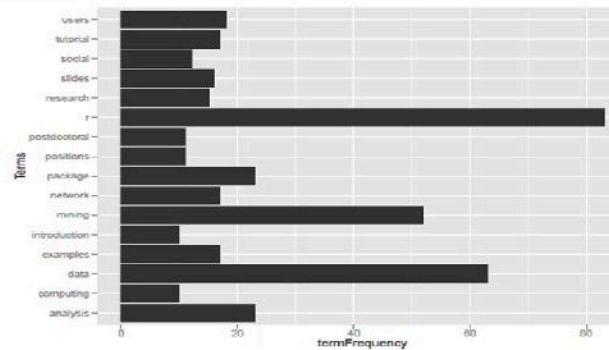


Step 2 – Data Processing

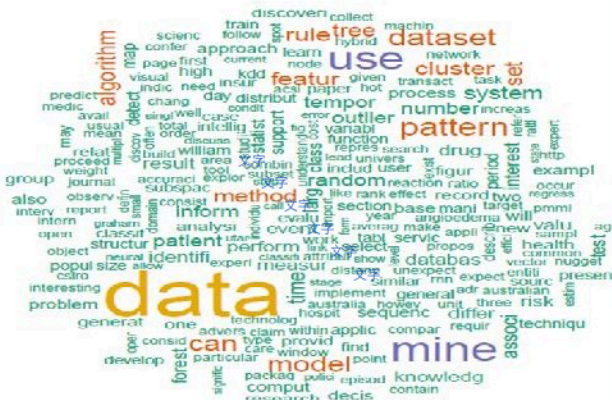
- 2A - Explore Corpus
- 2B - Convert text to lowercase
- 2C - Remove
 - a) Numbers(if required)
 - b) Punctuations
 - c) English stop words
 - d) Own stop words(if required)
 - e) Strip whitespace
 - f) Lemmatization/Stemming
 - g) Sparse terms
- 2D - Create document term matrix

Step 3 - Visualization

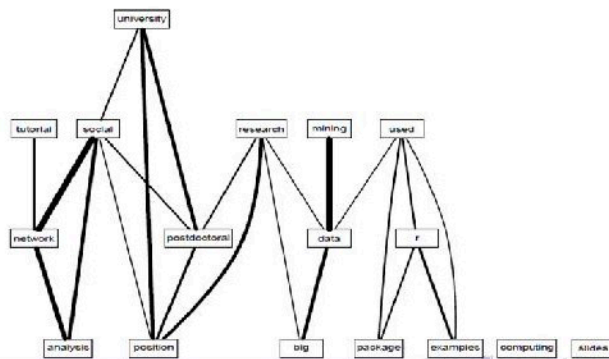
Frequency Items



Word Cloud



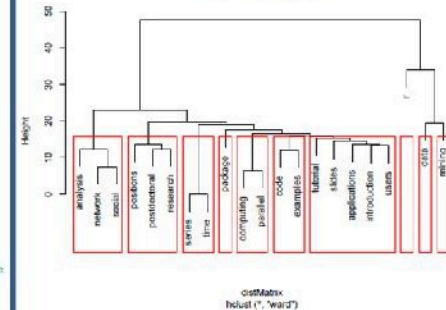
Correlation Plot



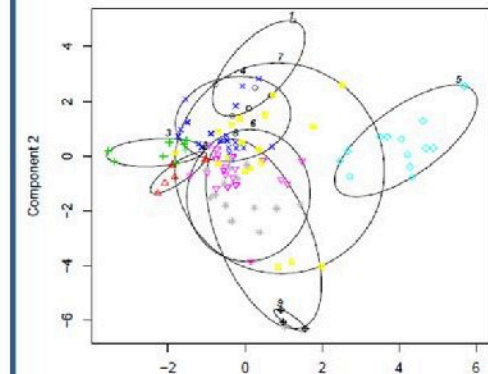
Step 4 – Run Model(s)

- ❖ Clustering
- ❖ Classification
- ❖ Sentiment Analysis

Cluster Dendrogram



```
clusplot(pam(x = sdata, k = k, diss = diss, metric = "manhattan"
```



Component 1
These two components explain 24.81 % of the point variability.