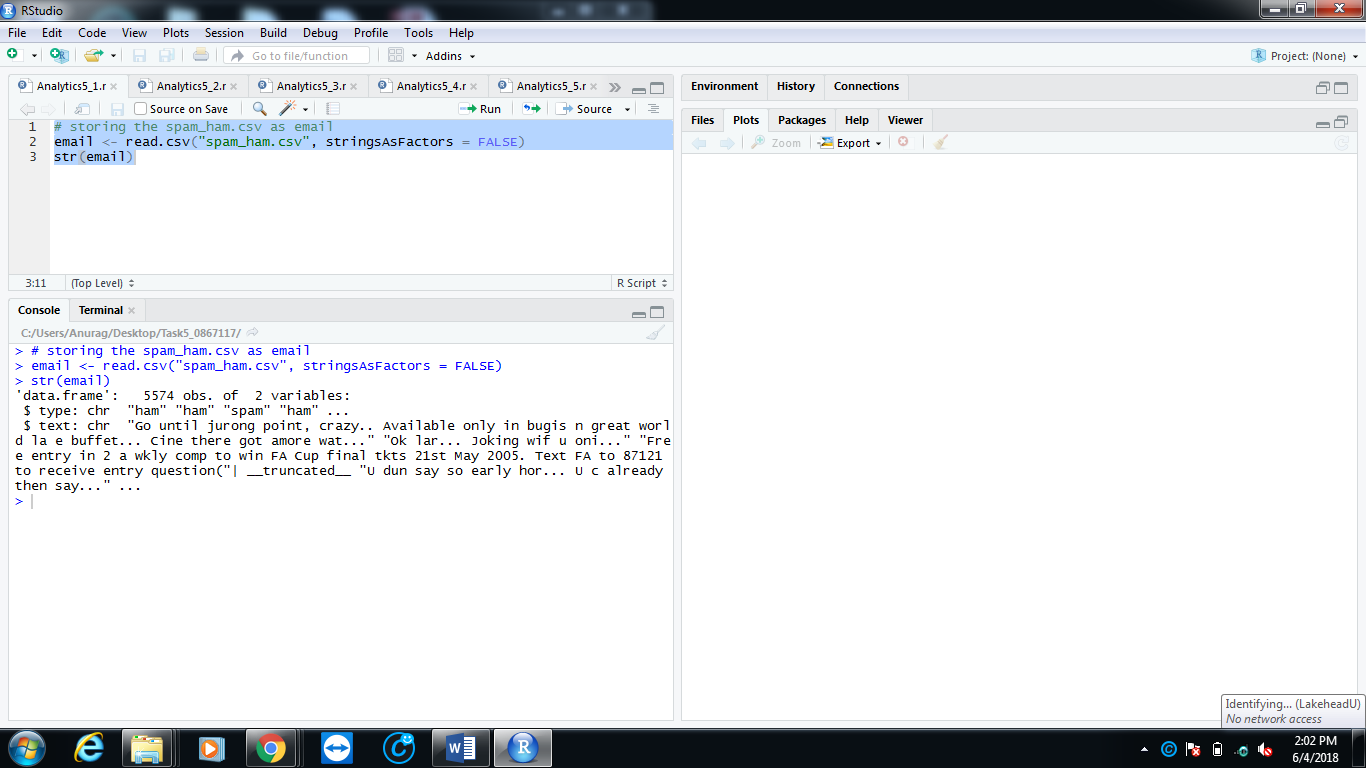
**Task-5**

**Analytics5\_1:**

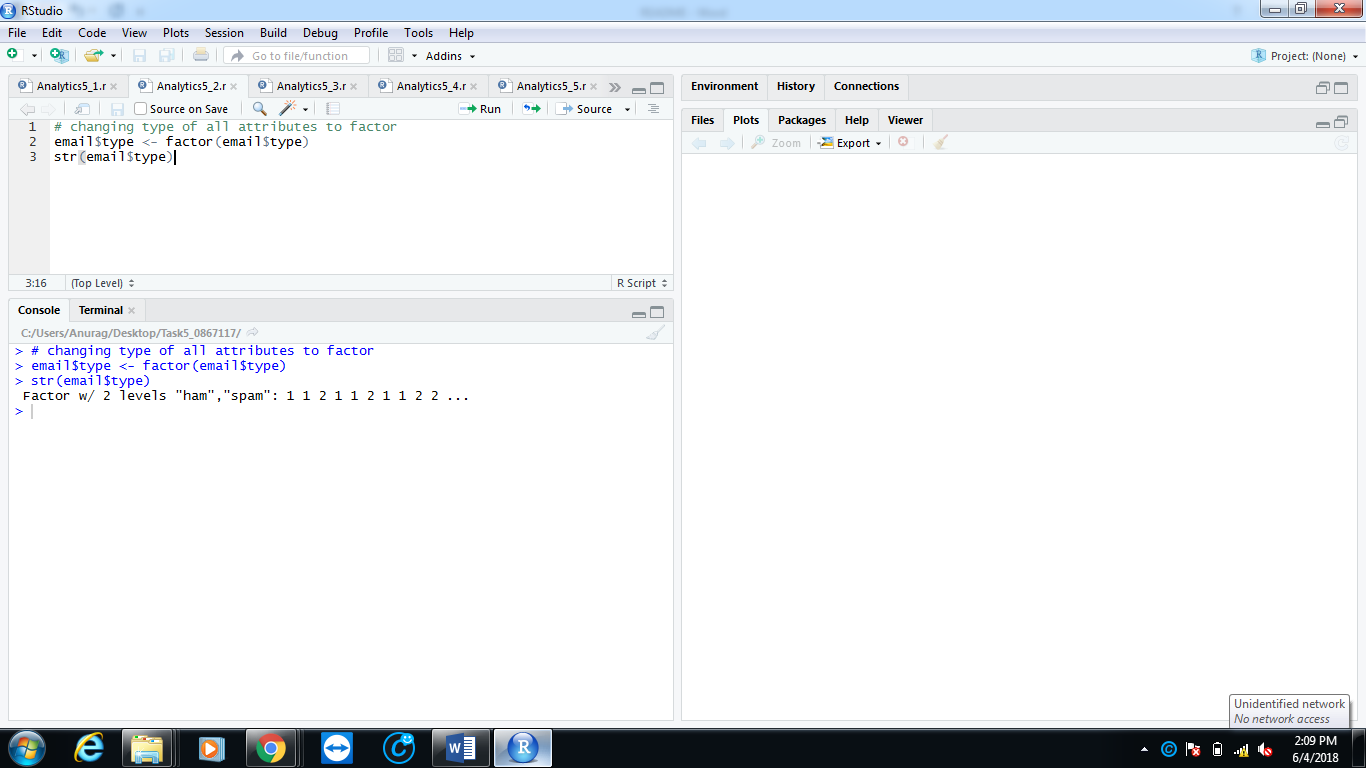
Code:

Code for storing spam\_ham.csv as email



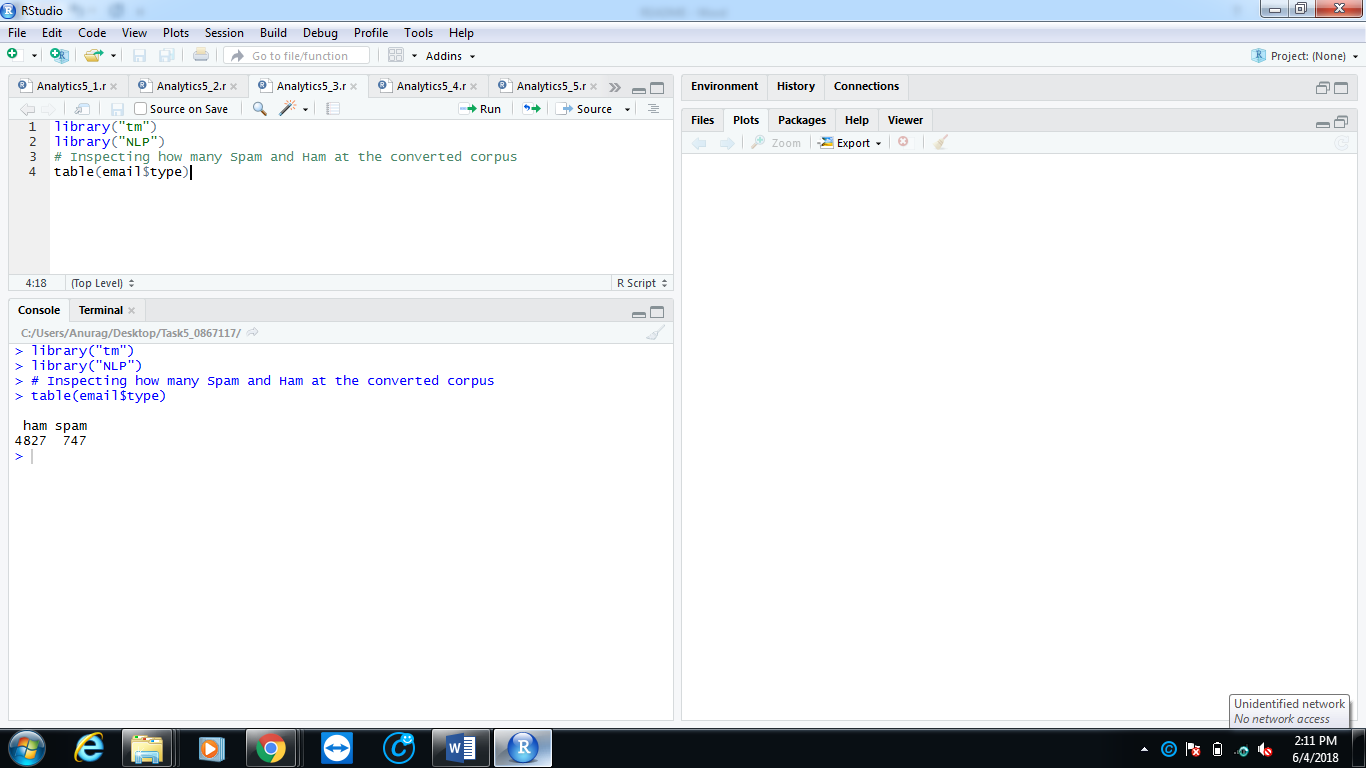
**Analytics5\_2:**

Code:

Code for changing type of all attributes to factor.

**Analytics5\_3:**

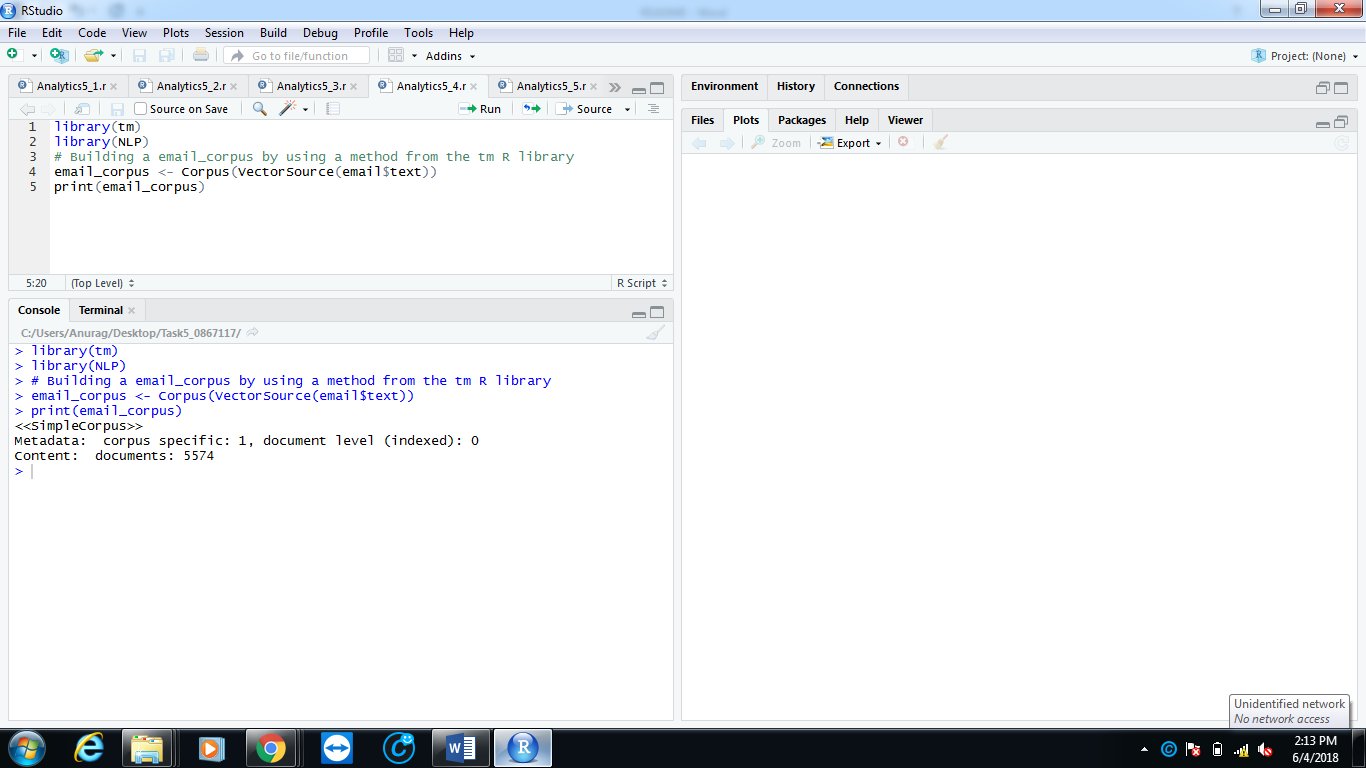
Code:

Code for inspecting how many spam and ham.

**Analytics5\_4:**

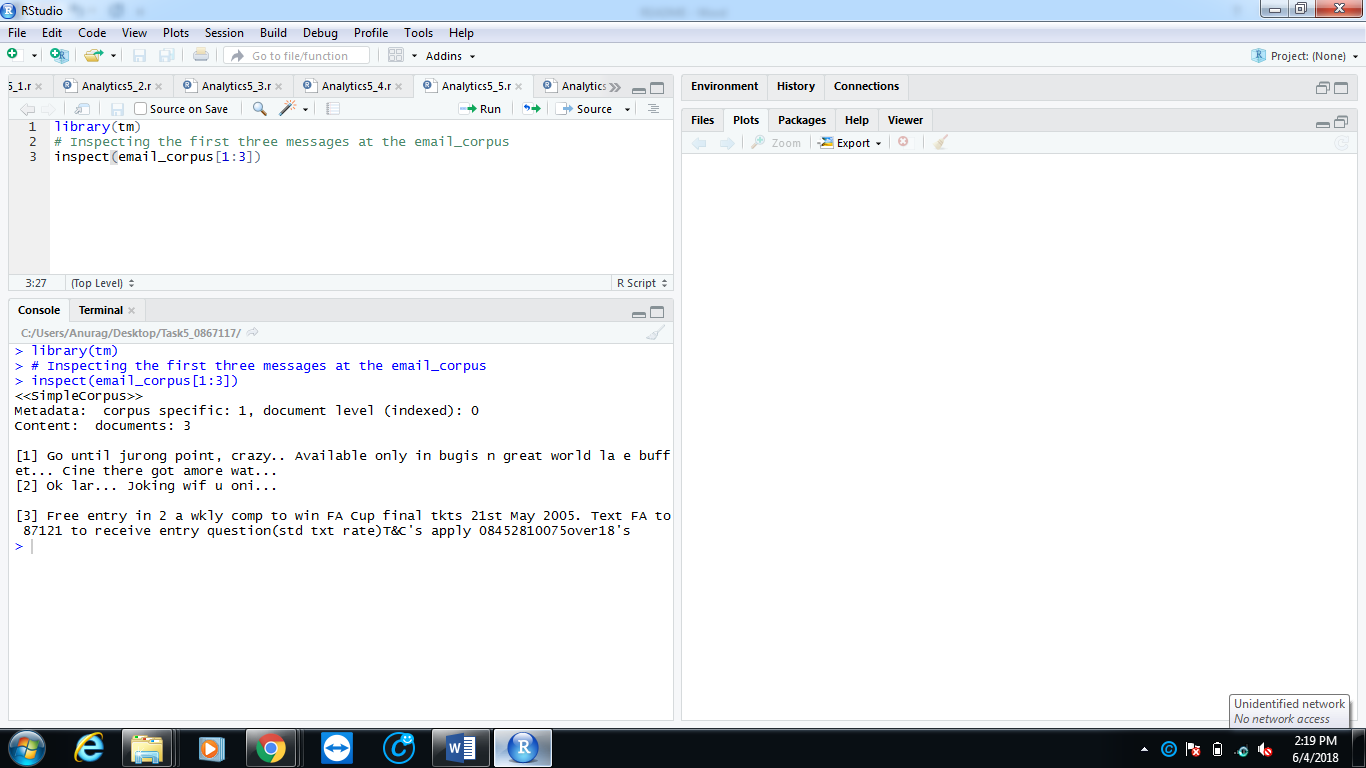
Code:

Code for building a email\_corpus using method from tm library.



**Analytics5\_5:**

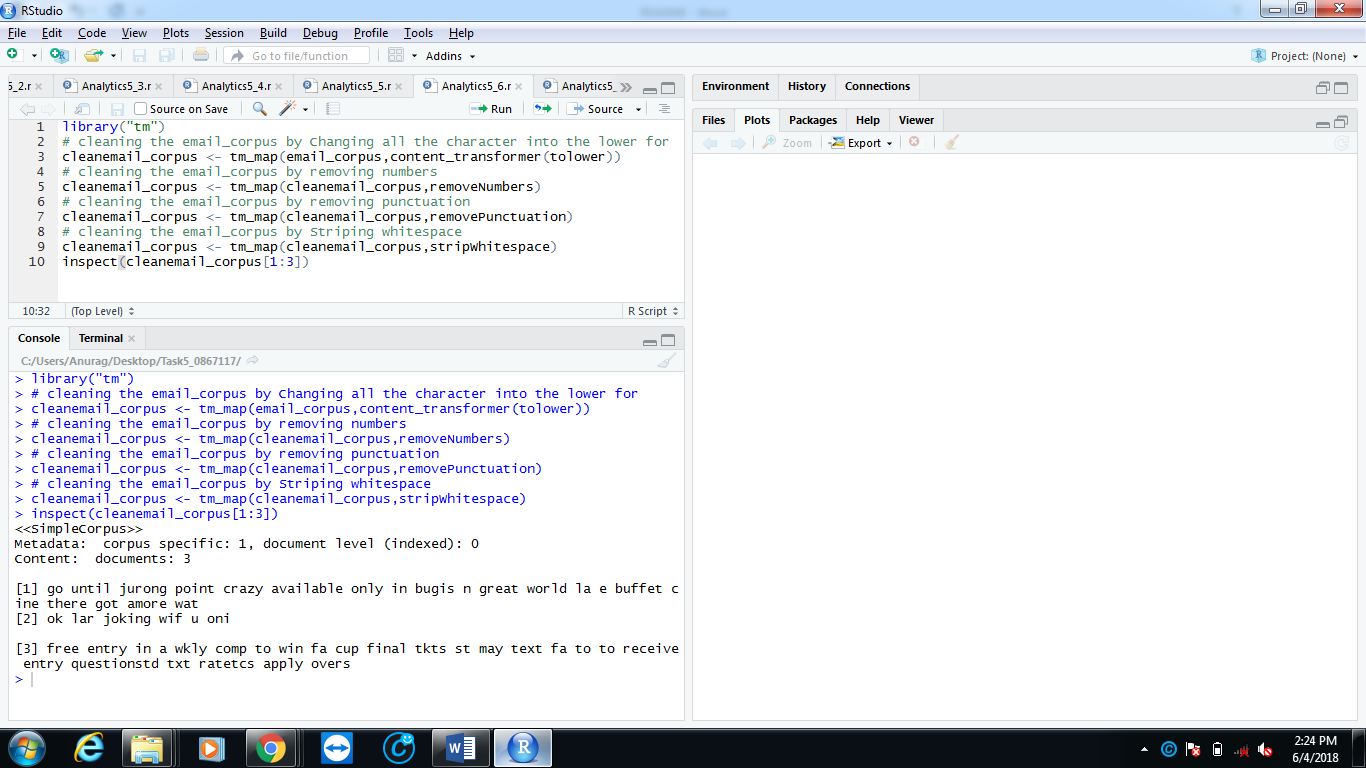
Code:

Code for inspecting first three messages at the email\_corpus.

**Analytics5\_6:**

Code:

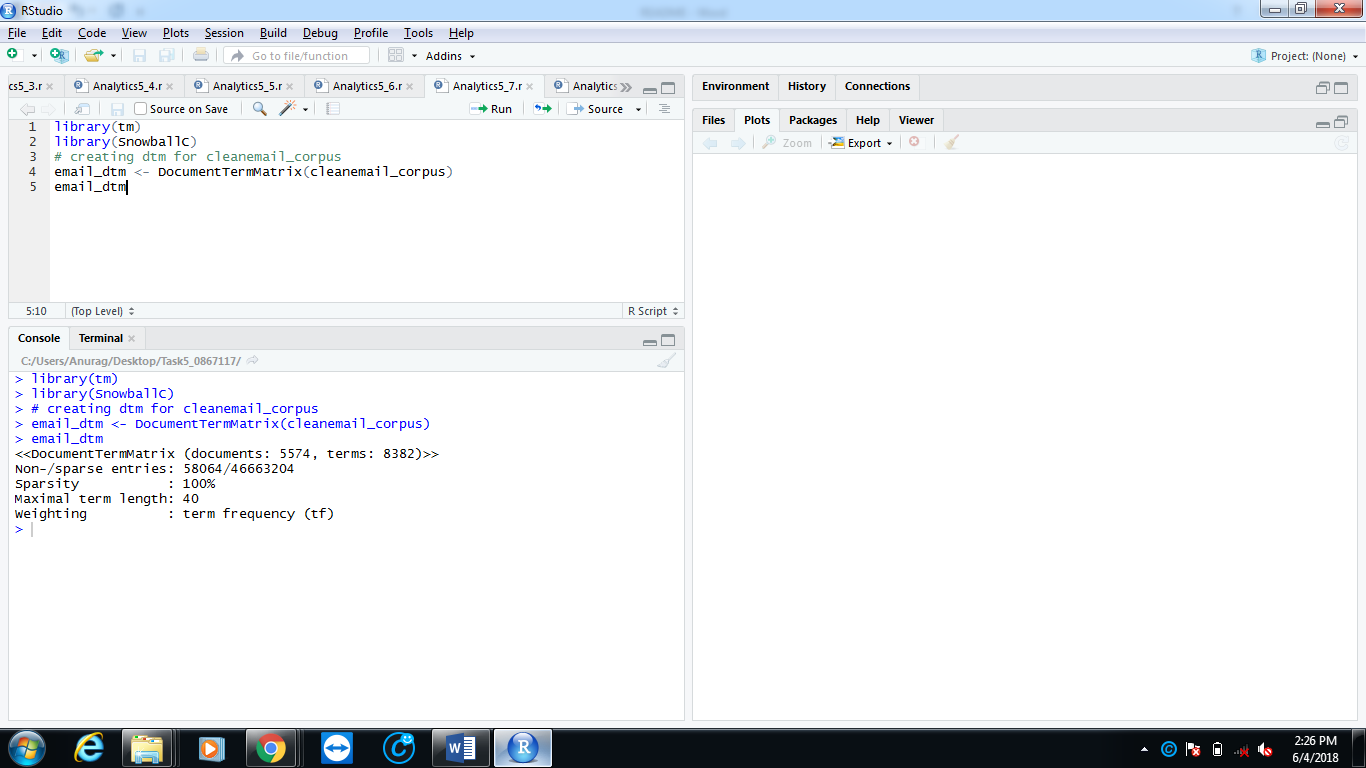
Code for cleaning corpus by changing alphabets to lower case, remove numbers, punctuations and strip whitespaces.



**Analytics5\_7:**

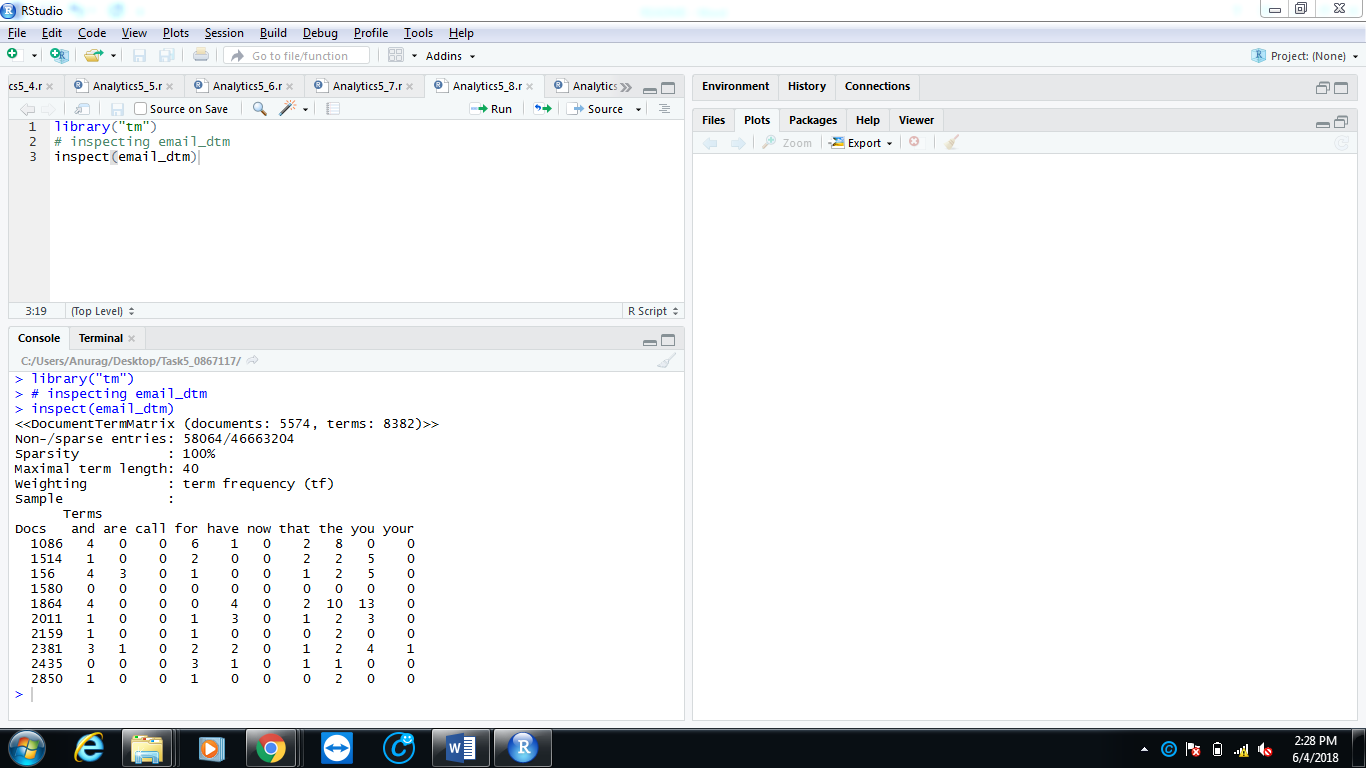
Code:

Code for creating email\_dtm for cleanemail\_corpus.



**Analytics5\_8:**

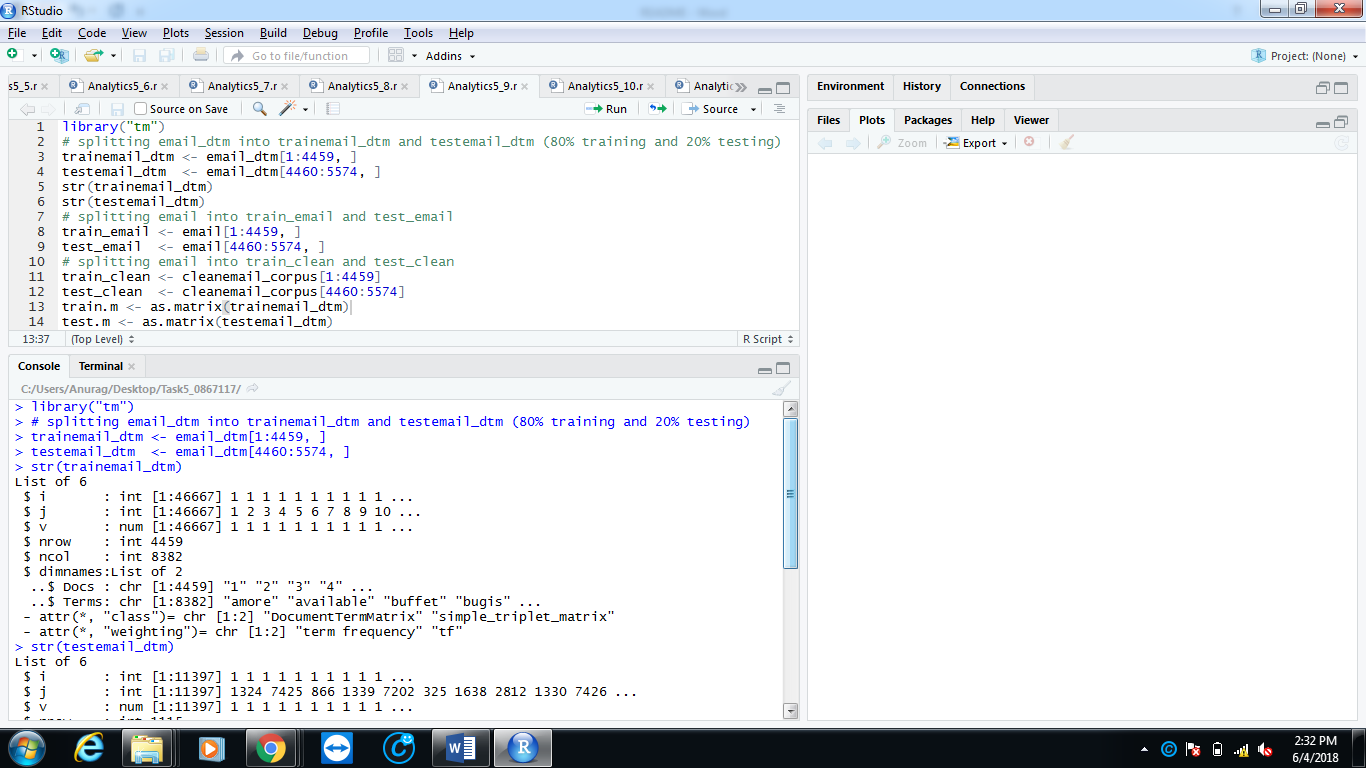
Code:

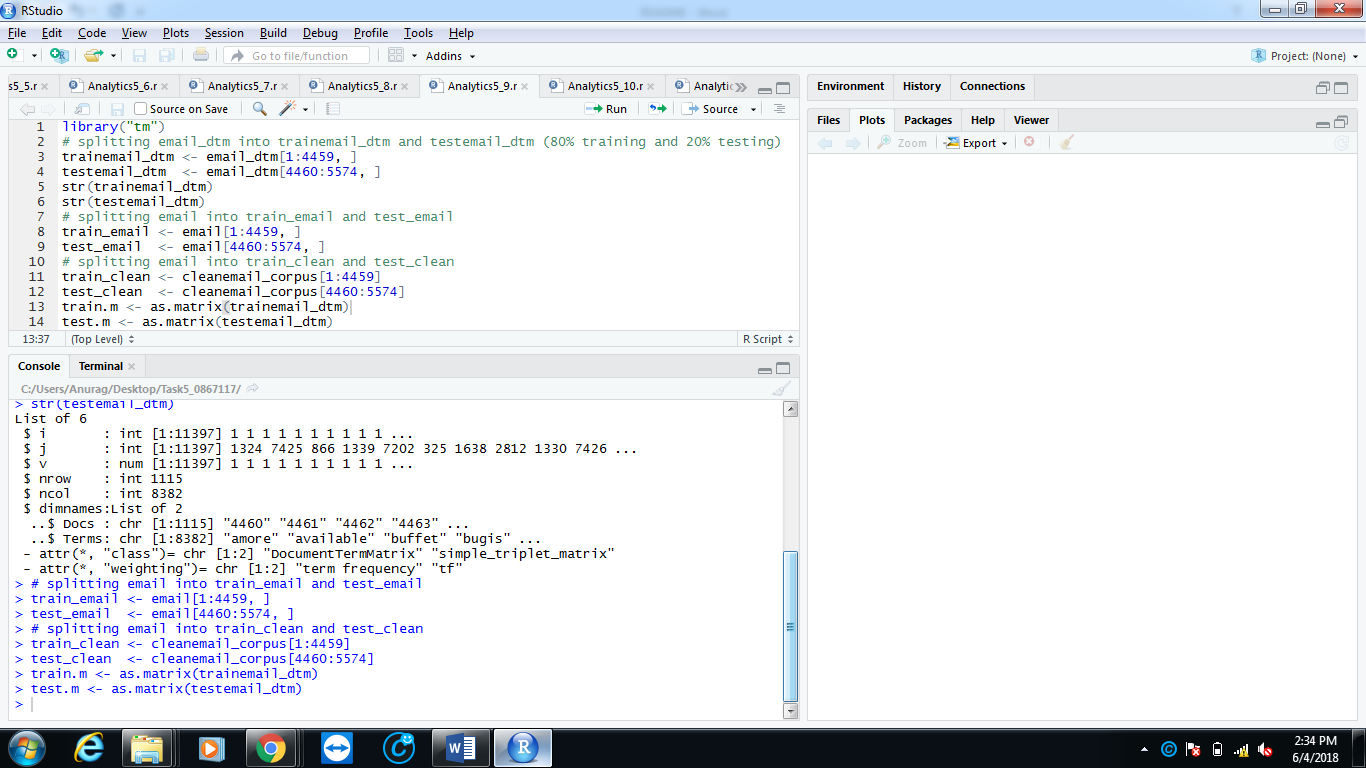
Code for inspecting email\_dtm.

**Analytics5\_9:**

Code:

Code for splitting email\_dtm into trainemail\_dtm and testemail\_dtm (80% training and 20% testing).



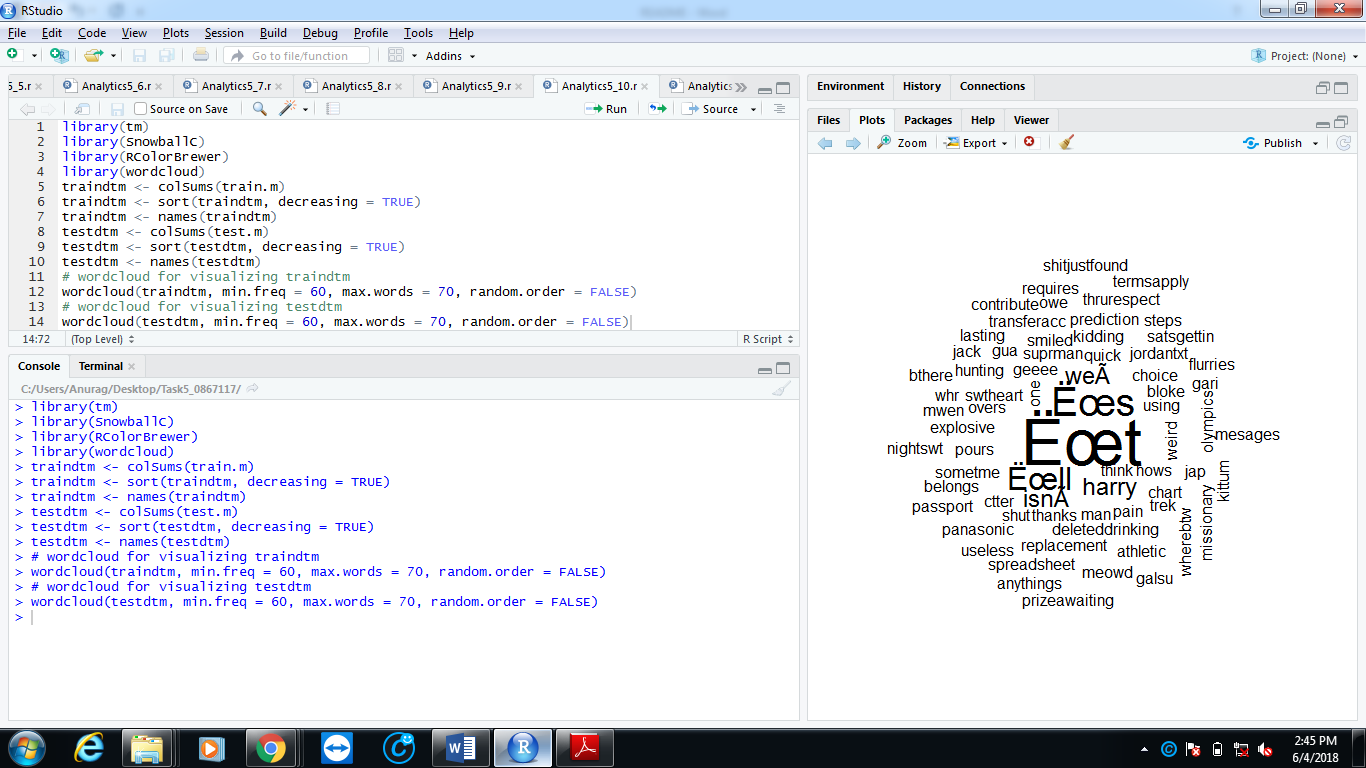


**Analytics5\_10:**

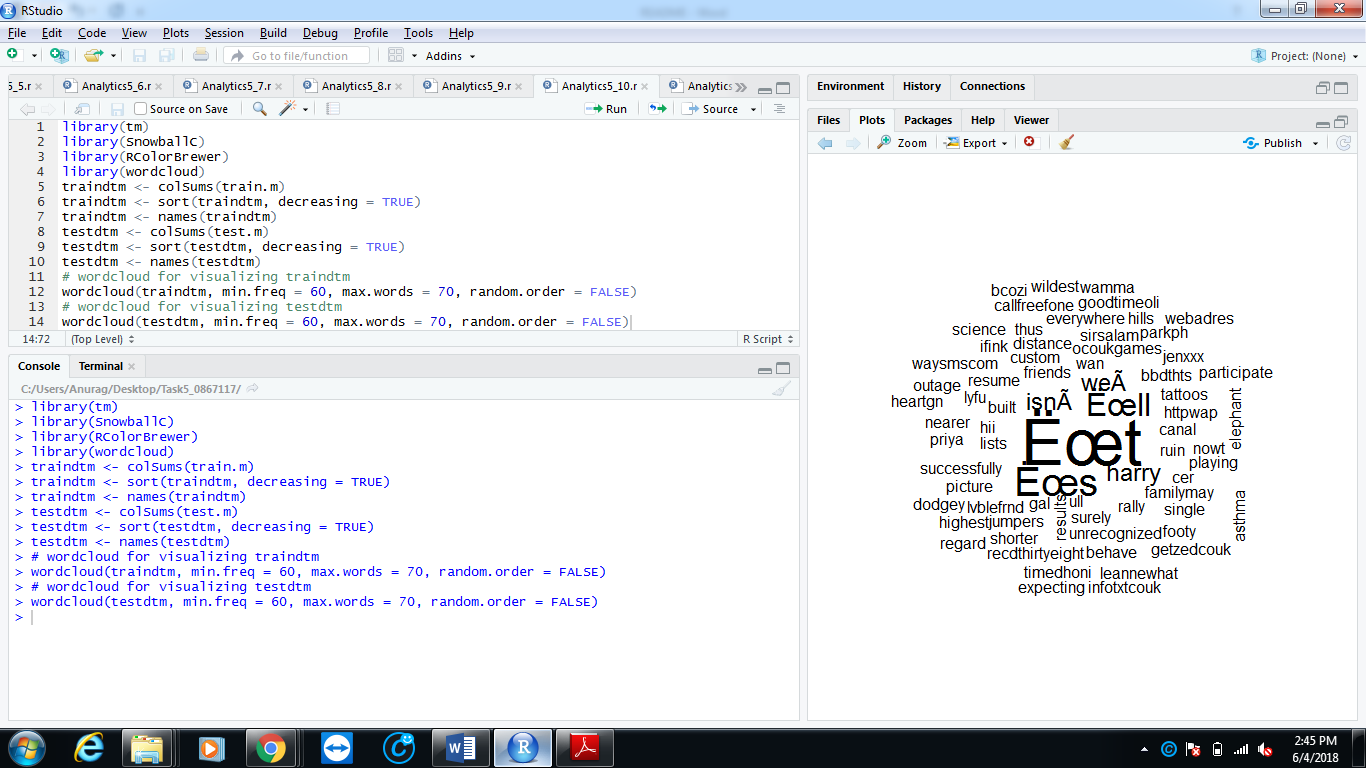
Code:

Code for visualizing traindtm and testdtm using library wordcloud.

For testdtm wordcloud:

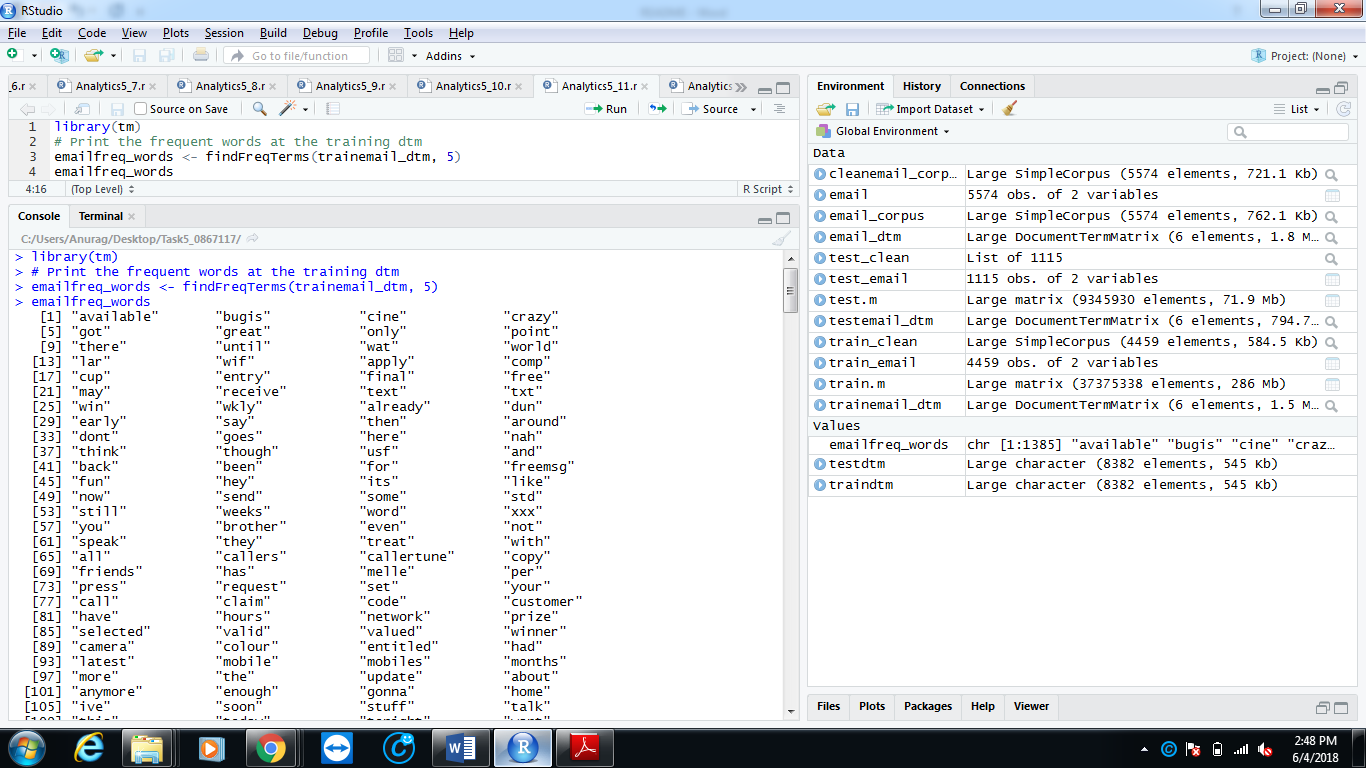


For testdtm wordcloud:-



**Analytics5\_11:**

Code:

Code for printing frequent words at training dtm.

**Analytics5\_12:**

Code:

Code for training a naive bayes classifier and predict its outcome by testedtm.

