Part2 Analysis Part1

2.1.1 Task Description

In this part we were focusing on the original data set, trying to retrieve the insight of the data set. What information can we obtain from the data set, what’s the overall percentage sentiment (negative, neutral, positive percentage), the percentage of tweets per airline, what is the Proportion of negative, neutral or positive sentiment tweets per airline, what words cause the

negative sentiment tweets, tweet location exploration and the tweet time zone study.

2.1.2 Techniques and data flow

We introduced Apache Pig to retrieve information from the data set, and Tableau to visualize the result

2.1.3

1. overall sentiment analysis

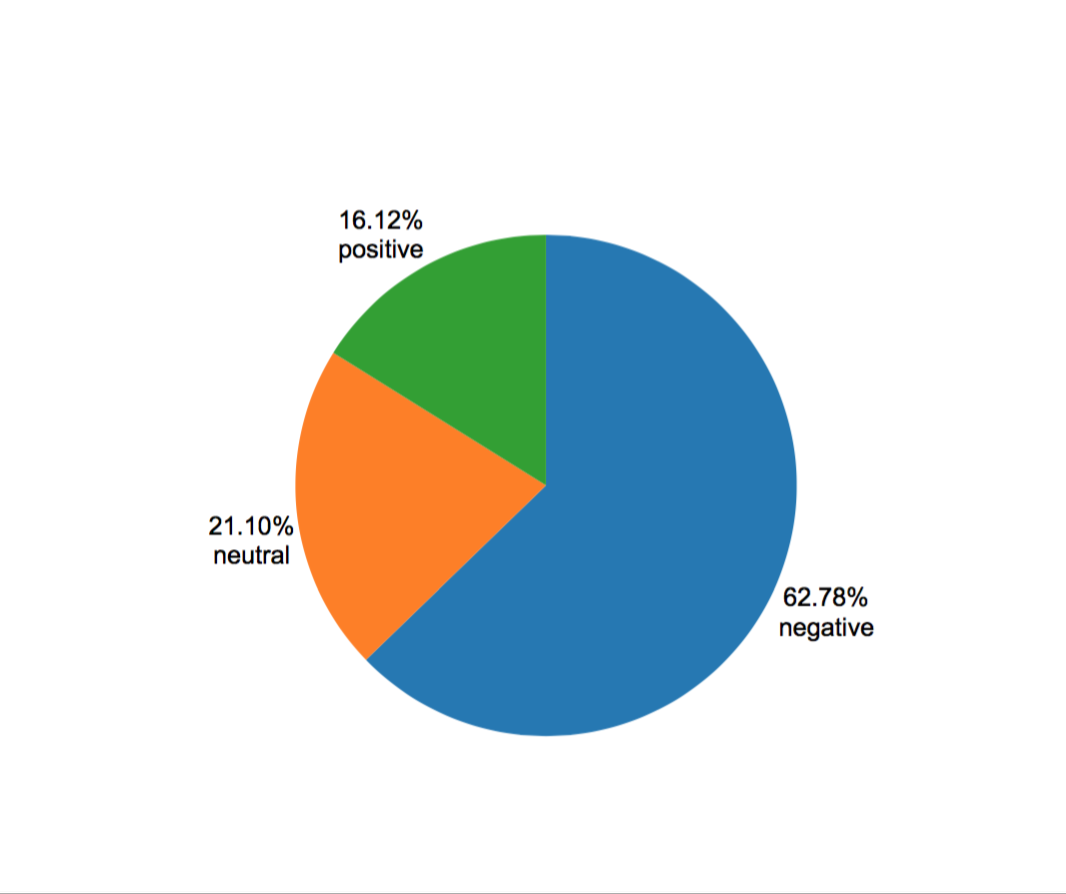
This part we are focusing on the proportion of overall tweets

Figure1. Overall sentiment

We can see from above that negative tweets take the lead which has 62.78% of the total tweets.

2. Percentage of tweets per airline

Now let’s look at each airline, which airline has most tweets

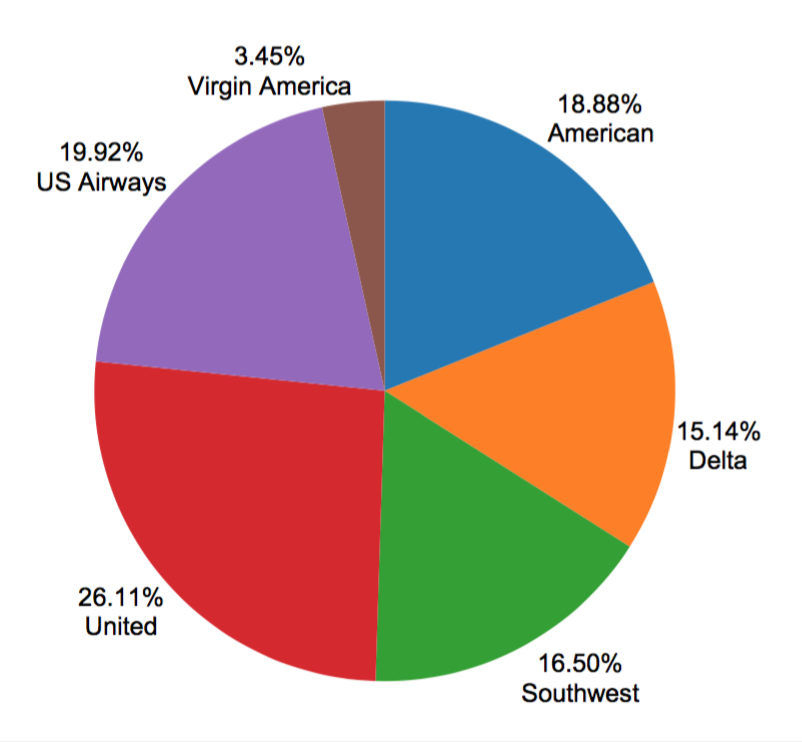


figure2. Percentage of tweets per airline

from the chart above we can see that United Airline has the most tweets regarding their services, US Airways takes the second place which has 20% tweets, and then American Airline in the third place with 19% tweets.

3. Proportion of negative, neutral positive sentiment tweets per airline

Now let’s look at the details:

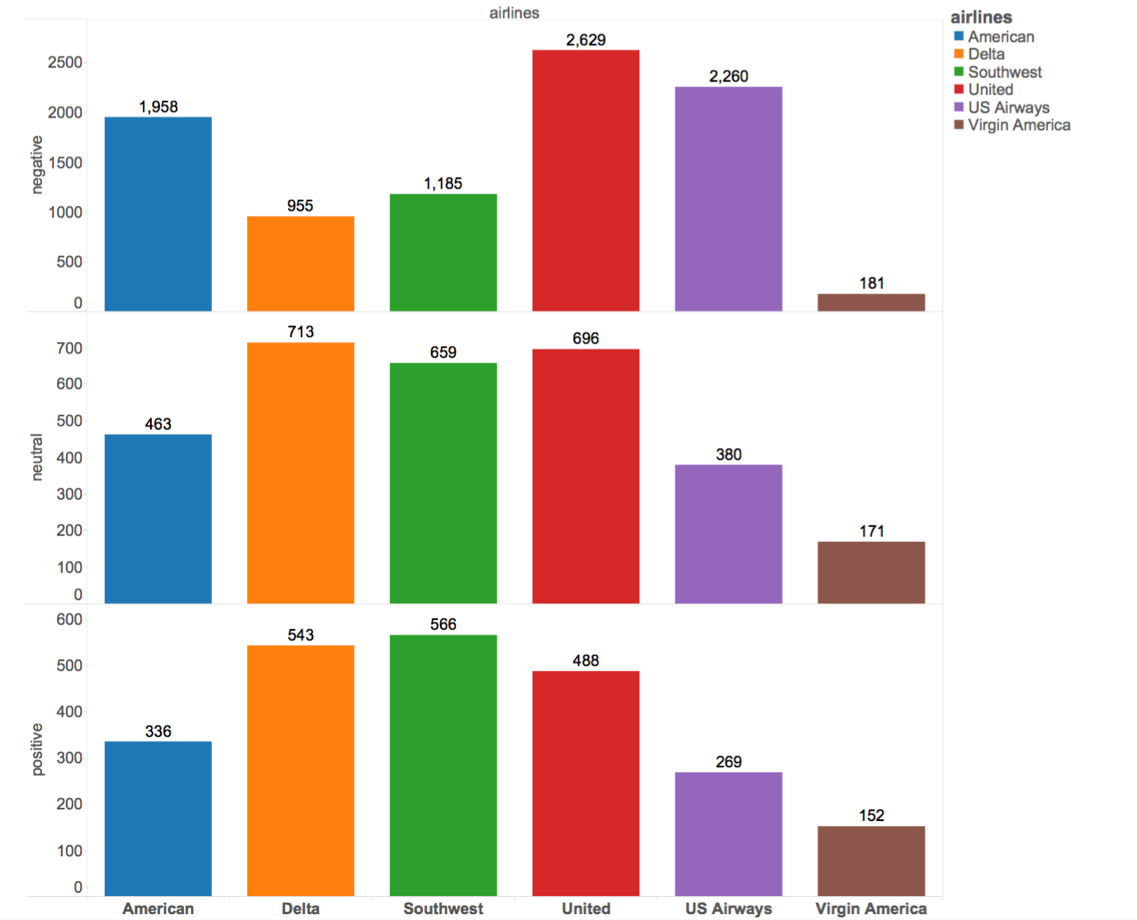


Figure3. Proportion of negative, neutral positive sentiment tweets per airline

Observing from the output, we can see that United Airline received the highest negative tweets which has 2629 out of 14500 tweets about their services, and Southwest has the highest positive tweets.

Also, an interesting point is that all companies have negative tweets more than positive tweets

we can see graph below

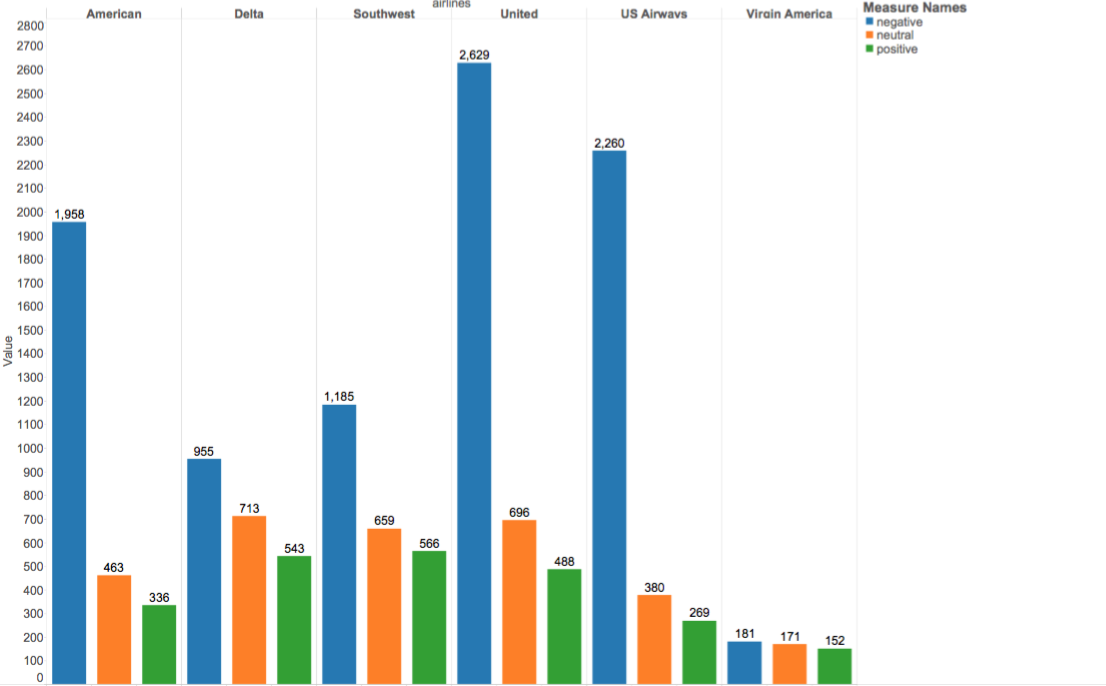


Figure4. Proportion of negative, neutral positive sentiment tweets per airline

4. Negative sentiment

what kind negative sentiment that most twitter user have been talking about? In this part we are focusing on negative sentiment

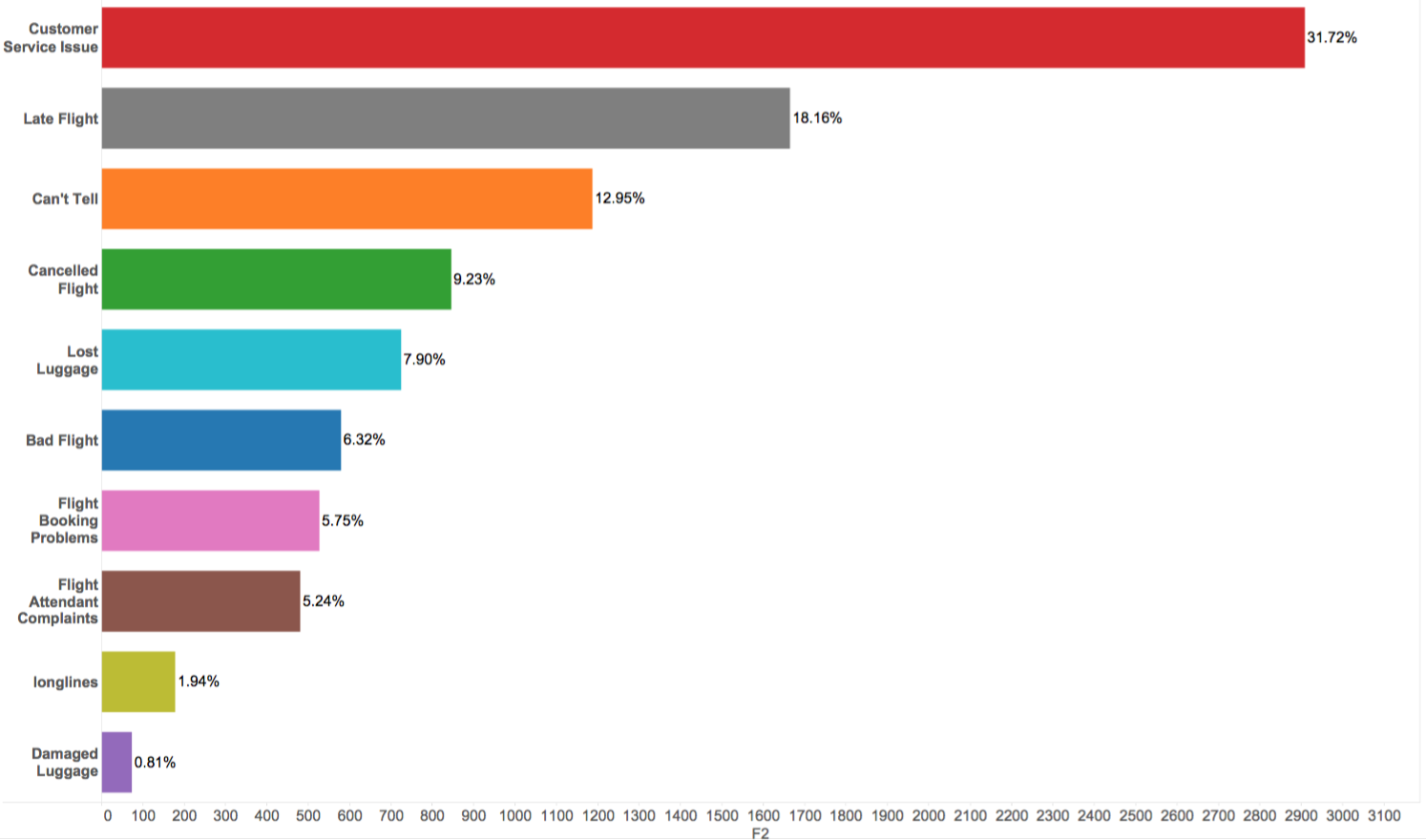


Figure5. Negative Sentiment

We can see from above that most tweets are about customer service issue, and light delay is the second important negative sentiment

5. Location Study

Now we move to the fact that where are the tweets coming from, which people in which city was tweeting about airline. Observing from the map below we can see that most tweets are coming from east and west region, and Chicago is another region has most tweets.

Airline companies should focus on flights between west and east.

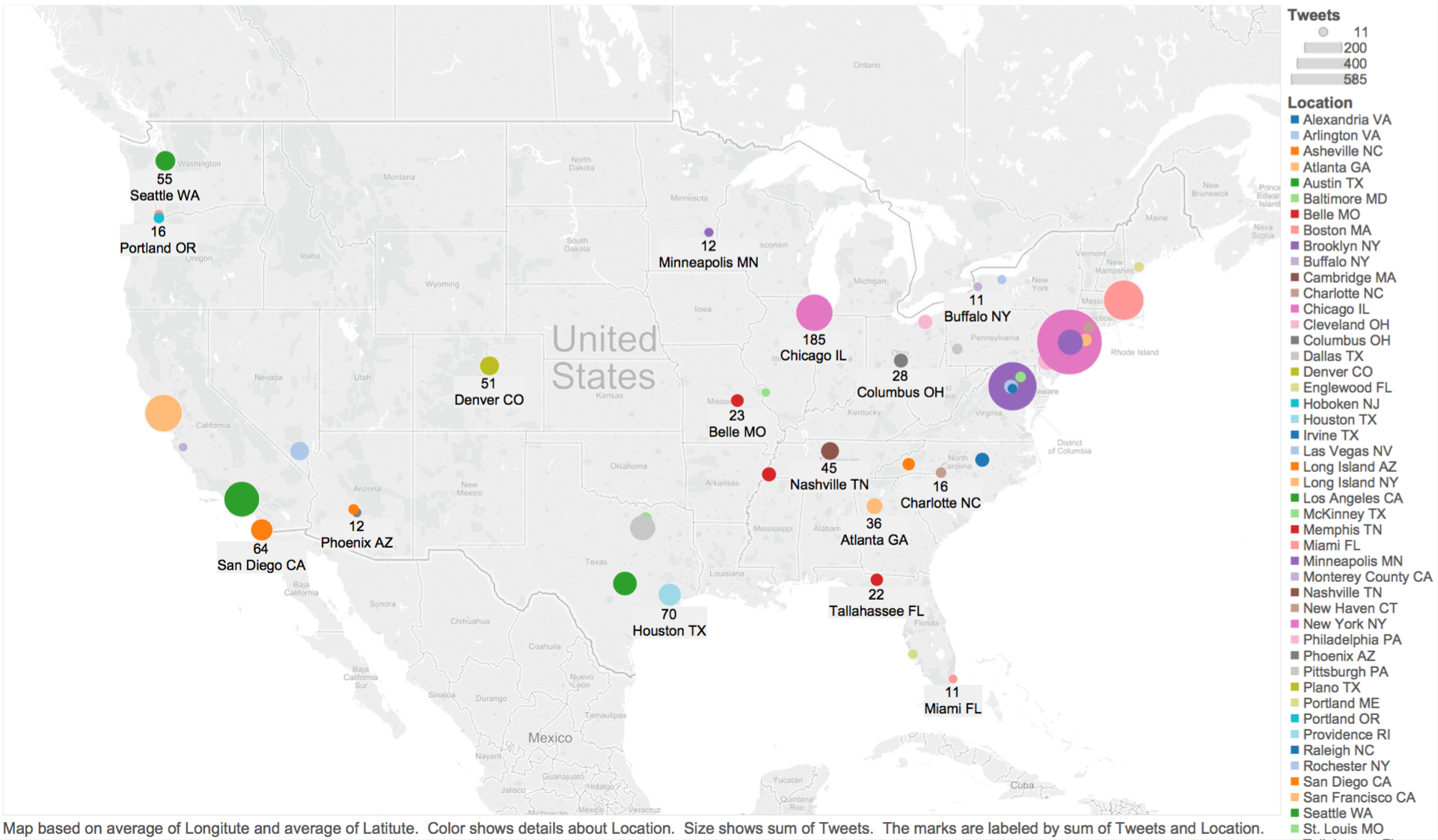


Figure6. Map Visualization