

# Sen Dai sd 3227 Project1

Nowadays, everyone loves to listen to music. They not only love the melody, but also love the lyrics. However, different genres of music have different type of lyrics. The interesting thing is, that we can get lots of hints from analyzing the lyrics instead of listening the music. To get the secret of the lyrics, I decide to analysis some lyrics of different genres from 1970s to 2010s by using R. I only focus on three aspect:

- 1.What's the popular genres of music in different decades?
- 2.What's the difference of lyrics in different genres of music?
- 3.What's the habit of the artist to write the lyrics?Are they Shakespeare or Dumas?

I analysis the data by different methods and get lots of interesting results.I'll show what i did step by step and tell a fantastic "data story".

Firstly, I load all the required libraries.(step0)

## Step0 - Load all the required libraries

Secondly, I load and clean the data for future use.(step1-3)

## Step1 - Load the processed data and artist to use

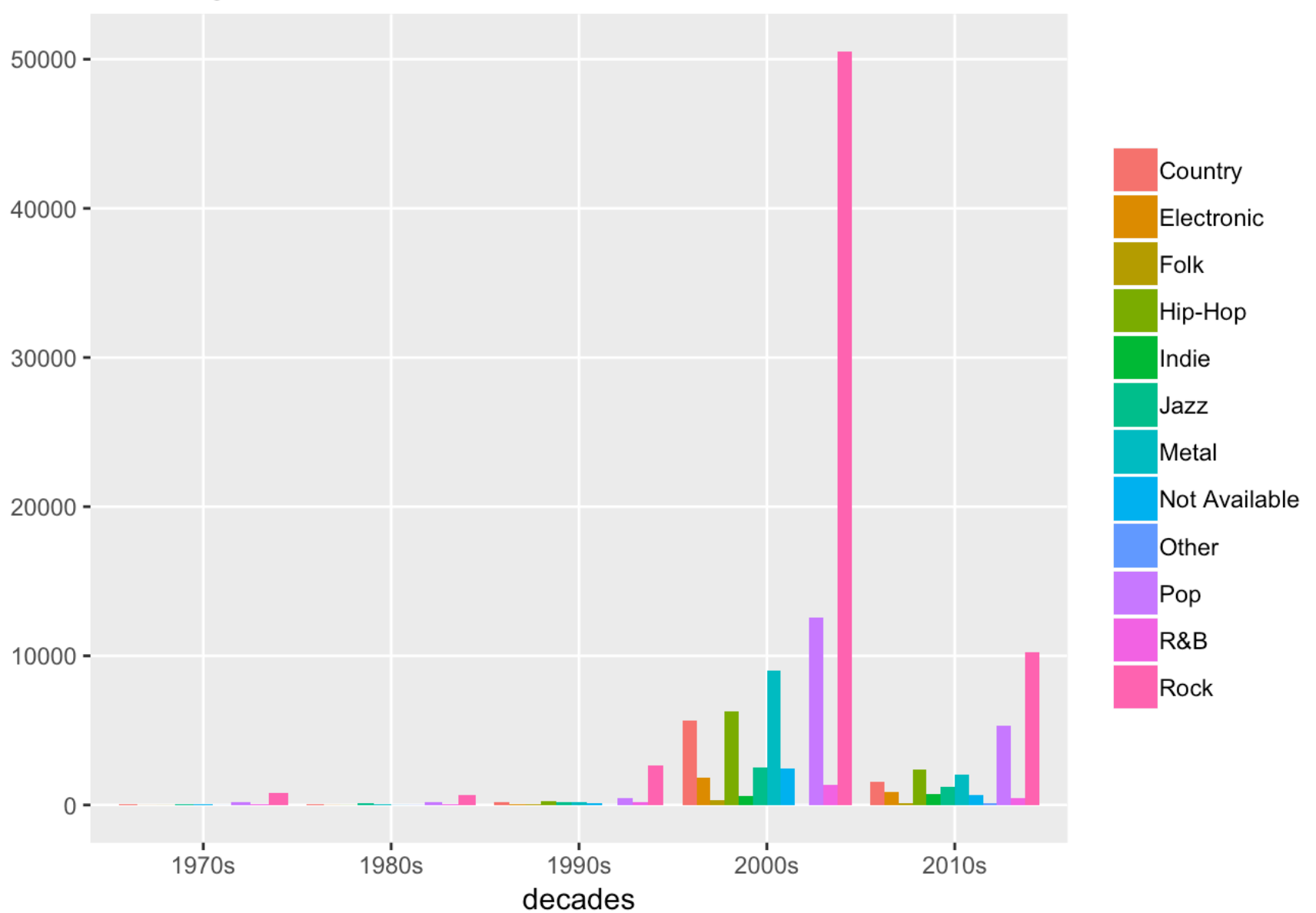
## Step2 - Preparation of data for further use

## Step3 - Create the decade column for the data

## Step4 - Analysis data: Song stats graphs of different genres in each decades

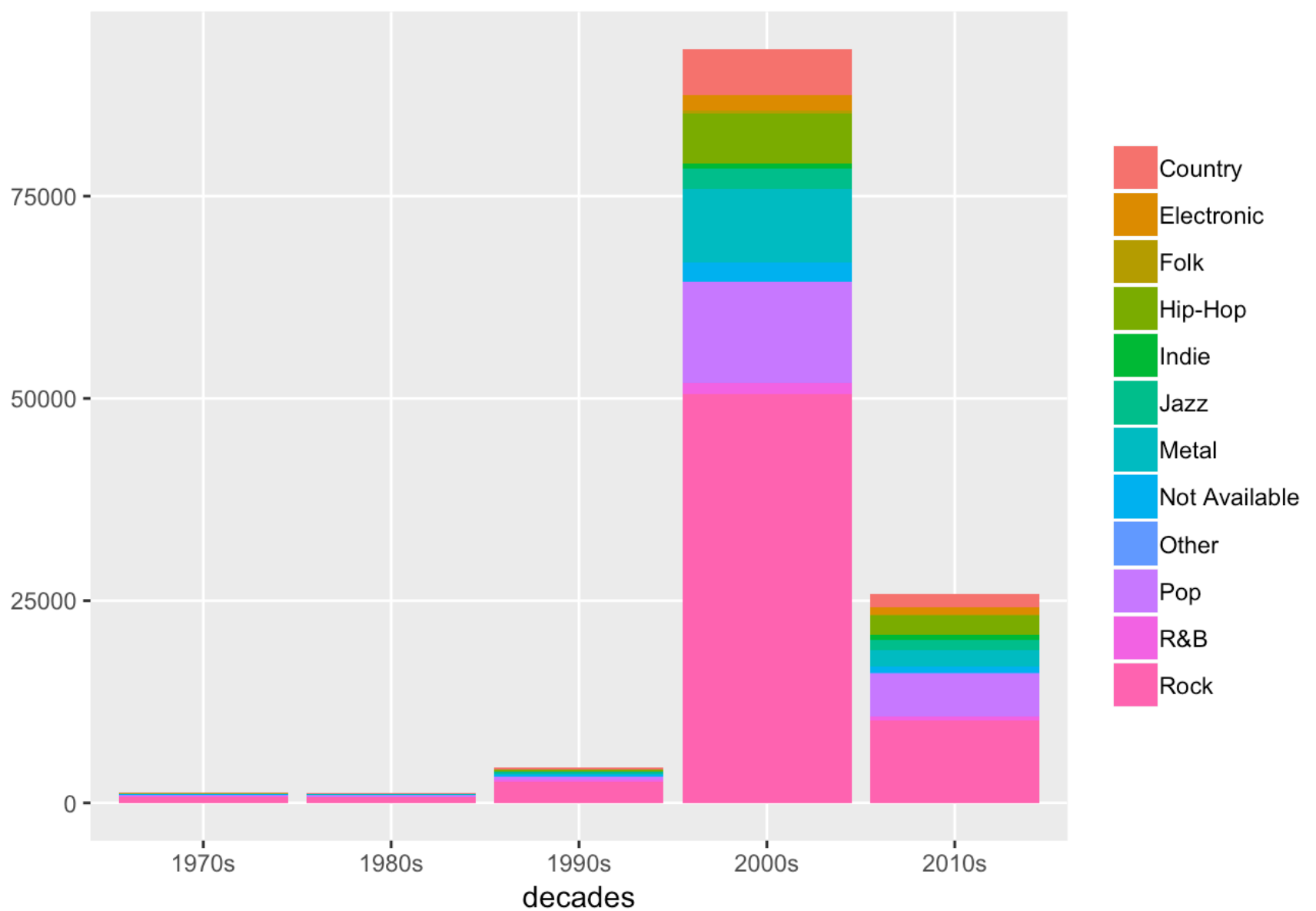
Rightnow,I get a question. I really want to know what kind of music do people love at differet decades. To solve this question, I plot some graphs to get the answer. From Fig1.1, we can see that different type of music at different decades.It's difficult to know what kind of music do people like from 1970s to 1990s since the data is not enough. We can easily find out Rock is the most popular music in 2000s, the second popular one is Pop and the third one is Metal. However, in 2010s, the third popular changed from Metal to Hip Hop. Hence we can see that with the time, listener change their preference, but not a lot.

Fig1.1



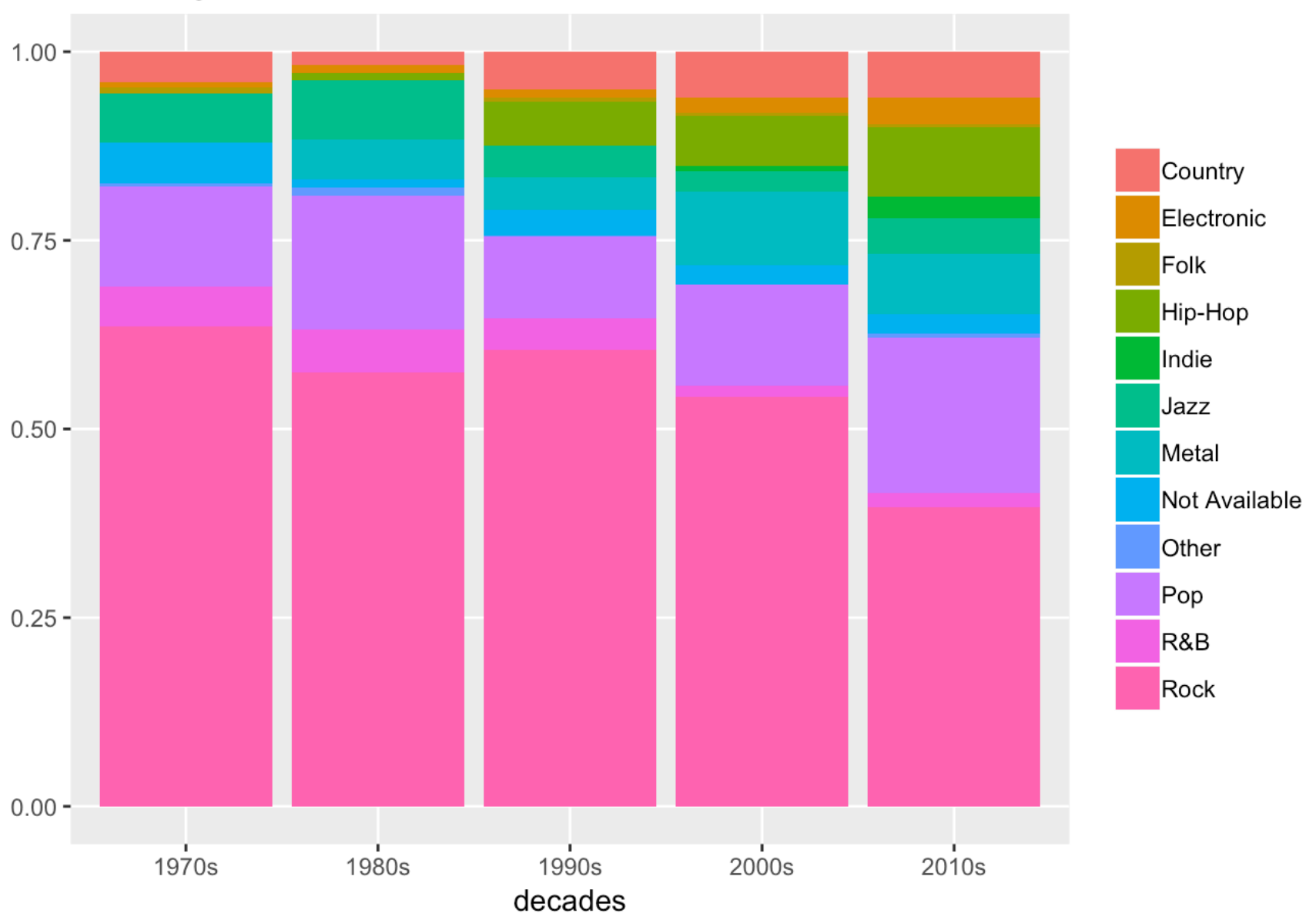
From Fig1.2, we can see that before 2000s, the number of music is just a little. However there are lots of music released in 2000s, and the number of music decrease in 2010s. Maybe there were something happen in 2000s that stimulated the music market. For my own experience, I begin to listen American music in 2005. I think this rocket in 2000s because of the Internet. People can get the music easily on the Internet.

Fig1.2



From Fig1.3, we can get some general information about each genres of music. For example, Rock is the most popular type of music no matter in which decades; More and more people begin to listen Hip-Hop, Electronic and Metal; Recently, R&B is not as popular as before; Just a few people is interested in Indie no matter in which decades.

Fig1.3



After this three figures, we can get some basic information about the lyrics from 1970s to 2010s. However, to get more interesting things, we need to do more analyzing.

#### Step5 - Analysis data: Word frequency in different genre

Suddenly, another question pop in my mind. What are the lyrics really talk about? Do the different genres tell the same kind of story of music or different? To solve this problem, I begin a new analyzing.

To get what kind of word is the popular word in different genres, I try to get the 50 top high frequency word for each genres. In this way we can analysis the high frequency word and compare with each other to find the secret behind the lyrics. Right now, I'll post the top 50 frequency word for each genres music:

## the top 50 high frequency word for Country

word	n
<chr>	<int>
love	13308
time	6004

ill	4770
youre	4521
heart	4462
ive	4232
day	3800
night	3548
baby	3425
home	2853
1-10 of 50 rows	<div> Previous 1 2 3 4 5 Next </div>

# the top 50 high frequency word for Electronic

word	n
<chr>	<int>
love	4865
time	2259
youre	2059
world	1582
baby	1542
night	1447
ill	1363
life	1353
heart	1343
day	1245
1-10 of 50 rows	<div> Previous 1 2 3 4 5 Next </div>

# the top 50 high frequency word for Folk

word	n
<chr>	<int>
love	578
time	403

day	337
ill	243
ive	225
night	183
home	177
heart	176
youre	173
eyes	163
1-10 of 50 rows	
Previous 1 2 3 4 5 Next	

# the top 50 high frequency word for Hip-Hop

word	n
<chr>	<int>
love	15234
shit	14150
ya	12531
time	12176
girl	11887
baby	10728
yo	10574
niggas	10441
em	10216
bitch	10152
1-10 of 50 rows	
Previous 1 2 3 4 5 Next	

# the top 50 high frequency word for Indie

word	n
<chr>	<int>
love	1669
youre	1062

time	964
ill	881
ive	758
day	646
heart	598
home	539
eyes	517
life	467
1-10 of 50 rows	
Previous 1 2 3 4 5 Next	

## the top 50 high frequency word for Jazz

word	n
<chr>	<int>
love	9409
youre	2628
baby	2555
heart	2488
day	2321
time	2270
ill	2188
dream	1795
night	1747
ive	1590
1-10 of 50 rows	
Previous 1 2 3 4 5 Next	

## the top 50 high frequency word for Metal

word	n
<chr>	<int>
time	8462
life	8271
die	5253
eyes	5237

world	5193
youre	5171
live	5025
love	4870
day	4557
soul	4029
1-10 of 50 rows	
Previous 1 2 3 4 5 Next	

# the top 50 high frequency word for Pop

word	n
<chr>	<int>
love	53150
baby	21657
youre	21582
time	19103
heart	14129
ill	14061
girl	11908
life	10949
day	10942
ive	10106
1-10 of 50 rows	
Previous 1 2 3 4 5 Next	

# the top 50 high frequency word for R&B

word	n
<chr>	<int>
love	6055
baby	4128
time	2352
youre	2170
girl	1640



ill	1526
heart	1516
day	1244
night	1100
life	1076
1-10 of 50 rows	
Previous 1 2 3 4 5 Next	

# the top 50 high frequency word for Rock

word	n
<chr>	<int>
love	77803
time	56019
youre	50548
ill	33160
day	32694
ive	30295
baby	29851
life	27917
night	26968
heart	25737
1-10 of 50 rows	
Previous 1 2 3 4 5 Next	

From the data, we can find lots of interesting things. Firstly, we can see that love is the most common word in each type of music. I think this is because music is a way for people to express the emotion. Artists like to show what they love in the song (a person, a luxury car, a kind of spirit, etc). Secondly, baby is a very interesting word in the lyrics, lots of music like this word. We can see that baby is not in the top 50 words only at Metal. Hence we can surmise that Metal music talks more about spirit and soul instead of love, friend and family. Thirdly, lyrics can be the icon for different types of music. I want to talk two typical examples, one is about Hip-Hop music. We know the Hip-Hop music is about sex, drug and money. Hence there are lots of nasty and violent words in Hip-Hop music: shit, bitch, ass, beat, money and so on. It's a kind of icon for Hip-Hop because it's impossible for people to get "bitch" and "ass" from Country music. For Metal music, it's similar: hell, fire, burn, blood, black, kill are the typical words in Metal. Metal music is kind of aggressive and loud. It's really strange for people to get "blood" and "hell" from Pop music. Fourthly, there are lots of things we can talk about the word frequency in different genres of music. However, what I want to talk is about the prediction of the creation of lyrics. I get some inspiration from this analyzing, in the future, we can use programming to create lyrics by machine learning or something else. In this way, maybe we can get some more beautiful lyrics created by AI instead of human being.

#### Step 6 - Analysis data: Lexical Diversity & Lexical Density

After this analyzing, I still get one more question. What's the level and habit of these artists? Are they clever writers or just middle school level? To get the answer, I plan to analyze the lexical diversity and lexical density for these lyrics. Lexical diversity is to use measure the variety of vocabulary. The more varied vocabulary in the lyrics, the higher lexical diversity.

Fig2.1 is the graph about the index of lexical diversity of each song in each year. We can see that before 2000, the lexical diversity is very low. However, after 2000, the lexical diversity becomes higher than before. We can also see the trend of the blue line is going up slightly. In another word, artists open their mind and use more new words in their songs. There is a rocket in 2007 and it back to an average level. I think at this time, more and more artists try the new style of their songs and there be something that push the development of music from 2000s to 2010s. I think it's good for the development of American music because more and more artists try the new things and make progress.

Fig2.1

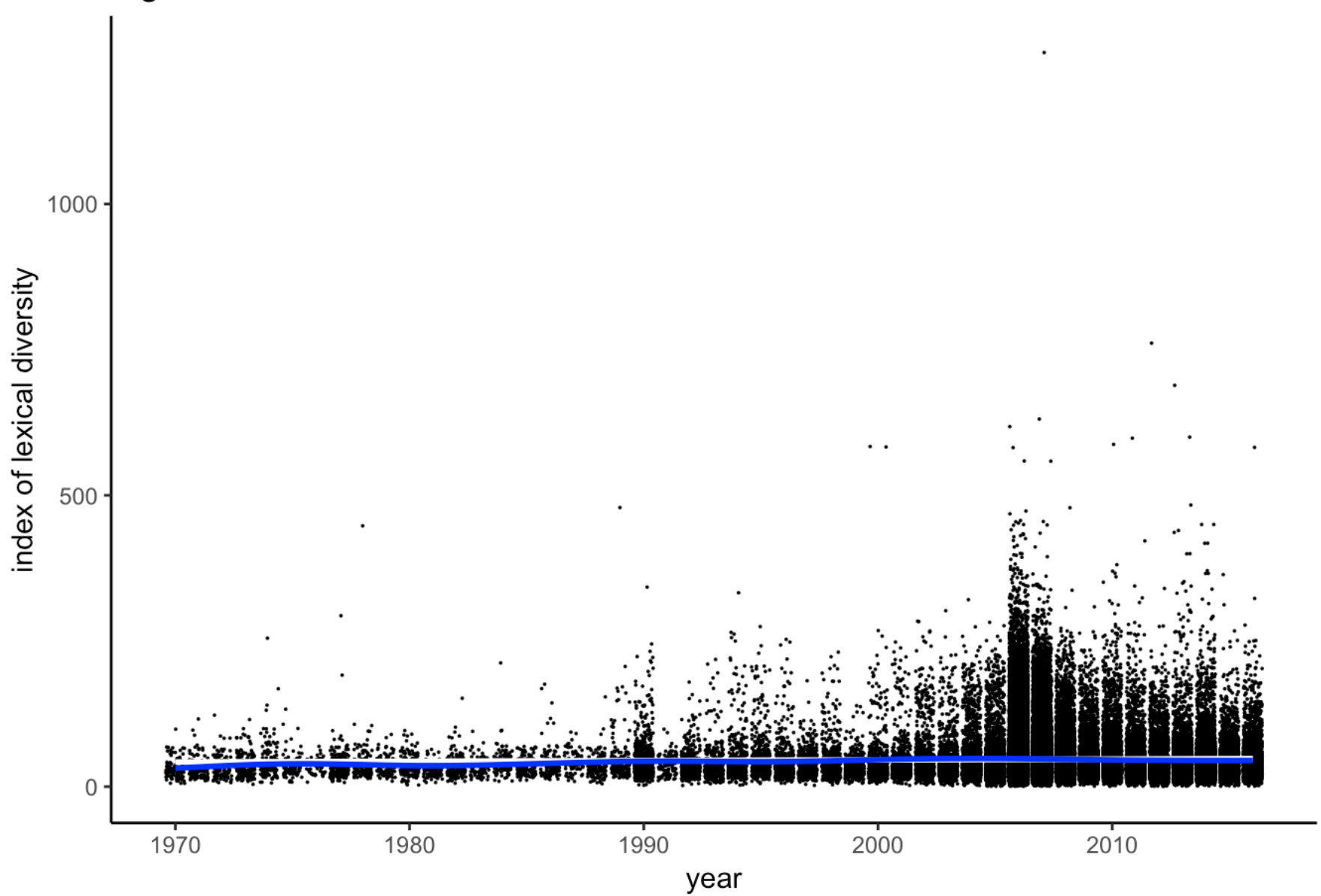
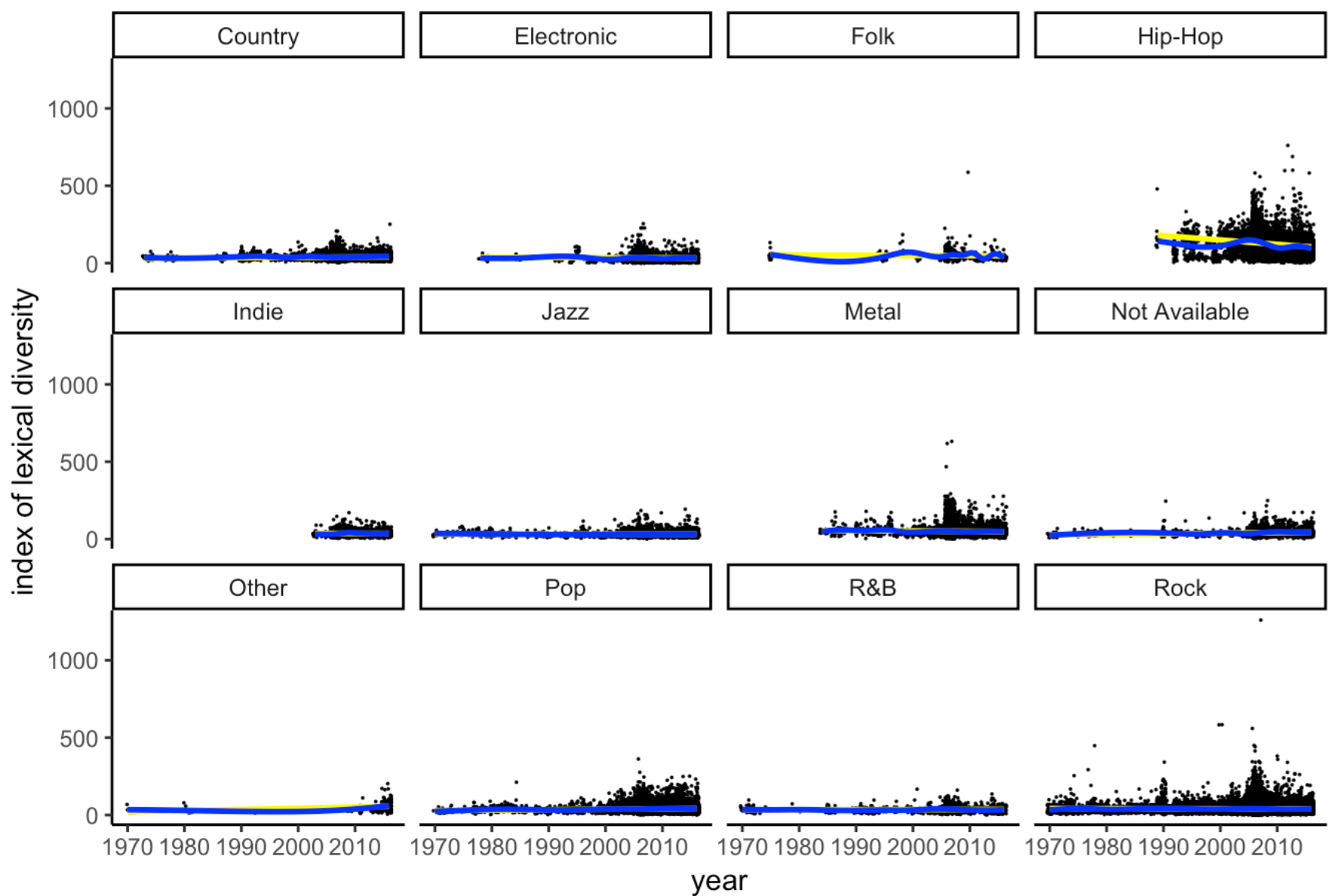


Fig2.2 is the graph about the index of lexical diversity of each genres of songs in each year. To compare each artists in different genres, we can find lots of interesting things. For Country, Folk, R&B and Jazz, the lexical diversity of them are relative low. I think this is because the style of these music is mature. I can't find the big difference from listening Country music between 1980s and 2010s. However, for Hip-Hop, Pop, and Rock, these kind of music refresh every year. It's just like the fashion, artists should update their lyrics day by day to keep their songs in New Wave. It's really make sense why lexical diversity of these kind of music is high.

Fig2.2



Rightnow I want to know more about the artists in another way. After talking about the lexical diversity, I want to talk about lexical density. Lexical density is defined as the number of unique words/the number of total words. In this way we can know the repetition of the word. If lexical density decrease, the repetition rate increase.

Fig2.3 is the graph about the index of lexical density of each songs in each year. We can see that the trend of lexical density goes down, which means the repetition goes up. It's hard to say good or bad for music quality, but in my own idea, I think it means the change of the writing style of lyrics from artists. For example, In 1970s, maybe the whole lyrics narrate a story once in the songs; In 2010s, maybe the whole lyrics narrate a story but repeat it twice even three times in the songs. For another example, maybe with the development of music, people focus on melody instead of the lyrics. Hence artists focus more on melody instead of lyrics.

Fig2.3

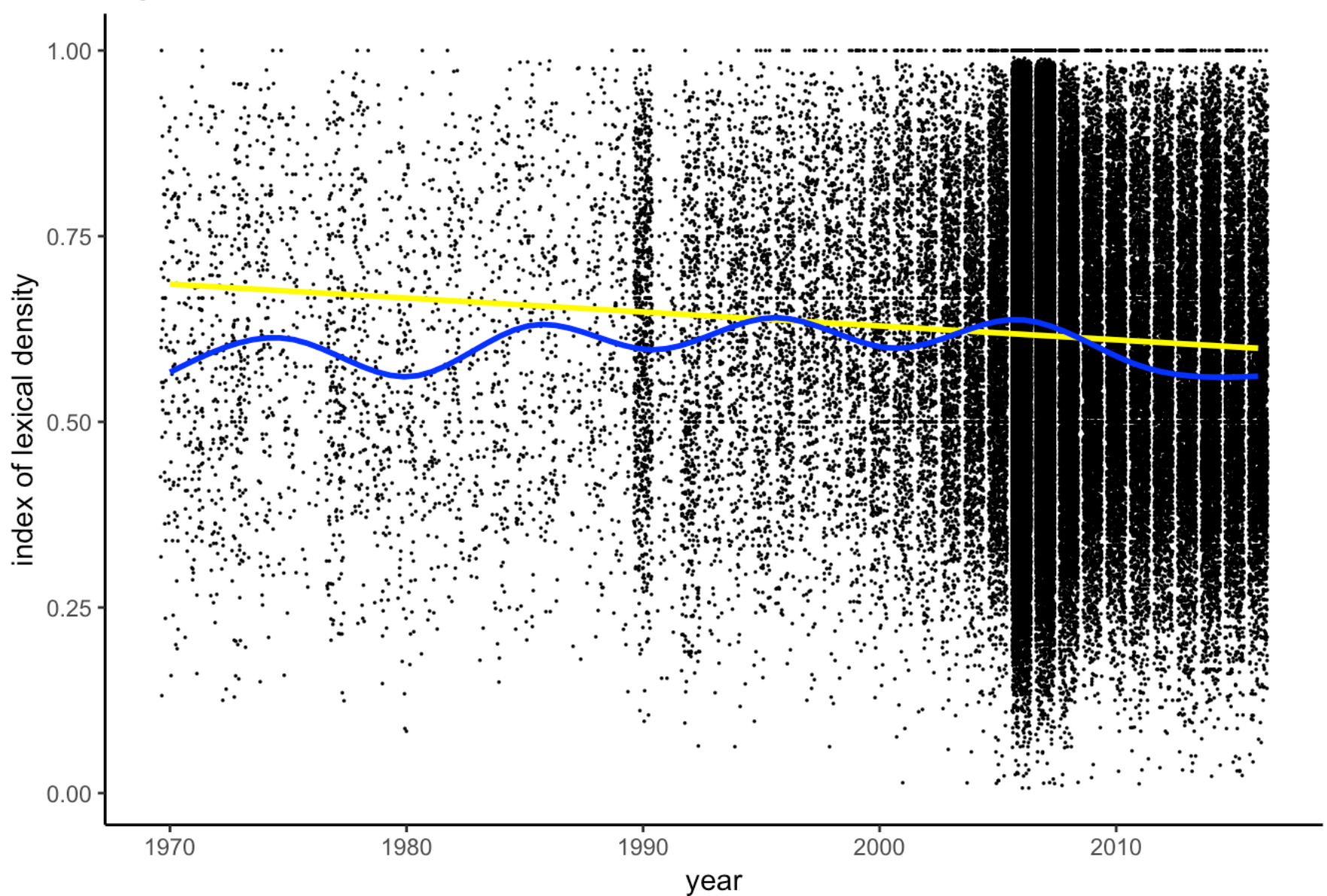
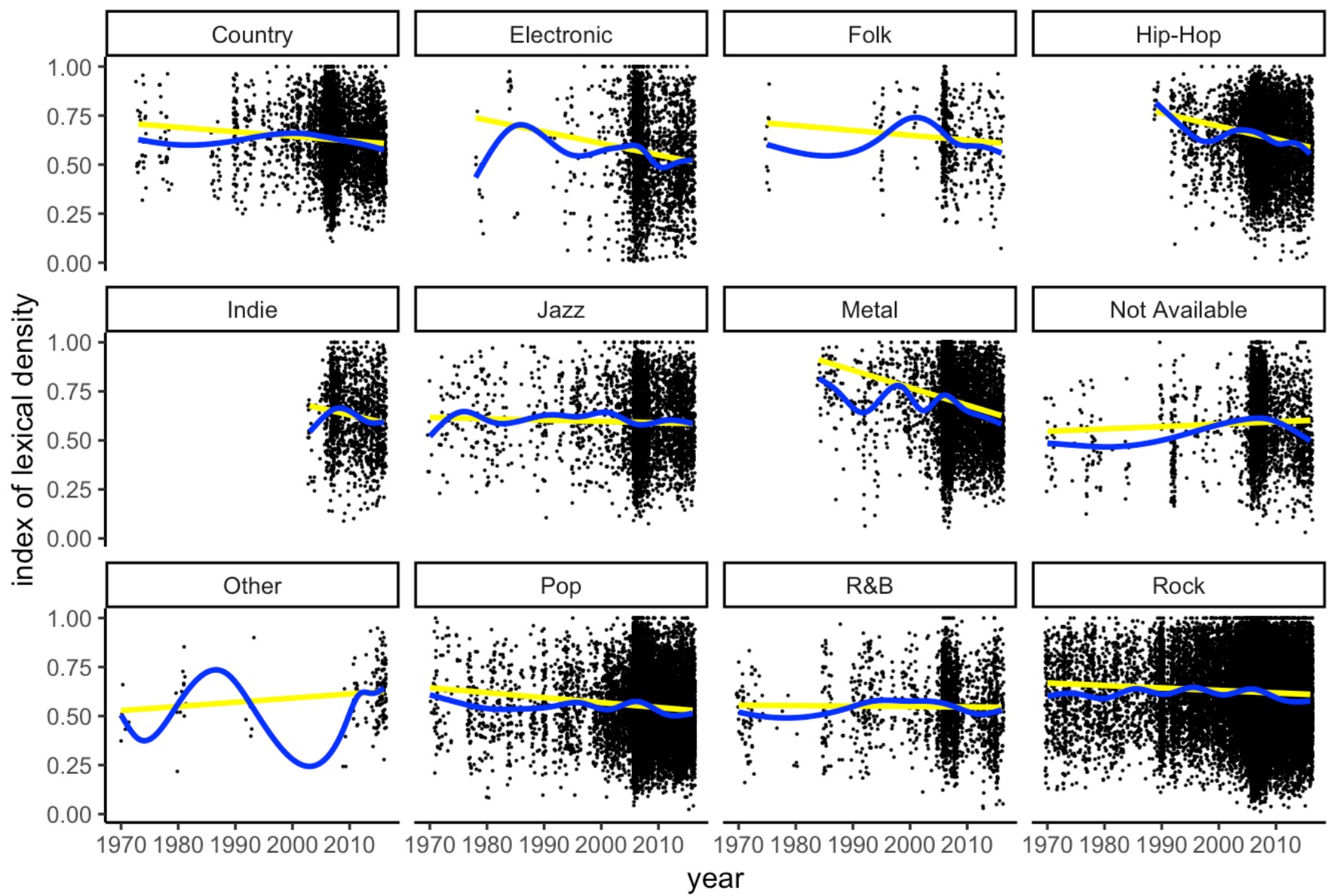


Fig2.4 is the graph about the index of lexical density of each genres of songs in each year. We can get some interesting results from the graph with real life. For Hip-Hop and Metal, we can see a big going-down trend here, which means the repetition rate of these two goes up. Does it mean the quality of music go down? However, from the above analyzing, we can get the Hip-Hop and Metal become more and more popular. Why this happening? Why lyrics become more repetitive and become more popular? The answer is easy: people love it!!! Lots of Hip-Hop and Metal artists perform their songs at live house. To warm up, audience always sing the song with artists together. If the lyrics are too complex, people who never listen the music don't know how to sing. Hence, to solve this problem, artists use some simple and repetitive word in their lyrics that help people to sing the songs with them together. For example, the hip-hop music "My name is " from Eminem have 48-times repetitive lyrics "my name is..." in the song!!!

Fig2.4



After some analyzing, I think I get the answer of all my questions. However, I think it's just a beginning of the analyzing, there are lots of things that we can find by using data science. It's really interesting to get the secret behind the lyrics from natural language processing data mining. Thank you!!!