Problem 1: Data Acquisition:

Data collection is important during any analysis as the accuracy and the efficiency of the results obtained are constrained to the quality of data collected.

A few details about the data set are as given below.

Data source: All of the data that was collected was done using the twitterAPI. The data was collected in the form of tweets in JSON format.

Data Structure: The tweets that were collected is in the JSON format. All the data is in a nested JSON object. The JSON object consists of 15 fields, namely, "text", "favorited", "favoriteCount", "created", "truncated", "replyToSID","id","replyToUID", "statusSource", "screenname", "retweetCount", "isRetweet", "retweeted", "longitude" and "latitude". These fields either hold text value, numeric value, or Boolean values.

Data Collection Time Period: The tweets were collected every day over a week.

Packages: The packages that were imported during the collection and the import of data and the details of them are given below:

ROAuth -> used for authentication and OAuth authorization.

streamR-> The filterStream method of the streamR package was used to search twitter based on search terms and get data.

rjson and jsonlite -> These packages were used to import the JSON file into R.