

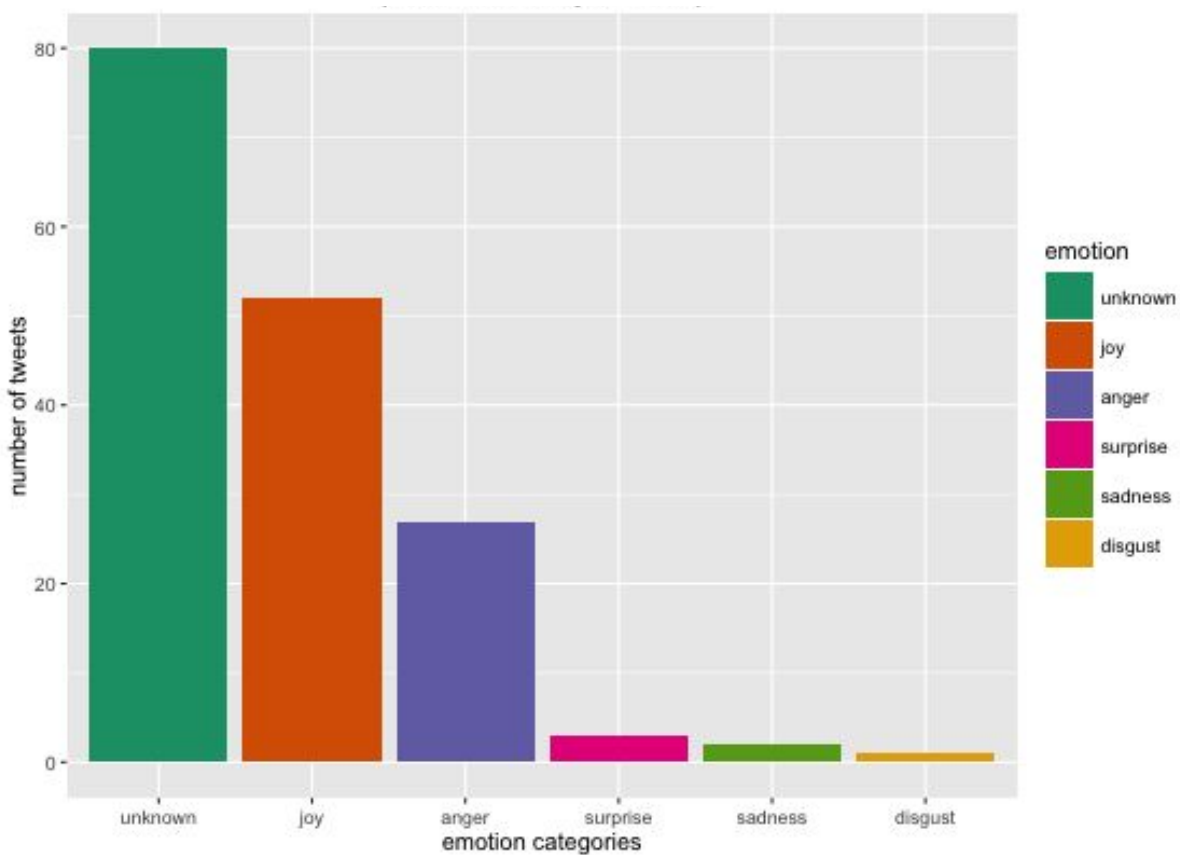
Problem 4: Statistical analysis to support new data product

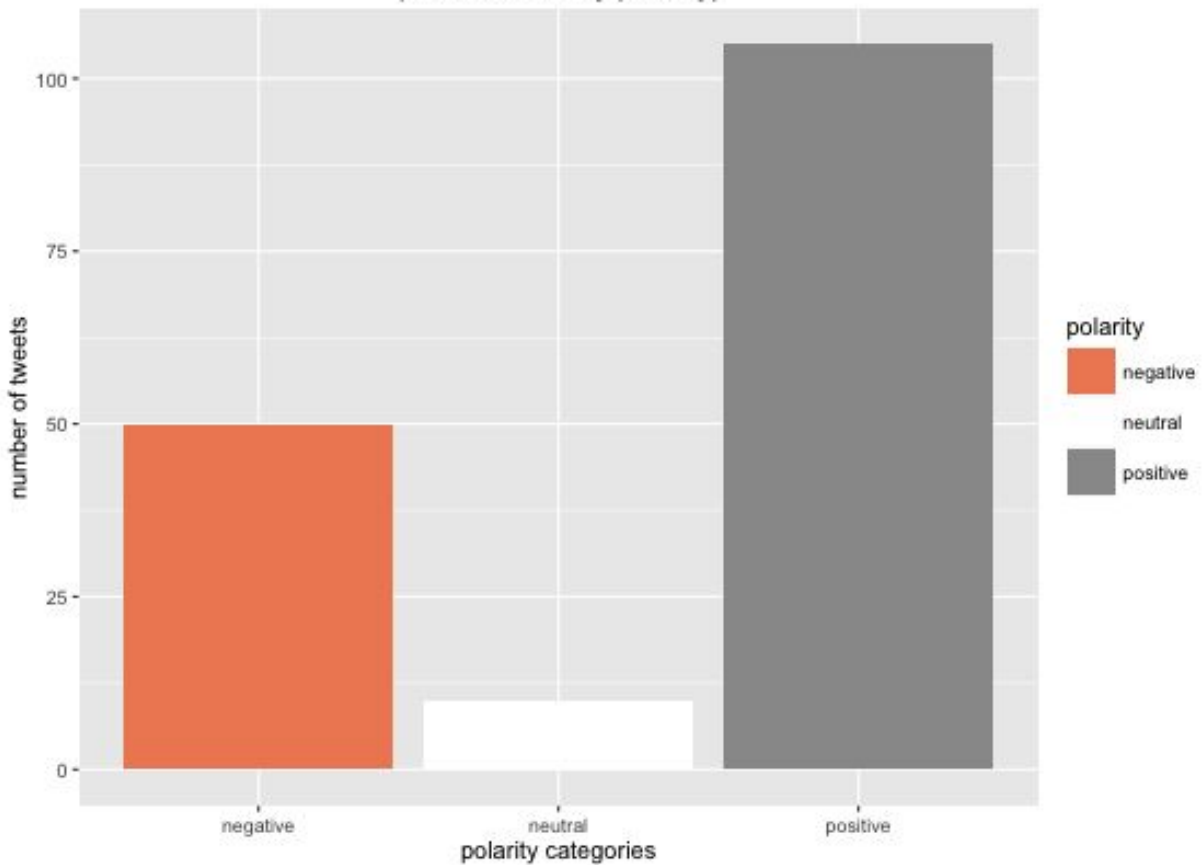
Real Direct is an online vendor where users can buy or sell apartments and contact agents in New York city. First of all, New York being the financial capital of the country offers many jobs to people, hence a lot people migrate to New York in search of jobs. Second of all, since many people migrate to New York it's becoming crowded day by day, hence people are starting to move to outer parts of New York like Jersey City after living temporarily for some time.

Considering above two points, Real Direct can improve its profits by launching new product for apartment rental. To support this, analysis of users' view has been performed through tweets being posted. Tweets from 02/23/2016 to 03/4/2016 has been collected and a sentiment analysis has been performed (getting help from the code posted by Colin Priest which can be viewed at

<http://colinpriest.com/2015/07/04/tutorial-using-r-and-twitter-to-analyse-consumer-sentiment/>).

By analysing this data it is inferred that a most of the people talk about renting an apartment in a good way. This can be summarised by following plots:





Apart from users' view, to support apartment rental purchase cost of New York homes has also been analysed. Purchase cost of most of the apartments sold lies between \$300,000-\$700,000, average being \$595,121 which is almost double the average of the houses sold in the whole nation i.e. \$321,200 (according to a report by US government which can be viewed at <https://www.census.gov/construction/nrs/pdf/uspricemon.pdf>). Following is the plot showing the price range vs frequency:

