

W200 | Project 2 | Summer 2018
The 2016 Presidential Candidates, As Viewed by Hip Hop
Zeliang Xu | Katie Mo | John Boudreaux

Introduction

The 2016 US presidential election proved to be an interesting event on the world stage, and has propagated ripple effects through many parts of society both domestically and globally. Public opinion about the election and the results has been divided. While the news presents a very distinct view of current events such as these, a different perspective can be seen in the artistic and musical world. In particular, the hip hop genre tends to have a way of not holding back any type of opinion, and will “tell it like it is.” For the 2016 election, there were some prominent examples of the sentiments in the rap community such as “F*** Donald Trump” by YG feat G-Eazy & Macklemore which made news headlines and were representative of particular audiences’ views of the candidates. Talking about politics and candidates in hip hop is nothing new, however, and some of the candidates have been talked about for years before their presidential bids. This led to our research group to ask the question:

How were the candidates in the 2016 US presidential election portrayed by hip hop artists?

The nature of this question is very broad, and will be addressed in several sub-topic areas throughout the rest of this report. We will first start with a small introduction to the datasets used to conduct this analysis with some exploratory data analysis, and then go into examining different facets of the data. The first is to look at the mentions of the candidates over time leading up to the election, followed by a more detailed look into individual candidates and how positively they were portrayed. The final portion of the analysis shifts perspective into looking at how the artists themselves talked about the candidates. We end our report with our conclusions about what we can learn about the candidates and rap music from these analyses.

The data

We used four datasets from different locations for our analysis: FiveThirtyEight Hip Hop Candidate Lyrics, FiveThirtyEight Hip Hop Candidate Lyrics Supplemental Data, RealClearPolitics Polling Numbers, and Google Trends Data. In this section, we describe each of these datasets in more detail, including our exploration and validation of the data.

FiveThirtyEight Hip Hop Candidate Lyrics

Our primary dataset for this analysis came from FiveThirtyEight, which was created by scraping song lyrics from genius.com for references to presidential candidates in the 2016

presidential election. It has 377 entries (missing entry id 97), and 8 columns (with an additional ID column) that contain the following information:

- **candidate**: the candidate referenced
- **song**: song name
- **artist**: the artist of the song
- **theme**: topic area that is referenced in the lyric, manually labeled by fivethirtyeight (many are left blank)
- **sentiment**: positive, negative, or neutral (based on manual labeling by fivethirtyeight)
- **album_release_date**: year of album release
- **line**: the specific lyric mentioning the candidate
- **url**: link to full song lyrics on genius.com

This data was checked for obvious errors, but is relatively clean. In particular, in the hip-hop genre it is common to feature collaborations between multiple artists for songs. Our dataset, with very few exceptions, only includes the primary artist. This allows for us to treat the artist as a categorical variable and not have to use regex to figure out where they might have been featured in a different track. Additionally, even though the sentiment classification was done manually, it is the opinion of the writers that this is decently accurate. Some of the neutrals seem to skew positive when taken in greater context. Random examples of each sentiment can be seen below.

sentiment	line
positive	Stephen Curry Basedgod, Donald Trump Basedgod
positive	In hot pursuit of Donald Trump rap loot
positive	It's like Trump in your crib countin' money wi...

sentiment	line
neutral	Look at Donald Trump!
neutral	Senseless, Apprentice I thought Da Don knew Tr...
neutral	I'm at the poker table like I'm Donald Trump

sentiment	line
negative	I fantasize shooting Trump down, A shot for ev...
negative	But Hillary, I promise to wipe my server clea...
negative	I want to buy Donald Trump and sell him to EI ...

It is notable that Hillary Clinton and Donald Trump make up a vast majority of the references, as shown in the table of unique artists and songs per candidate:

	song	artist
candidate		
Ben Carson	1	1
Bernie Sanders	2	2
Chris Christie	2	2
Donald Trump	266	192
Hillary Clinton	92	79
Jeb Bush	9	9
Mike Huckabee	2	2
Ted Cruz	1	1

Given that Hillary Clinton and Donald Trump have the most references, they also have the most entries with attributed themes. There are 79 entries that have no theme attributed. The rest are broken down by the following table:

candidate	theme	song
Bernie Sanders	political	2
Donald Trump	The Apprentice	17
	hotel	68
	money	102
	personal	15
	political	24
	power	2
	sexual	1
Hillary Clinton	money	5
	personal	29
	political	21
	sexual	3
Jeb Bush	personal	3
	political	4
Ted Cruz	political	1

FiveThirtyEight Hip Hop Candidate Lyrics Supplemental Data

Because an “url” column was included in the FiveThirtyEight data, the BeautifulSoup package was used to scrape more information from the genius.com web pages. Two additional columns were populated from parsing the web page HTML from the links provided:

- **lyrics_views**: the number of page views for the lyrics on genius.com
- **track_release_date**: the release date of the track formatted as month, day of month, and year

To check the quality of the data scraped from the web pages, the describe function was used to gather a summary of the columns.

For the “lyrics_views” column, the count gathered all 377 points, indicating that there were no missing values. The values ranged from 92 to 1,307,782 views as of August 5, 2018. Note that if the scraping function was recalled, the “lyrics_views” values may increase due to more people visiting these web pages since the last time the data was collected.

For the “track_release_date” column, the count only gathered 248 points, indicating that there were 129 missing values, which consisted about a third of the data. Furthermore, upon closer inspection, there were some discrepancies in the “track_release_date” and the “album_release_date”, the column curated by the FiveThirtyEight staff. For example, navigating to the url for the song None Shall Pass by Aesop Rock (<http://genius.com/Aesop-rock-none-shall-pass-lyrics>) showed that the the “album_release_date” of 2011 was nowhere to be seen.

Due to these discrepancies, two new columns were created. The first column “years_match” indicates whether the “album_release_date” and the year for “track_release_date” matched for rows without missing “track_release_date” data. The second column “corrected_release_year” corrects the discrepancies by using the year from “track_release_date”, and when there were no data for this column, “album_release_date” was used. The following table shows the 18 rows that were corrected with this method:

	song	artist	album_release_date	track_release_date	years_match	corrected_release_year
0	None Shall Pass	Aesop Rock	2011	2007-08-28	False	2007
22	We Can Get It On	Yo Gotti	2012	2010-12-11	False	2010
23	Where They Do That At	Yo Gotti	2015	2010-07-28	False	2010
32	Again (Stankonia Remix)	Big Boi	2001	2000-09-14	False	2000
52	Every Girl	Gudda Gudda	2010	2009-04-06	False	2009
53	Bad Bad Bad	Gucci Mane	2010	2009-12-08	False	2009
61	Bitch, Don't Kill My Vibe (Remix)	Jay Z	2012	2013-03-19	False	2013
82	Krazy	Pitbull	2009	2008-09-30	False	2008
95	Back Together	Nicki Minaj	2015	2016-08-06	False	2016
128	Ba Bump	Black Eyed Peas	2006	2005-05-25	False	2005
175	Playas Only	The Game	2006	2005-06-19	False	2005
186	Song for the Ville	J.Cole	2010	2011-04-09	False	2011
205	Trump	Jeezy	2011	2012-02-07	False	2012
223	On the Real	Nas	2011	2004-03-30	False	2004
251	Giv No Fuks	Migos	2014	2015-12-04	False	2015
288	10 2 10 (Remix)	Rick Ross	2014	2013-12-31	False	2013
350	CIT4DT	Dooley	2016	2015-03-27	False	2015
359	Look Me in My Eyes	Vado	2014	2013-12-30	False	2013

RealClearPolitics Polling Numbers

We used data for both republican and democratic primary campaigns from RealClearPolitics (realclearpolitics.com) in order to calculate how long each candidate was in the race for nomination. Both the data tables for republicans and democrats followed the same format, which contained the following columns:

- **Poll:** the poller
- **Date:** the dates the poll was conducted
- **Sample:** the sample size, with what type of sample (A: All Adults, LV: Likely Voters, RV: Registered Voters)
- **Candidate Specific Numbers:** the percentage for each individual candidate in that poll
 - Republicans had columns for **Cruz, Kasich, Trump**
 - Democrats had columns for **Clinton, Sanders**
- **Spread:** the candidate who is winning in this poll, and by how much

The dates from this data source required heavy preprocessing. They were originally given in the format of “5/13 - 6/15” without years. Given additional documentation from the source, we were able to find out that the first values in the table were from 2016 and the rest worked backwards. The dates were processed to only include the final date in the date range, and to include years (work can be seen using a combination of regex and looping in the repository in Appendix 1, in analysis/PreProcessingPolls.ipynb).

The candidate specific numbers originally had “--” values for when polls were conducted before/after a candidate was in the race. In pre-processing, these were changed to numpy NaN values. This was then queried to find the dates for when the candidate was in the race, and a length was generated by subtracting the minimum value from the maximum value as a timestamp. Although our analysis did not look into the difference between poll types, the spread, or the numbers themselves, we did not find any discrepancies in any of those fields that would affect our analysis.

Google Trends Data

We also incorporated the data on google Trends by using the candidate names as search keyword (<https://trends.google.com/trends/?geo=US>). The Google based online searches help us understand better how hiphop mentions are really indicating the public opinions of political candidates. Our expectation is to find certain correlation between the hiphop data and the Google search data, locate those spikes in mentions/searches that coincide, and then find what events gave the boost to internet searches and lyrics mentions.

	month	searches
0	2004-01-01	3.0
1	2004-02-01	2.0
2	2004-03-01	2.0
3	2004-04-01	4.0
4	2004-05-01	1.0

For each candidate, a CSV file is downloaded with two simple columns:

- **Date:** The date ranges from 2004-01-01 (the earliest available date for Google trend) to 2018-01-01
- **No. of searches:** the data here is not the real number of Google searches, but scaled data with the historical peak being 100.

The data in Google trend is very simple and clean, with only the column of time and number of Google searches. In exploring and validating the dataset, we did the following steps:

1. In some rows of data, when Google searches are smaller than 1% of the peak value, it is marked as "<1", which is insignificant for this project and creates complexity in data plotting. Therefore, we replaced all "<1" values with 0
2. Plotting the time history of Google searches of each candidate, checking for outliers and missing portion of the data

In data exploration and analysis, we plotted for each candidate, the data from Google Trend together with hip hop lyrics mentions. The purposes for doing so are:

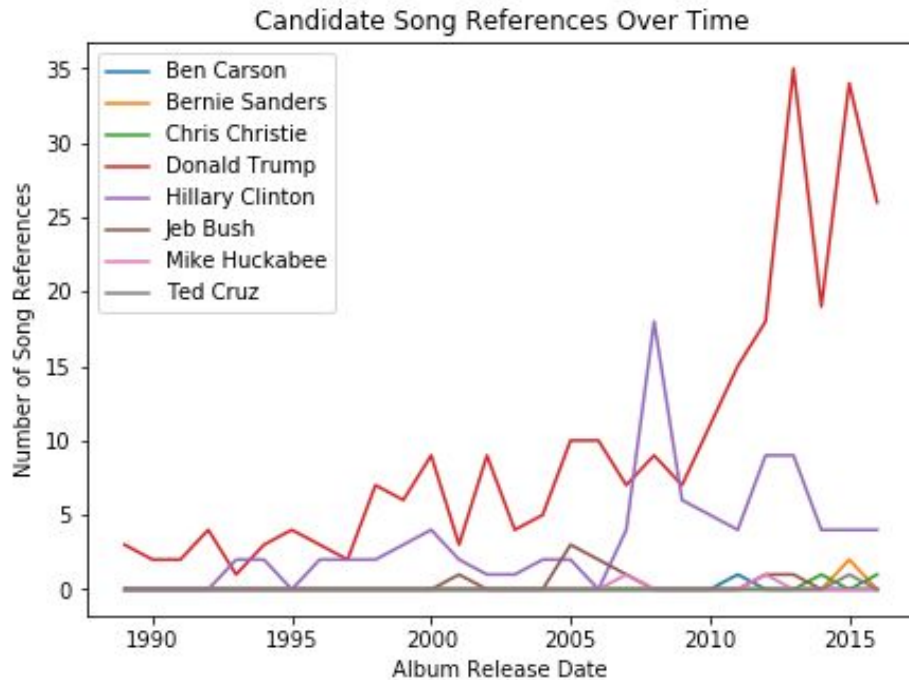
1. We are investigating the popularity of political candidates reflected in hip hop lyrics mentions. And we believe that internet searches usually reflect well the public attention (either good or bad) on the candidate. Together with hip hop lyrics mentions, we can identify some events that affect the popularity level of candidates
2. If we see matches of spikes in time history of both data set, we can pull out the hip hop songs written during that time and extract the lines where singers mentioned the candidates. This can help us understand better the sentiment behind the lyrics

Hip hop references and current events

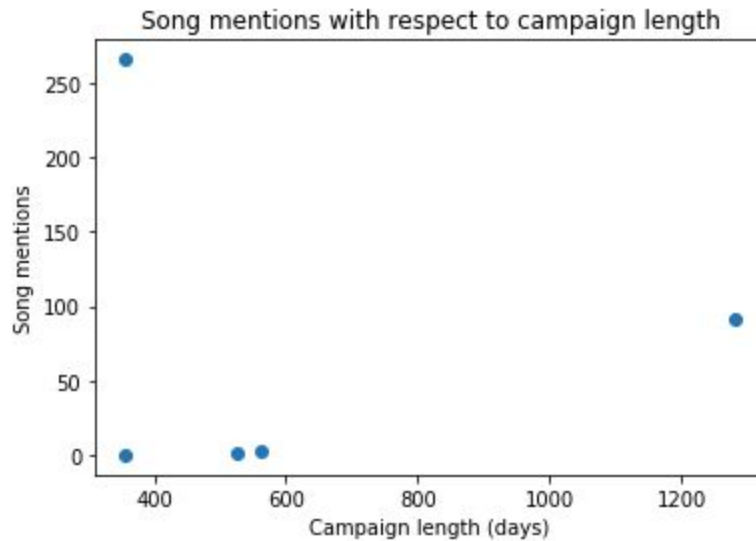
We wanted to explore how hip hop artists incorporate political current events into their work. We suspected that the most significant events would give rise to a larger number of references in the hip hop song lyrics. In this section, we investigate when political figures were referenced in song lyrics over time and what events occur that may provoke these references.

What events lead to song lyric mentions?

As we would expect for the two final 2016 presidential candidates, the graph below shows that Donald Trump and Hillary Clinton were the most referenced candidates in hip hop songs released between 1989 to 2016.



Before diving into the analysis of individual candidates, it is worth mentioning that for the 5 candidates we had polling data for, there appears to be no sort of relationship between campaign length and song mentions.



This suggests that artists may not be concerned so much about political issues, but more so about current events themselves. Hillary Clinton and Donald Trump top the song mentions, and for reasons that include their political careers but also many other factors having to do with them being in the public eye for a while.

Donald Trump

Starting in 1989, Donald Trump had a few hip hop songs which reference him by name. This could be due to his first appearance on the cover of Time magazine. References to Trump continued throughout the next decade. From 1996 to 2009, Donald Trump's fame increased with his purchase of the Miss Universe organization, the premiere of his television show "The Apprentice", and his receipt of a star on the Hollywood Walk of Fame. His increased fame may have led to more frequent references in hip hop lyrics. By 2012, Donald Trump became more involved in the political environment as part of the "birther" movement, which sought to cast doubt on Barack Obama's legitimacy as president, and also had publicly considered running for president. References to Trump within hip hop lyrics were significantly higher from 2012 onwards when compared to the previous two decades. In 2015, Donald Trump announced his bid to be the President of the United States as a very polarizing figure.

To better understand what type of words are used most frequently in the referenced line of hip hop lyrics, the 8 most frequent notable words from the column "line" in the FiveThirtyEight data set is shown below. Notable words did not include common pronouns, prepositions, conjunctions, and non-descriptive verbs such as "the", "like", "in", "and", "you", or "got", among others.

8 Most Frequent Notable Words in Referenced Line of Lyrics

rank	word	count	frequency
1	trump	274	0.1023
2	donald	143	0.0534
3	tower	22	0.0082
4	money	18	0.0067
5	f***	14	0.0052
6	towers	13	0.0049
7	fire	12	0.0045
8	call	12	0.0045

Donald Trump was more commonly referenced by his last name. He was associated with the word "tower", which refers to the Trump Tower in New York City opened in 1983. The plural "towers" was also frequently used, which could refer to his line of hotels, which are named Trump International Hotel and Tower. "Money" was also commonly used, which is not surprising as Donald Trump frequently makes public boasts about how much money he has. The 7th most frequent notable word, "fire", could be related to his television show "The Apprentice" in which his catchphrase is "You're fired!". The most notable frequent words seem to be more related to his businesses rather than his political career.

Hillary Clinton

In 1993, Hillary Clinton was first referenced in the lyrics of hip hop songs of the data set. In this same year, Hillary Clinton became the First Lady of the United States when her husband, Bill Clinton, was elected president. Over the next eight years as First Lady, Hillary Clinton was occasionally referenced by hip hop artists. In the year 2000, Clinton received a small uptick in mentions, which corresponds to her election as Senator for New York. In 2007 and 2008, references to Hillary Clinton increased, with a significant spike in 2008. This could be explained by her campaign for the Democratic Presidential nomination in 2007 and her subsequent appointment as Secretary of State under President Barack Obama in 2008. References to Hillary Clinton continued throughout her time as a public political figure, with another increase in references in the years 2012 and 2013. At this time, Barack Obama was re-elected, but Clinton chose not to pursue a second term as the Secretary of State. Although she announced her run for President of the United States in 2015, the number of references to Hillary Clinton did not increase as significantly as her previous campaign in 2008. This may be due to Donald Trump's overwhelming number of references, "stealing the spotlight" as a controversial presidential candidate.

The 8 most frequent notable words from the column "line" in the FiveThirtyEight data set was also found for Hillary Clinton, in a similar manner as described above for Donald Trump.

8 Most Frequent Notable Words in Referenced Line of Lyrics

rank	word	count	frequency
1	hillary ⁺	88	0.0818
2	clinton	35	0.0325
3	bill	24	0.0223
4	rodham	12	0.0112
5	b****	8	0.0074
6	obama	7	0.0065
7	f***	6	0.0056
8	she	6	0.0056

⁺The word counts for "hillary" and "hilary" were combined. Because genius.com relies on crowdsourcing to populate lyrics, the misspelled name is an artifact of the process and not necessarily from the hip hop artists themselves.

In contrast to Donald Trump, Hillary Clinton was more commonly referenced by her first name. On occasion, her maiden last name "rodham" was used. She was more frequently associated with other political figures, with "bill" referring to her husband Bill Clinton and "obama" referring to Barack Obama, the President in which she served under as Secretary of State. Other

frequent notable words, “b****” and “she”, refer to her gender, with the former having a strong derogatory connotation.

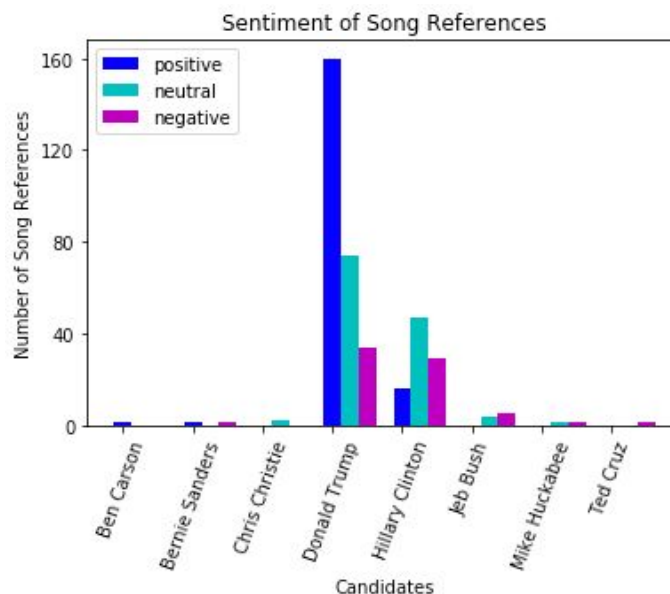
Other candidates

Of the other candidates in the 2016 election cycle, Jeb Bush had a distant third place in number of references. Hip hop artists mentioned Jeb Bush in their lyrics around the year 2005. At this time, Jeb Bush was the Governor of Florida, completing his term in 2006. In addition, his brother, George W. Bush, was President of the United States. As a public political figure, Jeb Bush accumulated some references during this time. Other candidates involved in the 2016 Presidential election were referenced by hip hop artists only once or twice.

The view of the candidates

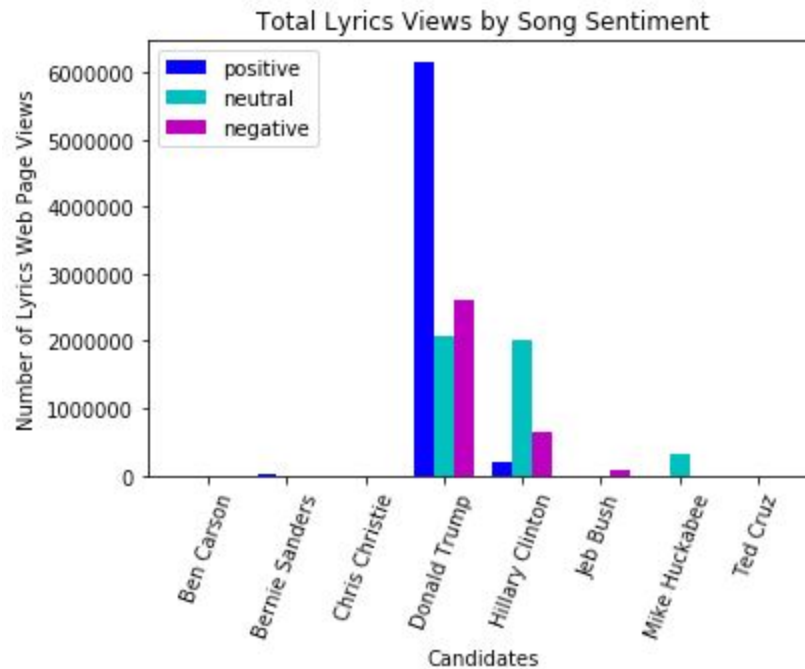
In this section, we dive deeper into the sentiment behind the hip hop songs that reference the 2016 presidential candidates as well as their perceived popularity inferred by song views and google searches.

Are some politicians viewed more positively or negatively than others?

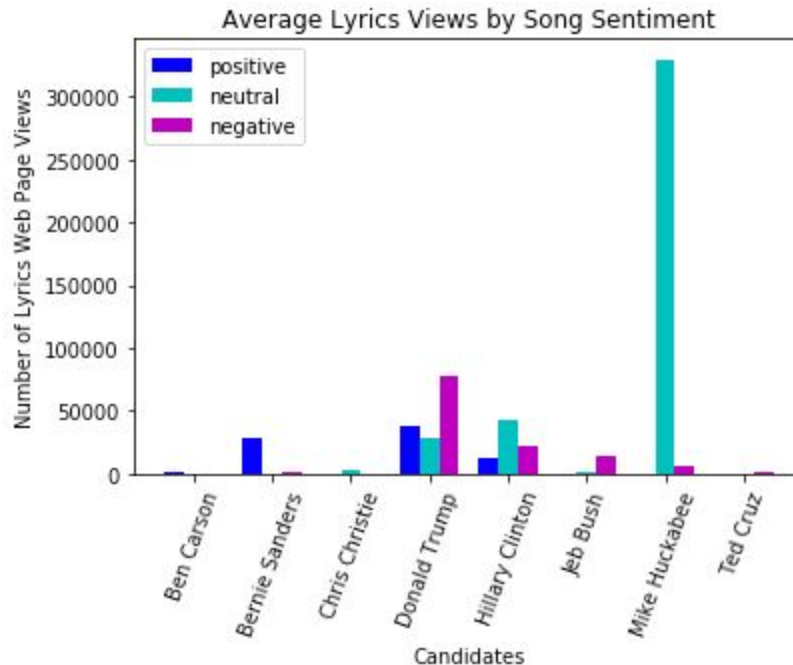


The graph above describes the total number of references made to a political candidate broken down by the sentiment of the reference. Based on the total number of song references, we can see that Donald Trump has been mentioned in a positive manner over the historical span of the data set. Trump's total number of song references dwarfs those of the other candidates, including Hillary Clinton, which suggests that Trump has historically been the most famous

figure of all the political candidates. Unlike Trump, Hillary Clinton was referenced with mixed sentiment by hip hop artists, with most references tending to be neutral. The data was further explored from different perspectives to gain a better understanding of the sentiment in mentioning both of these political figures.



This graph shows the total number of lyrics views associated with each sentiment is displayed for each candidate. This graph corresponds heavily to the previous graph, but the songs that were more popular contributed more weight to the sentiment ratings. Popular songs which reference Donald Trump tended to skew towards either positive or negative sentiment, which is reflected by his lengthy time as a public figure. As shown in the "Donald Trump: lyric mentions" graph, Donald Trump references were largely positive for the time period for the first 25 years of lyrics captured in this data set. In contrast, the last 3 to 5 years of the data set contain an increase number of negative Donald Trump song lyric references, which could be due to his controversial campaign for the United States presidency.



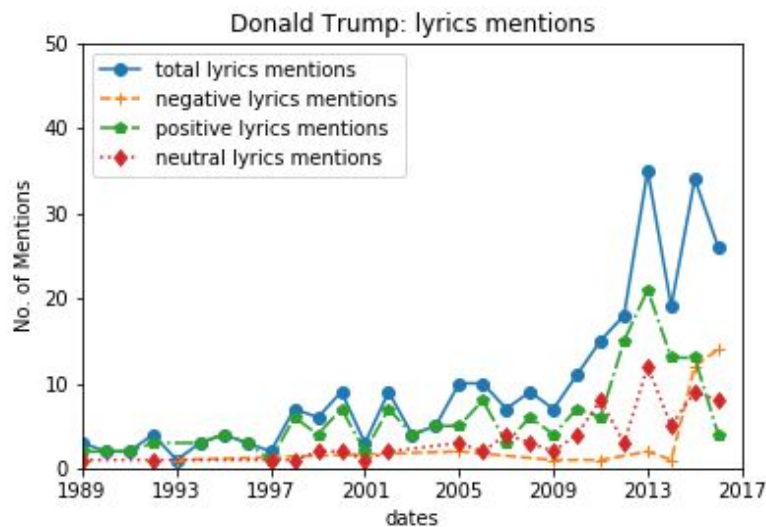
In this third graph, we balance the weight of songs with high number of views by looking at the average number of lyrics views by sentiment for each candidate. On this graph, the first data point that stands out is Mike Huckabee's high average lyrics views for the neutral sentiment. Due to Huckabee's low number of song references, the average number of lyrics views here reflects only one song, "None Shall Pass" by Aesop Rock, which had 329,759 views.

When comparing Donald Trump's average song views by sentiment to the previous graphs, we notice that the distribution of sentiment is very different. Negative sentiment has a higher number of average song views, indicating that songs which reference Donald Trump negatively tended to be more popular. Although more songs with positive sentiment for Donald Trump have been created historically, the popularity of these songs is much lower.

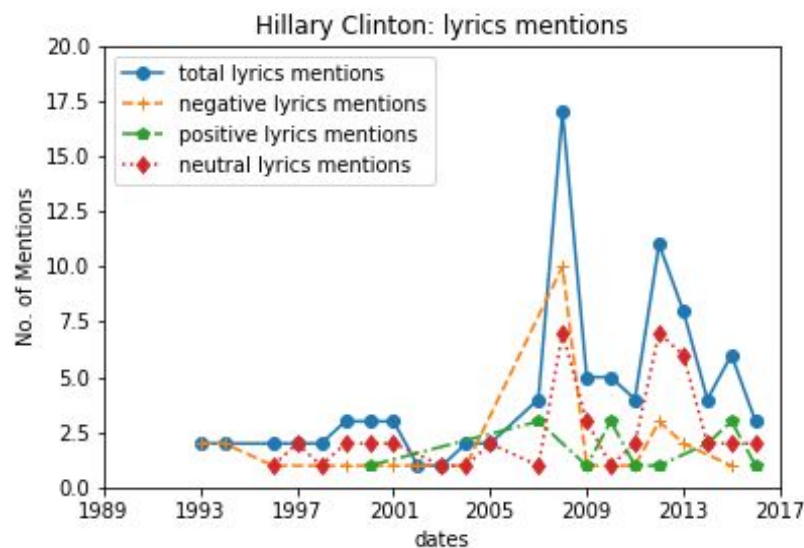
On the other hand, Hillary Clinton's average song views have a similar distribution of sentiment when compared to previous graphs. Neutral sentiment songs referencing Hillary Clinton are more popular.

How do the opinion of politicians change over time?

In this section, we split up the lyrics mentions based on the sentiment expressed through the lyrics, we are able to find more interesting results. We are only mentioning two candidates here (Donald Trump and Hillary Clinton) since those two have enough mentions for us to see a complete variation over time, as shown below:



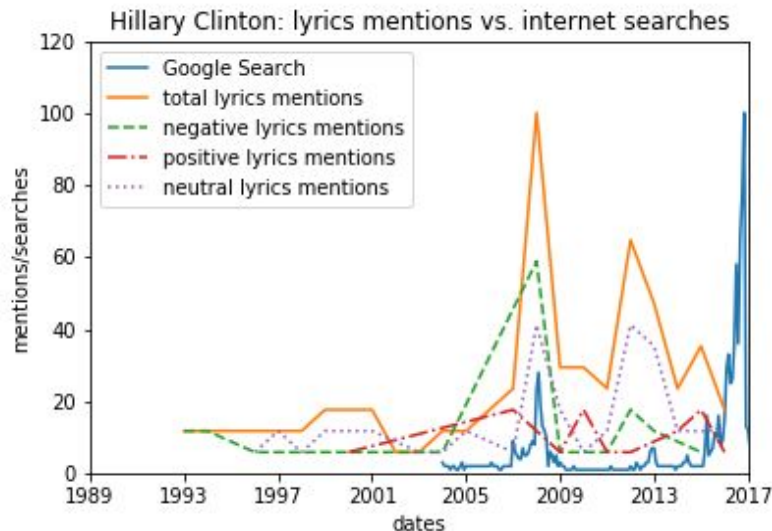
Take Donald Trump first, he not only holds the No.1 position in the number of lyrics mentions, but also enjoyed an overall "healthy" growth of over time. When we add in sentiments in the equation, however, the trend does not look "healthy" anymore. For lyrics mentions with "negative" sentiment, Donald Trump has a good record of keeping it low for decades until after 2014 when it sees a historical jump. It is still rapidly growing towards the end of our dataframe time limit. At the same time, his positive mentions in lyrics went exactly the other way, reaching a very low point and still dropping towards the end of time limit of our data.



Now let's take a look at Hillary Clinton. Her lyrics mentions have more spikes through history timelines, with the biggest spike happening around 2008, the second between 2012 and 2013. In the first spike, both negative mentions and neutral mentions contribute to the total number of mentions, while in the second spike, neutral mentions dominate.

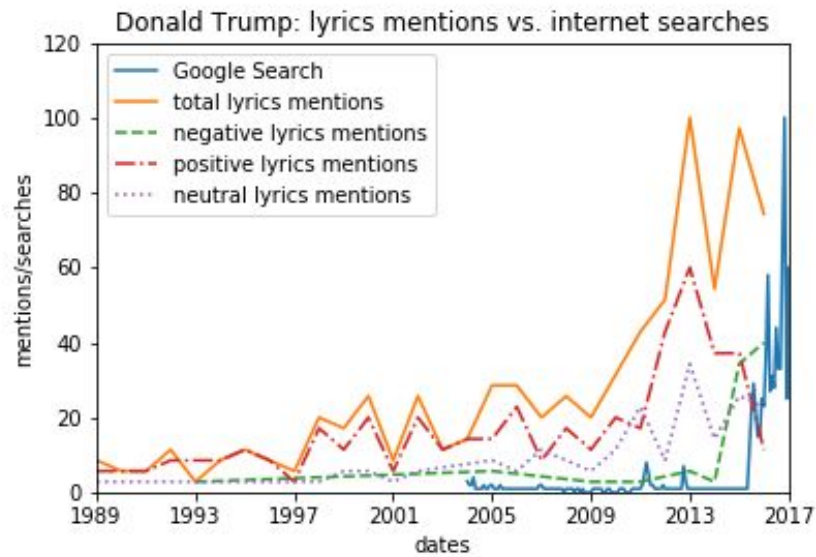
To help us filter out some coincidental spikes and focus on some big spikes in lyrics mentions, we bring in another data set from Google Trends, which shows the historical changes of the

number of internet searches of certain keyword. We typed in the name of each candidate and plotted those data against their lyrics mentions.



Taking Hillary as an example. The plot above shows that the Google searches peaked once around 2008, which matches her first peak in lyrics mentions as well. Even without the history book, we can take a wild guess that it was her election year in 2008 against Obama within their party. With this assumption in mind, we went back to the lyrics mentions dataset, and filtered out the songs that mentioned her during 2008. It turned out that a number of songs are featuring the singers' dissatisfaction with the "traditional" politicians (like Hillary) and their support for a Obama to be the president. One example is the song "politics as usual" by singer Ludacris, who openly supported Obama during his 2008 election. Several other songs featured the same sentiment around the same time, resulting in this huge spike of negative mentions of Hillary Clinton due to her election campaign against Obama.





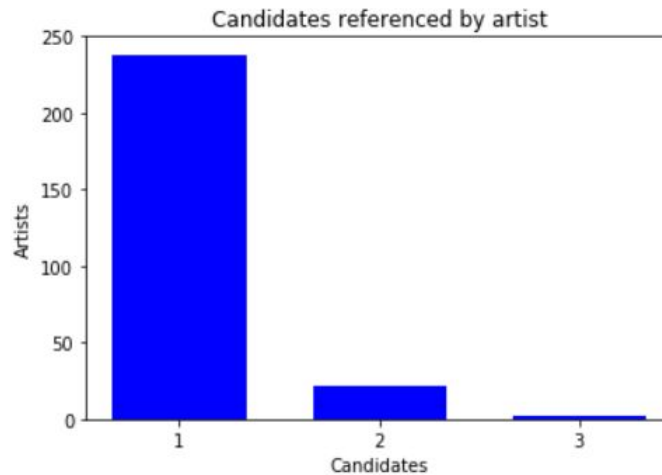
As for Donald Trump. Compared to his very active history of lyrics mentions, his Google searches are relatively quiet, but grew abruptly before 2016. This jump is following well with the path of his negative lyrics mentions around the same time. It apparently is associated with his election during 2016 and a series of campaign over a long time range.

For other candidates, we also found that Jeb Bush had a coincidence of lyrics mention peak and Google search peak around 2005, which was during his term as governor of Florida. However, we found that the content in lyrics mentions are more related to Bush family in general, rather than talking about a particular event that happened in the year.

The artists' view

An additional view of our data that could be interesting might be to look at it from the perspective of the artists, and broader trends of how politics are referenced in hip hop. Although our data only presents a small subset of the overall work that each of the included artists have produced, we can still glean some insights. In this section we present multiple questions that might be answered by adopting the view of the artists.

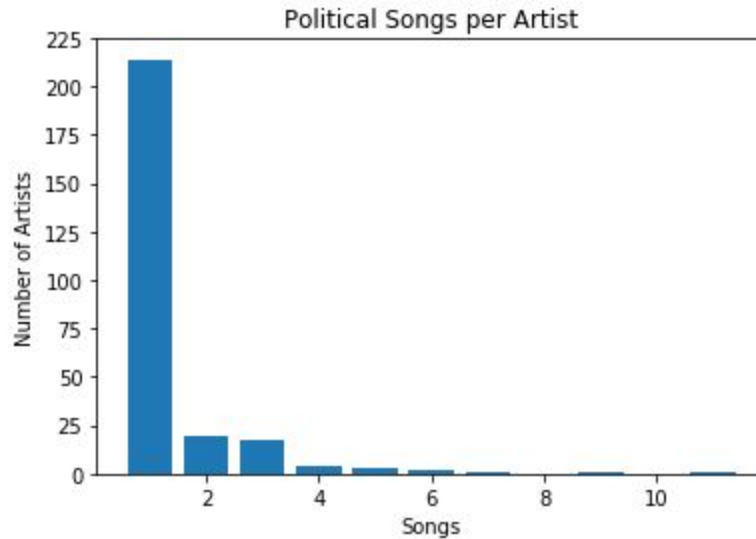
Do rappers talk about one candidate? Or multiple?



Very simply, rap artists tend to only talk about one political candidate. Even though some artists may talk about multiple candidates, they may talk about them in the same song. Four out of our 23 artists are on the list simply by referencing multiple candidates in a single song.

candidates_referenced	song	artist
2	Datpiff (freeverse series Ep. 4)	Chris Webby
2	I Shoulda Tried Harder	Wax
2	Money Man	Gorilla Zoe
2	PNT	Apollo Brown & Ras Kass

It seems that artists tend to not focus on a variety of candidates, which might mean that they aren't concerned about painting a broad picture of the political landscape. Additional evidence to this hypothesis is given when looking at the number of songs mentioning the political candidates. Most artists only mention one candidate, with a small minority mentioning more than one. Rick Ross and Nas both stand out as exceptions, with 11 and 9 candidate mentions respectively. It is interesting to note that 9 of the 11 mentions of Rick Ross are explicitly Donald Trump Mentions, with 8 of those 9 being about the Trump Tower or Trump-like money.



Overall, it is probably fair to say that with the heavy skew of our dataset towards mentions of Donald Trump for his money or real estate ventures, hip hop artists don't seem terribly interested in painting a comprehensive view of the US political landscape. A given rapper can put out hundreds of songs during their career, which brings perspective to how low these candidate mentions are. There are instances of direct political commentary though, even with rappers like Rick Ross, dropping lines in his song "Free Enterprise" about his discontent with the Trump candidacy and racial divides in the country: "Assassinate Trump like I'm Zimmerman."

Conclusions

- Increases and peaks in hip hop references correspond to current events affecting the candidate, whether the events are political or not.
- Hip hop singers are brief but expressive about their political views, especially during major political events like elections, which created jumps in lyrics mentions of key political candidates. In this sense, hip hop lyrics mentions can reflect in some degree how much attention the public is paying to certain candidate.
- In most cases, the politicians are mentioned in lyrics when they become election candidates. In the case of Hillary Clinton, not matter how much spotlight she got for the last decades, lyrics mentions of her two out of three times due to her presidential election in 2008 and 2016.
- Hip hop lyrics' view about political candidates are usually not changing overtime. As the positive mentions of candidates are only having small fluctuations overtime, but negative ones can have take sudden jumps due to certain political events.
- However, there are cases like Donald Trump, who enjoyed a lot of positive lyrics mentions during his previous job as a businessman. However, ever since he becomes a serious candidate in the election process, his positive mentions took a fall and negative mentions soared.

- Most rappers do not talk about multiple candidates or have multiple songs mentioning political candidates. This suggests that the primary goal of rappers is not political discourse, even though it still happens.

Appendices

Appendix 1: The git repository

The corresponding code for the analyses conducted in this project, along with the raw data used, can be found in the following Github repository:

<https://github.com/boudrejp/PythonProj2>

There is additional documentation in the README files throughout the repository, including the sources of the data themselves.