

# DATA 643 Final Project Planning Document

*Dan Fanelli*

*July 4, 2017*

## Music Recommender with User Filters

### System I Plan to Build Out:

- Music listeners have different moods, and often want suggestions within a specific genre.
- In my previous work with music and musical artist recommenders, I often found results from a genre that I was not hoping for (ie - when predicting for a user who had only specified 5 rap artists, the Beatles would be recommended).
- To accommodate for those 2 problems, I plan to implement a recommender that accepts an initial filter, namely the music genre (via tags?), from the user, and then makes recommendations ***only within that domain***.
- My single machine work in R lacked the power to process the data in the past, so multi-machine multi-processor work should allow for usage of the entire data set.

### Data Set:

Kaggle's "Million Song Dataset Challenge" (<https://www.kaggle.com/c/msdchallenge#description>) contains "the full listening history for 1M users".

### Tools:

This will be a large dataset, and so that I do not run into multi-hour waits as in past music recommender homeworks, I plan to use Apache Spark to spread the work among slave computers and allow for larger data sets to be fully analyzed.

### Data:

Either grouplens.org, kaggle, or another large music recommender data set.