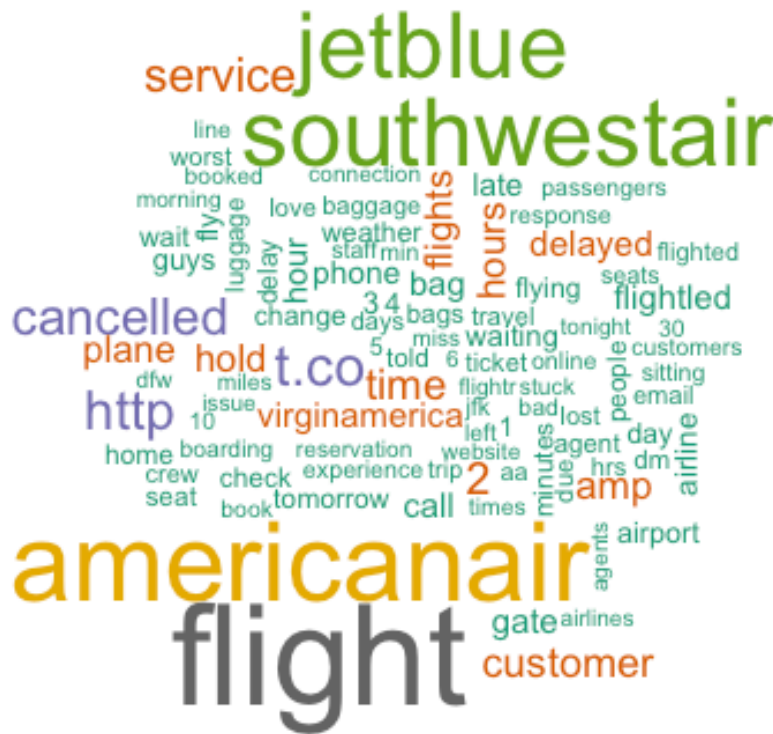


Youbeen Shim
Prof. Jordan Rodu
2017.12.12

Word Cloud

Part 1



Notes

What does `unnest_tokens()` function do?

The `unnest_tokens()` addresses a lot of the common pitfalls of word clouds while making it possible for R to produce a word cloud. First, it takes the data in the `tweets.Rdata` file, and parses so that we can obtain the items we want -the individual words. Deriving from the resulting word cloud, it also removes capitalizations and punctuations so that there are no repetitive/overrepresentation of words.

Variables in the dataset

*Note: Columns not particularly relevant or not complete is left out.

tweet_id: ID of the twitter who wrote the tweet.

airline_sentiment: Sentiment towards the airline. Three levels of positive, negative, and neutral.

airline_sentiment_confidence: Confidence of the sentiment

airline: Type of airline
name: User name
text: Physical text on the tweet
tweet_created: Time the tweet was created
tweet_location: Location of the tweet, approximate

Interpretation

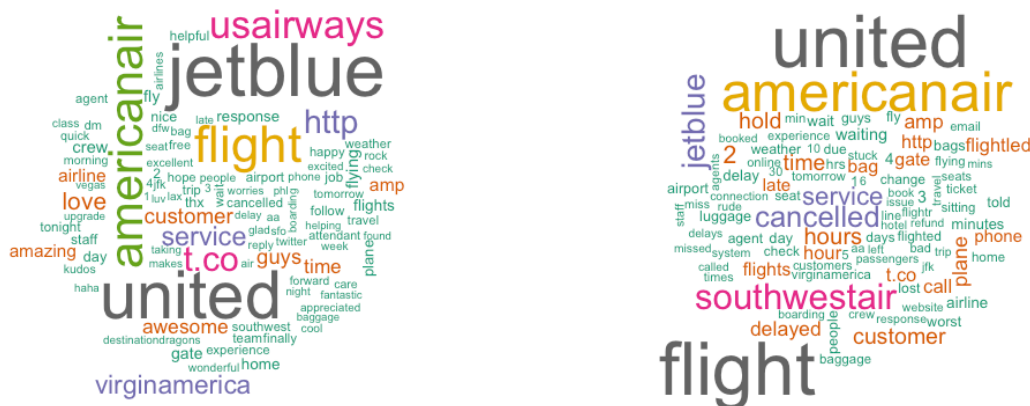
Going off how the word cloud is presented and my knowledge of the variables that are collected in the data, few interpretations can be made -but all none with guarantees. Based the the airline_sentiment data that is recorded, it appears that the word cloud goes beyond just calculating the frequency but also takes in to account the sentiment towards the airline. I cannot observe how the sentiment data is represented in the word cloud, however, color -especially red and green- is suspicious but ultimately vague. While the word cloud represents what words are said and how frequent they are said, too many confounding variables exist to tell an effective story.

Part 2

Airline Sentiment:

Positive

Negative



A simple observation reveals that tweets carrying a positive sentiment mentions jetBlue, United, AmericanAir, and USAirways while tweets carrying a negative sentiment frequently contains again, United and American Air, accompanied by Southwest Airlines and jetBlue. At the smallest level, the green words are general words that are found along with the sentiment. Intuitively, words such as awesome, love, or quick follows the positive and words such as delay or hours follows the negative sentiment. While the word cloud suggests that airlines such as jetBlue is a good airline, airlines such as Southwest Airlines is a negative airline. However, the presence of the United Airlines and American Air in both word clouds undermine the point that such interpretation makes. It is also important to note that standards, target consumer, and expectations all vary from airline to airline and that ultimately, any interpretation should be left for a formal analysis.