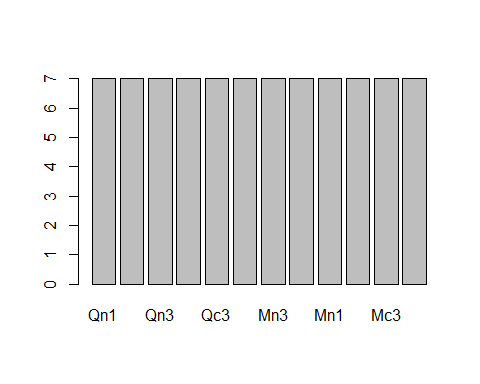
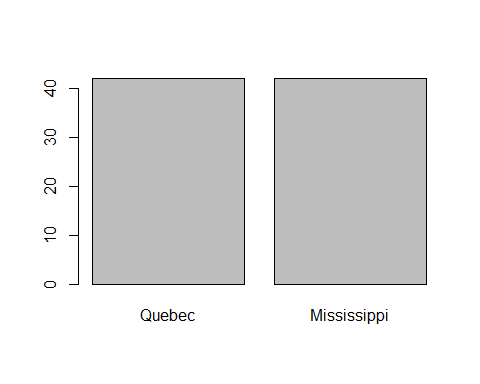
# Part1 Data visualization with base R

## Create bar graph of plant, type and treatment variable.

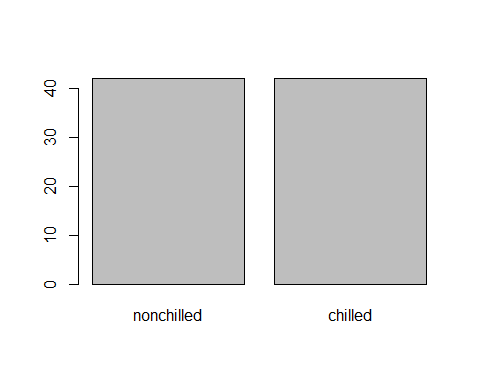
df <- data.frame(CO2)  
frequency\_Of\_plant <- table(df$Plant)  
barplot(frequency\_Of\_plant)



frequency\_Of\_type <- table(df$Type)  
barplot(frequency\_Of\_type)

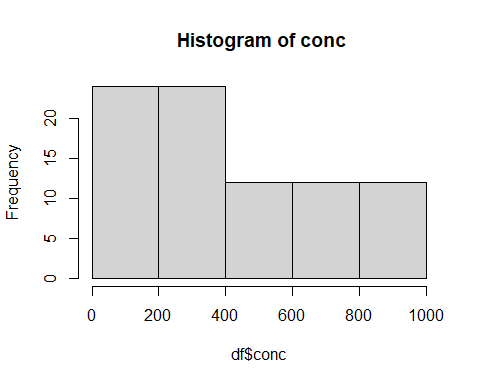


frequency\_of\_treatment <- table(df$Treatment)  
barplot(frequency\_of\_treatment)

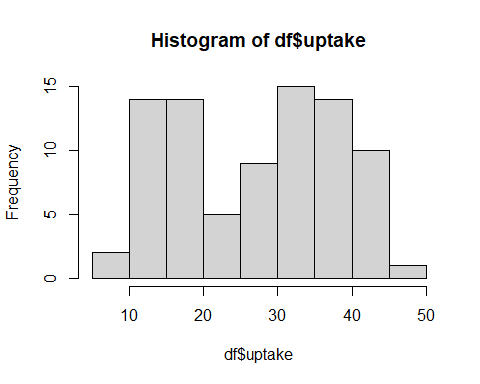
 `

## Create histogram of conc and uptake variable.

hist(df$conc, breaks = 5, main = " Histogram of conc")

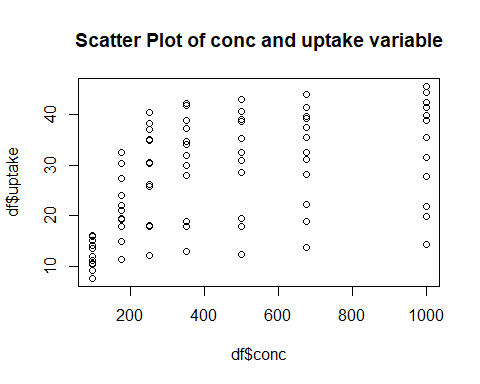


hist(df$uptake)



## Create the scatterplot of conc and uptake variables

plot(df$conc,df$uptake,main = "Scatter Plot of conc and uptake variable")



There is positive association between conc and uptake variable. Also our plot is not linear so approprate correlation coefficent is spearman correlation coefficent.

## 5.Compute the best correlation coefficent for conc and uptake variable and interpret the result carefully.

cor.test(x=df$conc,y=df$uptake, methods= "spearman")

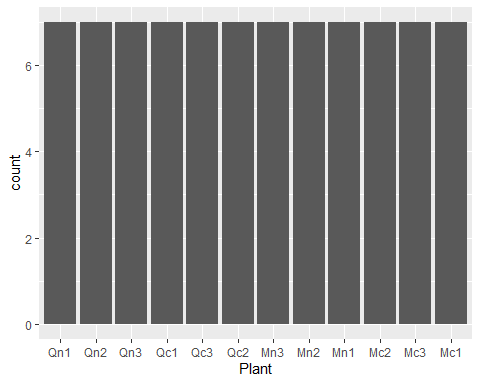
##   
## Pearson's product-moment correlation  
##   
## data: df$conc and df$uptake  
## t = 5.0245, df = 82, p-value = 2.906e-06  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3022189 0.6336595  
## sample estimates:  
## cor   
## 0.4851774

There is positive correlation between conc and upkate variables.

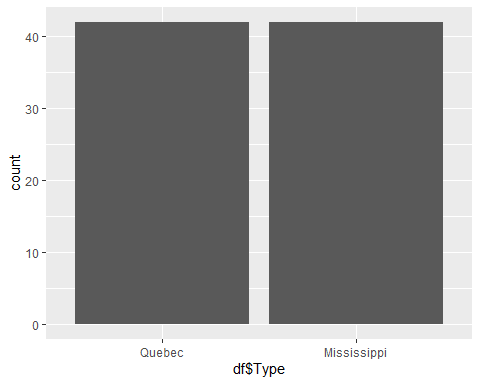
# Part2 from ggplot2

## create the bar diagram of plant, type and treatment variable

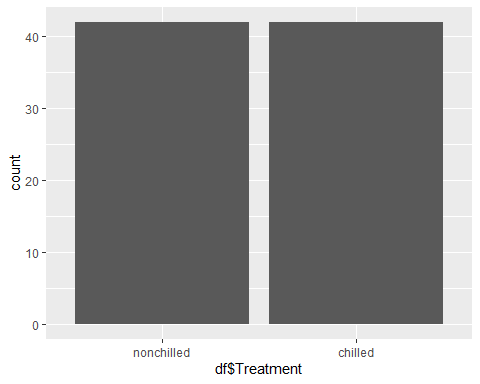
library(ggplot2)  
ggplot(CO2, aes(Plant), main = "Bar plot of Plant variable") + geom\_bar()



ggplot(CO2, aes(df$Type), main = "Bar plot of Type variable")+ geom\_bar()



ggplot(CO2, aes(df$Treatment), main = "Bar plot of Treatment variable") + geom\_bar()

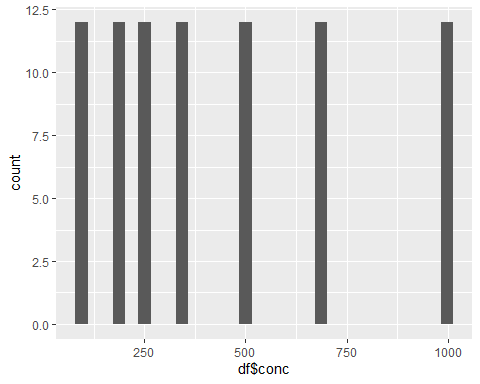


## Create the histograms of conc and uptake variables

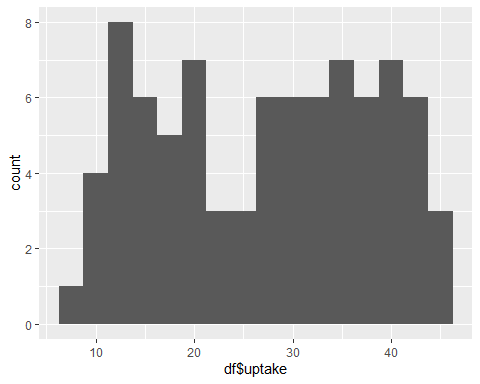
ggplot(CO2, aes(df$conc), main = "Histogram of conc variable")+ geom\_histogram(bin= 5)

## Warning: Ignoring unknown parameters: bin

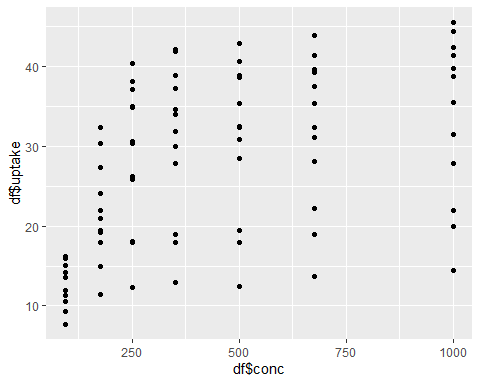
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



ggplot(CO2, aes(df$uptake), main = "Histogram of uptake variable") + geom\_histogram(binwidth = 2.5)



ggplot(CO2,aes(df$conc, df$uptake)) + geom\_point()

 There is positive association between conc and uptake variables. The approprate measure of correlation is spearman correlation.

cor.test(df$conc,df$uptake, method = "spearman")

## Warning in cor.test.default(df$conc, df$uptake, method = "spearman"): Cannot  
## compute exact p-value with ties

##   
## Spearman's rank correlation rho  
##   
## data: df$conc and df$uptake  
## S = 41483, p-value = 7.39e-09  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.5800041

We saw there is positive correlation between conc and uptake variables.

# Part3.Explain the advantages and limitations of using base R graphics package and ggplot2 package for doing visualization in R

Advantages of Base R

* R is free, open-source code. R is available under an open-source license, which means that anyone can download and modify the code.
* R runs anywhere.
* R supports extensions.
* R provides an engaged community.
* R connects with other languages.

Limitation of Base R

* Weak Origin. R shares its origin with a much older programming language “S”.
* Data Handling. In R, the physical memory stores the objects.
* Basic Security. R lacks basic security.
* Complicated Language. R is not an easy language to learn.
* Lesser Speed.
* Spread Across various Packages.

Advantages of ggplot2 \* It can do quick-and-dirty and complex

* The default colors and other aesthetics are nicer.
* Never again lose an axis title (or get told your pdf can’t be created) due to misspecified outer or inner margins.
* We can save plots (or the beginnings of a plot) as objects.
* Multivariate exploration is greatly simplified through faceting and coloring.
* Easily build plots in layers to tell a more complete story.
* Let our plots evolve (or devolve) with minimal changes to code

Disadvantages of ggplot2

* It is often slower than base R graphic.

library(tm)

## Loading required package: NLP

##   
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':  
##   
## annotate

library(twitteR)  
library(igraph)

##   
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':  
##   
## decompose, spectrum

## The following object is masked from 'package:base':  
##   
## union

library(wordcloud)

## Loading required package: RColorBrewer

load("F:/MDS R/termDocMatrix.rdata")

## Convert this data as matrix

m <- as.matrix(termDocMatrix)

## Get the term frequencies

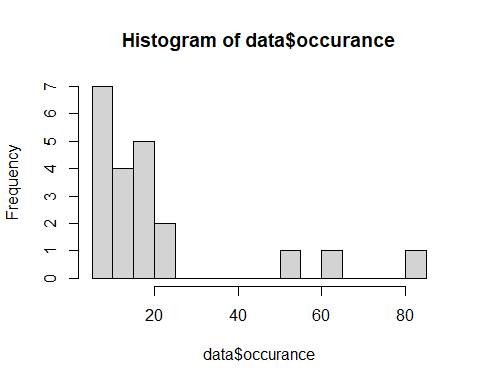
freq <- sort(rowSums(m), decreasing=T)  
freq

## r data mining analysis package users   
## 83 63 52 23 23 18   
## examples network tutorial slides research social   
## 17 17 17 16 15 12   
## positions postdoctoral computing introduction applications code   
## 11 11 10 10 9 9   
## parallel series time   
## 8 8 8

r Word has frequency 83 which is maximun among other words, data and mining has frequency 63 and 52 respectively.

## Create the histogram of the term frequencies

data <- data.frame(term = names(freq), occurance = freq)  
hist(data$occurance, breaks= 20)

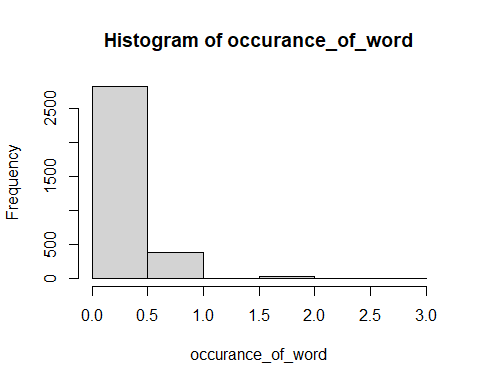


## Create the histogram of the terms with frequencies of 5 and more

m<- as.matrix(termDocMatrix)  
occurance\_of\_word<- m[rowSums(m)> 5,]  
occurance\_of\_word

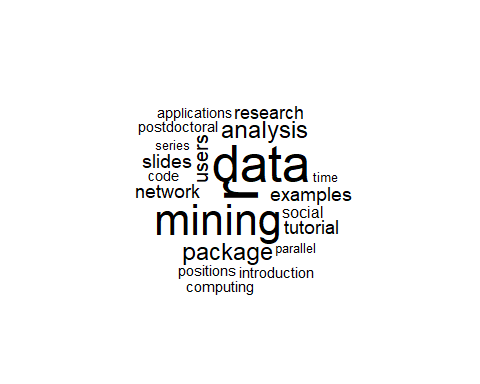
## Docs  
## Terms 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24  
## analysis 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 1 0 0 1  
## applications 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## code 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## computing 0 0 1 1 0 1 1 1 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## data 1 1 0 0 2 0 0 0 0 0 1 2 1 1 1 0 1 0 0 0 0 0 0 0  
## examples 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## introduction 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## mining 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0 0 0 0 0 0  
## network 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 1 1 0 1  
## package 0 0 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## parallel 0 0 1 1 0 1 1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## positions 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1  
## postdoctoral 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1  
## r 0 0 1 1 1 1 1 1 1 1 0 1 0 0 0 0 0 0 1 0 1 1 0 0  
## research 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0  
## series 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## slides 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0  
## social 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 1 1 0 0 1  
## time 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## tutorial 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 1 0 0  
## users 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## Docs  
## Terms 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45  
## analysis 0 0 1 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## applications 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0  
## code 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## computing 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## data 0 0 0 1 0 0 0 1 0 0 1 1 0 0 0 0 0 1 0 0 1  
## examples 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## introduction 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## mining 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 1 0 1  
## network 0 0 1 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## package 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0  
## parallel 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## positions 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 1 1  
## postdoctoral 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1  
## r 1 0 0 1 0 1 0 0 1 0 1 1 0 0 0 0 1 1 0 0 0  
## research 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0  
## series 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## slides 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0  
## social 0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## time 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## tutorial 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## users 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## Docs  
## Terms 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66  
## analysis 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## applications 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## code 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## computing 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0  
## data 1 1 0 0 1 0 0 0 1 1 0 0 1 1 0 1 0 1 0 1 1  
## examples 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## introduction 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## mining 1 0 0 0 1 0 0 0 0 1 0 0 0 1 0 1 1 1 0 1 1  
## network 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## package 0 0 0 0 0 0 0 0 0 1 0 0 0 0 2 0 0 0 0 0 0  
## parallel 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## positions 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## postdoctoral 0 0 0 0 0 0 0 1 0 0 0 1 1 1 0 0 0 1 0 0 0  
## r 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 1 0 0  
## research 0 0 0 0 0 0 0 1 2 0 0 0 3 1 0 0 0 0 0 0 0  
## series 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## slides 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## social 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## time 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## tutorial 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0  
## users 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0  
## Docs  
## Terms 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87  
## analysis 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## applications 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## code 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0  
## computing 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## data 0 0 0 1 0 0 0 1 0 0 0 2 0 0 0 0 1 0 0 0 0  
## examples 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0  
## introduction 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## mining 0 0 0 1 0 0 1 0 0 1 0 1 0 1 0 0 1 0 0 1 1  
## network 0 0 1 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0  
## package 0 0 0 0 1 0 0 0 1 0 0 1 0 0 0 0 0 1 0 0 0  
## parallel 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## positions 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## postdoctoral 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## r 2 1 0 0 0 0 1 1 1 0 1 1 0 0 0 1 0 1 1 0 0  
## research 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## series 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## slides 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
## social 0 0 1 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0  
## time 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## tutorial 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0  
## users 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## Docs  
## Terms 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106  
## analysis 0 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 0 0  
## applications 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## code 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0  
## computing 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## data 1 0 1 0 0 1 0 0 0 1 0 0 0 0 0 0 3 0 1  
## examples 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 1 0  
## introduction 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## mining 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 3  
## network 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## package 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0  
## parallel 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## positions 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## postdoctoral 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## r 0 0 0 1 2 0 2 1 0 0 0 0 1 1 1 0 0 2 0  
## research 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0  
## series 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## slides 0 0 0 0 0 1 0 0 0 0 0 0 1 1 1 1 0 0 0  
## social 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## time 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## tutorial 0 0 0 0 1 2 1 0 0 0 0 0 0 0 0 0 0 0 0  
## users 0 0 0 0 0 0 1 0 0 0 0 0 1 0 1 0 0 0 0  
## Docs  
## Terms 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122  
## analysis 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0  
## applications 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## code 0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0  
## computing 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## data 0 0 1 0 0 0 0 1 0 0 1 0 1 0 1 0  
## examples 0 1 0 0 0 1 0 0 0 0 1 0 1 0 1 0  
## introduction 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0  
## mining 0 1 0 0 0 0 0 0 0 0 1 0 1 1 0 0  
## network 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## package 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## parallel 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## positions 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## postdoctoral 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## r 0 1 1 1 0 1 2 1 0 0 1 1 1 1 1 1  
## research 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## series 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0 0  
## slides 1 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1  
## social 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## time 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0 0  
## tutorial 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## users 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1  
## Docs  
## Terms 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138  
## analysis 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0  
## applications 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## code 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0  
## computing 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## data 1 1 0 0 0 0 0 0 0 0 1 0 0 0 0 1  
## examples 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0  
## introduction 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
## mining 0 1 0 1 0 1 1 0 0 1 1 0 0 0 0 1  
## network 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## package 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## parallel 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## positions 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## postdoctoral 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## r 0 0 0 2 0 2 1 0 1 0 0 0 0 1 2 1  
## research 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## series 0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0  
## slides 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## social 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## time 0 0 0 1 1 1 0 0 0 0 0 0 0 1 0 0  
## tutorial 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0  
## users 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0  
## Docs  
## Terms 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154  
## analysis 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0  
## applications 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1  
## code 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## computing 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## data 2 0 0 0 2 0 2 1 0 0 2 0 1 1 0 2  
## examples 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## introduction 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## mining 2 0 0 0 1 0 0 0 0 0 2 0 1 1 0 2  
## network 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0  
## package 0 0 1 2 0 0 1 0 0 0 0 0 0 1 0 1  
## parallel 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## positions 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## postdoctoral 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## r 0 2 2 2 1 0 1 1 1 0 1 0 1 1 0 2  
## research 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## series 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## slides 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## social 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## time 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## tutorial 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## users 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 1

hist(occurance\_of\_word, breaks = 10)



## Create the words cloud of term frequency

m <- as.matrix(termDocMatrix)  
freq <- sort(rowSums(m), decreasing=T)  
wordcloud(words=names(freq), freq=freq, min.freq=4,  
random.order=F)

 From above words clouds we see that the word r has maximum frequency hence it is appeared in big size. Similarly data and mining have also maximum frequency

## Perform social network analysis of the termDocumentMatrix data and interpret it carefully

termM <- m %\*% t(m)  
termM[1:10,1:10]

## Terms  
## Terms analysis applications code computing data examples introduction  
## analysis 23 0 1 0 4 4 2  
## applications 0 9 0 0 8 0 0  
## code 1 0 9 0 1 6 0  
## computing 0 0 0 10 2 0 0  
## data 4 8 1 2 85 5 3  
## examples 4 0 6 0 5 17 2  
## introduction 2 0 0 0 3 2 10  
## mining 4 7 3 1 50 5 3  
## network 12 0 1 0 0 2 2  
## package 2 1 0 2 12 2 0  
## Terms  
## Terms mining network package  
## analysis 4 12 2  
## applications 7 0 1  
## code 3 1 0  
## computing 1 0 2  
## data 50 0 12  
## examples 5 2 2  
## introduction 3 2 0  
## mining 64 1 6  
## network 1 17 1  
## package 6 1 27

Now we have built a term-term adjacency matrix, where the rows and columns represents terms, and every entry is the number of co-occurrences of two terms. Next we can build a graph with graph.adjacency() from package igraph.

g <- graph.adjacency(termM,weighted = T,mode = 'undirected')  
g

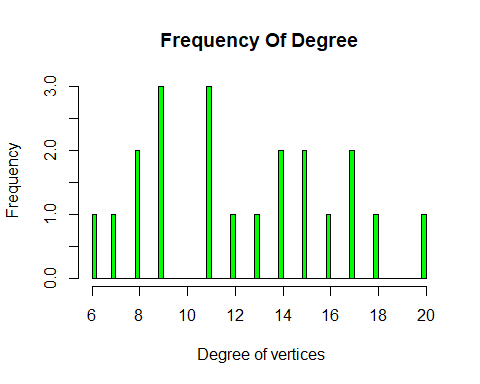
## IGRAPH a7a54ae UNW- 21 151 --   
## + attr: name (v/c), weight (e/n)  
## + edges from a7a54ae (vertex names):  
## [1] analysis --analysis analysis --code   
## [3] analysis --data analysis --examples   
## [5] analysis --introduction analysis --mining   
## [7] analysis --network analysis --package   
## [9] analysis --positions analysis --postdoctoral  
## [11] analysis --r analysis --research   
## [13] analysis --series analysis --slides   
## [15] analysis --social analysis --time   
## + ... omitted several edges

g <- simplify(g)  
g

## IGRAPH a7a8839 UNW- 21 130 --   
## + attr: name (v/c), weight (e/n)  
## + edges from a7a8839 (vertex names):  
## [1] analysis --code analysis --data   
## [3] analysis --examples analysis --introduction  
## [5] analysis --mining analysis --network   
## [7] analysis --package analysis --positions   
## [9] analysis --postdoctoral analysis --r   
## [11] analysis --research analysis --series   
## [13] analysis --slides analysis --social   
## [15] analysis --time analysis --tutorial   
## + ... omitted several edges

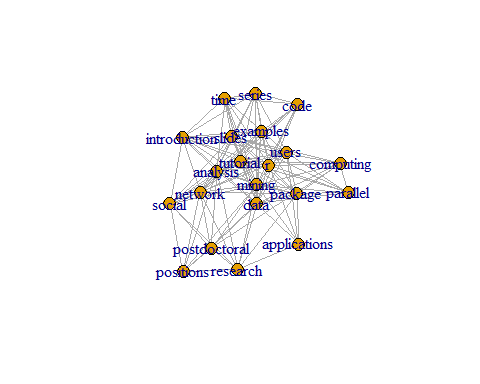
V(g)$label <- V(g)$name  
V(g)$degree <- degree(g)

hist(V(g)$degree, breaks = 100,col = 'green', main = "Frequency Of Degree", xlab = " Degree of vertices", ylab = " Frequency")

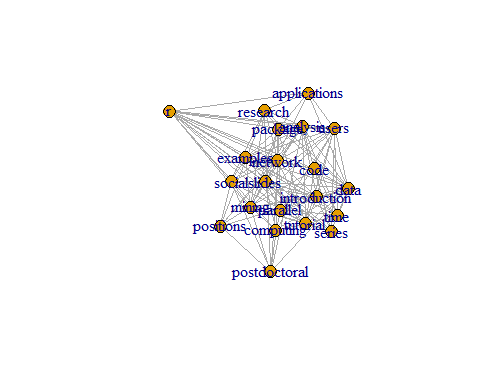


Plot of graph

set.seed(3952)  
  
layout1 <- layout.fruchterman.reingold(g)  
  
plot(g, layout=layout1)

 A different layout can be generated with the first line of code below. The second line produces an interactive plot, which allows us to manually rearrange the layout

plot(g, layout=layout.kamada.kawai)

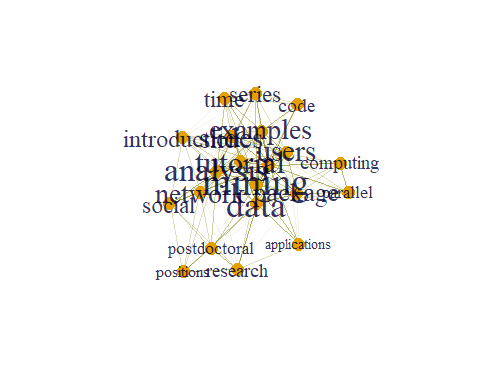


tkplot(g, layout=layout.kamada.kawai)

## [1] 1

Make this look better

V(g)$label.cex <- 2.2 \* V(g)$degree / max(V(g)$degree)+ .2  
  
V(g)$label.color <- rgb(0, 0, .2, .8)  
  
V(g)$frame.color <- NA  
  
egam <- (log(E(g)$weight)+.4) / max(log(E(g)$weight)+.4)  
  
E(g)$color <- rgb(.5, .5, 0, egam)  
  
E(g)$width <- egam  
  
# plot the graph in layout1  
  
plot(g, layout=layout1)

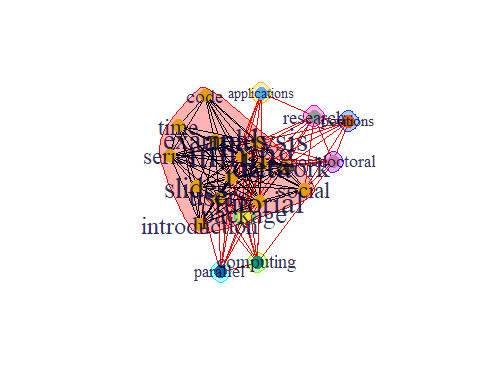
 ###Community detection

comm <- cluster\_edge\_betweenness(g)

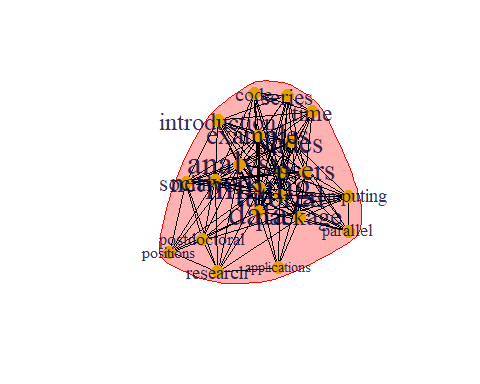
## Warning in cluster\_edge\_betweenness(g): At community.c:461 :Membership vector  
## will be selected based on the lowest modularity score.

## Warning in cluster\_edge\_betweenness(g): At community.c:468 :Modularity  
## calculation with weighted edge betweenness community detection might not make  
## sense -- modularity treats edge weights as similarities while edge betwenness  
## treats them as distances

plot(comm,g)

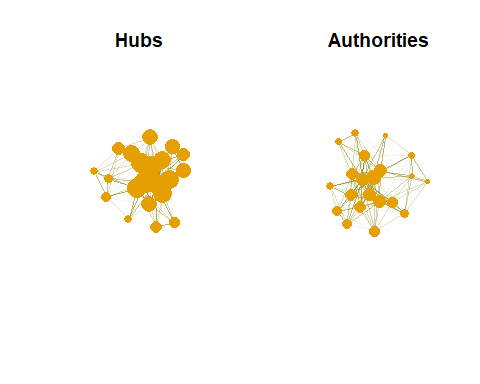
 There are dense connection within the group within the community the connection is sparse.

prop <- cluster\_label\_prop(g)  
plot(prop, g)

 This is another algorithms for community detection we get another type graph which is different from previous one.

hs <- hub\_score(g,weights = NA)$vector  
as <- authority\_score(g, weights = NA)$vector

par(mfrow = c(1,2))  
plot(g,vertex.size= hs\*50, main = "Hubs",  
 vertex.label = NA,  
 vertex.colour = rainbow(50))  
plot(g,vertex.size= as\*30, main = "Authorities",  
 vertex.label = NA,  
 vertex.colour = rainbow(50))

 Hubs are expected to contain large number of outgoing link. And authorities are expected to contaion large number of incoming link from hubs.