Word Cloud Lab

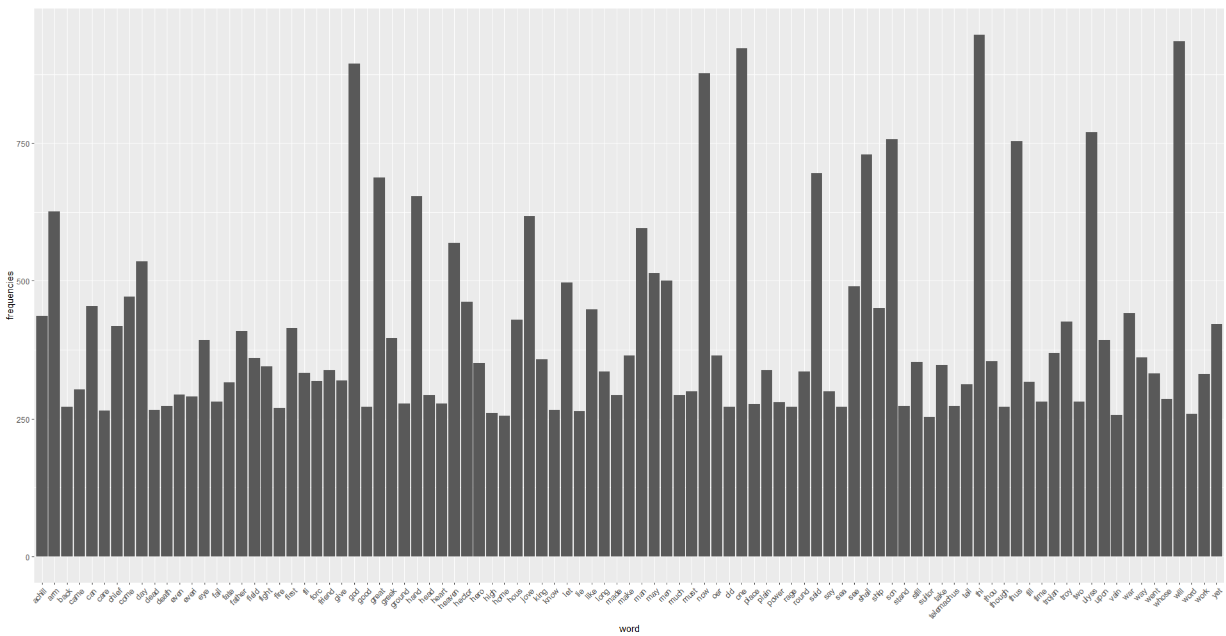
I have uploaded the R file for the Word Cloud Lab in the *Program* folder. Please check it. Below files are the generated images after performing the basic text mining for the '*iliad'* and '*odyssey'* text files.

PROCEDURE:

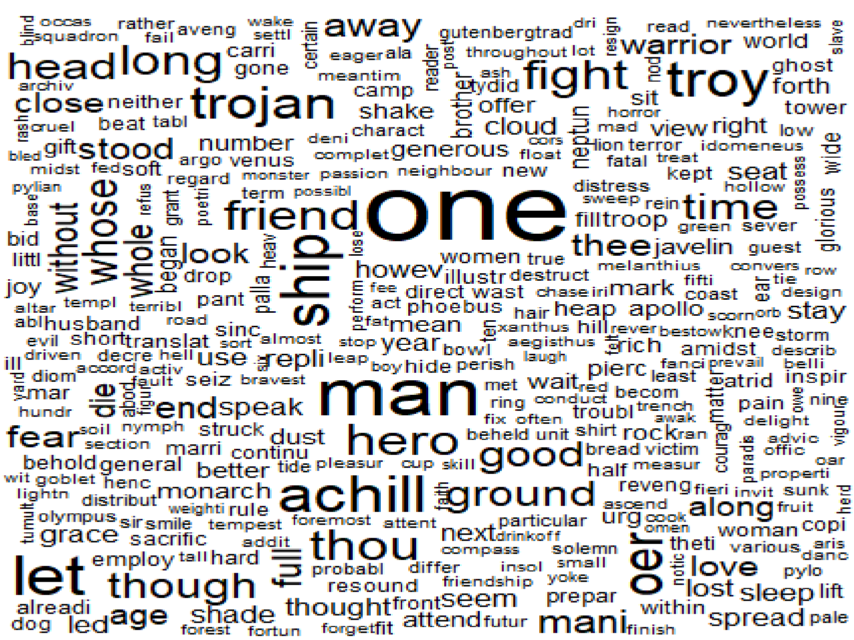
1. Installing the necessary required libraries
2. Load the two text files in the C:/texts path location
3. Preprocessing the data:
    - Removing Punctuations
    - Removing Special characters like "/", "@", "\\|"
    - Removing Numbers
    - Changing to Lower characters
    - Removing the Stop words in English
    - Removing the particular words like "department" and "email"
    - Combining the words which are meant to specify in single word
    - Removing the common ending words like "ing","es","s"
    - Stripping the unnecessary white spaces
4. Staging the data:
    - Creating a document term matrix
    - Creating a term document matrix (Transpose)
    - Checking their dimensions
5. Exploring the data:
    - Organizing the terms based on their frequencies
    - Exporting the matrix to an Excel file
    - Removing the sparse terms
    - Creating the data frame based on the word frequency
    - Plotting the word frequency by using '*ggplot'* library
6. Finding the relationships between the respective words
7. Visualizing using the Word Cloud
8. Performing Hierarchical clustering
9. Performing k-means clustering

GENERATED REPORTS:

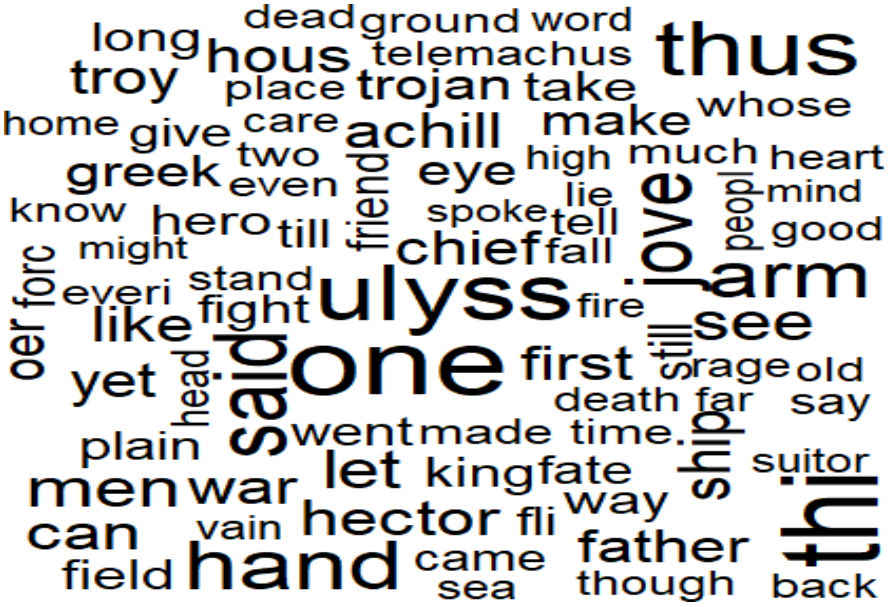
1. Plotting Word Frequencies:



2. Word Cloud with minimum frequency 25(*min.freq=25*):



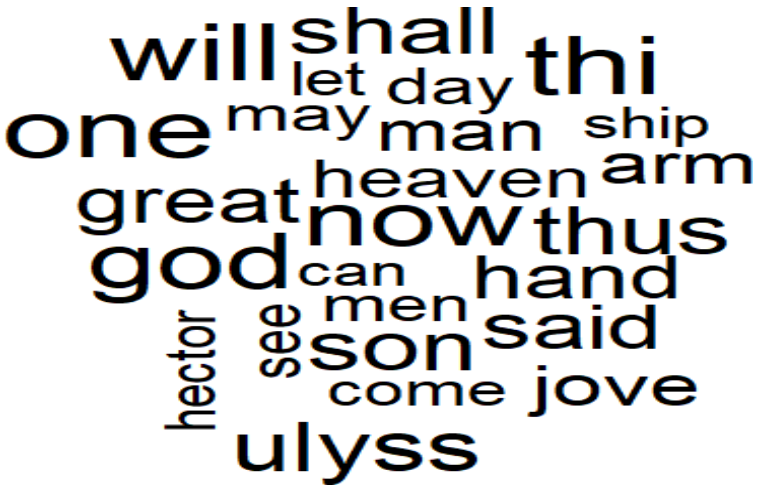
3. Word Cloud with maximum words 100 (*max.words=100*):



4. Word Cloud with minimum frequency 250 along with color Dark2:

5. Word Cloud with minimum frequency 250 and maximum words 200:

6. Word Cloud with minimum frequency 500 and maximum words 25:



7. Cluster Dendrogram: