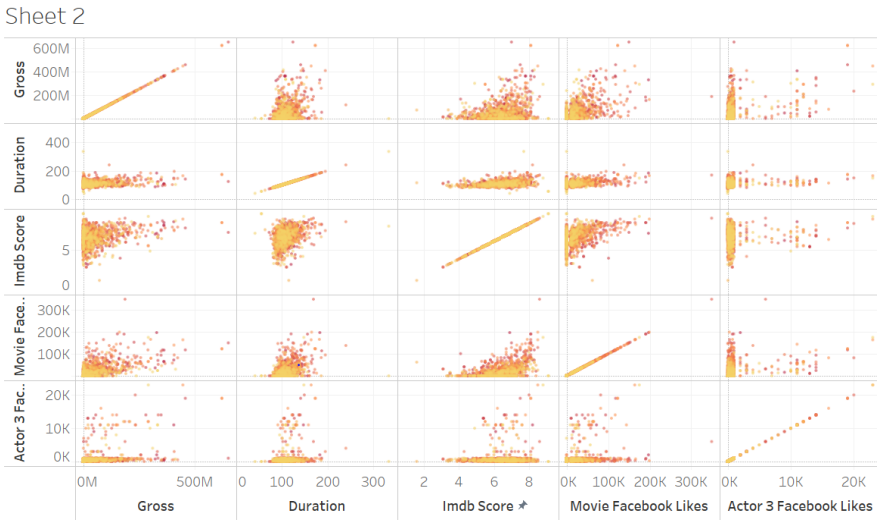
Alan Matias

CS 4460

Lab 7

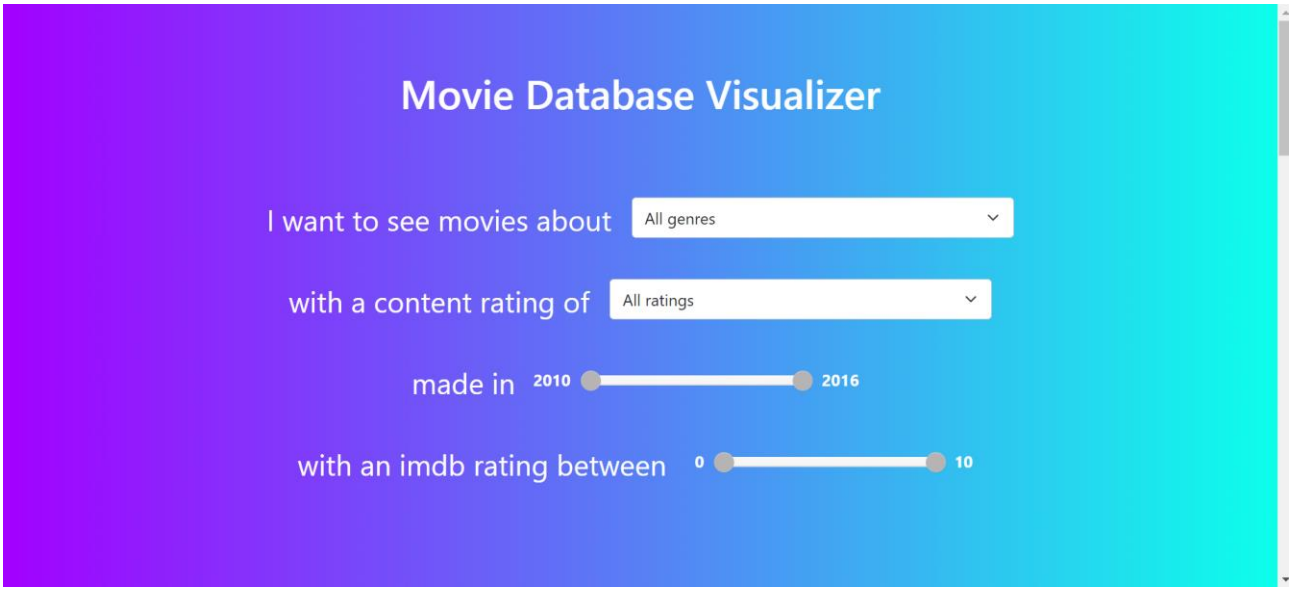
For this lab, I have chosen the **movies.csv** dataset. To preface the following features I have implemented for this lab, I will explain the users it is intended for. After looking at a scatterplot matrix in Tableau of the data, I found no interesting / surprising trends:

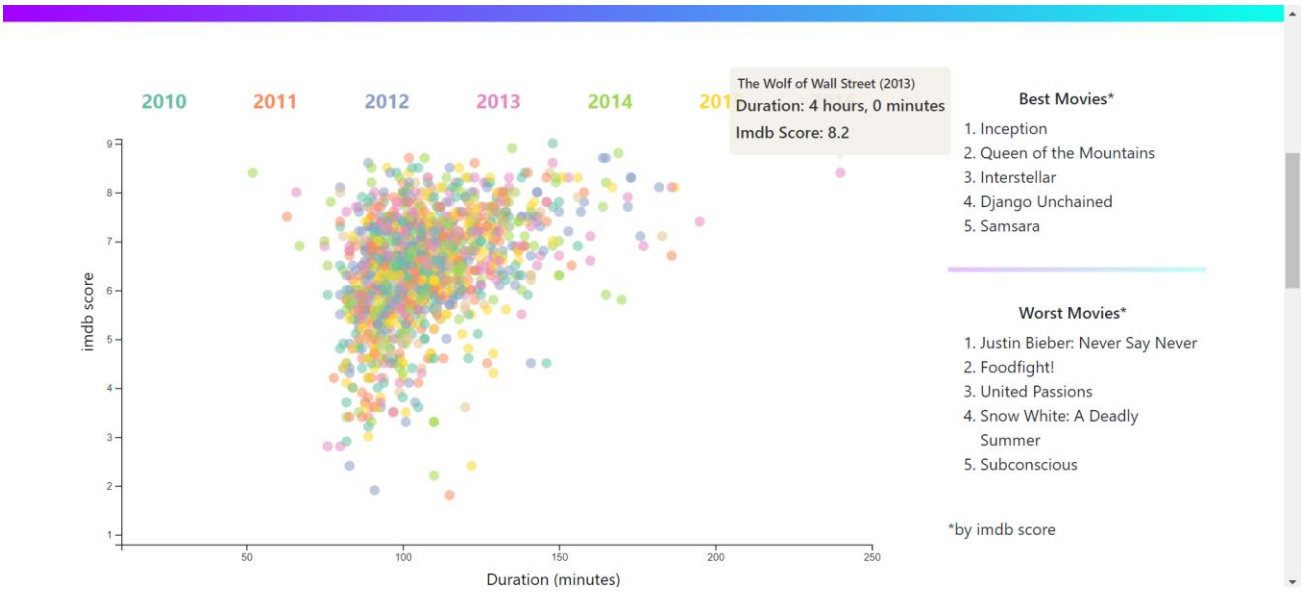


Additionally, aggregated data by year showed similar averages like avg imdb score or gross. Thus, instead of showing trends in data (of which are unsurprising), I decided to create a visualization like Zillow that lets the user explore data (movies) based on dynamic queries.

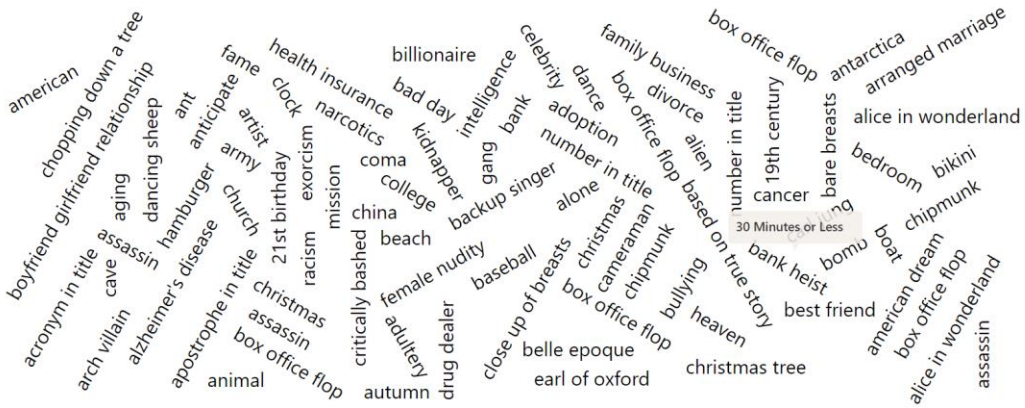
The main objective for the user of this viz is someone looking to decide what movie to watch. The user is able to select a genre, a content rating, a range of years, and imdb rating. Based on these parameters, the user will know: the top 5 and bottom 5 movies based on imdb ratings, the top 20 movies in sorted order by Facebook likes, and a word cloud of a random sample of plot keywords. There is also a scatterplot of movies with the duration on the x axis and

the imdb on the y axis. Thus, for example a user with a child can put on the longest movie with a PG rating. Or a group of friends can put on the worst imdb scoring Horror movie for the duration that they want to hang out. Or a user who is bored can look at the word cloud and see something they find interesting like “archery” and can hover over that word and see that the associated movie is “Brave”.





Random sample of plot keywords:



Top 20 movies based on facebook likes