Introduction

For this project, we generated two graphs: one graph that represents a small random sample of public tweets (Figure 1), and another graph that represents tweets filtered by the word “tired” (Figure 2). Both graphs were grouped by time zone to represent the tweet volumes of the four American time zones: Eastern, Central, Mountain and Pacific. Figure 1 displays the variation in the volume of tweets over time. Figure 2 presents the variation in the volume of tweets containing the word “tired” over time. We chose to filter tweets according to the word “tired” because we expected to see meaningful trends in the data throughout a day, such as an increase in frequency after dinner and before bedtime.

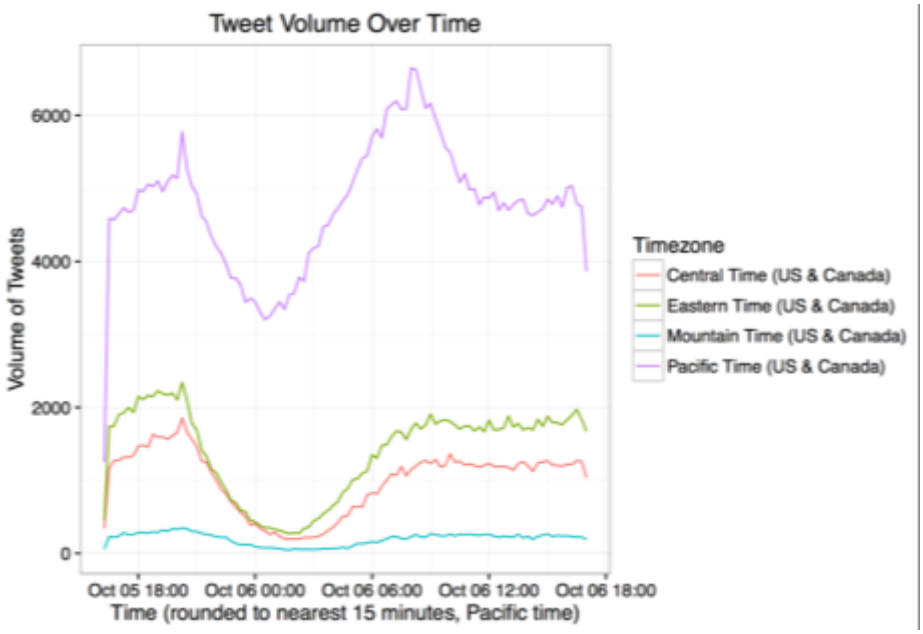

Figure 1.

Analysis of “Tweet Volume Over Time” Graph

As evidenced by Figure 1, all line plots have peaks (or local maximums) at 8pm. This 8pm peak is the global maximum for all plots except for that of Pacific Time, which has its global maximum at around 8am. The steepest decline in tweets occurs after 8pm, which terminates in the global minimum for all time zones at 1am. Then, we note the most significant rise in tweets from 1am to 8am for all time zones. With the exception of the Pacific Time Zone, the above graph shows a plateau in volume of tweets from 8am to 4 pm, which correlated to the typical workday. Additionally, the y-values of the different time zones suggest that there are relatively more Tweeters in the Pacific Time Zone, and relatively fewer in the Mountain Time Zone. We can also observe that the line corresponding to the Mountain Time Zone has less variation, while the line corresponding to the Pacific Time Zone has a larger variation than the others.

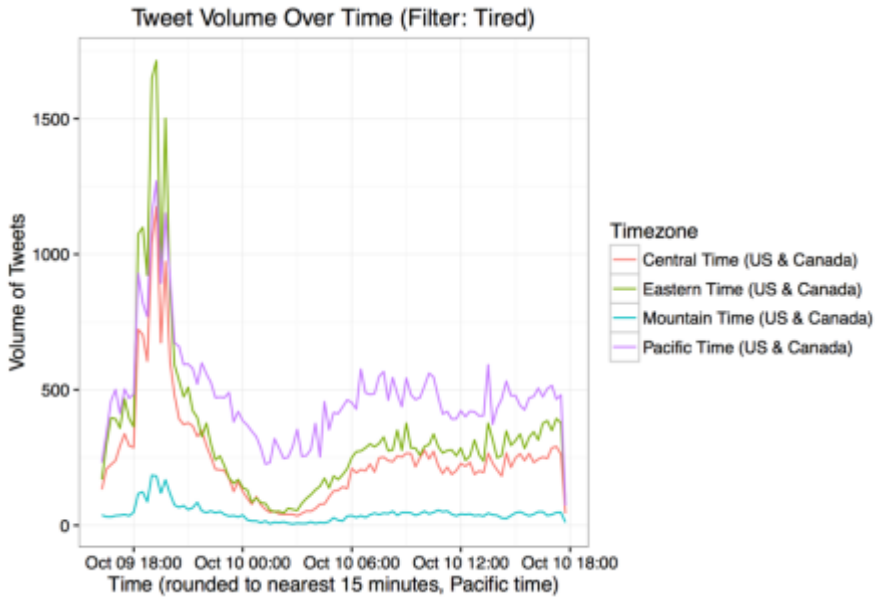

Figure 2

Analysis of filtered plot

As evidenced by Figure 2, the difference in volume between time zones is smaller than in Figure 1. From 5pm to 8pm we observe a steep increase in the volume of tweets containing the word 'tired' for all time zones. This graph suggests that people in the US start feeling tired around 5pm and the level of tiredness peaks at 8pm. The frequency of tweets with the word "tired" then drops dramatically between 8pm and 2am as Americans begin to go to bed. Then, between 2am and 7am we see a small but steady increase in the frequency of "tired" tweets, which correlates with people's wake up time. Similar to Figure 1, the frequency of tweets containing the word "tired" plateaus between 7am and 5pm, which corresponds to the typical American workday. It is important to note that the frequency of the word "tired" in tweets from the Eastern Time Zone surpasses that of the Pacific Time Zone between 6pm and 8pm, suggesting that a higher proportion of people in the Eastern Time Zone are feeling tired at this time. This point is emphasized when looking back at Figure 1, where the volume of tweets from the Pacific Time Zone exceeds that of the Eastern Time Zone by about a factor of 2. Lastly, we note that there is a proportionally larger variance in the volume of tweets containing the word "tired" in the Eastern Time Zone than in the other time zones.

Compare and Contrast

Figure 1 manifests large differences in the volume of tweets across time zones while Figure 2 shows smaller differences, especially amongst the Pacific, Eastern and Central Time Zones. Both figures contain peaks at 8pm – this suggests that the peak at 8pm for tweets containing "tired" is partially due to the sheer volume of tweets at that time, and not necessarily because of people being more tired. However, the slopes from 6pm to 8pm and 8pm to 1am are much more pronounced in Figure 2, suggesting that the increase is largely due to people's tiredness levels, and not just the sheer amount of increase in tweets. The Figure 2 Eastern Time Zone slope is especially pronounced in the filtered graph, suggesting that people in the Eastern Time

Zone tweet about tiredness proportionally more often than those in other time zones. In regards to the Pacific Time Zone plot in Figure 1, we note a steep slope leading up to a global maximum at 8am. These characteristics are not present in Figure 2; instead, it plateaus. This suggests that the proportion of tweets containing the word “tired” is actually markedly low around 8am. There are two effects that contribute to the volume of tweets that contain the word “tired”: (1) the total number of tweets created, and (2) the proportion of people tweeting the word “tired”. To get a better idea of how these factors interact, we created another graph (Figure 3) that is used to compare the proportion of “tired” tweets across time zones, adjusting for the total number of tweets created.

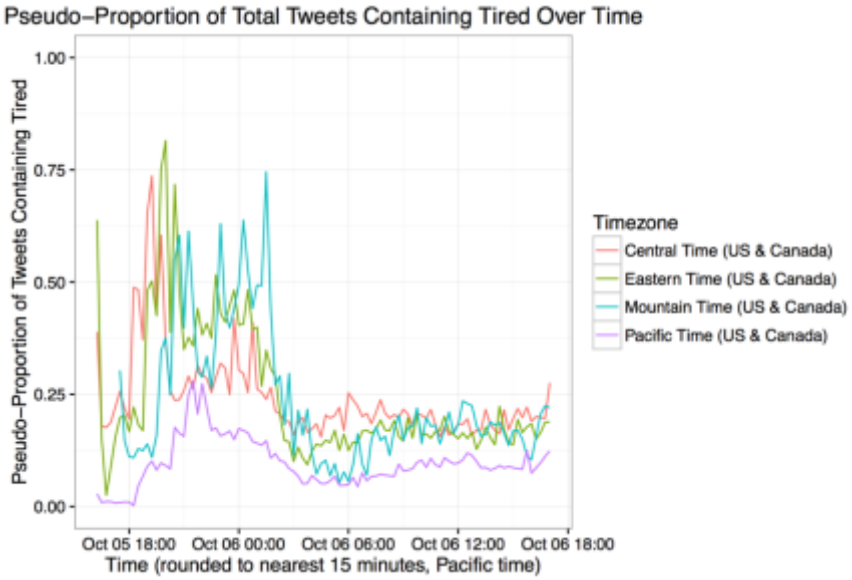

Figure 3

Here, we used the word pseudo-proportion because while it’s not the actual proportion, it gives us a way to compare Figures 1 and 2 within the same graph (calculated as the number of “tired” tweets divided by the number of total tweets at each time point). Figure 3 shows that between 7pm and 2am there is a high proportion of “tired” tweets, especially for Central, Eastern, and Mountain Time Zone lines. This suggests that the peaks in Figure 2 in these time zones is actually due to the proportion of tweets containing the word “tired” than just the absolute amount of tweets in Figure 1. Additionally, the low plateau of tweets from 2am to 7pm suggests that the proportion of tweets containing the word “tired” is not as impactful. Furthermore, as the Pacific Time plot has the lowest pseudo-proportions, we can infer that the proportion of people feeling tired is relatively smaller than in other time zones.

Limitations

As we took a random selection of tweets from Twitter and not the actual whole API repository, we cannot be certain that the representation of tweet fluctuation and distribution between time zones is accurate. Furthermore, since we collected tweets by filtering for the word “tired”, we cannot be certain if the proportion of tweets containing the word “tired” truly represents

the proportion of people feeling tired in each time zone. The set of tweets containing the word "tired" could also contain tweets about people not being tired, or using tired in a different way (e.g. people being "tired of Trump" or "tired of cats"). Although Figure 3 helps us understand the true level of tiredness across the country, it is ultimately difficult to distinguish if we can interpret variation in volume of "tired" tweets as how tired Americans are feeling.