Questions

1. You have shown that this analysis could be performed after the fact. Could this be done in real time as well?
    a. Absolutely. The original method of data collection scraped the Twitter API. A similar method could be used that then passes the tweet text to the TextBlob.
2. Why a TextBlob?
    a. It did everything we needed to do and built on the back of the NLTK. All we had to do was pass it a string.
3. Why did you choose the election for the analysis?
    a. Quite simply, the polarization of the election. With sentiment analysis, it is much easier to identify the extremes than the slightly leaning text.
4. Is there any analysis you would have liked to perform but were unable to?
    a. I would have loved to have seen better geographic data. Unfortunately, the only data we had scraped was when the user provided geographic location. This could be fictional places, or nowhere near they were, so it would be hard to use that data conclusively.
5. Why Twitter data specifically?
    a. It was relatively easy to collect hundreds of thousands of data points over a specified time period. Compared to directly polling people, the amount of data we obtained was far greater and in a much shorter time frame. Plus, we can easily adjust the parameters to run the same methods on different time periods and subjects.
6. Were the results what you expected? Did you anticipate the spikes?
    a. I was hoping for spikes! It was also a lot easier than I expected to match the spikes in sentiment polarity. The spikes were large, and the headlines followed.
7. What challenges were you able to overcome?
    a. I initially tried to visualize the data using .CSV files as the inputs to Tableau, but the data was not importing correctly. I ended up switching to Excel, and instead of using the full data file I only exported the four columns I needed. This slashed the file size significantly and made it much easier to work with. Plus, the data imported correctly!
8. Why Tableau for visualizations and not say GGPlot?
    a. I used Tableau in our Data Visualization class and liked using it. It also made it very easy to consolidate the points by hourly and minutely values.
9. Any improvements you wish you could have made?
    a. Reviewing some of the tweets, I would not necessarily agree completely with the sentiment assigned. It is hard to classify sarcasm. It would be nice to tweak some of the default weights or even use a corpus trained on political tweets in the first place instead of a general corpus.
10. What is your biggest takeaway?
    a. I was surprised at how overwhelmingly positive our searches were. It is likely that the negative responses are the vocal minority. But for the most part, there were very few time periods where the overall sentiment was negative!