

A Data Story on Lyrics

Out[1]: [Click here to toggle on/off the raw code.](#)

Overview

By applying the starter code provided by the instructor, we obtain a dataset that contains the song name, artist, genre, lyrics and stemmed-words of over 100000 songs.

Out[3]:

	Unnamed: 0	song	year	artist	genre	lyrics	id	stemmedwords
0	1	when-you-were-with-me	2009	a	Hip-Hop	I stopped by the house we called our home\r\nI...	1	stop house call home rundown grass overgrown s...
1	2	careless-whisper	2009	a	Hip-Hop	I feel so unsure\r\nAs I take your hand and le...	2	unsure hand lead dance floor music die eyes ca...
2	3	2-59	2007	a	Hip-Hop	Mark:] Sunday football I got boot off the pitc...	3	mark sunday football boots pitch people gamble...
3	4	power-of-desire	2007	a	Hip-Hop	[Chris:] Fallin' for a fantasy\r\nI threw away...	4	chris fallin fantasy threw destiny stop feelin...
4	5	you-re-not-in-love	2007	a	Hip-Hop	something in the way we touch\r\nyou hold my h...	5	touch hold hand hold somethings change somethi...

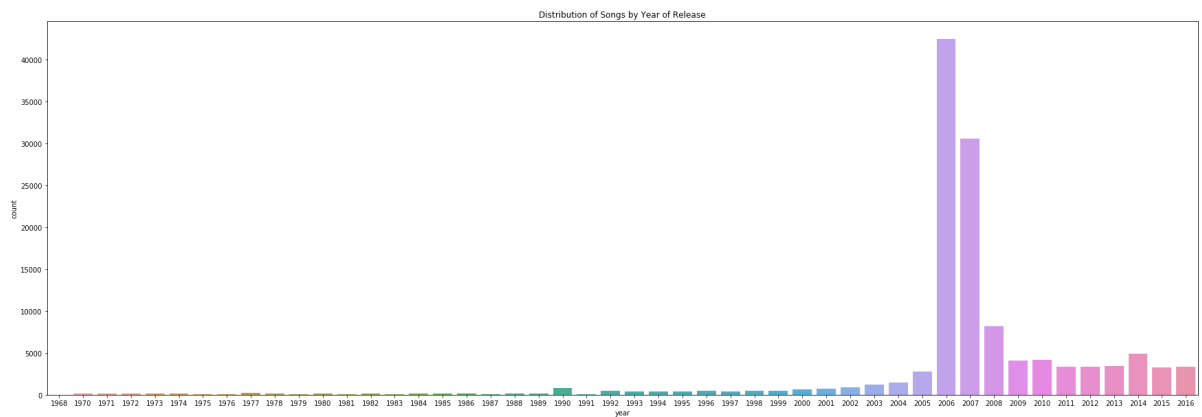
Basic Data visualization

Prior to performing any analysis, we would like to do some basic EDA and data visualization. We start by taking a look at the distribution of the songs by their year of release.

Year

From the below results we can see that the songs included in the dataset are released from 50 different years with most songs from 2006 and 2007. It is shown that 42457 and 30600 of the songs in the dataset are from the year 2006 and 2007 respectively. It is also shown two songs from the year 702 and 112 appears in the dataset so I decided to exclude them.

Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x1672ced5630>



Genres

Next we want to take a look at the genres of the songs. It is shown that the songs come from 12 different genres with 64845 Rock songs and 18697 Pop songs ranking first and second.

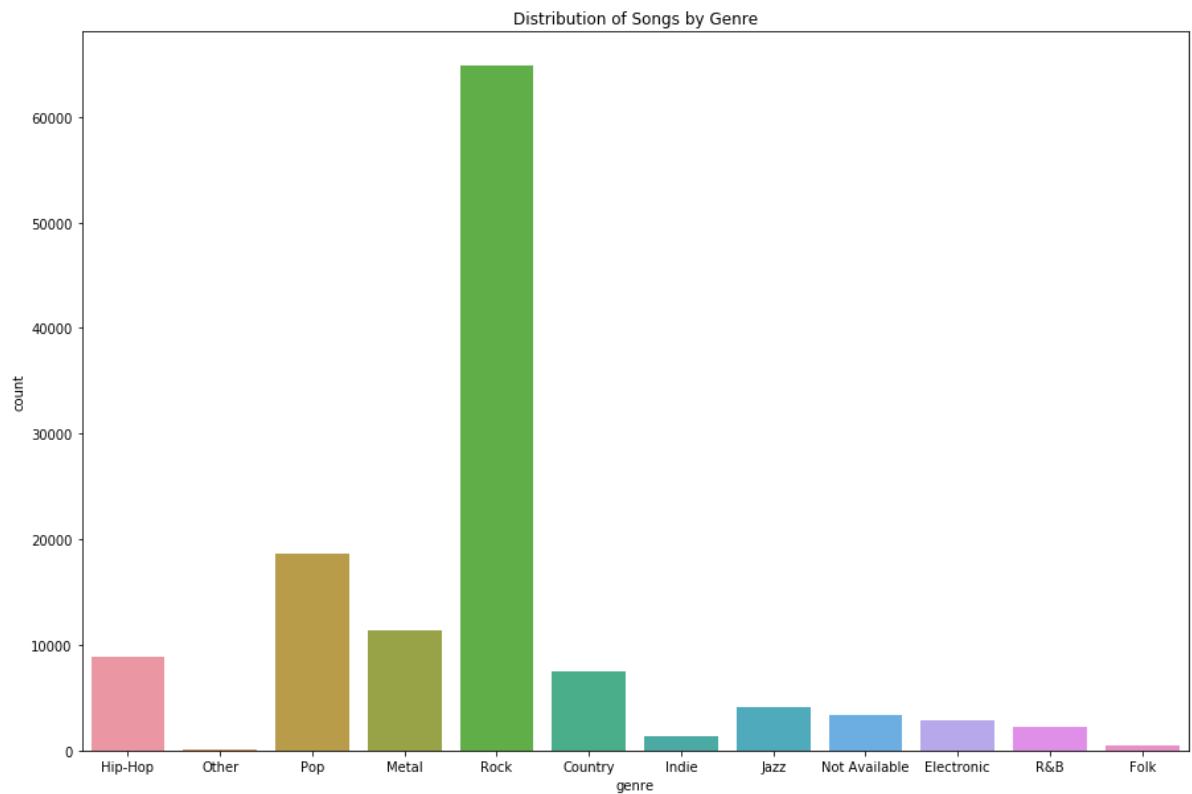
Out[9]: 12

Out[10]:

Rock	64845
Pop	18697
Metal	11311
Hip-Hop	8903
Country	7534
Jazz	4122
Not Available	3325
Electronic	2812
R&B	2174
Indie	1318
Folk	514
Other	147

Name: genre, dtype: int64

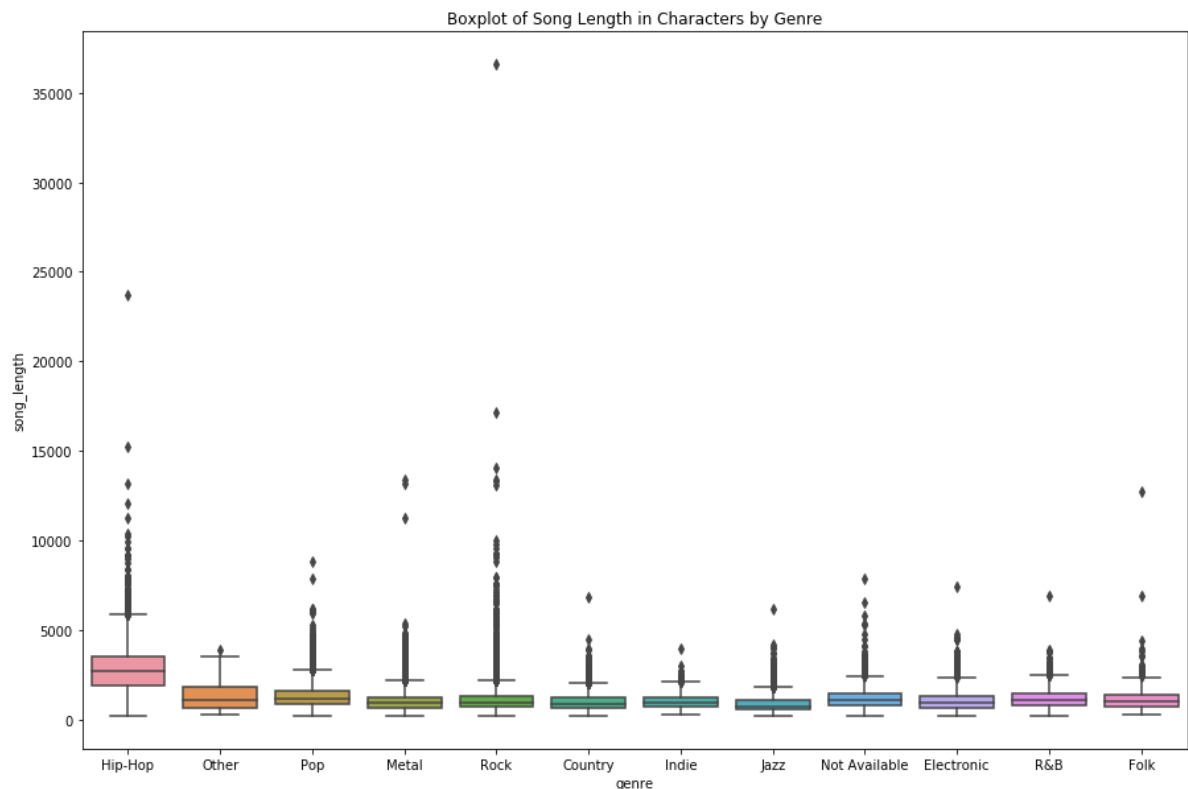
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x16716682f28>



Song Length

Songs of different genres tend to have different lengths. Here we obtain the length of the lyrics in characters and display the boxplot of these lengths by their genres. It is clear that songs from the Hip-Hop genre have the longest lyrics. Does that mean Hip-Hop songs contain the most information?

Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x16716682908>



Topic Modelling

Being a huge Hip-Hop fan myself, I am very curious about the themes and topics depicted by rap songs. Thus I conducted topic modelling on lyrics of the Hip-Hop genre with respect to the years of their release. Both the number and percentage of songs in each topic are generated and plotted. From the tables and graphs, we are able to see how Hip-Hop song themes evolve over time. In order to conduct topic modelling, I employed the LatentDirichletAllocation package from `sklearn.decomposition` and `CountVectorizer` from `sklearn.feature_extraction`. By applying the lda topic model on Hip-Hop lyrics, I am able to identify 5 topics based on 10 key-words each. Based on deep personal understanding of Hip-Hop, I assigned the following themes to each topic: 'rapper swag', 'street life', 'female', 'society' and 'party'. The lda model returned scores on each topic for each song and we need to determine a threshold to classify these topics. Here I assign the song to a topic if the score is greater than 0.1. After the songs are assigned to different topics, both the number and percentage of songs in each topic are generated and plotted. Here we are able to see that in the 90s, the majority of Hip-Hop music are focused on society and as time goes by, more songs about partying and swag are written.

Out[48]:

	Unnamed: 0	song	year	artist	genre	lyrics	id	stemmedwords	song_length
0	1	when-you-were-with-me	2009	a	Hip-Hop	I stopped by the house we called our home\r\nI...	1	stop house call home rundown grass overgrown s...	522
1	2	careless-whisper	2009	a	Hip-Hop	I feel so unsure\r\nAs I take your hand and le...	2	unsure hand lead dance floor music die eyes ca...	1564
2	3	2-59	2007	a	Hip-Hop	Mark:] Sunday football I got boot off the pitc...	3	mark sunday football boots pitch people gamble...	1813
3	4	power-of-desire	2007	a	Hip-Hop	[Chris:] Fallin' for a fantasy\r\nI threw away...	4	chris fallin fantasy threw destiny stop feelin...	785
4	5	you-re-not-in-love	2007	a	Hip-Hop	something in the way we touch\r\nyou hold my h...	5	touch hold hand hold somethings change somethi...	611

C:\Users\lyc07\AppData\Local\Continuum\anaconda3\lib\site-packages\sklearn\decomposition\online_lda.py:314: DeprecationWarning: n_topics has been renamed to n_components in version 0.19 and will be removed in 0.21
 DeprecationWarning)

Topic #0: rock rhyme rap hit stop move time beat black kids
 Topic #1: hand dance run lord hear music people sing song air
 Topic #2: girl baby love time night tonight call shake boy body
 Topic #3: time life love day live world mind die stay heart
 Topic #4: money ass hit hoes boy game real thug ride bout

Out[50]:

	rapper swag	street life	female	society	party
0	0.009695	0.009617	0.009783	0.961212	0.009694
1	0.003056	0.470011	0.003127	0.520756	0.003050
2	0.003454	0.230236	0.226987	0.440254	0.099069
3	0.006039	0.228444	0.299467	0.460113	0.005938
4	0.006205	0.006187	0.785930	0.006264	0.195413

Out[52]:

	Unnamed: 0	song	year	artist	genre	lyrics	id	stemmedwords	song_length	rap sv
0	1	when-you-were-with-me	2009	a	Hip-Hop	I stopped by the house we called our home\r\nI...	1	stop house call home rundown grass overgrown s...	522	0.0091
1	2	careless-whisper	2009	a	Hip-Hop	I feel so unsure\r\nAs I take your hand and le...	2	unsure hand lead dance floor music die eyes ca...	1564	0.0031
2	3	2-59	2007	a	Hip-Hop	Mark:] Sunday football I got boot off the pitc...	3	mark sunday football boots pitch people gamble...	1813	0.0031
3	4	power-of-desire	2007	a	Hip-Hop	[Chris:] Fallin' for a fantasy\r\nI threw away...	4	chris fallin fantasy threw destiny stop feelin...	785	0.0061
4	5	you-re-not-in-love	2007	a	Hip-Hop	something in the way we touch\r\nyou hold my h...	5	touch hold hand hold somethings change somethi...	611	0.0061

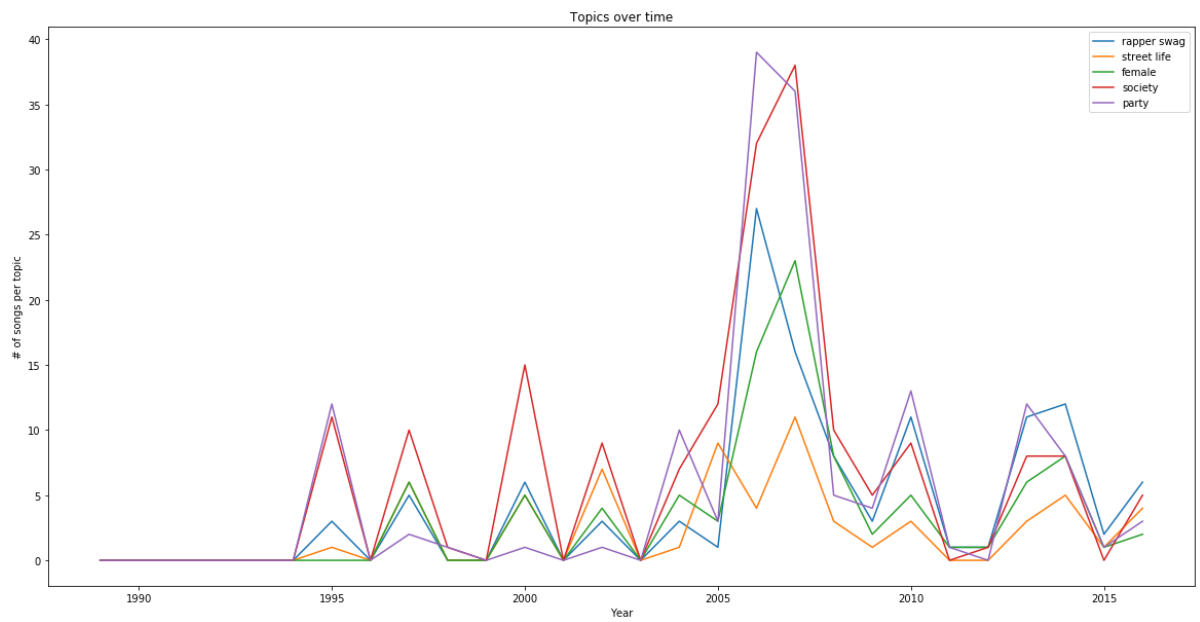
Out[55]:

	Unnamed: 0	song	year	artist	genre	lyrics	id	stemmedwords	song_length	rappe swa
0	1	when- you- were- with-me	2009	a	Hip- Hop	I stopped by the house we called our home\r\nI...	1	stop house call home rundown grass overgrown s...	522	0.1
1	2	careless- whisper	2009	a	Hip- Hop	I feel so unsure\r\nAs I take your hand and le...	2	unsure hand lead dance floor music die eyes ca...	1564	0.1
2	3	2-59	2007	a	Hip- Hop	Mark:] Sunday football I got boot off the pitc...	3	mark sunday football boots pitch people gamble...	1813	0.1
3	4	power- of-desire	2007	a	Hip- Hop	[Chris:] Fallin' for a fantasy\r\nI threw away...	4	chris fallin fantasy threw destiny stop feelin...	785	0.1
4	5	you-re- not-in- love	2007	a	Hip- Hop	something in the way we touch\r\nyou hold my h...	5	touch hold hand hold somethings change somethi...	611	0.1

Out[57]:

	year	Unnamed: 0	id	song_length	rapper swag	street life	female	society	party
0	1989	285230	285252	32504	0.0	0.0	0.0	0.0	0.0
1	1992	630423	630483	47543	0.0	0.0	0.0	0.0	0.0
2	1993	750294	750366	25597	0.0	0.0	0.0	0.0	0.0
3	1994	1540446	1540568	108353	0.0	0.0	0.0	0.0	0.0
4	1995	3444257	3444580	135964	3.0	1.0	0.0	11.0	12.0
5	1996	3173795	3174074	97622	0.0	0.0	0.0	0.0	0.0
6	1997	474	474	14558	5.0	6.0	6.0	10.0	2.0
7	1998	1937109	1937293	71914	0.0	0.0	0.0	1.0	1.0
8	1999	2219424	2219607	72821	0.0	0.0	0.0	0.0	0.0
9	2000	1742965	1743118	145939	6.0	5.0	5.0	15.0	1.0
10	2001	3325384	3325698	142455	0.0	0.0	0.0	0.0	0.0
11	2002	5533001	5533554	299171	3.0	7.0	4.0	9.0	1.0
12	2003	4524106	4524539	168473	0.0	0.0	0.0	0.0	0.0
13	2004	13915352	13916690	563900	3.0	1.0	5.0	7.0	10.0
14	2005	15716821	15718185	756177	1.0	9.0	3.0	12.0	3.0
15	2006	209028437	209047145	8776967	27.0	4.0	16.0	32.0	39.0
16	2007	98868404	98877473	4186553	16.0	11.0	23.0	38.0	36.0
17	2008	44845076	44849238	1710592	8.0	3.0	8.0	10.0	5.0
18	2009	29156019	29158766	1031984	3.0	1.0	2.0	5.0	4.0
19	2010	36222808	36226246	1238230	11.0	3.0	5.0	9.0	13.0
20	2011	27980444	27983209	890647	1.0	0.0	1.0	0.0	1.0
21	2012	27427261	27429986	801414	1.0	0.0	1.0	1.0	0.0
22	2013	20553045	20555109	774839	11.0	3.0	6.0	8.0	12.0
23	2014	34053315	34056637	1025001	12.0	5.0	8.0	8.0	8.0
24	2015	17234402	17236133	524032	2.0	1.0	1.0	0.0	1.0
25	2016	21981538	21983748	706232	6.0	4.0	2.0	5.0	3.0

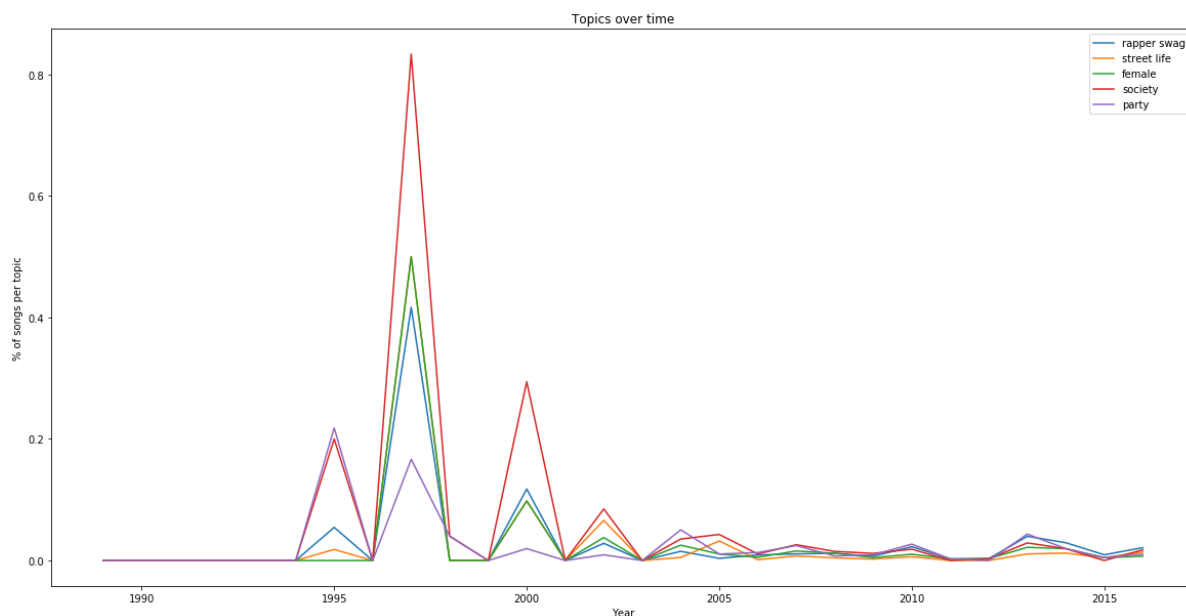
Out[58]: <matplotlib.legend.Legend at 0x16739c4d6a0>



Out[61]:

	year	Unnamed: 0	id	song_length	rapper swag	street life	female	society	party
0	1989	285230	285252	32504	0.000000	0.000000	0.000000	0.000000	0.000000
1	1992	630423	630483	47543	0.000000	0.000000	0.000000	0.000000	0.000000
2	1993	750294	750366	25597	0.000000	0.000000	0.000000	0.000000	0.000000
3	1994	1540446	1540568	108353	0.000000	0.000000	0.000000	0.000000	0.000000
4	1995	3444257	3444580	135964	0.054545	0.018182	0.000000	0.200000	0.218182
5	1996	3173795	3174074	97622	0.000000	0.000000	0.000000	0.000000	0.000000
6	1997	474	474	14558	0.416667	0.500000	0.500000	0.833333	0.166667
7	1998	1937109	1937293	71914	0.000000	0.000000	0.000000	0.040000	0.040000
8	1999	2219424	2219607	72821	0.000000	0.000000	0.000000	0.000000	0.000000
9	2000	1742965	1743118	145939	0.117647	0.098039	0.098039	0.294118	0.019608
10	2001	3325384	3325698	142455	0.000000	0.000000	0.000000	0.000000	0.000000
11	2002	5533001	5533554	299171	0.028302	0.066038	0.037736	0.084906	0.009434
12	2003	4524106	4524539	168473	0.000000	0.000000	0.000000	0.000000	0.000000
13	2004	13915352	13916690	563900	0.015152	0.005051	0.025253	0.035354	0.050505
14	2005	15716821	15718185	756177	0.003571	0.032143	0.010714	0.042857	0.010714
15	2006	209028437	209047145	8776967	0.009106	0.001349	0.005396	0.010793	0.013153
16	2007	98868404	98877473	4186553	0.010951	0.007529	0.015743	0.026010	0.024641
17	2008	44845076	44849238	1710592	0.011976	0.004491	0.011976	0.014970	0.007485
18	2009	29156019	29158766	1031984	0.007059	0.002353	0.004706	0.011765	0.009412
19	2010	36222808	36226246	1238230	0.022822	0.006224	0.010373	0.018672	0.026971
20	2011	27980444	27983209	890647	0.002618	0.000000	0.002618	0.000000	0.002618
21	2012	27427261	27429986	801414	0.003077	0.000000	0.003077	0.003077	0.000000
22	2013	20553045	20555109	774839	0.039855	0.010870	0.021739	0.028986	0.043478
23	2014	34053315	34056637	1025001	0.029268	0.012195	0.019512	0.019512	0.019512
24	2015	17234402	17236133	524032	0.009434	0.004717	0.004717	0.000000	0.004717
25	2016	21981538	21983748	706232	0.021201	0.014134	0.007067	0.017668	0.010601

Out[62]: <matplotlib.legend.Legend at 0x1672e61e320>

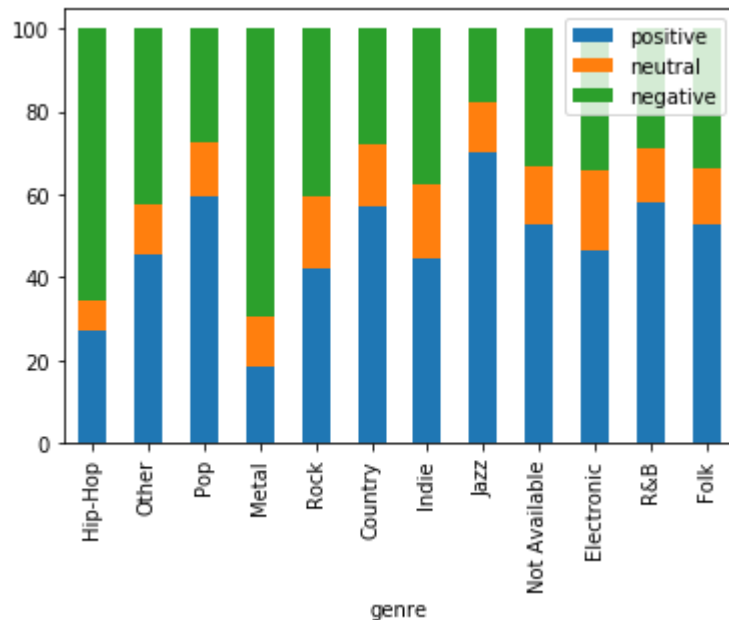


Word Cloud

By generating wordcloud on lyrics, we can see intuitively what are the concentrations and focuses of these lyrics. Here I employed the wordcloud package and generated wordcloud on the Hip-Hop genre to identify the most frequently used words and compared them with wordclouds of Rock and Pop genre. However, after removing certain stop words such as certain slangs and swear words, the wordcloud of Hip-Hop music and Rock music and Pop music does not exhibit much difference.

Based on the graph and table below, it is clear that most songs in the Hip-Hop genre convey negativity. In fact, around 65% of Hip-Hop lyrics are categorized as negative, 27% are positive and 7.2% are neutral. However, there is one genre with a higher portion of negative lyrics and that is Metal with 69.4%.

<Figure size 2160x1080 with 0 Axes>



Out[44]:

	genre	positive	neutral	negative
0	Hip-Hop	27.159384	7.222285	65.618331
1	Other	45.578231	12.244898	42.176871
2	Pop	59.501524	13.130449	27.368027
3	Metal	18.627884	11.970648	69.401468
4	Rock	42.134320	17.225692	40.639988
5	Country	57.366605	14.773029	27.860366
6	Indie	44.385432	17.905918	37.708649
7	Jazz	69.941776	12.518195	17.540029
8	Not Available	52.992481	13.804511	33.203008
9	Electronic	46.728307	19.238976	34.032717
10	R&B	57.957682	13.293468	28.748850
11	Folk	52.723735	13.424125	33.852140

Summary

- Songs from different genres have different lengths in their lyrics. Hip-Hop music tends to have the longest lyrics.
- The theme of Hip-Hop music changes with respect to time. In the 90s, the dominant theme was about society whereas at present more songs about partying are written.
- Hip-Hop music tends to contain negative sentiments. Listening to Hip-Hop music could be a good way to lash out negative emotions.