

Name: Alex Link  
Computing ID: acl5mt

## **Billboard Hot 100 Lyrics Through the Decades**

### **Intro:**

Culturomics is loosely defined as making inferences about cultural trends through the quantitative analysis of digitized texts. Although there are hundreds, if not thousands, of different forms of texts, I believe different ones better capture the cultural attitude and viewpoints of a society at certain time periods. By this, I mean that there are prevailing forms of media given a society's location and point in history which best express the views of the overall public. For example, one could argue that currently social media giants (Facebook, Instagram, Twitter, etc.) are this outlet, as it gives any individual with internet access the ability to share their views at any point in time. Before this, one could argue the prior hundred years had been completely dominated by newspapers, and before that novels. However, I believe one of the most telling and lasting forms of a society's views and trends is its choice in music. Wing-Wah Law and Wai-Chung Ho, professors at the University of Hong Kong, concluded in their 2011 study (first published in February 2015), that a society's popular music reflects the ideals that are prevalent at the time it is performed or published. Following this theory, I decided it would be interesting to see the evolution of American ideals/views over the past 50 years by analyzing the annual *Billboard* Hot 100 charts of this time period. These year-end charts are calculated using the cumulative total of yearlong record sales, streaming and airplay (radio) points. By analyzing the lyrics contained in the most listened to songs each year in the United States, via Topic Modeling and Sentiment Analysis, I hoped to glean a bit of insight into the prevailing viewpoints of the American public at these points in time, and see possible shifts in these sentiments throughout the past five decades.

### **Data Collection:**

Since there was no readily available text or spreadsheet data on all these songs' lyrics, I had to utilize a combination of various packages to get the data necessary to conduct this analysis. The BeautifulSoup web-scraping package was utilized to scrap each year-end charts' info, specifically the song title and respective artist. From there, this data was feed into a package called LyricsGenius, designed by John Miller to specifically pull song and artist information from the website *Genius*, which claims to have the most extensive collection of song lyrics online (they are my personal go-to when I am trying to see what a song's lyrics are). LyricsGenius is a fantastic package, with a variety of functions that were especially useful in gathering metadata to assist further with my analysis, however, it is not one hundred percent perfect, so a little more manual cleaning of my source file was required after the web-scraping function finished running. After adding an additional data point to identify which decade each song charted, I converted my data to the required F2 table formats and was ready to begin my analysis.

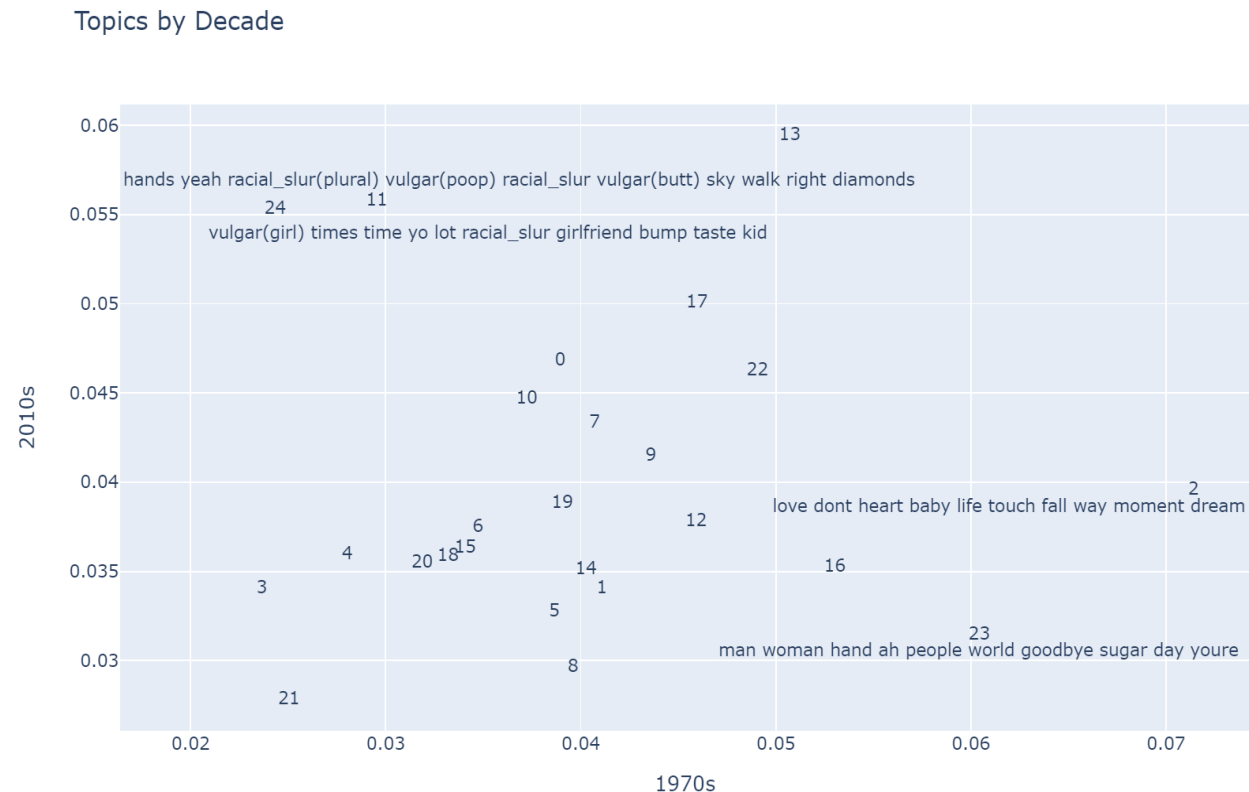
## Topic Modeling:

I began my analysis by utilizing LDA to create topic models across my corpus & see which topics were most associated with each decade. The below results represent a model generated using 5,000 terms (classified as nouns by NLTK) in the count vectorization and 25 topics in the LDA algorithm, sorted by topics most associated with the past decade (2010's).

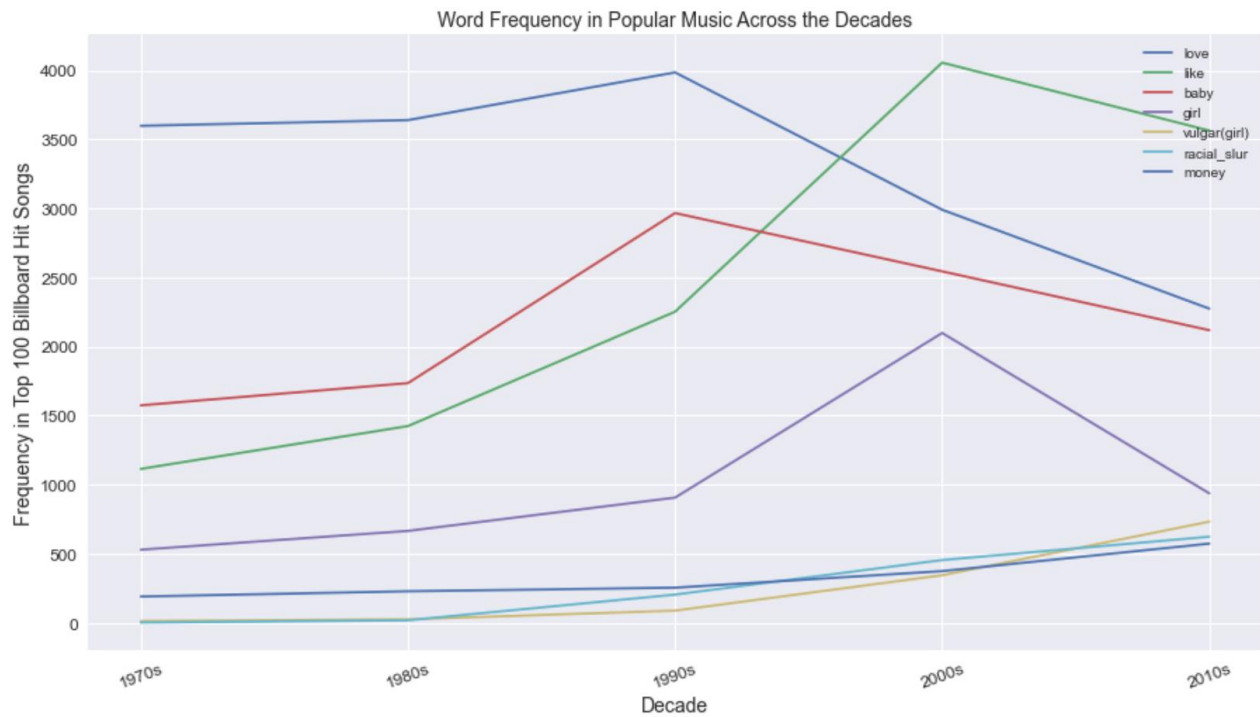
decade	1970s	1980s	1990s	2000s	2010s	topterms
topic_id						
13	0.050713	0.057716	0.055577	0.049146	0.059415	yeah heart beat door soul nights star way thunder low
11	0.029538	0.023304	0.038641	0.048835	0.055766	hands yeah racial_slur(plural) vulgar(poop) racial_slur vulgar(butt) sky walk right diamonds
24	0.024326	0.024512	0.032970	0.043120	0.055312	vulgar(girl) times time yo lot racial_slur girlfriend bump taste kid
17	0.045946	0.072068	0.066258	0.051791	0.050065	time body heart mind way matter ma pain life eyes
0	0.038945	0.045711	0.037349	0.038858	0.046812	oh eyes dreams place wind vulgar(poop) world ready bass vulgar(sex)
22	0.049058	0.037905	0.035136	0.048028	0.046251	thing life money rock time roll gimme sorry car dont
10	0.037228	0.036897	0.033816	0.045677	0.044649	tonight everybody ya head right floor yeah help reason stay
7	0.040704	0.041408	0.053515	0.040695	0.043300	baby yeah oh time ride somethin thing fallin turn dont
9	0.043582	0.037907	0.040091	0.050537	0.041456	way dance things uh lets friends club hmm care problem
2	0.071412	0.081098	0.072692	0.039848	0.039553	love dont heart baby life touch fall way moment dream
19	0.039045	0.044524	0.041807	0.042208	0.038791	mind girls cause lady dream shes truth hot heat end
12	0.045905	0.044399	0.044863	0.043217	0.037779	day world babe music life wont thats light yeah way
6	0.034725	0.036818	0.036216	0.032056	0.037441	somebody yeah lover ayy city round arms power things stars
15	0.034070	0.040292	0.038639	0.052361	0.036285	girl talk world kiss yeah lips groove whos sunshine lights
4	0.028026	0.029256	0.033664	0.040182	0.035919	youre whoa wanna crazy nah gon yeah freak booty thang
18	0.033194	0.029973	0.027367	0.035534	0.035823	chance la face rain ho game ohoh shake dat kids
20	0.031864	0.025958	0.022763	0.026277	0.035459	ooh hey doo bit town ill yeah fine aha hearts
16	0.053024	0.054292	0.045637	0.035938	0.035223	night sun morning day choo woo train words light moon
14	0.040261	0.039795	0.034783	0.036598	0.035077	home look baby yes life kind line world im dont
1	0.041074	0.040386	0.035052	0.033901	0.034011	gonna boy friend run bad change chick boys fun feelings
3	0.023650	0.027943	0.033415	0.030164	0.034009	da work dont mmm play middle friends lovin tell gonna
5	0.038641	0.028909	0.030961	0.032159	0.032726	song huh days summer breath lies bop songs mistake feel
23	0.060427	0.043838	0.044649	0.042876	0.031423	man woman hand ah people world goodbye sugar day youre
8	0.039600	0.030941	0.032664	0.030987	0.029632	air ha ba jump fool party da dee time ring
21	0.025040	0.024152	0.031473	0.029006	0.027822	alright na step minute du angel wit feels boom ahh

As can be seen in the above, the topic most associated with the past decade is topic #13, which also happens to have a fairly high association with the all the other decades of music. Given the top terms in this topic, I would wager to say this topic is indicative of the rhythm and blues genre of music. Words such as “heart”, “beat” and “soul” have always been strongly identified with R&B music, but the base genre of blues, which R&B is derived from, strongly incorporates other musical forms, such as chants (“yeah”) and spirituals, which invoked the poignant lyrics of sentimental ballads to express one’s desperate emotional state (“low”). Consequently, the genres of music considered most popular throughout the decades in the analysis are all derivatives of R&B. Rock and roll, and its subsequent forms (such as metal, punk and alternative), is widely considered the most popular musical genre throughout the 1970’s up until the mid-nineties. At this point, rap, hip-hop and contemporary R&B started seeing a rise in popularity with the American public; a trend which continues to this day.

Most telling from this analysis is when we look at the second and third topics most associated with the past decade. They greatly differ in terms of association with the pre-2000 decades, specifically the seventies and eighties. I decided to take a closer look at the evolution of popular topics by plotting those of the seventies against those of the past decade.

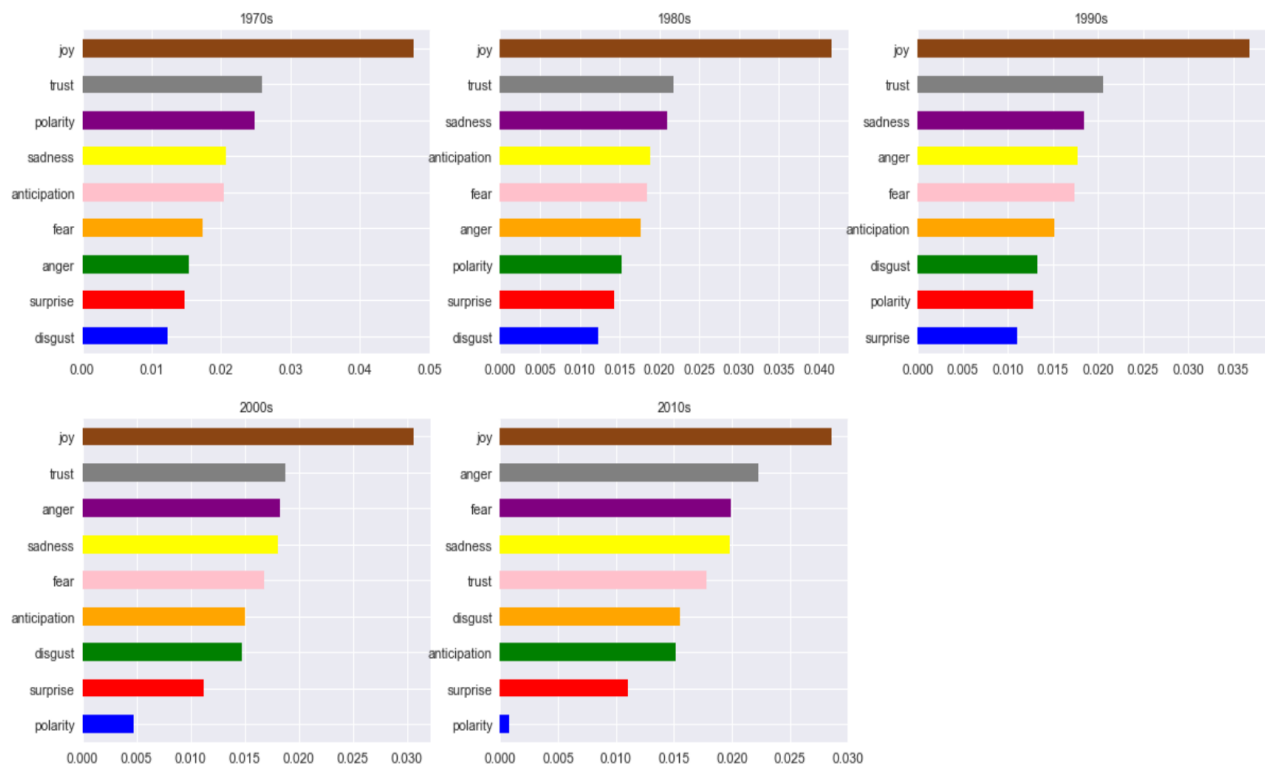


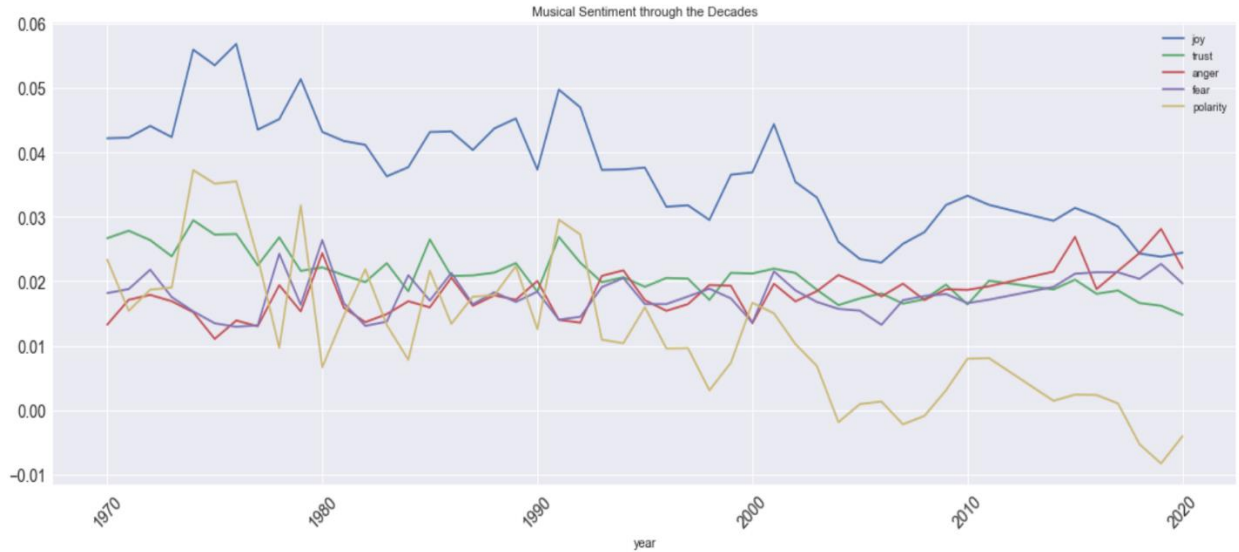
The most evident difference when looking at the above scatterplot is the increase in explicit lyrical content over the past fifty years. Music in the seventies appeared to be more emotionally loving, as evident by many of the tender terms in topics 2 (“love”, “heart”, “baby”) and 23 (“sugar”), whereas the 2010’s is associated with swear words (“a\*\*” and “sh!t”). The seventies also utilized, what many would consider, appropriate nouns for referencing others (“man”, “woman”, “people”), whereas the past decade utilizes slang, sometimes considered crass or inappropriate, language for addressing other people (“b!tch”, “kid”, n-word). Finally, it appears lyrics in the seventies were more introspective (“dream”, “life”, “moment”), but the 2010’s are more materialistic, as evident by topic 11 terms such as “diamonds” and “money” (which is the 15<sup>th</sup> top term in the topic). The below chart further illustrates the transition from loving, tender terms, to more explicit ones, as the terms “love” and “baby” have significantly decreased in lyrical usage over the past two decades, while the word “b!tch” (which is used almost as frequently as “girl” now) and the n-word have seen a steady increase over the same time period.



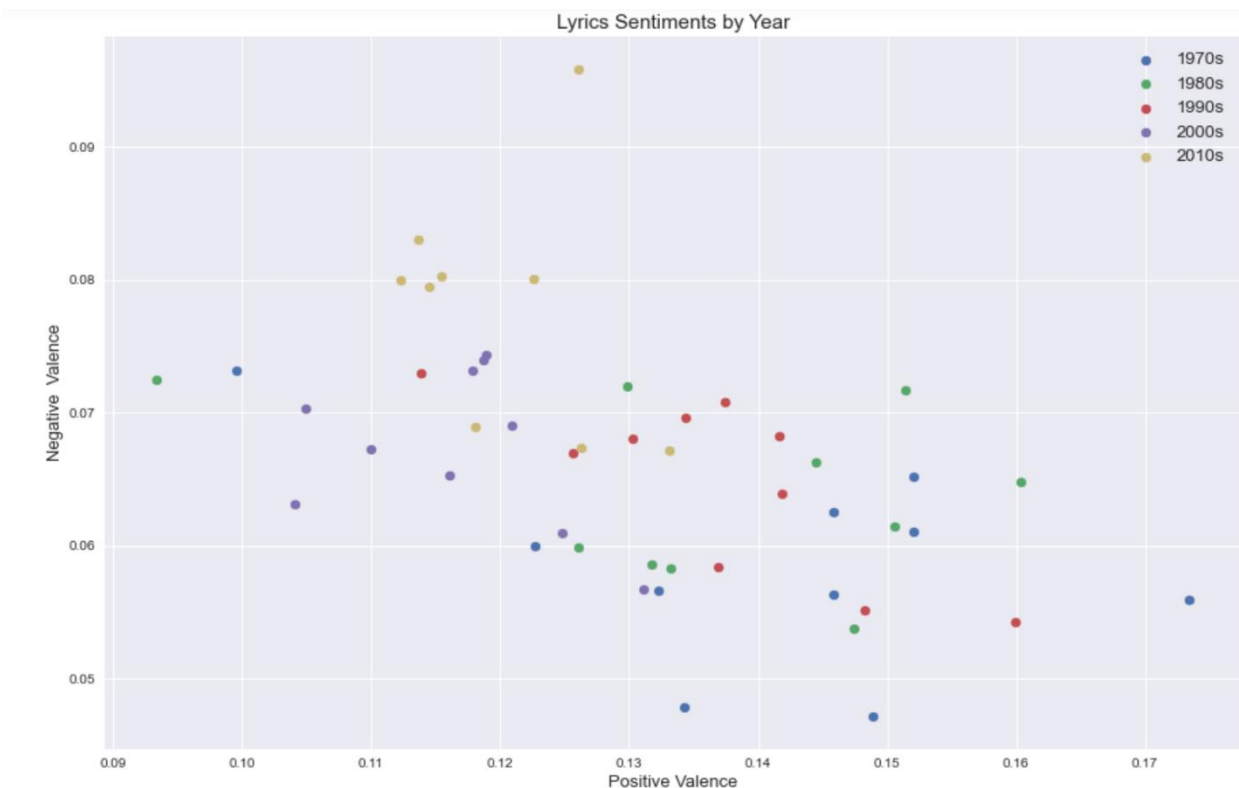
### Sentiment Analysis:

Delving further into the lyrics, I performed sentiment analysis to see what the most listened to songs could reveal about the mood of the American population over the course of the past fifty years, and how this mood may have shifted across the decades.

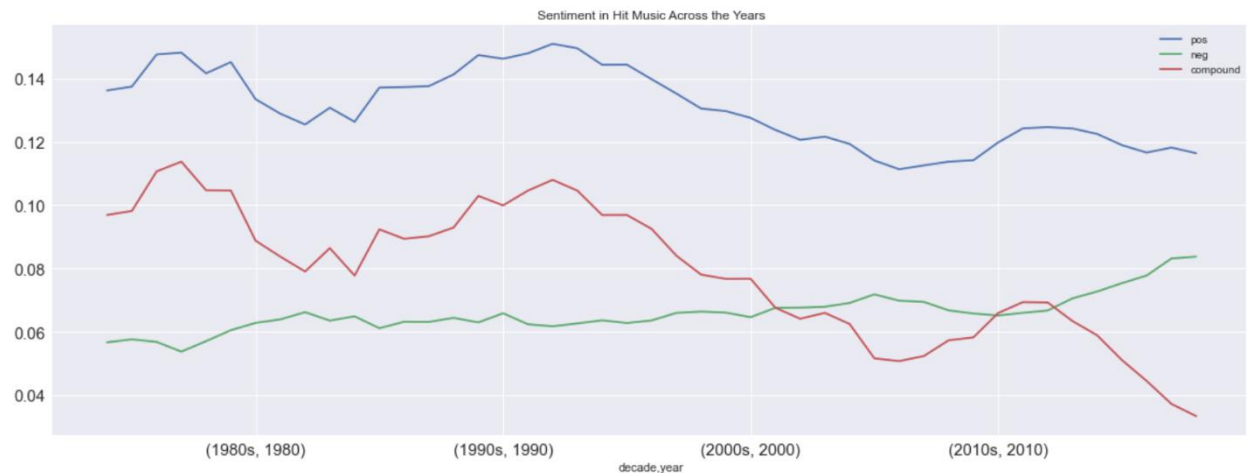




The above charts show that while joy has been the predominant emotion expressed in popular music over the past 50 years, it has seen an overall decline in each decade since the 1970's. Meanwhile, more negative emotional terms, specifically those of anger and fear, have seen steady increases in popular music, so much so that anger and fear were the second and third most dominant emotions in music of the past decade, respectively. Negative terms were so dominant in music over the past several years, that the overall polarity of the most popular music indicates a general negative emotional mood in the American population. This conclusion is further illustrated in the below scatterplot.



As can be seen in the above chart, the top six most negative “emotional years” in popular music have all been from the past decade, while the seven through nine spots are occupied by years from the decade before that. These same data points are all also on the lower half of the positive sentiment axis, contributing to the decrease in the overall polarity of popular music since the turn of the millennium. The below graph further illustrates that since the mid to late-nineties (when the rise in gangsta rap, grunge and alternative rock began), there has been a steady decrease in positive sentiment in popular music, met with a corresponding steady increase in negative sentiment.

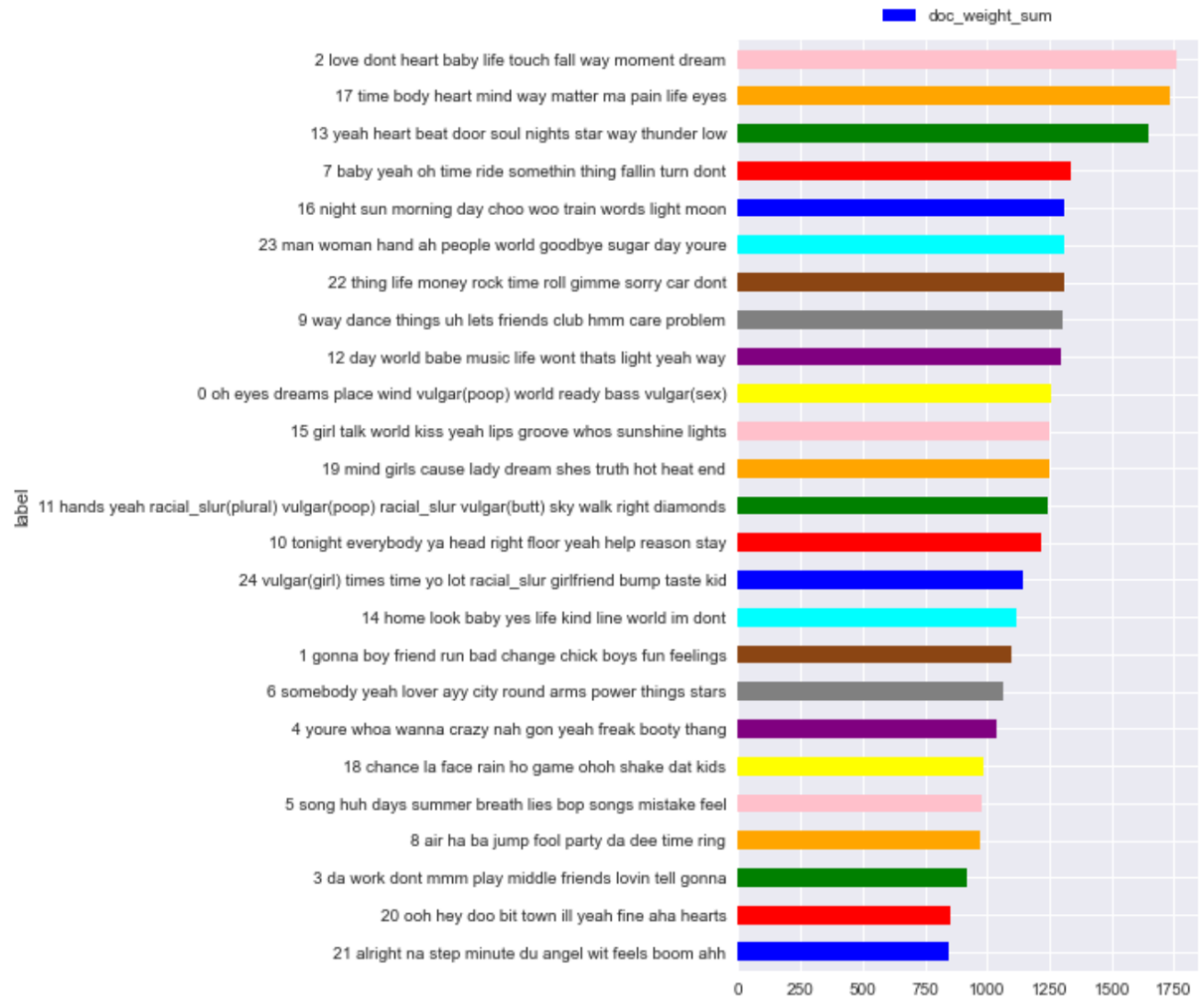


### **Conclusion:**

If Law and Ho’s theory that a society’s popular music reflects the ideals that are prevalent at the time is to be believed, then, given the results of this analysis, it’s fair to say that American society has seen a gradual decline in positivity over the past fifty years, while negative sentiments, specifically those of anger and fear, have seen gradual increases, especially in the past decade. While I cannot confirm that these results one hundred percent accurately reflect the overall mood in America during the studied period, as music and its lyrics are not the only indicators of cultural trends, I do believe they provide some insight into the overall cultural attitude. Specifically, events of the past several years (and especially the one in DC at the start of this year), indicate a growing fear and anger in America, a trend accurately captured by these results. I only hope that we can see the compound valence/sentiment of the red line in the above graph start increasing again.

## Appendix of Additional Visuals:

Overall Topics in the Data





20: ooh hey doo bit town ill yeah fine aha hearts  
 0: oh eyes dreams place wind vulgar(poop) world ready bass vulgar(sex)  
 21: alright na step minute du angel wit feels boom ahh  
 18: chance la face rain ho game ohoh shake dat kids  
 9: way dance things uh lets friends club hmm care problem  
 5: song huh days summer breath lies bop songs mistake feel  
 2: love dont heart baby life touch fall way moment dream  
 16: night sun morning day choo woo train words light moon  
 22: thing life money rock time roll gimme sorry car dont  
 12: day world babe music life wont thats light yeah way  
 23: man woman hand ah people world goodbye sugar day youre  
 15: girl talk world kiss yeah lips groove whos sunshine lights  
 19: mind girls cause lady dream shes truth hot heat end  
 4: youre whoa wanna crazy nah gon yeah freak booty thang  
 10: tonight everybody ya head right floor yeah help reason stay  
 14: home look baby yes life kind line world im dont  
 7: baby yeah oh time ride somethin thing fallin turn dont  
 8: air ha ba jump fool party da dee time ring  
 3: da work dont mmm play middle friends lovin tell gonna  
 1: gonna boy friend run bad change chick boys fun feelings  
 24: vulgar(girl) times time yo lot racial\_slur girlfriend bump taste kid  
 17: time body heart mind way matter ma pain life eyes  
 13: yeah heart beat door soul nights star way thunder low  
 11: hands yeah racial\_slur(plural) vulgar(poop) racial\_slur vulgar(butt) sky walk right diamonds  
 6: somebody yeah lover ayy city round arms power things stars

(Notice Decrease in tender terms such as “love” and “baby”)

