Pitchfork Data Analysis 

Data Acquisition:

All data was acquired from pitchfork.com, a popular music review website based in Chicago, Illinois. Using the R libraries rvest, XML, and magritr, as well as the Google Chrome extension SelectorGadget, reviews dating back to the website’s inception in 1999 were scraped. The information scraped from these reviews are as follows: the name of the artist, the name of the album, the album’s score, the name of the reviewer, the record label, the date of the review, and the genre. The scraping file can all be found here: <https://github.com/mikehalverson11/Pitchfork/blob/master/Scraping.R>. A csv containing a complete (up to June 10th, 2017) can be found here: <https://github.com/mikehalverson11/Pitchfork/blob/master/Pitchfork2.csv>.

Data Cleaning:

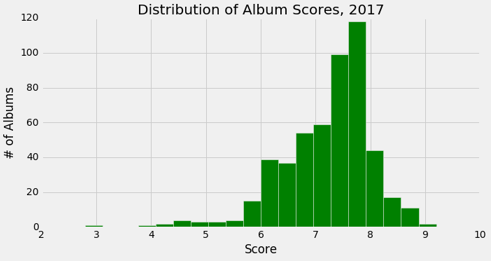
The data, as presented by the Pitchfork website, had several issues that required fixing in it. Among them was reviews of classic albums being listed as under the year they were review, not the year they were released. Other cleaning included marking reissued records as such, given that they do not represent the actual quality of the records released that year. Other manipulations including extracting the features for models, such as the previous scores for albums made by the artist, the amount of years between releases, and the genres the album lies under. The file for the data clean-up can be found here: <https://github.com/mikehalverson11/Pitchfork/blob/master/Dataslicing.ipynb>.

Summary Statistics:

2017 Summary:

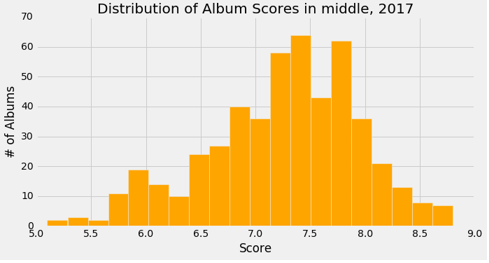
Plots for code and significance testing can be found here. Reviews are up to an include June 10th, 2017.

Many critics have lamented the lack of upper-echelon records to come out this year. While several stellar albums have been released, few of them have been seen by the press as instant classics as Kendrick Lamar’s *To Pimp A Butterfly* or Frank Ocean’s *Blonde* had been in year’s past.

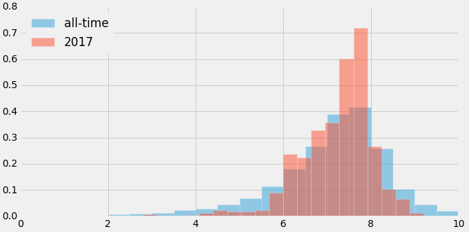


The distribution of album scores bears this out; only two albums cracked the 9.0 mark (Kendrick Lamar’s *Damn*, and Mount Eerie’s *A Crow Looked At Me)*. 2016, 2015, and 2014 all had a great deal over that mark; 2016 had 10, 2015 had 6, and 2014 had 6. There’s plenty of noteworthy releases to come (Arcade Fire being the most notable) of the bunch, but thus far 2017 has lacked the upper-echelon albums.

More interesting is the shrinking of the scale that Pitchfork uses. Of the ~500 albums released this year and reviewed by Pitchfork, two score above a 9 (as mentioned), and only 10 score below a 5. More damning, however, is the artists scoring below a 5, all low hanging fruit like Ed Sheeran, John Mayer, The Chainsmokers, and classic Pitchfork whipping boys like Joan of Arc and Alt-J. The most controversial ‘pans’ have been of Tory Lanez (not saying much) and Diet Cig, which somehow ignited a mini controversy of its own. Zooming in gives a clearer glimpse:

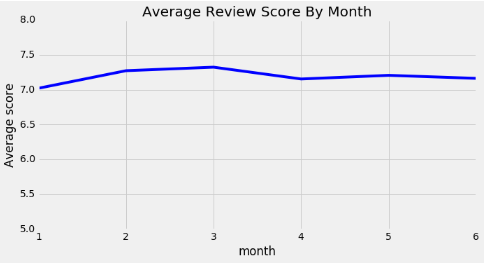


Almost everything falls between the high 6s and the high 7s; going below indicates an album is outright bad. Looking at how the publication has done in non-2017 years:



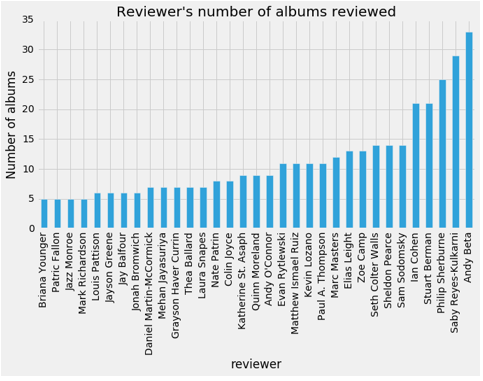
This all occurs despite the fact that the averages pre-2017 and 2017 are roughly identical (7.03 vs. 7.20). This difference is not insignificant, however; using Welch’s T test (due to the clearly unequal variances displayed by the above histogram), we found a very small p value (<10-6).

At least among music fans I’ve been listening to, there seems to be a perception that the year has picked up as time passed. There appears to be a slight trend indicating this:

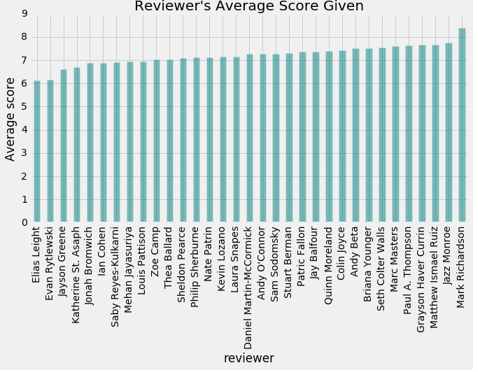


The difference is slight, and not significant at the .01 level (again, using Welch’s Test), but there does appear to some improvement as the year goes on. Only one of the year’s 20 highest rated albums was reviewed in the month of January (although Priests’ *Nothing Feels Natural* was released in January and reviewed in February), so there is some evidence to the upper-echelon records coming out later on in the year.

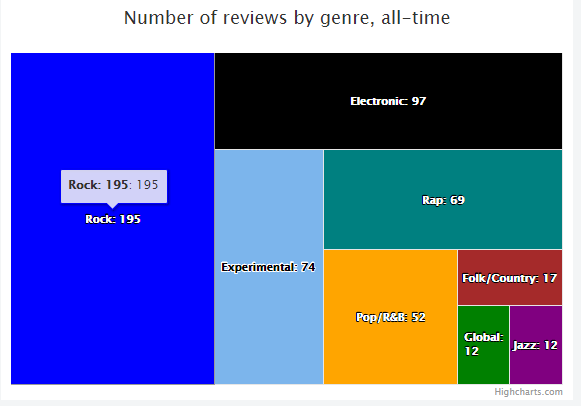
Let’s take a look at the reviewers:



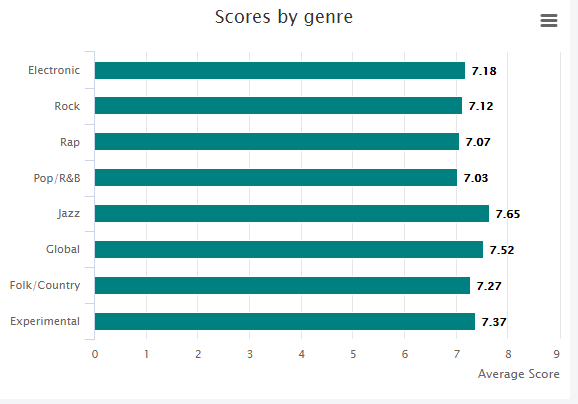
We see that relatively new personalities Andy Beta and Saby Reyes-Kulkarni have a relatively commanding lead in albums reviewed. Old stand-bys like Mark Richardson and Jayson Greene appear to have fallen off in quantity, for two writers who still stand in the website’s top 10 in all-time writing reviews. But how have they scored?



Richardson, perhaps because of his longevity, has by far the highest score, giving two “Best New Music” tags in his 5 reviews. One has to assume he picks and chooses what he reviews based on what he personally enjoys. Meanwhile Elias Leight and Evan Rytlewski stick out as outliers in the opposite direction. Leight, a newcomer, has yet to get a BNM review nod, while reviewing three of the year’s ‘worst’ albums (the two Tory Lanez albums and the Flume’s *Skin Companion EP II*). Rytlewski has been around far longer, but seems to just enjoy the negative review game a little more than most.



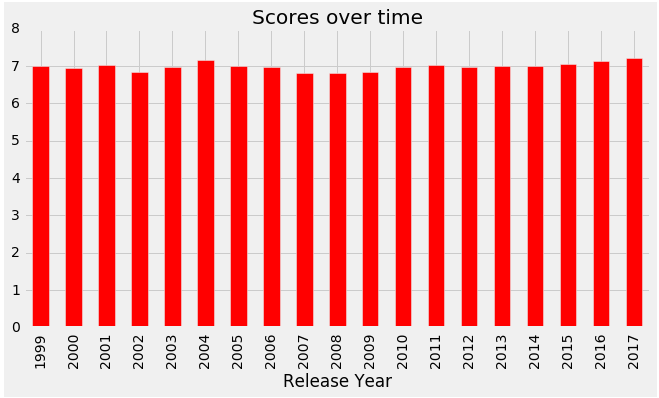
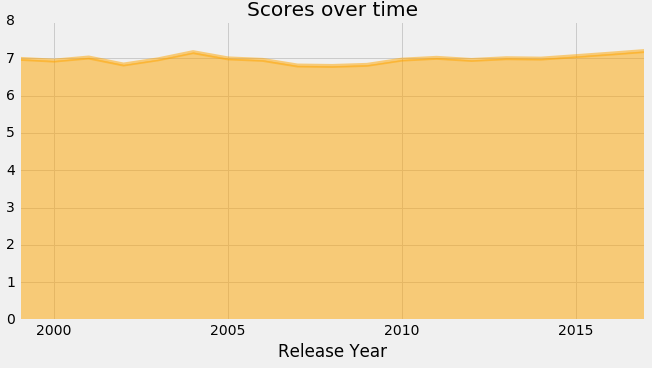
Genre wise, we see that rock is still Pitchfork’s dominant genre, but in-roads have been made by experimental, electronic, R&B, and rap to take up a sizable amount of the writing space. Folk takes up a shockingly small amount of the market here, at the same level as Pitchfork’s largely ignored country.



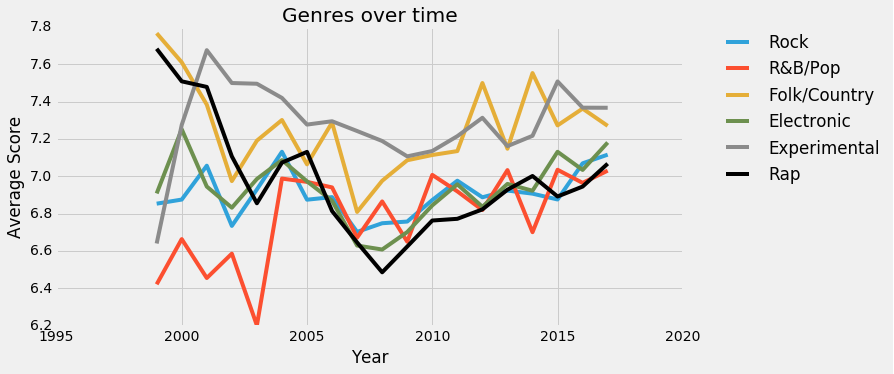
Little difference is seen here in terms of score. Sure jazz, global, folk, and country have higher scores, but that’s largely due to Pitchfork selecting albums based on them being good, whereas they’ll review pretty much any rock record that comes out. Experimental does surge forward, while rap and R&B do worse than rock and electronic, but the differences are not large.

Historical Summary:

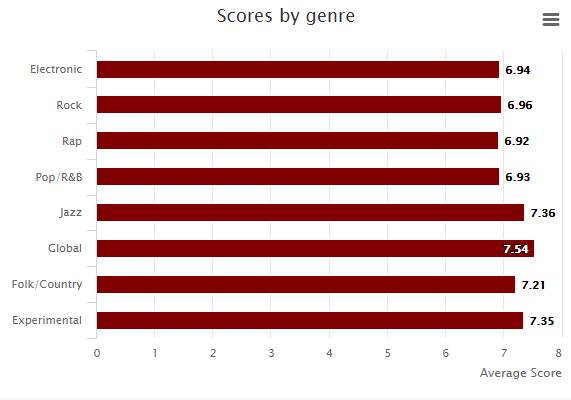
Now we go a little more ranging with our view, to the all-time look. First and foremost, we want to see how the scores have changed over time:

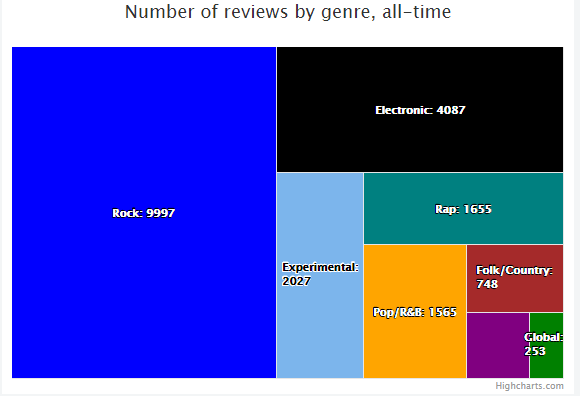


The change over time isn’t huge, that’s for sure, but there’s some interesting information to be gleaned here. For one, we see 2004 and 2011 stand out as the best years; while 2015-2017 have higher or comparable averages, it seems to be as much a function of time as the music itself. Also interesting is how Pitchfork’s cynicism/negativity has ebbed and flowed; at its inception, the site was not the hub of cynicism and negativity that it often portrayed as. As a matter of fact, it hit its peak in negativity in 2007-2009 before coming out of the hole at the turn of the decade. Perhaps more interesting is the genre breakdown:

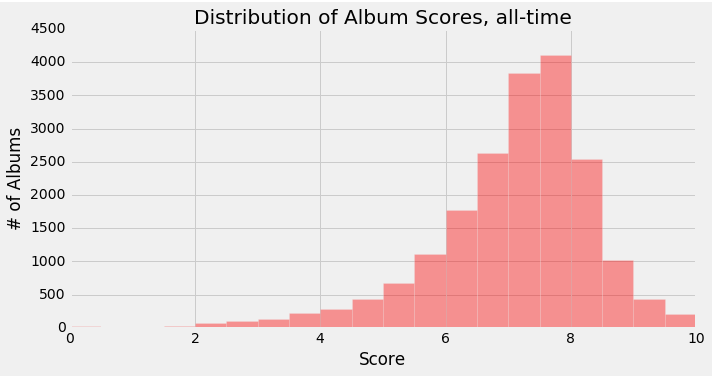


One can see any number of macro trends here. The decline of rap around the ringtone rap era of 2006-2009 before it’s recovery, Pitchfork increasing acceptance of pop music, the general increase of scores over time, and electronic’s following a similar phase system to rap. One thing to notice is that Pitchfork’s reception of rock has not really changed; the average score now is the same as it was at the site’s inception.

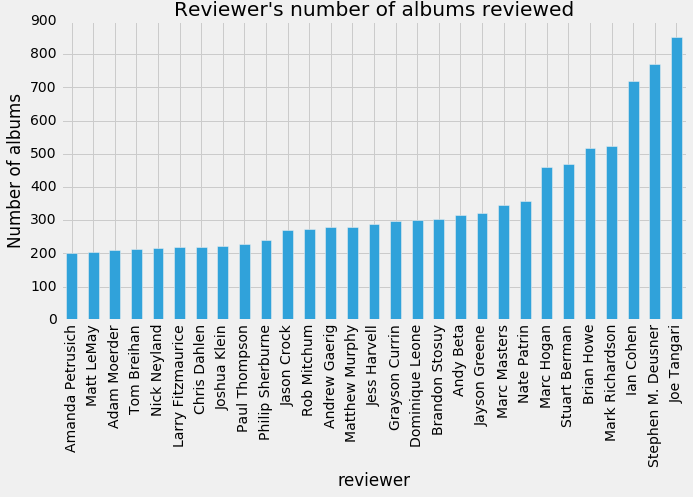




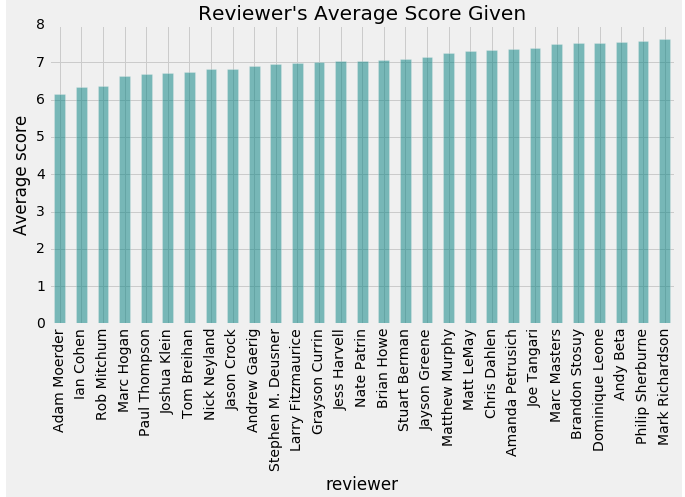
The global and jazz genres succeed largely due to the selection bias; jazz and global music don’t get reviewed unless they’re exceptional. Among the more reviewed albums, we see Pitchfork doesn’t give much preference, except for to experimental records. The count look bears out Pitchfork rock origins; over twice as many rock records have been reviewed as any other genre.



As we saw with the overlayed distribution in the 2017 section, the distribution is still centered in the 6-7 range, but it has much smoother and less abrupt tails. The 10s are more prominent, as are the sub 4 reviews.



Looking at the most reviews by reviewer, we see familiar names like Cohen and Richardson. Interestingly enough, it’s Joe Tangari who tops the list, who I wouldn’t associate with the strongest of Pitchfork brands. Note the gap after Tangari, Stephen Deusner, and Cohen; these three lead the way by a *significant* amount.



We see some predictable trends in their average scores as well. Cohen lives up to his reputation as a negative reviewer, although not the most negative; Adam Moerder is the worst of the bunch. Richardson continues to be positive from his 2017 trend, and for the most part he’s an outlier with plenty of reviewers clustered in the middle.

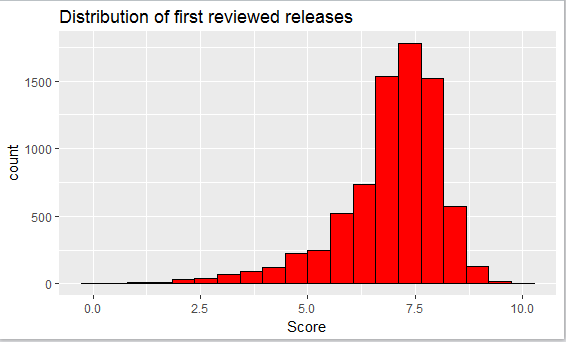
Models

Now we turn to attempting to predict the scores of the album. Now, without much to go off of in terms of information on these albums, these predictions are going to be poor, but they will (hopefully) give some insight into how these albums turn out. All code for model construction can be found here: <https://github.com/mikehalverson11/Pitchfork/blob/master/modelBuilding.R>.

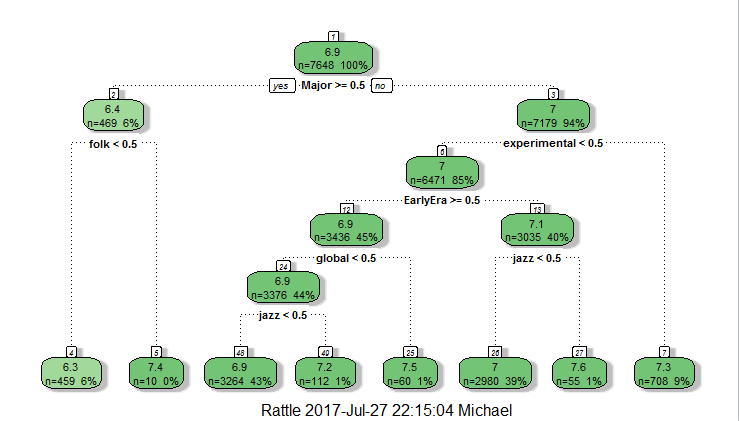
Debuts

First, we must tackle the ones with the least amount of information available, the debut albums (or at least the ones without previous information available). Since we want to mainly look at albums that are truly debuts, or ones that are the first relevant for an artist, we’ll look at only those reviewed after 2001, when the site has some time under its belt to review artists.

These are how the debuts distribute themselves:

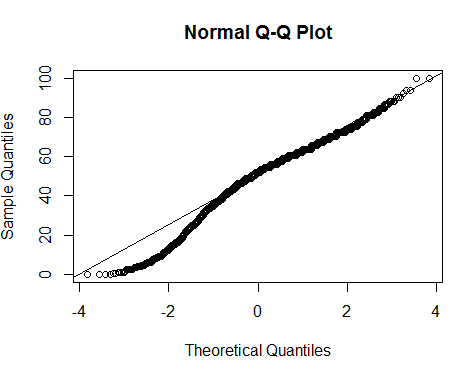
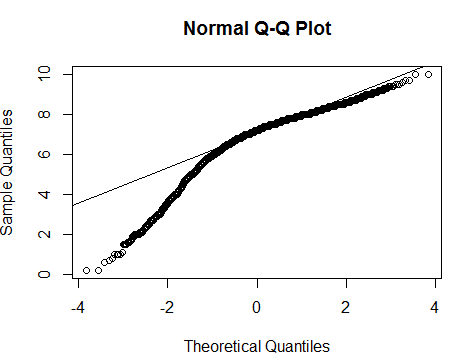


Pretty similar to the distribution of all reviews, although it does look like the tail for bad reviews is a bit larger, and the tail for high scored reviews seems a bit smaller. So what information do we have to distinguish between the mass in the middle? We turn to the more easily interpretable decision tree for some insight:



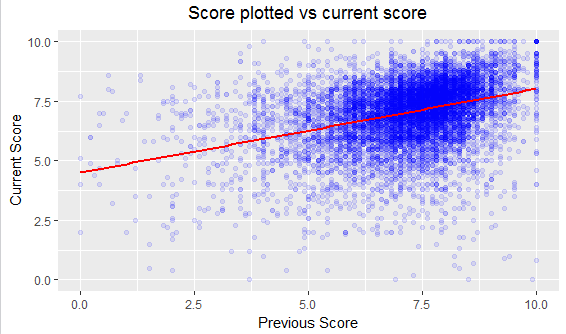
As you can tell by the scores on the bottom, the model struggles to distinguish between any of the model features. Of importance, though, is that first split, which indicates that major label releases struggle a great deal more than releases by more independent labels. As we saw earlier, jazz, experimental, and global music all tended to score higher. The significance of the Era for independent releases (Early Era indicates a release before 2009) shows a slight shift in the site’s taste over time, as they now are less friendly.

One area of concern is the non-normality of the data. The left plot shows the non-adjusted data; the best possible alteration was to square the review score which gives the second plot. It’s not quite perfect, but it’s a much closer approximation of the normal distribution than the unadjusted data. Smaller order transformations (log, square root) just made the issue worse.

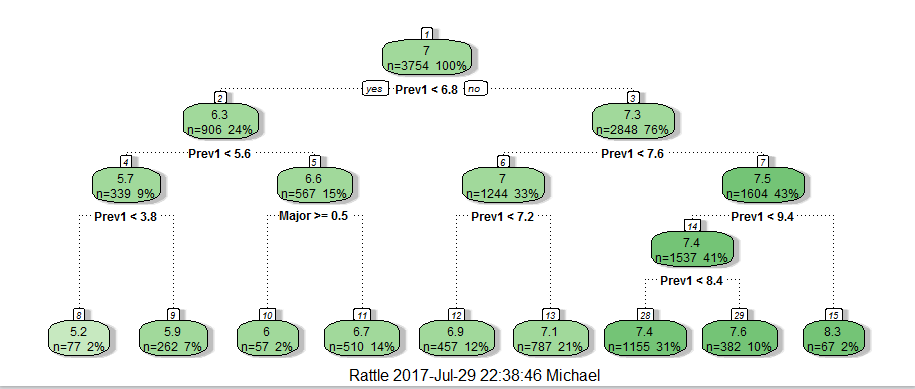


Sophomores

The accuracy of the debut album model was always going to be spotty, just because of the lack of data available. With subsequent albums, we might have a better chance. The scatterplot of previous album score and the followup score gives some hope:



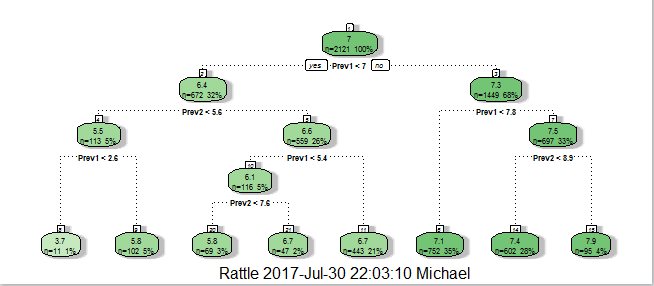
The relationship is mostly positive, which is what we would expect. There is a slight regression toward the mean on both ends; great rated albums tend to have a worse next record, whereas horrible records tend to not have a follow-up reviewed or they do slightly better with their next effort.



The decision tree now can pick out upper and lower tiers off relatively simple decisions, assuming that the next album will be slightly closer to the mean than the previous. The one interesting thing to note here is again, poorly received major label releases tend to rebound less strong than indie releases. Also of note is the sophomore slump for bands that go above a 9.4 with their first release- they go all the way back down to an 8.3 average for their second.

Third Albums

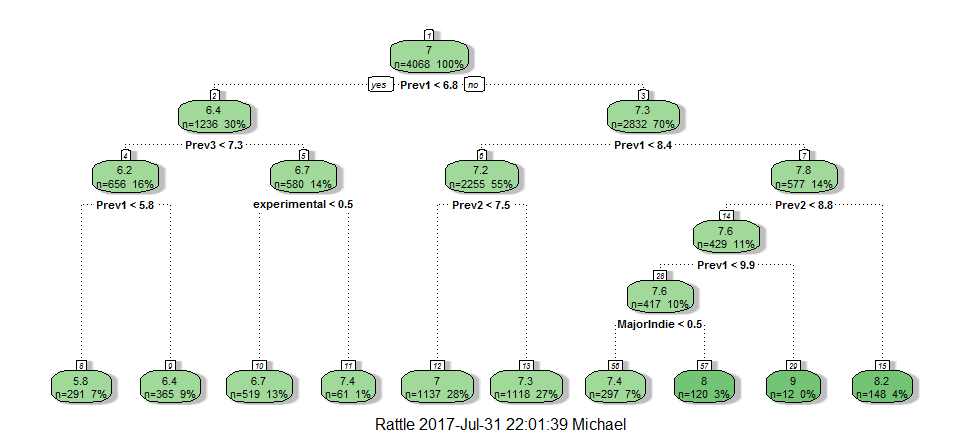
We now expand to the third albums:



Adding another album gives a slight better look. The right hand of the tree shows that yet again, bands that put out great records continue to do so. Although the furthest right shows that two great records to start off only leads to an average of 7.9; far less than a debut of 8.9 and a most recent greater than 7.8. The biggest gap we see is of artists who release a decent debut (>5.6) but a poor most recent album (<5.4). If their debut was good (>7.6), they rebound and average a 6.7; if they’ve done bad work twice in a row, they average 5.8.

All Others

Now we move onto the final set, all other albums. Since we want to grab all other albums, we’ll restrict ourselves just to the past three records. We don’t want to overfit, so this help prevents this as well. Again, we start with the decision tree for some preliminary insights:



Future Avenues of Exploration

The work here is far from comprehensive, and may be returned to it at a later date. With review data added for other websites, interesting relationships between the tastes of various other sites and Pitchfork could be explored. From the site itself, looking at the coverage of the site prior to releases (interviews, track reviews and other pieces) could also help us get more accuracy. While it would not necessarily be predictive, analyzing the relationship between the textual content of a review and its eventual score could be another useful tool for understanding the review process.