

Written by Gurpreet Singh (Master's Student UCD) for Text Analytics.

Q1: Word Clouds

a. *Commands for the word cloud in R:*

Answer:

Before this let us see what a word cloud means: It's an image composed of words used in a particular text or subject, in which the size of each word indicates its frequency or importance.

Before executing the below commands install the required packages, i.e. 'wordcloud' and 'tm'.

Execute Commands:

```
> library(wordcloud)
Loading required package: RColorBrewer
> library(tm)
Loading required package: NLP
> wordcloud("May our children and our children's
+ children to a thousand generations, continue to enjoy
+ the benefits conferred upon us by a united country,
+ and have cause yet to rejoice under those glorious
+ institutions bequeathed us by Washington and his
+ compeers.",
+ colors=brewer.pal(6, "Dark2"), random.order=FALSE)
```

Output:



b. *Analysis on the output of the program:*

Answer:

List of words that got included are: Children, country, rejoice, yet glorious, compeers, childrens, bequeathed, upon, cause, continue, thousand, institutions, united, enjoy, may, continue, generations, benefits.

List of words that are not included are: and, the, us, a, by, his, Washington, yet, to, have.

Yes only some of the words were included in the cloud as it checks for the frequency and removes the articles such as a, an, the, it, they, them.

But the word Washington also got removed with a warning message as : "washington could not be fit on page. It will not be plotted."

c. *Theory about the understanding:*

Answer:

The word cloud is the method of visualisation method that displays how frequently the words appear in a given body of text, by making the size of each word proportional to its frequency. Usually all the words are arranged in the form of a cloud.

While working on the word cloud I found that a large word was not plotted even though when I increased its frequency, I received a warning message every time: "<15 letter word> could not be fit on page. It will not be plotted." Then I installed some go-lot packages to plot but no help.

Here is the text used by me: "Peter Piper picked a peck of peepppplyyyyyy pickled peppers A peck of pickled peppers Peter Piper picked peepppplyyyyyy If Peter Piper picked a peck of pickled peepppplyyyyyy peppers Wheres the peck of pickled peppers Peter Piper picked?"

output: (made random.order=TRUE)



d. Again, using your word-list add more repeated words and see what happens? Can you change the package's to make it more inclusive of the words in the word-list?

Answer:

Upon changing the word count of the repeated word I found that the size of the repeated word was maximum. That is size of the word was dependent on its frequency.

Upon increasing the word "Peter" from 4 to 10 there was a huge difference in the output, that can be clearly seen below:

