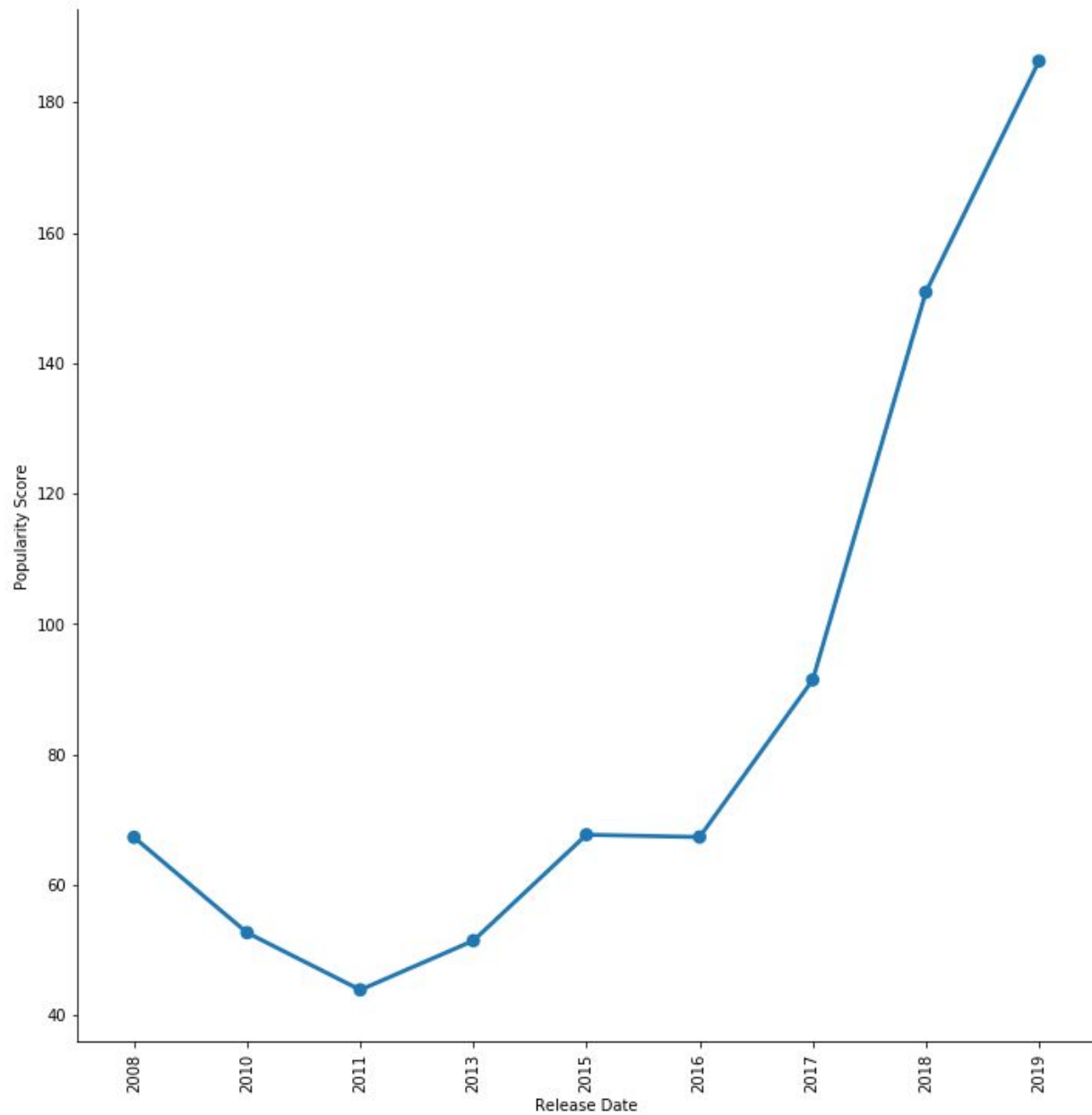I ended up doing two somewhat separate analyses. First, I extracted data from The Movie Database's API to find the popularity scores of every marvel movie. I then compared that to the csv file I found in kaggle, which included all of the marvel movies' data from IMDB. In comparing the average popularity score of marvel movies by year from the API and review ratings of marvel movies by year, I was able to visualize the history of marvel studios from the audience's perspective. It looks like in the beginning (circa 2008), Marvel movies were mildly popular. Then, from around 2011 to 2013 there was a major dip in popularity, meaning that they had a particular stretch of movies that were received poorly. Then in 2014 the popularity increased. This is where the two sets of data diverge, as the popularity scores steadily increase till present day, while IMDB reviews fluctuate until hitting an incline in 2017 with the release of Avengers: infinity war. This movie was wildly popular as it was a two part storyline and provided the ending to many character's story arcs. The main struggle I had working with this data was finding that my data was not in the right data types within my csv file. For example, I had to convert release dates into datetime objects in order to filter my data by year.

 On the flipside, I also did a deepdive into dialogue textfiles of 17 of Marvel's films (the files of the other films were unreadable). Looking at the dialogue of each of the films, I was curious to see if the average number of words said in the films would have any correlation to the story I derived from the review and popularity data. I didn't see one in particular, as my data was hard to work with in terms of years because I had to leave out the films which had corrupted text files. I did note, though, that there was no clear correlation between the ratings of the movies and the length of their text files. For example, Iron man 2 was rated relatively low, yet it had the third most words in it's text file. I also noted that based on the movies with the most words in their text files, it seems that runtime and words spoken do not correlate. It is known that the more recent marvel movies have longer runtimes than their predecessors (though unfortunately I did not compile the data for this and add it to my dataframe), yet the number of words in marvel films seems to stay within a relatively close range of around 40,000 to 60,000 words. For example, the movie with the longest runtime is Avengers: Endgame at 3 hours long, yet Iron man 3, which is only 2 hours and 11 minutes long has almost 10,000 more words in it's text file. An explanation of why this could be might have to do with the amount of non-dialogue scenes in the films. Maybe scenes without dialogue (ie. action scenes perhaps?) take up a larger sum of the runtime of the newer movies as opposed to more dialect-heavy scenes. This is an interesting and unexpected finding which I think would be interesting to explore further. If I were to continue with this analysis, I would look at comparing runtime and words spoken to see if my hypothesis is correct.

Average IMDB Review Rating by Year for Marvel Movies
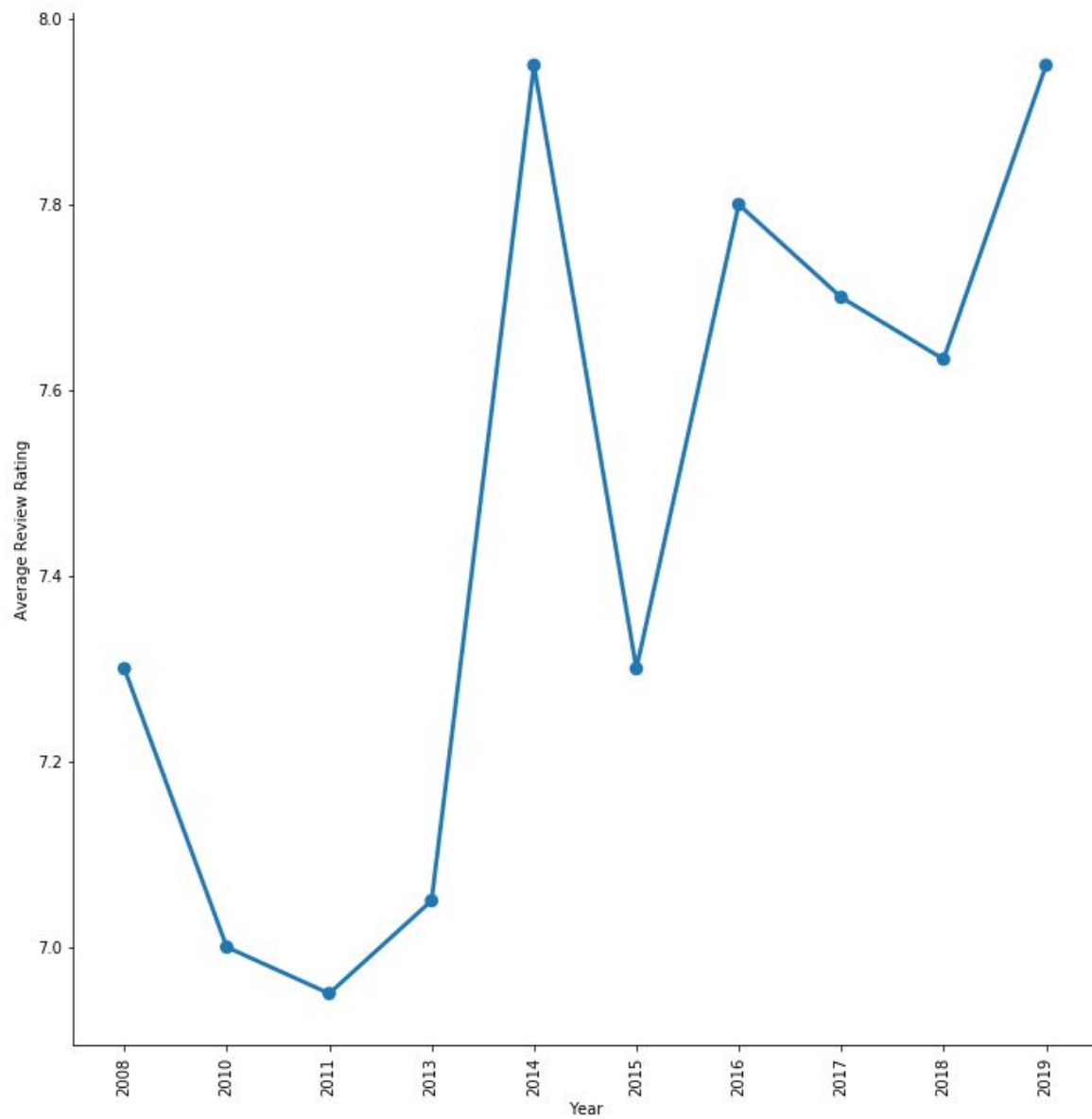
Average TMDB popularity Score by Year for Marvel Movies

Marvel Movies Dialogues Analysis:

Number of Words in Marvel Movies

Average Number of Words in Marvel Movies by Year