Project -1

Problem 5 – Rshiny App to analyse live tweets

CSE – 587

Instructor – Bina Ramammurthy

Submitted By:-

NALIN KUMAR

Person - 50170479

US Presidential Election Trends 2016 – RSHINY Application

Shiny is an R package that makes it easy to build interactive web applications straight from R. Apps built using Shiny have essentially two components namely a user interface script and a server script. User interface is just a web document that the user gets to see, it's HTML that we write using Shiny's functions. The user interface is responsible for creating the layout of the app and telling Shiny exactly where things go. The server is responsible for the logic of the app. It's actually the set of instructions that tell the web page what to show when the user interacts with the page. The application which I built in this part is based on the tweet sentiment analysis of US Presidential elections 2016 for four different candidates namely Donald Trump, Bernie Sanders, Hillary Clinton, Ted Cruz. While building the UI of this app, we essentially provide two select input dropdowns, one to select and view the sentiments related to a specific presidential candidate and the other which prompts the user to select weekly or current based on whether he wishes to view tweet analysis for tweets collected for the past seven days or for live tweets using Twitter Streaming API. Based on whatever option user has chosen in the two dropdowns, we show him the respective results. The results are displayed on a bar plot where Number of Tweets are plotted on y-axis and tweet sentiment score is plotted on x-axis. This tweet sentiment score is calculated for all of the tweets based on whether those tweets contained one or more words which are present in positive.txt or negative.txt files. The function which checks whether a particular tweet should be assigned a positive score or a negative score does many things. Initially it cleans up the data based on certain regular expressions and subsequently checks whether tweets contain any words which match the words present positive.txt or negative.txt files which primarily contain words based on a universal set of positive emotions and negative emotions depicting words. Subsequent to this analysis, every tweet is assigned a score as some negative or a positive value and the results are plotted on a bar plot as can be seen in the following sections.

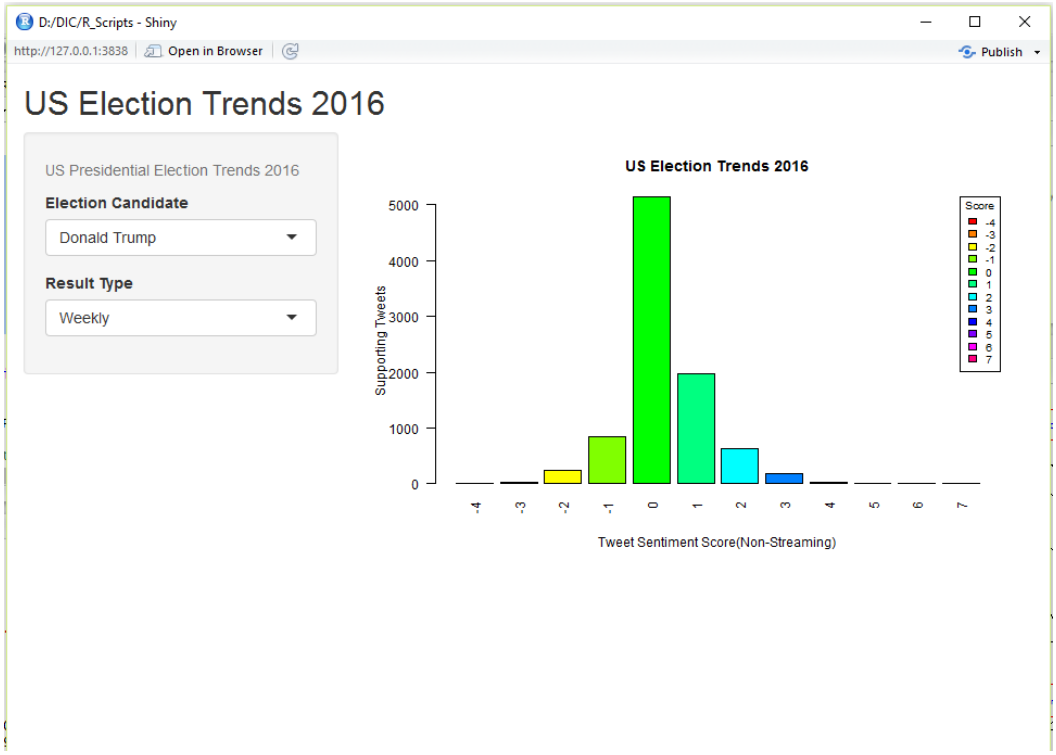
Plots corresponding to weekly collected tweets

Donald Trump Weekly

The first plot corresponds to the case where user has chosen Donald Trump as Presidential Candidate and Weekly as result type from the two dropdowns. As can be seen in the plot, there are certain tweets which have a positive score and some others are neutral or possess a negative score. Net tweet sentiment score can be defined as by calculating the positive tweets and negative tweets and subsequently subtracting negative tweets from positive tweets(ignore neutral tweets). From the graph plotted below, we can visualize that net tweet score for Donald Trump based on Weekly tweets is approximately 1700.

```
Listening on http://127.0.0.1:3838
[1] "Donald Trump"
[1] "Weekly"
Read 2006 items
Read 4783 items

  -4   -3   -2   -1    0    1    2    3    4    5    6    7
   7   27  239  843 5149 1981  624  178   25    1    3    1
```

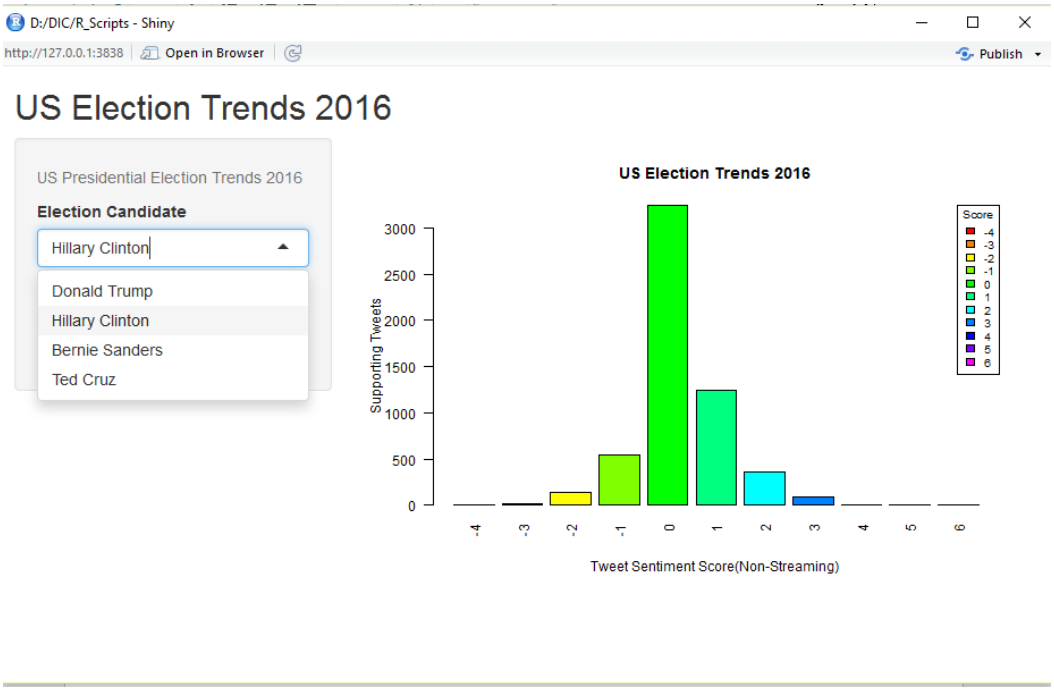


Hillary Clinton Weekly

The second plot corresponds to the case where user has chosen Hillary Clinton as Presidential Candidate and Weekly as result type from the two dropdowns. As can be seen in the plot, there are certain tweets which have a positive score and some others are neutral or possess a negative score. Form the graph plotted below, we can visualize that net tweet score for Hillary Clinton based on Weekly tweets is approximately 1000.

```
[1] "Hillary Clinton"
[1] "Weekly"
Read 2006 items
Read 4783 items

  -4   -3   -2   -1    0    1    2    3    4    5    6
   6   23  142  542 3248 1238  359   86    8    2    1
```

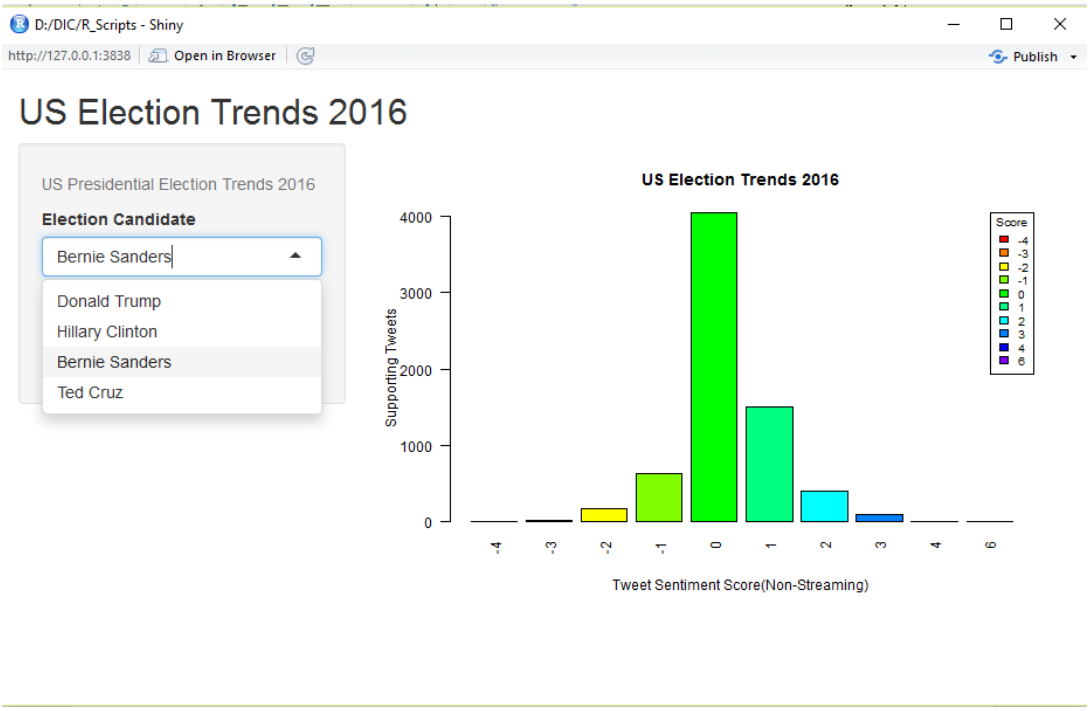


Bernie Sanders Weekly

The third plot corresponds to the case where user has chosen Bernie Sanders as Presidential Candidate and Weekly as result type from the two dropdowns. As can be seen in the plot, there are certain tweets which have a positive score and some others are neutral or possess a negative score. Form the graph plotted below, we can visualize that net tweet score for Bernie Sanders based on Weekly tweets is approximately 1150.

```
[1] "Bernie Sanders"
[1] "Weekly"
Read 2006 items
Read 4783 items

  -4   -3   -2   -1    0    1    2    3    4    6
   6   20  181  633 4051 1501  409   94   11    1
```

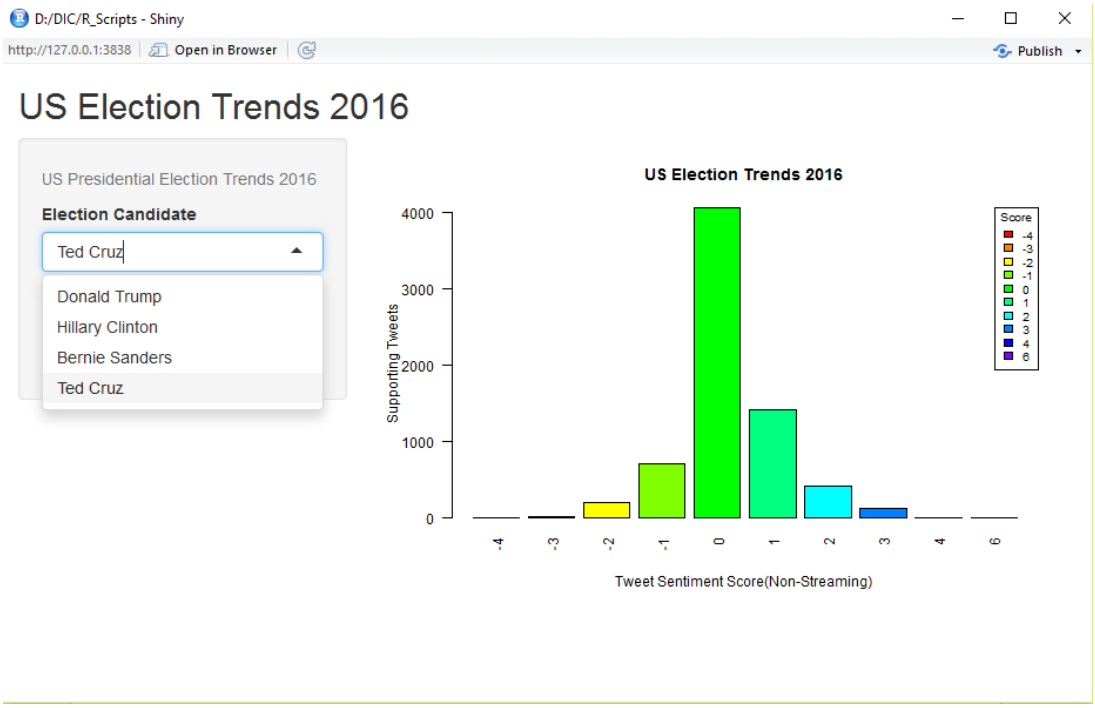


Ted Cruz Weekly

The last plot corresponds to the case where user has chosen Ted Cruz as Presidential Candidate and Weekly as result type from the two dropdowns. As can be seen in the plot, there are certain tweets which have a positive score and some others are neutral or possess a negative score. Form the graph plotted below, we can visualize that net tweet score for Ted Cruz based on Weekly tweets is approximately 1000.

```
[1] "Ted Cruz"
[1] "Weekly"
Read 2006 items
Read 4783 items

  -4   -3   -2   -1    0    1    2    3    4    6
   7   20  213  709 4063 1412  414  126    6    2
```



Conclusions based on Weekly Tweets

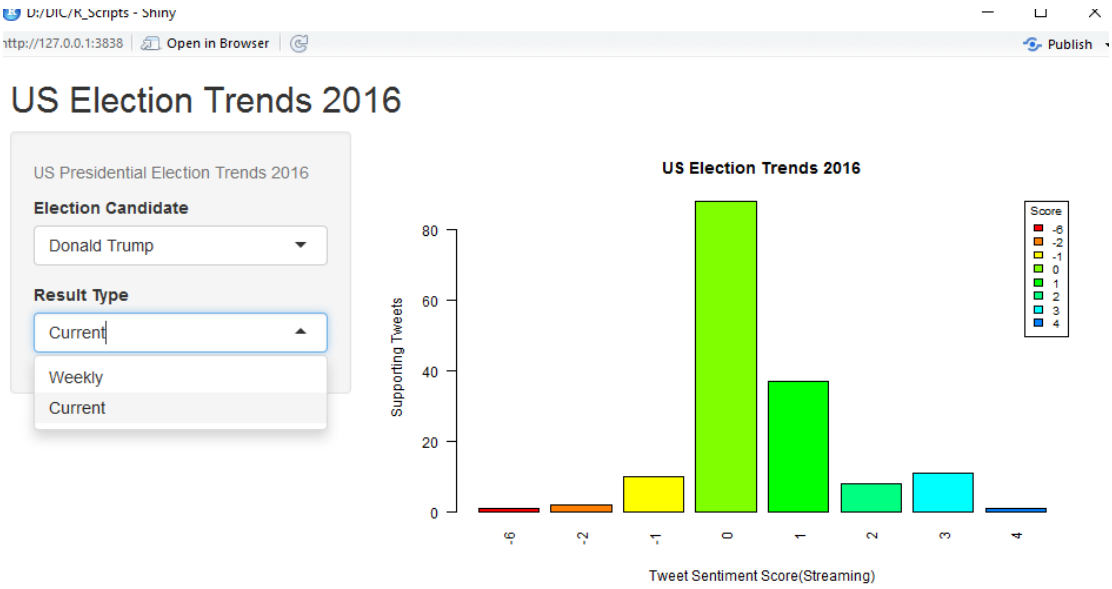
Based on the results plotted above, we can clearly see that Donald Trump has the maximum number of net positive tweets (1700) as compared to other presidential candidates by a significant margin. Also, the range of positive tweets is more in the case of Donald Trump as compared to other presidential candidates. But besides this, there are certain other things to observe while arriving at this conclusion. It's the fact that the number of tweets fetched for Donald Trump were more as compared to other candidates on every single day since Donald Trump recently was on a winning streak in 3 constituencies. Hence, we can say that based on this fact there were more total and positive tweets for Donald Trump as opposed to other candidates

Donald Trump Current

The first plot corresponds to the case where user has chosen Donald Trump as Presidential Candidate and Current as result type from the two dropdowns which essentially means that we will be fetching live tweets using Twitter Streaming API. The timeout value used is 15 seconds. As can be seen in the plot, there are certain tweets which have a positive score and some others are neutral or possess a negative score. Form the graph plotted below, we can visualize that net tweet score for Donald Trump based on current tweets is 44.

```
[1] "Donald Trump"
[1] "Current"
[1] "Donald Trump"
Capturing tweets...
Connection to Twitter stream was closed after 15 seconds with up to 331 tweets downloaded.
158 tweets have been parsed.
Read 2006 items
Read 4783 items

-6 -2 -1  0  1  2  3  4
 1  2 10 88 37  8 11  1
```

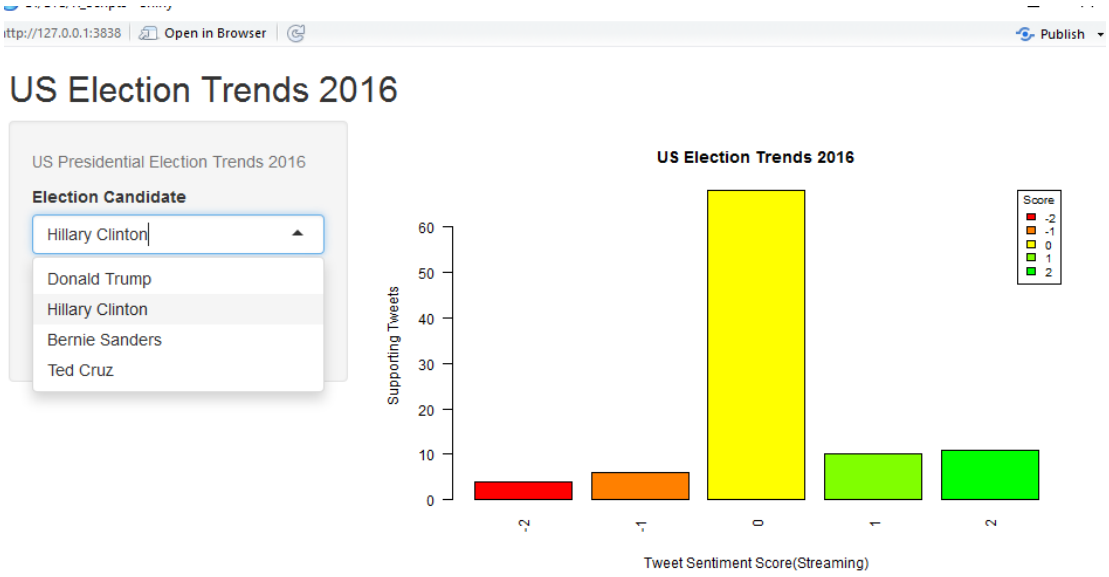


Hillary Clinton Current

The second plot corresponds to the case where user has chosen Hillary Clinton as Presidential Candidate and Current as result type from the two dropdowns which essentially means that we will be fetching live tweets using Twitter Streaming API. The timeout value used is 15 seconds. As can be seen in the plot, there are certain tweets which have a positive score and some others are neutral or possess a negative score. Form the graph plotted below, we can visualize that net tweet score for Hillary Clinton based on current tweets is 11.

```
99 tweets have been parsed.
Read 2006 items
Read 4783 items

-2 -1  0  1  2
 4  6 68 10 11
```

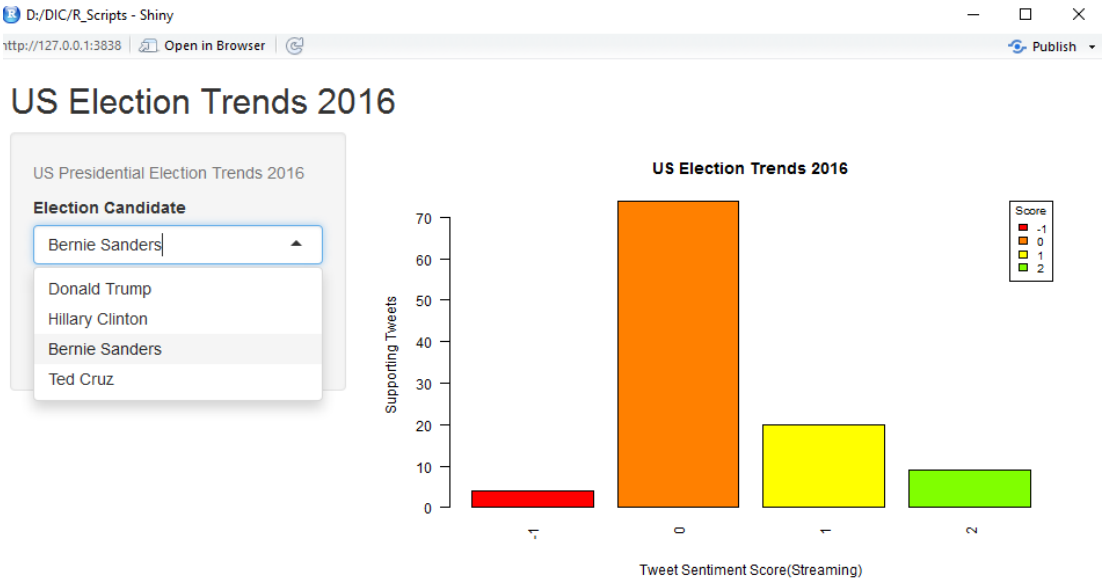


Bernie Sanders Current

The third plot corresponds to the case where user has chosen Bernie Sanders as Presidential Candidate and Current as result type from the two dropdowns which essentially means that we will be fetching live tweets using Twitter Streaming API. The timeout value used is 15 seconds. As can be seen in the plot, there are certain tweets which have a positive score and some others are neutral or possess a negative score. Form the graph plotted below, we can visualize that net tweet score for Bernie Sanders based on current tweets is 25.

```
Connection to Twitter stream was closed after 15 seconds with up to 264 tweets downloaded.
107 tweets have been parsed.
Read 2006 items
Read 4783 items

-1  0  1  2
 4 74 20  9
```

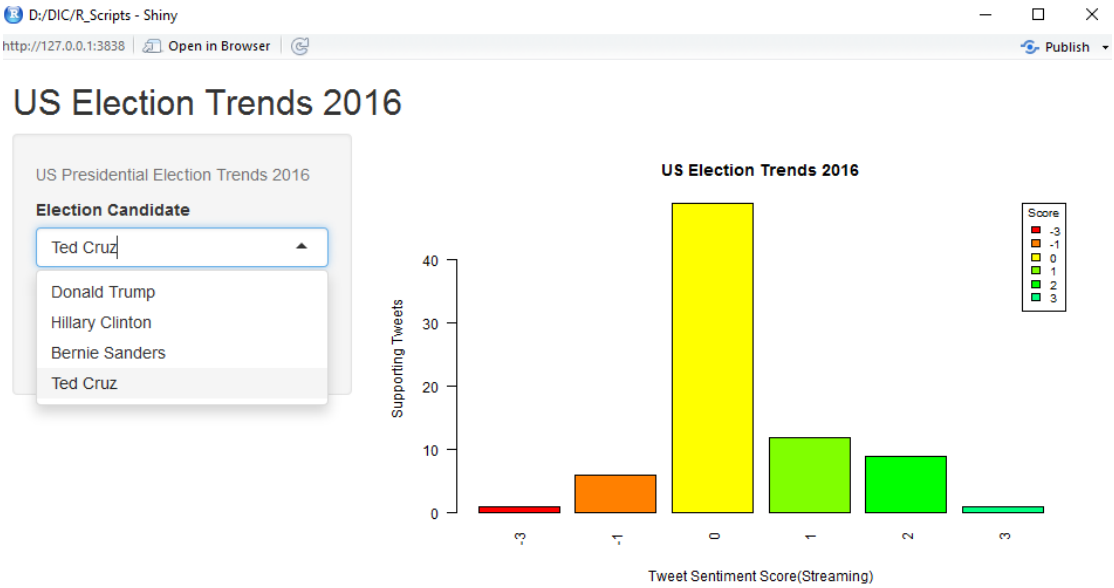


Ted Cruz Current

The third plot corresponds to the case where user has chosen Bernie Sanders as Presidential Candidate and Current as result type from the two dropdowns which essentially means that we will be fetching live tweets using Twitter Streaming API. The timeout value used is 15 seconds. As can be seen in the plot, there are certain tweets which have a positive score and some others are neutral or possess a negative score. Form the graph plotted below, we can visualize that net tweet score for Bernie Sanders based on current tweets is 15.

```
[1] "Ted Cruz"
Capturing tweets...
Connection to Twitter stream was closed after 15 seconds with up to 198 tweets downloaded.
78 tweets have been parsed.
Read 2006 items
Read 4783 items

-3 -1  0  1  2  3
 1  6 49 12  9  1
```



Final Conclusions based on Live Tweets and Weekly Tweets

Since the timeout was kept at 15 seconds, we only got few live tweets to compare among the four presidential candidates. But again in this, Donald Trump received the maximum positive tweet score (44) despite the fact that the total tweets collected through streaming API is only a small number. This clearly reflects the fact that Donald Trump is receiving the maximum support from the people since he received the maximum positive tweets score for both weekly as well as current tweets.

References :-

1) https://github.com/mjhea0/twitter-sentiment-analysis