

The Language of Hip-Hop

Introduction



The word ‘hip’ was a part of the African American Vernacular English language since 1904. It meant ‘current’. The South Bronx community leader Afrika Bambaataa used the term ‘Hip-Hop’ for the first time. In the 1970s, the hip-hop music came up. Hip-hop is more than a music genre, it is a culture that has shaped America over the past three decades. Its influence has spread across the United States, impacting politics and media culture. So today, my project is focused on Hip-Hop music.

First Look

```
```{r}
hiphop <- dt_lyrics[dt_lyrics$genre=="Hip-Hop",]

clean the data and make a corpus
corpus <- VCorpus(VectorSource(hiphop$stemmedwords))

tdm <- TermDocumentMatrix(corpus)
m <- as.matrix(tdm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)

wordcloud(d$word,d$freq,random.order = FALSE, rot.per = 0.3, scale=c(4,.5),colors=brewer.pal(8,
```

```

A good starting point when exploring text data is looking at the word frequencies of the whole text. I will use word cloud to visualize this text. When I removed all the most common words, the below is the output for the entire hip-hop lyrics corpus:



The top 16 most commonly used words are

"love", "shit", "time", "girl", "baby", "niggas", "bitch", "you're", "ill", "chorus", "y'all", "life", "money", "gotta", "hit" and "ass". At the first glance, which is surprising me that "LOVE" is the most common words in hip-hop lyrics, appears 83 times for every 10000 words. But at the same time, what doesn't shock me is that some words like "shit", "niggas" and "bitch" appear in Hip-hop's lyrics very frequently. Overall, this is the first look that Hip-hop left for me, and then I will analyze the lyrics of Hip-hop further.

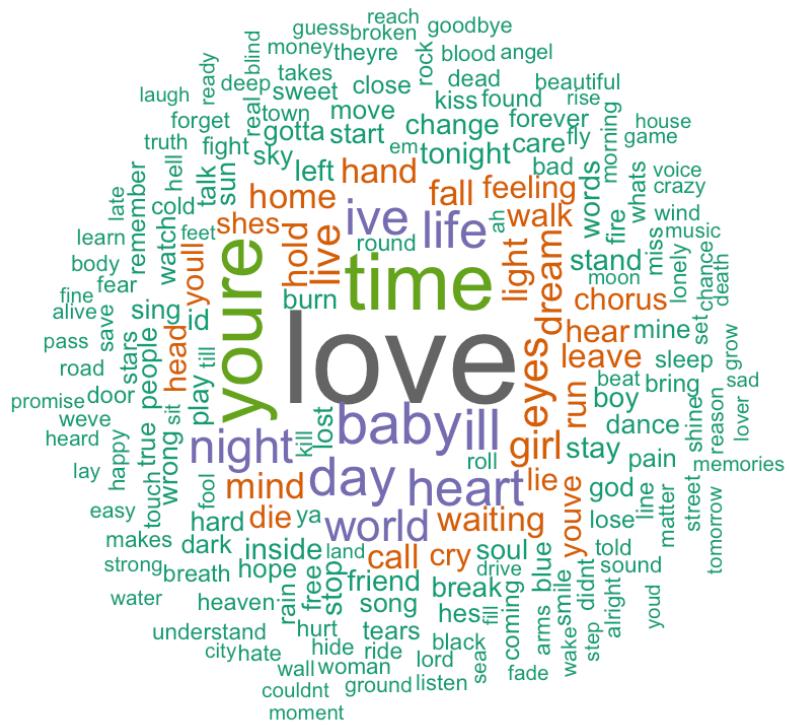
What Words Are “Most Hip-Hop”

```
```{r,warning=FALSE}
others <- dt_lyrics[dt_lyrics$genre!="Hip-Hop",]
corpus_others <- VCorpus(VectorSource(others$stemmedwords))
word_tibble <- tidy(corpus_others) %>%
 select(text) %>%
 mutate(id = row_number()) %>%
 unnest_tokens(word, text)

d_others <- count(filter(word_tibble, id %in% which(dt_lyrics$genre != "Hip-Hop")), word, sort = TRUE)
wordcloud(d_others$word,d_others$n,random.order = FALSE, rot.per = 0.3, scale=c(4,.5),colors=brewer.pal(8,"Dark2"),max
```

```

The below plot is the wordcloud plot for all the words from other genres. I used lyrics from 116799 songs, about 25 million words, except hip hop. From the below plot, we can find that the word “LOVE” is still the most common word, about 230 times for every 10000 words. Actually, the word “LOVE” is not the target I want to analyze, because I want to analyze the unique words which have the high frequency in Hip-Hop but low in other genres.



Then let's compare the frequency of words between Hip-Hop and others.

```

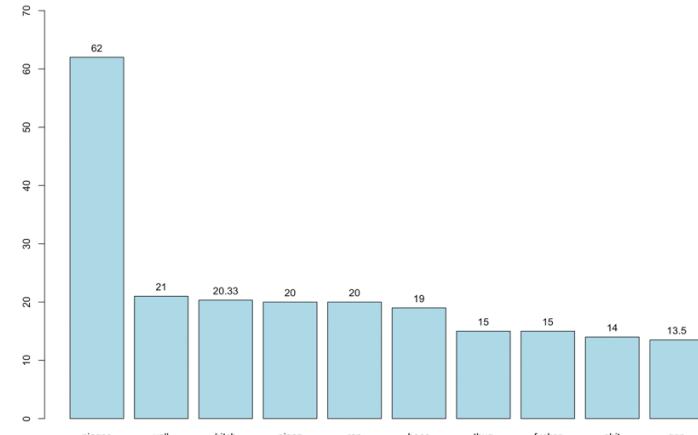
```{r}
d$class <- "HipHop"
d_others$class <- "Others"
d$prop <- round((d$freq/sum(d$freq))*10000,0)
d_others$prop <- round((d_others$n/sum(d_others$n))*10000,0)

word_list <- d[1:200,]$word
c <- rep(0,length(word_list))
for(i in 1:length(word_list)){
 c[i] <- d_others[d_others$word==word_list[i],]$prop
}
new_table <- data.frame(word_list, d[1:length(word_list),]$prop)
new_table$others_prop <- c
colnames(new_table) <- c("Word","Hiphop_prop","Others_prop")
new_table$ratio <- round(new_table$Hiphop_prop/new_table$Others_prop,2)
new_table[46,]$ratio =20
new_table[56,]$ratio =19
new_table[84,]$ratio =15
new_table[115,]$ratio =13
new_table[129,]$ratio =12
new_table[155,]$ratio =10
new_table[194,]$ratio =9
new_table[198,]$ratio =8
sort_table<-new_table[order(new_table$ratio,decreasing = TRUE),]
ggplot(data=new_table,aes(x=HipHop_prop,y=Others_prop,color=ratio, label=Word)) +
 geom_point()+
 geom_text(aes(label=ifelse(Hiphop_prop>30,as.character(Word),'')),hjust=0.5,vjust=1.2)+
 labs(x="Hip-Hop",y="Others",title="Most Hip Hop (every 10000 words)")

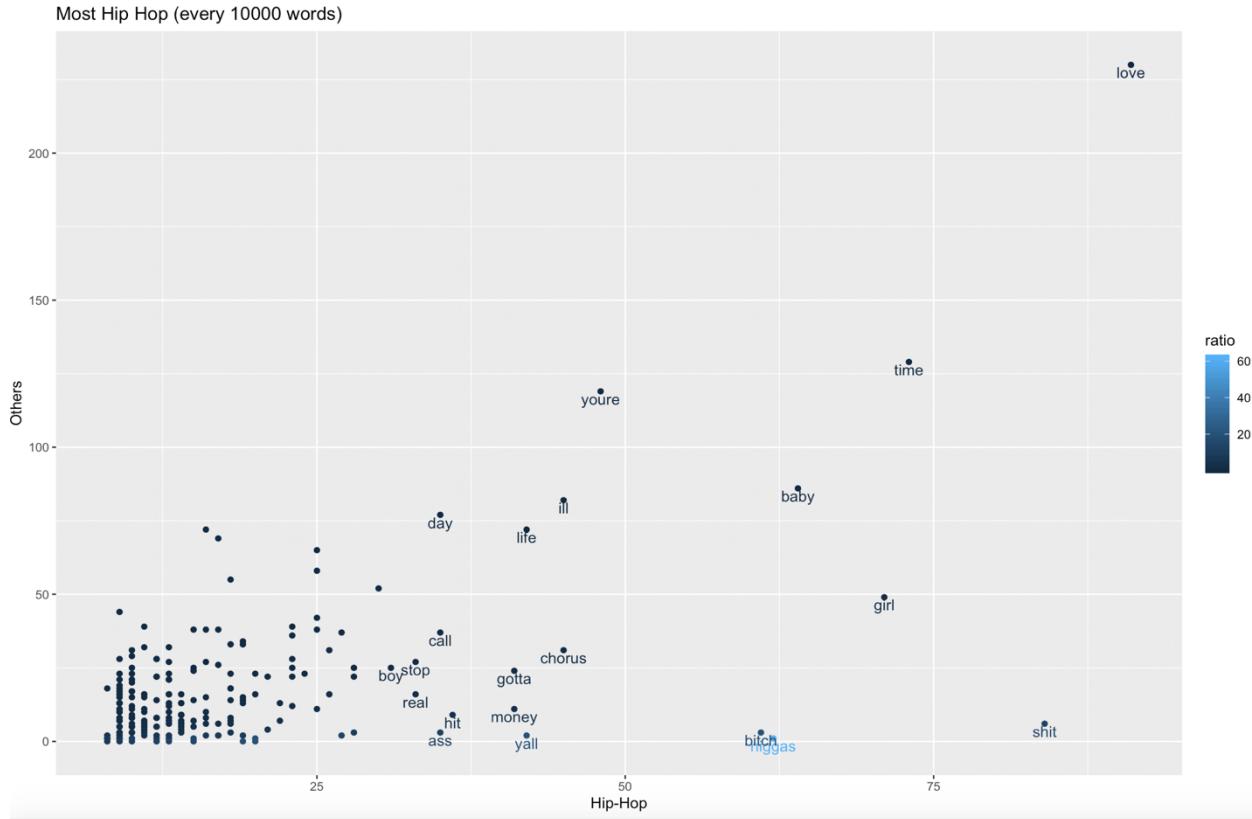
xx<-barplot(sort_table[1:10,4],names.arg=c("niggas","yall","bitch","nigaz","rap","hoes","thug","fucker","shit","gon"),col="lightblue",
text(x=xx,y=sort_table[1:10,4],label=sort_table[1:10,4],pos=3)
```

```

| Word | Hiphop_prop | Others_prop | ratio |
|--------------|-------------|-------------|-------|
| niggas | 62 | 1 | 62.00 |
| yall | 42 | 2 | 21.00 |
| bitch | 61 | 3 | 20.33 |
| niggaz | 20 | 0 | 20.00 |
| rap | 20 | 1 | 20.00 |
| hoes | 19 | 0 | 19.00 |
| thug | 15 | 0 | 15.00 |
| motherfucker | 15 | 1 | 15.00 |
| shit | 84 | 6 | 14.00 |
| gon | 27 | 2 | 13.50 |
| bust | 13 | 1 | 13.00 |
| dem | 13 | 0 | 13.00 |
| dick | 13 | 1 | 13.00 |
| hood | 12 | 1 | 12.00 |
| yuh | 12 | 0 | 12.00 |
| ass | 35 | 3 | 11.67 |
| mic | 10 | 0 | 10.00 |
| crew | 10 | 1 | 10.00 |



Now let's see the above table and plot. Let me explain them. From the left table, I bring the top 18 “Most Hip Hop” list. For example, the word “niggas” appears 62 times for every 10000 words in Hip-Hop, but only appears 1 time for every 10000 words in other genres. Then we can say “niggas” is the “Most Hip Hop” word. From the right barplot, we can also find the likelihood that word appears in hip-hop vs. other genres.



Then let me introduce this dot plot. From this plot, the dots farther to the right are more popular in hip-hop, and farther up-top are more popular in genres. What's more, the words are common in hip-hop but rare in other genres appear more light blue. For example, “niggas” is the “Most Hip-hop” word, so it appears in the lightest blue. On the other hand, the words are common in other genres than hip-hop, appears in darker blue.

Except the word “LOVE”. In terms of the number of occurrences, “shit” is used the most. But in terms of the probability, “N-word” is the first place, the probability of appearing “N-word” is 62 times than in other songs. This is an insulting vocabulary with racism, but today many blacks use this word to greet with each other. When other genres’ singers perform this type of songs, they often use other words to replace it or skip this word.

Consistent with our original imagination, the “Most Hip-hop” words are bad language or slang. And actually, there are more violence and insults to women in hip-hop songs. But these words appear in hip-hop songs, have their unique social and culture background. Netflix’s previous episode “The Get Down” is a good example of the struggles and dreams of slum teenagers in the Bronx, in 1970s.

Different Hip-Hops of Different Ages

```

```{r,warning=FALSE}
hiphop90 <- hiphop[hiphop$year>=1990 & hiphop$year <= 1999,]
hiphop80 <- hiphop[hiphop$year>=1980 & hiphop$year <= 1989,]
hiphop00 <- hiphop[hiphop$year>=2000 & hiphop$year <= 2009,]
hiphop10 <- hiphop[hiphop$year>=2010 & hiphop$year <= 2019,]

#80s
corpus_80s <- VCorpus(VectorSource(hiphop80$stemmedwords))

tdm_80s <- TermDocumentMatrix(corpus_80s)
m_80s <- as.matrix(tdm_80s)
v_80s <- sort(rowSums(m_80s),decreasing=TRUE)
d_80s <- data.frame(word = names(v_80s),freq=v_80s)

wordcloud(d_80s$word,d_80s$freq,random.order = FALSE, rot.per = 0.3, scale=c(4,.5),colors=brewer.pal(8,"Dark2"))

#90s
corpus_90s <- VCorpus(VectorSource(hiphop90$stemmedwords))

tdm_90s <- TermDocumentMatrix(corpus_90s)
m_90s <- as.matrix(tdm_90s)
v_90s <- sort(rowSums(m_90s),decreasing=TRUE)
d_90s <- data.frame(word = names(v_90s),freq=v_90s)

wordcloud(d_90s$word,d_90s$freq,random.order = FALSE, rot.per = 0.3, scale=c(4,.5),colors=brewer.pal(8,"Dark2"))

#00s
corpus_00s <- VCorpus(VectorSource(hiphop00$stemmedwords))

tdm_00s <- TermDocumentMatrix(corpus_00s)
m_00s <- as.matrix(tdm_00s)
v_00s <- sort(rowSums(m_00s),decreasing=TRUE)
d_00s <- data.frame(word = names(v_00s),freq=v_00s)

wordcloud(d_00s$word,d_00s$freq,random.order = FALSE, rot.per = 0.3, scale=c(4,.5),colors=brewer.pal(8,"Dark2"))

```

```

#10s
corpus_10s <- VCorpus(VectorSource(hiphop10$stemmedwords))

tdm_10s <- TermDocumentMatrix(corpus_10s)
m_10s <- as.matrix(tdm_10s)
v_10s <- sort(rowSums(m_10s),decreasing=TRUE)
d_10s <- data.frame(word = names(v_10s),freq=v_10s)

wordcloud(d_10s$word,d_10s$freq,random.order = FALSE, rot.per = 0.3, scale=c(4,.5),colors=brewer.pal(8,"Dark2"))

ggplot(data=d_80s[1:20,], aes(x=word,y=freq,group=1))+

 geom_line(linetype="dashed")+
 geom_point()+
 labs(title = "Top 20 Common Words under 80s")

ggplot(data=d_90s[1:20,], aes(x=word,y=freq,group=1))+

 geom_line(linetype="dashed")+
 geom_point()+
 labs(title = "Top 20 Common Words under 90s")

ggplot(data=d_00s[1:20,], aes(x=word,y=freq,group=1))+

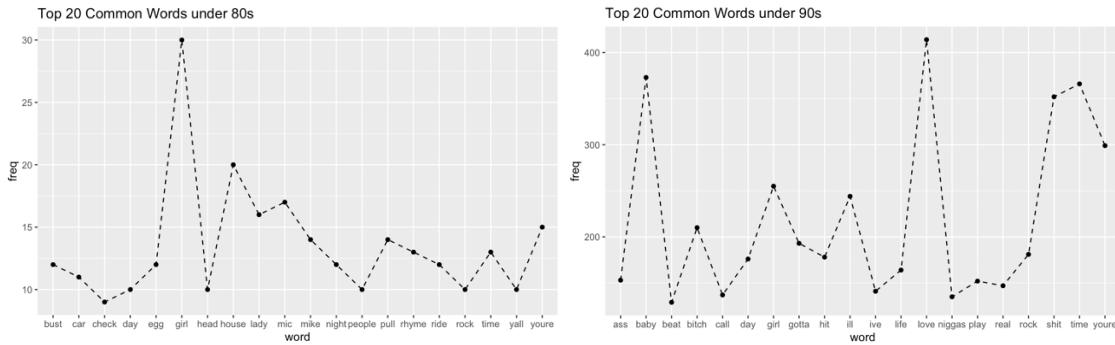
 geom_line(linetype="dashed")+
 geom_point()+
 labs(title = "Top 20 Common Words under 00s")

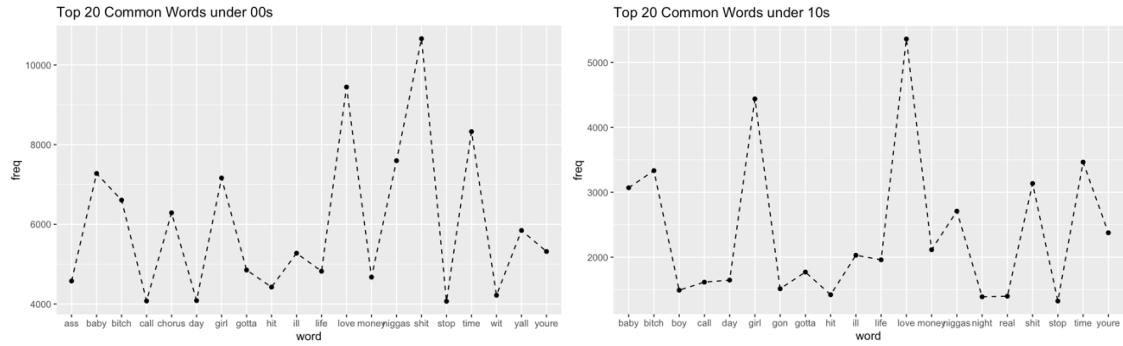
ggplot(data=d_10s[1:20,], aes(x=word,y=freq,group=1))+

 geom_line(linetype="dashed")+
 geom_point()+
 labs(title = "Top 20 Common Words under 10s")

```

I have revealed the “Most Hip-Hop” words above. Now, let’s focus on the different ages. Based on the dataset, I split them into 4 ages: 80s (1980-1989), 90s (1990-1999), 00s (2000-2009) and 10s (2010-2019).





From the above plots, we can see that in 80s, the most popular words are “girl”, “house”, “mic”, “lady”. Then in 90s, they become to “love”, “baby”, “time” and “shit”. Now let’s see the most common words in 00s, they are “shit”, “love”, “time” and “niggas”. In 10s, they are “love”, “girl”, “time” and “bitch”.

Nowadays, the hip-hop is divided into old school and new school styles. In 1980s at a time where drugs and violence were prominent in society. This was also a time where African Americans were being discriminated against which caused them to form their own alliances amongst their own people. At that time, hip-hop was about collectivism. From the popular lyrics in 80s, we can find that people wanted to own a house or have a girlfriend, which they don’t have at that time. Then in 90s, rappers who wanted to own a house or a girlfriend had already married. Then the most popular words became to “love” and “baby”. But at the same time, people not only got to mingle and enjoy the good music but also began to battle amongst themselves. So some words like “shit” began to appear in their lyrics. In the mid 2000’s, we entered the age of internet, the music became much more easier to access. So the new school appeared. At that time, the hip-hop got worse and worse, most of the artist were just doing it for the money. So the content of lyrics has become very unnutritive. Today’s music is an offspring of the early 2000’s and pinch of the 90’s.

Although the garbage trap is very popular nowadays and very simple melodies can make a song, money, sex and violence never represents all Hip-hop. The 2pac’s lyrics have the feeling of resistance and the pursuit of freedom. Jay-Z tries to sort out himself on his new albums and talks about family, children, friends, society, careers and other topics. Nicki Minaj keeps telling her fans not to drop from school.

## The Only Unchanged Word



After analyzing so many lyrics for all kinds of songs, we can find that there is only one unchanged word all the time for all genres. That is “LOVE”. Whether it’s hip-hop or other forms of music, the central theme people want to express is always love.