1. The analysis I did was primarily based around finding out what words occurred most often in the review data. Because of the sheer volume of data, this process took some time and ended up being quite computationally intensive. Additionally, due to my relative lack of experience with sentiment analysis, I was unable to filter words such as “the” and “I”. Regardless, I was able to find the most commonly used words across the reviews. The corresponding visualization for this is a pie chart depicting the 10 most used words in the dataset (words.pdf). The other analysis I did was finding the average review for every day in the review data. This involved a non-zero amount of pre-processing as data needed to be cleaned and I had to convert Unix timestamps into dates. The corresponding visualizations for this analysis are line plots (Appliances.csv.pdf and Magazine_Subscription.csv.pdf). Due to the sheer amount of data in these plots, neither plot is easily readable, but I intend to remedy this problem in my final submission. Finally, I plotted Amazon’s adjusted closing price versus time (AMZN.pdf).
2. For advanced visualizations, I made some relatively simple interactive line graphs (in Bonus.ipynb). Being able to mouseover stock price and other timeseries data and see values at any given point is a standard feature of any respectable trading platform or the like. Because my data is so heavily focused on timeseries data, this visualization is useful to aid the user with reading and understanding individual pieces of data as opposed to looking at overall trends. In the case of the ratings data, it is still relatively hard to read, but it becomes much more accessible to the user with this visualization technique.