

Ashley Bruce
 MAT 259A
 Winter 2021

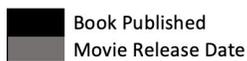
Assignment 1: Data Mining and Knowledge

During this year’s winter break, my sister had recommended that I read the Maze Runner series, as it was one of her favorite trilogies. After reading all three books, we decided to have a movie marathon and watch all three of the respective movies. This got me thinking about how movie adaptations of book series’ affected the popularity of a series over time.

I specifically wanted to explore books that were part of a series, to see if interest in the book spiked after the respective movie came out (ie. checkouts of book 2 in a series spiked after movie 2 came out). In order to investigate this trend, I had to find books in a series that had a movie counterpart for each book. This limits the books being looked at to those that were released / had movies come out while the database has been active (2006 on). Books that fit this criteria included the Maze Runner series, the Twilight series, the Hunger Games series, and the Divergent series. The following table shows the publishing dates of each book in each series and the respective release dates of the movies.

Table 1. Year the book was published and year the respective movie was released.

	Maze Runner	Scorch Trials	Death Cure	Twilight	New Moon	Eclipse	Breaking Dawn	Hunger Games	Catching Fire	Mockingjay	Divergent	Insurgent	Allegiant
2005													
2006													
2007													
2008													
2009													
2010													
2011													
2012													
2013													
2014													
2015													
2016													
2017													
2018													
2019													
2020													



In order to count the number of books checked out for each year, I first looked at all the entries that were brought up when searching for the desired title name. The query for this call can be seen in Figure 1 below.

```
1 select *
2 from spl_2016.inraw
3 where (
4     LOWER (title) like '%maze runner%'
5 )
6 limit 100
```

Figure 1. Query to list the entries with the title “Maze Runner.” This call was used to obtain the ‘callNumber’ entry of the desired book and to filter out any other entries with a similar title. In this case, the desired callNumber is ‘YA DASHNER.’

Out of the entries generated, I looked for the entry that was applicable to the book desired and made note of the unique callNumber. The call numbers for each series is shown in Table 2. This call number was then used to count the relevant number of checkouts for each book over the years, as seen in Figure 2. Output for the call in Figure 2 can be seen in Figure 3.

Table 2. Queries in Figure 1 were made to the respective series, column 1, to get the callNumber, column 2.

Series	CallNumber
Maze Runner	YA DASHNER
Twilight	YA MEYER
Hunger Games	YA COLLINS
Divergent	YA ROTH

```
1 • select
2     count(title) as checkouts,
3     year(cout) as years
4   from spl_2016.outraw
5  where (
6     LOWER (title) like '%maze runner%'
7     and callNumber like '%YA DASHNER%'
8  )
9   GROUP BY year(COUT)
10
```

Figure 2. Query for seeing the number of checkouts of *Maze Runner* over the years.

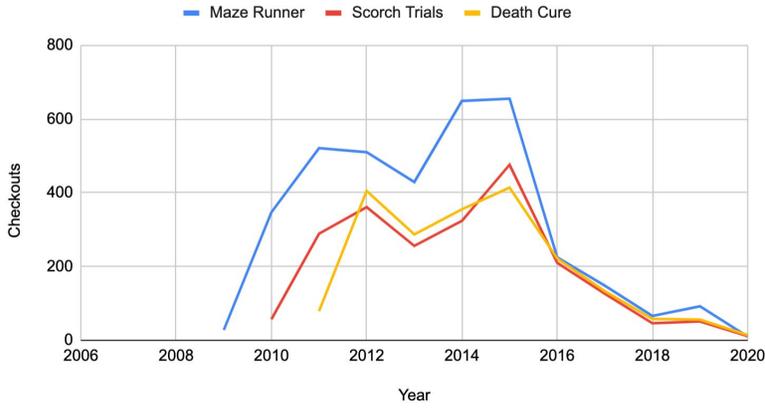
	checkouts	years	
▶	28	2009	
	347	2010	
	521	2011	
	510	2012	
	429	2013	
	649	2014	
	655	2015	
	225	2016	
	148	2017	
	66	2018	
	92	2019	
	10	2020	

Figure 3. Search results for the query in Figure 3.

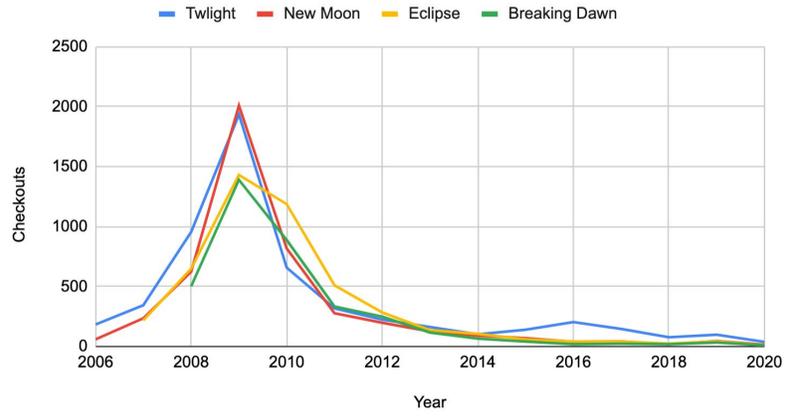
This query was repeated for each book of each series with its respective call number. Data for each call is shown in the following charts.

Figure 4. Charts showing checkouts a year for each book in their respective series.

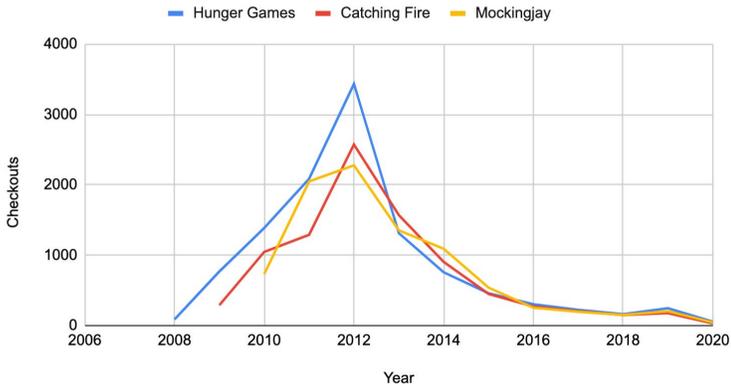
Maze Runner Series



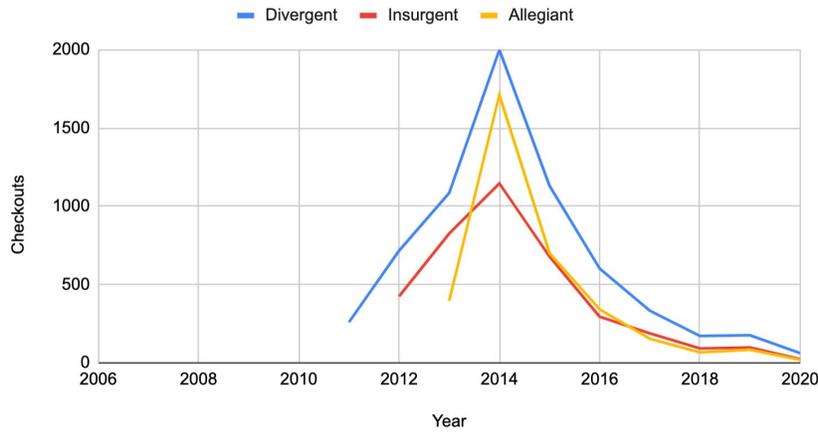
Twilight Series



Hunger Games Series



Divergent Series



As seen in the graphs above, with the exception of Maze Runner, there is only one significant spike in each series. This increase in checkouts for each book in the series appears to be associated with the release of the first movie in the series. What I thought was interesting was that the subsequent release of movies related to the series didn't spike additional checkouts within the series, like I thought it might. One possible explanation for this could be that because all the books in the series are published before the first movie is released, those who are intrigued by the movie can go and read all the books in the series within the same year. Subsequent movie releases no longer trigger the same increase in book checkouts because those who were interested enough in the series read it after hearing about it after the first movie.

```
1 • select
2     count(title) as checkouts,
3     month(cout) as month
4   from spl_2016.outraw
5  where (
6     LOWER (title) like '%twilight%'
7     and callNumber like '%YA MEYER%'
8     and year (cout) = 2010 )
9  GROUP BY month(COUT)
```

Figure 5. Query to see the number of checkouts a month for a given year for one of the books in the series.

checkouts	month	
122	1	
91	2	
83	3	
79	4	
79	5	
83	6	
101	7	
90	8	
47	9	
59	10	
43	11	
43	12	

Figure 6. Search results for the query in Figure 5.

Twilight Series

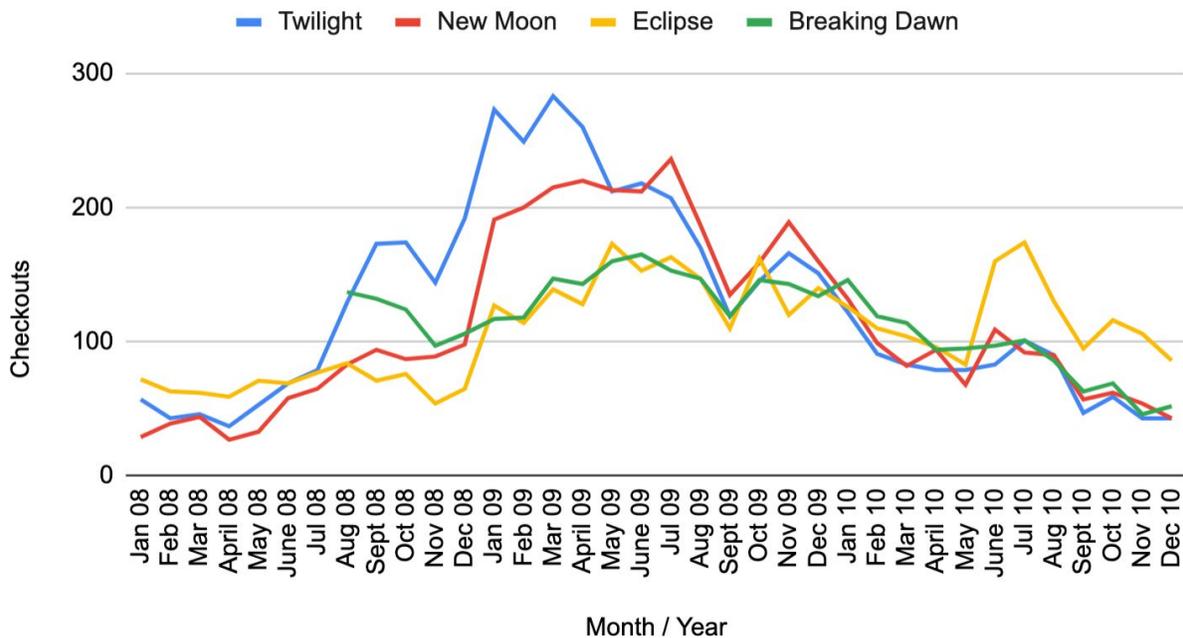


Figure 7. Graph of the checkouts per month over 3 years for each book in the twilight series.

After looking at the charts in Figure 4 and seeing how the max number of checkouts for each book was during the same year, I wanted to get a better breakdown of how the checkouts changed during the course of that year. I picked one series, Twilight, and looked at a month by month breakdown during the year where the max checkouts was (2009) as well as the years before and after. To do this I used the query seen in Figure 5 and changed the title and year as appropriate.

Twilight, the first movie, came out on November 21, 2008. As seen on the graph, there is a huge spike in the number of checkouts for Twilight, New Moon, and Breaking Dawn in December of 2008, which is just after the first movie came out.

New Moon, the second movie, came out on November 20, 2009. After the first movie was released in 2008, the checkouts slowly declined, but there is a small spike following the release of the second movie, especially in checkouts of New Moon.

Eclipse, the third movie, came out on June 24, 2010. A small spike can be seen in checkouts of Eclipse in June and July of 2010.

On the graphs that had years as a scale, it appeared that the number of checkouts was only correlated with the date the first movie was released. But, if breaking it down to a month by month basis, there does appear to be a small increase in the number of checkouts for the books during the dates subsequent movies are released. Due to this trend, I decided to increase the timeframe being looked at in Figure 7 to include the dates that the Breaking Dawn movies were released.

Twilight Series

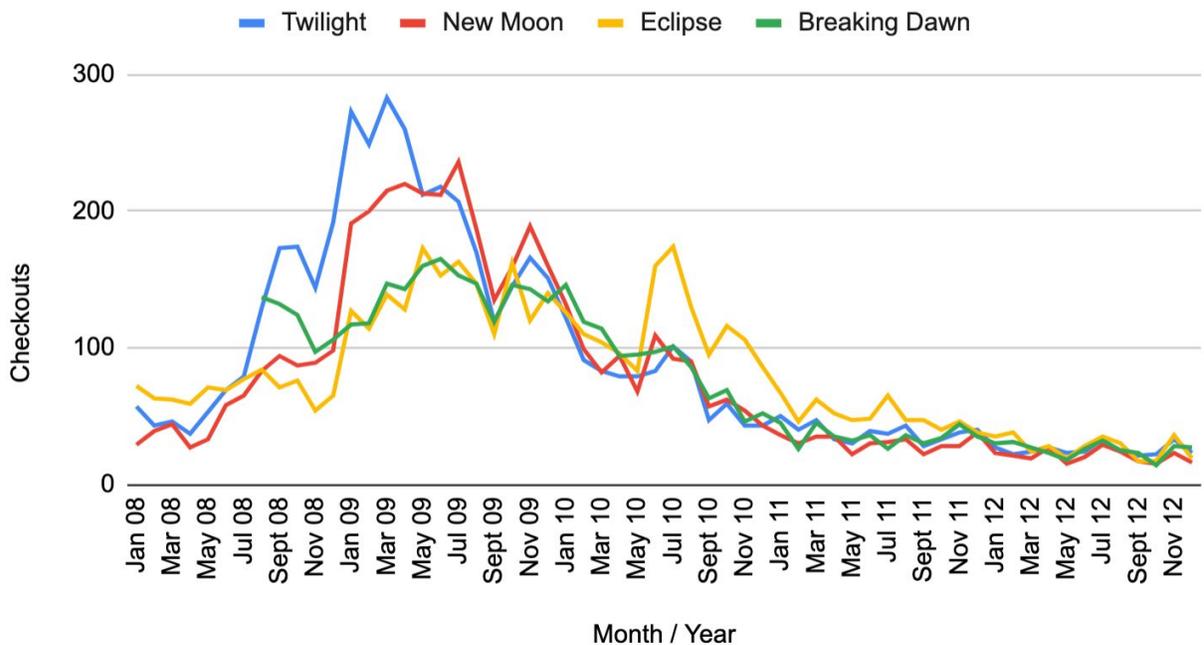


Figure 8. Increased timeline of Figure 7 to include dates that the Breaking Dawn movies were released.

Breaking Dawn Part 1 was released on November 18, 2011. It doesn't appear to have the same spike during this release date as the other books did.

If time permitted, I would have done this same month-to-month comparison for all the series I looked at above and see if this trend held for the other books. Even without this comparison for each series though, it seems that book popularity of a series is most related to the release of the first movie in the series.