Problem 5 Stream Processing

For getting streaming twitter data for 1 week I used two approaches as:

I. Rshiny integration with stream data fetched via streamR api.
II. Rshiny integration with data fetched by twitterR api which allowed to get data for 1 week which was not possible via streamR api.

I. Rshiny integration with stream data fetched via streamR api

In this approach I used streamR api which helps in fetching live tweets in real time by allowing api to be opened for a specific period of time. For example if I need to fetch live tweets where keyword “Donald Trump” is specified and tweets should be from specific area for example “New York” my streamR query would be:

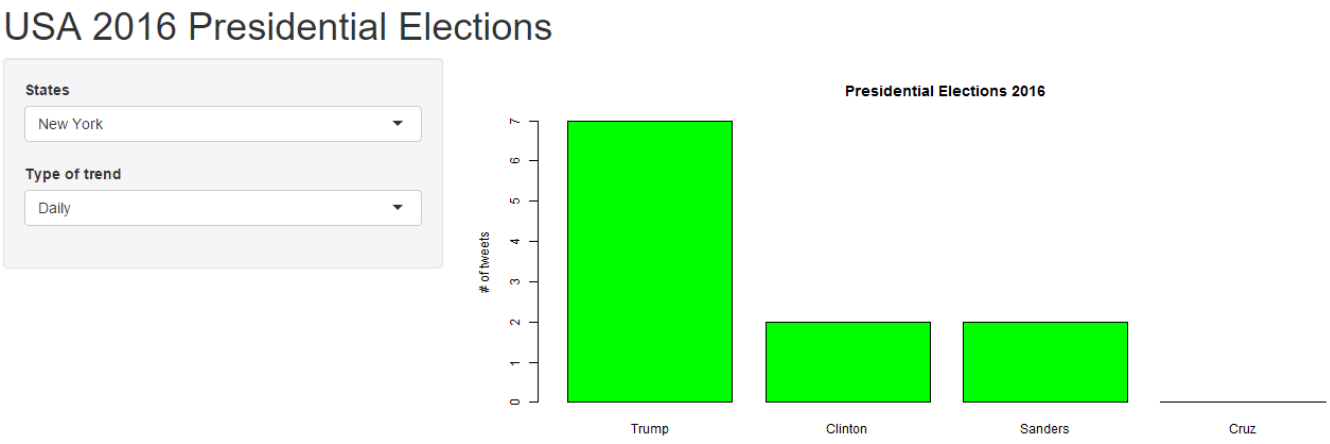
filterStream (file.name="tweets_re3.json", track="Donald Trump", tweets=100000, oauth=credential, timeout=120, lang='en')

where track specifies keyword which we are searching in tweets. Attribute timeout specific seconds up to which streaming of data should happen.

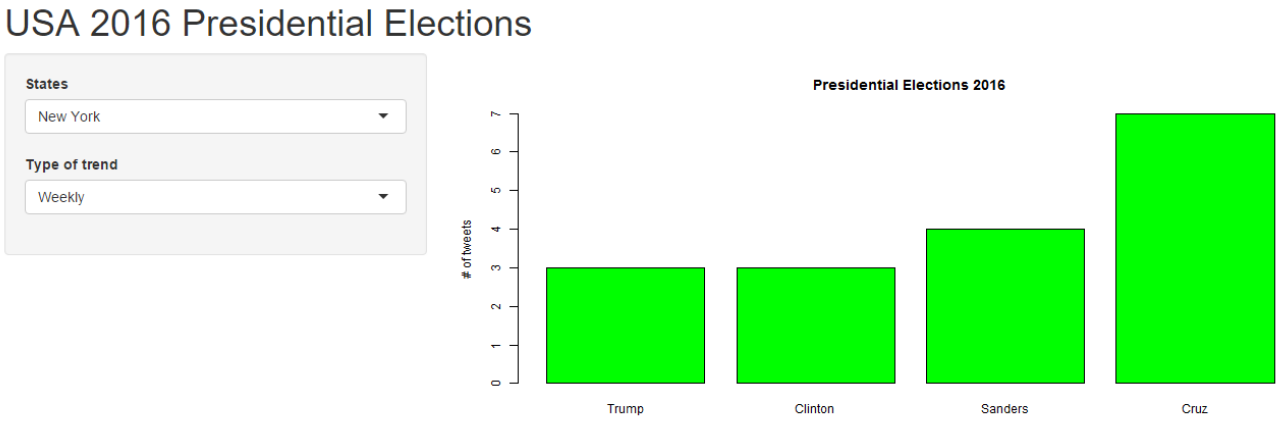
For UI purpose I used library shiny which provides link of UI between user and R. After setting up initial configurations of shiny framework I provided UI asking user for State user want data from and duration of data like “Daily” or “Weekly”. After user's selection shiny's reactive framework updates data by fetching it for recent queries.

For example results I got when user selected various options from State and Duration are as follow:

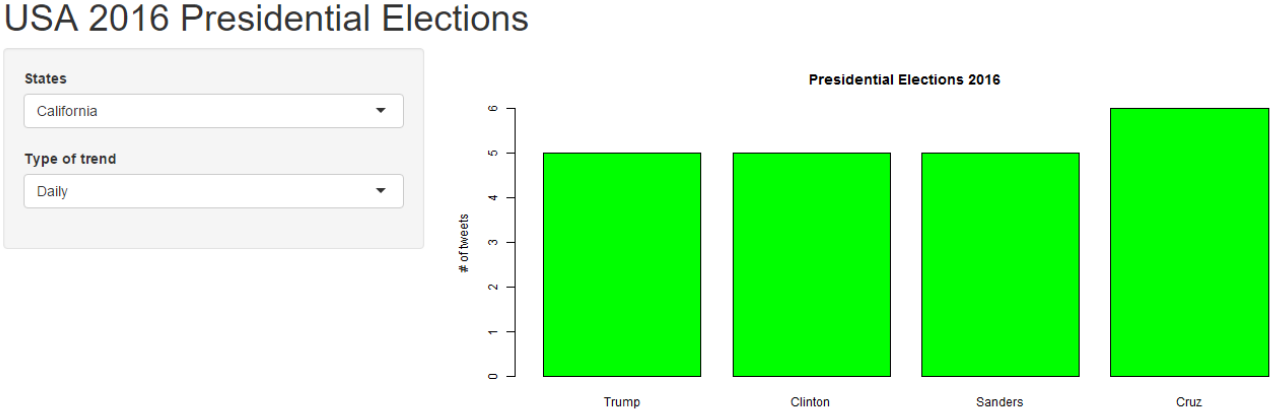
State: New York
Day/Week: Daily



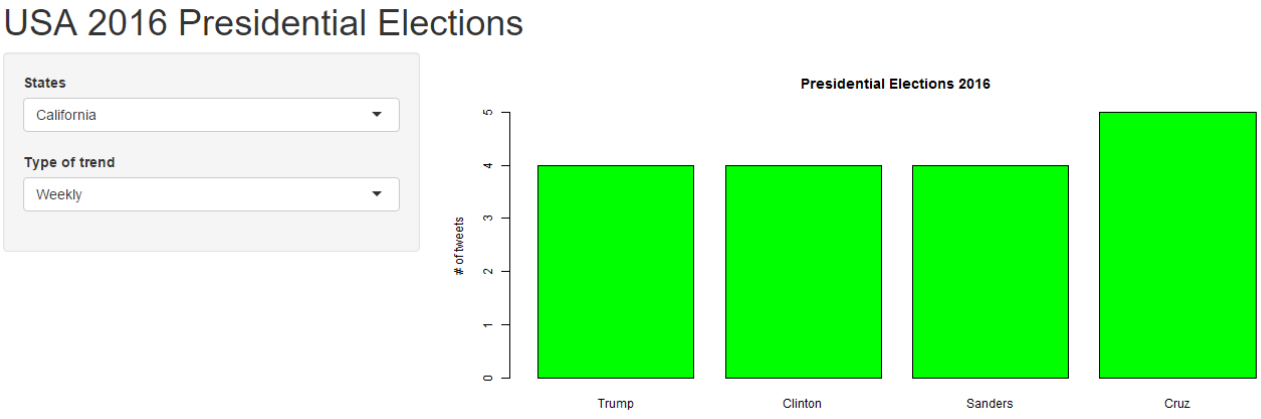
State: New York
Day/Week: Weekly



State: California
Day/Week: Daily



State: California
Day/Week: Weekly

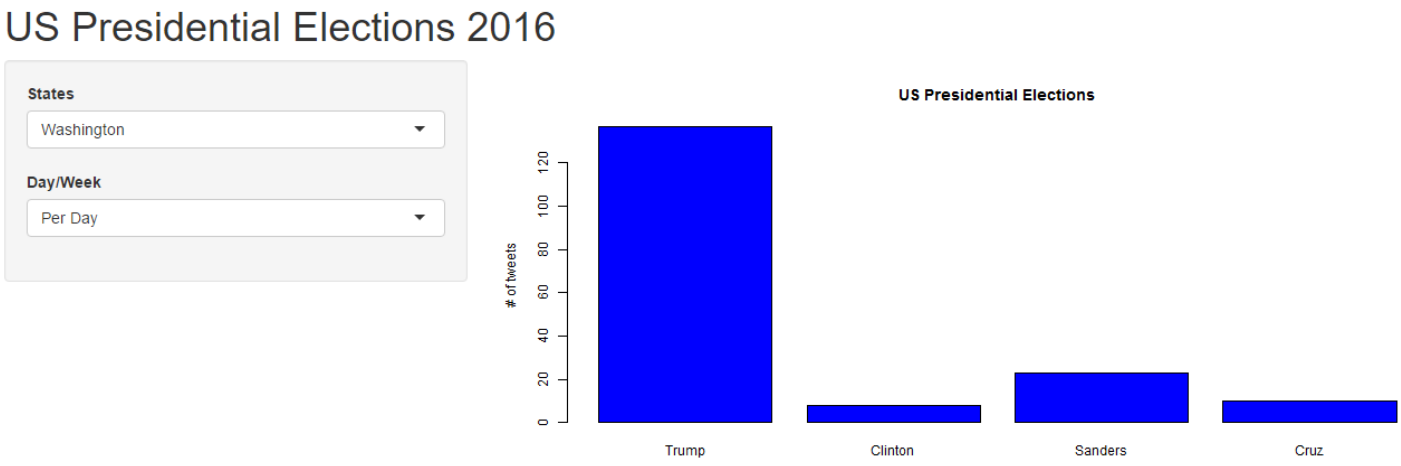


II. Rshiny integration with data fetched by twitterR api

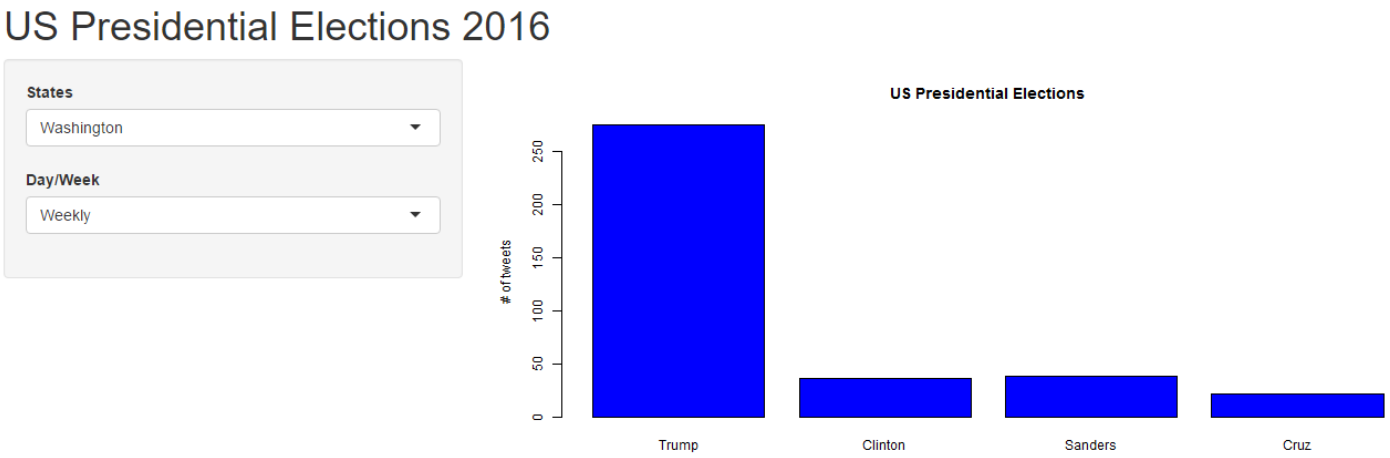
I took this additional approach as it was not feasible streamR to get data for a week. With twitterR api we can use attribute since which let us define data after which we need to pick tweets till today. To use this api I set some of it's attributes like since to set since date, geocode to provide latitude and longitude of place from where we need to fetch our tweets. Its complete query looks something like this:

trump_tweets<- searchTwitter("Donald Trump", n=2000, lang="en", since = as.character(Sys.Date()-1), geocode= '43.06888777,-75.19042969,300mi')

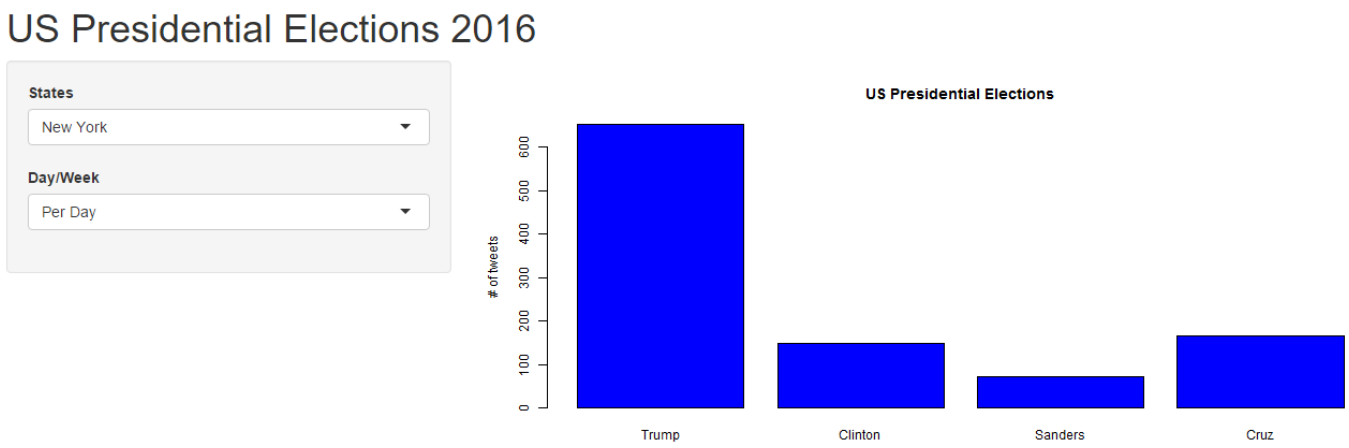
Results I got after integration of twitterR with Rshiny are as follow:
Selected Parameters:
State: Washington
Day/Week: Per Day



State: Washington
Day/Week: Weekly



State: New York
Day/Week: Per Day



State: New York
Day/Week: Weekly

US Presidential Elections 2016

States

New York

Day/Week

Weekly