Question of interest: Can we use song feature data to predict the year of a song?

Data: ~500,000 songs from Million Song Dataset

* Year
* 12 mean timbre features
* 78 covariances

No variables labeled

Background

(insert table of data distribution by decade)

Approach: Principal components analysis

1. Take smaller random subset of 10,000
2. Split into 7,500 training, 2,500 test set
3. Run PCA on correlation matrix of 90 predictors
4. Used first 8 species