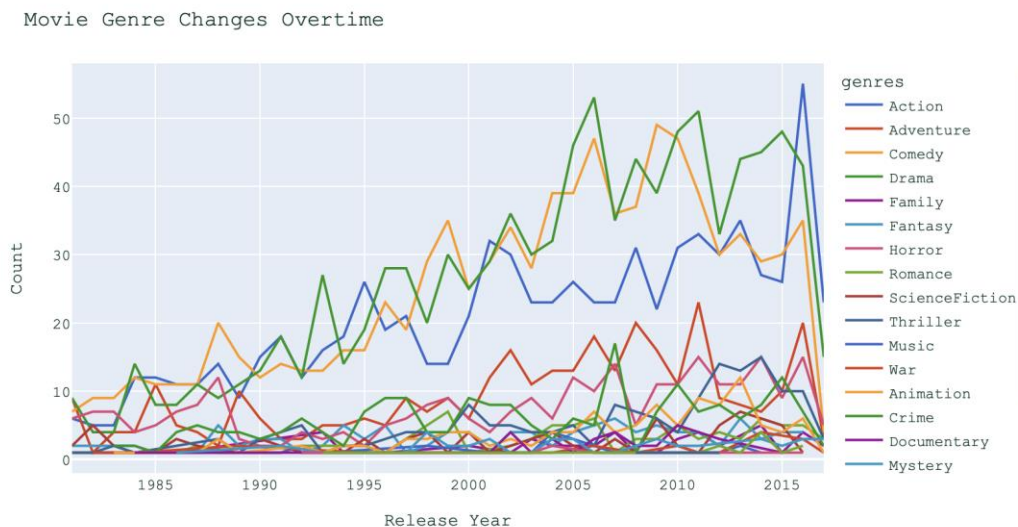


Group H Process Book: Discarded plots

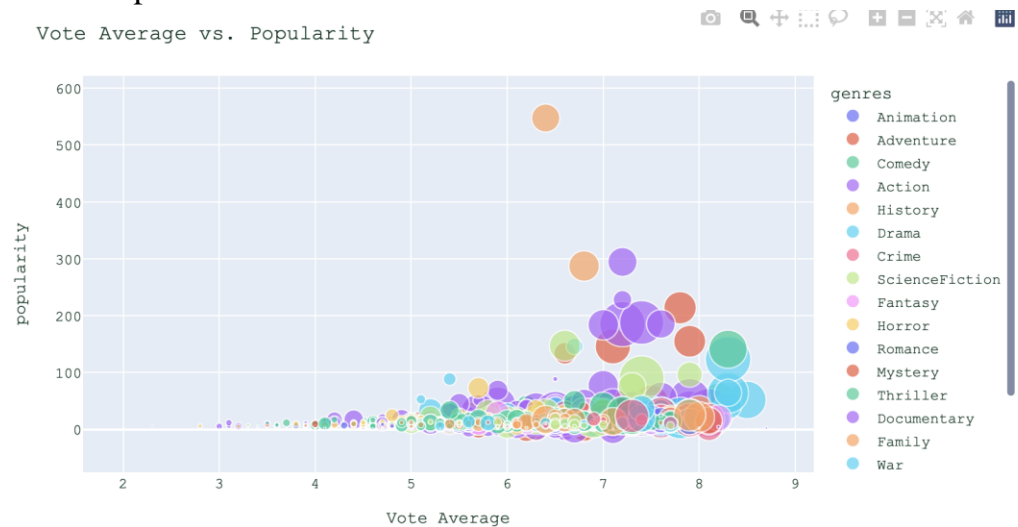
Decided to only use the first genre since we didn't want duplicated count in the part chart and bar chart. Then I made a plotly showing the release year and movie count grouped by movie genres in a line chart, but decided to drop that and do an animated bar chart instead since the original plot was not clearly showing the change in my opinion.

Discarded plot 1:

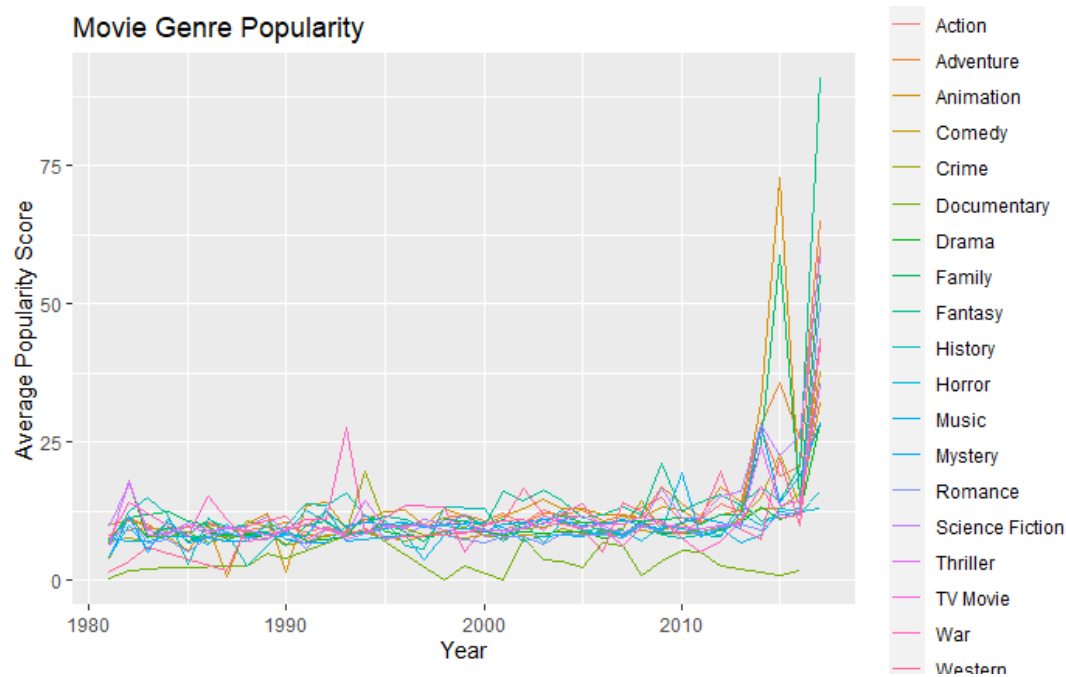


For the scatter plot, I initially did vote average vs. popularity with size equal to vote count, but decided to do vote average vs. revenue with size equal to popularity instead since I thought revenue is a more important metric compared to popularity.

Discarded plot 2:

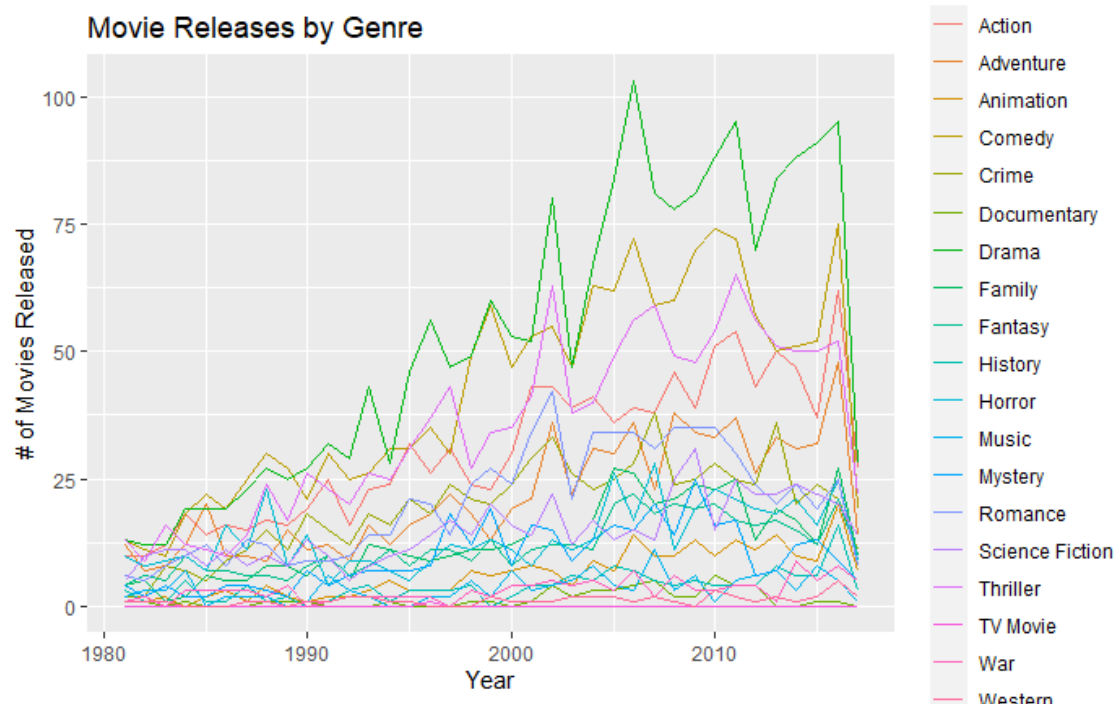


Popularity. This line was created in ggplot2 and maps the average IMDb of each genre. This plot was discarded because it was difficult to discern the y-value for any given line. There were too many genres included in the plot to show any insights

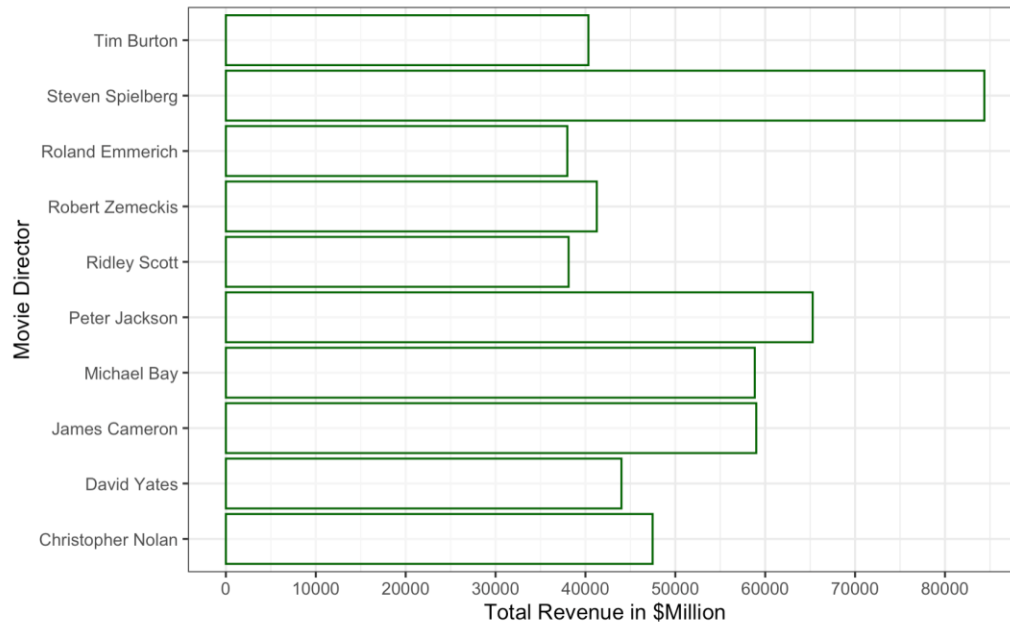


Movie releases

This was a similar plot to the one above, but for counts of movies released. It suffers from the same visual flaws. We also decided to use a cumulative plot instead of an annual production.



Initially default bar chart for the top directors by revenue and found the echarts4r package that will make the visualization more interesting. Can use icons instead of the common bar shape



Changing the color scheme for the keywords on genres. The initial color scheme is Dark 2 where the frequency is shown by the size of the word in the wordcloud. Change them to single colors scheme and differ among the genres under comparison. The size and color saturation indicate the frequency of the keywords appeared.

