 shows the expected revenues for a single that does or does not reach the Top 40 chart position. The horizontal lines show the different scenarios and the vertical lines show the estimated revenues. The difference in the estimated revenues shows how important it is to be able to predict if a song will be a hit or not (Elberse et al., 2005).



Graph 1. *Expected revenues album success*



Graph 2. *Expected revenues single success*

Most senior music company managers estimate that one out of five musicians are commercially successful (International Federation of the Phonographic Industry, 2014). There are different meanings of commercial success in the music industry. It can be a Gold or a Platinum status of a release or the advance for the musician has been earned back or the musician is satisfied and all the costs are covered (International Federation of the Phonographic Industry, 2014).

There are no publicly known criteria for artists to get signed, but a few top employees of record companies have spoken about why they would give an artist a record deal. Afo Verde, who is the chairman of Sony Music and CEO for the Latin Region, Spain and Portugal, says that music needs to create emotions. Glen Barros, president chief executive of Concord Music group says that he wants to see empirical evidence that shows that there is a connection between the audience and the artist. Teresa LaBarbera Whites, who is senior Vice President of A&R at Columbia Records says that an artist needs to have star appeal. Andreas Weitkämper of Warner Music Group says that an artist needs a unique voice, the drive to

make a success of the career and a good persona (International Federation of the Phonographic Industry, 2014).

Local musicians are an important part of albums sales, as local musicians released the majority of top selling albums of 2013 in many markets (International Federation of the Phonographic Industry, 2014). An example of this is France where seventeen of the twenty best selling albums of 2013 albums were from local musicians. IFP saw this in thirteen selected markets where consumer's first language was not English (International Federation of the Phonographic Industry, 2014).

Twenty-three percent of the signed musicians are seen as new musicians, this means that they are signed in the last twelve months (Shuker 2001). Radio stations play established artists that already have a fan base more often than new artists until their songs become hits. This is because they do not want to risk losing listeners because of unpopular songs (Shuker 2001). This indicates that it is difficult for twenty-three percent of the artists to get radio airplay, which means that it is important for them to know how they improve their music to get radio airplay. They can do this by releasing songs as singles that have lyrics that are about popular themes. In turn, radio stations can play songs of new artists that are about popular themes, this decreases the chance of losing listeners.

Listeners

Music preferences of English adolescents are influenced by a high amount by their social class (Murdock & Phelps, 1973). Age and gender also influence preferences of music listeners. Younger teenagers, especially girls, like to listen to commercial pop music most. Older teenagers like to listen more to progressive music (Marshall 1997). Listeners in their late teens and early twenties prefer genres that are less commercial and popular. This is a way for them to distance themselves from adolescence. When listeners get older, they are more likely to listen to new genres and music that is less commercial. An example of this is that

'world music' is mostly listened to by middle-class listeners, who are in their thirties and forties (Shuker 2001). However, a lot of adults do not change their music preferences and keep listening to commercial music. They keep listening to the same musicians or they stop buying music or stop visiting clubs and concerts (Shuker, 2001). Music preference is also influenced by friends, since friends usually listen to the same kind of music (Ball, 1981).

Psychological effects of music

It is important to understand what the effect of music itself has on listeners to understand the effect that lyrics have on listeners. This paragraph shows the effect of music and the importance for its listeners. Lyrics are of course a part of music and so the discussed importance of music can be (partially) due to the lyrics.

Music has three psychological functions: cognitive, social and emotional functions (Hargreaves & North, 1999). The reactions that listeners have to music are based on different factors such as, the mood that they are in, their preferences, how familiar they are with the music and music training (Abeles & Chung, 1996). Ruud (1997) opinions that music has an effect on the quality of life of the listener. He suggests that music has a contribution to the quality of life in four different ways:

1. The ability to experience emotions.
2. Music gives the listeners a way to empower themselves.
3. Music makes the listener belong to a certain social group.
4. Music gives value and reason to the life of the listener.

Music is used to help patients recover from their illnesses. Patients who listen to music through headphones have a reduction of among others, stress hormone levels, pain, heart rates, anxiety, lower blood pressure rates and lower pulse (Steelman, 1990; Good, Anderson & Stanton-Hicks, 2002). Music therapy can also help patients to recover from chronic stress (Aldridge & Brandt, 1991).

Music is also used to create positive emotions for senior patients. The music is used to remember moments in their lives, to use these memories for their recovery. This shows the impact that songs can have on their listeners (Wigram, Nygaard Pedersen & Bonde, 2002).

Music components

Gómez, Klapuri and Meudic (2003) suggest that melody, rhythm, harmony, timbre and spatial are the main six dimensions for sound description. Lyrics and melody are independently of each other processed by the listener (Bonnel, Faita, Peretz & Besson, 2001).

The dimension melody effects how listeners recall lyrics. People are better at recalling the lyrics of some songs when they hear the melody, than without (Rubin, 1977; Bartlett & Snelus, 1980). Other research showed that short, unknown lyrics are better recognized when listeners hear the original melody than when they hear it with a new melody or when they hear an earlier heard melody with a different text (Crowder, Serafine & Repp, 1984; Serafine et al., 1986; Crowder, Serafine & Repp, 1990; Samson & Zatorre, 1991; Wallace, 1994). Simple symmetrical melodies creat a better text recall (Wallace, 1994). Selfridge-Field (1998) suggests that melody makes listeners able to distinguish one song from another, because melody makes music memorable.

Music influences the meaning and emotion that listeners give to lyrics. There are three different outlooks on how lyrics and music are synchronized (Thompson & Russo, 2004).

1. Assimilation: Music is dominant over lyrics, the lyrics are just a part of the music (Langer, 1957).
2. Independence: Lyrics and music are not equal to each other, because lyrics are build upon a symbolic system and music not. This means that the two components cannot be integrated together as one part (Benveniste, 1985).
3. Interaction: Music and lyrics are independent to each other, but there is also overlap between the two components, which creates interaction. This means that there is

overlap between the structural characteristics of melody and rhythm and this can increase the chance that one component can influence the interpretation of the other component (Crowder et al., 1984; Juslin & Laukka, 2003).

Effect of lyrics

There have been studies that showed the effects that lyrics have on listeners. A study done by Ali and Peynircioglu (2006) showed that lyrics influence the emotions that listeners get when listening to music. It makes it easier to convey positive and negative emotions (Ali et al., 2006). The impact of lyrics on young adolescents has been studied. This showed that most teenagers focus on the music while a minority paid more attention to the lyrics (Prinsky et al., 1987; Arnett 1991).

Lyrics are not always as strong when they are presented outside of a musical context, even when the quality is high. The musical context gives a stronger meaning to the text than when it is just text (Galizio & Hendrick, 1972). Galizio et al. (1972) suggested that this was because listeners were in a good mood, because they were listening to music and this gave them a more positive attitude towards the lyrics.

Listeners are able to identify lyrical themes even when the themes can be difficult to identify (Hansen & Hansen, 1991). Another study states that younger listeners do not recognize what the lyrics mean (Prinsky et al, 1987). Students who enrolled in a course about mass communication, did not know the meaning of songs that made references to mass media (Edwards & Singletary, 1984). Psychology undergraduates were able to correctly identify homicidal (ninety-five percent) and violent (ninety-one percent) lyrics of rap and heavy metal songs, but they were not able to correctly identify suicidal lyrics in most cases (thirty-eight percent). The subjects classified sixty-three percent of the non-violent lyrics as non-violent and thirty-seven percent classified the lyrics as homicidal. Children in the ages ten through thirteen said that lyrics are the least important factors in songs along with some other factors,

they found the beat of the song to be the most important factor (Christenson, DeBenedittis & Lindlof, 1985).

Content of lyrics

The majority of music has been about love (Friedlander, 1996). The theme love has been written about by musicians from all different genders, races and ages (Hatfield & Rapson, 1987). The theme love was the most used theme in lyrics in popular music in the early twenties (Horton, 1957). The amount of love songs did not change significantly over the time periods 1958–1972, 1976–1984 and 1991–1998. Sex was most sung about during the time period 1976-1984, during this time women sang about sex five times more than men. Men sang more about sex during 1991-1998 than women (Bisel, Borega, Dukes, Lobato & Owens, 2003). Political protest entered music in the sixties, which was a reflection of that time. Songs that were in the Top 40 were not a lot about sex and drugs (Hirsch, 1971).

Women are frequently portrayed as secondary to men in songs (Tuchman, 1978). This has changed somewhat over time since Tuchman's research (Van Zoonen, 1994). Although women nowadays are often portrayed with more diversity, more complicity and more diverse messages and the role that women have in society, it is still uncommon that women are presented as intelligent, independent, inventive or superior to men (Cooper, 1999). Country music is a genre that also shows women in a degrading matter. A study of videos of male performers showed that two-thirds of the videos degraded women (they were shown in a patronizing manner or in a traditional role). Just nine percent of the videos showed women as fully equal to men (Andsager & Roe, 1999). Fifty percent of the videos of the female country musicians showed females as fully equal to men. Country songs and videos are however not violent against women and they do not portray women as prostitutes or strippers. This is because the country music industry would not allow this (Andsager et al., 1999).

Music genres

Rock music

Rock music began to surface in the mid-1950s as rock 'n' roll (Grossberg, 1992). The baby boom after the Second World War is seen as a very important reason why the genre was able to get so popular at the time (Grossberg, 1992). The baby boom created a large youth market (fourty percent of the population of the United States was younger than twenty years) (Grossberg, 1992). These American teenagers had a high level of wealth, which made it very appealing to target them with the music they preferred (Welsh, 1990).

Rock music is a genre that consists of a wide spectrum of subgenres, a few of those are metal, hard rock and alternative rock. Rock music was the most popular genre in 2008 in the United States (Recording Industry Association of America, 2008). A big part of lyrical themes of rock music are about defiance, rebellion and the expression of youthful angst (Bleich et al., 1991). An analysis of rock music videos found that seventy-five percent of the videos showed women in a degrading manner and twenty-five percent showed women in a traditional gender role. Just fourteen percent of the rock music videos showed the women in the videos as equals to men (Vincent, Davis & Boruszkowski, 1987). Another study showed that fifty-seven percent of the examined rock videos portrayed women in a conventional manner, a third portrayed women as strong and independent (Alexander, 1999). A study done with sixty-two music videos (that were mostly rock videos) shown in 1984 on MTV showed that most shown themes were visual abstraction, sex, dance, violence, and crime (Baxter, Riemer, Landini, Leslie & Singletary, 1985).

Former research has documented that a rock music video gets more appealing for viewers when sex is shown, while violence makes them less appealing for viewers (Hansen & Hansen, 1990). The use of deliberate abuse of women is uncommon in heavy metal, which is a subgenre of rock music (Walser, 1993).

Rap music

Rap music is a genre that originated in the mid-1970s in New York City on the streets and in the clubs (Pareles & Romanwoski, 1983). Rap was first recorded in the late 1970s (Pareles et al., 1983). Afro-American youth started to listen to the genre in the early 1980s (Pareles et al., 1983). The lyrics were getting more complex and subtle over the years since rap was evolving (Simpson, 1990). Rap was becoming a part of the musical mainstream by the mid-1980s. Musicians who do not have the means to buy relatively cheap instruments or instruments at all are more likely to make music, like rap, which does not require instruments (Blake, 1992). This has resulted in underprivileged urban blacks in the 1980s that made rap music with their voices, turntables and cheap drum machines. Usually richer black musicians who are able to pay for relatively expensive instruments prefer to make jazz-funk (Blake, 1992).

Rap is a genre that, like rock music consists of a wide spectrum of subgenres: crossover, parody, rock, booty, pimp, don, jazz, new jack swing, race, G funk, dirty south, east coast and west coast gangsta (Lena, 2006). Rap was the second most popular music genre in 2008 in the United States behind rock music (Recording Industry Association of America, 2008). A study done by Mcfarland (2003) showed that the main themes in Chicano rap lyrics from 1999 to 2002 were: endorsing superiority of men over women and criticism towards racism (Mcfarland, 2003). A study was conducted in 2001 with about 490 rap lyrics during 1987-1993. This showed that twenty-two percent of the lyrics were violent towards women, the violence consisted of assault, rape and murder (Armstrong, 2001). Watkins (2001) says that the lyrics of rap songs are mostly about violence and misogyny, because these themes sell best, although Watkins does not show how he came to these findings. Rap is a way for listeners, especially White youth, to get to know the ghetto (Quinn, 2005). The major music labels released up to five and a half times the amount of hardcore rap singles than all the independent music labels combined (Lena, 2006). Rap artists are stimulated by producers to

make more hardcore music and rap artists who are not hardcore often get rejected (Weitzer, 2009). Many rappers that have political and social messages have started to rap about money and sex because of this (Powell, 2000).

A difference between heavy metal and rap is that the lyrics of heavy metal do not deliberately discuss abuse towards women in contrary to rap music, which contains a lot of misogyny. A study about rap and heavy metal songs from 1985 to 1990 showed that rap lyrics were more sexually explicit and graphic, while heavy metal was more about male domination and sexual acts (Binder, 1993).

The described previously conducted content analyses showed that rock lyrics were mostly about defiance, rebellion, misogyny and expression of youthful angst (Bleich et al., 1991). Rap lyrics are mostly about misogyny, racial injustice / inequality and violence (Watkins, 2001; Mcfarland, 2003). This gives the indication that these themes are popular among music listeners of these genres and so the following statements can be formulated and will be examined:

1. The lyrical themes defiance, rebellion, misogyny and expression of youthful angst have positive effects on the popularity of a rock song.
2. The lyrical themes misogyny, racial injustice / inequality and violence have positive effects on the popularity of a rap song.

Methodology

Research design

The data set consists of songs released in 2013 that are streamed on Last.fm. The data consists of music from two genres, rock music and rap music. The goal of this research is to examine

the effect of lyrical themes on song popularity of the mentioned genres, rock music and rap music.

Procedure

The popularity of the selected songs was measured with the amount of listeners of a song on Last.fm. Last.fm is a music service that lets listeners stream music, it is used by listeners all over the world. The music that was used needed to be released in 2013 as a part of an album. Only songs that can be listened to on Last.fm were used for this study. The list of albums released in 2013 was used to create a list of musicians. A list on Wikipedia with released albums in the United States in 2013 was used for this. This list was filtered by two criteria: rock music and rap music. Two separate lists were made: rock and rap music lists. These lists were put in a randomizer provided by random.org to create a random sample since not all of the albums will be used for the analysis.

Dataset

The sample size was determined with a calculation from Krejcie and Morgan (1970).

The following equation and variables were used to determine the sample size:

$$s = X^2NP(1-P) \div d^2\,(N-1) + X^2P\,(1-P)$$

X2 = The value of chi-square with a degree of freedom of 1 at the desired confidence level (3.841)

N = the population size (1,000,000+)

P = the population proportion (.50)

d = the degree of accuracy stated as a proportion (.05)

There was no indication of how many rock and rap songs annually get released by musicians worldwide. This meant that it was not possible to use an estimation to calculate the sample

size. Krejcie and Morgan describe that the sample size does not increase when the population size is higher than 1,000,000. Since it was not very time consuming to collect the data and there were no costs involved, the highest population size was used to be sure that a sufficient sample size was used. Krejcie and Morgan say in their article that a population proportion of .50 generates the highest sample size, so this population proportion was used to generate a sufficient sample size. A degree of accuracy of .05 was used.

This resulted in the following equation and a required sample size of 384 songs per genre:

$$s = 3.841^2 \times 1{,}000{,}000 \times .50(1-.50)\ .05^2(1{,}000{,}000 - 1) + 3.841^2 \times .50\ (1 - .50) = 384.$$

A list of 268 rock albums and 240 rap albums could be obtained. The researched songs were collected from these albums. The most listened songs on the albums were selected. Two songs per rock album were used for the research, to use an equal amount of songs per album. The amount of used songs per rap album ranged between 2 and 3, because not all of the 240 rap albums were listed on Last.fm. This means that the amount of albums was not large enough to collect 384 songs by selecting 2 songs per album. This means that the maximum of 2 songs per rap album had to be raised. The rap songs were eventually selected from 144 rap albums. The lyrics were collected from the websites rapgenius.com, metrolyrics.com, sing365.com, azlyrics.com and lyricsmode.com.

The following variables were used for the study:

Dependent variables

1.The amount of listeners of a song on Last.fm.

Independent variables

1. Lyrical themes. The paragraph 'Text mining' in this chapter shows how the themes of the lyrics were determined.

2. Discography. The more albums a musician releases the more known the musician will become and this can lead to higher popularity (Adler, 1985). The information was collected from the musicians personal website, compilation albums and mix tapes were included *(Control variable)*.

3. Listeners per musician: Already existing popularity of the musician could effect the number of listenings. The number of listeners that the musician has on Last.fm was collected for this variable *(Control variable)*.

4. Scrobbles per musician: Every time a song of the musician is listened to, the song is added to the listeners music profile, this is called a scrobble. This means that the variable shows the popularity of the musician. This variable could have the same effect as the control variable 'listeners per musician' *(Control variable)*.

5. Month. There is a possibility that older songs are listened to more because they have been longer on Last.fm. This information was collected from Last.fm. The months were coded as dummies (1 it is the given month and 0 it is not the given month), which means that there were 12 variables. One month per genre was excluded because the model would get overidentified if all the month variables were included. The month April was the month with the highest frequency in the rock dataset and in the rap dataset and so this month was excluded from both datasets *(Control variable)*.

6. Position on album: It is possible that songs that are on the beginning of the album get listened to more because listeners do not have the time to listen to the entire album. Each song was given the number that it had on the track list of the album on Last.fm *(Control variable)*.

7. Track list length: A song can be listened to more times when an album is shorter. A listener can repeat an album more often, which means that the songs will be played more. The variable was measured at a continues level *(Control variable)*.

8. Song length: It is possible that listeners prefer songs of a certain length. The variable was measured at a continues level *(Control variable)*.

The descriptive statistics of the variables can be seen in 'Appendix three: Descriptive statistics'.

Text mining

Text mining is defined as a process where important information is extracted from unstructured text (Dörre, 1999; Netzer, Feldman, Goldenberg & Fresko, 2012). Text mining is useful when there is a huge amount of information that needs to be analysed (Cavusgil, 2003). The value of text mining increases when the amount of unstructured data increases (Fan, Wallace, Rich, Zhang, 2006). Text mining is useful for a wide spectrum of research fields (Fan et al., 2006). Examples of these are, intelligence departments of governments and security agencies that need to analyse security threats like terrorist warnings. Another example are businesses that have a very high amount of information. They can use text mining to analyse, for example, their competition, marketing strategies and customers (Fan et al., 2006). Text mining was first done at tag level and was later also used on term level (Feldman, et al., 1998). This research also applies to text mining on term level. Text mining is a novel approach to marketing research that has gained popularity in the last couple of years (Lee & Bradlow, 2011). A few examples of studies that have used this new research technique are: Ludwig, Ruyter, Friedman, Brüggen (2013), they have used text mining to find changes in semantics in customer reviews on books on Amazon.com. Ghose, Ipeirotis, Li (2012) have used text mining to analyse user generated content to be able to rank products to

give customers the best recommendations. Decker and Trusov (2010) have used text mining to examine the effect of brand names and product attributes on the evaluation that customers give to products. Eliashberg, Hui and Zhang (2007) used text mining to predict movie returns based on movie scripts.

The text mining was done with the program Rapidminer. Figure 1 shows the main process of the text mining. The two text files were separately entered in Rapidminer with the 'Read Excel' operator. The operator 'Process Documents from Data' processed the lyrics and prepared it for analysis. The operator used Term Frequency-Inverse Document Frequency (TF-IDF) to create vectors based on the importance of a word. TF-IDF is determined by multiplying TF with IDF. TF is determined by counting the number of times a term is used in a lyric and IDF is determined with the following formula: IDF(t) = log (Total number of lyrics / Number of lyrics with term t present). An absolute prune method with a minimum of 2 and a maximum of 99 was used to ignore words that are used to frequent or to infrequent. There are words that are used in most of the songs and that are used frequently and so these words could give insignificant information, because they would not help in distinguishing between lyrical themes. This could 'hide' more important information.

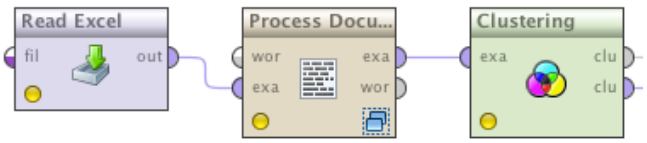


Figure 1. Main process text mining.

Figure 2 shows the process of 'cleaning' the data to be able to cluster the lyrics. The lyrics were first tokenized to divide the text into groups of basic elements, examples of these are separators and words (Cavusgil, 2003). 'Filter Stopwords' was used to filter English stopwords, the stopwords were build in by Rapidminer. Stopwords are words that are

frequently used, like the word ‘the’. The ‘Stem-Snowball’ operator makes it able for Rapidminer to recognize different forms of a word as the same word (Hu, 2005). The ‘Transform Cases’ operator was used to transform all the words in the lyrics to lower cases. The ‘Filter tokens’ operator filtered all the words that had less than 2 characters or more than 99 characters.

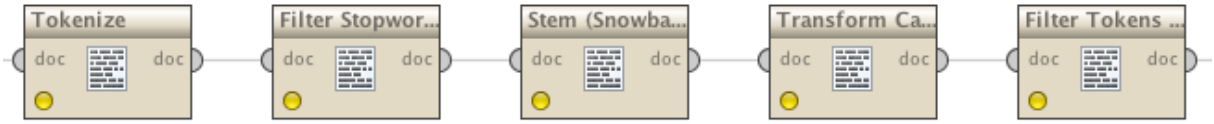


Figure 2. Transformation process.

Similarities between groups can be found by using clustering (Fan et al., 2006, see for example Archak, Ghose, Ipeirotis, 2011). This is why the lyrics were clustered, they were clustered by using K-means, a numerical measure type was used with Cosine similarity (see for example Lee et al., 2011). Cosine similarity measures how similar two vectors are based on a scale of 0-1. The number of clusters was determined by searching for the highest R^2 value, which was calculated by using linear regression with Hubert-White estimators (see for example Lee et al., 2011). The best amount of clusters for the rock lyrics was 7 with a R^2 value of .367 and the best amount of clusters for the rap lyrics was 13 with a R^2 value of .372. Table 1 shows the different amounts of clusters that were created to determine the highest R^2 value for the rock clusters and Table 2 shows this for the rap clusters. Rapidminer showed words that were frequently used in each of the clusters, all the words were given a frequency score. The clusters were named based on these words and their frequency scores.

Table 1. *Maximisation Rock Clusters*

Number of clusters	c
2	.202
5	.340
6	.357
7*	.367
8	.359

* 7 Clusters created the highest R^2 value and so 7 themes were used in the analysis.

Table 2. *Maximisation Rap Clusters*

Number of clusters	R^2 value
2	.051
5	.181
8	.250
10	.310
12	.324
13*	.372
14	.365

* 13 Clusters created the highest R^2 value and so 13 themes were used in the analysis.

The themes are mutually exclusive, which means that a song can only contain one theme and there is no overlap. The themes and the accompanying words can be seen in 'Appendix One: Theme clusters'.

Results

Data inspection

The rock and rap data were analysed separately since not all predictor variables were the same per genre and this would influence the results. First, the outliers were detected with casewise diagnostics with 2 standard deviations. This showed 16 outliers in the rock dataset and 20 outliers in the rap dataset, the outliers were excluded from the analysis. Secondly, data inspection was done to check if the data met the assumptions so that the regression model could be generalized. The assumptions discussed by Andy Field (2009) were analysed.

- **Linearity:** Linearity means that the dependent variable is a linear function of the independent variables (Darlington, 1968). A scatterplot of residuals from the linear regression was used to test if there was a linear relationship between the dependent variable and the independent variables. A diagonal line was drawn and it was checked if the residuals were around that line. The independent variables of both genres did not have a linear effect on the dependent variables since most of the values of the dependent variable were around the diagonal line. The dependent variables of both genres were log transformed to make the effect linear; Chintagunta, Gopinath and Venkataraman (2010) used a similar approach when they analysed box office performances in the movie industry. The scatterplots showed that this made the effect linear since most of the values of the dependent variables were not around the diagonal line. The scatterplots can be seen in 'Appendix two: Log-transformed scatterplots'.
- **Variable type:** The independent variables needed to be quantitative or categorical (with a maximum of 2 categories) and the outcome variable needed to be quantitative, unbounded and continues. The variable MONTH was changed to twelve different variables (1 it is the given month and 0 it is not the given month) to make the variable

categorical with two categories. The lyrical variables also consisted of two categories (1 the theme is present and 0 the theme is not present). Some variables were excluded because the model would get overidentified if all the month variables and all the lyrics variables were included. The variables *LYRICSheartbreak* and *MONTHApril* had the highest frequency in the rock dataset and so these variables were excluded. The variables *LYRICSsex* and *MONTHApril* had the highest frequency in the rap dataset and so these variables were excluded. The variables TRACK LIST LENGTH, POSITION ON THE ALBUM, DISCOGRAPHY, SONG LENGTH, LISTENERS PER MUSICIAN and SCROBBLES PER MUSICIAN were quantitative. The outcome variables NUMBER OF LISTENINGS were quantitative, continuous and unbound. The outcome variables were log transformed to make the effect linear.

- **Independence:** The values of the dependent variables needed to be independent from each other, which means that they needed to be from separate entities (Field, 2009). The values of the outcome variables all came from different songs, which guaranteed the independence.
- **Non zero-variance:** The predictor variables needed to have non-zero variance, which means that they needed to have variances that are not equal to zero (Field, 2009). This variance was measured with a descriptive test that showed the variances of all the predictor variables. All the predictor variables had a variance that was not equal to zero. The variances can be seen in 'Appendix three: Descriptive statistics'.
- **Predictors are uncorrelated with 'external variables:** This assumption cannot be tested, but there are no variables that are expected to correlate with the internal variables.
- **Independent errors**: The residual terms of any two observations should be independent, which means that they need to be uncorrelated. All the observations are

independent and there is no connection between observations besides genre and year of release (2013), which are not variables in this research. This means that the assumption is met.

- **Normally distributed errors:** The residuals in the model needed to be random and normally distributed with a mean of zero (Field, 2009). This was measured with the residual statistics. The residuals statistics showed that both datasets had a mean score of zero, which means that the assumption was not violated.
- **No perfect multicollinearity:** The predictor variables needed to be uncorrelated (Darlington, 1968). This was measured with collinearity diagnostics. The variance inflation factor (VIF) will show whether there is a strong linear relationship between the predictors. There is not a rule about the VIF score, but Myers (1990) states that a score higher than 10 is too high. Field (2009) states that a tolerance score below .10 is a problem.

  Rock dataset: The lowest tolerance score was .204 and the highest VIF score was 4.136 for the variable SCROBBLES PER MUSICIAN. The highest tolerance score was .959 and the lowest VIF score was 1.043 for the variable SONG LENGTH.

  Rap dataset: The variable *MONTHFebruary* had the lowest tolerance score with .527 and the highest VIF score with 1.897. The variable SONG LENGTH had the highest tolerance score with .951 and the lowest VIF score with 1.051. The scores can be seen in 'Appendix four: No perfect multicollinearity test'.
- **Homoscedasticity:** A scatterplot (Field, 2009) and Breusch-Pagan test (Breusch and Pagan, 1979) were used to test if there was homoscedasticity. The scatterplots of both genres showed a low spread at one end of the plot and a wide spread at the other end, this means that there was heteroscedasticity instead of homoscedasticity. The scatterplots can be seen in 'Appendix two: Log-transformed scatterplots'. The

Breusch-Pagan test of both genres had a score of zero, which means that there was heteroscedasticity. This matched with the scatterplot.

Huber-White estimators were used to account for the heteroscedasticity in the data and to create a regression analysis (Basu, Mazumdar and Raj, 2003).

Data-Analysis

Lyrical themes

Table 3 shows the themes that were present in the rock songs and the frequency of occurrence as derived from the content analysis. The data shows that there were 7 themes present in the rock songs. The most common themes in the rock lyrics were *heartbreak* (22.1%) and *empowerment* (22.1%). *Violence* (14.6%), *abandonment* (17.4%) and *honesty* (10.7) were also frequently discussed. The themes *hell* (7.8%) and *surviving* (5.2%) were discussed the least.

Table 3. *Rock Themes*

Themes	**Frequency**	**Percentage**
Heartbreak	85	22.1%
Empowerment	85	22.1%
Abandonment	67	17.4%
Violence	56	14.6%
Honesty	41	10.7%
Hell	30	7.8%
Surviving	20	5.2%

Table 4 shows the themes that were present in the rap lyrics and the frequency of occurrence as derived from the content analysis. The data shows that there were 13 themes present in the rap songs. The most common themes in the rap lyrics where *sex* (16.7%), *ghetto* (14.1%) and *hitting rock bottom* (11.5%). The other themes were *bragging* (9.4%), *wealth and status* (8.1%), *women* (7.6%), *being the best* (6.5%), *punishment* (5.5%), *drugs* (5.2%), *prostitution* (5.2%) and *gangsta life* (3.6%).

Table 4. *Rap Themes*

Themes	**Frequency**	**Percentage**
Sex	64	16.7%
Ghetto	54	14.1%
Rock bottom	44	11.5%
Bragging	36	9.4%
Wealth and status	31	8.1%
Women	30	7.8%
Being the best	25	6.5%
Punishment	21	5.5%
Drugs	20	5.2%
Prostitution	20	5.2%
Gangsta life	14	3.6%
Empowerment	13	3.4%
Shooting	12	3.1%

Comparing rock themes to rap themes

The most noticeable difference between the two genres is that there are more themes discussed in rap music than in rock music (13 > 7). The theme *empowerment* was present in the lyrics of both genres, although it was much more discussed in rock music (22.1% > 3.4%). There were no other themes that overlapped although *shooting* (rap theme) could be seen as a part of the theme *violence* (rock theme).

Effect of lyrical themes on song popularity

Table 5 shows that only one of the themes in the rock songs had a significant effect on song popularity namely, *empowerment*. It had an effect of .395, which means that the theme had a positive effect. The other variables that had a significant effect were *listeners per musician* (.001) and s*crobbles per musician* (.001). Both these variables show existing popularity of the musicians on Last.fm. *Listeners per musician* does this by showing how many people have listened to the musician and *scrobbles per musician* does this by showing how many times a song of the musician has been added to a listeners music profile.

Table 5. *Effect Of Rock Variables On The Amount Of Listeners Per Rock Song*

Variable	B	SE	Sig
(Intercept)	8.832	.420	.000
Lyrical theme			
Violence	.058	.209	.780
Hell	-.500	.266	.061
Empowerment	.395	.194	.042
Surviving	.487	.285	.089
Abandonment	.173	.202	.393
Honesty	.302	.245	.220
Song length	-.001	.001	.508
Position album	.027	.016	.085
Track list length	-.009	.016	.575
Discography	.004	.010	.676
Listeners per musician	.001	.001	.000
Scrobbles per musician	.001	.001	.003
January	-.077	.327	.814
February	.038	.269	.889
March	.186	.317	.559
May	-.370	.315	.241
June	.102	.269	.706
July	-.045	.294	.879
August	.034	.369	.926
September	.014	.302	.962
October	-.071	.249	.776
November	-.187	.286	.513
December	.289	.294	.326

* R^2 = .367

** When *HEARTBREAK* was included in the model and *EMPOWERMENT* was excluded, *HEARTBREAK* had B value of -.451, a significance score of .149, with a SE value of .228.

*** When *JUNE* was included in the model and *APRIL* was excluded, *JUNE* got a B value of -.140 and a significance score of .673, with a SE value of .331.

Table 6 shows that the theme *gangsta life* is the only theme that had a significant effect on popularity of rap music. The theme had an effect of -1.279, which means that it had a negative effect in comparison to the other themes. The other variables that had a significant effect were *track list length* (.034), *discography* (-.048), *scrobbles per musician* (.001), *listeners per musician* (.001), *January* (-.769), *July* (-.853), *August* (-.841), *December* (-.935).

Table 6. *Effect Rap Variables On The Amount Of Listeners Per Rap Song*

Variable	B	SE	Sig.
(Intercept)	7.599	.360	.000
Empowerment	-.867	.560	.122
Being The Best	-.129	.347	.710
Wealth And Status	-.625	.385	.105
Gangsta Life	-1.276	.416	.002
Women	.072	.321	.822
Bragging	.392	.279	.161
Drugs	.157	.309	.612
Ghetto	.085	.297	.775
Rock Bottom	-.140	.273	.608
Shooting	-.697	.579	.229
Prostitution	.161	.329	.625
Punishment	.294	.281	.297
Song length	-.001	.001	.350
Position on album	-.024	.025	.342
Track list length	.034	.013	.010
Discography	-.048	.013	.000
Listeners per musician	.001	.001	.011
Scrobbles per musician	.001	.001	.010
January	-.769	.359	.033
February	-.520	.298	.081
March	.060	.412	.884
May	-.039	.329	.907
June	.389	.334	.245
July	-.853	.384	.027
August	-.841	.379	.027
September	-.521	.384	.176
October	-.066	.326	.839
November	-.592	.379	.119
December	-.935	-.441	.035

* $R^2 = .372$

** When *SEX* was included in the model and *GHETTO* was excluded, *SEX* got a B value of .448 and a significance score of .250, with a SE value of .389.

*** When *APRIL* was included in the model and JUNE was excluded, APRIL got a B value of -3.82 and a significance score of .370, with a SE value of .425.

Conclusion and discussion

Conclusion

The goal of this report was to examine which themes in rock music and rap music had positive effects on rock and rap popularity, this resulted in the following research question:

What is the difference of the effect of lyrical themes on the popularity of a song between rock and rap music?

The following statements, based on past literature, were examined as sub-research statements:

1. The lyrical themes defiance, rebellion, misogyny and expression of youthful angst have positive effects on the popularity of a rock song.
2. The lyrical themes misogyny, racial injustice / inequality and violence have positive effects on the popularity of a rap song.

First, the results of this research showed that the mentioned themes in the statements were not present in the current rap and rock lyrics. Although the rock themes *violence* and *surviving* can be seen as a part of *defiance* and *rebellion*. Rap themes like *shooting* and *punishment* can be seen as a part of *violence*. Women are also discussed in the rap songs but the results did not show misogyny. Because the themes were not directly present in the content analysis it was not possible to test if the themes in the statements had a positive effect on popularity. This is why the effects of the themes that are present in the current rock and rap lyrics are being discussed.

Second, the content analysis of the rock lyrics showed that there are seven themes in rock lyrics. The most common themes are *empowerment* and *heartbreak*. Other discussed themes are *abandonment*, *violence*, *honesty*, *hell* and *surviving*. *Empowerment* is the only theme that has a significant effect on popularity. *Empowerment* has a positive effect on

popularity.

Third, the content analysis of the rap lyrics showed that there are thirteen themes in rap lyrics. The most common themes are *ghetto*, *sex* and *hitting rock bottom*. The other themes were *bragging*, *wealth and status*, *women*, *being the best*, *punishment*, *drugs*, *prostitution*, *gangsta life*, *empowerment* and *shooting*. *Gangsta life* is the only theme that has a significant effect on popularity. *Gangsta life* has a negative effect on popularity.

Fourth, the difference between the two genres is that they both have mostly different themes and that the theme *empowerment* (which is present in both genres) only has a significant effect on rock popularity. Another difference is that the explanatory value of the R-square was the highest at seven themes for rock music and at thirteen themes for rap music.

Discussion

Content

The content analysis of both genres showed differences with research that has been done regarding rock and rap lyrics. Bleich et al., (1991) found that defiance, rebellion, misogyny and expression of youthful angst were the most discussed themes in rock music. These themes were not found in the rock lyrics. The differences could be existing because they did their research in 1991 and the dataset of this research is from 2013, this would mean that the themes of rock lyrics have changed over time. The content analysis also showed strong differences with content analysis of music videos (Baxter et al., 1985). The content analysis showed the presence of the themes *violence*, *visual abstraction*, *sex*, *dance* and *crime*. Only *violence* was also present in the lyrics of this research. The difference probably exists because the themes were based on sixty-two music videos, which were not all rock videos and because the music videos are from 1984. A similarity was found with Walser`s research (1995). Walser found that *misogyny* was not present in heavy metal (a subgenre of rock), this was also not found in the content analysis of the rock lyrics.

The rap lyrics showed a similarity with the findings of Lena (2006). The theme *sex* was also present in the findings of Lena (2006). The other themes were not present in the findings. A possible explanation is that the themes have changed over the years. It could also be because of another method of naming the themes, but the method is not discussed in the article, so it is not possible to compare the methods. Mcfarland (2008) found that Chicano rap songs were mostly about *endorsing superiority of men over women* and *criticism towards racism*. These themes were not found in the results, although *women* are the main theme in 7.8 per cent of the rap songs. This difference is probably because this article discusses rap lyrics and Mcfarland`s article only the subgenre Chicano rap.

Music in general is said to be mostly about love (Horten, 1957; Friedlander, 1996; Bisel et al, 2003). Sex is also a much discussed topic in music (Bisel et al., 2003). The results of this research show that s*ex* was the most discussed theme in rap music, but love was not present in both genres, which shows a big difference in content.

It was expected that themes that were mentioned in other researches where themes that had a positive effect on popularity. The fact that almost all of the themes were not present in the analysed songs shows that lyrical content has changed over time.

Popularity

Watkins stated that *violence* and *misogyny* had positive effects on rap sales based on the opinions of industry members. This research shows that *violence* did not have a significant effect on popularity and *misogyny* was not present in the songs of this sample (although *women* was a theme in the rap songs). This could mean that preferences of listeners have changed over the years. This research also shows that *empowerment* was the only rock theme that had a significant effect on popularity, the effect was positive. The only rap theme that had a significant effect on popularity was *gangsta life*, the effect of *gangsta life* was negative.

The fact that just two of the themes of the two genres did have a significant effect on

popularity means that there are other factors that have an effect on popularity besides the significant control variables in this research. These findings agree with Langer's (1957) findings, which said that music is dominant over lyrics when it comes to influencing listeners. This research did not examine if music is dominant over music but the fact that just two themes had a significant effect, could mean that music has a stronger effect than lyrics.

The findings also show similarities with (Christenson et al., 1985) who found that children in the ages ten through thirteen said that lyrics are the least important factors in songs along with some other factors.

Managerial implications

Empowerment is the most discussed theme in rock music (tie with the theme *heartbreak*). This is positive for popularity and it could be stimulated to release songs with *empowerment* as singles or put a high amount of songs with this theme on albums. *Gangsta life* is already a not regularly discussed theme in rap music, which should be maintained since it has a negative effect on popularity. The fact that the other themes did not have a significant effect implies that it is not extra beneficial to invest in songs / musicians that discuss these themes. The same thing applies for radio stations, they should also play rock music with the theme *empowerment* and they should not play rap music with the theme *gangsta life*. These implications are for organizations that aim to reach a mass market since this research did not look at effects of lyrical themes on popularity of subgenres.

Limitations

A limitation of the research method was that the labels of the clusters were based on personal interpretation of the given words by the software program Rapidminer. Another person could interpret the words differently and give the clusters different labels. The words that were used to name the clusters are listed in 'Appendix One: Theme clusters', this gives other researchers

the chance to evaluate if the clusters have been given the right labels and perhaps rename them.

Another limitation is that the analysis does not show subtleties in the lyrics, it cannot be said if every theme is positive or negative. Future analysis should show semantics to create a better view on the meaning of the songs.

A third limitation is that mutually exclusive themes were used, which means that a song can only contain one theme. While, this does not always have to be the case. The lyrics of a song can contain multiple themes.

A fourth limitation is that the variables 'listeners per musician' and 'scrobbles per musician' were collected after the songs were released so these variables could have been influenced by the dependent variable 'number of listenings'.

Future research

Future research should look at the content and effects of lyrics of the subgenres of rock and rap music, since this research did not do this. Future research could use the themes described in this research to see which subgenres contain which themes.

Future research could also replicate this research for other genres to know how lyrical themes effect the popularity of other genres.

This research used mutually exclusive themes, future research could latent themes, because the lyrics of a song can contain multiple themes.

A last recommendation for future research is based on the fact that most of the themes did not have a significant effect, this can mean that lyrical content does not have an effect on listeners preferences. It needs to be researched if listeners have preferences in themes. This research looked at what the songs were about and then looked if these themes were popular. Future research could ask listeners if they find lyrical content important and what themes they prefer.

References

Abeles, H. F. & Chung, J. W. (1996), Responses to music. In D. A. Hodges (Ed.), Handbook music psychology, San Antonio: University of Texas.

Adler, M. (1985), Stardom and Talent, American Economic Review 75: 208–212.

Aldridge, D. & Brandt G. (1991), Music therapy and inflammatory bowel disease, The Arts and Psychotherapy, 18, 113–121.

Alexander, S. (1999), The gender role paradox in youth culture: An analysis of women in music videos, Michigan Sociological Review 13:46-64.

Ali, S. O. & Peynircioglu Z. F. (2006), Songs and emotions: are lyrics and melodies equal partners?, Psychology of music, 34: 511.

Andsager, J. & Kimberly R. (1999), Country music video in country's year of the woman, Journal of Communication 49:69-82.

Archak, N., Ghose A., Ipeirotis P.G., (2011), Deriving the pricing power of product features by mining consumer reviews, Management Science, 57:1485–1509.

Armstrong, G. (2001), Gangsta misogyny: A content analysis of the portrayals of violence against women in rap music, 1987-1993. Journal of Criminal Justice and Popular Culture, 8:96-126.

Arnett, J. (1991), Heavy Metal Music and Reckless Behavior among Adolescents, Journal of Youth and Adolescents, 20: 573–92.

Ball, S. J. (1981), Beachside Comprehensive, Cambridge: Cambridge University Press.

Bartlett, J. C. & Snelus P. (1980), Lifespan memory for popular songs, American Journal of Psychology, 93, 551-560.

Basu, A., Mazumdar T. & Raj S. P. (2003). Indirect network externality effects on product attributes, Marketing Science, 22, 209-221.

Baxter, R. L., De Riemer C., Landini A., Leslie L. & Singletary M. W. (1985), A content analysis of music videos, Journal of Broadcasting & Electronic Media, 3, 333-340.

Benveniste, E. (1985). The semiology of language. In: R.E Innis (Ed,) Semiotics: An introductory anthology, Bloomington: Indiana University Press, 226-246.

Binder, A. (1993), Media depictions of harm in heavy metal and rap music, American Sociological Review, 58:753-67.

Bisel, T. M., Borega K. N., Dukes R. L., Eligio A. L. & Owens M. D. (2003), Expressions of love, sex, and hurt in popular songs: a content analysis of all-time greatest hits, The Social Science Journal, 40, 643–650.

Blake, A. (1992) The Music Business, London: Batsford.

Bleich, S., Zillmann D. & Weaver, J. (1991), Enjoyment and consumption of defiant rock

music as a function of adolescent rebelliousness, Journal of Broadcasting and Electronic Media, 35, 351-366.

Bonnel, A. M., Faita F., Peretz I. & Besson M. (2001), Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing, Perception & Psychophysics, 63, 1201-1213.

Breusch, T. S. & Pagan A. R. (1979), A simple test for heteroscedasticity and random coefficient variation, Econometrica, 47, 1287–1294.

Christenson, P., DeBenedittis P. & Lindlof T. (1985), Children's use of audio media. Communication Research, 12, 327-345.

Cavusgil, S.T., Chan, K. & Zhang, C. (2003), Strategic orientations in export pricing: a clustering approach to create firm taxonomies, Journal of International Marketing, 11, 47-72.

Chintagunta, P.K., Gopinath S. & Venkataraman S. (2010), The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. Marketing Science, 29, 944–957.

Cooper, B. L. (1999), From Lady Day to Lady Di: Images of women in contemporary recordings,1938-1998, International Journal of Instructional Media, 26, 353-58.

Crowder, R. G., Serafine M. L. & Repp B. H. (1984), Integration of melody and text in memory for songs, Cognition, 16, 285-303.

_____, _____ & _____ (1990), Physical interaction and association by contiguity in memory for the words and melodies of songs, Memory & Cognition, 18, 469-476

Darlington, R. B. (1968), Multiple regression in psychological research and practice, Psychological Bullitin, 69, 161-182.

Decker, R. & Trusov M. (2010), Estimating aggregate consumer preferences from online product reviews, International journal of research marketing, 27(4), 293–307.

Dörre, J., Gerstl P. & Seiffert R. (1999), Text mining: finding nuggets in mountains of textual data, In Proc. 5th ACM Int. Conf. on knowledge discovery and data mining ACM press, New York, US, 398–401.

Edwards, E., & Singletary M. (1984), Mass media images in popular music. Popular Music and Society, 9, 17-26.

Elberse, A., Eliashberg J. & Villanueva J. (2005), Polyphonic HMI: Mixing music and math, Harvard Business School Case Study #HBS-9-506-009.

Eliashberg J, Hui S.K. & Zhang J.Z. (2007), From story line to box office: A new approach for green-lighting movie scripts, Management science, 53(6), 881–893.

Gómez, E., Klapuri A. & Meudic B. (2003), Melody description and extraction in the context of music content processing, Journal of New Music Research, 32, 23-40.

Fan, W., Wallace L., Rich S., Zhang Z. (2006), Tapping the power of text mining, Communication of the ACM, 49(9), 77-82.

Feldman, R. Fresko M., Kinar Y., Lindell Y., Liphstat O., Rajman M., Schler Y. & Zamir O. (1998), Text mining at the term level, Principles of data mining and knowledge discovery, 1510, 56–64.

Field, A. (2009), Discovering statistics using SPSS, Sage Publications LTD.

Friedlander, P. (1996). Rock-and-roll: A social history, Boulder, CO: Westview Press.

Galizio, M & Hendrick C. (1972), Effect of music accompaniment on attitude: The guitar as a prop for persuasion, Journal of applied social psychology, 2, 350-359.

Ghose, A., Ipeirotis P. G. & Li (2012), Designing ranking systems for hotels using UGC, Marketing science, 31(3), 493–520.

Good, M., Anderson G. C. & Stanton-Hicks M. (2002), Relaxation and music reduce pain after gynaecologic surgery, Pain Management Nursing, 3, 61–70.

Grossberg, L. (1992), We Gotta Get Out of This Place: Popular Conservatism and Postmodern Culture, New York: Routledge.

Hansen, C. H. & Hansen R. D. (1990), The influence of sex and violence on the appeal of rock music videos, Communication Research, 17: 212.

_____ & _____ (1991). Schematic information processing of heavy metal lyrics, Communication research, 18, 357-369.

Hargreaves, D. J. & Adrian C. (1999), The functions of music in everyday life: redefining the social in music psychology, North Psychology of Music 27, 71.

Hatfield, E. & Rapson R. (1987). Passionate love: New directions in research, Advances in personal relationships, 1, 109–139, Greenwich, CT: JAI Press.

Hirsch, P. (1971). Sociological approaches to the popular music phenomenon, American Behavioral Scientist, 74, 371-388.

Hong, S. H. (2013), Measuring the effect of napster on recorder music sales: Difference-in differences estimates under compositional changes, Journal of Applied Econometrics, 28, 297–324.

Horton, D, (1957), The dialogue of courtship in popular songs, American Journal of Sociology, 17, 569-578.

Hu, X., Downie J. S., West K. & Ehmann A. (2005), Mining music reviews: promising preliminary results, in proceedings of the sixth joint conference on digital libraries, 536-539.

International Federation of the Phonographic Industry (2014), Digital music report 2014, available at http://www.ifpi.org/downloads/Digital-Music-Report-2014.pdf.

_____ (2014), Investing in music, available at http://www.ifpi.org/content/library/investing_in_music.pdf.

Juslin, P. N. & Laukka P. (2003), Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? Psychological Bulletin, 129, 770–814.

Krejcie, R. & Morgan D. W. (1970), Determining sample size for research activities, Educational and Psychological Measurement 30, 607-610.

Langer, S (1957). Philosophy in a new key: A study in the symbolism of reason, rite, and art. Cambridge, MA: Harvard University Press.

Lee, T. Y. & Bradlow E. T. (2011), Automated marketing research using online customer reviews, Journal of marketing research, 48, 881-894.

Lena, J. C. (2006), Social context and musical content of rap music, 1979-1995, Oxford University Press, 85 479-495

List of 2013 albums (2013), available at http://en.wikipedia.org/wiki/List_of_2013_albums.

Ludwig, S., de Ruyter K., Friedman M., Brüggen E. C., Wetzels I. & Pfann G. (2013), More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates, Journal of Marketing, 77, 87-103

Negus, K. (1996), Popular music in theory, Cambridge: Polity Press.

Netzer, O., Feldman R., Goldenberg J., Fresko M. (2012), Mine your own business: Market structure surveillance through text mining, Marketing science, 31, 521-543.

Marshall, P.D. (1997), Celebrity and Power: Fame in Contemporary Culture, Minneapolis, MN/London: University of Minnesota Press.

McFarland, P (2003), Challenging the contradictions of Chicanismo in Chicano rap music and male culture, Race, Gender, and Class, 10, 92-107.

Myers, R. (1990), Classical and modern regression with applications, Boston, MA: Duxbury.

Murdock, G. & Phelps, G. (1973), Mass media and the secondary school. London: MacMillan.

Quin, E. (2005), Nuthin' but a "G" Thang: The culture and commerce of gangsta rap popular cultures, everyday lives, Columbia University Press.

Pareles, J & Romanowski P., eds (1983), The Rolling Stone Encyclopedia of Rock & Roll,

New York: Rolling Stone Press/Summit Books.

Powell, K. (Ed.). (2000). Step Into a World: A Global Anthology of the New Black Literature, New York: John Wiley & Sons, Inc.

Prinsky, L. E. & Rosenbaum J. L. (1987), 'Leer-ics' or lyrics: Teenage impressions of rock 'n' roll, Youth & Society, 18(4), 384-397.

Rubin, D. C. (1977), Very long-term memory for prose and verse, Journal of verbal learning and verbal behavior, 16, 611-621.

Ruud, E. (1997). Music and quality of life. Nordic Journal of Music Therapy, 6, 86–91.

Samson, S, Zatorre R. J. (1991), Recognition memory for text and melody of songs after unilateral temporal lobe lesion: evidence for dual encoding, Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 793-804.

Selfridge-Field, E. (1998). Conceptual and representational issues in melodic comparison, melodic similarity – concepts, procedures, and applications, Cambridge, Massachusetts: MIT Press.

Serafine, M. L., Davidson J., Crowder R. & Repp B. (1986), On the nature of melody – text integration in memory for songs, Journal of memory and language 25, 123-135.

Shuker, R. (2001). Understanding popular music, New York, Routledge.

Simpson, J. C. (1990,), Yo! Rap gets on the map: Led by groups like Public Enemy, it socks a Black message to the mainstream, Time, 60-62.

Steelman, V. M. (1990). Intraoperative music therapy. Effects on anxiety, blood pressure. AORN journal, 52, 1026–1034.

Stefanac, S, (1986), Rock n' roll repression. Maximum Rock n' Roll, September, 18-27,

The Recording Industry Association of America (2008), 2008 Consumer Profile, available at http://76.74.24.142/CA052A55-9910-2DAC-925F-27663DCFFFF3.pdf.

Thompson, W. F. & Russo F. A. (2004), The attribution of emotion and meaning to song lyrics, Polskie forum psychologiczne, 9(1), 51-62.

Tuchman, G. (1978), Making news: a study in the construction of reality, New York: Free Press.

Vincent, R., Davis D. & Boruszkowski L. (1987), Sexism on MTV: the portrayal of women in rock videos, Journalism quarterly, 64, 750-55.

Wallace, W. T. (1994), Memory for music: Effect of melody on recall of text, Journal of experimental psychology: Learning, memory and cognition, 20, 1471-1485.

Walser, R. (1993), Running with the Devil: Power, gender, and madness in heavy metal music, Hanover, NH: Wesleyan University Press.

Watkins, S. C. (2001) A nation of millions: Hip hop culture and the legacy of black nationalism, The Communication Review 4, 373-98.

Weitzer, R. & Kubrin C. E. (2009), Misogyny in rap music: A content analysis of prevalence

and meanings, men and masculinities, 12, 3-29.
Welsh, R. (1990), Rock 'n' roll And social change', History today, February, 32–9.
Wigram, T., Pedersen N, I. & Bonde L. O. (2002). A comprehensive guide to music therapy. London: Jessica Kingsley Publishers.
Van Zoonen, L. (1994), Feminist media studies, Thousand Oaks: Sage.
Zentner, A. (2006), Measuring the effect of file sharing on music purchases, Journal of law and economics, 49, 63-90.

Appendix One: Theme clusters

Rock clusters

Cluster 1 rock: Heartbreak

girl	.034
damn	.026
trouble	.023
friend	.022
break	.021
lose	.020
miss	.020
baby	.018

Cluster 2 rock: Violence

suck	.027
hate	.027
murder	.024
blood	.021
heart	.021
fall	.021
kill	.020
cold	.020
live	.018
silence	.017

Cluster 3 Rock: Hell

live	.056
hell	.054
earth	.037
protect	.036
commit	.036
force	.030
fuck	.029
power	.024
sacrifice	.028
burn	.023
devil	.022
soul	.022
prey	.021
heart	.019
virus	.018
evil	.018

Cluster 4 Rock: Empowerment

want	.047
fire	.039
breath	.034
alright	.025
hold	.024
start	.022
wake	.022
lost	.021
live	.017
believe	.016

Cluster 5 Rock: Surviving

pray	.127
fight	.038
survive	.037
drive	.035
forever	.032

Cluster 6 Rock: Abandonment

stay	.045
move	.035
remember	.034
journey	.028
call	.022
keep	.021
gone	.020
hurt	.019

Cluster 7 Rock: Honesty

please	.073
believe	.046
truth	.042
brave	.026
honest	.025
heart	.020

Rap clusters

Cluster 1 Rap: Empowerment

power	.229
coward	.070
alive	.043
rule	.040
fire	.026
stronger	.022

Cluster 2 Rap: Being the best

king	.063
shine	.032
lord	.024
crown	.022
skyscraper	.021

Cluster 3 Rap: Wealth and status

girl	.031
porsche	.030
hollywood	.019
respect	.019
women	.019

Cluster 4 Rap: Gangsta life

nigger	.210
thiev	.043
motherfucker	.038
thief	.033
sword	.032
compton	.024
street	.022
kill	.021
devil	.021

Cluster 5 Rap: Women

girl	.053
body	.040
dirty	.035
party	.034
pussy	.032

Cluster 6 Rap: Bragging

dippin	.028
walk	.028
shout	.023
scream	.021
holdin	.020
riding	.018
club	.016

Cluster 7 Rap: Drugs

fiend	.040
thirsty	.037
weed	.032
rush	.023
lust	.020

Cluster 8 Rap: Ghetto

gang	.018
black	.016
nothing	.016
shot	.015
hood	.015
living	.015
killer	.014
street	.014
block	.014

Cluster 9 Rap: Rock bottom

help	.027
fall	.027
bottom	.021
grave	.018
dark	.017
kill	.017
thought	.016
resist	.015

Cluster 10 Rap: Sex

tonight	.042
lady	.036
dream	.033
night	.030
bang	.028
baby	.027

Cluster 11 Rap: Shooting

power	.178
coward	.058
immortal	.040
nine	.034
backup	.031
fire	.022
stronger	.018

Cluster 12 Rap: Prostitution

gucci	.044
golden	.040
babi	.039
mane	.037
dope	.031
dress	.028
shorty	.028
trick	.028
girl	.026
sell	.024
dick	.024

Cluster 13 Rap: Punishment

sorry	.095
whip	.040
court	.028
problem	.028
trouble	.026

Appendix two: Log-transformed scatterplots

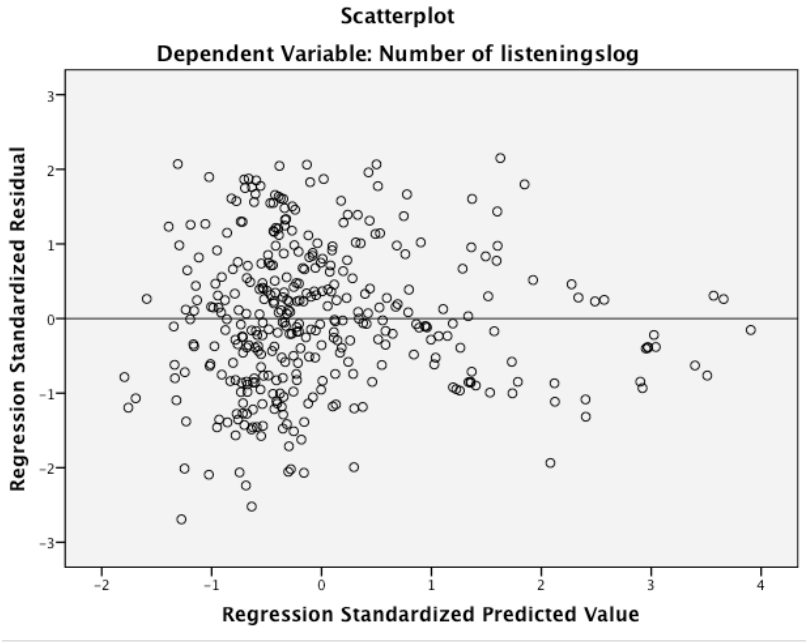


Figure 3. *Rock scatterplot*

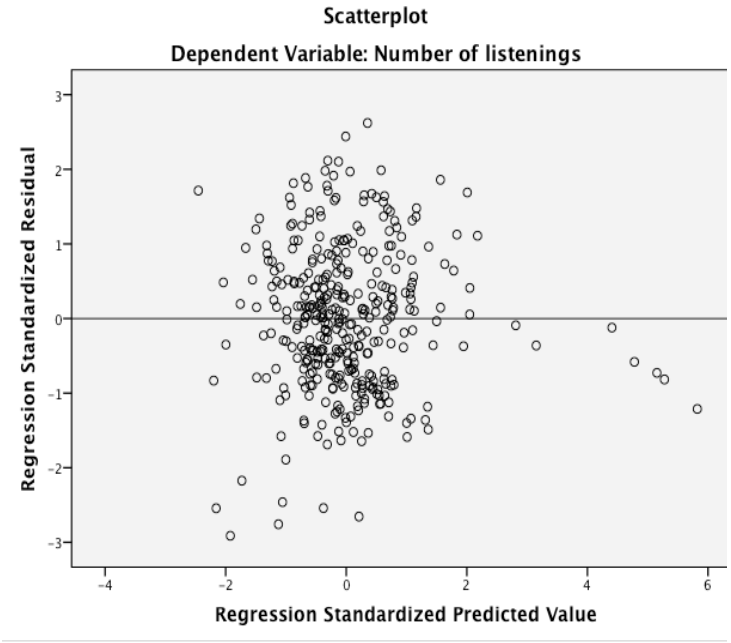


Figure 4. *Rap scatterplot*

Appendix three: Descriptive statistics

Table 7. *Rock Data*

Variables	Minimum	Maximum	Mean	Std. Deviation	Variance
Number of listenings	15	897867	41245.40	87966.548	7738113643.928
Number of listenings (log)	2.71	13.71	9.3941	1.73058	2.995
Song length	0:30	7:37	3:28	0:58	12515133.853
Position on the album	1	19	4.69	3.817	14.570
Track list length	2	45	14.73	5.626	31.650
Discography	1	37	11.20	7.139	50.972
Listeners per musician	0	3116113	620235.28	742929.133	551943696813.246
Scrobbles per musician	0	252517461	23816226.3	39475239.107	1558294502540675
Heartbreak	0	1	.22	.416	.173
Violence	0	1	.15	.353	.125
Hell	0	1	.08	.269	.072
Empowerment	0	1	.22	.416	.173
Surviving	0	1	.05	.222	.049
Abandonment	0	1	.17	.380	.144
Honesty	0	1	.10	.306	.094
January	0	1	.06	.238	.057
February	0	1	.11	.310	.096
March	0	1	.07	.256	.066
April	0	1	.11	.316	.100
May	0	1	.07	.261	.068
June	0	1	.11	.319	.102
July	0	1	.08	.277	.077
August	0	1	.05	.217	.047
September	0	1	.06	.238	.057
October	0	1	.11	.310	.096
November	0	1	.09	.292	.085
December	0	1	.07	.252	.063

Table 8. *Rap Data*

Variables	Minimum	Maximum	Mean	Std. Deviation	Variance
Number of listenings	2	242343	10179.87	29318.801	859592071.448
Number of listenings (log)	.69	12.40	7.3058	2.12647	4.522
Song length	0:30	6:57	3:27	0:57	11820125.228
Position album	1	19	4.72	3.814	14.547
Track list length	2	45	14.66	5.435	29.535
Discography	1	37	11.14	7.083	50.168
Scrobbles per musician	118	155301607	7388887.15	19439843.602	377907519252476
Listeners per musician	273	4017983	377047.37	643407.229	413972862777.19
Empowerment	0	1	.03	.181	.033
Being The Best	0	1	.06	.242	.059
Wealth and status	0	1	.08	.273	.074
Gangsta Life	0	1	.04	.188	.035
Women	0	1	.08	.269	.072
Bragging	0	1	.09	.292	.085
Drugs	0	1	.05	.228	.052
Ghetto	0	1	.14	.348	.121
Rock Bottom	0	1	.12	.322	.104
Sex	0	1	.16	.371	.138
Shooting	0	1	.03	.174	.030
Prostitution	0	1	.05	.222	.049
Punishment	0	1	.05	.228	.052
January	0	1	.06	.242	.059
February	0	1	.11	.309	.096
March	0	1	.07	.247	.061
April	0	1	.11	.316	.100
May	0	1	.08	.265	.070
June	0	1	.11	.319	.102
July	0	1	.08	.277	.077
August	0	1	.05	.222	.049
September	0	1	.06	.233	.054
October	0	1	.11	.313	.098
November	0	1	.09	.292	.085
December	0	1	.07	.252	.063

Table 9a. *Rock: Correlation Between Variables*

Variables	Number of listenings	Number of listeningslog	Song length	Position on the album	Track list length
Number of listenings	1	.617**	.005	.103*	-.008
Number of listeningslog	.617**	1	-.047	.081	-.062
Song length	.005	-.047	1	.000	-.117*
Position on the album	.103*	.081	.000	1	.172**
Track list length	-.008	-.062	-.117*	.172**	1
Discography	.035	.036	.008	.115*	.114*
Listeners per musician	.494**	.522**	-.054	.049	.001
Scrobbles per musician	.544**	.502**	-.024	.042	-.020
Heartbreak	-.011	-.014	-.016	.035	.018
Violence	-.050	-.036	.088	.003	.030
Hell	-.051	-.089	-.052	.029	.002
Empowerment	.026	.051	.021	-.072	-.039
Surviving	.045	.021	-.010	-.027	.034
Abandonment	.005	.016	-.045	-.022	-.040
Honesty	.045	.032	.000	.048	.018
January	-.063	-.059	.001	.041	.198**
February	-.023	.027	.076	-.107*	-.184**
March	.020	.021	-.032	.060	.111*
April	-.053	-.013	-.026	-.037	.025
May	.133**	-.064	.074	-.022	-.088
June	-.044	-.017	-.016	-.069	-.026
July	-.020	-.043	-.046	.052	.117*
August	.016	.042	-.047	.000	.092
September	-.014	.051	.012	.023	.049
October	.005	-.007	-.033	-.021	-.075
November	.063	.030	.002	.071	-.016
December	-.004	.043	.033	.055	-.124*

**. Correlation is significant at the .01 level (2-tailed).
*. Correlation is significant at the .05 level (2-tailed).

Table 9b. *Rock: Correlation Between Variables*

Variables	Discography	Listeners per musician	Scrobbles per musician	Heartbreak	Violence	Hell
Number of listenings	.035	.494**	.544	-.011*	-.050	-.051
Number of listeningslog	.036**	.522	.502	-.014	-.036	-.089
Song length	.008	-.054	-.024	-.016	.088*	-.052
Position on the album	.115*	.049	.042	.035	.003**	.029*
Track list length	.114	.001	-.020*	.018**	.030	.002*
Discography	1	.043	.010	-.097*	.061*	-.026
Listeners per musician	.043**	1**	.863	.082	-.055	-.070
Scrobbles per musician	.010**	.863**	1	.115	-.035	-.067
Heartbreak	-.097	.082	.115	1	-.220	-.155
Violence	.061	-.055	-.035	-.220	1	-.120
Hell	-.026	-.070	-.067	-.155	-.120	1
Empowerment	.020	.006	-.019	-.284	-.220	-.155
Surviving	-.006	.011	.052	-.125	-.097	-.068
Abandonment	.006	.034	-.010	-.245	-.190	-.134
Honesty	.035	-.047	-.055	-.182	-.141	-.099
January	.016	-.099	-.086	-.002	.051**	-.073
February	.098	-.047	-.026	.019*	.024**	-.006
March	.261	-.031	-.012	.050	-.056*	-.004**
April	-.037	.048	.018	-.034	.014	.078
May	-.005**	-.010	.011	-.029	-.002	.068
June	-.060	-.006	-.038	.045	-.033	-.044
July	-.004	-.051	-.051	-.025	.009*	.018
August	-.040	.074	.021	.023	.110	-.022
September	-.001	.084	.077	.024	.020	.049
October	-.156	-.025	.015	-.002	-.071	-.006**
November	-.015	.058	.051	-.085	-.006	.006
December	-.023	.016	.027	.031	-.023*	-.078

**. Correlation is significant at the .01 level (2-tailed).
*. Correlation is significant at the .05 level (2-tailed).

Table 9c. *Rock: Correlation Between Variables*

Variables	Empowerment	Surviving	Abandonment	Honesty	January	February	March
Number of listenings	.026	.045**	.005	.045*	-.063	-.023	.020**
Number of listeningslog	.051**	.021	.016	.032	-.059	.027	.021**
Song length	.021	-.010	-.045	.000	.001*	.076	-.032
Position on the album	-.072*	-.027	-.022	.048	.041**	-.107*	.060
Track list length	-.039	.034	-.040*	.018**	.198	-.184*	.111
Discography	.020	-.006	.006	.035*	.016*	.098	.261
Listeners per musician	.006**	.011**	.034	-.047	-.099	-.047	-.031
Scrobbles per musician	-.019**	.052**	-.010	-.055	-.086	-.026	-.012**
Heartbreak	-.284	-.125	-.245	-.182	-.002	.019	.050
Violence	-.220	-.097	-.190	-.141	.051	.024	-.056
Hell	-.155	-.068	-.134	-.099	-.073	-.006	-.004
Empowerment	1	-.125	-.245	-.182	-.002	-.002	-.024
Surviving	-.125	1	-.108	-.080	.089	-.043	.027
Abandonment	-.245	-.108	1	-.157	-.029	-.003	.035
Honesty	-.182	-.080	-.157	1	-.014	-.007	-.027
January	-.002	.089	-.029	-.014	1**	-.087	-.069
February	-.002	-.043	-.003	-.007*	-.087**	1	-.095
March	-.024	.027	.035	-.027	-.069*	-.095**	1
April	.045	-.048	-.079	.038	-.091	-.124	-.099
May	.043**	.024	-.023	-.063	-.071	-.097	-.077
June	-.074	.026	.050	.038	-.091	-.124	-.099
July	-.047	.014	.060	-.010	-.076*	-.104	-.083
August	-.006	.001	-.042	-.078	-.058	-.079	-.063
September	-.029	.040	-.029	-.086	-.064	-.087	-.069
October	-.002	-.043	.041	.075	-.087	-.120**	-.095
November	.087	-.035	-.030	.066	-.081	-.111	-.088
December	.006	-.017	.040	.010	-.068*	-.093	-.074

**. Correlation is significant at the .01 level (2-tailed).
*. Correlation is significant at the .05 level (2-tailed).

Table 9d. *Rock: Correlation Between Variables*

Variables	April	May	June	July	August	September	October	November
Number of listenings	-.053	.133**	-.044	-.020*	.016	-.014	.005**	.063**
Number of listeningslog	-.013**	-.064	-.017	-.043	.042	.051	-.007**	.030**
Song length	-.026	.074	-.016	-.046	-.047*	.012	-.033	.002
Position on the album	-.037*	-.022	-.069	.052	.000**	.023*	-.021	.071
Track list length	.025	-.088	-.026*	.117**	.092	.049*	-.075	-.016
Discography	-.037	-.005	-.060	-.004*	-.040*	-.001	-.156	-.015
Listeners per musician	.048**	-.010**	-.006	-.051	.074	.084	-.025	.058**
Scrobbles per musician	.018**	.011**	-.038	-.051	.021	.077	.015**	.051
Heartbreak	-.034	-.029	.045	-.025	.023	.024	-.002	-.085*
Violence	.014	-.002	-.033	.009	.110	.020	-.071	-.006
Hell	.078	.068	-.044	.018	-.022	.049	-.006	.006
Empowerment	.045	.043	-.074	-.047	-.006	-.029	-.002	.087
Surviving	-.048	.024	.026	.014	.001	.040	-.043	-.035
Abandonment	-.079	-.023	.050	.060	-.042	-.029	.041	-.030
Honesty	.038	-.063	.038	-.010	-.078	-.086	.075	.066
January	-.091	-.071	-.091	-.076	-.058**	-.064	-.087	-.081
February	-.124	-.097	-.124	-.104*	-.079**	-.087	-.120	-.111
March	-.099	-.077	-.099	-.083	-.063*	-.069**	-.095	-.088
April	1	-.101	-.129	-.108	-.082	-.091	-.124	-.116
May	-.101**	1	-.101	-.085	-.064	-.071	-.097	-.090
June	-.129	-.101	1	-.108	-.082	-.091	-.124	-.116
July	-.108	-.085	-.108	1	-.069*	-.076	-.104	-.097
August	-.082	-.064	-.082	-.069	1	-.058	-.079	-.073
September	-.091	-.071	-.091	-.076	-.058	1	-.087	-.081
October	-.124	-.097	-.124	-.104	-.079	-.087**	1	-.111
November	-.116	-.090	-.116	-.097	-.073	-.081	-.111	1
December	-.097	-.076	-.097	-.081	-.061*	-.068	-.093	-.087

**. Correlation is significant at the .01 level (2-tailed).
*. Correlation is significant at the .05 level (2-tailed).

Table 9e. *Rock: Correlation Between Variables*

Variables	December
Number of listenings	-.004
Number of listeningslog	.043**
Song length	.033
Position on the album	.055*
Track list length	-.124
Discography	-.023
Listeners per musician	.016**
Scrobbles per musician	.027**
Heartbreak	.031
Violence	-.023
Hell	-.078
Empowerment	.006
Surviving	-.017
Abandonment	.040
Honesty	.010
January	-.068
February	-.093
March	-.074
April	-.097
May	-.076**
June	-.097
July	-.081
August	-.061
September	-.068
October	-.093
November	-.087
December	1

**. Correlation is significant at the .01 level (2-tailed).
*. Correlation is significant at the .05 level (2-tailed).

Rap correlations

Table 10a. *Rap: Correlation Between Variables*

Variables	Number of listenings	Number of listenings	Length	Position album	Track list length	Discography
Number of listenings	1	.571**	.073	.027	-.024	-.043
Number of listenings	.571**	1	.050	-.077	.004	-.155**
Song length	.073	.050	1	-.002	-.086	.022
Position album	.027	-.077	-.002	1	.187**	.107*
Track list length	-.024	.004	-.086	.187**	1	.087
Discography	-.043	-.155**	.022	.107*	.087	1
Listeners per musician	.650**	.430**	.083	.052	-.037	-.022
Scrobbles per musician	.728**	.414**	.068	.041	-.048	-.006
Empowerment	.082	-.028	.021	.048	-.031	-.057
Being The Best	-.052	.002	.021	-.032	.028	-.069
Wealth and status	.177**	.013	.002	.052	.068	.000
Gangsta Life	-.044	-.071	.057	-.029	-.052	-.008
Women	.036	.062	-.035	-.088	-.055	-.033
Bragging	.046	.094	.044	.014	-.072	-.003
Drugs	-.044	.005	.026	.072	.002	-.014
Ghetto	-.071	-.078	.076	.053	-.034	.024
Rock Bottom	.025	-.001	-.103*	.025	.081	-.028
Sex	-.064	-.023	-.088	-.017	.064	.073
Shooting	-.042	-.072	.055	-.038	-.011	.134**
Prostitution	-.004	.016	.030	.005	.028	.002
Punishment	-.028	.063	-.029	-.093	-.074	-.037
January	-.064	-.064	.011	.028	.185**	.015
February	-.102*	-.067	.045	-.101*	-.170**	.104*
March	-.009	.022	-.009	.075	.048	.236**
April	-.003	.061	-.031	-.032	.045	-.033
May	-.050	.045	.078	-.028	-.082	.011
June	.151**	.164**	-.014	-.072	-.023	-.058
July	-.087	-.134**	-.045	.049	.125*	-.002
August	-.050	-.044	-.044	.008	.095	-.015
September	.112*	.047	.014	-.005	.057	-.025
October	-.034	.056	-.038	-.003	-.073	-.158**
November	.052	-.036	.004	.068	-.013	-.013
December	.080	-.094	.036	.053	-.125*	-.021

**. Correlation is significant at the .01 level (2-tailed).
*. Correlation is significant at the .05 level (2-tailed).

Table 10b. *Rap: Correlation Between Variables*

Variables	Listeners per musician	Scrobbles per musician	Empowerment	Being The Best	Wealth and status	Gangsta Life
Number of listenings	.650	.728**	.082	-.052	.177	-.044
Number of listenings	.430**	.414	-.028	.002	.013	-.071**
Song length	.083	.068	.021	.021	.002	.057
Position album	.052	.041	.048	-.032	.052**	-.029*
Track list length	-.037	-.048	-.031	.028**	.068	-.052
Discography	-.022	-.006**	-.057	-.069*	.000	-.008
Listeners per musician	1**	.888**	.055	.001	.099	.067
Scrobbles per musician	.888**	1**	.092	-.025	.060	.032
Empowerment	.055	.092	1	-.048	-.055	-.036
Being The Best	.001	-.025	-.048	1	-.077	-.050
Wealth and status	.099**	.060	-.055	-.077	1	-.058
Gangsta Life	.067	.032	-.036	-.050	-.058	1
Women	.051	.056	-.054	-.075	-.086	-.057
Bragging	.012	.016	-.060	-.083	-.095	-.063
Drugs	-.018	-.024	-.045	-.062	-.071	-.047
Ghetto	-.014	-.045	-.076	-.104	-.120	-.079
Rock Bottom	.037	.043	-.068*	-.094	-.108	-.071
Sex	-.123	-.087	-.083	-.114	-.131	-.086
Shooting	-.064	-.027	-.034	-.046	-.053	-.035**
Prostitution	-.045	-.031	-.044	-.061	-.069	-.046
Punishment	-.014	-.007	-.045	-.062	-.071	-.047
January	-.106	-.084	.011	-.022	.002**	.007
February	-.083*	-.075	-.065	.050*	-.010**	.068*
March	-.037	-.005	-.049	-.025	.038	.005**
April	-.015	-.015	.025	-.058	.016	.019
May	.015	-.030	.001	.089	.060	-.003
June	.053**	.079**	.023	-.025	-.017	-.026
July	-.022	-.042**	-.056	.078	-.020*	-.059
August	-.071	-.051	-.044	-.012	-.026	-.046
September	.087*	.074	.078	-.064	.009	.072
October	.006	.001	-.019	.047	-.043	.021**
November	.015	.025	-.011	-.009	.003	-.063
December	.158	.121	.121	-.070	-.004*	.003

**. Correlation is significant at the .01 level (2-tailed).
*. Correlation is significant at the .05 level (2-tailed).

Table 10c. *Rap: Correlation Between Variables*

Variables	Women	Bragging	Drugs	Ghetto	Rock Bottom	Sex	Shooting	Prostituti
Number of listenings	.036	.046**	-.044	-.071	.025	-.064	-.042**	-.004**
Number of listenings	.062**	.094	.005	-.078	-.001	-.023**	-.072**	.016**
Song length	-.035	.044	.026	.076	-.103	-.088	.055	.030
Position album	-.088	.014	.072	.053	.025**	-.017*	-.038	.005
Track list length	-.055	-.072	.002	-.034**	.081	.064	-.011	.028
Discography	-.033	-.003**	-.014	.024*	-.028	.073	.134	.002
Listeners per musician	.051**	.012**	-.018	-.014	.037	-.123	-.064	-.045**
Scrobbles per musician	.056**	.016**	-.024	-.045	.043	-.087	-.027**	-.031
Empowerment	-.054	-.060	-.045	-.076	-.068	-.083	-.034	-.044
Being The Best	-.075	-.083	-.062	-.104	-.094	-.114	-.046	-.061
Wealth and status	-.086**	-.095	-.071	-.120	-.108	-.131	-.053	-.069
Gangsta Life	-.057	-.063	-.047	-.079	-.071	-.086	-.035	-.046
Women	1	-.094	-.070	-.118	-.106	-.129	-.052	-.068
Bragging	-.094	1	-.077	-.130	-.117	-.142	-.058	-.075
Drugs	-.070	-.077	1	-.097	-.088	-.107	-.043	-.056
Ghetto	-.118	-.130	-.097	1	-.147	-.179	-.073	-.095
Rock Bottom	-.106	-.117	-.088*	-.147	1	-.161	-.065	-.085
Sex	-.129	-.142	-.107	-.179	-.161	1	-.080*	-.104
Shooting	-.052	-.058	-.043	-.073	-.065	-.080**	1	-.042
Prostitution	-.068	-.075	-.056	-.095	-.085	-.104	-.042	1
Punishment	-.070	-.077	-.058	-.097	-.088	-.107	-.043	-.056
January	-.035	-.046	.033	.019	.040**	-.056	.015*	.085
February	.025*	-.024	-.046	-.067*	-.047**	-.017*	.035	.147
March	-.037	-.012	-.017	.045	-.096	.054**	.074	.033
April	-.042	.027	-.049	-.001	-.027	.088	-.016	-.083
May	-.046	.010	-.069	.026	-.043	.006	.005	-.067
June	.139**	-.032**	.021	-.051	.021	-.005	.029	-.011
July	.053	.000**	.093	.041	-.022*	-.032	.000	-.071
August	-.068	-.035	-.056	.040	.133	-.009	-.042	.103
September	-.030*	-.041	.039	-.035	.015	.012	.020	-.007
October	-.009	.002	.062	.026	.054	-.065**	-.063	-.045
November	.040	.080	.001	-.002	-.006	.026	-.058	-.075
December	-.040	.056	-.019	-.020	-.002*	-.007	.011**	.030*

**. Correlation is significant at the .01 level (2-tailed).
*. Correlation is significant at the .05 level (2-tailed).

Table 10d. *Rap: Correlation Between Variables*

Variables	Punishment	January	February	March	April	May	June	July
Number of listenings	-.028	-.064**	-.102	-.009	-.003	-.050	.151**	-.087*
Number of listenings	.063**	-.064	-.067	.022	.061	.045**	.164**	-.134*
Song length	-.029	.011	.045	-.009	-.031	.078	-.014	-.045
Position album	-.093	.028	-.101	.075	-.032**	-.028*	-.072	.049
Track list length	-.074	.185	-.170	.048**	.045	-.082	-.023	.125
Discography	-.037	.015**	.104	.236*	-.033	.011	-.058	-.002
Listeners per musician	-.014**	-.106**	-.083	-.037	-.015	.015	.053	-.022*
Scrobbles per musician	-.007**	-.084**	-.075	-.005	-.015	-.030	.079**	-.042
Empowerment	-.045	.011	-.065	-.049	.025	.001	.023	-.056
Being The Best	-.062	-.022	.050	-.025	-.058	.089	-.025	.078
Wealth and status	-.071**	.002	-.010	.038	.016	.060	-.017	-.020
Gangsta Life	-.047	.007	.068	.005	.019	-.003	-.026	-.059
Women	-.070	-.035	.025	-.037	-.042	-.046	.139	.053
Bragging	-.077	-.046	-.024	-.012	.027	.010	-.032	.000
Drugs	-.058	.033	-.046	-.017	-.049	-.069	.021	.093
Ghetto	-.097	.019	-.067	.045	-.001	.026	-.051	.041
Rock Bottom	-.088	.040	-.047*	-.096	-.027	-.043	.021	-.022
Sex	-.107	-.056	-.017	.054	.088	.006	-.005*	-.032
Shooting	-.043	.015	.035	.074	-.016	.005**	.029	.000
Prostitution	-.056	.085	.147	.033	-.083	-.067	-.011	-.071
Punishment	1	-.015	.028	-.017	.060	.018	-.051	-.031
January	-.015	1	-.089	-.068	-.092**	-.074	-.093*	-.078
February	.028*	-.089	1	-.091*	-.123**	-.099*	-.124	-.104
March	-.017	-.068	-.091	1	-.094	-.075**	-.095	-.080
April	.060	-.092	-.123	-.094	1	-.101	-.128	-.107
May	.018	-.074	-.099	-.075	-.101	1	-.103	-.086
June	-.051**	-.093**	-.124	-.095	-.128	-.103	1	-.108
July	-.031	-.078**	-.104	-.080	-.107*	-.086	-.108	1
August	-.005	-.061	-.081	-.062	-.083	-.067	-.084	-.071
September	-.010*	-.064	-.085	-.065	-.088	-.070	-.089	-.074
October	.026	-.090	-.121	-.092	-.124	-.100**	-.126	-.106
November	.001	-.083	-.111	-.085	-.114	-.092	-.116	-.097
December	-.019	-.070	-.093	-.071	-.096*	-.077	-.097**	-.081*

**. Correlation is significant at the .01 level (2-tailed).
*. Correlation is significant at the .05 level (2-tailed).

Table 10e. *Rap: Correlation Between Variables*

Variables	August	September	October	November	December
Number of listenings	-.050	.112**	-.034	.052	.080
Number of listenings	-.044**	.047	.056	-.036	-.094
Song length	-.044	.014	-.038	.004	.036
Position album	.008	-.005	-.003	.068	.053**
Track list length	.095	.057	-.073	-.013**	-.125
Discography	-.015	-.025**	-.158	-.013*	-.021
Listeners per musician	-.071**	.087**	.006	.015	.158
Scrobbles per musician	-.051**	.074**	.001	.025	.121
Empowerment	-.044	.078	-.019	-.011	.121
Being The Best	-.012	-.064	.047	-.009	-.070
Wealth and status	-.026**	.009	-.043	.003	-.004
Gangsta Life	-.046	.072	.021	-.063	.003
Women	-.068	-.030	-.009	.040	-.040
Bragging	-.035	-.041	.002	.080	.056
Drugs	-.056	.039	.062	.001	-.019
Ghetto	.040	-.035	.026	-.002	-.020
Rock Bottom	.133	.015	.054*	-.006	-.002
Sex	-.009	.012	-.065	.026	-.007
Shooting	-.042	.020	-.063	-.058	.011
Prostitution	.103	-.007	-.045	-.075	.030
Punishment	-.005	-.010	.026	.001	-.019
January	-.061	-.064	-.090	-.083	-.070**
February	-.081*	-.085	-.121	-.111*	-.093**
March	-.062	-.065	-.092	-.085	-.071
April	-.083	-.088	-.124	-.114	-.096
May	-.067	-.070	-.100	-.092	-.077
June	-.084**	-.089**	-.126	-.116	-.097
July	-.071	-.074**	-.106	-.097	-.081*
August	1	-.058	-.082	-.075	-.063
September	-.058*	1	-.086	-.079	-.066
October	-.082	-.086	1	-.113	-.094
November	-.075	-.079	-.113	1	-.087
December	-.063	-.066	-.094	-.087	1*

**. Correlation is significant at the .01 level (2-tailed).
*. Correlation is significant at the .05 level (2-tailed).

Appendix four: No perfect multicollinearity test

Table 11. *Rock Collinearity Diagnostics*

Variables	Tolerance	VIF
Song length	.959	1.043
Position on the album	.925	1.081
Track list length	.834	1.198
Discography	.859	1.164
Listeners per musician	.240	4.163
Scrobbles per musician	.243	4.110
Violence	.680	1.471
Hell	.770	1.299
Empowerment	.618	1.618
Surviving	.840	1.190
Abandonment	.662	1.511
Honesty	.724	1.380
January	.654	1.530
February	.541	1.850
March	.607	1.648
May	.636	1.571
June	.543	1.840
July	.609	1.643
August	.710	1.409
September	.679	1.473
October	.554	1.806
November	.598	1.672
December	.640	1.562

Table 12. *Rap Collinearity Diagnostics*

Variables	Tolerance	VIF
Song length	.945	1.058
Position album	.906	1.103
Track list length	.825	1.212
Discography	.872	1.147
Listeners per musician	.199	5.019
Scrobbles per musician	.193	5.180
Empowerment	.818	1.223
Being The Best	.731	1.368
Wealth and status	.702	1.425
Gangsta Life	.811	1.233
Women	.697	1.434
Bragging	.682	1.467
Drugs	.756	1.323
Ghetto	.597	1.676
Rock Bottom	.628	1.593
Shooting	.841	1.188
Prostitution	.750	1.334
Punishment	.775	1.291
January	.649	1.542
February	.525	1.903
March	.632	1.581
May	.624	1.604
June	.537	1.863
July	.598	1.673
August	.685	1.459
September	.686	1.458
October	.545	1.833
November	.587	1.702
December	.631	1.586