Additional information about the way we intend to visualize the project data with the use of Shiny in RStudio.

**1)Data Exploration:**

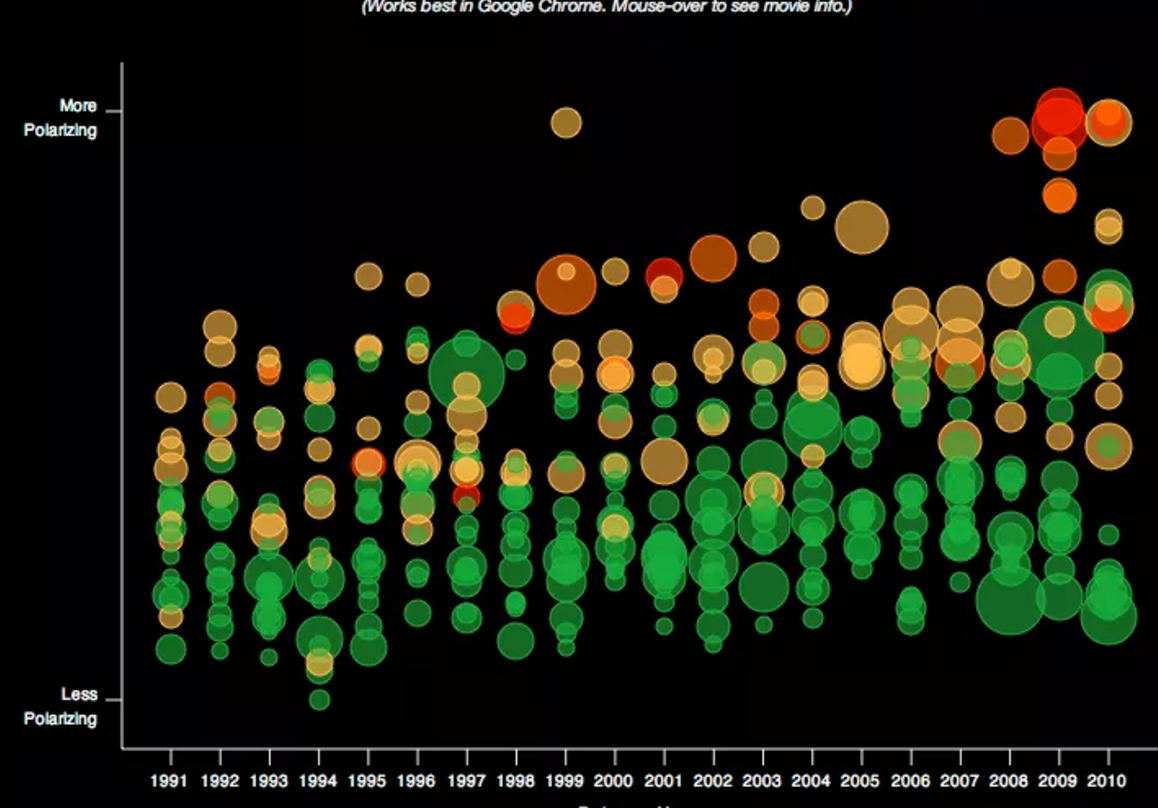
Idea1: A graph detection of years of great release/production. For a selected genre, a graph indicates the quality of releases over the years.

To do that, we will consider category (ex: fiction) and period(ex: 10-year) of release. For each of those 10 years, the 20 most popular movies(ex: in fiction category) is considered. We plot their standard deviation of rating (“polarization” ) in a graph.

Reasoning:

Movies with wider spread of ratings are more polarized. Movies that are more polarized are movies that some love and some hate, “which is the likely profile of a bad movie that’s safely manufactured for an existing fanbase”.

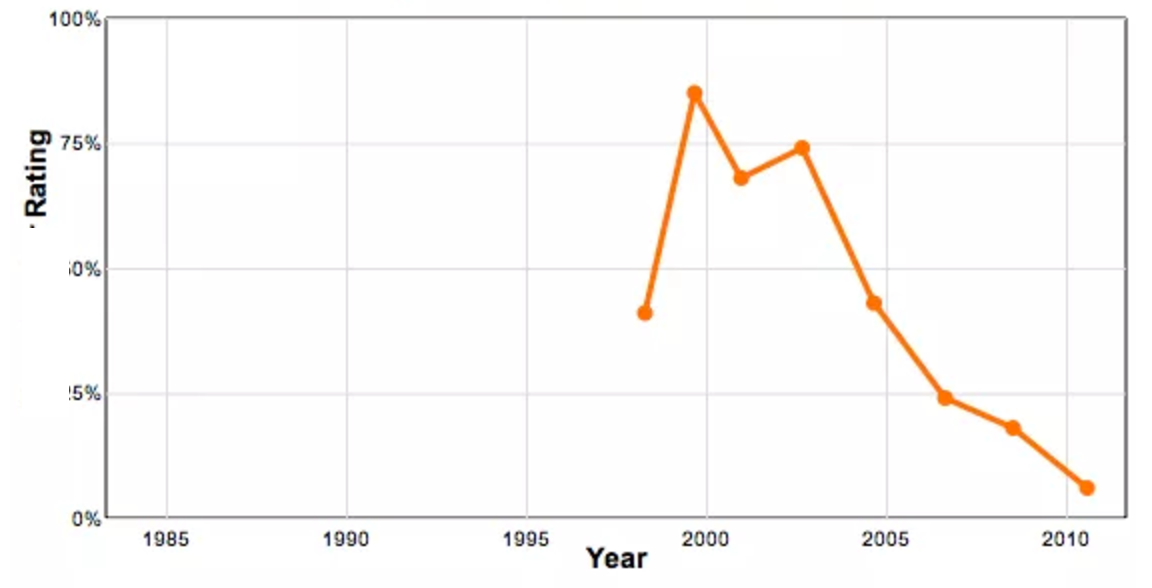
Wider spread of ratings are indicative of non-consensus in a specific year for a considered category. Which could be considered as year of mediocre movies release compare to other years with narrower spread of ratings for the same category.

Visualization template:

Idea 2: Average ratings for films , actors, and directors.

Punch in a name of film (director or actor) and see the ratings over time.

Visualization template:

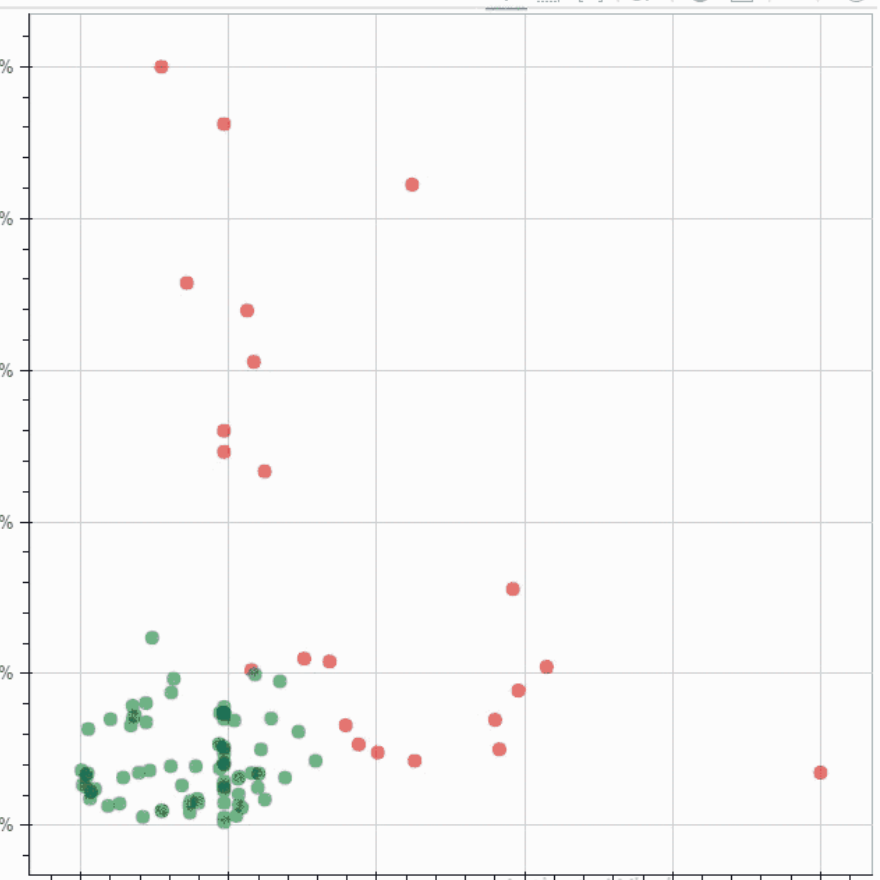


**2) Clustering:**

Idea1: A map of movie ratings classified as: Excellent, good, bad.

The map has 2 dimensions: budget and gross. In the below template the x-axis will be budget and the y-axis will be gross.

Green, red and blue points will differentiate Excellent ratings from good and bad. Rating bucket creation will be based on the Netflix and IMDB data source.

Visualization template:

Ps: Genre and Release year could be considered as dimensions for another map

**3) Prediction:**

Scoring of linear regression model built to predict number “movie\_facebook\_likes”.

Details:

Principal Component Analysis will be used to Identify which features of movies have the most influence on ratings (or "movie\_facebook\_likes" ).

1. We will use those features as the basis of movie recommendation.
2. We will use those features as the basis of “movie\_facebook\_likes” prediction.

Visualization template (1 and 2):

