The worried parent paradox -
an investigation of profanity in music

Morgan Heggland & Kjartan Krange
05.08.2020

Many fear society is becoming more extreme and explicit. Parents worry kids play too many violent video games and listen to rappers singing about drugs and crime. But is the music released nowadays that much worse than what older generations played during their own glory days? How has the profanity in popular genres evolved over the years?

In order to investigate this, we have used the Million Song Dataset, which contains metadata about 1 million songs. The songs were selected on the basis that they were added to musicbrainz.org by 2011, a website described as "the open music encyclopedia". Thus, the dataset contains a wide variety of songs spread throughout different genres and years (1920s to 2011). Linking this set with other data sources we base our conclusions on over 900.000 data points. We have used the Facebook bad words list - a 1700+ word list used by Facebook to filter bad words - to identify profane words in song lyrics.

The increasing vulgarity

So, what does our dataset say about profanity in music over the years?
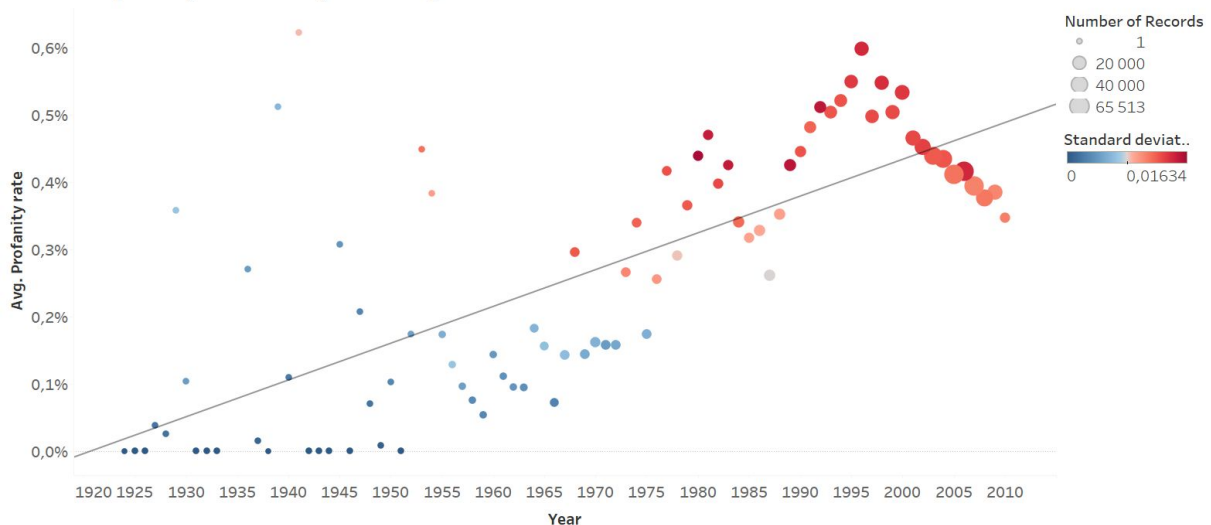

Figure 1: Average profanity rate across all songs for each year of the dataset. The profanity rate is calculated as the number of profane words divided by the total number of words in the song. The straight line represents the linear regression trend line of the datapoints. The sizes of the data points are determined by the number of songs released that year, and the color gradient is determined by the standard deviation. 100 842 songs without year data have been excluded. N = 804 751

Figure 1 shows a clear upwards trend since the late 50s. It seems the mothers' concerns are somewhat justified after all. This trend is also evident by the slope of the regression line. While this may indicate that music has become more profane over time, some of the trend may be explained by the reduced

barriers to releasing songs as the years go by, as seen by the increasing number of records. Another factor could be that the words considered profane by the Facebook dataset does not capture the profanity of years gone by, however we believe this effect to be rather insignificant. In 1995, when profanity was at its peak, if you were to turn on your radio to the average music station, every 166th word would be a cuss word. Since then, the last 15 years of lyrics has become less explicit. Profanity is down to the levels seen in the 1980s, the time where many current parents of teenagers themselves were teenagers. That's something to tell the nagging parents!

Some genres are less explicit than others

So far we have examined songs in general, but surely various music genres should be explicit to different extents. So how does profanity vary within genres? We selected a number of popular genres to investigate:
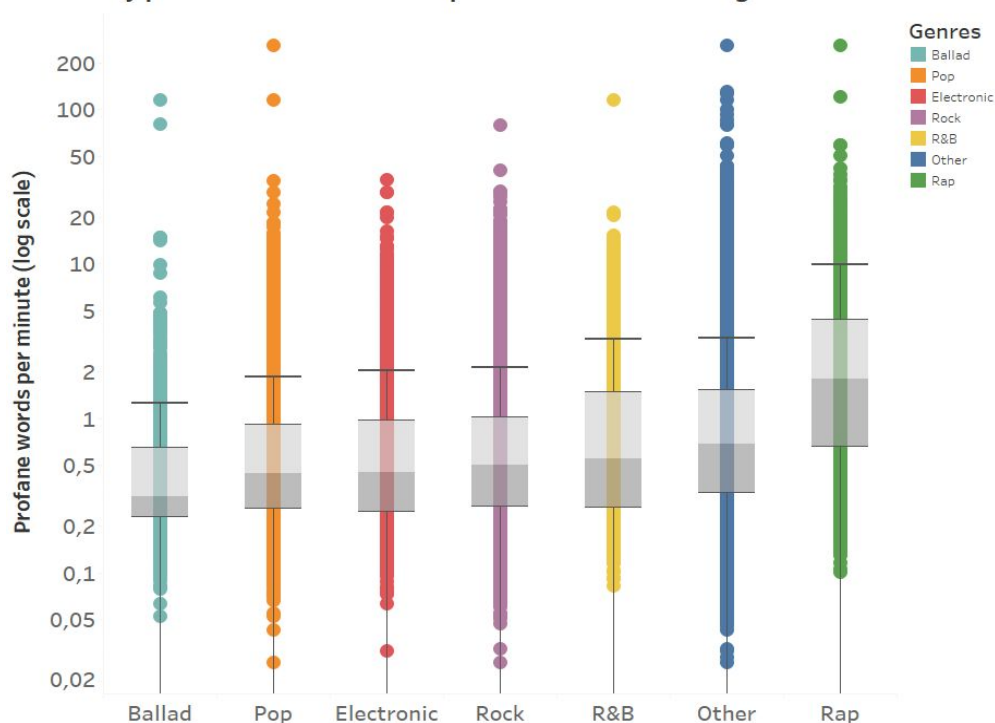

Figure 2: Profane words per minute in songs for selected genres. The median is represented by the transition from dark to light grey in the boxes. The boxes represent the interquartile range, meaning it represents the middle 50% of the data, from the 25th to 75th percentile. The whiskers extend to 1.5 times the interquartile range. The Y-axis uses a logarithmic scale to emphasize the interquartile range and account for large outliers. Songs with no profane words are not shown, but are included in the calculations with a value of zero profane words per minute. The "other" genre represents all other genres in the data. N = 905 593.

As the worried parent might expect, we see swearing is most frequent in rap songs. This is indicated by the high median combined with the high value for the upper whisker. Perhaps more surprising, R&B is another explicit genre, even more profane on average than rock, however less so than rap. On the contrary, many might expect pop to be almost as "clean" as ballads. However, pop music seems to me more diverse, and ballads are consistently cleaner. While calculating the profanity rate per minute

allows us to standardize the songs for comparison, this approach might be biased towards songs with more words in them, such as for example rap songs.

The odd one out

Has rap always been much more explicit than other genres?

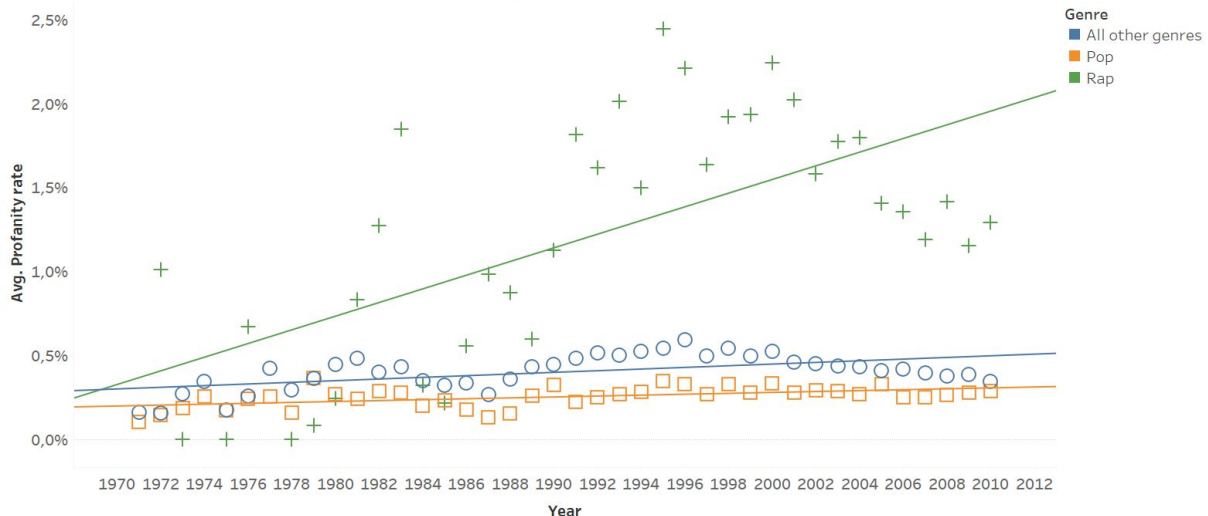**How frequent are profane words in rap and pop songs compared to other genres?**

Figure 3: Average profanity rate per year broken down by genres. The lines represent the linear regression lines for each group. All songs released before 1971 (when rap music originated) and songs without year data have been excluded. N = 771049

It is evident that rap, since its origin in the early 70s, has been more profane than the rest of the genres. We see that profanity was at its peak in rap in the 90s. In the early days, rap songs appear to have been less profane, something hip-hop-heads may suspect. While one might question the fit of a linear regression line in this case, it is still valuable for examining the average change over time without overfitting. Relatively speaking, rap looks to be about 3-4 times as profane as other branches of music. While this is only an increase from .5% to around 2 %, every 400th word is a bad one in the average genre, and every 50-67th word is a curse word in rap. On the other hand, profanity in pop has remained steady and below the average of the other genres combined. Furthermore, it seems that changes in profanity in rap has driven the overall changes seen since the 70s, as the averages of all other genres have been rather stable compared to rap. Thus, the recent trend towards cleaner rap music may explain the development seen in Figure 1 where profanity today has decreased to levels seen in the 80s.

Profanity overall has in recent times decreased to the levels of the 80s. Rap is quite explicit, nevertheless only about one word of every hundred spoken in rap today is profane. If parents are still concerned about their children's purity, the data seems to suggest they should only let them listen to cleaner than average genres like pop, or even better yet: Ballads only.

Sources:
- Million song dataset
  - Metadata link: http://millionsongdataset.com/
  - Musixmatch lyrics: http://millionsongdataset.com/musixmatch/

- - Lastfm genre data: http://millionsongdataset.com/lastfm/
- Facebook list of bad words:
  https://www.freewebheaders.com/bad-words-list-and-page-moderation-words-list-for-facebook/
- Data retrieval and cleanup notebook:
  https://github.com/MorganFH/CS10-final-project/blob/master/Data_retrieval_and_cleanup.ipynb