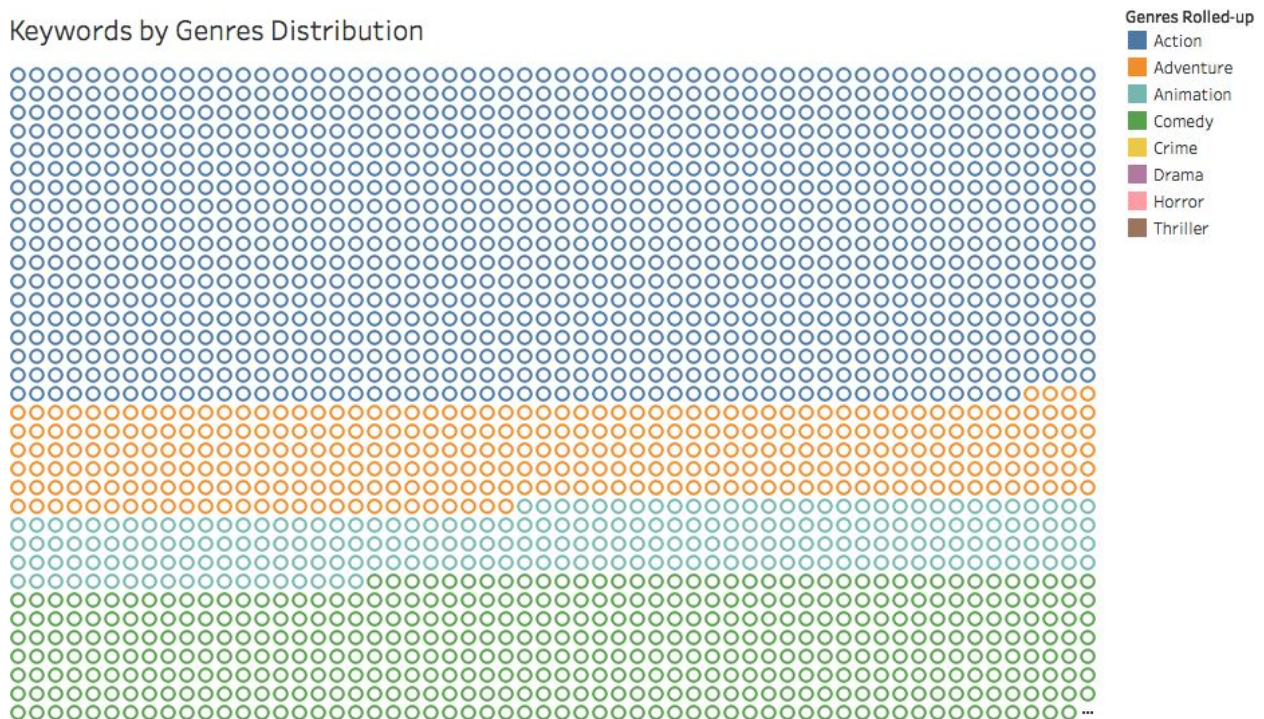


Misclass plots show that for poorly predicted genres Comedy and Drama the algorithm had a harder time searching for an optimal value of  $\lambda$ - can be seen by comparing to the the misclass plots of genres with high accuracies, Adventure and Animation. The classification error rises towards higher values of  $\lambda$  for both of the poorly predicted genres, and the ranges of the minimum  $\lambda$  values are either too far or too close to be impactful.

### Keywords by Genres Distribution



A stylization of the keywords describing movies in the IMBD 5000 database used for exploratory data analysis. The keywords are colored by genres to show the relative distribution of keywords available to describe each genre. Shows that predicting genres using a bag-of-words approach may be fruitful.