Springboard Capstone - Statistics Exercise

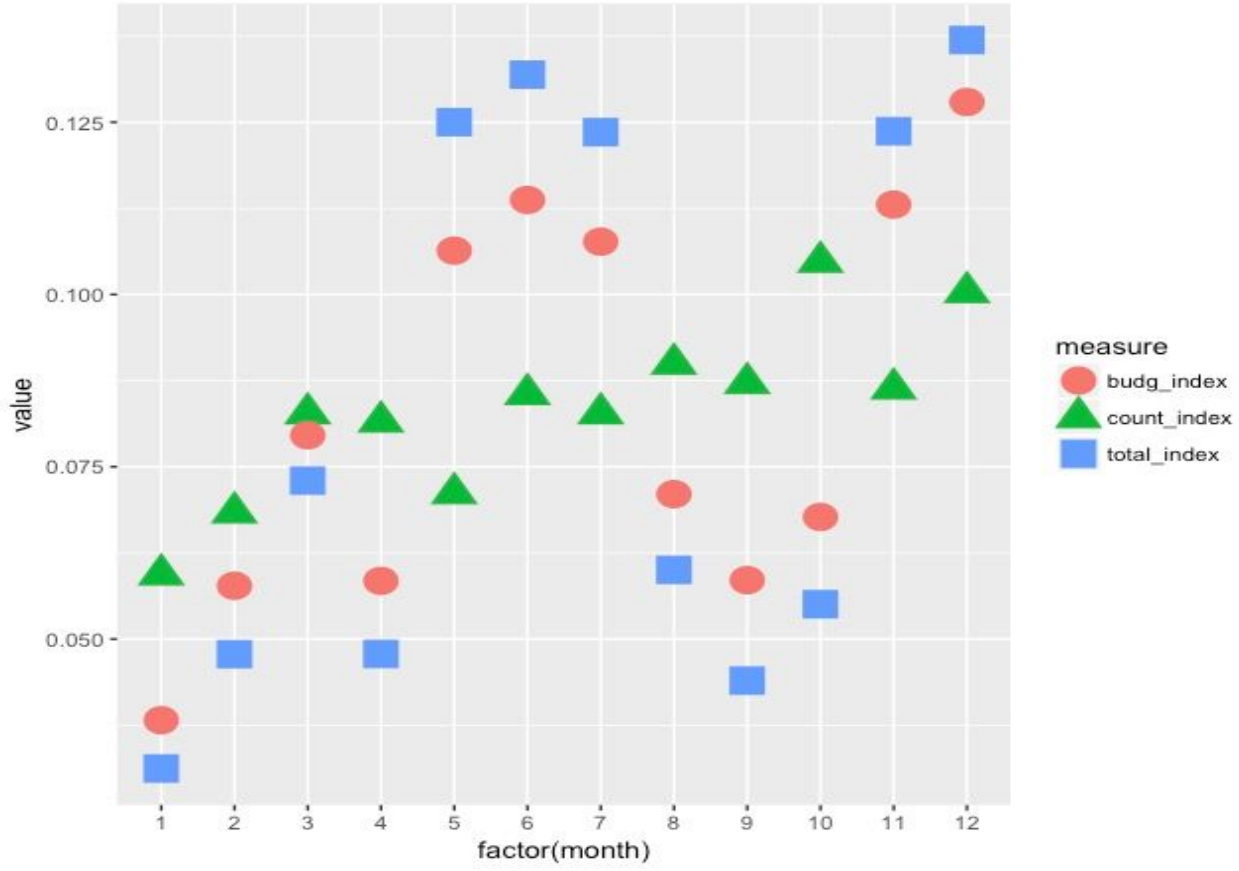
*[R code for the below is here.]*

Interesting count - I decided to see how many films in my dataset would be considered a commercial success at different performance ratios (where performance ratio = total box office / production budget)

- At perf_ratio = 3, 1625 or 5003 films (32.46%) are considered successful
- At perf_ratio = 4, 1157 of 5003 films (23.12%) are considered successful
- At perf_ratio = 5, 892 of 5003 films (17.8%) are considered successful

Given these counts (and an anecdotal idea of approximately the proportion many films are considered successful within in the industry), I choose a perf_ratio of 4 for the below analysis.

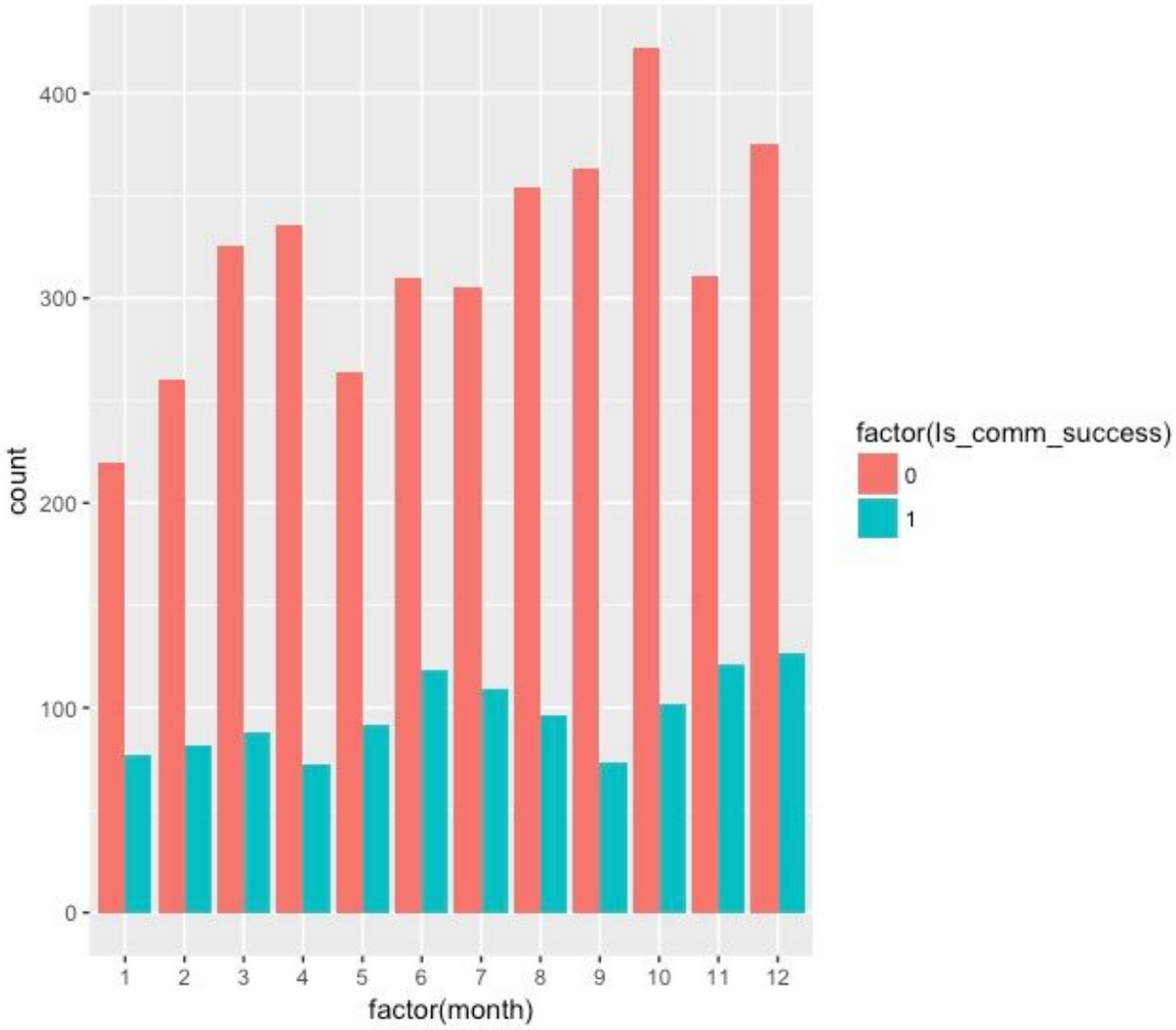
Trends - I looked at the relationship between release month, production budget, total box office, and the supply of movies.




-

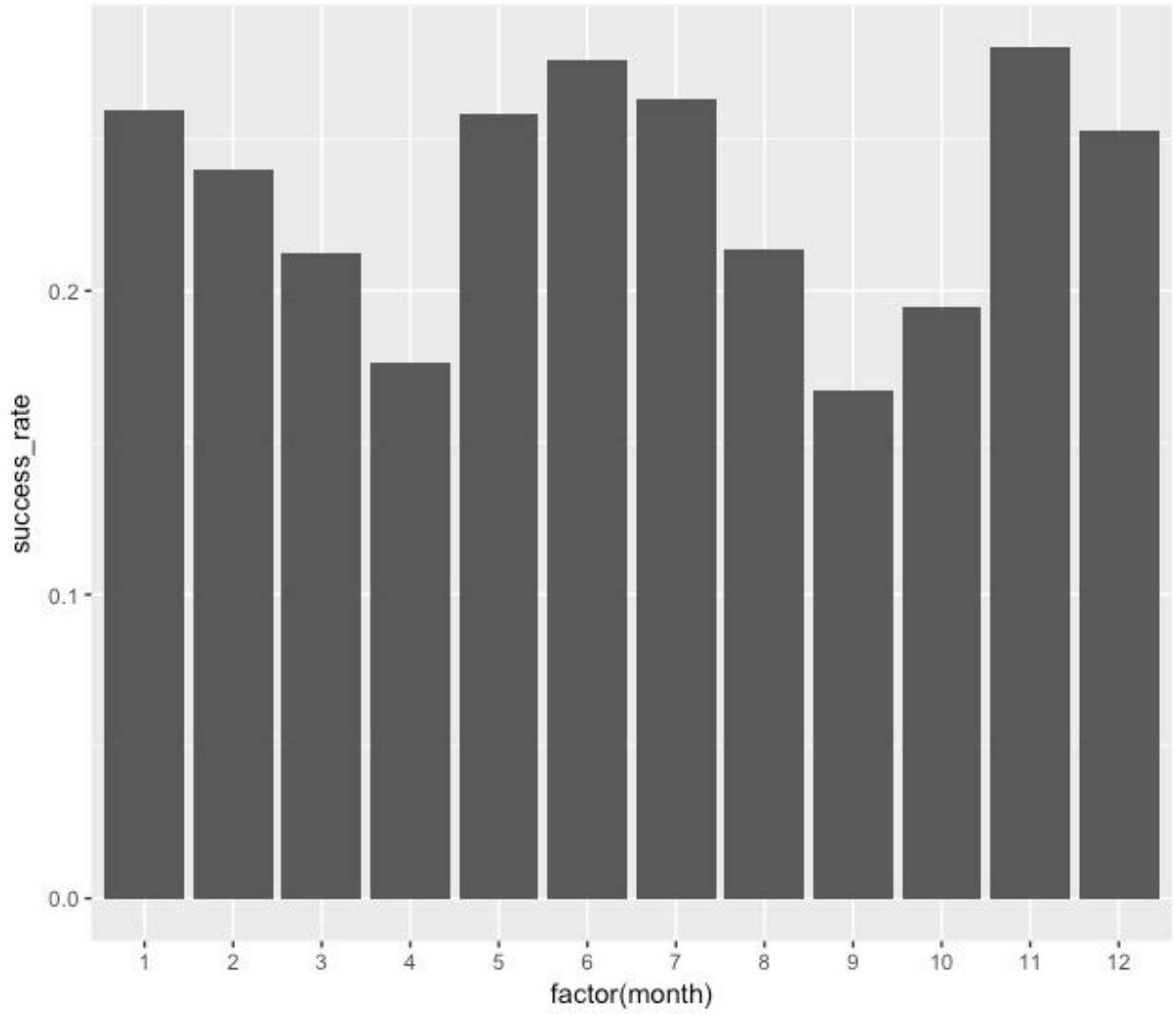
- It appears to me to look like box office (total_index) is much more related to production budget (budg_index) than to supply of movies at a given time (count_index).

Obviously, there are months that are prone to lower box office (January, September) and months that are prone to higher box office (May through July, November, December). This made me curious if there are more commercial successes in those months than in the lower performing months.




- Again, there are more total commercial successes in the higher performing months, but there are also month total films.

The question became whether or not there is a higher ratio of commercial success in the high performing months.




- It does appear as though some of the higher performing months have a higher rate of success, but Feb and Jan are obvious counterexamples. January (the month with the lowest total box office) has a higher rate of success than December (the month the highest total box office).

Given all of the above, I am still interested in examining seasonality for film, specifically as it relates to traditional 'underperforming' months (like September). Do September films underperform and have a lower success rate because consumers aren't as interested in entertainment in September, or because studios knowingly release films that are less likely to do well in that window? What really drives box office, supply or demand? To what extent is it both?