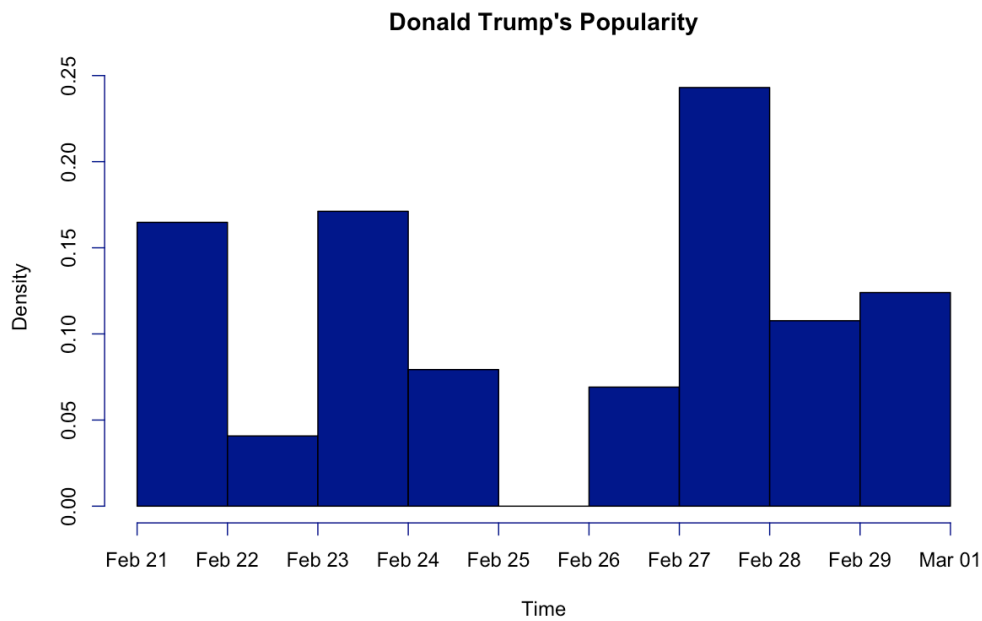


Problem 5

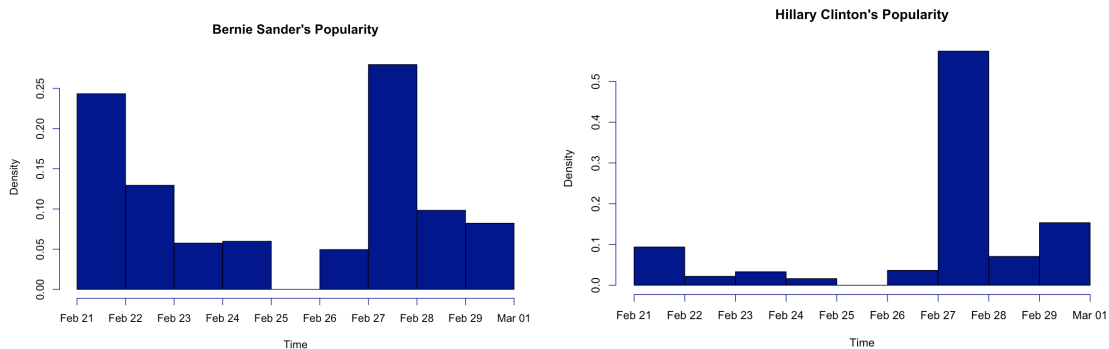
Problem Description:

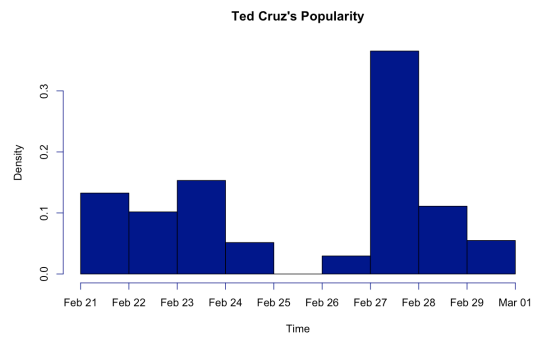
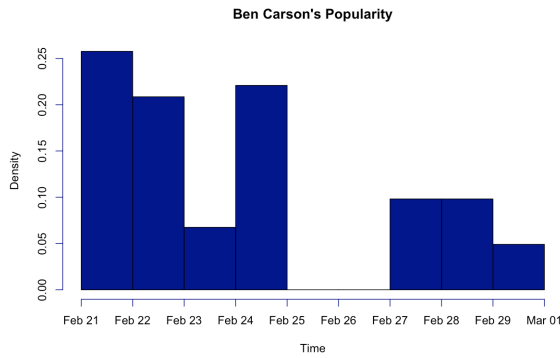
In this problem, twitter data was collected randomly using the keywords “US+elections”. These tweets were stored as a JSON file. The JSON file is read and cleaned. The text column is converted to lower case and all unnecessary special characters are removed. The date field was formatted to contain only the date and time was removed. The tweets were then classified based on the keywords contained in the text field.

Analysis of Plots:

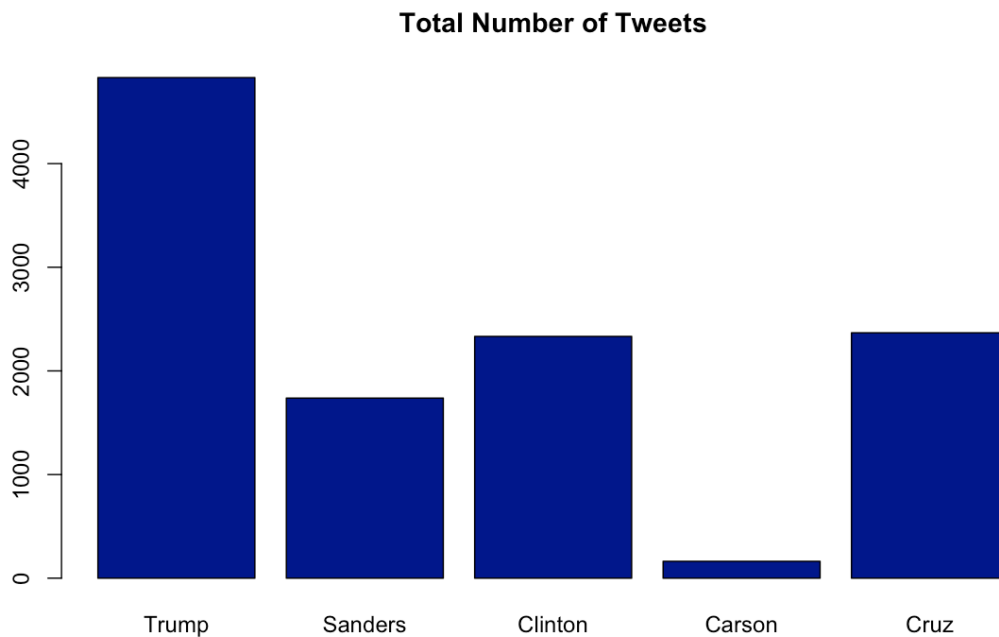


The histogram above shows the variation in popularity daily over a 10-day period. Similarly data is plotted for all other candidates using R Shiny.





The final bar plot below shows the total number of tweets containing keywords related to each of the candidates. It can be seen that the popularity of Donald Trump is much higher than any other candidate.



Streaming Data

The script also contains a button to fetch and plot live data from Twitter, when the button is clicked, 100 most recent tweets are fetched from twitter using keywords "US+elections" and we search through those tweets to find keywords matching the candidates' names.