Name:Suhani Vinod Prajapati

Roll No:17BCS030

Text-Analysis on Demonetization in India

**Dataset**:The dataset (containing 14940)was obtained from Kaggle and R was used for the analysis.

**Introduction**:

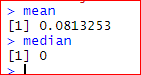
Sentiment Analysis (SA) or Opinion Mining (OM) is the computational investigation of individuals’ sentiments, states of mind and feelings toward an element.

**Analysis**(Step involved):

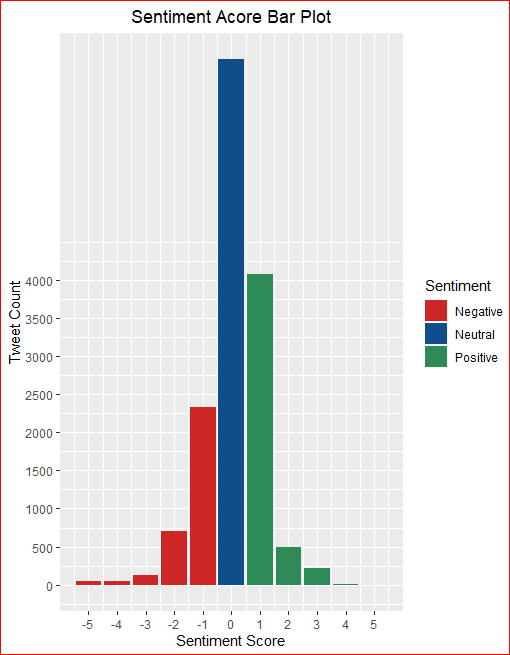
* The sentiment score of each tweet was analyzed using ‘bag of words’ method.
* The sentences is broken into words which are treated as token.
* The text of each tweet is pulled out,cleaned and preprocessed.
* Lexicon is a list of words tagged positive and negative.Each tweet was parsed and scanned to the presence of words matching the words from the list of the positive and negative words from the Lexicon.
* The score of the tweet was the net score of all its words.This process is repeated for all 14940 tweets.
* Based on the score,the tweet was identified as positive(>0),negative(<0) or neutral(=0).

**Results**:

Mean Score:

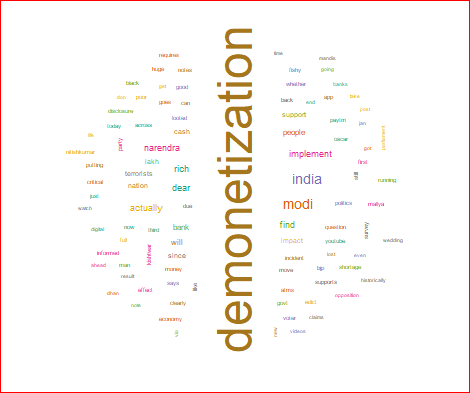


Clearly,the sentiment of most of the tweets are neutral. Also, more number of people had positive sentiments regarding demonetization



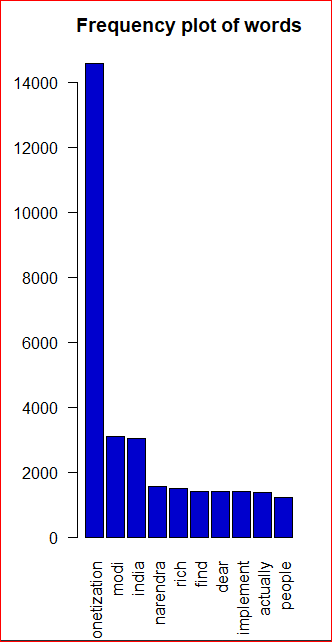


**Wordcloud**



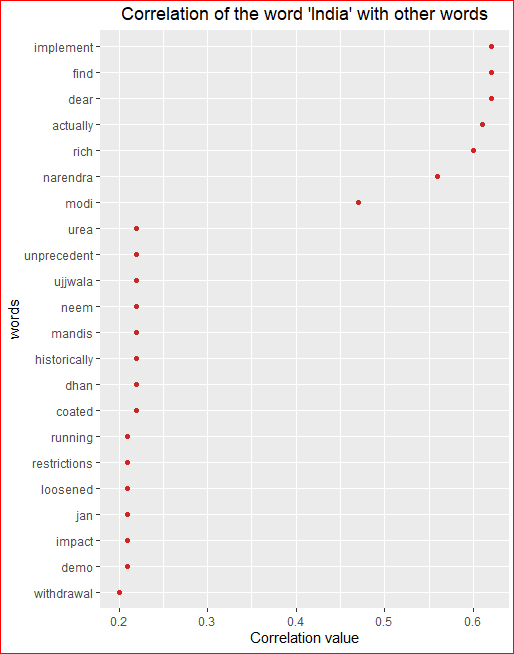
From wordcloud obtained,it is clear that many people support demonetization.

**Frequency of occurrence of words**:The most frequent words are modi,India,rich,narendra,find,dear, implement,actually,people.



**Word Association Plot:**

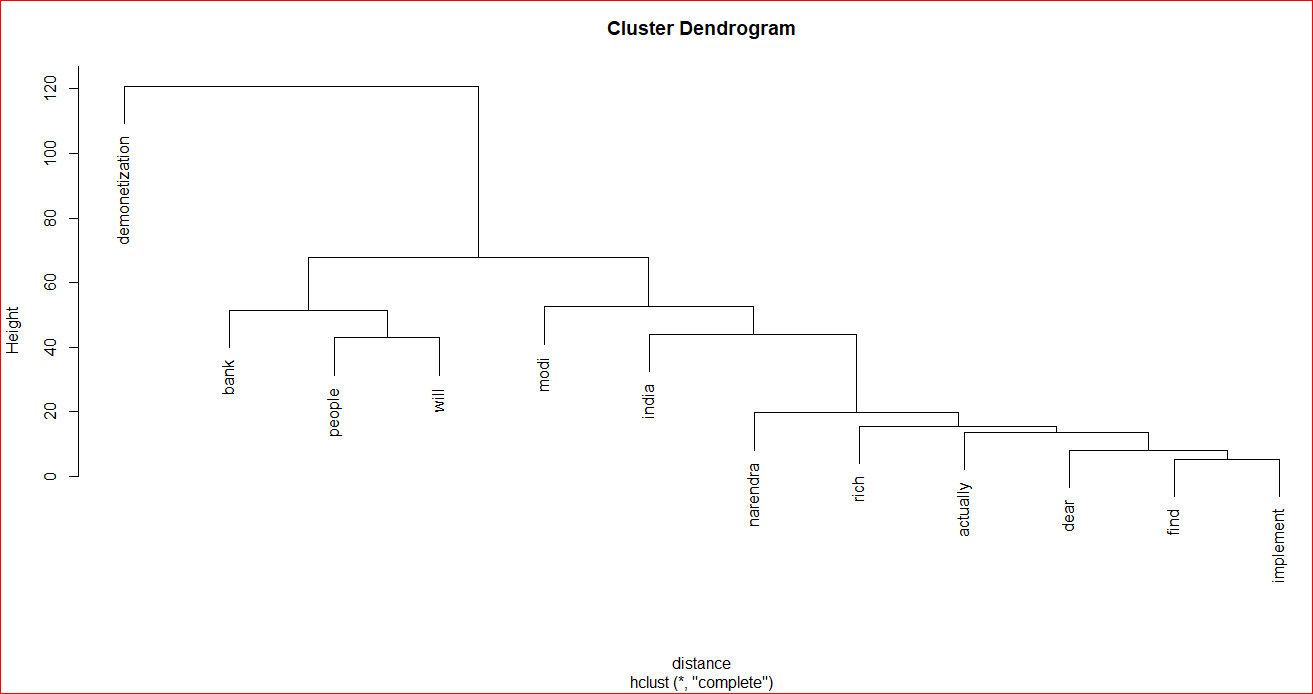
The word association,inspected for ’India’ i.e.which words are closely associated with India.



As from the above plot, implement and find have highest association with India.

**Dendrogram:**

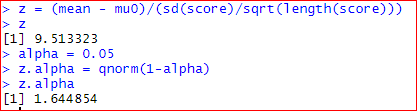
Dendrogram is used to visualize hierarchical clustering of word frequency distances. ( Shows how the fall in the various clusters.



**Hypothesis Testing:**

Let H0: Demonetization had majority popular support.

H1 (Alternative Hypothesis):Demonetization do not had majority popular support.



At alpha = 0.05 the calculated-value < tabulated-value i.e. 1.644 < 1.96.Hence null hypothesis is accepted.Hence, demonetization had majority popular support.

**THE END**