
Evaluating Patterns in Critically Acclaimed Music

Paul Bartlett (# 250753008)
The University Of Western Ontario

PBARTLE7@UWO.CA

Abstract

The purpose of this analysis is to identify relationships between musical genre of critically acclaimed albums and time. The dataset used for this analysis contains over 18,000 reviews from Pitchfork from January 5th, 1999 to January 8th, 2017. It contains important data including release year, artist name, genre, and a score ranging from 0.0-10.0. The findings may be useful for determining what the most successful genre of critically acclaimed music is for each of the last 18 years and what is going to be the most successful in the future.

1. Description of Applied Problem

1.1. Existing solutions to similar problems

The trends of popular music can easily be attained through the various Billboard charts that have existed since 1955. A group of scientists from the University of London analyzed around 17,000 songs that charted on the U.S. Billboard Hot 100 over the last 50 years and created a visualization of the popularity of musical genres over time (Matthias Mauch and Leroi, 2015). The problem with getting data from these charts is that popular music generally isn't critically acclaimed, and is therefore not as interesting as data from sources that evaluate music more objectively. Another source that uses visualization of this problem well is musicmap (Crauwels, 2016). The website contains information about hundreds of genres of music and their history. It provides a great overview of all the popular strands of music, but doesn't go into too much depth about specific artists or albums. It does provide a good overview of all genres regardless of popularity, but I'm more interested in evaluating the history of genres from the best albums created by artists.

Project report for CS4437/CS9637: Intro to Data Science.
University of Western Ontario, Winter 2017.

1.2. Pitchfork solution

As various breakthroughs in music happen, there is generally a shift in the type of genres that become popular. Artists get influenced by other talented artists and adapt part of their style into their own music. In addition, a good Pitchfork review can have a significant impact on an album's popularity. This is very important for independent artists because they don't have the resources or backing of a large label to get their name out there. Using a dataset that includes over 18,000 reviews from Pitchfork, I will be going through the data to find how critically acclaimed music has changed over time. In addition, I will be looking into applications of machine learning for predicting an album's score to find out what influences this.

2. Description of Available Data

2.1. Pitchfork

The dataset that I will be using is taken from Pitchfork. Pitchfork is an online magazine that focuses on reviewing both popular and independent music. It is one of the most popular platforms for users interested in finding higher quality music. The data set for Pitchfork Reviews from January 5th, 1999 to January 8th, 2017 is available on kaggle (Conaway, 2017). There are 18,393 reviews that include important data including release year, artist name, genre, and a score ranging from 0.0-10.0. Considering that Pitchfork is one of the longest running online review sites, it makes it a primary choice for useful data. There may be some bias in review scores, notably staff preference, but Pitchfork does cover a lot of genres with ratings similar to many other music review platforms. Through looking at 18 years of data, we should be able to find some notable trends.

3. Analysis

3.1. Genre

To analyze this data properly, we first clean the data to prepare it for processing. By running a summary

of our dataset, we can see all 22,690 records there, with scores in between 0 and 10 with a mean of 6.991 and published year between 1999 and 2017. To check for empty fields in the data we run a missmap and see that genre has many fields that are empty, so we replace every instance with “none” for easier data processing and better-looking labels. Running unique on genre in the dataset shows that there are 10 different genres we are working with: electronic, metal, rock, none, rap, experimental, pop/r&b, folk/country, jazz, global. These genres are fairly broad and hopefully will prevent some genres from having too little review data to work with. To check this, we make a frequency table shown in figure 1 of the number of reviews per genre. For some of the earlier years where there are less reviews, it is harder to see the trends, so we make a proportional table shown in figure 2 to also highlight how each the proportion of each genre changes over time. From this, we can see that genres such as global, folk/country and jazz have a much lower proportion than the other genres. Another observation of interest from this table is how the proportion of “none” increases in the middle of 2010-2015 and then rapidly reduces in 2015. Although the difference isn’t too great in comparison to other years, it still takes away from properly classifying almost 20% of the reviews for a couple years.

The next part of the data we want to check is the consistency of the review scores. If the median of the score for a year is significantly different than other years, it means that their process of reviewing isn’t normalized and it will influence the patterns we see each genre over time. In figure 3 we have a boxplot of scores for each year. From this, we can see that the median of the scores from each year are approximately the same, at around 7.0. This means that we shouldn’t have to normalize the scores for patterns and we can move on. Interestingly, there are almost no extremely low score outliers after the year 2009. This could be because reviews with a score lower than 2.0 are considered to be uninteresting to readers, or that the staff thought that it was unnecessary to waste time writing reviews for albums that they strongly disliked.

Figure 4 shows a boxplot of the scores for each genre that was reviewed. Once again, the medians of the boxplots appear to be approximately equal. Interestingly, the boxplot with the highest median comes from the global genre, which was one of the least reviewed genres. This could be because the staff don’t review something from the global genre unless they know it’s going to be at least somewhat decent, or that the staff don’t listen to enough global music to give out poor reviews. We can see that there is only one outlier for

the global genre, and it’s still above 2.0 when most of the other genres have outliers that are below 0, but this could just be because we don’t have enough review data for the genre. From this table, we can also see that the “none” genre appears to also be approximately the same as the other genres. This is important to identify because there can be many reasons a review may not have a genre tag. In the worst case, a reviewer might not bother to put a tag for poor reviews because reviews for higher rated albums will generally have priority for bringing users to their site. We can see that this is most likely not the case, which is great because the data that’s missing seems random, rather than missing data about low scoring albums.

After verifying that the data are reliable enough to do some analysis, we can finally apply some regression. In figure 5 we have a table of the mean scores throughout the year with LOESS smoothing as a way of seeing the relationship between each genre and time (Burgoine, 2017). We can see that most genres are noticeably increasing in 2016 and 2017. The only genre that seems to have a poor outlook is rap, with none and pop/r&b also noticeably lower than the rest of the genres. This result is not too surprising, because rap has become a lot more popular in the last couple years and there are now a lot of artists that produce low quality music in an attempt to easily make money. From looking at the proportions of each genre in the previous figure, we can assume that the genres with a higher proportion are generally more popular on the site. From this we can guess that if an independent artist wanted to make a successful album, they would choose for it to be rock or electronic. However, choosing other genres like experimental and metal that are also increasing would also have a greater chance of creating a successful album.

3.2. Score Prediction

For some extra analysis, I wanted to apply machine learning methods to my data to create a predictor of score. To do this, we start by choosing the features that seem the most beneficial in our model. We then create a list of 80% of the rows in the original dataset for training and testing the models, and select 20% of the data for validation. We split the input and the output and run a scatterplot matrix shown in figure 6. Unfortunately, we do not have many features to work with, but we still choose best new music, publication year, and score to check if there are any patterns. From this we can see that there isn’t really anything that is beneficial, other than score and best new music. Since best new music is a Boolean value that is generally only given 1 when above a certain score, it is a good

predictor for deciding if an album is above or below around a score of 8. The problem with this is that several of the earlier years don't have a single best new music album, meaning that it was likely a feature that was added by the site in the early 2000's. Since there aren't enough other features for our model, it is unlikely that it will be successful. I attempted to run a random forest on the dataset using 10-fold cross validation, but the results were as poor as expected. After this I considered adding my own features, and one that I thought would be highly beneficial would be album release number. Artists tend to have worse releases after their first 2-3, so using this information would help predict score. Unfortunately, the dataset doesn't include a way to get this data because they don't review every album from an artist.

4. Results

The best prediction model we could get from the dataset was the regression of mean scores of each year with LOESS smoothing shown in figure 5. From this, we were able to see some clear patterns in the genres and make the conclusion that rock and electronic music would be the most likely to be successful for independent artists. We could also conclude that their methods of reviewing were fairly normalized for each year and each genre overall. Although we could not successfully apply any other methods of machine learning to our dataset, we were able to identify the problems with our dataset, specifically the lack of quality features, that prevented us from creating a successful model. Overall, the findings were still useful, and if we were given access to more useful features we could be more successful in creating a successful predictor of album scores on Pitchfork.

References

- Paul Burgoine. Pitchforks favourite genre by year, 2017. URL <https://www.kaggle.com/burgoine/pitchforks-favourite-genre-by-year>.
- Nolan Conaway. 18,393 pitchfork reviews, 2017. URL <https://www.kaggle.com/nolanbconaway/pitchfork-data>.
- Kwinten Crauwels. musicmap, 2016. URL <https://musicmap.info/>.
- Mark Levy Matthias Mauch, Robert M. MacCallum and Armand M. Leroi. The evolution of popular music: Usa 19602010. 2015. URL <http://pubmedcentralcanada.ca/pmcc/articles/PMC4453253/>.

Figures

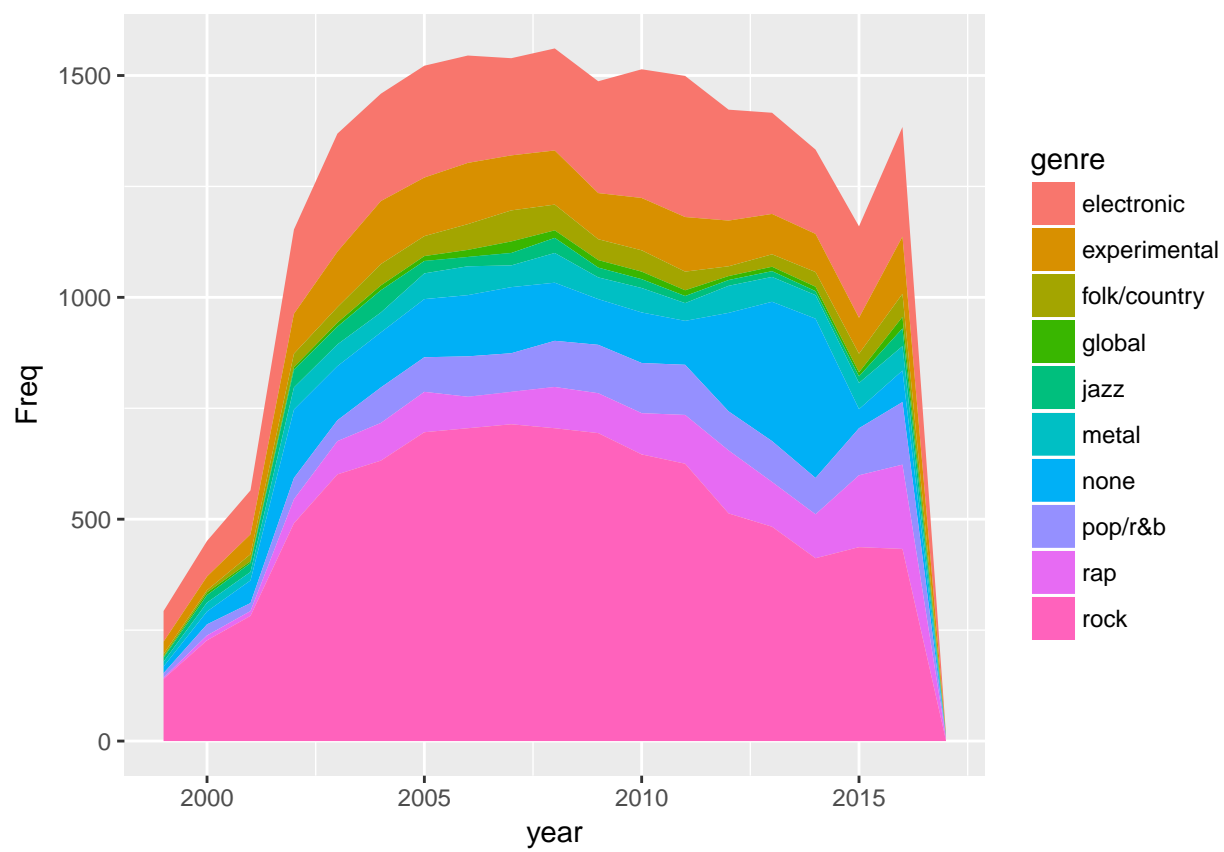


Figure 1: Frequency table of genre reviews per year

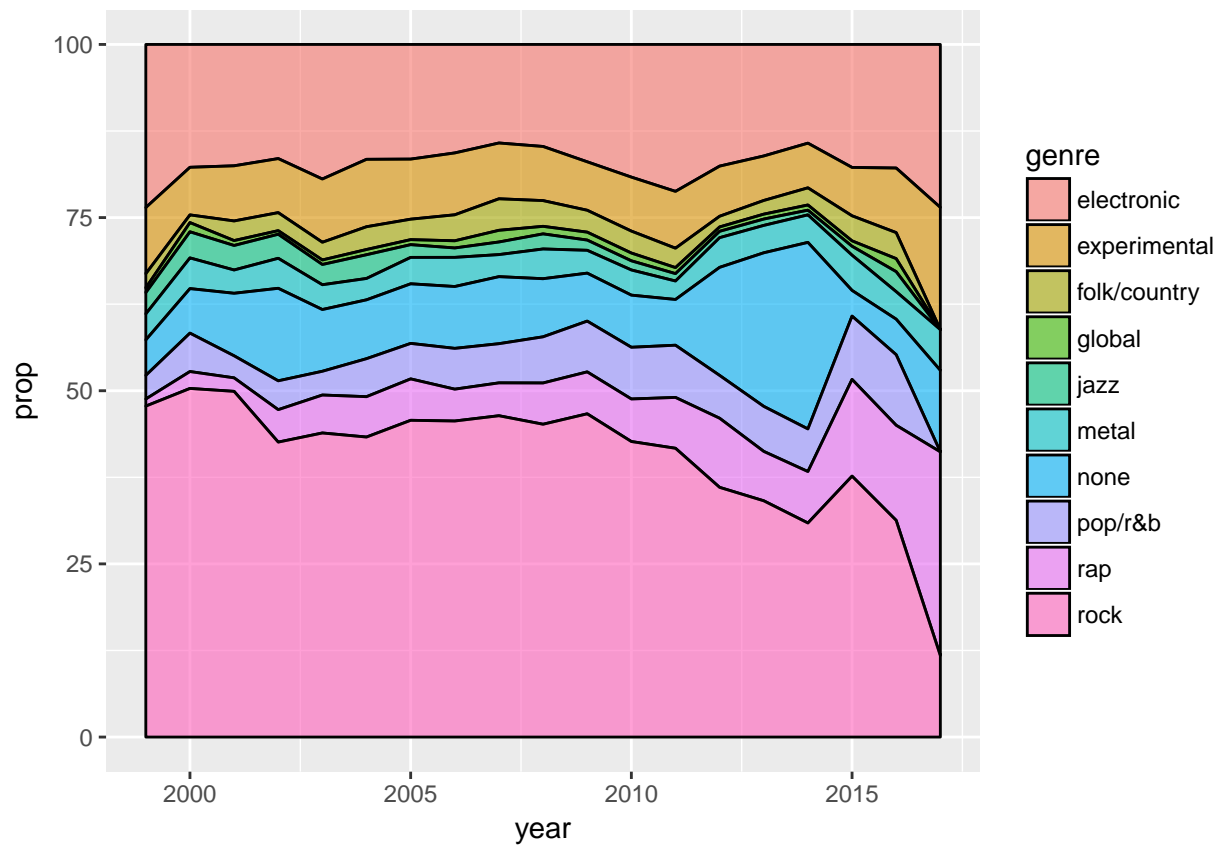


Figure 2: Proportional table of genre reviews per year

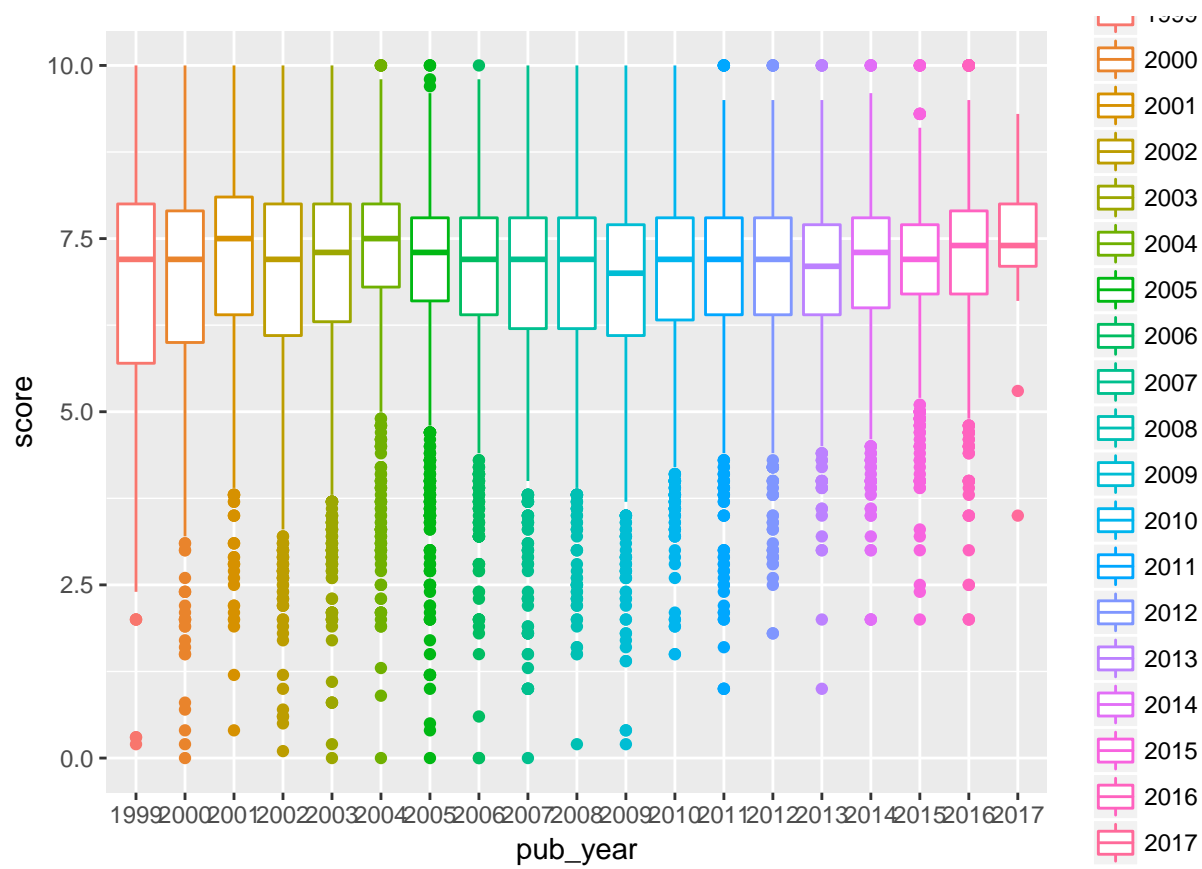


Figure 3: Boxplots of score by year

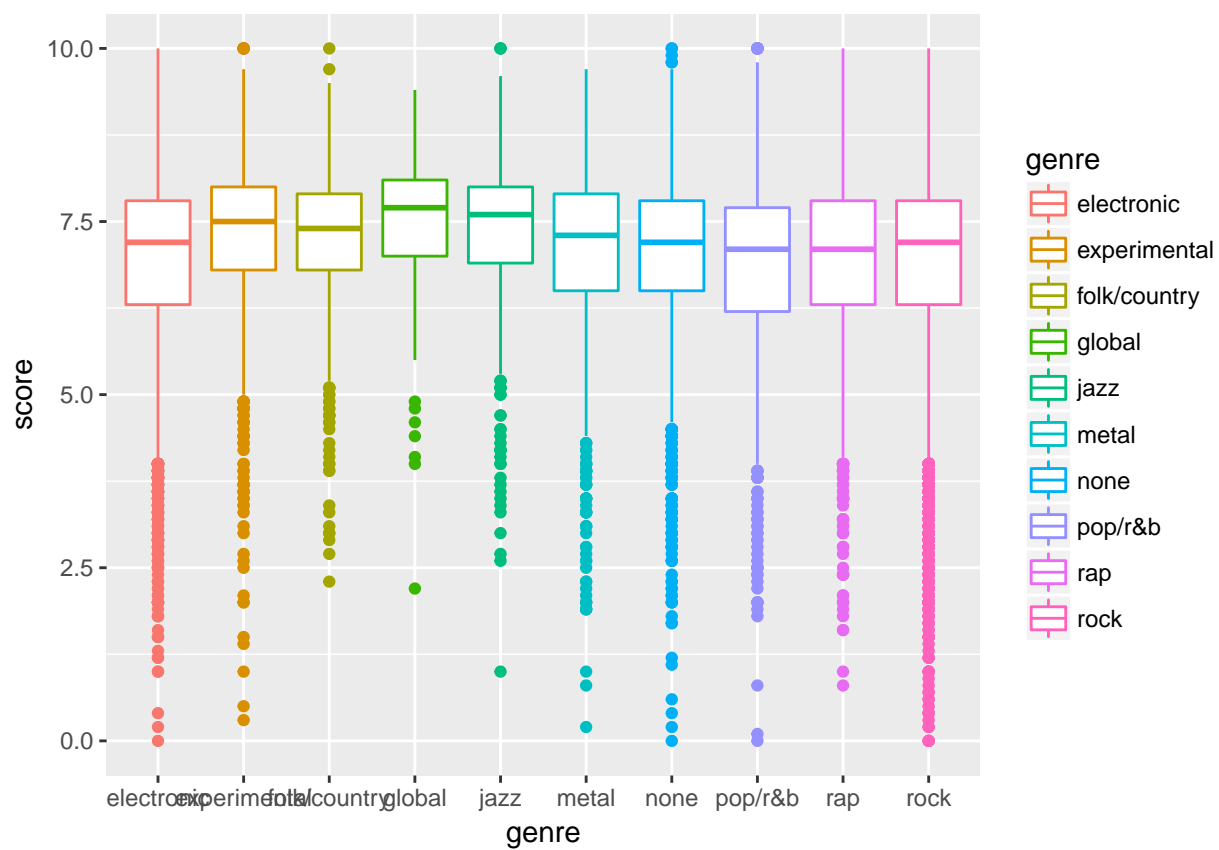


Figure 4: Boxplots of score by genre

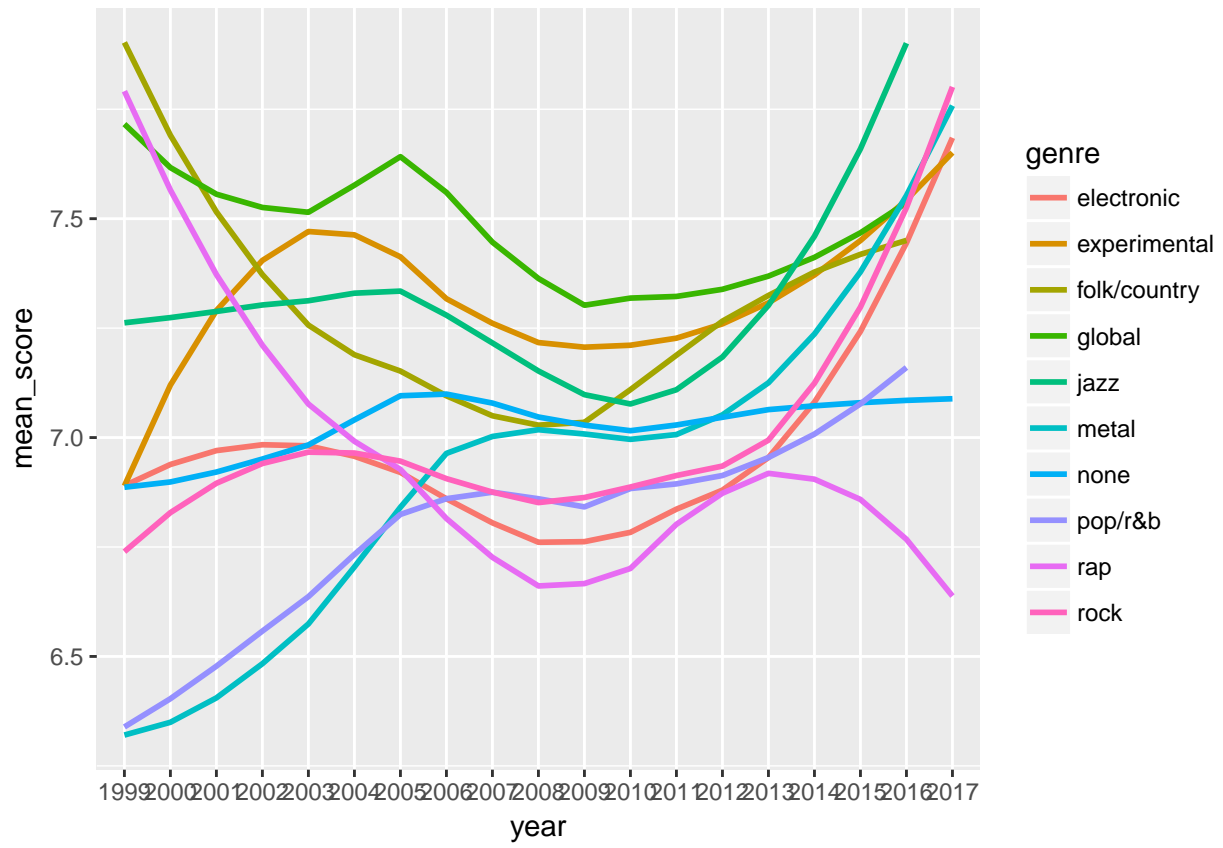


Figure 5: Regression of mean scores of each year with LOESS smoothing

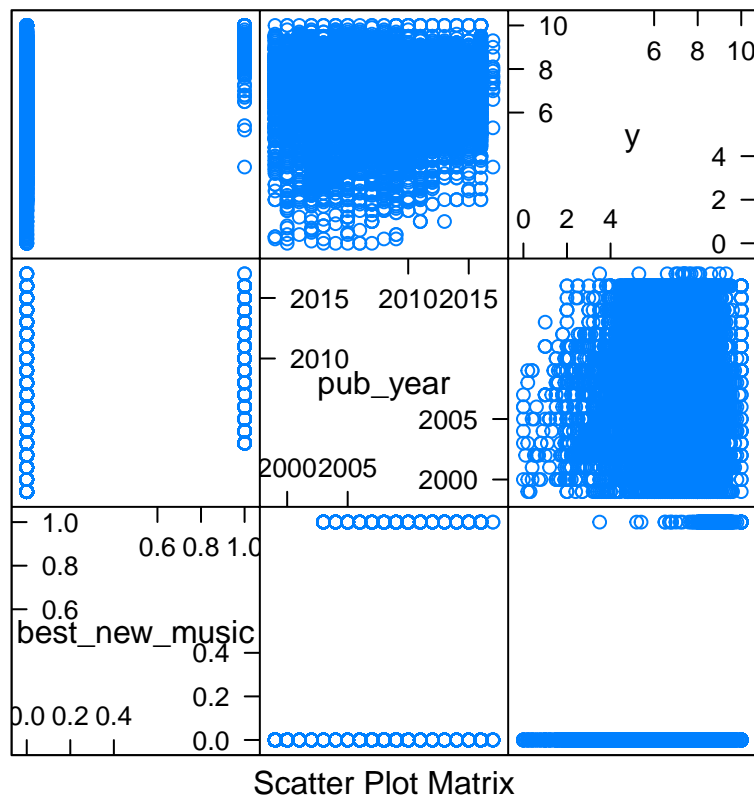


Figure 6: Scatterplot matrix of features