

Problem 5: Stream Processing

In this problem, a live streaming web app was created using Shiny based on live streaming of election data. A dataset containing ten days of tweets (about 30,000) is used as a base. They are stored in such a way that each tweet has a field for mention of any of the three famous presidential candidates (Donald Trump, Hillary Clinton & Bernie Sanders). This is how number of mentions of each candidate is measured and used later for comparison in popularity. For the purpose of this assignment, I have hidden the other attributes of the tweets for the time being and used on the basic ones.

	text	Trump	Hillary	Sanders
column 0: rownames	Donald Trump" are fighting words<ed...	1	0	0
16	RT @PizzaPartyBen: Bernie supporters Vs. Trump sup...	1	0	1
17	RT @Zacharycohen: Hillary can't beat trump. Bernie ca...	1	1	1
18	@imcrews So, remember that convo we had a while b...	1	0	0
19	@knuckleitis @MrGrahamClegg saw this thought of y...	0	0	0

I have created a function “Collecting_Live_Tweets” which uses Twitter Streaming API (from package `twitteR`) to collect live tweets and store them in the format I want. This means that to every tweet I add columns “Trump”, “Hillary” and “Sanders” whose value becomes 1 if any of these were mentioned in the tweet. I use “`grepl`” function to find mentions of these presidential candidates in the tweets. Next these tweets are passed on to the server.R of my app which refreshes every 10s. This means that new live tweets are reflected on the web page & calculations every 10s. In server.R, the latest tweets are plotted to

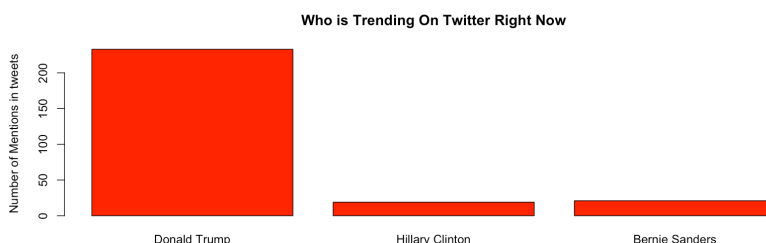
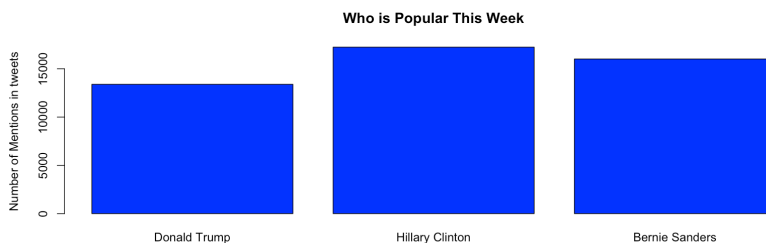
<http://127.0.0.1:7486> [Open in Browser](#) [Publish](#)

Analysis of Live Twitter Data on Elections

Donald Trump was mentioned 13392 times

Hillary Clinton was mentioned 17243 times

Bernie Sanders was mentioned 16013 times



Some of the tweets are:

The anti #DonaldTrump vote currently split 3 ways. Worst thing that can happen 4 Trump is #MarcoRubio drops out, those votes go to #TedCruz

Donald Trump projected winner of Kentucky's Republican presidential caucuses - AP, CNN via @Breaking @KNX1070

RT @AP: BREAKING: Vermont Sen. Bernie Sanders wins Democratic caucuses in Nebraska. @AP race call at 9:04 p.m. EST. #Election2016 #APracec...

RT @AP: BREAKING: Donald Trump wins the Republican caucuses in Kentucky. @AP race call at 10:46 p.m. EST. #Election2016 #APracecall

RT @EsotericCD: Trump wins both KY and LA, but it's surprisingly close in LA. Cruz had a magnificent election-day showing.

find currently trending candidate. Then these latest tweets get appended to the already present twitter dataset and a weekly analysis is done and the bar plot of that is plotted on the web page.

On the web application, I am displaying how many times a candidate has been mentioned in the whole dataset, a bar plot of who is popular this week and another bar plot of who is trending right now (daily statistics). In addition, on the side panel I have displayed the some of the latest collected tweets.

Note: When starting the app, the data and distribution is displayed after 10 seconds since the app refreshes every 10 seconds and at the beginning it is still collecting data at the first 10 seconds.

