

EDA-Innovation-Naresh

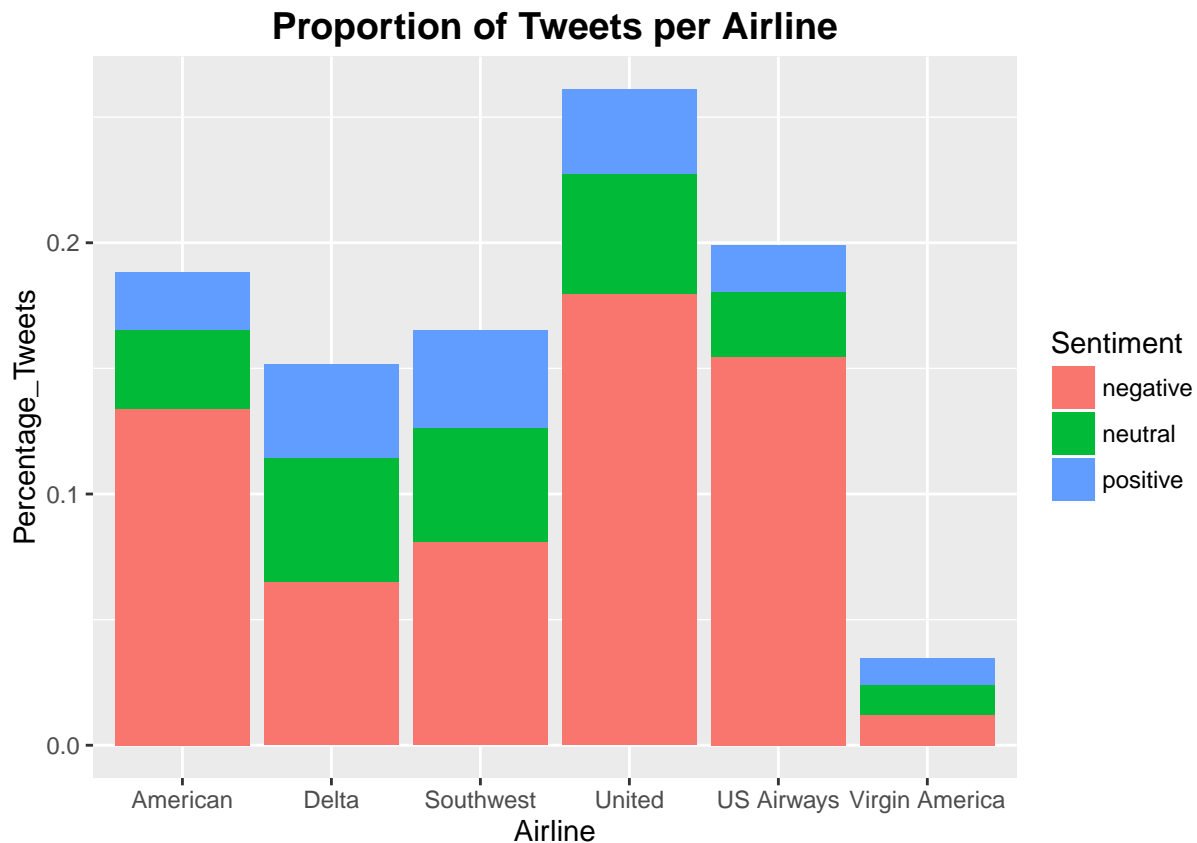
Naresh

April 29, 2016

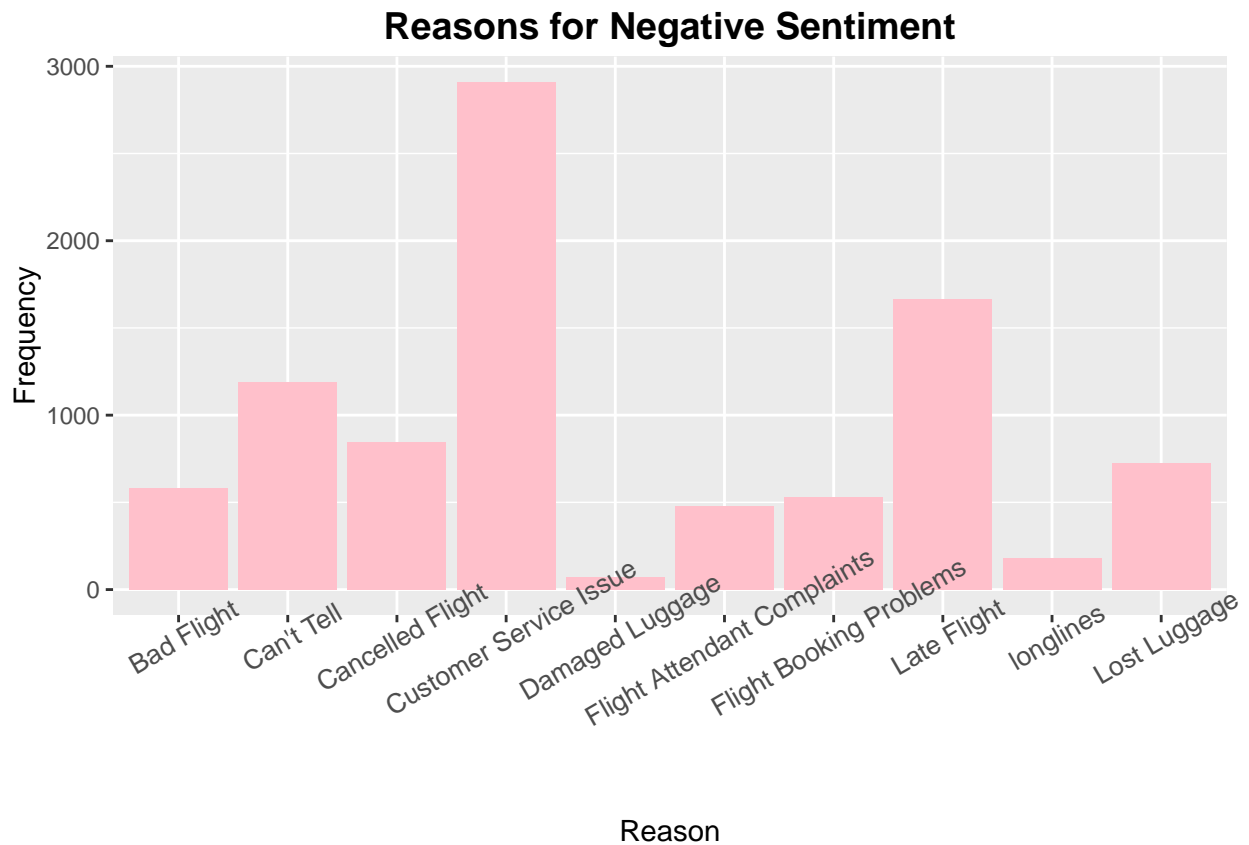
Dataset Selected: Twitter US Airline Dataset

Data has been loaded!

```
##
##           American      Delta  Southwest      United US Airways
## negative 0.13387978 0.06523224 0.08101093 0.17984973 0.15457650
## neutral  0.03162568 0.04938525 0.04535519 0.04760929 0.02602459
## positive 0.02295082 0.03715847 0.03893443 0.03360656 0.01837432
##
##           Virgin America
## negative      0.01236339
## neutral       0.01168033
## positive      0.01038251
```



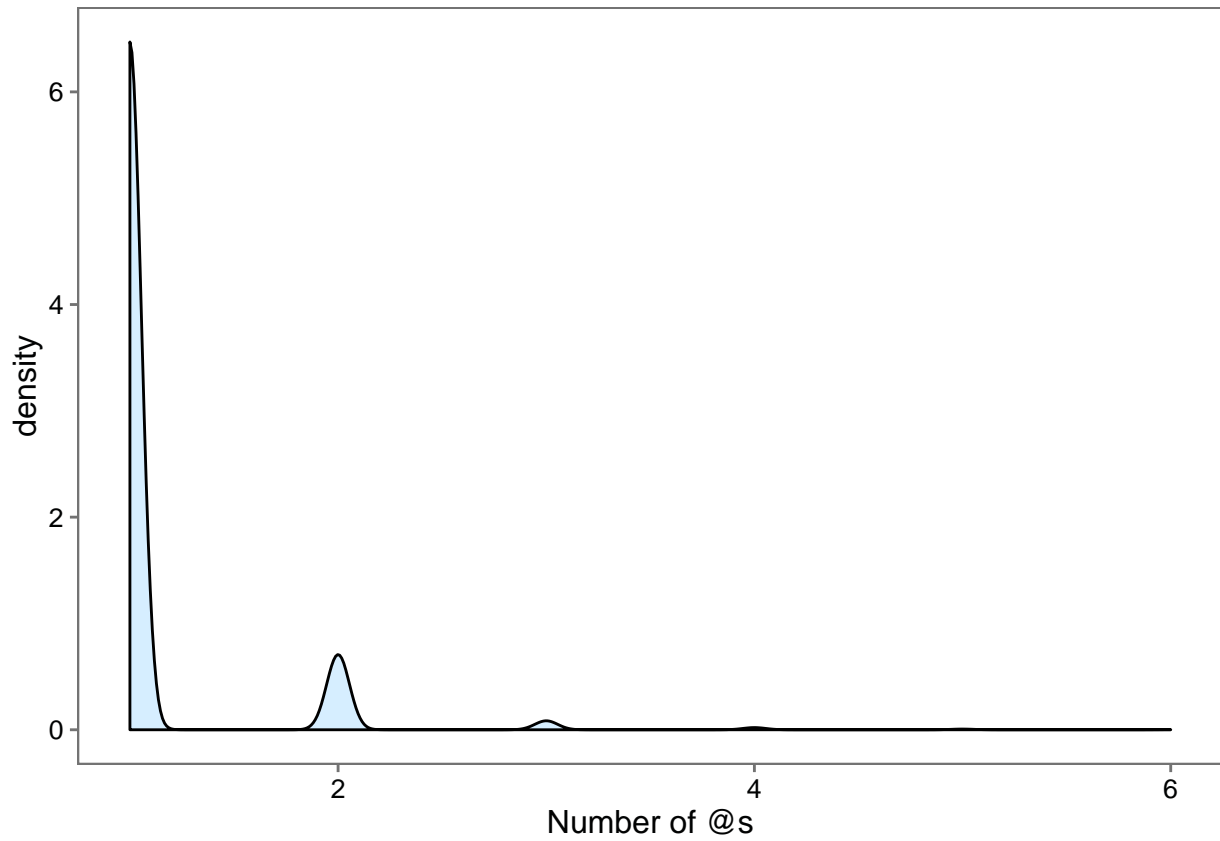
Clearly it seems that United Airlines needs to work on improving its services. It receives far too many negative tweets.



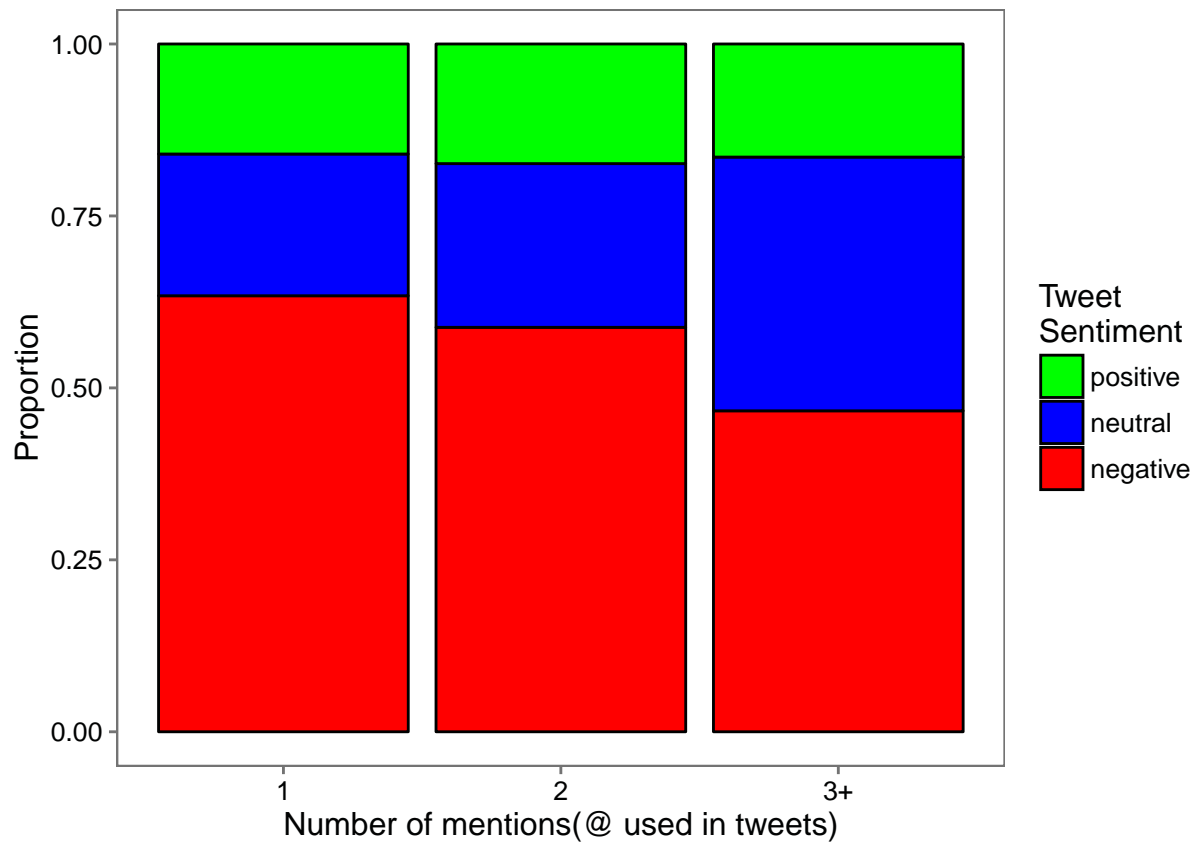
Digging deeper it is obvious that Customer Service is a pain point for most of the negative tweets.

We now prepare the data with respect to user mentions within the tweets. Perhaps another aspect to look at is a focus on the customers who are more likely to spread information via twitter as a social network. We will look at this from two points of view: Via user_mentions and via retweet counts. These customers(with high number of user_mentions) are likely to be central figures or atleast well connected entities within the social network.

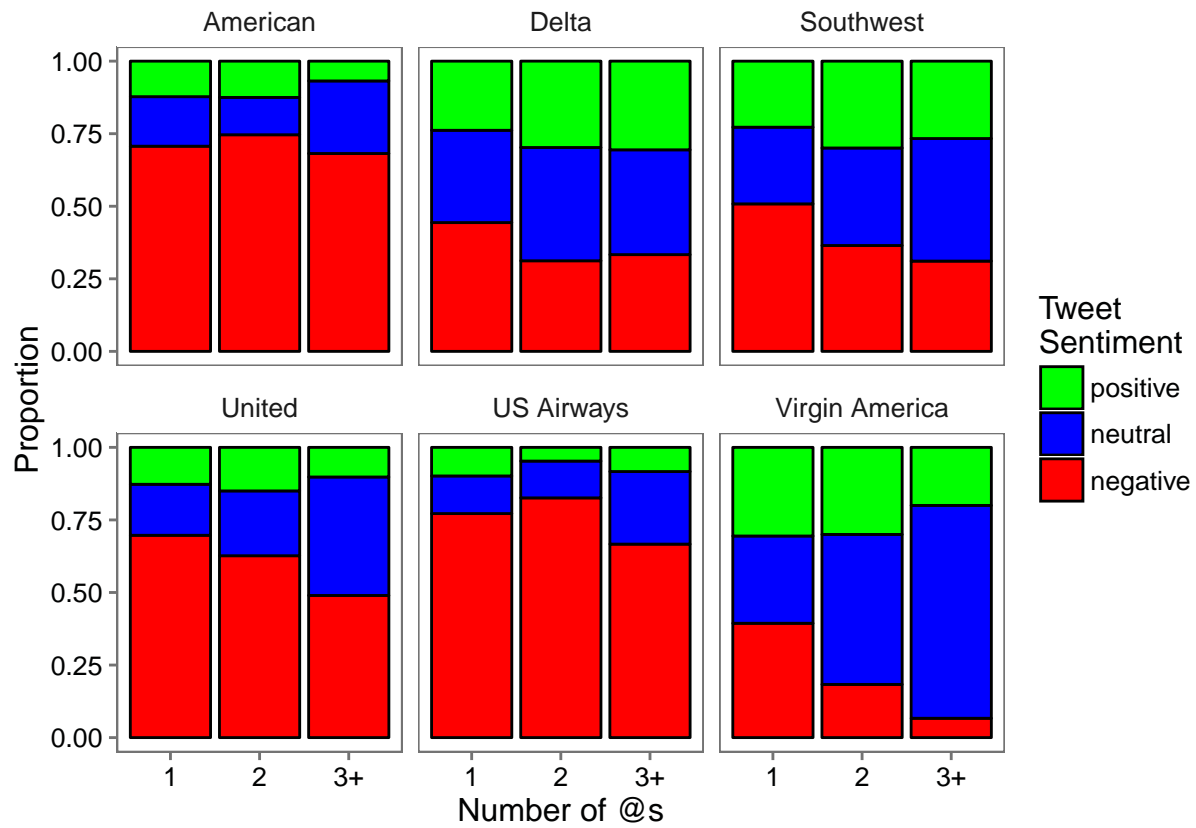
Here's a basic visualization of how many people were mentioned in the twitter dataset and the density distribution of user_mentions.



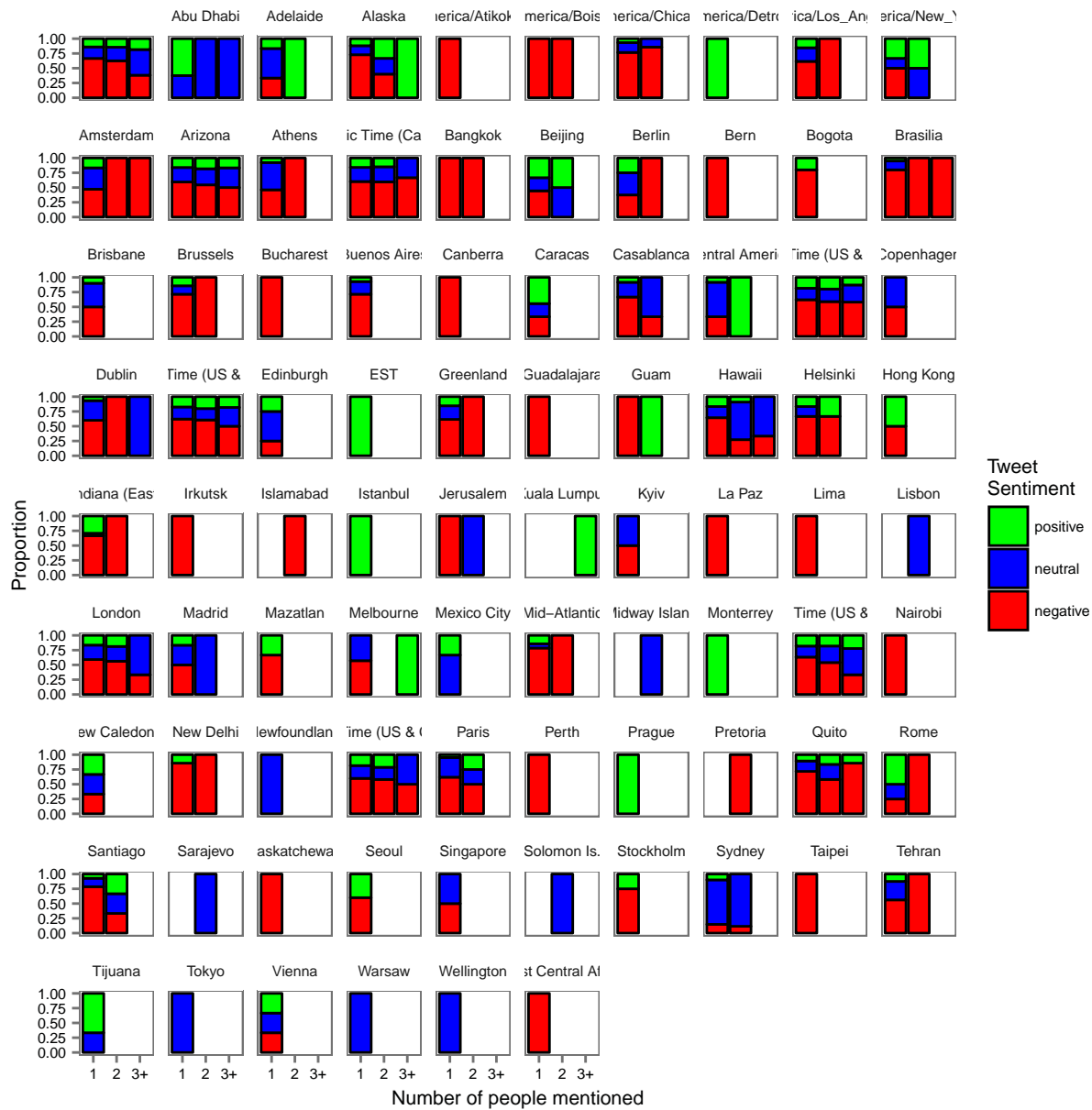
Setting a fixed color of red for negative, blue for neutral and green for negative sentiments in tweets. We look at the correlation between the negative sentiment tweets and how they spread(as a function of user_mentions)



Now we try to look at whether there is a significant difference between airline with respect to how they treat influencers(people with high number of user_mentions in their tweets)?

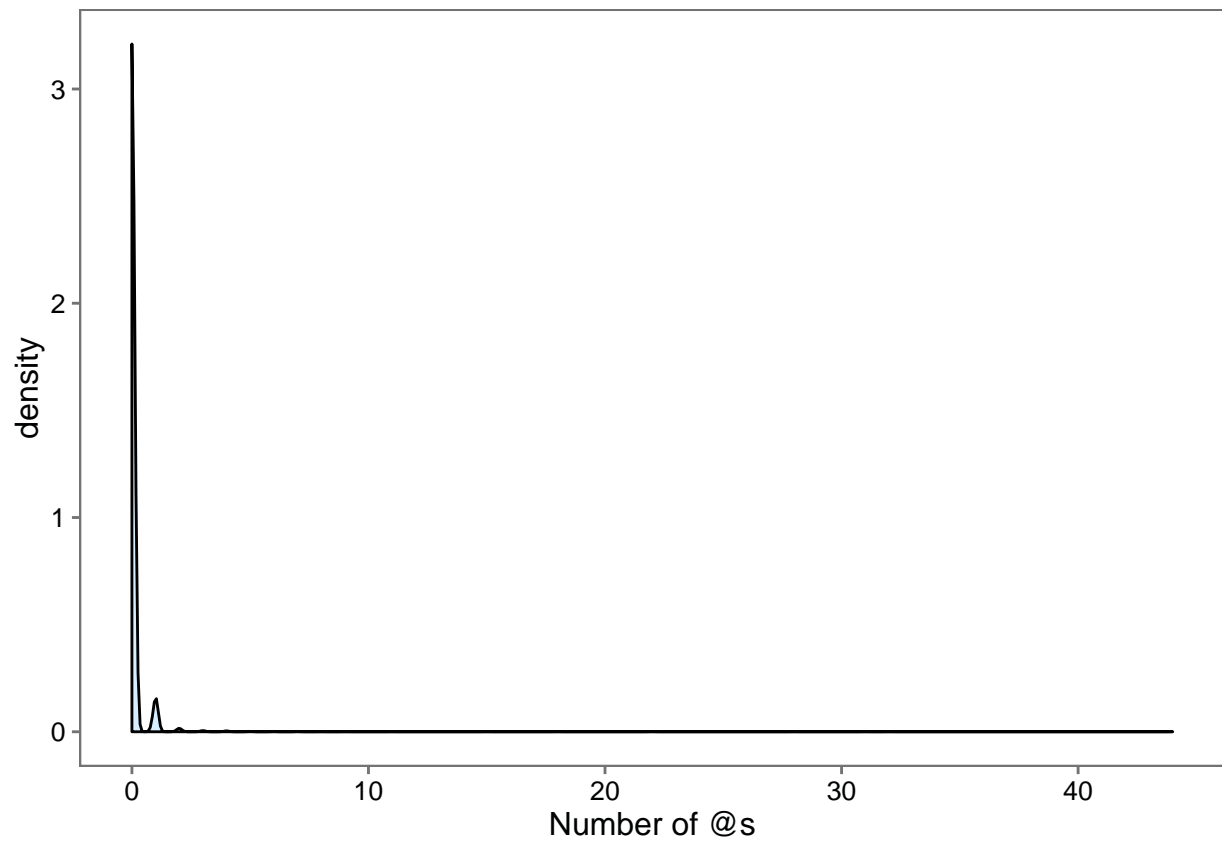


It seems like the airlines do indeed treat the people who are influencers much better on an average. With the only exception being Delta airlines and even that with a very low difference.

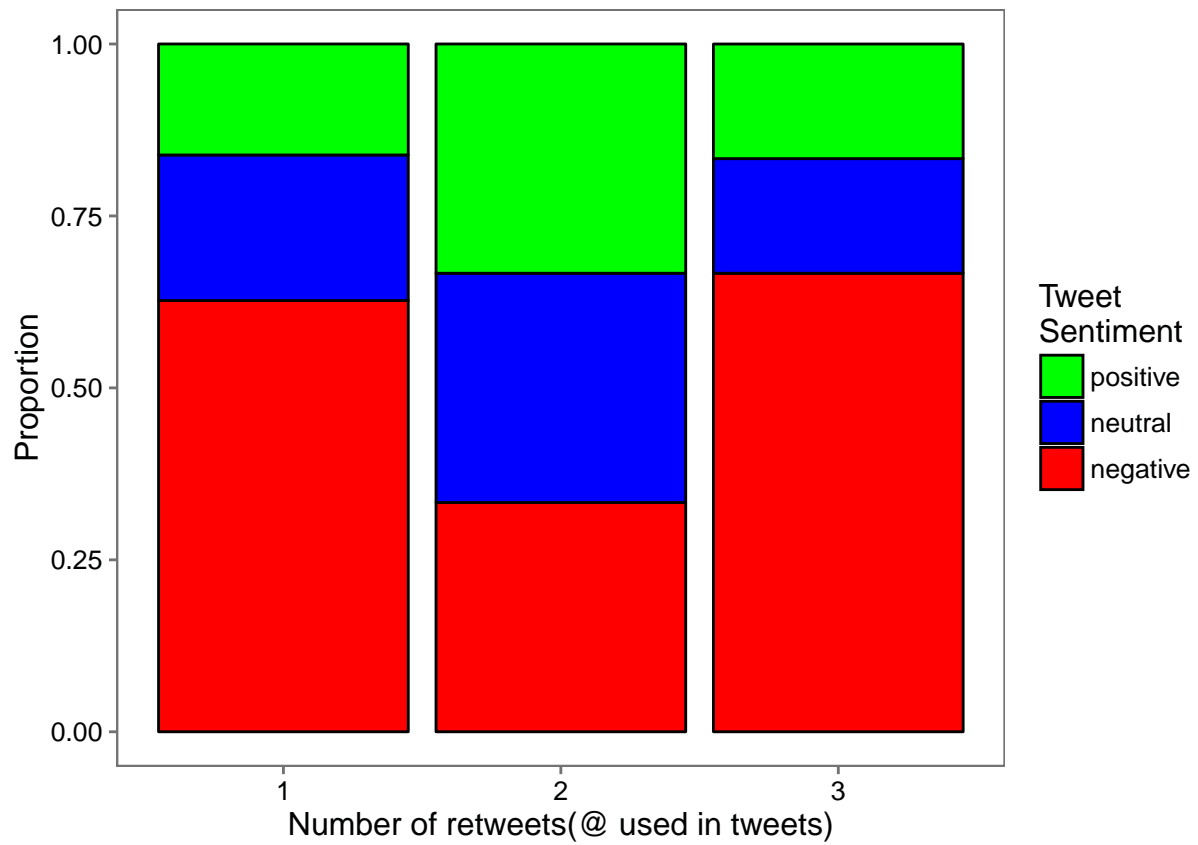


Is the number of user mentions dependent on where the user tweeted from? We assume here that users tweet from their home city. Therefore if this assumption holds true, we can understand two things: The general attitude of people from certain places (generally positive/negative/neutral tweets). This will allow the airline companies to form a matrix to cater specifically to people from certain countries and focus on treating them better if they form a significant amount of their userbase. This can also allow companies to re-evaluate their crew performance on particular flight routes.

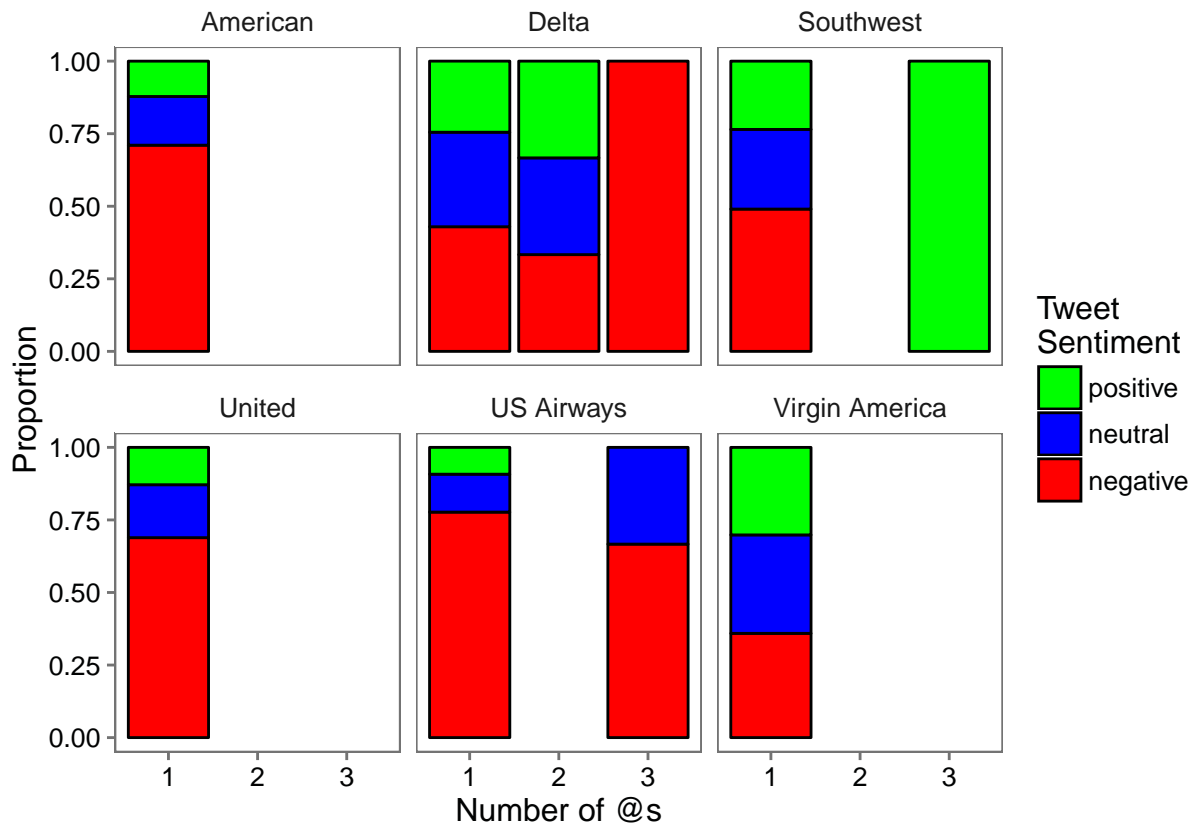
Second, companies can perhaps focus a bit more on routes where the information spread is far more quicker or rather cities where a lot more people are likely to receive poor sentiments about the airline if there is a negative tweet about the airline.



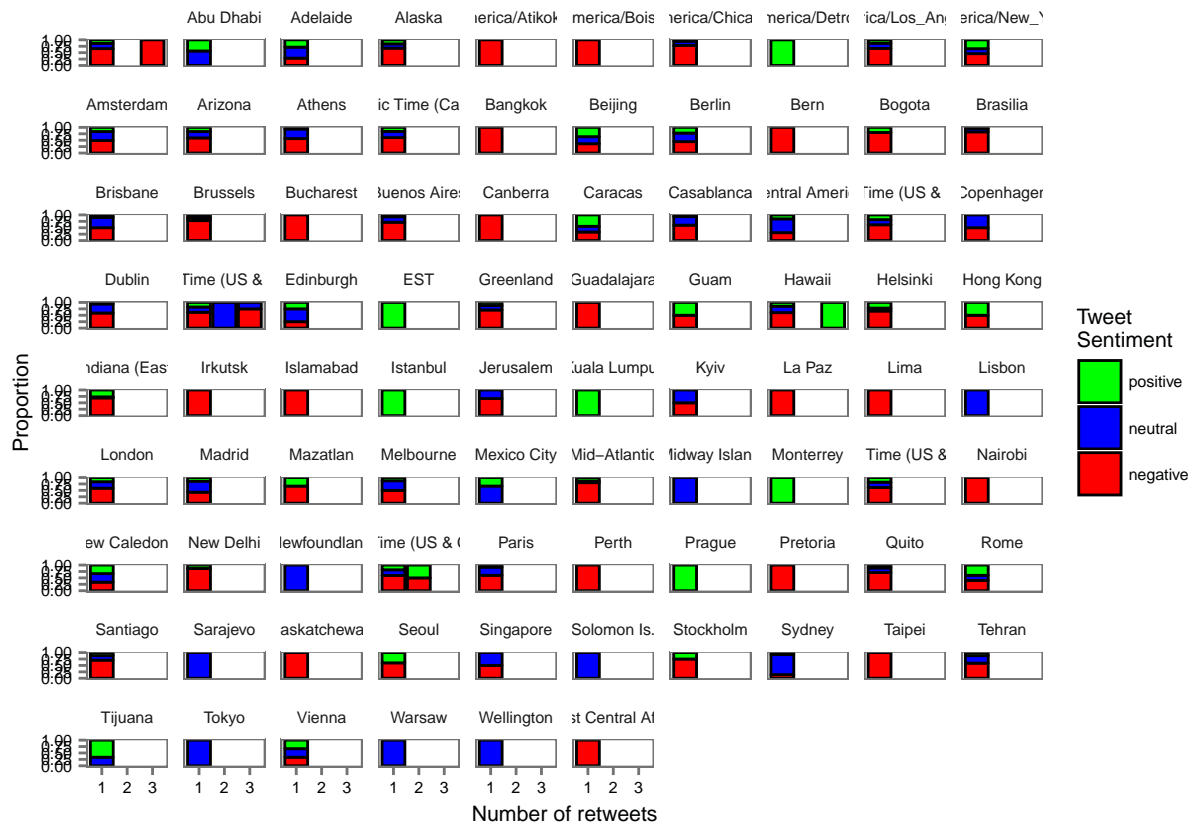
If a post had a high retweet count, it means that it was likely to have been a common problem/point of pain for many passengers.



Here 1 indicates low, 2 indicates medium and 3 indicates high retweet count.



We look at the retweet count per airline and try to see if there are any significant differences.



There seem to be some retweets which received a lot of retweets. We look at why these tweets struck a chord with the community.

```
##
```

```
##      0      1      2      3      4      5      6      7      8      9     11     15
## 13873  640   66   22   17    5    3    3    1    1    1    1
##     18   22   28   31   32   44
##      1      2      1      1      1      1
```

```
## [1] "@US Airways 5 hr flight delay and a delay when we land . Is that even real life ? Get me off this
```

```
## [1] " "
```

```
## [1] "@US Airways of course never again tho . Thanks for tweetin ur concern but not Doin anythin to fi
```

```
## [1] " "
```

```
## [1] "STOP. USING.THIS.WORD. IF. YOU'RE. A. COMPANY. RT @JetBlue: Our fleet's on fleek. http://t.co/F
```

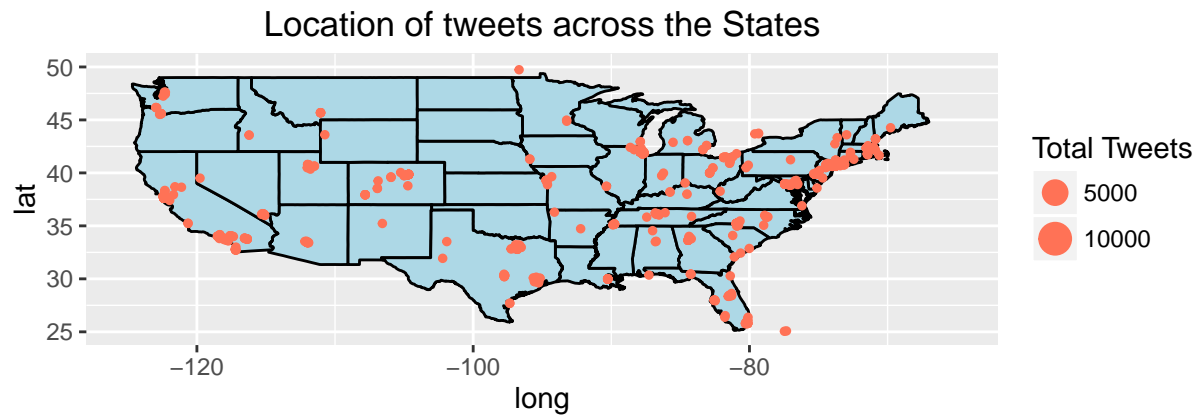
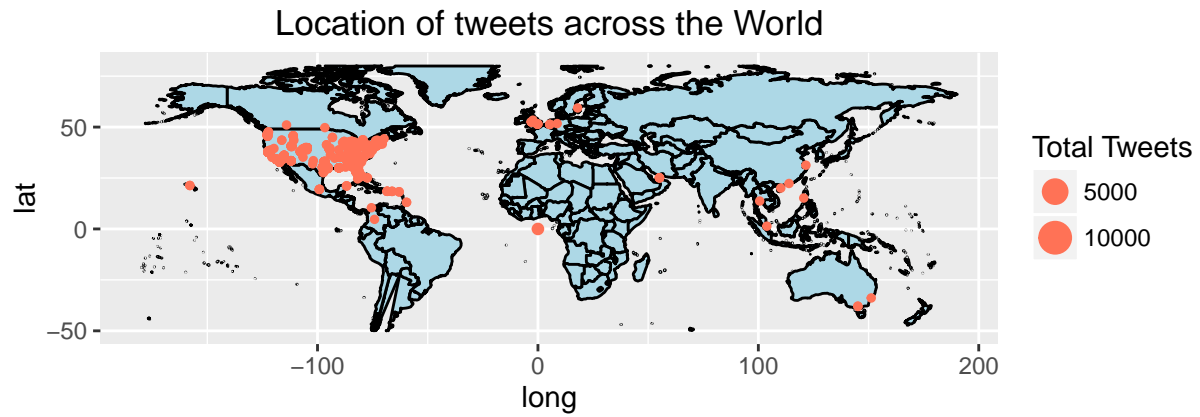
```
## [1] " "
```

```
## [1] "@US Airways with this livery back in the day. http://t.co/EEqWVAMmiy"
```

3 of these tweets were negative issues about problems faced by a particular influencer as a customer. The 4th article is one on nostalgia. It is quite obvious but perhaps leads to advertising ideas to build brand loyalty for the airline company by reminiscing nostalgic events from the past.

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 57 rows containing missing values (geom_point).
```



Lastly, we look at tweets from across the world and the across the states and try to find major points where people are likely to comment or communicate more to the airlines about what they want/expect from the airline companies.