CSE 587: Project 1
Problem 1 -Data Acquisiton

Lalith Vikram Natarajan, UBIT ID: lalithvi, Person No: 50169243

March 5, 2016

1 Objective

Collect data and import it in R

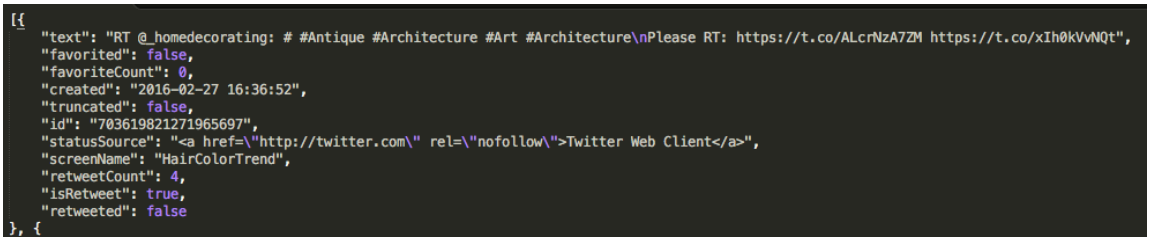
- Collect data of interest from Social Media (Twitter)
- Import saved json script

2 Data Source

Twitter data was used as the source of data for this exercise. I came up with a list of keywords that would help narrow the search down. They are real estate, housing, residential, realtor, house, foreclosure, condos, apartments, fsbo, buy home., etc. New York, NYC, Manhattan were the other words added along with the above mentioned words. Also, a search was made with the location set to New York's co-ordinates.The data was steadily collected over 10 days. But in retrospect, it would have made more sense to use Facebook for data retrieval since more people list their properties over there.

3 Data Structure

twitteR package was used to collect the tweets. This package had a method called searchTwitter which accepted parameters like text, number of maximum tweets per call, location, since and until. This function returns a data frame of results. This is then converted to a JSON for easy access and retrieval. For retrieval, we used a fromJSON method specified in the jsonlite package. The Twitter API was accessible by setting up an OAuth using the setup_twitter_oauth method found in the same twitteR package.



4 Period of Time

The tweets were collected from **February 23rd, 2016 - March 5th,2016**

5 Implementation

twitteR package provides a wrapper around the Twitter API and is flexible enough to allow searching using keywords, location, since date and until date. Duplicate tweets were removed by comparing existing tweets. A cronjob was written to activate this every hour and this helped in collecting the specified no. of tweets over a period of time.