---

### 0.0.1 Question 1d

Suppose we are interested in using the news to predict future stock values. What additional data would we need to predict stock prices, and how could we connect that data to news articles? In addition, what attributes or characteristics of the news might help predict the stock value?

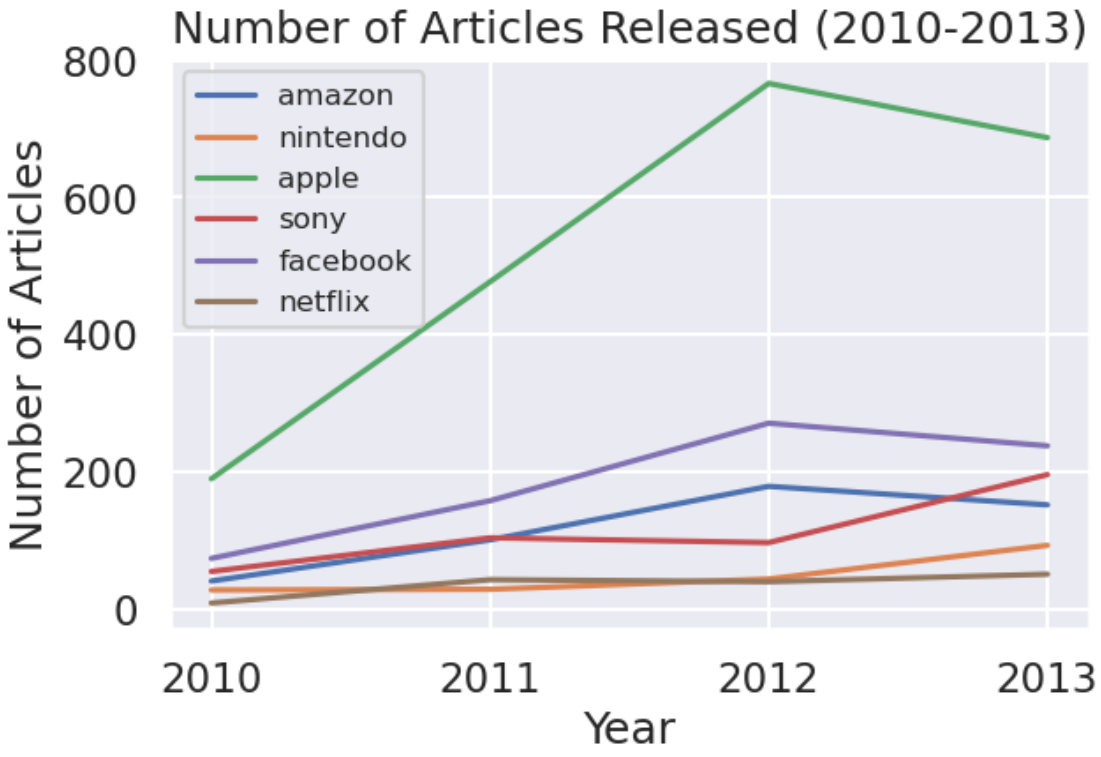
We would need data with company names, their corresponding stock symbol, their corresponding stock price, and the dates/times for those stock prices. We could connect that data to the news articles by parsing the content or title of the news articles for the stock symbol or company name, for example "(OPERA"), and creating a new column in the news dataframe with the stock symbols that are referred to in the article. Then, we can merge based off the stock symbol.

In addition, the released_at attribute of the news would be helpful to predict stock value, as we would like to see the stock prices before and after the released_at date of the news to see if there is a trend between the news and the stock prices

**Question 2d, Part ii** Given your code in the previous part is correct, after running the cell below, you should be able to see the number of articles released mentioning `companies` for each year. The plot should look like this:

```
In [18]: plt.figure(figsize=(6, 4))

         for company in companies:
             sns.lineplot(data=year_news.reset_index(),
                          x="Year",
                          y=company,
                          label=company)
         plt.legend(fontsize="12")
         plt.xticks(np.arange(2010, 2014), np.arange(2010, 2014))
         plt.ylabel("Number of Articles")
         plt.xlabel("Year")
         plt.title("Number of Articles Released (2010-2013)");
```




What trends do you notice in the plot above? Feel free to reference or Google any events to explain the trends seen in the graph. What are some limitations of using data and the corresponding plot to analyze the performance of different companies or trends?

**Hint:** Remember the source of the articles and the subset of the articles we are analyzing in this assignment.

Through the years 2010-2012, there was a large increase in the number of articles mentioning the company Apple. Facebook and Amazon also had a small increase over these years, but not nearly as much as Apple. Upon Googling Steve Jobs, he died in 2011, and presumably left a lot of turmoil or speculation about the company's future. Thus, it makes sense that the number of articles mentioning Apple skyrocketed. However, just because they were mentioned more times, this does not necessarily indicate a positive trend in Apple's performance. As I recall, a lot of articles questioned how the company would function in the future, and thus not all mentions were good mentions in this scenario

**Question 3c, Part ii** Below we have provided a plot looking at these differences. Comment on why we see differences when calculating the sentiment of an article as the sentiment of the first sentence mentioning "microsoft" or "msft" in the article versus the sentiment of the entire article itself. How does context play a role when evaluating the sentiment of a text?

```
In [ ]: sns.kdeplot(msft_scores_2010['sentiment_difference'])
        plt.xlabel('Sentiment difference')
        plt.title('Difference between full and approximate sentiment scores');
```

Perhaps the article as a whole has a negative sentiment, as in if the market overall is trending downwards. However, even if the overall market trend is trending downwards, sometimes the Microsoft stock is individually performing well, or vice versa. Therefore, the overall sentiment of the article may not match the sentiment of the Microsoft mentions within that article.