Evaluating Patterns in Critically Acclaimed Music

Paul Bartlett (*# 250753008*) PBARTLE7@UWO.CA
The University Of Western Ontario

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.

```
## Warning: running command 'ls ../input' had status 127
```

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, 2015a). 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.

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 inpact on an album's popularity. This is very important for independant 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, 2017a). 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.

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

3. Analysis and Visualization

3.1. Genre

To analyze this data properly, we must use a method to take out only the best reviews from the dataset. Fortunately, Pitchfork has a system to distinguish the best albums called "Best New Music". Albums that receive this tag are guaranteed to be of higher quality and therefore valid for our analysis. Unfortunately, this feature launched in 2003 so using it would leave out all the music before it was launched. By looking at the data, we will be able to find the typical rating for an album that gets the "Best New Music" tag and use that rating to take all albums from the data set that are higher than the threshold. From this, we should be able to classify each album that meets the requirement by year and genre so that it can be used for visualization.

3.1.1. Visualization

For visualization, I would like to do something similar to what was done for visualizing the U.S. Billboard Hot 100 over the last 50 years (Matthias Mauch and Leroi, 2015b). The chart features spindles for each genre that run vertically with the width of the spindles proportional to the frequency of the genre. The y-axis contains the year so the viewer can easily compare between each year to see what genre is the most popular or the least popular. Since I do not have much experience with visualization, it is possible that doing something similar will be too difficult to achieve. A simple way to visualize this in a similar way would be to use a line graph, with each line representing a genre, the x-axis covering each year, and the y-axis covering the frequency.

3.2. Release

3.2.1. Analysis

To analyze what release number is considered to be the best critically, we will have to query all artists that have multiple releases on the site. A similar analysis and visualization was done on kaggle by the author of the data (Conaway, 2017b). The author only covered the first and last album, but did complete an analysis on the number of reviews for each artist. After this, we need to get the review scores of each album from each of the artists with multiple releases. There might be some bias in how release numbers are scored, mainly that a poor release could potentially prevent the staff from Pitchfork from reviewing later releases.

3.2.2. Visualization

A Box plot would be perfectly suited for this data. Being able to visualize the entire range of score for the releases while also seeing the median and upper and lower quartiles is highly beneficial. There would be a separate box plot for each release number up to the maximum reviewed releases by an artist.

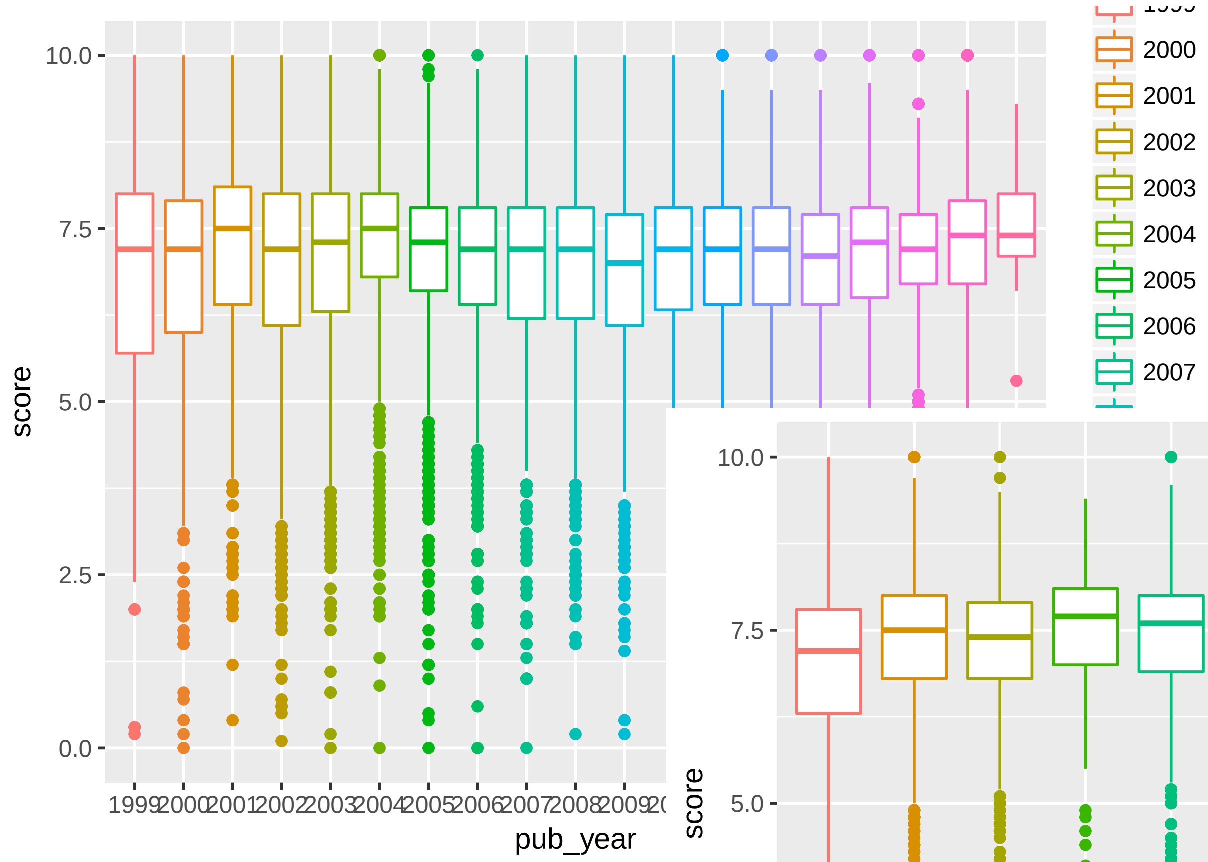

Figure 1. Boxplots of score by year

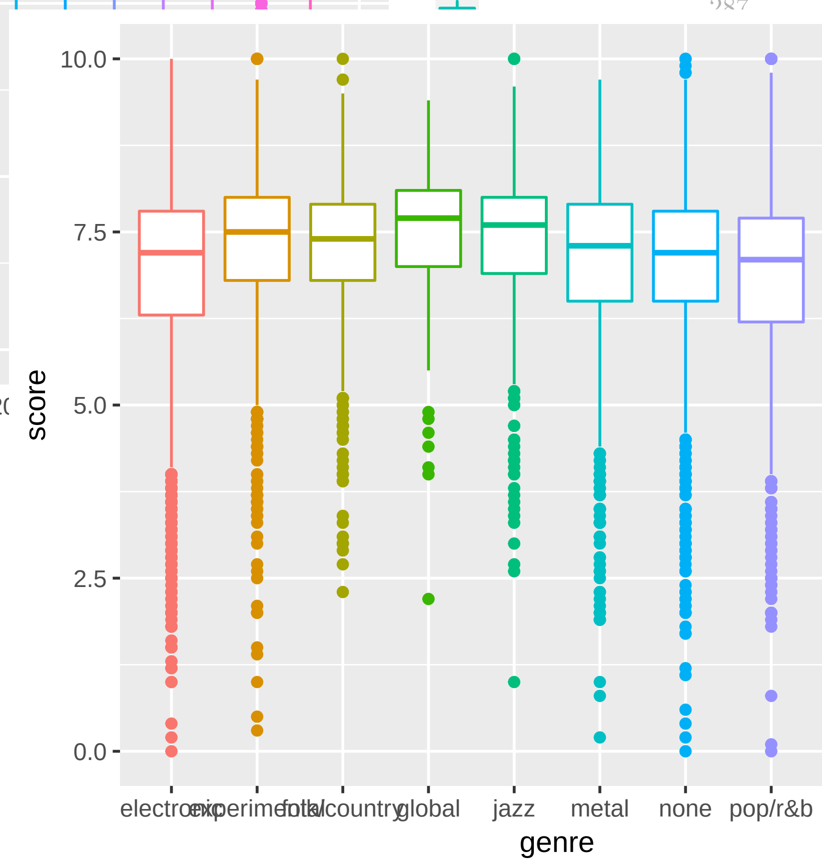

Figure 2. Boxplots of score by genre

References

Nolan Conaway. 18,393 pitchfork reviews, 2017a. URL https://www.kaggle.com/nolanbconaway/pitchfork-data.

Nolan Conaway. Is the first album the best?, 2017b. URL https://www.kaggle.com/nolanbconaway/is-the-first-album-the-best.

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. 2015a. URL http://pubmedcentralcanada.ca/pmcc/articles/PMC4453253/.

Mark Levy Matthias Mauch, Robert M. MacCallum and Armand M. Leroi. Figure 3., 2015b. URL http://pubmedcentralcanada.ca/pmcc/articles/PMC4453253/figure/RSOS150081F3/.

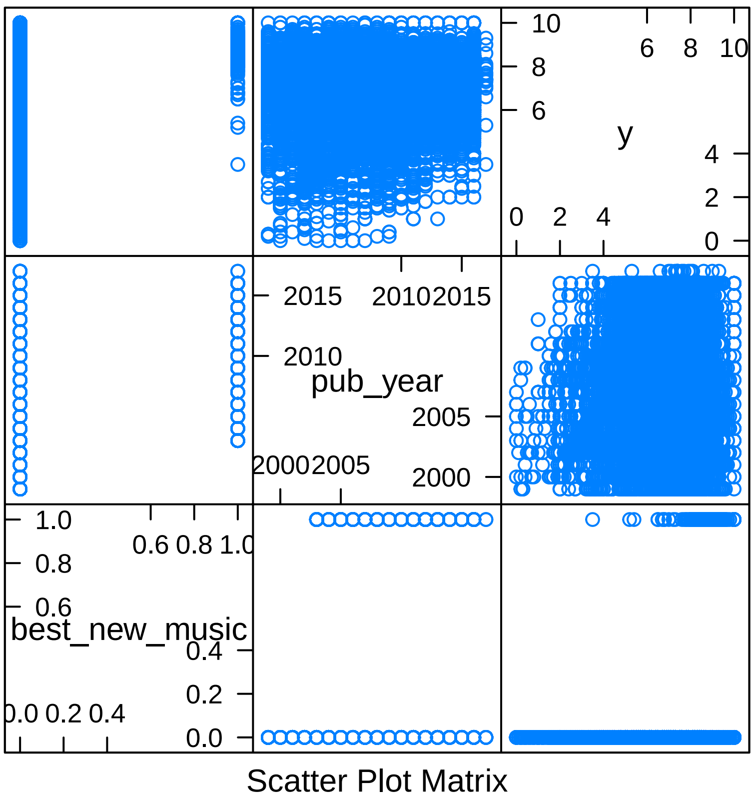

Figure 3. Scatterplot matrix of features