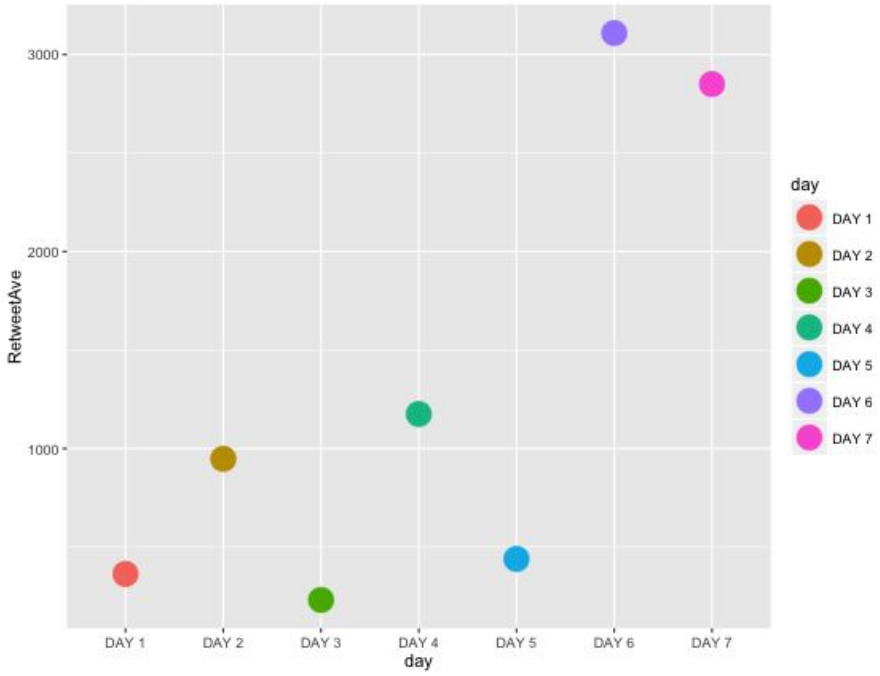
Problem 1: Data Acquisition

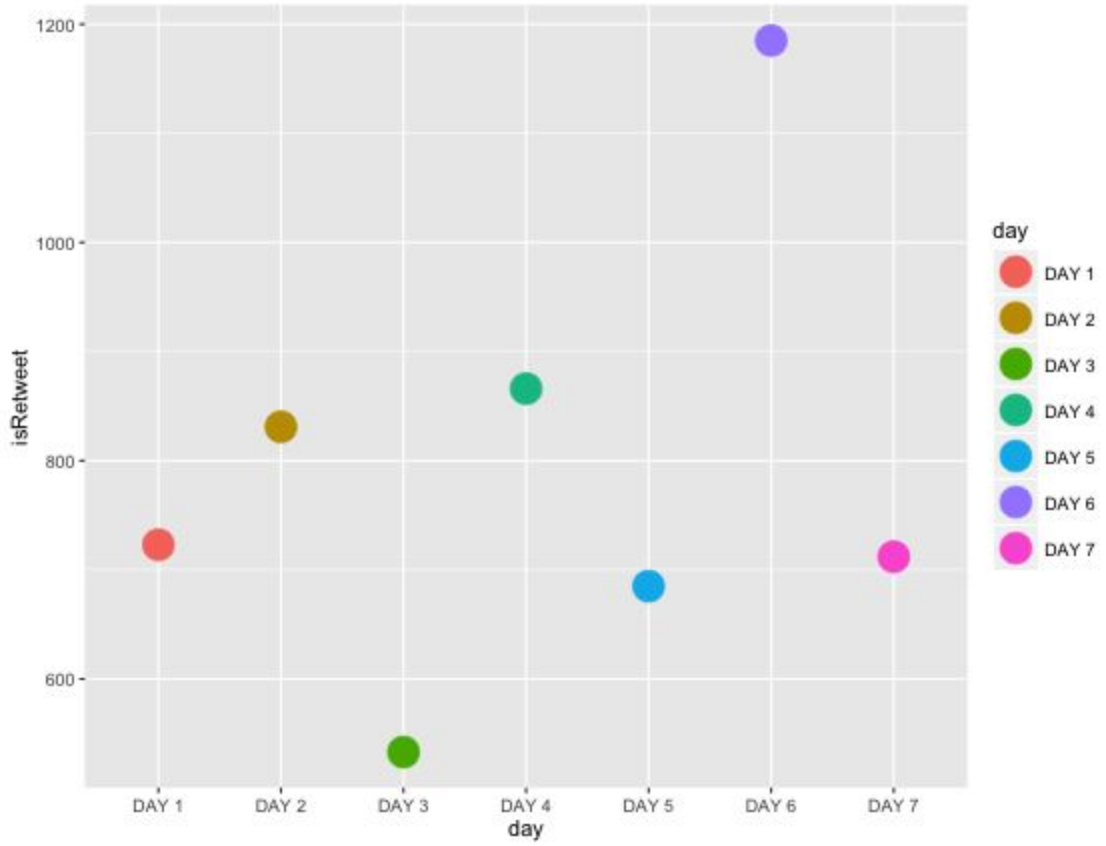
For this problem tweets are collected through twitter search API in R called twitteR. To collect tweets from twitter an authorization token need to be received by creating an app on twitter's dev site. This process returns four keys Consumer key (API key), Consumer secret key (API secret), Access token and Access token secret. Through these four keys twitter OAuth is established through setup_twitter_oauth() method. OAuth is like a handshake which checks the authorization and gives permission to download twitter data. Now that OAuth is in place twitter can be searched through searchTwitter() method by some keywords, or certain geo location, or for a certain period. The disadvantage that comes along with this method is that it allows to collect data only for seven previous days. This method returns a status list of tweets, this list is converted to a dataframe and thereby to JSON format through RJSONIO package. This can be written and hence saved to drive in a file.

The JSON file created can be imported through fromJSON() method in rjson package which return a list. This list can be converted to a dataframe.

Through this method tweets are collected over a period of one week (02/23/2016 to 02/29/2016) and saved in the JSON format.

Below are some of the charts created from the data collected:




1200
1000
800
600
isRetweet
DAY 1
DAY 2
DAY 3
DAY 4
DAY 5
DAY 6
DAY 7
day
day
DAY 1
DAY 2
DAY 3
DAY 4
DAY 5
DAY 6
DAY 7