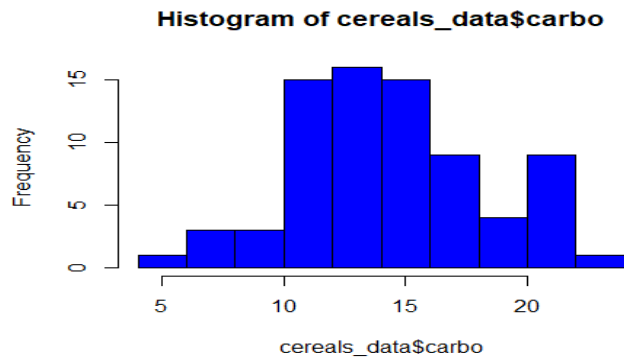


R markdown- TASK1

(Developing an appropriate approach to address missing values)

```
library(readr)
cereals_data <- read.csv("C:/Users/hp/Desktop/R- document/project2/cereals_data.csv")

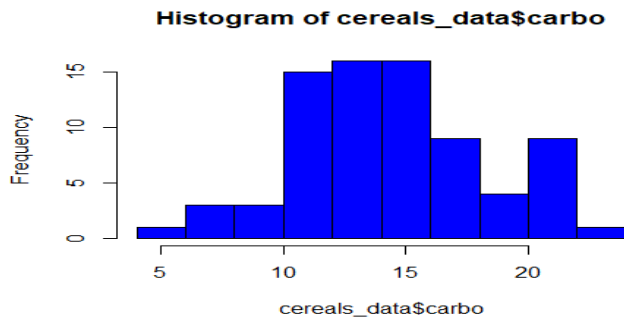
#missing value of carbo
hist(cereals_data$carbo,col= "BLUE") # the graph is almost normally distributed
```



```
mean(cereals_data$carbo,na.rm = T)

## [1] 14.80263

# replacing NA value with the mean of the variable
cereals_data$carbo[is.na(cereals_data$carbo)]<-mean(cereals_data$carbo,na.rm = T)
hist(cereals_data$carbo,col= "BLUE")
```

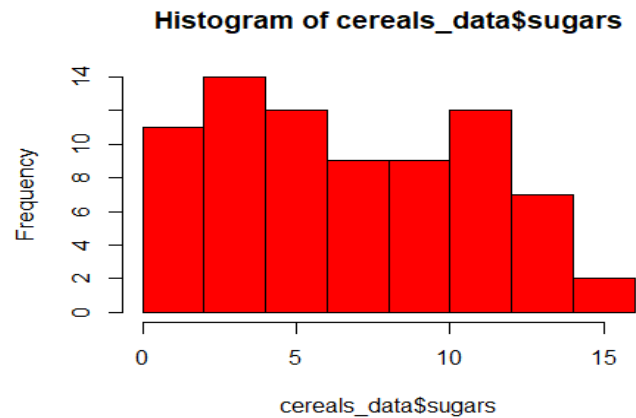


```
summary(cereals_data$carbo)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       5.0   12.0   14.8   14.8   17.0   23.0
```

```
# missing value of sugars
```

```
hist(cereals_data$sugars, col="red") # the graph is not normally distributed
```



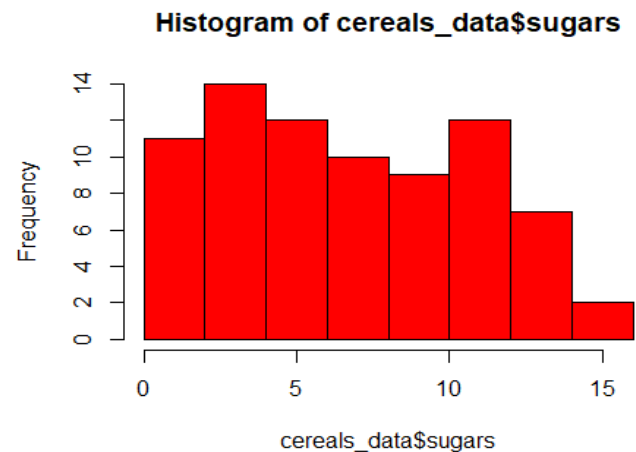
```
median(cereals_data$sugars, na.rm = T)
```

```
## [1] 7
```

```
#replacing the nA with median of the variable
```

```
cereals_data$sugars[is.na(cereals_data$sugars)]<- median(cereals_data$sugars,  
na.rm = T)
```

```
hist(cereals_data$sugars, col="red")
```

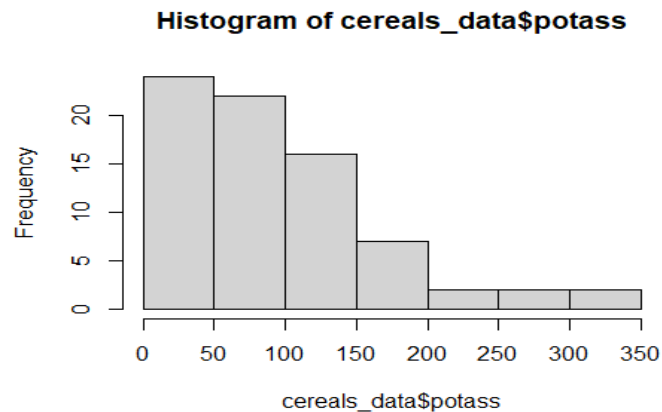


```
summary(cereals_data$sugars)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.000   3.000   7.000   7.026  11.000  15.000
```

```
# missing value of pottass
```

```
hist(cereals_data$potass)# the graph is not normally distributed
```



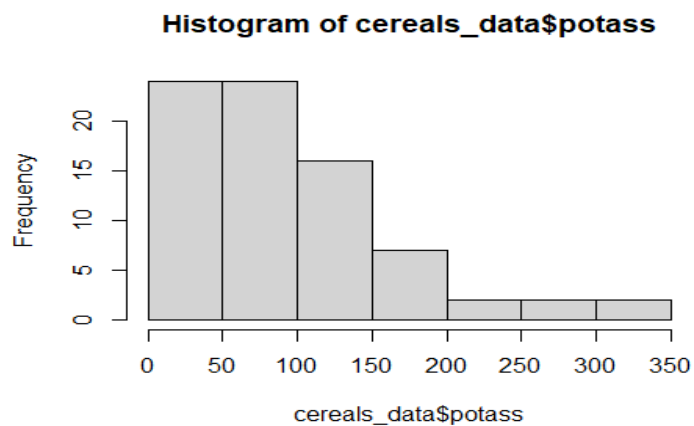
```
median(cereals_data$potass,na.rm = T)
```

```
## [1] 90
```

```
#replacing NA with the median
```

```
cereals_data$potass[is.na(cereals_data$potass)]<- median(cereals_data$potass,  
na.rm = T)
```

```
hist(cereals_data$potass)
```



```
summary(cereals_data$potass)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    15.00   45.00   90.00   98.44  120.00  330.00
```

R markdown- TASK2

(Evaluating all 16 Variables and Understanding and observe interesting findings)

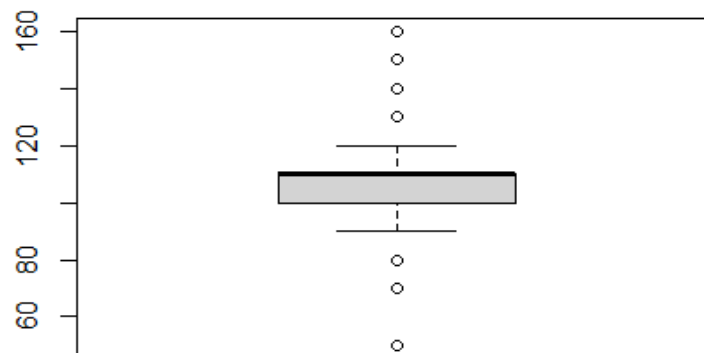
```
cereals_data <- read.csv("C:/Users/hp/Desktop/R- document/project2/cereals_data.csv")
```

```
library(psych)
```

```
describe(cereals_data$calories) # Non normal, Leptokurtic
```

```
##      vars  n  mean    sd median trimmed  mad min max range  skew kurtosis   se
## X1      1 77 106.88 19.48   110   107.3 14.83  50 160   110 -0.43     2.01
##      2.22
```

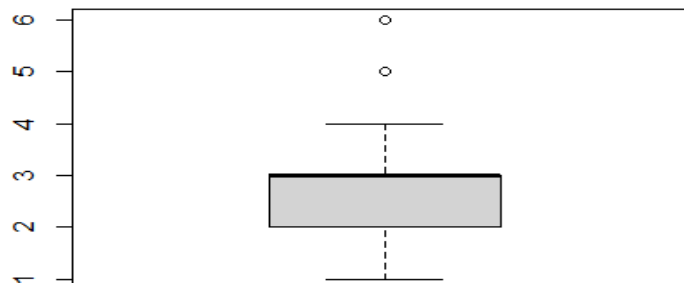
```
boxplot(cereals_data$calories)# 7 Outlier exists
```



```
describe(cereals_data$protein) # Normal, Leptokurtic
```

```
##      vars  n mean    sd median trimmed  mad min max range skew kurtosis   se
## X1      1 77 2.55 1.09     3    2.48 1.48   1  6    5 0.72     0.93 0.12
```

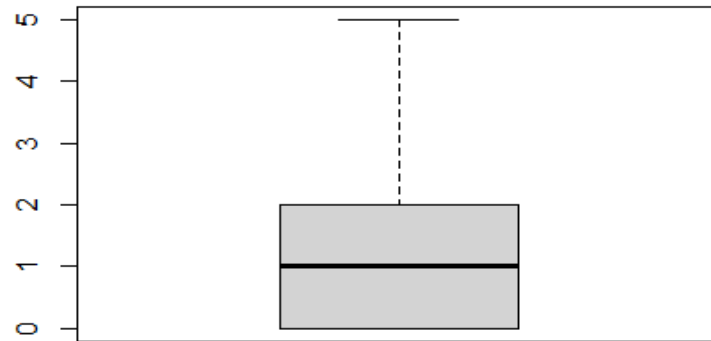
```
boxplot(cereals_data$protein) # 2 outlier exists
```



```
describe(cereals_data$fat) # Normal, Leptokurtic
```

```
##      vars  n mean   sd median trimmed  mad min max range skew kurtosis   se  
## X1      1 77 1.01 1.01      1    0.89 1.48   0  5    5 1.12     1.71 0.11
```

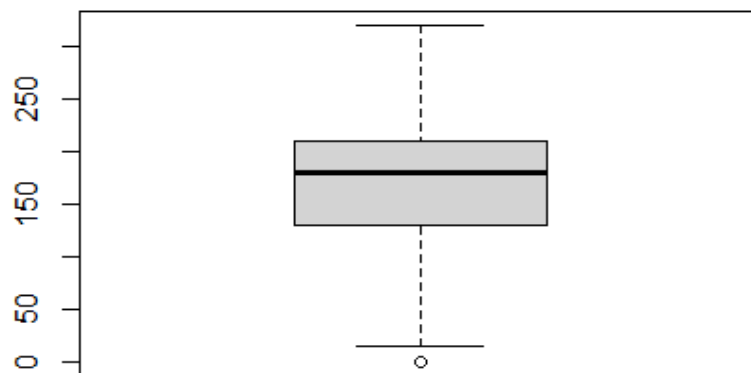
```
boxplot(cereals_data$fat)# 0 outlier exists
```



```
describe(cereals_data$sodium) # Normal, Platykurtic
```

```
##      vars  n  mean   sd median trimmed  mad min max range  skew kurtosis  
se  
## X1      1 77 159.68 83.83    180   163.25 59.3   0 320   320 -0.55    -0.47 9  
.55
```

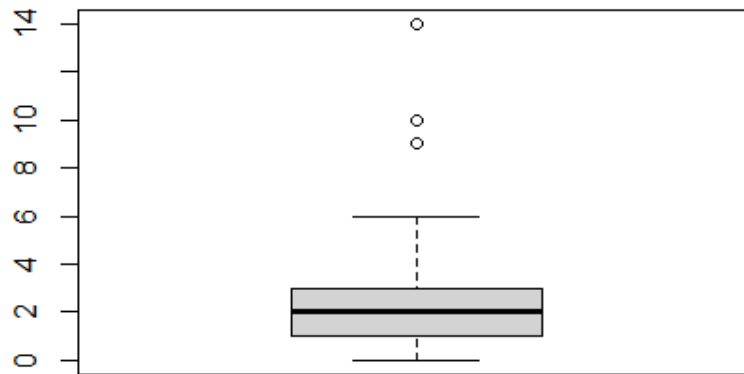
```
boxplot(cereals_data$sodium)# 1 outlier exists
```



```
describe(cereals_data$fiber) # Non normal, Leptokurtic
```

```
##      vars  n mean   sd median trimmed  mad min max range skew kurtosis   se  
## X1      1 77 2.15 2.38      2    1.77 1.48   0  14    14 2.34     7.73 0.27
```

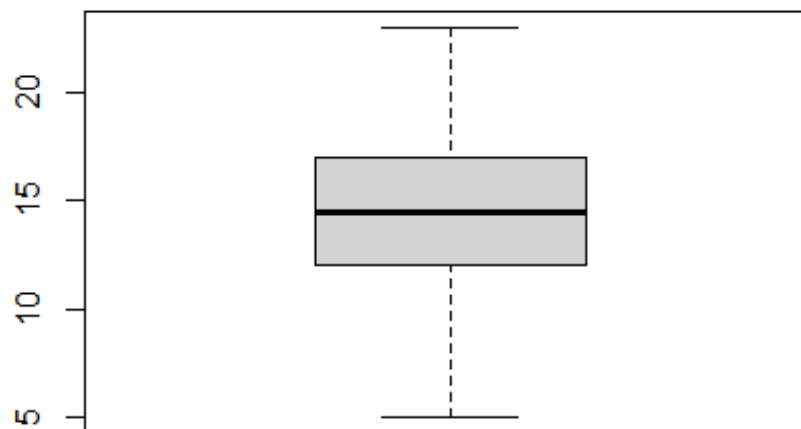
```
boxplot(cereals_data$fiber)# 3 outlier exists
```



```
describe(cereals_data$carbo) # Normal, Platykurtic
```

```
##      vars  n mean   sd median trimmed  mad min max range skew kurtosis   se  
## X1      1 76 14.8 3.91    14.5   14.79 3.71   5  23    18 0.11     -0.46 0.45
```

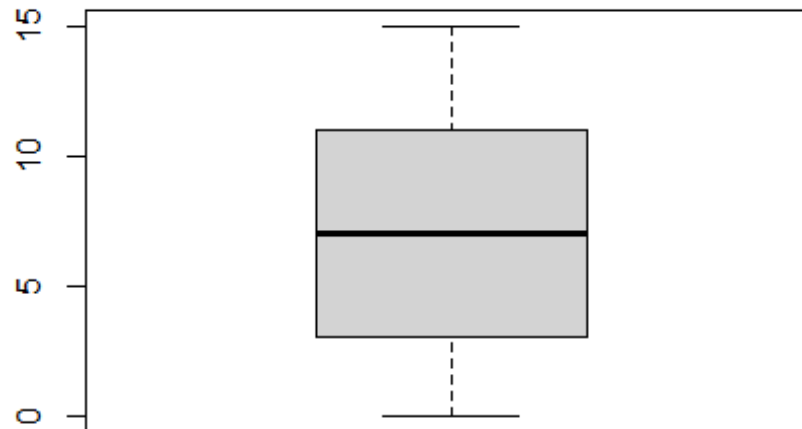
```
boxplot(cereals_data$carbo)# 0 outlier exists
```



```
describe(cereals_data$sugars) # Normal, Platykurtic
```

```
##      vars  n mean   sd median trimmed  mad min max range skew kurtosis  se
## X1      1 76 7.03 4.38      7    7.03 5.93   0 15   15 0.04    -1.2 0.5
```

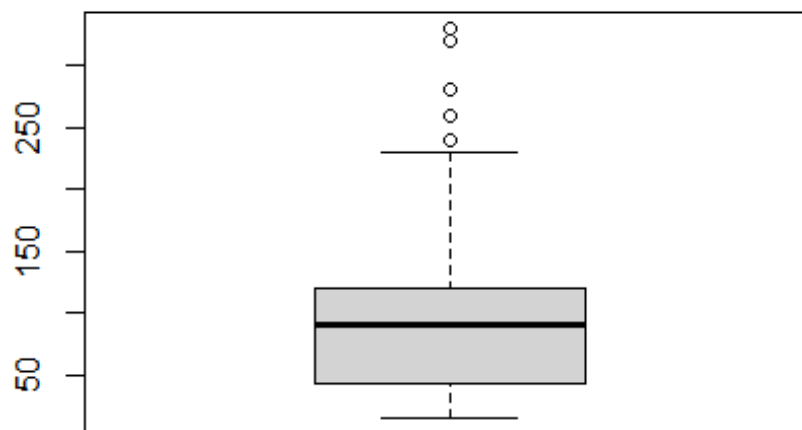
```
boxplot(cereals_data$sugars)# 0 outlier exists
```



```
describe(cereals_data$potass) # Normal, Leptokurtic
```

```
##      vars  n mean   sd median trimmed  mad min max range skew kurtosis  se
## X1      1 75 98.67 70.41     90    88.11 66.72  15 330   315 1.34    1.63 8.13
```

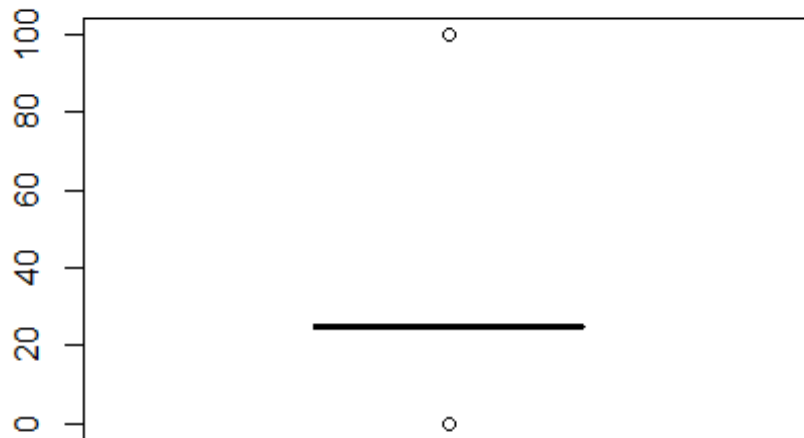
```
boxplot(cereals_data$potass)# 5 outlier exists
```



```
describe(cereals_data$vitamins) # Non normal, Leptokurtic
```

```
##      vars  n mean    sd median trimmed mad min max range skew kurtosis   se  
## X1      1 77 28.25 22.34     25    24.6   0   0 100   100 2.37     5.55 2.55
```

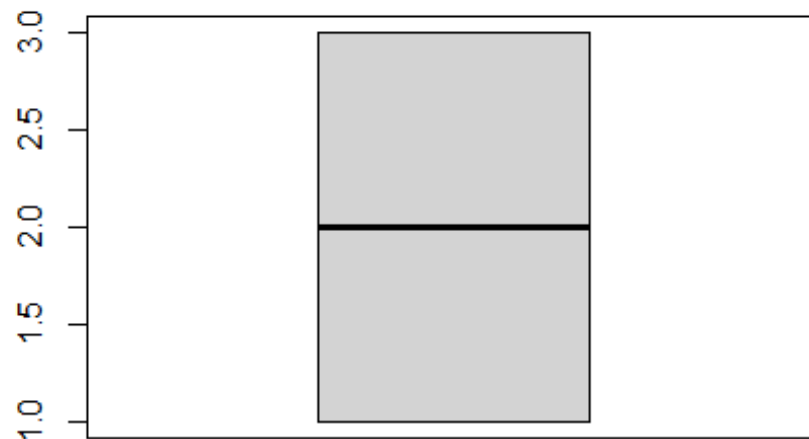
```
boxplot(cereals_data$vitamins)# 2 outlier exists
```



```
describe(cereals_data$shelf)
```

```
##      vars  n mean    sd median trimmed mad min max range skew kurtosis   se  
## X1      1 77 2.21 0.83      2    2.25 1.48   1   3     2 -0.39    -1.47 0.09
```

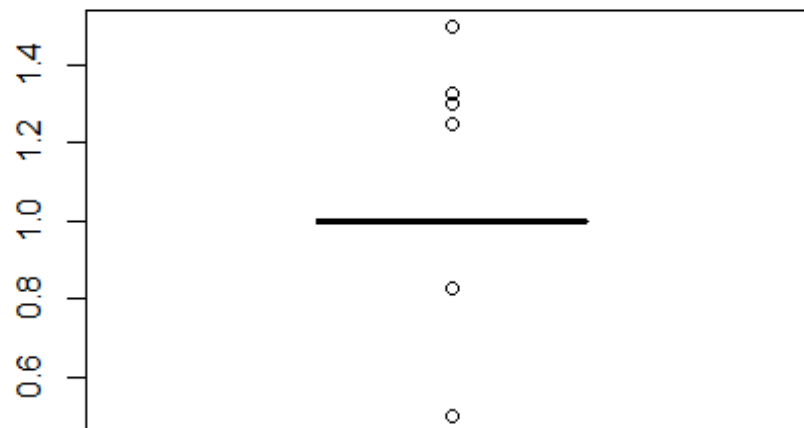
```
boxplot(cereals_data$shelf)
```




```
describe(cereals_data$weight)
```

```
##      vars  n mean   sd median trimmed mad min max range skew kurtosis   se  
## X1      1 77 1.03 0.15      1    1.01   0 0.5 1.5     1  0.3     4.79 0.02
```

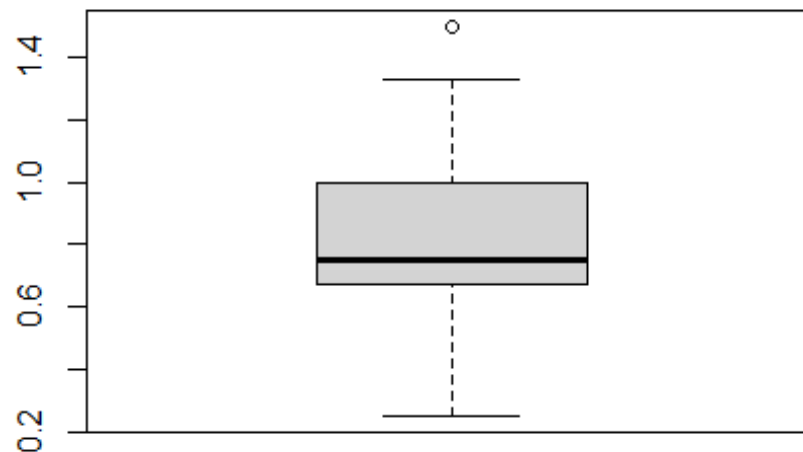
```
boxplot(cereals_data$weight)
```



```
describe(cereals_data$cups)
```

```
##      vars  n mean   sd median trimmed mad min max range skew kurtosis   se  
## X1      1 77 0.82 0.23   0.75    0.83 0.37 0.25 1.5    1.25 -0.1     0.17 0.03
```

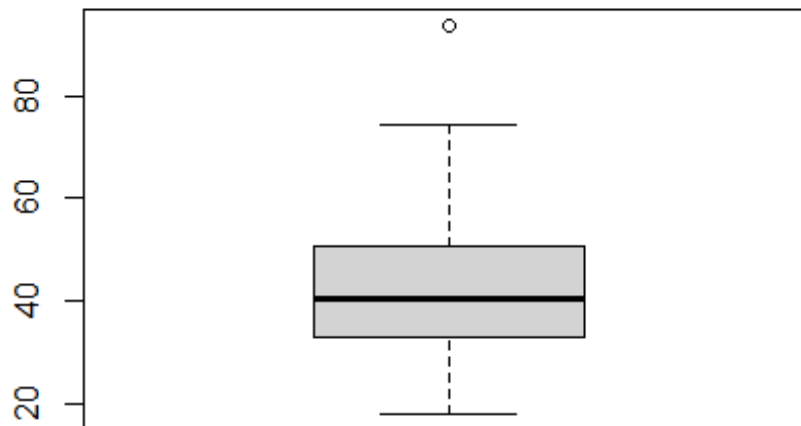
```
boxplot(cereals_data$cups)
```



```
describe(cereals_data$rating) # Normal, Leptokurtic
```

```
##      vars  n mean    sd median trimmed  mad   min  max range skew kurtosis  
se  
## X1      1 77 42.67 14.05   40.4   41.66 13.6 18.04 93.7 75.66 0.88     1.05  
1.6
```

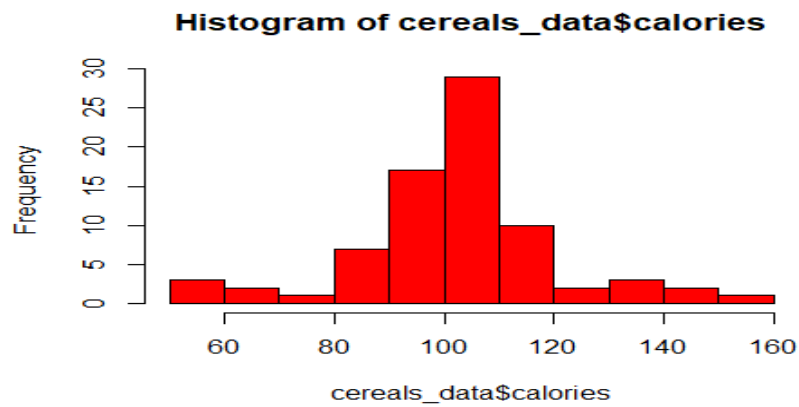
```
boxplot(cereals_data$rating)# 1 outlier exists
```



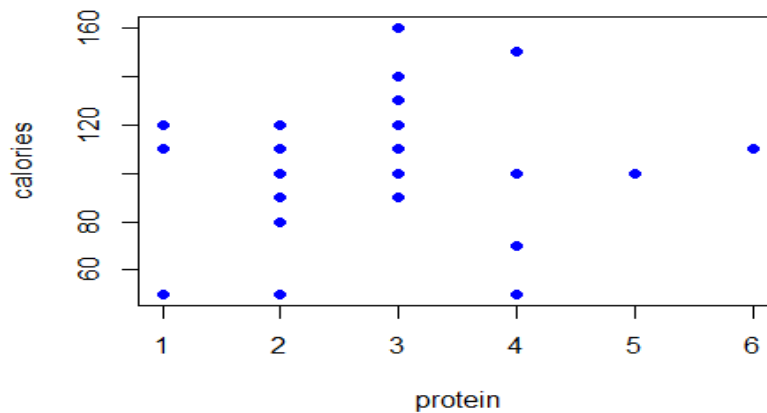
R markdown- TASK3

(Data Visualisation)

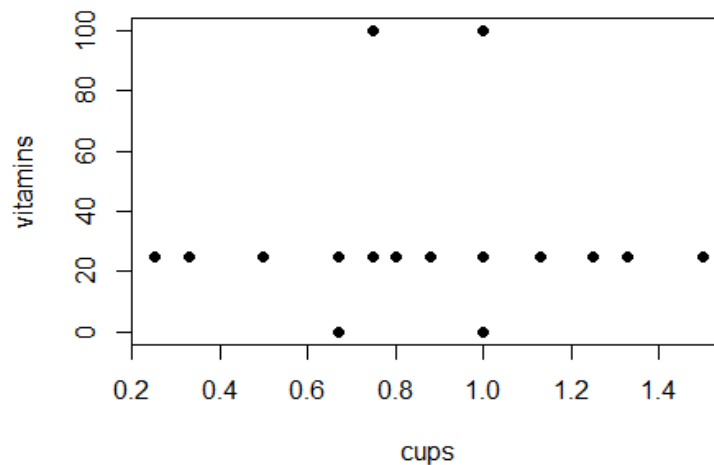
```
cereals_data <- read.csv("C:/Users/hp/Desktop/R- document/project2/cereals_data.csv")  
hist(cereals_data$calories, col = "red")
```



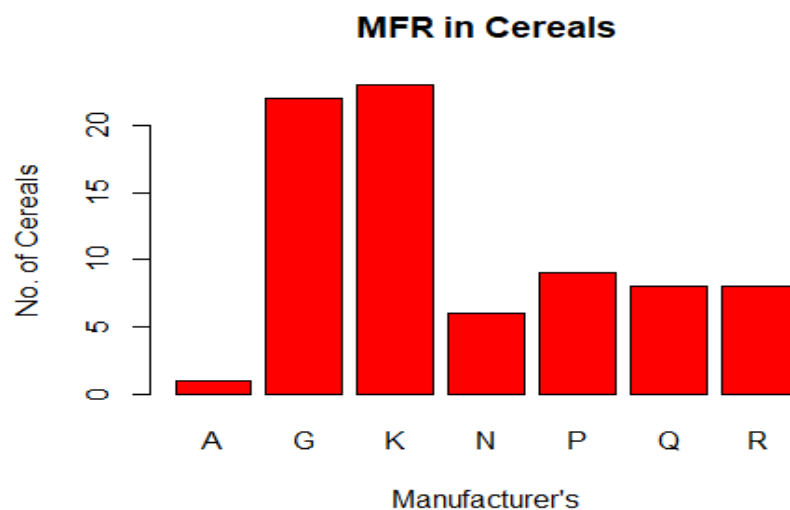
```
#PLOT BETWEEN CALORIES AND PROTIEN  
plot(calories ~ protein, data = cereals_data, pch=16, col = "Blue")
```



```
#PLOT BETWEEN VITAMINS AND cups SERVED
plot(vitamins ~ cups, data = cereals_data, pch=16)
```



```
#PLOT FOR NO. OF CEREALS PER MANUFACTURER
Mfr_counts = table(cereals_data$mfr)
mfr_viz= barplot(Mfr_counts, main = "MFR in Cereals",
                 xlab = "Manufacturer's", ylab = "No. of Cereals", col = "Red
")
```



```
library(ggplot2)

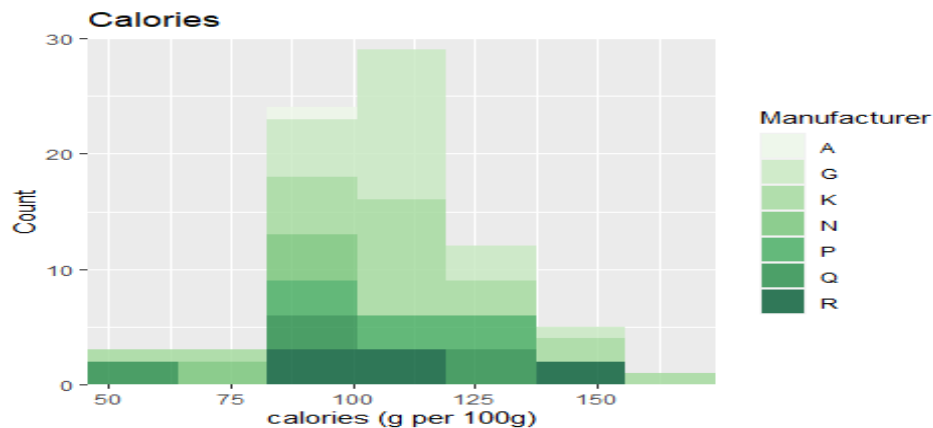
#Calories with count by mfr
ggplot(data = cereals_data,
```

```

aes(x=calories, fill = mfr)) +
geom_histogram(alpha = 0.8, bins = 7) +
scale_x_continuous(name = "calories (g per 100g)", expand = c(0,0)) +
scale_y_continuous(name = "Count", expand = c(0,0),
                    limits = c(0, 30), breaks = seq(0, 30, 10)) +
scale_fill_brewer(palette = "set2") +
labs(fill = "Manufacturer", title = "Calories")

```

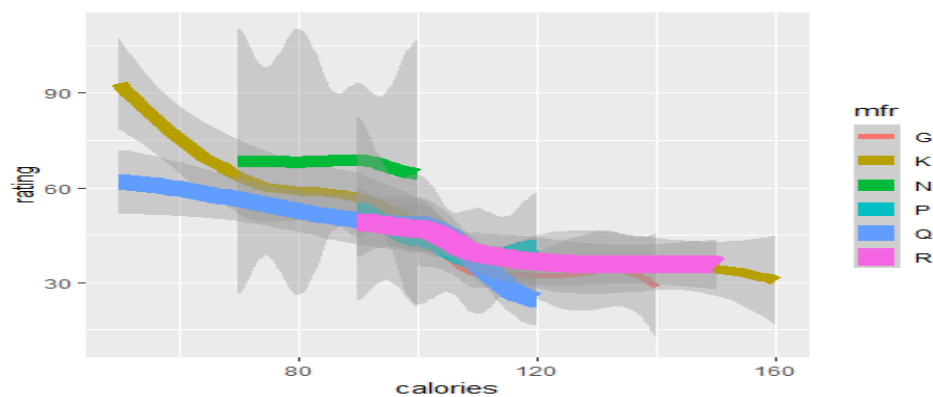
Warning in pal_name(palette, type): Unknown palette set2



```

#Ratings for calories with mfr
ggplot(data = cereals_data) +
  geom_smooth(mapping = aes(x= calories,
                           y= rating,
                           xlab = "Calories",
                           ylab = "Rating",
                           color = mfr, size = mfr,
                           Title = "Ratings for Calories with Manufacturer"))
)

```



```

#Corrplot
a = cereals_data[c(4:11,16)]
cereals_data

```

```

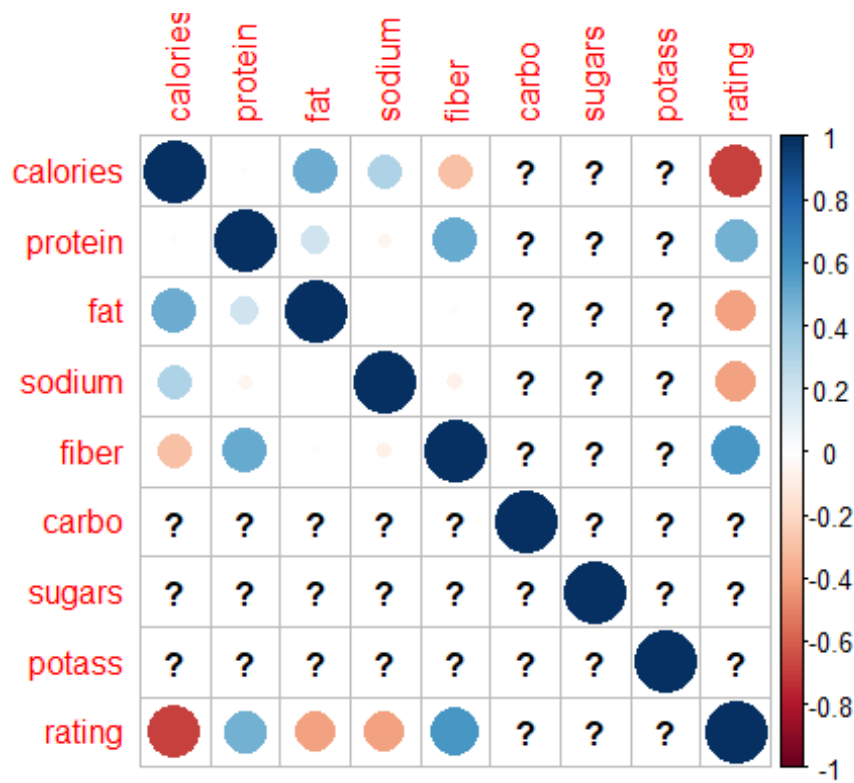
a = cor(a)
library(corrplot)

## Warning: package 'corrplot' was built under R version 4.0.2

## corrplot 0.84 loaded

corrplot(a)

```



R markdown- TASK4

(SLICING AND DICING FOR DEEPER ANALYSIS)

```
library(psych)
describe(cereals)

View(cereals)
summary(cereals)
```

##	name	mfr	type	calories
##	Length:77	Length:77	Length:77	Min. : 50.0
##	Class :character	Class :character	Class :character	1st Qu.:100.0
##	Mode :character	Mode :character	Mode :character	Median :110.0
##				Mean :106.9
##				3rd Qu.:110.0
##				Max. :160.0
##	protein	fat	sodium	fiber
##	Min. :1.000	Min. :0.000	Min. : 0.0	Min. : 0.000
##	1st Qu.:2.000	1st Qu.:0.000	1st Qu.:130.0	1st Qu.: 1.000
##	Median :3.000	Median :1.000	Median :180.0	Median : 2.000
##	Mean :2.545	Mean :1.013	Mean :159.7	Mean : 2.152
##	3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.:210.0	3rd Qu.: 3.000
##	Max. :6.000	Max. :5.000	Max. :320.0	Max. :14.000
##	carbo	sugars	potass	vitamins
##	Min. : 5.0	Min. : 0.000	Min. : 15.00	Min. : 0.00
##	1st Qu.:12.0	1st Qu.: 3.000	1st Qu.: 45.00	1st Qu.: 25.00
##	Median :14.8	Median : 7.000	Median : 90.00	Median : 25.00
##	Mean :14.8	Mean : 7.026	Mean : 98.44	Mean : 28.25
##	3rd Qu.:17.0	3rd Qu.:11.000	3rd Qu.:120.00	3rd Qu.: 25.00
##	Max. :23.0	Max. :15.000	Max. :330.00	Max. :100.00
##	shelf	weight	cups	rating
##	Min. :1.000	Min. :0.50	Min. :0.250	Min. :18.04
##	1st Qu.:1.000	1st Qu.:1.00	1st Qu.:0.670	1st Qu.:33.17
##	Median :2.000	Median :1.00	Median :0.750	Median :40.40
##	Mean :2.208	Mean :1.03	Mean :0.821	Mean :42.67
##	3rd Qu.:3.000	3rd Qu.:1.00	3rd Qu.:1.000	3rd Qu.:50.83
##	Max. :3.000	Max. :1.50	Max. :1.500	Max. :93.70

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
## filter, lag
```

```
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union
```

summarising the data for the manufactrer AH - American Home Food Products

```
cereal_AH<- subset(cereals, mfr == 'A')
View(cereal_AH)
summary(cereal_AH)
```

```
##      name           mfr           type           calories
## Length:1          Length:1          Length:1          Min.    :100
## Class :character   Class :character   Class :character   1st Qu.:100
## Mode  :character   Mode  :character   Mode  :character   Median :100
##                                     Mean   :100
##                                     3rd Qu.:100
##                                     Max.   :100
##      protein        fat          sodium        fiber        carbo        sugars
## Min.    :4         Min.    :1         Min.    :0         Min.    :0         Min.    :16         Min.    :3
## 1st Qu.:4         1st Qu.:1         1st Qu.:0         1st Qu.:0         1st Qu.:16         1st Qu.:3
## Median :4         Median :1         Median :0         Median :0         Median :16         Median :3
## Mean    :4         Mean    :1         Mean    :0         Mean    :0         Mean    :16         Mean    :3
## 3rd Qu.:4         3rd Qu.:1         3rd Qu.:0         3rd Qu.:0         3rd Qu.:16         3rd Qu.:3
## Max.    :4         Max.    :1         Max.    :0         Max.    :0         Max.    :16         Max.    :3
##      potass        vitamins        shelf        weight        cups        rating
## Min.    :95         Min.    :25         Min.    :2         Min.    :1         Min.    :1         Min.    :54.
85
## 1st Qu.:95         1st Qu.:25         1st Qu.:2         1st Qu.:1         1st Qu.:1         1st Qu.:54.
85
## Median :95         Median :25         Median :2         Median :1         Median :1         Median :54.
85
## Mean    :95         Mean    :25         Mean    :2         Mean    :1         Mean    :1         Mean    :54.
85
## 3rd Qu.:95         3rd Qu.:25         3rd Qu.:2         3rd Qu.:1         3rd Qu.:1         3rd Qu.:54.
85
## Max.    :95         Max.    :25         Max.    :2         Max.    :1         Max.    :1         Max.    :54.
85
```

summarising the data for the manufactrer GM -General Milks

```
cereal_GM<- subset(cereals, mfr == 'G')
View(cereal_GM)
summary(cereal_GM)
```

```
##      name           mfr           type           calories
## Length:22          Length:22          Length:22          Min.    :100.0
## Class :character   Class :character   Class :character   1st Qu.:110.0
```



```
## Mode :character Mode :character Mode :character Median :110.0
## Mean :111.4
## 3rd Qu.:110.0
## Max. :140.0
## protein fat sodium fiber
## Min. :1.000 Min. :1.000 Min. :140.0 Min. :0.000
## 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:180.0 1st Qu.:0.000
## Median :2.000 Median :1.000 Median :200.0 Median :1.500
## Mean :2.318 Mean :1.364 Mean :200.5 Mean :1.273
## 3rd Qu.:3.000 3rd Qu.:2.000 3rd Qu.:217.5 3rd Qu.:2.000
## Max. :6.000 Max. :3.000 Max. :290.0 Max. :4.000
## carbo sugars potass vitamins
## Min. :10.50 Min. : 1.000 Min. : 25.00 Min. : 25.00
## 1st Qu.:12.00 1st Qu.: 3.750 1st Qu.: 55.00 1st Qu.: 25.00
## Median :14.25 Median : 8.500 Median : 80.00 Median : 25.00
## Mean :14.73 Mean : 7.955 Mean : 85.23 Mean : 35.23
## 3rd Qu.:16.75 3rd Qu.:10.000 3rd Qu.:108.75 3rd Qu.: 25.00
## Max. :21.00 Max. :14.000 Max. :230.00 Max. :100.00
## shelf weight cups rating
## Min. :1.000 Min. :1.000 Min. :0.500 Min. :19.82
## 1st Qu.:1.250 1st Qu.:1.000 1st Qu.:0.750 1st Qu.:27.96
## Median :2.000 Median :1.000 Median :0.875 Median :36.18
## Mean :2.136 Mean :1.049 Mean :0.875 Mean :34.49
## 3rd Qu.:3.000 3rd Qu.:1.000 3rd Qu.:1.000 3rd Qu.:39.59
## Max. :3.000 Max. :1.500 Max. :1.500 Max. :51.59
```

summarising the data for the manufactrer KE - KelLoggs

```
cereal_KE<- subset(cereals, mfr == 'K')
```

```
View(cereal_KE)
```

```
summary(cereal_KE)
```

```
## name mfr type calories
## Length:23 Length:23 Length:23 Min. : 50.0
## Class :character Class :character Class :character 1st Qu.:100.0
## Mode :character Mode :character Mode :character Median :110.0
## Mean :108.7
## 3rd Qu.:115.0
## Max. :160.0
## protein fat sodium fiber
## Min. :1.000 Min. :0.0000 Min. : 0.0 Min. : 0.000
## 1st Qu.:2.000 1st Qu.:0.0000 1st Qu.:132.5 1st Qu.: 1.000
## Median :3.000 Median :0.0000 Median :170.0 Median : 1.000
## Mean :2.652 Mean :0.6087 Mean :174.8 Mean : 2.739
## 3rd Qu.:3.000 3rd Qu.:1.0000 3rd Qu.:225.0 3rd Qu.: 3.000
## Max. :6.000 Max. :3.0000 Max. :320.0 Max. :14.000
## carbo sugars potass vitamins
## Min. : 7.00 Min. : 0.000 Min. : 20 Min. : 25.00
## 1st Qu.:12.00 1st Qu.: 3.000 1st Qu.: 35 1st Qu.: 25.00
```

```
## Median :15.00 Median : 7.000 Median : 60 Median : 25.00
## Mean :15.13 Mean : 7.565 Mean :103 Mean : 34.78
## 3rd Qu.:19.00 3rd Qu.:12.000 3rd Qu.:145 3rd Qu.: 25.00
## Max. :22.00 Max. :15.000 Max. :330 Max. :100.00
## shelf weight cups rating
## Min. :1.000 Min. :1.000 Min. :0.3300 Min. :29.92
## 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:0.6700 1st Qu.:34.48
## Median :3.000 Median :1.000 Median :0.7500 Median :40.56
## Mean :2.348 Mean :1.078 Mean :0.7961 Mean :44.04
## 3rd Qu.:3.000 3rd Qu.:1.000 3rd Qu.:1.0000 3rd Qu.:50.01
## Max. :3.000 Max. :1.500 Max. :1.0000 Max. :93.70
```

summarising the data for the manufactrer N - Nabisdco

```
cereal_N<- subset(cereals, mfr == 'N')
```

```
View(cereal_N)
```

```
summary(cereal_N)
```

```
## name mfr type calories
## Length:6 Length:6 Length:6 Min. : 70.00
## Class :character Class :character Class :character 1st Qu.: 82.50
## Mode :character Mode :character Mode :character Median : 90.00
## Mean : 86.67
## 3rd Qu.: 90.00
## Max. :100.00
## protein fat sodium fiber
## Min. :2.000 Min. :0.0000 Min. : 0.00 Min. : 1.00
## 1st Qu.:2.250 1st Qu.:0.0000 1st Qu.: 0.00 1st Qu.: 3.00
## Median :3.000 Median :0.0000 Median : 7.50 Median : 3.00
## Mean :2.833 Mean :0.1667 Mean : 37.50 Mean : 4.00
## 3rd Qu.:3.000 3rd Qu.:0.0000 3rd Qu.: 63.75 3rd Qu.: 3.75
## Max. :4.000 Max. :1.0000 Max. :130.00 Max. :10.00
## carbo sugars potass vitamins
## Min. : 5.00 Min. :0.000 Min. : 90.00 Min. : 0.000
## 1st Qu.:15.25 1st Qu.:0.000 1st Qu.: 91.25 1st Qu.: 0.000
## Median :17.50 Median :0.000 Median :107.50 Median : 0.000
## Mean :16.00 Mean :1.833 Mean :135.83 Mean : 8.333
## 3rd Qu.:19.75 3rd Qu.:3.750 3rd Qu.:135.00 3rd Qu.:18.750
## Max. :21.00 Max. :6.000 Max. :280.00 Max. :25.000
## shelf weight cups rating
## Min. :1.000 Min. :0.8300 Min. :0.3300 Min. :59.36
## 1st Qu.:1.000 1st Qu.:1.0000 1st Qu.:0.6700 1st Qu.:65.46
## Median :1.500 Median :1.0000 Median :0.8350 Median :68.32
## Mean :1.667 Mean :0.9717 Mean :0.7783 Mean :67.97
## 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:71.70
## Max. :3.000 Max. :1.0000 Max. :1.0000 Max. :74.47
```

```
# summarising the data for the manufactrer P- Post
cereal_P<- subset(cereals, mfr == 'P')
summary(cereal_P)
```

```
##      name          mfr          type          calories
## Length:9      Length:9      Length:9      Min.   : 90.0
## Class :character Class :character Class :character 1st Qu.:100.0
## Mode  :character Mode  :character Mode  :character Median :110.0
##                                         Mean  :108.9
##                                         3rd Qu.:120.0
##                                         Max.   :120.0
##      protein      fat      sodium      fiber
## Min.   :1.000    Min.   :0.0000    Min.   : 45.0    Min.   :0.000
## 1st Qu.:2.000    1st Qu.:0.0000    1st Qu.:135.0    1st Qu.:0.000
## Median :3.000    Median :1.0000    Median :160.0    Median :3.000
## Mean   :2.444    Mean   :0.8889    Mean   :146.1    Mean   :2.778
## 3rd Qu.:3.000    3rd Qu.:1.0000    3rd Qu.:180.0    3rd Qu.:5.000
## Max.   :3.000    Max.   :3.0000    Max.   :210.0    Max.   :6.000
##      carbo      sugars      potass      vitamins      shelf
## Min.   :11.00    Min.   : 3.000    Min.   : 25.0    Min.   :25    Min.   :1.0
## 1st Qu.:12.00    1st Qu.: 5.000    1st Qu.: 40.0    1st Qu.:25    1st Qu.:2.0
## Median :13.00    Median :10.000    Median : 90.0    Median :25    Median :3.0
## Mean   :13.22    Mean   : 8.778    Mean   :113.9    Mean   :25    Mean   :2.4
## 3rd Qu.:14.00    3rd Qu.:12.000    3rd Qu.:190.0    3rd Qu.:25    3rd Qu.:3.0
## Max.   :17.00    Max.   :15.000    Max.   :260.0    Max.   :25    Max.   :3.0
##      weight      cups      rating
## Min.   :1.000    Min.   :0.2500    Min.   :28.03
## 1st Qu.:1.000    1st Qu.:0.6700    1st Qu.:35.25
## Median :1.000    Median :0.6700    Median :40.92
## Mean   :1.064    Mean   :0.7144    Mean   :41.71
## 3rd Qu.:1.000    3rd Qu.:0.8800    3rd Qu.:52.08
## Max.   :1.330    Max.   :1.3300    Max.   :53.37
```

```
View(cereal_P)
```

```
# summarising the data for the manufactrer QK -Quaker Oats
cereal_QK<- subset(cereals, mfr == 'Q')
summary(cereal_QK)
```

```
##      name          mfr          type          calories
## Length:8      Length:8      Length:8      Min.   : 50.0
## Class :character Class :character Class :character 1st Qu.: 87.5
## Mode  :character Mode  :character Mode  :character Median :100.0
```

```
##                                     Mean   : 95.0
##                                     3rd Qu.:120.0
##                                     Max.   :120.0
##      protein      fat      sodium      fiber      carbo
## Min.    :1.000    Min.    :0.00    Min.    : 0.0    Min.    :0.000    Min.    : 8
## 1st Qu.:1.000    1st Qu.:0.75    1st Qu.: 0.0    1st Qu.:0.750    1st Qu.:11
## Median :2.500    Median :2.00    Median : 75.0    Median :1.500    Median :12
## Mean   :2.625    Mean   :1.75    Mean   : 92.5    Mean   :1.337    Mean   :11
## 3rd Qu.:4.000    3rd Qu.:2.00    3rd Qu.:167.5    3rd Qu.:2.000    3rd Qu.:13
## Max.   :5.000    Max.   :5.00    Max.   :220.0    Max.   :2.700    Max.   :14
##      sugars      potass      vitamins      shelf
## Min.    : 0.00    Min.    : 15.00    Min.    : 0.0    Min.    :1.000
## 1st Qu.: 4.50    1st Qu.: 42.50    1st Qu.: 0.0    1st Qu.:2.000
## Median : 6.50    Median : 72.50    Median :12.5    Median :2.500
## Mean   : 6.25    Mean   : 74.38    Mean   :12.5    Mean   :2.375
## 3rd Qu.: 8.75    3rd Qu.:110.00    3rd Qu.:25.0    3rd Qu.:3.000
## Max.   :12.00    Max.   :135.00    Max.   :25.0    Max.   :3.000
##      weight      cups      rating
## Min.    :0.500    Min.    :0.5000    Min.    :18.04
## 1st Qu.:0.875    1st Qu.:0.6700    1st Qu.:30.96
## Median :1.000    Median :0.8750    Median :47.42
## Mean   :0.875    Mean   :0.8237    Mean   :42.92
## 3rd Qu.:1.000    3rd Qu.:1.0000    3rd Qu.:53.31
## Max.   :1.000    Max.   :1.0000    Max.   :63.01
```

[View\(cereal_QK\)](#)

summarising the data for the manufactrer R- Ralston Purina

```
cereal_R<- subset(cereals, mfr == 'R')
```

[View\(cereal_R\)](#)

[summary\(cereal_R\)](#)

```
##      name      mfr      type      calories
## Length:8      Length:8      Length:8      Min.    : 90
## Class :character Class :character Class :character 1st Qu.:100
## Mode  :character Mode  :character Mode  :character Median :110
##                                     Mean   :115
##                                     3rd Qu.:120
##                                     Max.   :150
##      protein      fat      sodium      fiber      carbo
## Min.    :1.00    Min.    :0.00    Min.    : 95.0    Min.    :0.000    Min.    :14.
## 00
```

```

## 1st Qu.:2.00 1st Qu.:0.00 1st Qu.:180.0 1st Qu.:0.750 1st Qu.:15.
75
## Median :2.00 Median :1.00 Median :200.0 Median :2.000 Median :16.
50
## Mean :2.50 Mean :1.25 Mean :198.1 Mean :1.875 Mean :17.
62
## 3rd Qu.:3.25 3rd Qu.:2.25 3rd Qu.:232.5 3rd Qu.:3.000 3rd Qu.:19.
00
## Max. :4.00 Max. :3.00 Max. :280.0 Max. :4.000 Max. :23.
00
## sugars potass vitamins shelf weight
## Min. : 2.000 Min. : 25.0 Min. :25 Min. :1 Min. :1
## 1st Qu.: 3.000 1st Qu.: 67.5 1st Qu.:25 1st Qu.:1 1st Qu.:1
## Median : 5.500 Median :102.5 Median :25 Median :2 Median :1
## Mean : 6.125 Mean :100.6 Mean :25 Mean :2 Mean :1
## 3rd Qu.: 8.750 3rd Qu.:136.2 3rd Qu.:25 3rd Qu.:3 3rd Qu.:1
## Max. :11.000 Max. :170.0 Max. :25 Max. :3 Max. :1
## cups rating
## Min. :0.6700 Min. :34.14
## 1st Qu.:0.7300 1st Qu.:36.45
## Median :0.8750 Median :41.72
## Mean :0.8712 Mean :41.54
## 3rd Qu.:1.0000 3rd Qu.:45.53
## Max. :1.1300 Max. :49.79

```

R markdown- TASK5

(CLUSTER ANALYSIS)

```
cereals_data <- read.csv("C:/Users/hp/Desktop/R- document/project2/cereals_data.csv")

library(colorspace)
library(grid)
library(VIM)

## Warning: package 'VIM' was built under R version 4.0.2

## VIM is ready to use.

## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues

##
## Attaching package: 'VIM'

## The following object is masked from 'package:datasets':
##
##      sleep

cereals_data<- kNN(cereals_data)
cereals<- cereals_data[1:16]

# choosing variables for clustering
View(cereals)
str(cereals)

## 'data.frame':   77 obs. of  16 variables:
## $ name      : chr  "100%_Bran" "100%_Natural_Bran" "All-Bran" "All-Bran_wit
h_Extra_Fiber" ...
## $ mfr       : chr  "N" "Q" "K" "K" ...
## $ type      : chr  "C" "C" "C" "C" ...
## $ calories: int   70 120 70 50 110 110 110 130 90 90 ...
## $ protein  : int   4 3 4 4 2 2 2 3 2 3 ...
## $ fat       : int   1 5 1 0 2 2 0 2 1 0 ...
## $ sodium   : int  130 15 260 140 200 180 125 210 200 210 ...
## $ fiber    : num   10 2 9 14 1 1.5 1 2 4 5 ...
## $ carbo    : num    5 8 7 8 14 10.5 11 18 15 13 ...
## $ sugars   : int    6 8 5 0 8 10 14 8 6 5 ...
## $ potass   : int  280 135 320 330 120 70 30 100 125 190 ...
## $ vitamins: int   25 0 25 25 25 25 25 25 25 ...
## $ shelf    : int    3 3 3 3 3 1 2 3 1 3 ...
## $ weight   : num    1 1 1 1 1 1 1 1.33 1 1 ...
## $ cups     : num   0.33 1 0.33 0.5 0.75 0.75 1 0.75 0.67 0.67 ...
## $ rating   : num  68.4 34 59.4 93.7 34.4 ...
```

```

df<- cereals[ ,c(4,5,9,10)]

View(df)
dim(df)

## [1] 77 4

#scaling data

df<- na.omit(df)
df<- scale(df)

str(df)

## num [1:77, 1:4] -1.893 0.673 -1.893 -2.919 0.16 ...
## - attr(*, "dimnames")=List of 2
## ..$ : chr [1:77] "1" "2" "3" "4" ...
## ..$ : chr [1:4] "calories" "protein" "carbo" "sugars"
## - attr(*, "scaled:center")= Named num [1:4] 106.88 2.55 14.79 6.97
## ..- attr(*, "names")= chr [1:4] "calories" "protein" "carbo" "sugars"
## - attr(*, "scaled:scale")= Named num [1:4] 19.48 1.09 3.88 4.37
## ..- attr(*, "names")= chr [1:4] "calories" "protein" "carbo" "sugars"

View(df)
head(df,3)

##      calories  protein    carbo    sugars
## 1 -1.8929836 1.3286071 -2.522067 -0.2226911
## 2  0.6732089 0.4151897 -1.749391  0.2345680
## 3 -1.8929836 1.3286071 -2.006950 -0.4513207

library(cluster)

## Warning: package 'cluster' was built under R version 4.0.2

library(ggplot2)
library(factoextra)

## Warning: package 'factoextra' was built under R version 4.0.2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

set.seed(123)
ss<-sample(1:50, 15)
head(df,3)

##      calories  protein    carbo    sugars
## 1 -1.8929836 1.3286071 -2.522067 -0.2226911
## 2  0.6732089 0.4151897 -1.749391  0.2345680
## 3 -1.8929836 1.3286071 -2.006950 -0.4513207

```

```

df.scaled<- scale(df)
head(df.scaled, 3)

##      calories  protein    carbo    sugars
## 1 -1.8929836  1.3286071 -2.522067 -0.2226911
## 2  0.6732089  0.4151897 -1.749391  0.2345680
## 3 -1.8929836  1.3286071 -2.006950 -0.4513207

#euclidean distance

dist.eucl_15<- dist(df.scaled, method='euclidean')
head(dist.eucl_15, 3)

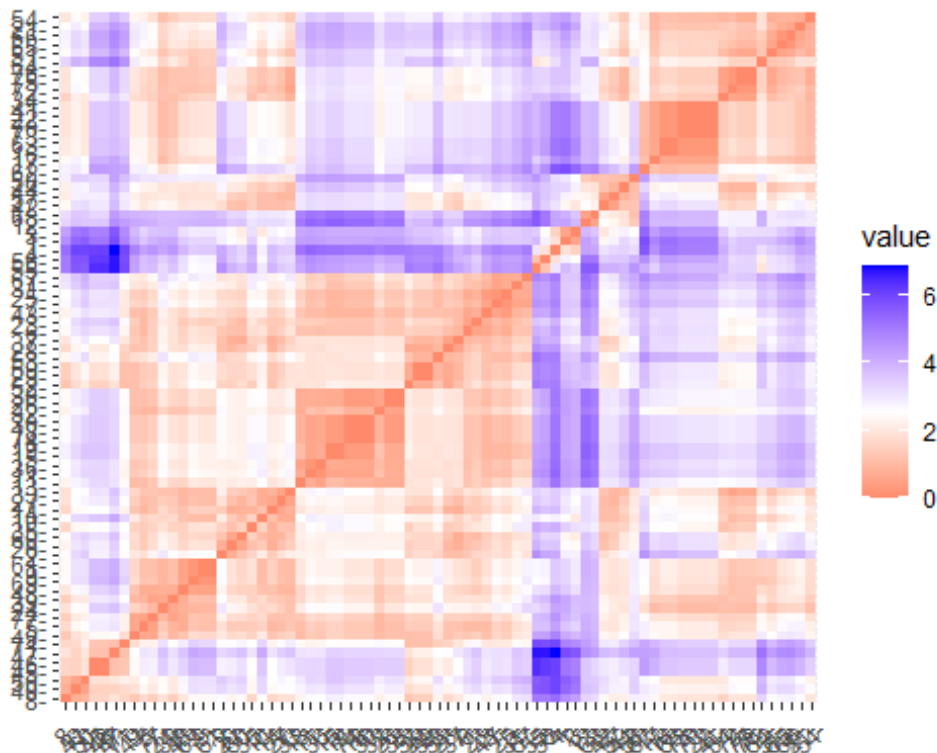
## [1] 2.8680636 0.5635753 1.8794828

round(as.matrix(dist.eucl_15)[1:3 , 1:3], 1)

##      1    2    3
## 1 0.0  2.9  0.6
## 2 2.9  0.0  2.8
## 3 0.6  2.8  0.0

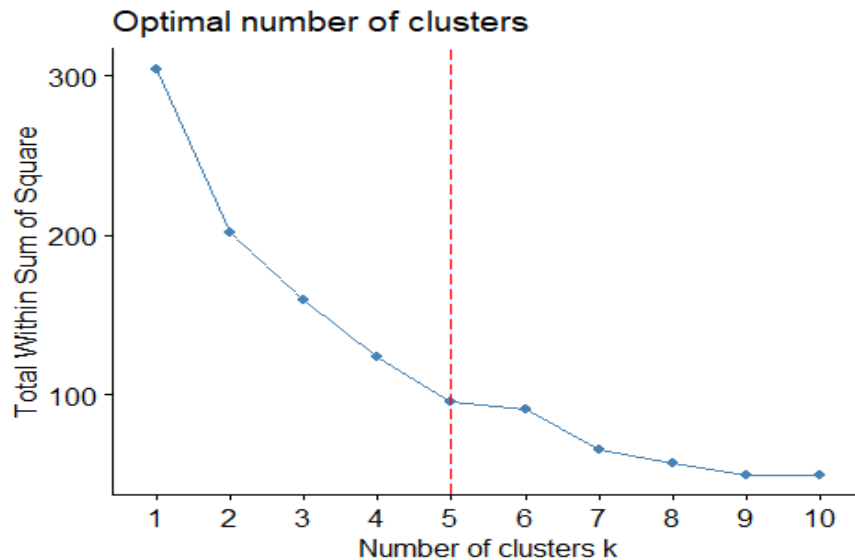
fviz_dist(dist.eucl_15)

```



#no of optimum clusters

```
fviz_nbclust(df, kmeans, method='wss')+  
  geom_vline(xintercept=5, linetype=5, col='red')
```



#K means clustering

```
set.seed(123)  
km.res<- kmeans(df,5, nstart=25)  
km.res  
  
## K-means clustering with 5 clusters of sizes 8, 5, 27, 24, 13  
##  
## Cluster means:  
##      calories  protein    carbo    sugars  
## 1  1.8279956  0.6435441  0.5847332  0.7775631  
## 2 -2.5088697  0.4151897 -1.5948562 -1.0914834  
## 3  0.2930323 -0.6673790 -0.6237650  0.9881987  
## 4 -0.3318831 -0.1937552  0.9979001 -0.8990535  
## 5 -0.1558686  1.1880814 -0.2931948 -0.4513207  
##  
## Clustering vector:  
## 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25  
## 26  
## 2  3  2  2  3  3  3  1  4  5  3  5  3  5  3  4  4  3  3  5  4  4  3  4  3  
## 3  
## 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51  
## 52  
## 5  3  3  3  3  3  5  4  5  3  3  3  4  1  4  5  3  5  1  1  1  4  3  1  4  
## 1  
## 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77  
## 3  4  2  2  5  5  3  5  4  4  4  4  4  4  3  5  4  4  1  4  4  3  4  4  3  
##
```

```

## Within cluster sum of squares by cluster:
## [1] 8.244739 12.330342 27.318865 25.885600 21.103942
## (between_SS / total_SS = 68.8 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withi
nss"
## [6] "betweenss"    "size"         "iter"         "ifault"

km.res$totss

## [1] 304

km.res$betweenss

## [1] 209.1165

aggregate(df, by = list(cluster=km.res$cluster), mean)

##   cluster  calories  protein  carbo  sugars
## 1      1  1.8279956  0.6435441  0.5847332  0.7775631
## 2      2 -2.5088697  0.4151897 -1.5948562 -1.0914834
## 3      3  0.2930323 -0.6673790 -0.6237650  0.9881987
## 4      4 -0.3318831 -0.1937552  0.9979001 -0.8990535
## 5      5 -0.1558686  1.1880814 -0.2931948 -0.4513207

#add cluster membership in data file

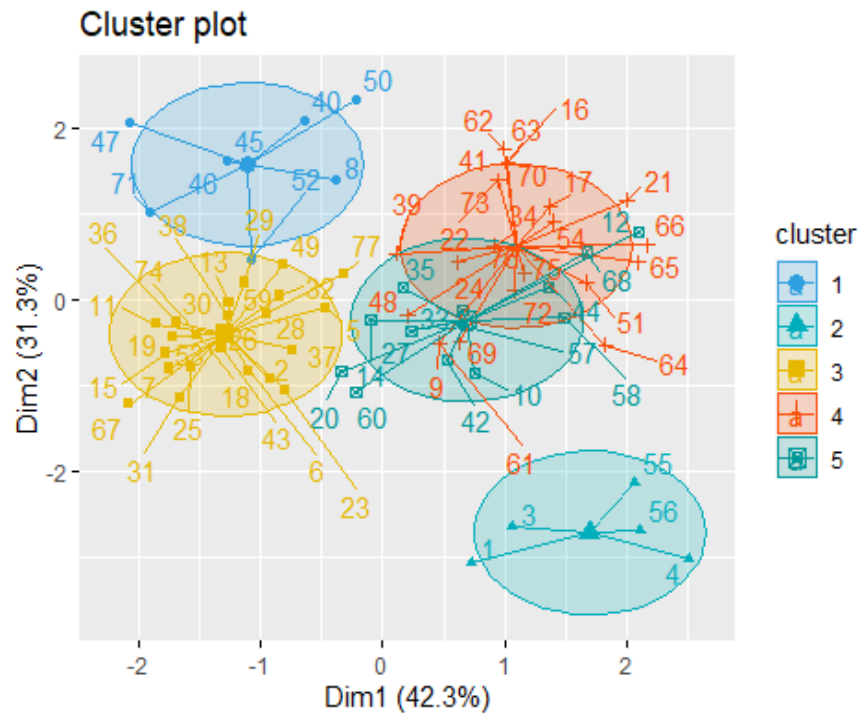
df_m<- cbind(df, cluster= km.res$cluster)
head(df_m)

##   calories  protein  carbo  sugars cluster
## 1 -1.8929836  1.3286071 -2.5220671 -0.2226911      2
## 2  0.6732089  0.4151897 -1.7493914  0.2345680      3
## 3 -1.8929836  1.3286071 -2.0069499 -0.4513207      2
## 4 -2.9194605  1.3286071 -1.7493914 -1.5944684      2
## 5  0.1599704 -0.4982277 -0.2040399  0.2345680      3
## 6  0.1599704 -0.4982277 -1.1054949  0.6918271      3

View(df_m)

fviz_cluster(km.res, data=df,
              palette= c('#2E9FDF', '#00AFBB', '#E7B800', '#FC4E07', '#009999'),
              ellipse.type='euclid',
              star.plot=TRUE,
              repel=TRUE,
              ggtheme=theme())

```



```
#dissimilarity matrix
```

```
res.dist<- dist(df, method='euclidean')
head(res.dist)
```

```
## [1] 2.8680636 0.5635753 1.8794828 3.6241249 3.2241225 3.6449185
```

```
round(as.matrix(res.dist)[1:3 , 1:3], 1)
```

```
##      1      2      3
```

```
## 1 0.0 2.9 0.6
```

```
## 2 2.9 0.0 2.8
```

```
## 3 0.6 2.8 0.0
```

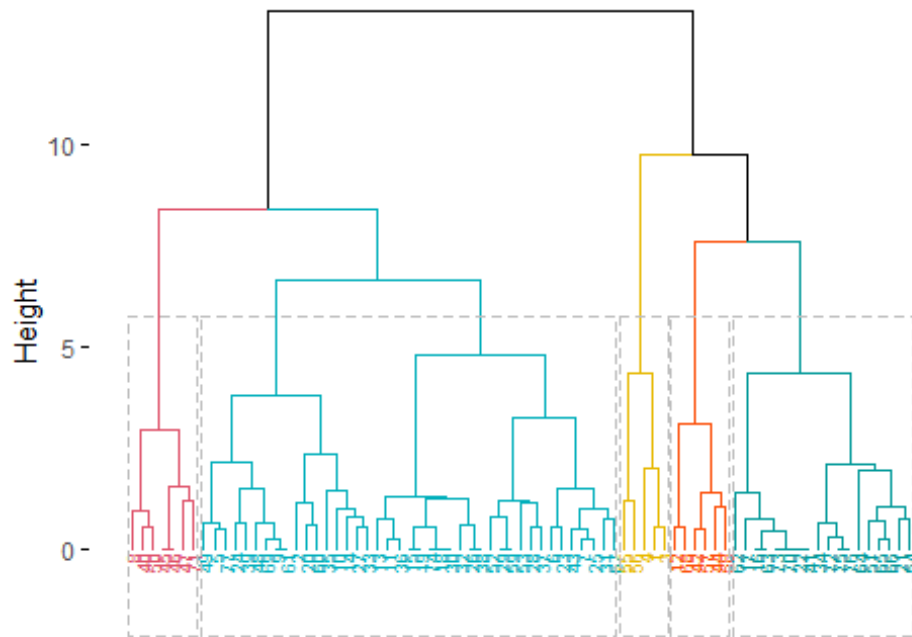
```
fviz_dist(res.dist)
```



```
#colorful dendrogram
```

```
fviz_dend(res.hc, k=5,  
  cex=0.5,  
  k_colors = c('2E9FDF', '#00AFBB', '#E7B800', '#FC4E07', '#009999'),  
  color_labels_by_k= TRUE,  
  rect=TRUE)
```

Cluster Dendrogram



```
#more beautiful diagram
```

```
#cut tree
```

```
grp<- cutree(res.hc, k=5)  
head(grp, n=4)
```

```
## 1 2 3 4
```

```
## 1 2 1 1
```

```
table(grp)
```

```
## grp
```

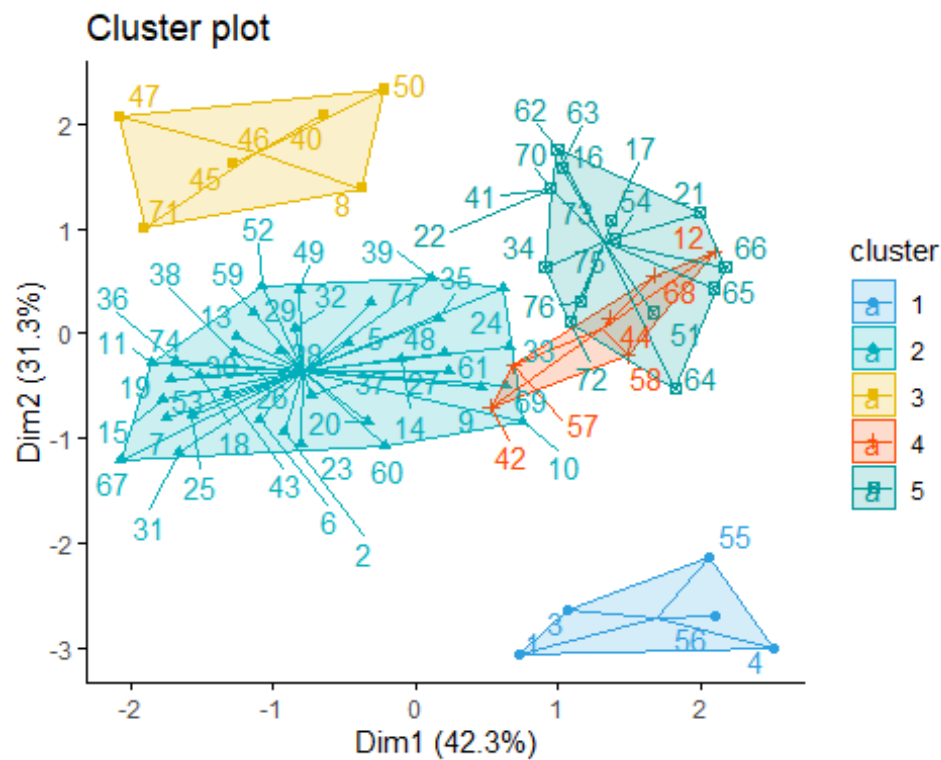
```
## 1 2 3 4 5
```

```
## 5 41 7 6 18
```

```
rownames(df)[grp==1]
```

```
## [1] "1" "3" "4" "55" "56"
```

```
fviz_cluster(list(data=df, cluster= grp),
              palette= c('#2E9FDF', '#00AFBB', '#E7B800', '#FC4E07', '#009999'),
              ellipse= TRUE,
              ellipse.type='convex',
              star.plot=TRUE,
              repel=TRUE,
              show.clust.cent=FALSE,
              ggtheme=theme_classic())
```



R markdown- TASK6

(WORD CLOUD)

```
library(NLP)
library(psych)
library(RColorBrewer)
library(wordcloud)
library(wordcloud2)
library('tm')

# Set the working directory, file path and read the information
setwd("C:/Users/HP/Desktop/cereal project")
filepath = "C:/Users/HP/Desktop/cereal project/word.txt"
text_file<-readLines(filepath)
head(text_file)

## [1] "Cereals have always held their share of the Indian breakfast platter
e.g. flattened rice flakes (chivda/poha) with"
## [2] "milk popular in western and central India, whole wheat grits (dalia)
in northern India, etc., alongside the regional"
## [3] "staples which have been traditionally preferred, e.g. paranthas in th
e north, and idli or dosa in the south. For the"
## [4] "most part, Indians have inclined towards consuming their quintessenti
al hot, cooked breakfast (naashta), and find"
## [5] "it difficult to give in to such breakfast options as cereals with mil
k. However, lifestyles are changing in tandem"
## [6] "with increasing spending power, greater time-poverty, higher need for
convenience, and health consciousness."

tail(text_file)

## [1] "Breakfast cereal manufacturers are likely to benefit from this trend
as health benefits are one of the factors due to"
## [2] "which consumers purchase cereals. The surge in demand will also drive
innovation in the breakfast cereal"
## [3] "market, with the introduction of better product offerings as well as
international brands in. Thus, although"
## [4] "Breakfast Cereals is an evolving market, as it evolves, it will creat
e scope to overcome challenges like product"
## [5] "awareness, brand loyalty, understanding consumer buying behavior, and
developing products which tantalize the"
## [6] "Indian consumers' palate."

# Step1 - Removing the spaced now and Cleaning the text
Remove_Spaces <- paste(text_file, collapse = " ")
text_clean<-tolower(Remove_Spaces)
head(text_clean)
```

[1] "cereals have always held their share of the indian breakfast platter e.g. flattened rice flakes (chivda/poha) with milk popular in western and central india, whole wheat grits (dalia) in northern india, etc., alongside the regional staples which have been traditionally preferred, e.g. paranthas in the north, and idli or dosa in the south. for the most part, indians have inclined towards consuming their quintessential hot, cooked breakfast (naashta), and find it difficult to give in to such breakfast options as cereals with milk. however, lifestyles are changing in tandem with increasing spending power, greater time-poverty, higher need for convenience, and health consciousness. these factors have encouraged indians, especially in urban areas, to opt for breakfast cereals. the influence of 'western' lifestyles and eating trends has played an effective role in opening the gateway for experimenting with different tastes and varying eating preferences. breakfast cereals is a growing market in india which was pegged at usd 157 million in 2013 with the promise of, double-digit growth over the next five years. within this, hot cereals and muesli have been the fastest growing product categories in the recent past. among hot cereals, oats has gained high acceptance and popularity; this can be attributed to the consumers' awareness of the grain's health benefits. breakfast cereal manufacturers have used the nutritional plank as a growth strategy. they have introduced healthy fortified options to suit the needs of different demographics, e.g. like kellogg's products in kids' category, all-family segment, and for aging adults. furthermore, offerings have been customized to suit the indian consumer's taste, e.g. mango, kesar-elaichi, etc. the breakfast cereal market already offers various product categories, e.g. cornflakes, oats, muesli, wheat flakes, etc. the most popular category in the indian market is cornflakes, which holds roughly over 50% of the market, followed by oats and muesli, which are on an upward stride vis-à-vis consumer acceptability. india, being a large market with growing middle-income group and double-income households, holds great potential for the breakfast cereal market, which in turn must provide healthy, convenient, and tasteful options. the success story of kellogg's, bagrry's, etc. reinforces their belief in the potential of the breakfast cereal market which, though at a growing stage with few national and international players, has captured a large share of the pie, witnessed healthy year-on-year growth in the past couple of years, and is positive about future business growth. in terms of product offerings, players will need to focus more the enhanced variety and innovation, and also on efforts to boost consumer trials. while there is no doubt about the potential of the breakfast cereal market in india, it is not going to be easy, with challenges abounding. the biggest challenge is competition, as there are more than 50 regional players in the segment apart from major players like kellogg's, pepsico's quaker oats, bagrry's, etc. the confidence of existing players in terms of enhancing their reach, as well as of new entrants like marico, heinz, etc. will add to the competition. competition can also be expected from other fmcg chains which are not currently focusing on packaged breakfast as their core product offering e.g. ready-to-eat players like britannia with its range of ready-to-cook upmas, porridge, and pahas under the healthy start brand. this provides consumers with varied, healthy options that both suit the indian palate and can be termed traditional as well. additionally, brands like mcdonald's are eyeing the breakfast market in india with their range of breakfast menus which cater to travelling professionals. market penetration, with better-yet-econ

omical distribution, poses another challenge for international players who hope to enhance their existing market reach and foray into new market segments like low income groups. although manufacturers are offering small packs at economical prices, widening the consumer base, however, requires that this economical range is available in categories other than oats. regional players have a competitive edge over bigger brands with their strong and robust local distribution network. regional manufacturers are targeting not only kirana stores but also using multilevel and consumer-toconsumer marketing approaches to deepen their market penetration. these players are not spending aggressively in branding; however, they offer relatively better margins to kirana owners to push their products. again, their products are relatively economical, compared to the bigger brands, and suit the pockets of the middle and lower economic classes. bigger brands will need to develop a wholly new business model that can extend their reach amid such cutting edge competition. inflation in raw materials' prices is another big hurdle not only because of increasing product prices, but also via limiting profit margins. this is increasing the cost passed on to the final consumer, who might be tempted to switch over to relatively economical breakfast options. howsoever, increasing urbanization, health consciousness, and incidences of such diseases as heart attacks, diabetes, and blood pressure are pushing consumers to switch eating preferences from indulgent to wholesome. breakfast cereal manufacturers are likely to benefit from this trend as health benefits are one of the factors due to which consumers purchase cereals. the surge in demand will also drive innovation in the breakfast cereal market, with the introduction of better product offerings as well as international brands in. thus, although breakfast cereals is an evolving market, as it evolves, it will create scope to overcome challenges like product awareness, brand loyalty, understanding consumer buying behavior, and developing products which tantalize the indian consumers' palate."

```
text_clean <- tolower(text_clean)
head(text_clean)
```

```
## [1] "cereals have always held their share of the indian breakfast platter
e.g. flattened rice flakes (chivda/poha) with milk popular in western and central india, whole wheat grits (dalia) in northern india, etc., alongside the regional staples which have been traditionally preferred, e.g. paranthas in the north, and idli or dosa in the south. for the most part, indians have inclined towards consuming their quintessential hot, cooked breakfast (naashta), and find it difficult to give in to such breakfast options as cereals with milk. however, lifestyles are changing in tandem with increasing spending power, greater time-poverty, higher need for convenience, and health consciousness. these factors have encouraged indians, especially in urban areas, to opt for breakfast cereals. the influence of 'western' lifestyles and eating trends has played an effective role in opening the gateway for experimenting with different tastes and varying eating preferences. breakfast cereals is a growing market in india which was pegged at usd 157 million in 2013 with the promise of , double-digit growth over the next five years. within this, hot cereals and muesli have been the fastest growing product categories in the recent past. among hot cereals, oats has gained high acceptance and popularity; this can be attributed to the consumers' awareness of the grain's health benefits. bre
```

breakfast cereal manufacturers have used the nutritional plank as a growth strategy. they have introduced healthy fortified options to suit the needs of different demographics, e.g. like kellogg's products in kids' category, all-family segment, and for aging adults. furthermore, offerings have been customized to suit the indian consumer's taste, e.g. mango, kesar-elaichi, etc. the breakfast cereal market already offers various product categories, e.g. cornflakes, oats, muesli, wheat flakes, etc. the most popular category in the indian market is cornflakes, which holds roughly over 50% of the market, followed by oats and muesli, which are on an upward stride vis-à-vis consumer acceptability. india, being a large market with growing middle-income group and double-income households, holds great potential for the breakfast cereal market, which in turn must provide healthy, convenient, and tasteful options. the success story of kellogg's, bagrry's, etc. reinforces their belief in the potential of the breakfast cereal market which, though at a growing stage with few national and international players, has captured a large share of the pie, witnessed healthy year-on-year growth in the past couple of years, and is positive about future business growth. in terms of product offerings, players will need to focus more on the enhanced variety and innovation, and also on efforts to boost consumer trials. while there is no doubt about the potential of the breakfast cereal market in india, it is not going to be easy, with challenges abounding. the biggest challenge is competition, as there are more than 50 regional players in the segment apart from major players like kellogg's, pepsico's quaker oats, bagrry's, etc. the confidence of existing players in terms of enhancing their reach, as well as of new entrants like marico, heinz, etc. will add to the competition. competition can also be expected from other fmcg chains which are not currently focusing on packaged breakfast as their core product offering e.g. ready-to-eat players like britannia with its range of ready-to-cook upmas, porridge, and pohas under the healthy start brand. this provides consumers with varied, healthy options that both suit the indian palate and can be termed traditional as well. additionally, brands like mcdonald's are eyeing the breakfast market in india with their range of breakfast menus which cater to travelling professionals. market penetration, with better-yet-economical distribution, poses another challenge for international players who hope to enhance their existing market reach and foray into new market segments like low income groups. although manufacturers are offering small packs at economical prices, widening the consumer base, however, requires that this economical range is available in categories other than oats. regional players have a competitive edge over bigger brands with their strong and robust local distribution network. regional manufacturers are targeting not only kirana stores but also using multilevel and consumer-to-consumer marketing approaches to deepen their market penetration. these players are not spending aggressively in branding; however, they offer relatively better margins to kirana owners to push their products. again, their products are relatively economical, compared to the bigger brands, and suit the pockets of the middle and lower economic classes. bigger brands will need to develop a wholly new business model that can extend their reach amid such cutting edge competition. inflation in raw materials' prices is another big hurdle not only because of increasing product prices, but also via limiting profit margins. this is increasing the cost passed on to the final consumer, who might be tempted to switch over to relatively economical breakfast options. howsoever, increasing urbanization, health

h consciousness, and incidences of such diseases as heart attacks, diabetes, and blood pressure are pushing consumers to switch eating preferences from indulgent to wholesome. breakfast cereal manufacturers are likely to benefit from this trend as health benefits are one of the factors due to which consumers purchase cereals. the surge in demand will also drive innovation in the breakfast cereal market, with the introduction of better product offerings as well as international brands in. thus, although breakfast cereals is an evolving market, as it evolves, it will create scope to overcome challenges like product awareness, brand loyalty, understanding consumer buying behavior, and developing products which tantalize the indian consumers' palate."

```
# Step 2 - Removing the punctuation using "\W" and digits using "\\d"
clean_text1=gsub(pattern = "\\W", replace = " ",text_clean)
head(clean_text1)
```

```
## [1] "cereals have always held their share of the indian breakfast platter e g flattened rice flakes chivda poha with milk popular in western and central india whole wheat grits dalia in northern india etc alongside the regional staples which have been traditionally preferred e g paranthas in the north and idli or dosa in the south for the most part indians have inclined towards consuming their quintessential hot cooked breakfast naashta and find it difficult to give in to such breakfast options as cereals with milk however lifestyles are changing in tandem with increasing spending power greater time poverty higher need for convenience and health consciousness these factors have encouraged indians especially in urban areas to opt for breakfast cereals the influence of western lifestyles and eating trends has played an effective role in opening the gateway for experimenting with different tastes and varying eating preferences breakfast cereals is a growing market in india which was pegged at usd 157 million in 2013 with the promise of double digit growth over the next five years within this hot cereals and muesli have been the fastest growing product categories in the recent past among hot cereals oats has gained high acceptance and popularity this can be attributed to the consumers awareness of the grains health benefits breakfast cereal manufacturers have used the nutritional plank as a growth strategy they have introduced healthy fortified options to suit the needs of different demographics e g like kellogg s products in kids category all family segment and for aging adults furthermore offerings have been customized to suit the indian consumer s taste e g mango kesar elaichi etc the breakfast cereal market already offers various product categories e g cornflakes oats muesli wheat flakes etc the most popular category in the indian market is cornflakes which holds roughly over 50 of the market followed by oats and muesli which are on an upward stride vis à vis consumer acceptability india being a large market with growing middle income group and double income households holds great potential for the breakfast cereal market which in turn must provide healthy convenient and tasteful options the success story of kellogg s bagrry s etc reinforces their belief in the potential of the breakfast cereal market which though at a growing stage with few national and international players has captured a large share of the pie witnessed healthy year on year growth in the past couple of years and is positive about future business growth in terms of product offerings players will need to f
```

focus more the enhanced variety and innovation and also on efforts to boost consumer trials while there is no doubt about the potential of the breakfast cereal market in india it is not going to be easy with challenges abounding the biggest challenge is competition as there are more than 50 regional players in the segment apart from major players like kellogg's pepsico's quaker oats bagrry's etc the confidence of existing players in terms of enhancing their reach as well as of new entrants like marico heinz etc will add to the competition competition can also be expected from other fmcg chains which are not currently focusing on packaged breakfast as their core product offering e.g ready to eat players like britannia with its range of ready to cook upmas porridge and pohas under the healthy start brand this provides consumers with varied healthy options that both suit the indian palate and can be termed traditional as well additionally brands like mcdonald's are eyeing the breakfast market in india with their range of breakfast menus which cater to travelling professionals market penetration with better yet economical distribution poses another challenge for international players who hope to enhance their existing market reach and foray into new market segments like low income groups although manufacturers are offering small packs at economical prices widening the consumer base however requires that this economical range is available in categories other than oats regional players have a competitive edge over bigger brands with their strong and robust local distribution network regional manufacturers are targeting not only kirana stores but also using multilevel and consumer to consumer marketing approaches to deepen their market penetration these players are not spending aggressively in branding however they offer relatively better margins to kirana owners to push their products again their products are relatively economical compared to the bigger brands and suit the pockets of the middle and lower economic classes bigger brands will need to develop a wholly new business model that can extend their reach amid such cutting edge competition inflation in raw materials prices is another big hurdle not only because of increasing product prices but also via limiting profit margins this is increasing the cost passed on to the final consumer who might be tempted to switch over to relatively economical breakfast options howsoever increasing urbanization health consciousness and incidences of such diseases as heart attacks diabetes and blood pressure are pushing consumers to switch eating preferences from indulgent to wholesome breakfast cereal manufacturers are likely to benefit from this trend as health benefits are one of the factors due to which consumers purchase cereals the surge in demand will also drive innovation in the breakfast cereal market with the introduction of better product offerings as well as in international brands in thus although breakfast cereals is an evolving market as it evolves it will create scope to overcome challenges like product awareness brand loyalty understanding consumer buying behavior and developing products which tantalize the indian consumers palate "

```
clean_text2 = gsub(pattern = "\\d", replace = " ", clean_text1)
head(clean_text2)
```

```
## [1] "cereals have always held their share of the indian breakfast platter
e.g flattened rice flakes chivda poha with milk popular in western and cen
tral india whole wheat grits dalia in northern india etc alongside the
```

regional staples which have been traditionally preferred e.g. paranthas in the north and idli or dosa in the south for the most part indians have inclined towards consuming their quintessential hot cooked breakfast naashta and find it difficult to give in to such breakfast options as cereals with milk however lifestyles are changing in tandem with increasing spending power greater time poverty higher need for convenience and health consciousness these factors have encouraged indians especially in urban areas to opt for breakfast cereals the influence of western lifestyles and eating trends has played an effective role in opening the gateway for experimenting with different tastes and varying eating preferences breakfast cereals is a growing market in india which was pegged at usd million in with the promise of double digit growth over the next five years within this hot cereals and muesli have been the fastest growing product categories in the recent past among hot cereals oats has gained high acceptance and popularity this can be attributed to the consumers awareness of the grains health benefits breakfast cereal manufacturers have used the nutritional plank as a growth strategy they have introduced healthy fortified options to suit the needs of different demographics e.g. like kellogg's products in kids category all family segment and for aging adults furthermore offerings have been customized to suit the indian consumer's taste e.g. mango kesar elaichi etc the breakfast cereal market already offers various product categories e.g. cornflakes oats muesli wheat flakes etc the most popular category in the indian market is cornflakes which holds roughly over of the market followed by oats and muesli which are on an upward stride vis à vis consumer acceptability india being a large market with growing middle income group and double income households holds great potential for the breakfast cereal market which in turn must provide healthy convenient and tasteful options the success story of kellogg's bagrry's etc reinforces their belief in the potential of the breakfast cereal market which though at a growing stage with few national and international players has captured a large share of the pie witnessed healthy year on year growth in the past couple of years and is positive about future business growth in terms of product offerings players will need to focus more the enhanced variety and innovation and also on efforts to boost consumer trials while there is no doubt about the potential of the breakfast cereal market in india it is not going to be easy with challenges abounding the biggest challenge is competition as there are more than regional players in the segment apart from major players like kellogg's pepsico's quaker oats bagrry's etc the confidence of existing players in terms of enhancing their reach as well as of new entrants like marico heinz etc will add to the competition competition can also be expected from other fmcg chains which are not currently focusing on packaged breakfast as their core product offering e.g. ready to eat players like britannia with its range of ready to cook upmas porridge and pohas under the healthy start brand this provides consumers with varied healthy options that both suit the indian palate and can be termed traditional as well additionally brands like mcdonald's are eyeing the breakfast market in india with their range of breakfast menus which cater to travelling professionals market penetration with better yet economical distribution poses another challenge for international players who hope to enhance their existing market reach and foray into new market segments like low income groups although manufacturers are offering small packs at economical

prices widening the consumer base however requires that this economical range is available in categories other than oats regional players have a competitive edge over bigger brands with their strong and robust local distribution network regional manufacturers are targeting not only kirana stores but also using multilevel and consumer to consumer marketing approaches to deepen their market penetration these players are not spending aggressively in branding however they offer relatively better margins to kirana owners to push their products again their products are relatively economical compared to the bigger brands and suit the pockets of the middle and lower economic classes bigger brands will need to develop a wholly new business model that can extend their reach amid such cutting edge competition inflation in raw materials prices is another big hurdle not only because of increasing product prices but also via limiting profit margins this is increasing the cost passed on to the final consumer who might be tempted to switch over to relatively economical breakfast options howsoever increasing urbanization health consciousness and incidences of such diseases as heart attacks diabetes and blood pressure are pushing consumers to switch eating preferences from indulgent to wholesome breakfast cereal manufacturers are likely to benefit from this trend as health benefits are one of the factors due to which consumers purchase cereals the surge in demand will also drive innovation in the breakfast cereal market with the introduction of better product offerings as well as international brands in thus although breakfast cereals is an evolving market as it evolves it will create scope to overcome challenges like product awareness brand loyalty understanding consumer buying behavior and developing products which tantalize the indian consumers palate "

#Step 3 - Removing the Stop words
stopwords()

## [1]	"i"	"me"	"my"	"myself"	"we"
## [6]	"our"	"ours"	"ourselves"	"you"	"your"
## [11]	"yours"	"yourself"	"yourselves"	"he"	"him"
## [16]	"his"	"himself"	"she"	"her"	"hers"
## [21]	"herself"	"it"	"its"	"itself"	"they"
## [26]	"them"	"their"	"theirs"	"themselves"	"what"
## [31]	"which"	"who"	"whom"	"this"	"that"
## [36]	"these"	"those"	"am"	"is"	"are"
## [41]	"was"	"were"	"be"	"been"	"being"
## [46]	"have"	"has"	"had"	"having"	"do"
## [51]	"does"	"did"	"doing"	"would"	"should"
## [56]	"could"	"ought"	"i'm"	"you're"	"he's"
## [61]	"she's"	"it's"	"we're"	"they're"	"i've"
## [66]	"you've"	"we've"	"they've"	"i'd"	"you'd"
## [71]	"he'd"	"she'd"	"we'd"	"they'd"	"i'll"
## [76]	"you'll"	"he'll"	"she'll"	"we'll"	"they'll"
## [81]	"isn't"	"aren't"	"wasn't"	"weren't"	"hasn't"
## [86]	"haven't"	"hadn't"	"doesn't"	"don't"	"didn't"
## [91]	"won't"	"wouldn't"	"shan't"	"shouldn't"	"can't"
## [96]	"cannot"	"couldn't"	"mustn't"	"let's"	"that's"
## [101]	"who's"	"what's"	"here's"	"there's"	"when's"

## [106]	"where's"	"why's"	"how's"	"a"	"an"
## [111]	"the"	"and"	"but"	"if"	"or"
## [116]	"because"	"as"	"until"	"while"	"of"
## [121]	"at"	"by"	"for"	"with"	"about"
## [126]	"against"	"between"	"into"	"through"	"during"
## [131]	"before"	"after"	"above"	"below"	"to"
## [136]	"from"	"up"	"down"	"in"	"out"
## [141]	"on"	"off"	"over"	"under"	"again"
## [146]	"further"	"then"	"once"	"here"	"there"
## [151]	"when"	"where"	"why"	"how"	"all"
## [156]	"any"	"both"	"each"	"few"	"more"
## [161]	"most"	"other"	"some"	"such"	"no"
## [166]	"nor"	"not"	"only"	"own"	"same"
## [171]	"so"	"than"	"too"	"very"	

```
clean_text3<-removeWords(clean_text2,stopwords())
head(clean_text3)
```

```
## [1] "cereals always held share indian breakfast platter e g flattened
rice flakes chivda poha milk popular western central india whole wheat
grits dalia northern india etc alongside regional staples tradition
ally preferred e g paranthas north idli dosa south part indians
inclined towards consuming quintessential hot cooked breakfast naashta
find difficult give breakfast options cereals milk however lifestyle
s changing tandem increasing spending power greater time poverty higher
need convenience health consciousness factors encouraged indians espec
ially urban areas opt breakfast cereals influence western lifestyles
eating trends played effective role opening gateway experimenting diffe
rent tastes varying eating preferences breakfast cereals growing market
india pegged usd million promise double digit growth next
five years within hot cereals muesli fastest growing product categorie
s recent past among hot cereals oats gained high acceptance popularity
can attributed consumers awareness grain s health benefits breakfast c
ereal manufacturers used nutritional plank growth strategy introduced
healthy fortified options suit needs different demographics e g like kel
logg s products kids category family segment aging adults furthermore
offerings customized suit indian consumer s taste e g mango kesar elai
chi etc breakfast cereal market already offers various product categories
e g cornflakes oats muesli wheat flakes etc popular category indian
market cornflakes holds roughly market followed oats muesli
upward stride vis à vis consumer acceptability india large market growin
g middle income group double income households holds great potential brea
kfast cereal market turn must provide healthy convenient tasteful optio
ns success story kellogg s bagrry s etc reinforces belief potential
breakfast cereal market though growing stage national international pl
ayers captured large share pie witnessed healthy year year growth pa
st couple years positive future business growth terms product offerin
gs players will need focus enhanced variety innovation also efforts
boost consumer trials doubt potential breakfast cereal market india
going easy challenges abounding biggest challenge competition"
```


regional players segment apart major players like kellogg s pepsico s quaker oats bagrry s etc confidence existing players terms enhancing reach well new entrants like marico heinz etc will add competition competition can also expected fmcg chains currently focusing packaged breakfast core product offering e g ready eat players like britannia range readyto cook upmas porridge pohas healthy start brand provides consumers varied healthy options suit indian palate can termed traditional well additionally brands like mcdonald s eyeing breakfast market india range breakfast menus cater travelling professionals market penetration better yet economical distribution poses another challenge international players hope enhance existing market reach foray new market segments like low income groups although manufacturers offering small packs economical prices widening consumer base however requires economical range available categories oats regional players competitive edge bigger brands strong robust local distribution network regional manufacturers targeting kirana stores also using multilevel consumer toconsumer marketing approaches deepen market penetration players spending aggressively branding however offer relatively better margins kirana owners push products products relatively economical compared bigger brands suit pockets middle lower economic classes bigger brands will need develop wholly new business model can extend reach amid cutting edge competition inflation raw materials prices another big hurdle increasing product prices also via limiting profit margins increasing cost passed final consumer might tempted switch relatively economical breakfast options howsoever increasing urbanization health consciousness incidences diseases heart attacks diabetes blood pressure pushing consumers switch eating preferences indulgent wholesome breakfast cereal manufacturers likely benefit trend health benefits one factors due consumers purchase cereals surge demand will also drive innovation breakfast cereal market introduction better product offerings well international brands thus although breakfast cereals evolving market evolves will create scope overcome challenges like product awareness brand loyalty understanding consumer buying behavior developing products tantalize indian consumers palate "

#Step 4 - Removing the single letter alphabets

```
Remove_alhabetbs = gsub(pattern = "\\b[A-z]\\b{1}", replace = " ", clean_text3)
head(Remove_alhabetbs)
```

```
## [1] "cereals always held share indian breakfast platter flattened
rice flakes chivda poha milk popular western central india whole wheat
grits dalia northern india etc alongside regional staples tradition
ally preferred paranthas north idli dosa south part indians
inclined towards consuming quintessential hot cooked breakfast naashta
find difficult give breakfast options cereals milk however lifestyle
s changing tandem increasing spending power greater time poverty higher
need convenience health consciousness factors encouraged indians espec
ially urban areas opt breakfast cereals influence western lifestyles
eating trends played effective role opening gateway experimenting diffe
rent tastes varying eating preferences breakfast cereals growing market
india pegged usd million promise double digit growth next
```


five years within hot cereals muesli fastest growing product categories recent past among hot cereals oats gained high acceptance popularity can attributed consumers awareness grain health benefits breakfast cereal manufacturers used nutritional plank growth strategy introduced healthy fortified options suit needs different demographics like Kellogg products kids category family segment aging adults furthermore offerings customized suit Indian consumer taste mango kesar elai chi etc breakfast cereal market already offers various product categories cornflakes oats muesli wheat flakes etc popular category Indian market cornflakes holds roughly market followed oats muesli upward stride vis à vis consumer acceptability India large market growing middle income group double income households holds great potential breakfast cereal market turn must provide healthy convenient tasteful options success story Kellogg Baggry etc reinforces belief potential breakfast cereal market though growing stage national international players captured large share pie witnessed healthy year year growth past couple years positive future business growth terms product offerings players will need focus enhanced variety innovation also efforts boost consumer trials doubt potential breakfast cereal market India going easy challenges abounding biggest challenge competition regional players segment apart major players like Kellogg Pepsico Quaker oats Baggry etc confidence existing players terms enhancing reach well new entrants like Marico Heinz etc will add competition competition can also expected FMCG chains currently focusing packaged breakfast core product offering ready eat players like Britannia Range Ready to cook Upma's porridge Pohas healthy start brand provides consumers varied healthy options suit Indian palate can termed traditional well additionally brands like McDonald eyeing breakfast market India range breakfast menus cater travelling professionals market penetration better yet economical distribution poses another challenge international players hope enhance existing market reach foray new market segments like low income groups although manufacturers offering small packs economical prices widening consumer base however requires economical range available categories oats regional players competitive edge bigger brands strong robust local distribution network regional manufacturers targeting kirana stores also using multilevel consumer to consumer marketing approaches deepen market penetration players spending aggressively branding however offer relatively better margins kirana owners push products products relatively economical compared bigger brands suit pockets middle lower economic classes bigger brands will need develop wholly new business model can extend reach amid cutting edge competition inflation raw materials prices another big hurdle increasing product prices also via limiting profit margins increasing cost passed final consumer might tempted switch relatively economical breakfast options howsoever increasing urbanization health consciousness incidences diseases heart attacks diabetes blood pressure pushing consumers switch eating preferences indulgent wholesome breakfast cereal manufacturers likely benefit trend health benefits one factors due consumers purchase cereals surge demand will also drive innovation breakfast cereal market introduction better product offerings well international brands thus although breakfast c

cereals evolving market evolves will create scope overcome challenges like product awareness brand loyalty understanding consumer buying behavior developing products tantalize indian consumers palate "

Step 5 - Removing white spaces

```
Remove_whitespace= stripWhitespace(Remove_alhabets)  
head(Remove_whitespace)
```

```
## [1] "cereals always held share indian breakfast platter flattened rice flakes chivda poha milk popular western central india whole wheat grits dalia northern india etc alongside regional staples traditionally preferred paranthas north idli dosa south part indians inclined towards consuming quintessential hot cooked breakfast naashta find difficult give breakfast options cereals milk however lifestyles changing tandem increasing spending power greater time poverty higher need convenience health consciousness factors encouraged indians especially urban areas opt breakfast cereals influence western lifestyles eating trends played effective role opening gateway experimenting different tastes varying eating preferences breakfast cereals growing market india pegged usd million promise double digit growth next five years within hot cereals muesli fastest growing product categories recent past among hot cereals oats gained high acceptance popularity can attributed consumers awareness grain health benefits breakfast cereal manufacturers used nutritional plank growth strategy introduced healthy fortified options suit needs different demographics like kellogg products kids category family segment aging adults furthermore offerings customized suit indian consumer taste mango kesar elaichi etc breakfast cereal market already offers various product categories cornflakes oats muesli wheat flakes etc popular category indian market cornflakes holds roughly market followed oats muesli upward stride vis à vis consumer acceptability india large market growing middle income group double income households holds great potential breakfast cereal market turn must provide healthy convenient tasteful options success story kellogg bagrry etc reinforces belief potential breakfast cereal market though growing stage national international players captured large share pie witnessed healthy year year growth past couple years positive future business growth terms product offerings players will need focus enhanced variety innovation also efforts boost consumer trials doubt potential breakfast cereal market india going easy challenges abounding biggest challenge competition regional players segment apart major players like kellogg pepsico quaker oats bagrry etc confidence existing players terms enhancing reach well new entrants like marico heinz etc will add competition competition can also expected fmcg chains currently focusing packaged breakfast core product offering ready eat players like britannia range readyto cook upmas porridge pahas healthy start brand provides consumers varied healthy options suit indian palate can termed traditional well additionally brands like mcdonald eyeing breakfast market india range breakfast menus cater travelling professionals market penetration better yet economical distribution poses another challenge international players hope enhance existing market reach foray new market segments like low income groups although manufacturers offering small packs economical prices widening consumer base however requires economical range available categories oats regional players competitive edge bigger brands strong robust local distribution network regional manufacturers targeting kirana s
```

tores also using multilevel consumer toconsumer marketing approaches deepen market penetration players spending aggressively branding however offer relatively better margins kirana owners push products products relatively economical compared bigger brands suit pockets middle lower economic classes bigger brands will need develop wholly new business model can extend reach amid cutting edge competition inflation raw materials prices another big hurdle increasing product prices also via limiting profit margins increasing cost passed final consumer might tempted switch relatively economical breakfast options hows oever increasing urbanization health consciousness incidences diseases heart attacks diabetes blood pressure pushing consumers switch eating preferences indulgent wholesome breakfast cereal manufacturers likely benefit trend health benefits one factors due consumers purchase cereals surge demand will also drive innovation breakfast cereal market introduction better product offerings well international brands thus although breakfast cereals evolving market evolves will create scope overcome challenges like product awareness brand loyalty understanding consumer buying behavior developing products tantalize indian consumers palate "

Step 6 - Split the words

```
Text_split = strsplit(Remove_whitespace, " ")
head(Text_split)
```

```
## [[1]]
## [1] "cereals"      "always"      "held"        "share"
## [5] "indian"       "breakfast"   "platter"     "flattened"
## [9] "rice"         "flakes"     "chivda"      "poha"
## [13] "milk"         "popular"     "western"     "central"
## [17] "india"        "whole"       "wheat"       "grits"
## [21] "dalia"        "northern"    "india"       "etc"
## [25] "alongside"    "regional"    "staples"     "traditionally"
## [29] "preferred"    "paranhas"    "north"       "idli"
## [33] "dosa"         "south"       "part"        "indians"
## [37] "inclined"     "towards"     "consuming"   "quintessential"
## [41] "hot"          "cooked"      "breakfast"   "naashta"
## [45] "find"         "difficult"   "give"        "breakfast"
## [49] "options"      "cereals"     "milk"        "however"
## [53] "lifestyles"   "changing"    "tandem"      "increasing"
## [57] "spending"     "power"       "greater"     "time"
## [61] "poverty"      "higher"      "need"        "convenience"
## [65] "health"       "consciousness" "factors"     "encouraged"
## [69] "indians"      "especially"   "urban"       "areas"
## [73] "opt"          "breakfast"   "cereals"     "influence"
## [77] "western"      "lifestyles"   "eating"      "trends"
## [81] "played"       "effective"    "role"        "opening"
## [85] "gateway"      "experimenting" "different"    "tastes"
## [89] "varying"      "eating"       "preferences" "breakfast"
## [93] "cereals"      "growing"     "market"      "india"
## [97] "pegged"       "usd"          "million"     "promise"
## [101] "double"       "digit"        "growth"      "next"
## [105] "five"         "years"        "within"      "hot"
```

## [109]	"cereals"	"muesli"	"fastest"	"growing"
## [113]	"product"	"categories"	"recent"	"past"
## [117]	"among"	"hot"	"cereals"	"oats"
## [121]	"gained"	"high"	"acceptance"	"popularity"
## [125]	"can"	"attributed"	"consumers"	"awareness"
## [129]	"grain"	"health"	"benefits"	"breakfast"
## [133]	"cereal"	"manufacturers"	"used"	"nutritional"
## [137]	"plank"	"growth"	"strategy"	"introduced"
## [141]	"healthy"	"fortified"	"options"	"suit"
## [145]	"needs"	"different"	"demographics"	"like"
## [149]	"kellogg"	"products"	"kids"	"category"
## [153]	"family"	"segment"	"aging"	"adults"
## [157]	"furthermore"	"offerings"	"customized"	"suit"
## [161]	"indian"	"consumer"	"taste"	"mango"
## [165]	"kesar"	"elaichi"	"etc"	"breakfast"
## [169]	"cereal"	"market"	"already"	"offers"
## [173]	"various"	"product"	"categories"	"cornflakes"
## [177]	"oats"	"muesli"	"wheat"	"flakes"
## [181]	"etc"	"popular"	"category"	"indian"
## [185]	"market"	"cornflakes"	"holds"	"roughly"
## [189]	"market"	"followed"	"oats"	"muesli"
## [193]	"upward"	"stride"	"vis"	"à"
## [197]	"vis"	"consumer"	"acceptability"	"india"
## [201]	"large"	"market"	"growing"	"middle"
## [205]	"income"	"group"	"double"	"income"
## [209]	"households"	"holds"	"great"	"potential"
## [213]	"breakfast"	"cereal"	"market"	"turn"
## [217]	"must"	"provide"	"healthy"	"convenient"
## [221]	"tasteful"	"options"	"success"	"story"
## [225]	"kellogg"	"bagrry"	"etc"	"reinforces"
## [229]	"belief"	"potential"	"breakfast"	"cereal"
## [233]	"market"	"though"	"growing"	"stage"
## [237]	"national"	"international"	"players"	"captured"
## [241]	"large"	"share"	"pie"	"witnessed"
## [245]	"healthy"	"year"	"year"	"growth"
## [249]	"past"	"couple"	"years"	"positive"
## [253]	"future"	"business"	"growth"	"terms"
## [257]	"product"	"offerings"	"players"	"will"
## [261]	"need"	"focus"	"enhanced"	"variety"
## [265]	"innovation"	"also"	"efforts"	"boost"
## [269]	"consumer"	"trials"	"doubt"	"potential"
## [273]	"breakfast"	"cereal"	"market"	"india"
## [277]	"going"	"easy"	"challenges"	"abounding"
## [281]	"biggest"	"challenge"	"competition"	"regional"
## [285]	"players"	"segment"	"apart"	"major"
## [289]	"players"	"like"	"kellogg"	"pepsico"
## [293]	"quaker"	"oats"	"bagrry"	"etc"
## [297]	"confidence"	"existing"	"players"	"terms"
## [301]	"enhancing"	"reach"	"well"	"new"
## [305]	"entrants"	"like"	"marico"	"heinz"

## [309]	"etc"	"will"	"add"	"competition"
## [313]	"competition"	"can"	"also"	"expected"
## [317]	"fmcg"	"chains"	"currently"	"focusing"
## [321]	"packaged"	"breakfast"	"core"	"product"
## [325]	"offering"	"ready"	"eat"	"players"
## [329]	"like"	"britannia"	"range"	"readyto"
## [333]	"cook"	"upmas"	"porridge"	"pohas"
## [337]	"healthy"	"start"	"brand"	"provides"
## [341]	"consumers"	"varied"	"healthy"	"options"
## [345]	"suit"	"indian"	"palate"	"can"
## [349]	"termed"	"traditional"	"well"	"additionally"
## [353]	"brands"	"like"	"mcdonald"	"eyeing"
## [357]	"breakfast"	"market"	"india"	"range"
## [361]	"breakfast"	"menus"	"cater"	"travelling"
## [365]	"professionals"	"market"	"penetration"	"better"
## [369]	"yet"	"economical"	"distribution"	"poses"
## [373]	"another"	"challenge"	"international"	"players"
## [377]	"hope"	"enhance"	"existing"	"market"
## [381]	"reach"	"foray"	"new"	"market"
## [385]	"segments"	"like"	"low"	"income"
## [389]	"groups"	"although"	"manufacturers"	"offering"
## [393]	"small"	"packs"	"economical"	"prices"
## [397]	"widening"	"consumer"	"base"	"however"
## [401]	"requires"	"economical"	"range"	"available"
## [405]	"categories"	"oats"	"regional"	"players"
## [409]	"competitive"	"edge"	"bigger"	"brands"
## [413]	"strong"	"robust"	"local"	"distribution"
## [417]	"network"	"regional"	"manufacturers"	"targeting"
## [421]	"kirana"	"stores"	"also"	"using"
## [425]	"multilevel"	"consumer"	"toconsumer"	"marketing"
## [429]	"approaches"	"deepen"	"market"	"penetration"
## [433]	"players"	"spending"	"aggressively"	"branding"
## [437]	"however"	"offer"	"relatively"	"better"
## [441]	"margins"	"kirana"	"owners"	"push"
## [445]	"products"	"products"	"relatively"	"economical"
## [449]	"compared"	"bigger"	"brands"	"suit"
## [453]	"pockets"	"middle"	"lower"	"economic"
## [457]	"classes"	"bigger"	"brands"	"will"
## [461]	"need"	"develop"	"wholly"	"new"
## [465]	"business"	"model"	"can"	"extend"
## [469]	"reach"	"amid"	"cutting"	"edge"
## [473]	"competition"	"inflation"	"raw"	"materials"
## [477]	"prices"	"another"	"big"	"hurdle"
## [481]	"increasing"	"product"	"prices"	"also"
## [485]	"via"	"limiting"	"profit"	"margins"
## [489]	"increasing"	"cost"	"passed"	"final"
## [493]	"consumer"	"might"	"tempted"	"switch"
## [497]	"relatively"	"economical"	"breakfast"	"options"
## [501]	"howsoever"	"increasing"	"urbanization"	"health"
## [505]	"consciousness"	"incidences"	"diseases"	"heart"

```
## [509] "attacks"      "diabetes"      "blood"         "pressure"
## [513] "pushing"      "consumers"     "switch"        "eating"
## [517] "preferences"  "indulgent"     "wholesome"     "breakfast"
## [521] "cereal"       "manufacturers" "likely"        "benefit"
## [525] "trend"        "health"        "benefits"      "one"
## [529] "factors"      "due"           "consumers"     "purchase"
## [533] "cereals"      "surge"         "demand"        "will"
## [537] "also"         "drive"         "innovation"    "breakfast"
## [541] "cereal"       "market"        "introduction"  "better"
## [545] "product"      "offerings"     "well"          "international"
## [549] "brands"       "thus"          "although"      "breakfast"
## [553] "cereals"      "evolving"      "market"        "evolves"
## [557] "will"         "create"        "scope"         "overcome"
## [561] "challenges"   "like"          "product"       "awareness"
## [565] "brand"        "loyalty"       "understanding" "consumer"
## [569] "buying"       "behavior"      "developing"    "products"
## [573] "tantalize"    "indian"        "consumers"     "palate"
```

#Step 7 - Check the word frequency now

```
word_freq=table(Text_split)
```

```
head(word_freq)
```

```
## Text_split
```

```
##          à      abounding acceptability      acceptance      add
##          1          1          1          1          1
## additionally
##          1
```

```
cbind(names(word_freq), as.integer(word_freq))
```

```
##          [,1]          [,2]
## [1,] "à"          "1"
## [2,] "abounding"   "1"
## [3,] "acceptability" "1"
## [4,] "acceptance"  "1"
## [5,] "add"         "1"
## [6,] "additionally" "1"
## [7,] "adults"      "1"
## [8,] "aggressively" "1"
## [9,] "aging"       "1"
## [10,] "alongside"  "1"
## [11,] "already"    "1"
## [12,] "also"       "5"
## [13,] "although"   "2"
## [14,] "always"     "1"
## [15,] "amid"       "1"
## [16,] "among"      "1"
## [17,] "another"    "2"
## [18,] "apart"      "1"
## [19,] "approaches" "1"
## [20,] "areas"      "1"
```

##	[21,]	"attacks"	"1"
##	[22,]	"attributed"	"1"
##	[23,]	"available"	"1"
##	[24,]	"awareness"	"2"
##	[25,]	"bagrry"	"2"
##	[26,]	"base"	"1"
##	[27,]	"behavior"	"1"
##	[28,]	"belief"	"1"
##	[29,]	"benefit"	"1"
##	[30,]	"benefits"	"2"
##	[31,]	"better"	"3"
##	[32,]	"big"	"1"
##	[33,]	"bigger"	"3"
##	[34,]	"biggest"	"1"
##	[35,]	"blood"	"1"
##	[36,]	"boost"	"1"
##	[37,]	"brand"	"2"
##	[38,]	"branding"	"1"
##	[39,]	"brands"	"5"
##	[40,]	"breakfast"	"17"
##	[41,]	"britannia"	"1"
##	[42,]	"business"	"2"
##	[43,]	"buying"	"1"
##	[44,]	"can"	"4"
##	[45,]	"captured"	"1"
##	[46,]	"categories"	"3"
##	[47,]	"category"	"2"
##	[48,]	"cater"	"1"
##	[49,]	"central"	"1"
##	[50,]	"cereal"	"7"
##	[51,]	"cereals"	"8"
##	[52,]	"chains"	"1"
##	[53,]	"challenge"	"2"
##	[54,]	"challenges"	"2"
##	[55,]	"changing"	"1"
##	[56,]	"chivda"	"1"
##	[57,]	"classes"	"1"
##	[58,]	"compared"	"1"
##	[59,]	"competition"	"4"
##	[60,]	"competitive"	"1"
##	[61,]	"confidence"	"1"
##	[62,]	"consciousness"	"2"
##	[63,]	"consumer"	"7"
##	[64,]	"consumers"	"5"
##	[65,]	"consuming"	"1"
##	[66,]	"convenience"	"1"
##	[67,]	"convenient"	"1"
##	[68,]	"cook"	"1"
##	[69,]	"cooked"	"1"
##	[70,]	"core"	"1"

##	[71,]	"cornflakes"	"2"
##	[72,]	"cost"	"1"
##	[73,]	"couple"	"1"
##	[74,]	"create"	"1"
##	[75,]	"currently"	"1"
##	[76,]	"customized"	"1"
##	[77,]	"cutting"	"1"
##	[78,]	"dalia"	"1"
##	[79,]	"deepen"	"1"
##	[80,]	"demand"	"1"
##	[81,]	"demographics"	"1"
##	[82,]	"develop"	"1"
##	[83,]	"developing"	"1"
##	[84,]	"diabetes"	"1"
##	[85,]	"different"	"2"
##	[86,]	"difficult"	"1"
##	[87,]	"digit"	"1"
##	[88,]	"diseases"	"1"
##	[89,]	"distribution"	"2"
##	[90,]	"dosa"	"1"
##	[91,]	"double"	"2"
##	[92,]	"doubt"	"1"
##	[93,]	"drive"	"1"
##	[94,]	"due"	"1"
##	[95,]	"easy"	"1"
##	[96,]	"eat"	"1"
##	[97,]	"eating"	"3"
##	[98,]	"economic"	"1"
##	[99,]	"economical"	"5"
##	[100,]	"edge"	"2"
##	[101,]	"effective"	"1"
##	[102,]	"efforts"	"1"
##	[103,]	"elaichi"	"1"
##	[104,]	"encouraged"	"1"
##	[105,]	"enhance"	"1"
##	[106,]	"enhanced"	"1"
##	[107,]	"enhancing"	"1"
##	[108,]	"entrants"	"1"
##	[109,]	"especially"	"1"
##	[110,]	"etc"	"6"
##	[111,]	"evolves"	"1"
##	[112,]	"evolving"	"1"
##	[113,]	"existing"	"2"
##	[114,]	"expected"	"1"
##	[115,]	"experimenting"	"1"
##	[116,]	"extend"	"1"
##	[117,]	"eyeing"	"1"
##	[118,]	"factors"	"2"
##	[119,]	"family"	"1"
##	[120,]	"fastest"	"1"

## [121,]	"final"	"1"
## [122,]	"find"	"1"
## [123,]	"five"	"1"
## [124,]	"flakes"	"2"
## [125,]	"flattened"	"1"
## [126,]	"fmcg"	"1"
## [127,]	"focus"	"1"
## [128,]	"focusing"	"1"
## [129,]	"followed"	"1"
## [130,]	"foray"	"1"
## [131,]	"fortified"	"1"
## [132,]	"furthermore"	"1"
## [133,]	"future"	"1"
## [134,]	"gained"	"1"
## [135,]	"gateway"	"1"
## [136,]	"give"	"1"
## [137,]	"going"	"1"
## [138,]	"grain"	"1"
## [139,]	"great"	"1"
## [140,]	"greater"	"1"
## [141,]	"grits"	"1"
## [142,]	"group"	"1"
## [143,]	"groups"	"1"
## [144,]	"growing"	"4"
## [145,]	"growth"	"4"
## [146,]	"health"	"4"
## [147,]	"healthy"	"5"
## [148,]	"heart"	"1"
## [149,]	"heinz"	"1"
## [150,]	"held"	"1"
## [151,]	"high"	"1"
## [152,]	"higher"	"1"
## [153,]	"holds"	"2"
## [154,]	"hope"	"1"
## [155,]	"hot"	"3"
## [156,]	"households"	"1"
## [157,]	"however"	"3"
## [158,]	"howsoever"	"1"
## [159,]	"hurdle"	"1"
## [160,]	"idli"	"1"
## [161,]	"incidences"	"1"
## [162,]	"inclined"	"1"
## [163,]	"income"	"3"
## [164,]	"increasing"	"4"
## [165,]	"india"	"6"
## [166,]	"indian"	"5"
## [167,]	"indians"	"2"
## [168,]	"indulgent"	"1"
## [169,]	"inflation"	"1"
## [170,]	"influence"	"1"

## [171,]	"innovation"	"2"
## [172,]	"international"	"3"
## [173,]	"introduced"	"1"
## [174,]	"introduction"	"1"
## [175,]	"kellogg"	"3"
## [176,]	"kesar"	"1"
## [177,]	"kids"	"1"
## [178,]	"kirana"	"2"
## [179,]	"large"	"2"
## [180,]	"lifestyles"	"2"
## [181,]	"like"	"7"
## [182,]	"likely"	"1"
## [183,]	"limiting"	"1"
## [184,]	"local"	"1"
## [185,]	"low"	"1"
## [186,]	"lower"	"1"
## [187,]	"loyalty"	"1"
## [188,]	"major"	"1"
## [189,]	"mango"	"1"
## [190,]	"manufacturers"	"4"
## [191,]	"margins"	"2"
## [192,]	"marico"	"1"
## [193,]	"market"	"15"
## [194,]	"marketing"	"1"
## [195,]	"materials"	"1"
## [196,]	"mcdonald"	"1"
## [197,]	"menus"	"1"
## [198,]	"middle"	"2"
## [199,]	"might"	"1"
## [200,]	"milk"	"2"
## [201,]	"million"	"1"
## [202,]	"model"	"1"
## [203,]	"muesli"	"3"
## [204,]	"multilevel"	"1"
## [205,]	"must"	"1"
## [206,]	"naashta"	"1"
## [207,]	"national"	"1"
## [208,]	"need"	"3"
## [209,]	"needs"	"1"
## [210,]	"network"	"1"
## [211,]	"new"	"3"
## [212,]	"next"	"1"
## [213,]	"north"	"1"
## [214,]	"northern"	"1"
## [215,]	"nutritional"	"1"
## [216,]	"oats"	"5"
## [217,]	"offer"	"1"
## [218,]	"offering"	"2"
## [219,]	"offerings"	"3"
## [220,]	"offers"	"1"

## [221,]	"one"	"1"
## [222,]	"opening"	"1"
## [223,]	"opt"	"1"
## [224,]	"options"	"5"
## [225,]	"overcome"	"1"
## [226,]	"owners"	"1"
## [227,]	"packaged"	"1"
## [228,]	"packs"	"1"
## [229,]	"palate"	"2"
## [230,]	"paranthas"	"1"
## [231,]	"part"	"1"
## [232,]	"passed"	"1"
## [233,]	"past"	"2"
## [234,]	"pegged"	"1"
## [235,]	"penetration"	"2"
## [236,]	"pepsico"	"1"
## [237,]	"pie"	"1"
## [238,]	"plank"	"1"
## [239,]	"platter"	"1"
## [240,]	"played"	"1"
## [241,]	"players"	"9"
## [242,]	"pockets"	"1"
## [243,]	"poha"	"1"
## [244,]	"pohas"	"1"
## [245,]	"popular"	"2"
## [246,]	"popularity"	"1"
## [247,]	"porridge"	"1"
## [248,]	"poses"	"1"
## [249,]	"positive"	"1"
## [250,]	"potential"	"3"
## [251,]	"poverty"	"1"
## [252,]	"power"	"1"
## [253,]	"preferences"	"2"
## [254,]	"preferred"	"1"
## [255,]	"pressure"	"1"
## [256,]	"prices"	"3"
## [257,]	"product"	"7"
## [258,]	"products"	"4"
## [259,]	"professionals"	"1"
## [260,]	"profit"	"1"
## [261,]	"promise"	"1"
## [262,]	"provide"	"1"
## [263,]	"provides"	"1"
## [264,]	"purchase"	"1"
## [265,]	"push"	"1"
## [266,]	"pushing"	"1"
## [267,]	"quaker"	"1"
## [268,]	"quintessential"	"1"
## [269,]	"range"	"3"
## [270,]	"raw"	"1"

## [271,]	"reach"	"3"
## [272,]	"ready"	"1"
## [273,]	"readyto"	"1"
## [274,]	"recent"	"1"
## [275,]	"regional"	"4"
## [276,]	"reinforces"	"1"
## [277,]	"relatively"	"3"
## [278,]	"requires"	"1"
## [279,]	"rice"	"1"
## [280,]	"robust"	"1"
## [281,]	"role"	"1"
## [282,]	"roughly"	"1"
## [283,]	"scope"	"1"
## [284,]	"segment"	"2"
## [285,]	"segments"	"1"
## [286,]	"share"	"2"
## [287,]	"small"	"1"
## [288,]	"south"	"1"
## [289,]	"spending"	"2"
## [290,]	"stage"	"1"
## [291,]	"staples"	"1"
## [292,]	"start"	"1"
## [293,]	"stores"	"1"
## [294,]	"story"	"1"
## [295,]	"strategy"	"1"
## [296,]	"stride"	"1"
## [297,]	"strong"	"1"
## [298,]	"success"	"1"
## [299,]	"suit"	"4"
## [300,]	"surge"	"1"
## [301,]	"switch"	"2"
## [302,]	"tandem"	"1"
## [303,]	"tantalize"	"1"
## [304,]	"targeting"	"1"
## [305,]	"taste"	"1"
## [306,]	"tasteful"	"1"
## [307,]	"tastes"	"1"
## [308,]	"tempted"	"1"
## [309,]	"termed"	"1"
## [310,]	"terms"	"2"
## [311,]	"though"	"1"
## [312,]	"thus"	"1"
## [313,]	"time"	"1"
## [314,]	"toconsumer"	"1"
## [315,]	"towards"	"1"
## [316,]	"traditional"	"1"
## [317,]	"traditionally"	"1"
## [318,]	"travelling"	"1"
## [319,]	"trend"	"1"
## [320,]	"trends"	"1"

```
## [321,] "trials"      "1"
## [322,] "turn"       "1"
## [323,] "understanding" "1"
## [324,] "upmas"      "1"
## [325,] "upward"     "1"
## [326,] "urban"      "1"
## [327,] "urbanization" "1"
## [328,] "usd"        "1"
## [329,] "used"       "1"
## [330,] "using"      "1"
## [331,] "varied"     "1"
## [332,] "variety"    "1"
## [333,] "various"    "1"
## [334,] "varying"    "1"
## [335,] "via"        "1"
## [336,] "vis"        "2"
## [337,] "well"       "3"
## [338,] "western"    "2"
## [339,] "wheat"      "2"
## [340,] "whole"      "1"
## [341,] "wholesome"  "1"
## [342,] "wholly"     "1"
## [343,] "widening"   "1"
## [344,] "will"       "5"
## [345,] "within"     "1"
## [346,] "witnessed"  "1"
## [347,] "year"       "2"
## [348,] "years"      "2"
## [349,] "yet"        "1"
```

```
head(word_freq)
```

```
## Text_split
##      à      abounding acceptability      acceptance      add
##      1      1      1      1      1
## additionally
##      1
```

```
View(word_freq)
```

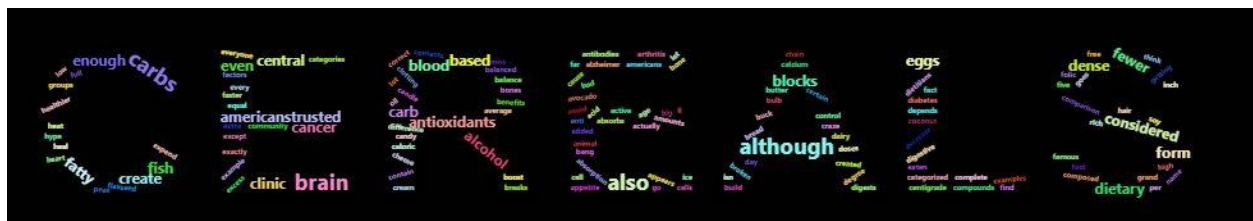
```
write.csv(word_freq, "wordfrequency.csv")
```

```
# Step 8 - Generate the word cloud
```

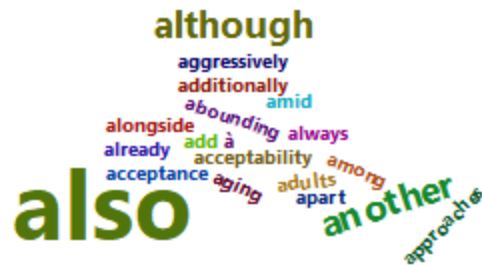
```
df <- data.frame(names(word_freq), as.integer(word_freq))
figPath <- "C:/Users/HP/Desktop/dh.png"
wordcloud2(df, figPath = figPath, size = 1.5, color = "green",
            backgroundColor = "white")
```



```
letterCloud(df, word = "CEREALS", color='random-light' ,
            backgroundColor="black")
```



```
df <- data.frame(names(word_freq), as.integer(word_freq))
wordcloud2(df, size = 0.9, shape = 'triangle')
```



7

```
# Step 9 - Time to organize the words as per wordcloud.  
# The class of the data needs to be characters.  
# we can achieve this by unlisting the list of words.
```

```
class(Text_split)  
## [1] "list"  
  
word_cloud1 = unlist(Text_split)  
View(word_cloud1)
```

```
# Step 11 - Add rainbow colors to the words. No. of colors in paranthesis  
wordcloud2(word_freq, color = "random-light", backgroundColor = "Black")
```

