

# Diversity in the World of Music

Johnny Lu

September 18, 2019

## Primer - Why?

I always wonder, with Trump being the President and the US waging trade wars to almost all major economies, are americans or even the people in this world more conservative in their culture?

Since I am a big fan of music and we are offered with data about song lyrics, I would like to give an answer from the perspective of music. A bit more about myself: I practiced vocal music for 7 years, played some Ukelele and, most importantly, am writing this report while listening to *Sekei ga Owaru Made wa*, SLAM DUNK ED, a masterpiece covered by **Ru’s Piano Ru**, a great Youtuber that does beautiful piano cover for many if not all of Anime music.

‘The end of words is music’. Ideally, we would get a more honest and sincere answer from music than from President’s Tweets. Let’s get started!

## Methodology - How?

In one sentence, I want to know how the diversity of the content of music (*aka, song lyrics*) is.

So, first and foremost, how to measure the **diversity**? For the sake of brevity and the efforts that I plan into Project 1, after several trials of different ideas, I decided on the following:

**Diversity of Music Content = The Dispersion of Different Topics in Song Lyrics**

So, next is to find a way to come up with the *different topics* that we need. As you may already thought of, I applied **topic modelling** on all song lyrics. The result is shown below:

### 11 Topics Generated from Topic Modelling on Song Lyrics

He/Ride
You/Leave
Death/Blood
She/Love
Shit/Money
Life/World
Chorus/Dance
Run/Burn

---

Day/Dream

---

Time/Heart

---

God/Soul

The two words you see are the top 2 most frequent words among all words that belong to that particular topic.

I intentionally set the number of topics to be 11, simply because we have 11 music genres in the dataset. It would be interesting to see how topics compare with genres. According to Wikipedia, music genres ([https://en.wikipedia.org/wiki/Music\\_genre](https://en.wikipedia.org/wiki/Music_genre)) are so-called *conventional categories* invented by human, while topics here represent categories by machine and algorithm.

Before moving on, let's remind ourselves, *Diversity of Music Content = The Dispersion of Different Topics in Song Lyrics*. We now have the topics, but **how exactly to quantify its dispersion?**

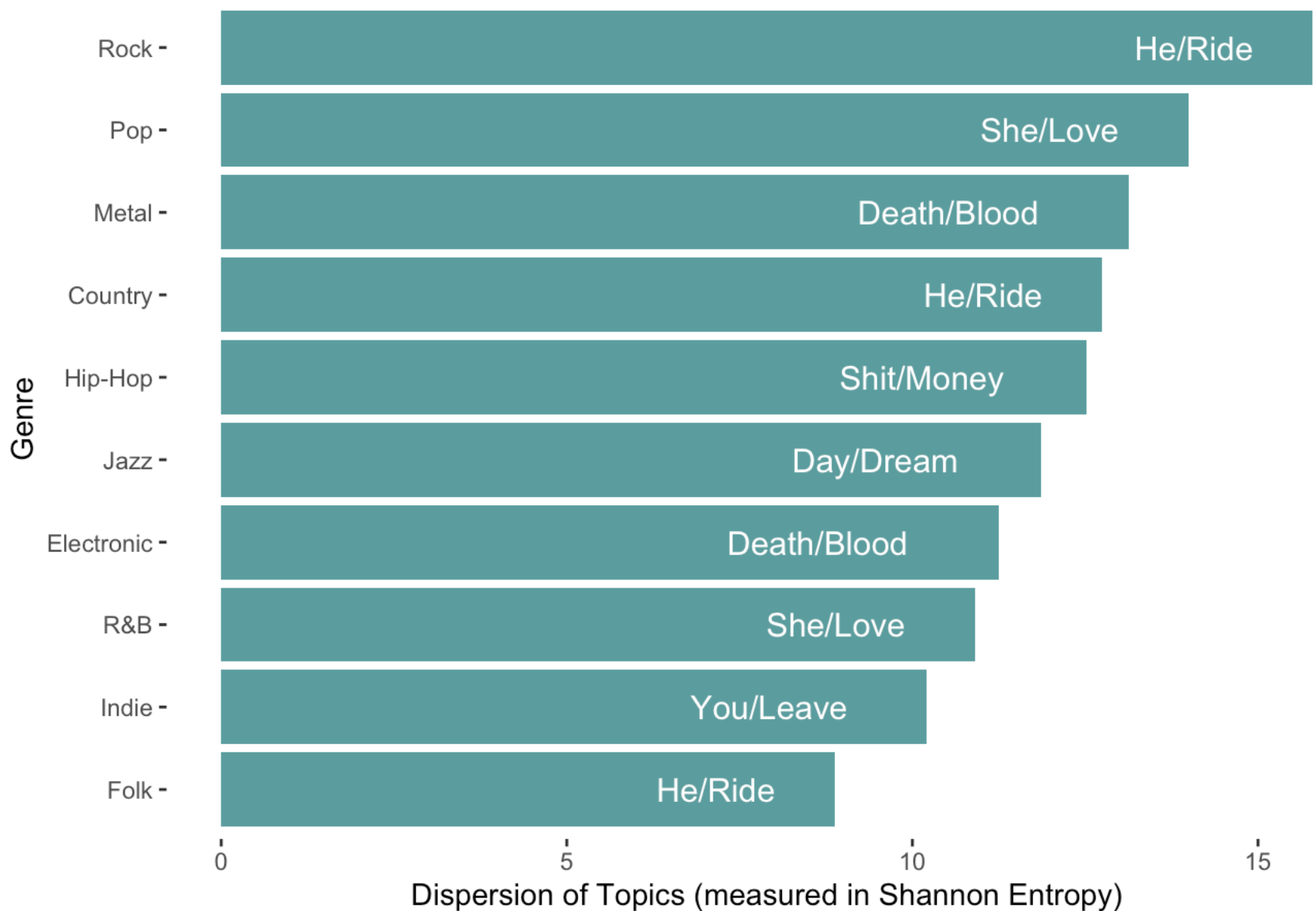
**Topics** are *factors*, or categorical data, making variance/standard deviation useless in this case. After a bit research, I chose **Shannon Entropy** ([https://en.wiktionary.org/wiki/Shannon\\_entropy](https://en.wiktionary.org/wiki/Shannon_entropy)) as the measure of dispersion, given as:  $H = - \sum_i^n p_i \log_b p_i$  where  $p_i$  is the probability of character number  $i$  appearing in the stream of characters of the message.

## Results - All the Eye Candies!

*Plot 1: What Is the Most Diverse Genre?*

Code

## Diversity of Content for 10 Music Genres



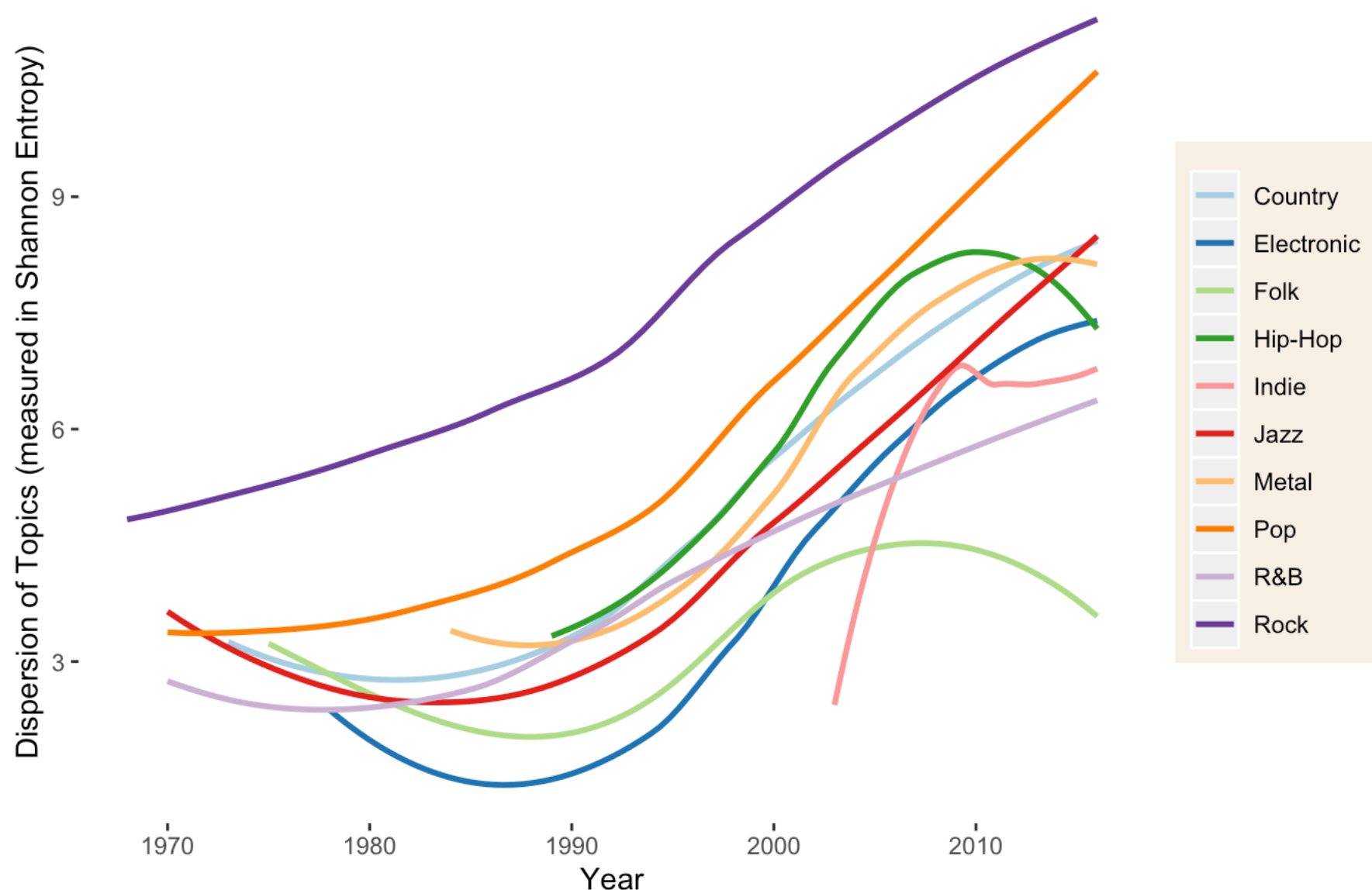
From the plot above, we see

- **Rock Music** wins the champion for diversity in music content; **Pop Music** comes next and **Metal Music** third.
- **Rock Music** most likely sings topic of “He/Ride”.
- The most frequent topic in each genre is labelled in white on its bar, and they are very likely to be different from others.
- The most frequent topic in each genre matched my impression (or sterotype lol) about each genre. Do they match yours?

*Plot 2: What is the Trend?*

Code

# Historical Trend of Diversity of Content for 10 Music Genres



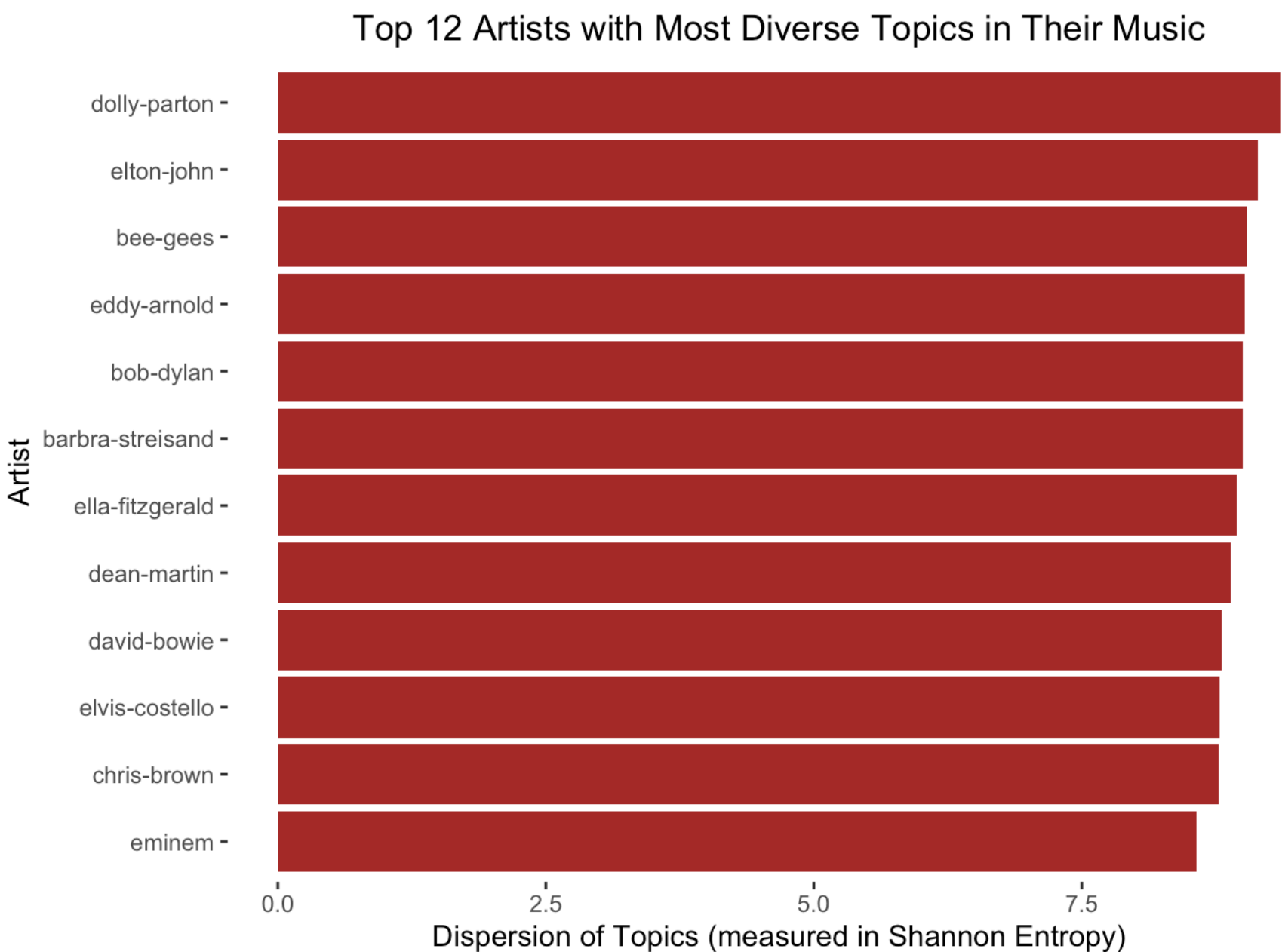
From this plot (the most exciting one in my opinion), we see

- Almost ALL music genres are getting more and more inclusive in terms of content!
- However, only **Hip-Hop Music** and **Folk Music** have a trend of decreasing diversity!

One might argue that this is actually due to the *word choice* in these 2 types of music do not really change, but not really because they tend to become less inclusive in terms of content/topics covered. I think it goes beyond the domain of science and enters what-we-call art. It really depends on what you believe and how you interpret, and that is exactly what makes this plot really exciting!

***Plot 3: Who Is the Most Diverse Artist in the History?***

Code

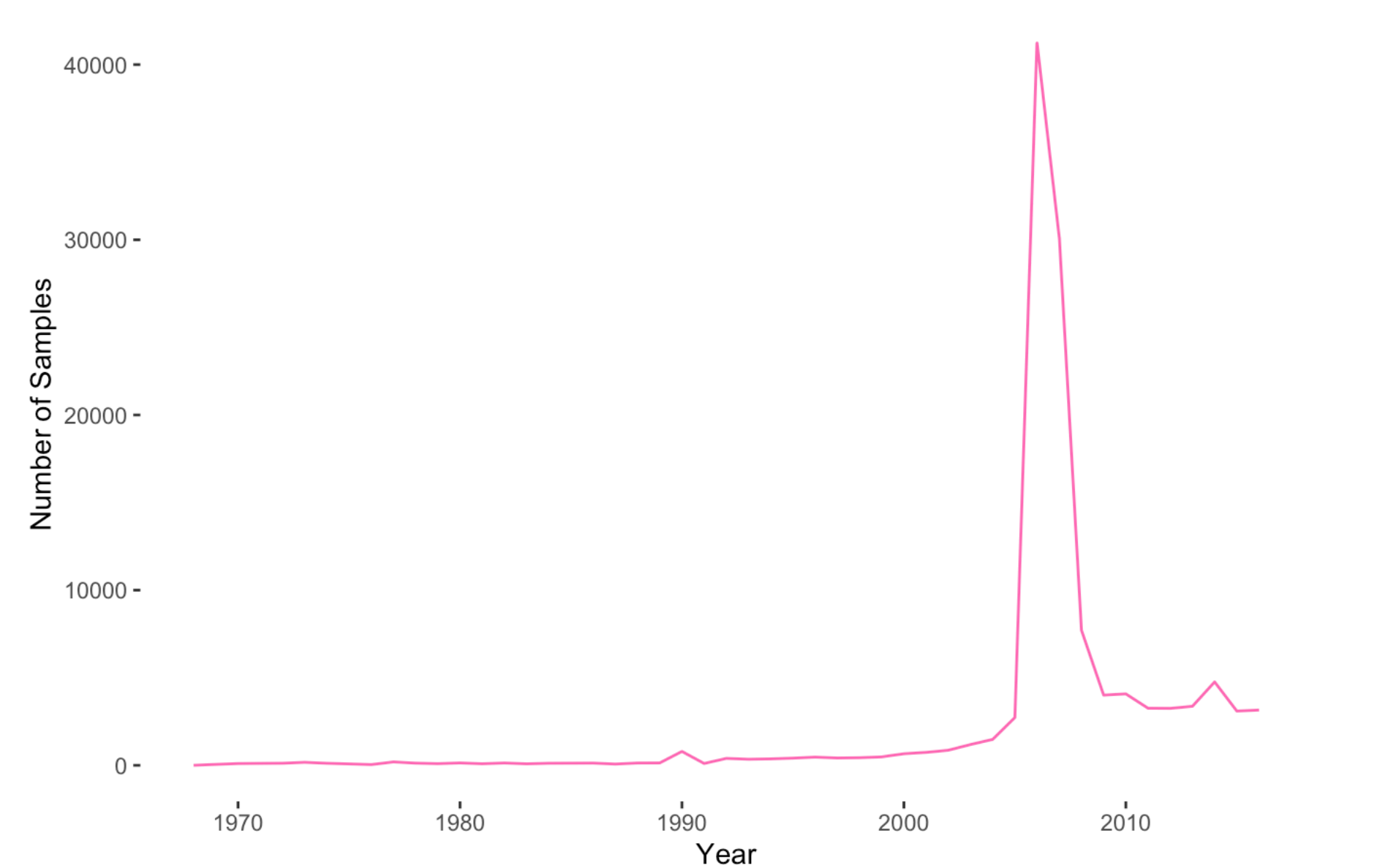


Since Shannon Entropy works meaningfully only when there are enough samples, I considered only artists that have released more than **500** songs in this competition. Thanks for the dataset, we have plenty of artists that meet this requirement. The top 12 artists are listed above. A few that I love the most, ***Elton John***, ***Bob Dylan*** and ***Eminem***!

***Plot 4: Caution!***

Code

Number of Samples Across Years



This is not the most aesthetically pleasing plot, I know, but a very important one. It helps us have a sense of **confidence** of what we found. Although the dataset contains a decent amount of samples anytime from 1970 to 2019, we need to be aware that we have implicitly put more weights on the more recent results and to be more precisely, results from 2005 to 2010, simply due to the more samples during these years.

## Conclusion

Music is getting more and more Diverse. We have plenty of great artists that are making music that cares about different groups of audience. This is not the end of world. Work hard and go get a job.

Processing math: 100%