Practicum 3

Smit Patil

—Problem 2—

1. Download the data set Plant Disease Data Set. Note that the data file does not contain header names; you may wish to add those. The description of each column can be found in the data set explanation. This assignment must be completed within a separate R Markdown Notebook. Use read.transaction() from the arules package to read the data.

```r
#Importing libraries
library(arules)

#Importing plant data
plant.data <- read.transactions("plants.data")

#Summarising plant data
summary(plant.data)
```

```
## transactions as itemMatrix in sparse format with
##  34781 rows (elements/itemsets/transactions) and
##  35463 columns (items) and a density of 6.717562e-05
##
## most frequent items:
##       var.       ssp.      carex astragalus  eriogonum    (Other)
##       5359       3069        694        615        436      72684
##
## element (itemset/transaction) length distribution:
## sizes
##     1     2     3     4     5     6
##  3383 23250     1  7865   181   101
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   2.000   2.382   2.000   6.000
##
## includes extended item information - examples:
##                   labels
## 1 ×abbeae,mi,mn,wi,on,sk
## 2          ×abbottiae,hi
## 3         ×abitibiana,qc
```

2. Explore the data set as you see fit and that allows you to get a sense of the data and get comfortable with it. Is there distributional skew in any of the features? Is there a need to apply a transform?

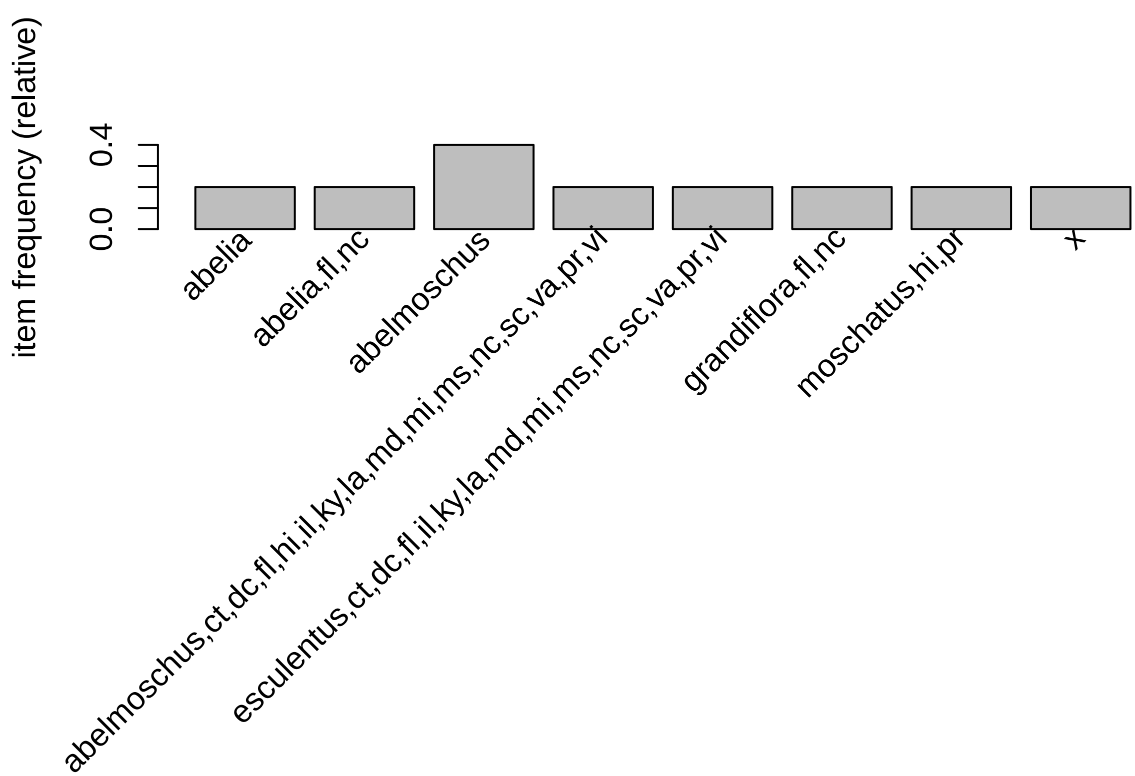
```r
#Inspecting first 5 rows of the plant data
inspect(plant.data[1:5])
```

```
##     items
## [1] {abelia,fl,nc}
## [2] {abelia,
##      grandiflora,fl,nc,
##      x}
## [3] {abelmoschus,ct,dc,fl,hi,il,ky,la,md,mi,ms,nc,sc,va,pr,vi}
## [4] {abelmoschus,
##      esculentus,ct,dc,fl,il,ky,la,md,mi,ms,nc,sc,va,pr,vi}
## [5] {abelmoschus,
##      moschatus,hi,pr}
```

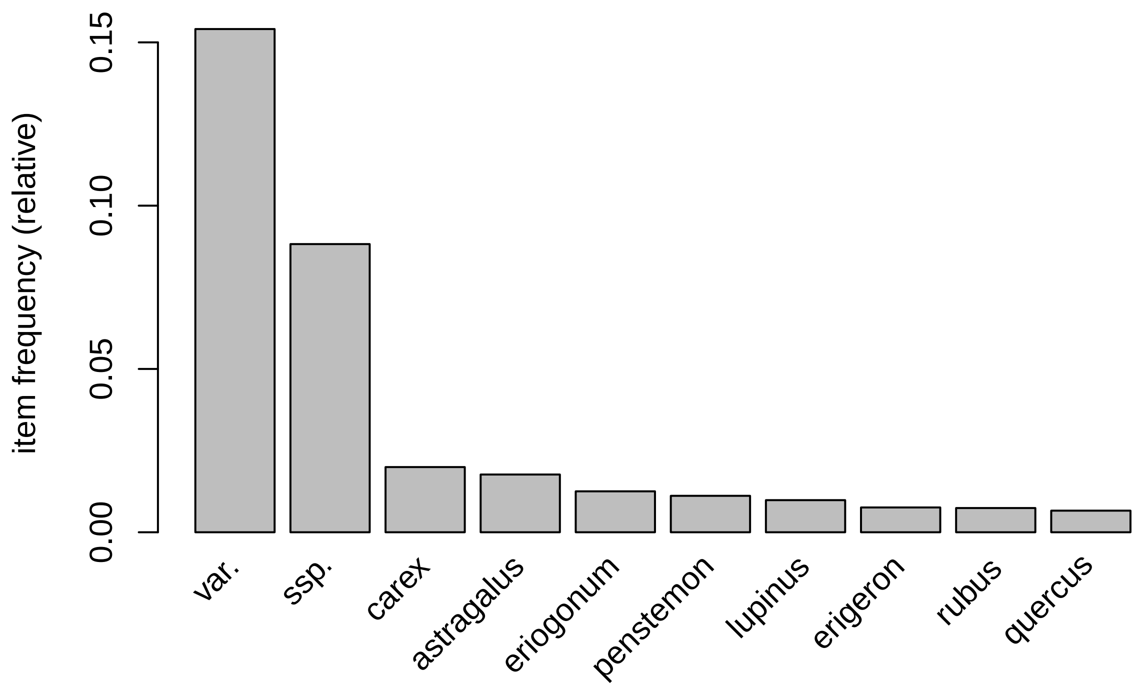
```r
#Calculating frequency for the first 5 rows
itemFrequency(plant.data[,1:5])
```

```
## ×abbeae,mi,mn,wi,on,sk          ×abbottiae,hi         ×abitibiana,qc
##           2.875133e-05           2.875133e-05           2.875133e-05
##         ×acadiensis,ns              ×achnella
##           2.875133e-05           2.875133e-05
```

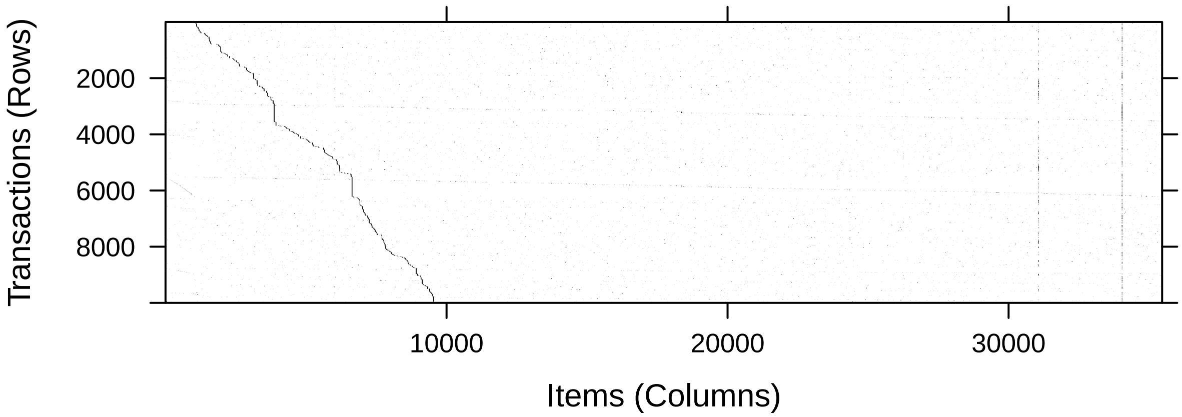
```r
#Plotting the ietms with atleast 10% support for first 5 species
itemFrequencyPlot(plant.data[1:5], support = 0.1)
```



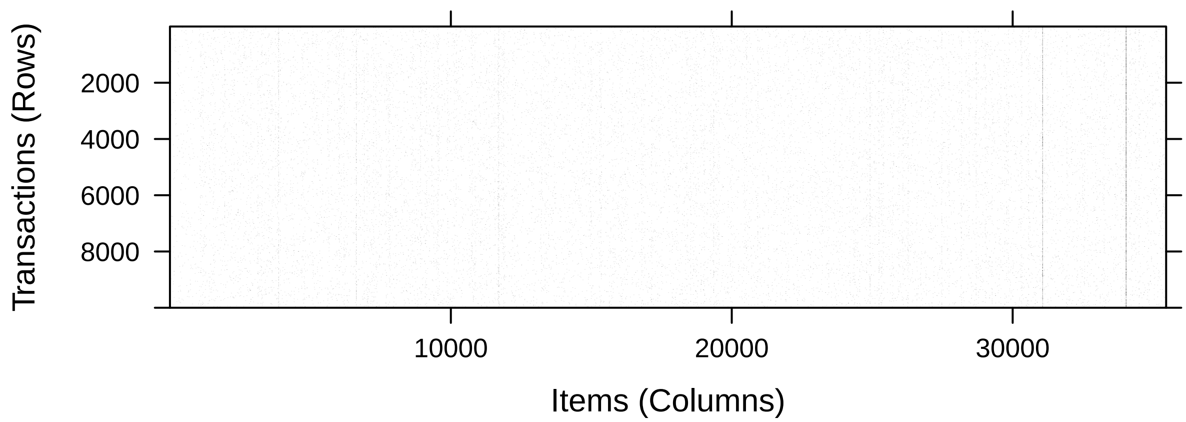
```r
#Plotting the top 10 species
itemFrequencyPlot(plant.data, topN = 10)
```




```r
#Creating a sparse matrix for the first 10k species
image(plant.data[1:10000])
```




```
#Creating a sparse matrix for a random sample of 10k species
image(sample(plant.data, 10000))
```



3. Use association rules to segment the data similar to what was done in Hämäläinen, W., & Nykänen, M. (2008, December). Efficient discovery of statistically significant association rules. In Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on (pp. 203-212). IEEE.

```r
#Calculating rules using the default values of the apriori function
apriori(plant.data)
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.8    0.1    1 none FALSE            TRUE       5     0.1      1
##  maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 3478
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[35463 item(s), 34781 transaction(s)] done [0.07s].
## sorting and recoding items ... [1 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
```

```
## checking subsets of size 1 done [0.00s].
## writing ... [0 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].

## set of 0 rules
```

```r
#Calculating rules using custom values
plant.rules <- apriori(plant.data, parameter = list(support = 0.0005, confidence = 0.9))
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.9    0.1    1 none FALSE            TRUE       5   5e-04      1
##  maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 17
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[35463 item(s), 34781 transaction(s)] done [0.05s].
## sorting and recoding items ... [456 item(s)] done [0.00s].
## creating transaction tree ... done [0.01s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [30 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```r
#summary of plant rules
summary(plant.rules)
```

```
## set of 30 rules
##
## rule length distribution (lhs + rhs):sizes
##  2  3  4
## 12 14  4
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.000   2.000   3.000   2.733   3.000   4.000
##
## summary of quality measures:
##     support            confidence        coverage              lift
##  Min.   :0.0005175   Min.   :0.9167   Min.   :0.0005175   Min.   :   5.949
##  1st Qu.:0.0006325   1st Qu.:1.0000   1st Qu.:0.0006325   1st Qu.:   6.490
##  Median :0.0006325   Median :1.0000   Median :0.0006900   Median :  11.333
##  Mean   :0.0007082   Mean   :0.9838   Mean   :0.0007197   Mean   : 207.709
##  3rd Qu.:0.0008410   3rd Qu.:1.0000   3rd Qu.:0.0008554   3rd Qu.: 363.177
##  Max.   :0.0010063   Max.   :1.0000   Max.   :0.0010063   Max.   :1449.208
##      count
```

```
##  Min.   :18.00
##  1st Qu.:22.00
##  Median :22.00
##  Mean   :24.63
##  3rd Qu.:29.25
##  Max.   :35.00
##
## mining info:
##        data ntransactions support confidence
##  plant.data         34781   5e-04        0.9
```

```r
#Inspecting the first 5 rules of the plant rules
inspect(plant.rules[1:5])
```

```
##     lhs               rhs        support      confidence coverage     lift
## [1] {breweri}     => {var.}     0.0005175239 1.0000000  0.0005175239   6.490203
## [2] {engelmannii} => {var.}     0.0005175239 1.0000000  0.0005175239   6.490203
## [3] {malviflora}  => {sidalcea} 0.0005462753 1.0000000  0.0005462753 543.453125
## [4] {malviflora}  => {ssp.}     0.0005462753 1.0000000  0.0005462753  11.333007
## [5] {canadensis}  => {var.}     0.0007762859 0.9310345  0.0008337886   6.042603
##     count
## [1] 18
## [2] 18
## [3] 19
## [4] 19
## [5] 27
```

```r
#Combining the sort function with the inspect function to find the best five rules according to the lif
inspect(sort(plant.rules, by = "lift")[1:5])
```

```
##     lhs                         rhs          support      confidence coverage
## [1] {ericameria,ssp.}      => {nauseosa}   0.0006900319 1          0.0006900319
## [2] {ericameria,ssp.,var.} => {nauseosa}   0.0006325293 1          0.0006325293
## [3] {malviflora}           => {sidalcea}   0.0005462753 1          0.0005462753
## [4] {malviflora,ssp.}      => {sidalcea}   0.0005462753 1          0.0005462753
## [5] {nauseosa}             => {ericameria} 0.0006900319 1          0.0006900319
##     lift      count
## [1] 1449.2083 24
## [2] 1449.2083 22
## [3]  543.4531 19
## [4]  543.4531 19
## [5]  457.6447 24
```

4. Are there clusters in the data? Can plants be segmented into groups? Build a k-means clustering model to investigate.

```r
library(klaR)
library(reshape2)

#Converting plant data to vector
plant.vector <- as.vector(unlist(plant.data@itemInfo))
```

```r
#Saving the vector file as a .csv file
write(x = plant.vector, "plant.vector.csv", sep = ",")

#Reading the plant vector csv file as a data frame
plant.vector.data <- read.csv("plant.vector.csv", header = F, col.names = 1:100)

#Removing empty columns from the data
plant.vector.data <- plant.vector.data[,which(!is.na(plant.vector.data[1,]))]

#Looking at the head and the structure of the data
head(plant.vector.data)
```

```
##             X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12 X13 X14 X15 X16 X17 X18 X19
## 1     ×abbeae mi mn wi on sk
## 2  ×abbottiae hi
## 3 ×abitibiana qc
## 4 ×acadiensis ns
## 5   ×achnella
## 6   ×achnella mt ne nd ab bc mb nt sk  yt
##   X20 X21 X22 X23 X24 X25 X26 X27 X28 X29 X30 X31 X32 X33 X34 X35 X36 X37 X38
## 1
## 2
## 3
## 4
## 5
## 6
##   X39 X40 X41 X42 X43 X44 X45 X46 X47 X48 X49 X50 X51 X52 X53 X54 X55 X56 X57
## 1
## 2
## 3
## 4
## 5
## 6
##   X58 X59 X60 X61 X62 X63 X64 X65 X66 X67 X68 X69 X70
## 1
## 2
## 3
## 4
## 5
## 6
```

```r
str(plant.vector.data)
```

```
## 'data.frame':    35463 obs. of  70 variables:
##  $ X1 : chr  "×abbeae" "×abbottiae" "×abitibiana" "×acadiensis" ...
##  $ X2 : chr  "mi" "hi" "qc" "ns" ...
##  $ X3 : chr  "mn" "" "" "" ...
##  $ X4 : chr  "wi" "" "" "" ...
##  $ X5 : chr  "on" "" "" "" ...
##  $ X6 : chr  "sk" "" "" "" ...
##  $ X7 : chr  "" "" "" "" ...
##  $ X8 : chr  "" "" "" "" ...
```

```
##  $ X9 : chr  "" "" "" "" ...
##  $ X10: chr  "" "" "" "" ...
##  $ X11: chr  "" "" "" "" ...
##  $ X12: chr  "" "" "" "" ...
##  $ X13: chr  "" "" "" "" ...
##  $ X14: chr  "" "" "" "" ...
##  $ X15: chr  "" "" "" "" ...
##  $ X16: chr  "" "" "" "" ...
##  $ X17: chr  "" "" "" "" ...
##  $ X18: chr  "" "" "" "" ...
##  $ X19: chr  "" "" "" "" ...
##  $ X20: chr  "" "" "" "" ...
##  $ X21: chr  "" "" "" "" ...
##  $ X22: chr  "" "" "" "" ...
##  $ X23: chr  "" "" "" "" ...
##  $ X24: chr  "" "" "" "" ...
##  $ X25: chr  "" "" "" "" ...
##  $ X26: chr  "" "" "" "" ...
##  $ X27: chr  "" "" "" "" ...
##  $ X28: chr  "" "" "" "" ...
##  $ X29: chr  "" "" "" "" ...
##  $ X30: chr  "" "" "" "" ...
##  $ X31: chr  "" "" "" "" ...
##  $ X32: chr  "" "" "" "" ...
##  $ X33: chr  "" "" "" "" ...
##  $ X34: chr  "" "" "" "" ...
##  $ X35: chr  "" "" "" "" ...
##  $ X36: chr  "" "" "" "" ...
##  $ X37: chr  "" "" "" "" ...
##  $ X38: chr  "" "" "" "" ...
##  $ X39: chr  "" "" "" "" ...
##  $ X40: chr  "" "" "" "" ...
##  $ X41: chr  "" "" "" "" ...
##  $ X42: chr  "" "" "" "" ...
##  $ X43: chr  "" "" "" "" ...
##  $ X44: chr  "" "" "" "" ...
##  $ X45: chr  "" "" "" "" ...
##  $ X46: chr  "" "" "" "" ...
##  $ X47: chr  "" "" "" "" ...
##  $ X48: chr  "" "" "" "" ...
##  $ X49: chr  "" "" "" "" ...
##  $ X50: chr  "" "" "" "" ...
##  $ X51: chr  "" "" "" "" ...
##  $ X52: chr  "" "" "" "" ...
##  $ X53: chr  "" "" "" "" ...
##  $ X54: chr  "" "" "" "" ...
##  $ X55: chr  "" "" "" "" ...
##  $ X56: chr  "" "" "" "" ...
##  $ X57: chr  "" "" "" "" ...
##  $ X58: chr  "" "" "" "" ...
##  $ X59: chr  "" "" "" "" ...
##  $ X60: chr  "" "" "" "" ...
##  $ X61: chr  "" "" "" "" ...
##  $ X62: chr  "" "" "" "" ...
```

```
##  $ X63: chr  "" "" "" "" ...
##  $ X64: chr  "" "" "" "" ...
##  $ X65: chr  "" "" "" "" ...
##  $ X66: chr  "" "" "" "" ...
##  $ X67: chr  "" "" "" "" ...
##  $ X68: chr  "" "" "" "" ...
##  $ X69: chr  "" "" "" "" ...
##  $ X70: chr  "" "" "" "" ...
```

```r
#Recasting the plant data
plant.new <- recast(plant.vector.data, X1 ~ value, id.var = "X1")

#Summary of the new plant data
summary(plant.new)
```

```
##       X1                 Var.2              ab                ak
##  Length:15078       Min.   :   2.0   Min.   : 0.000   Min.   : 0.0000
##  Class :character   1st Qu.:  67.0   1st Qu.: 0.000   1st Qu.: 0.0000
##  Mode  :character   Median :  68.0   Median : 0.000   Median : 0.0000
##                     Mean   : 143.6   Mean   : 0.212   Mean   : 0.1821
##                     3rd Qu.: 136.0   3rd Qu.: 0.000   3rd Qu.: 0.0000
##                     Max.   :4390.0   Max.   :22.000   Max.   :18.0000
##        al               ar                az                bc
##  Min.   : 0.000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.361   Mean   : 0.2903   Mean   : 0.4002   Mean   : 0.3016
##  3rd Qu.: 0.000   3rd Qu.: 0.0000   3rd Qu.: 1.0000   3rd Qu.: 0.0000
##  Max.   :32.000   Max.   :30.0000   Max.   :23.0000   Max.   :30.0000
##        ca                co                ct                dc
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.6297   Mean   : 0.3376   Mean   : 0.2786   Mean   : 0.1933
##  3rd Qu.: 1.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :43.0000   Max.   :28.0000   Max.   :47.0000   Max.   :30.0000
##        de               dengl               fl                fraspm
##  Min.   : 0.0000   Min.   :0.00000   Min.   : 0.0000   Min.   : 0.00000
##  1st Qu.: 0.0000   1st Qu.:0.00000   1st Qu.: 0.0000   1st Qu.: 0.00000
##  Median : 0.0000   Median :0.00000   Median : 0.0000   Median : 0.00000
##  Mean   : 0.2283   Mean   :0.02965   Mean   : 0.4134   Mean   : 0.07587
##  3rd Qu.: 0.0000   3rd Qu.:0.00000   3rd Qu.: 1.0000   3rd Qu.: 0.00000
##  Max.   :32.0000   Max.   :5.00000   Max.   :27.0000   Max.   :15.00000
##        ga               gl                 hi                ia
##  Min.   : 0.000   Min.   : 0.00000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.000   1st Qu.: 0.00000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.000   Median : 0.00000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.376   Mean   : 0.03442   Mean   : 0.2099   Mean   : 0.2294
##  3rd Qu.: 0.000   3rd Qu.: 0.00000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :38.000   Max.   :10.00000   Max.   :10.0000   Max.   :37.0000
##        id               il                in                ks
##  Min.   : 0.000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.000   Median : 0.0000   Median : 0.0000   Median : 0.0000
```

```
##  Mean   : 0.316   Mean   : 0.3263   Mean   : 0.2799   Mean   : 0.2414
##  3rd Qu.: 0.000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :38.000   Max.   :42.0000   Max.   :44.0000   Max.   :28.0000
##        ky                la                lb                 ma
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.00000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.00000   1st Qu.: 0.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.00000   Median : 0.0000
##  Mean   : 0.2882   Mean   : 0.3258   Mean   : 0.08894   Mean   : 0.3151
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.00000   3rd Qu.: 0.0000
##  Max.   :40.0000   Max.   :24.0000   Max.   :13.00000   Max.   :43.0000
##        mb                md                me                mi
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.1897   Mean   : 0.3232   Mean   : 0.2514   Mean   : 0.2987
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :28.0000   Max.   :43.0000   Max.   :39.0000   Max.   :44.0000
##        mn                mo                ms                mt
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.2482   Mean   : 0.2925   Mean   : 0.3041   Mean   : 0.2984
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :39.0000   Max.   :36.0000   Max.   :27.0000   Max.   :33.0000
##        nb                nc                nd                ne
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.1799   Mean   : 0.3761   Mean   : 0.1679   Mean   : 0.2053
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :37.0000   Max.   :42.0000   Max.   :26.0000   Max.   :27.0000
##        nf                nh                nj                nm
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.1372   Mean   : 0.2297   Mean   : 0.3061   Mean   : 0.3827
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :21.0000   Max.   :41.0000   Max.   :43.0000   Max.   :24.0000
##        ns                nt                nu                 nv
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.00000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.00000   1st Qu.: 0.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.00000   Median : 0.0000
##  Mean   : 0.1792   Mean   : 0.1252   Mean   : 0.06022   Mean   : 0.3381
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.00000   3rd Qu.: 0.0000
##  Max.   :32.0000   Max.   :13.0000   Max.   :10.00000   Max.   :30.0000
##        ny               oh               ok                on
##  Min.   : 0.000   Min.   : 0.000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.000   Median : 0.000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.367   Mean   : 0.302   Mean   : 0.2901   Mean   : 0.3201
##  3rd Qu.: 0.000   3rd Qu.: 0.000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :49.000   Max.   :46.000   Max.   :25.0000   Max.   :46.0000
##        or                pa                pe                pr
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
```

```
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.4203   Mean   : 0.3475   Mean   : 0.1159   Mean   : 0.2845
##  3rd Qu.: 1.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :45.0000   Max.   :48.0000   Max.   :25.0000   Max.   :14.0000
##        qc                ri                sc                sd
##  Min.   : 0.000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.269   Mean   : 0.2078   Mean   : 0.3438   Mean   : 0.1988
##  3rd Qu.: 0.000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :44.000   Max.   :37.0000   Max.   :34.0000   Max.   :27.0000
##        sk                tn                tx                ut
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.1788   Mean   : 0.3095   Mean   : 0.5085   Mean   : 0.3675
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 1.0000   3rd Qu.: 0.0000
##  Max.   :19.0000   Max.   :43.0000   Max.   :31.0000   Max.   :32.0000
##        va                vi              vt                wa
##  Min.   : 0.0000   Min.   :0.000   Min.   : 0.000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.:0.000   1st Qu.: 0.000   1st Qu.: 0.0000
##  Median : 0.0000   Median :0.000   Median : 0.000   Median : 0.0000
##  Mean   : 0.3575   Mean   :0.135   Mean   : 0.235   Mean   : 0.3481
##  3rd Qu.: 0.0000   3rd Qu.:0.000   3rd Qu.: 0.000   3rd Qu.: 0.0000
##  Max.   :46.0000   Max.   :8.000   Max.   :42.000   Max.   :44.0000
##        wi                wv                wy                yt
##  Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.0000
##  1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000
##  Median : 0.0000   Median : 0.0000   Median : 0.0000   Median : 0.0000
##  Mean   : 0.2728   Mean   : 0.2567   Mean   : 0.2917   Mean   : 0.1295
##  3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.0000
##  Max.   :44.0000   Max.   :45.0000   Max.   :27.0000   Max.   :13.0000
```

```r
#Creating five clusters for plant data
plant.cluster <- kmeans(plant.new[,-1], 5)

#Printing the plant cluster sizes
plant.cluster$size
```

```
## [1]   72 5315 8555  880  256
```

```r
#Looking at the plant clusters along with their centers
plant.cluster$centers
```

```
##       Var.2        ab        ak        al        ar        az        bc
## 1 2017.3472 4.7361111 3.8888889 8.0277778 6.5277778 8.7500000 6.6111111
## 2  152.7091 0.1403575 0.1266228 0.2600188 0.1954845 0.3298213 0.2101599
## 3   66.3723 0.1149036 0.1042665 0.1893629 0.1664524 0.1830508 0.1576856
## 4  437.8830 0.7193182 0.5875000 1.1840909 0.9068182 1.3284091 0.9931818
## 5  994.9062 1.9296875 1.5000000 3.2070312 2.5234375 3.5781250 2.8554688
##           ca        co        ct        dc        de      dengl        fl
## 1 11.0277778 7.8194444 5.8055556 3.6111111 4.7222222 0.63888889 8.2638889
```

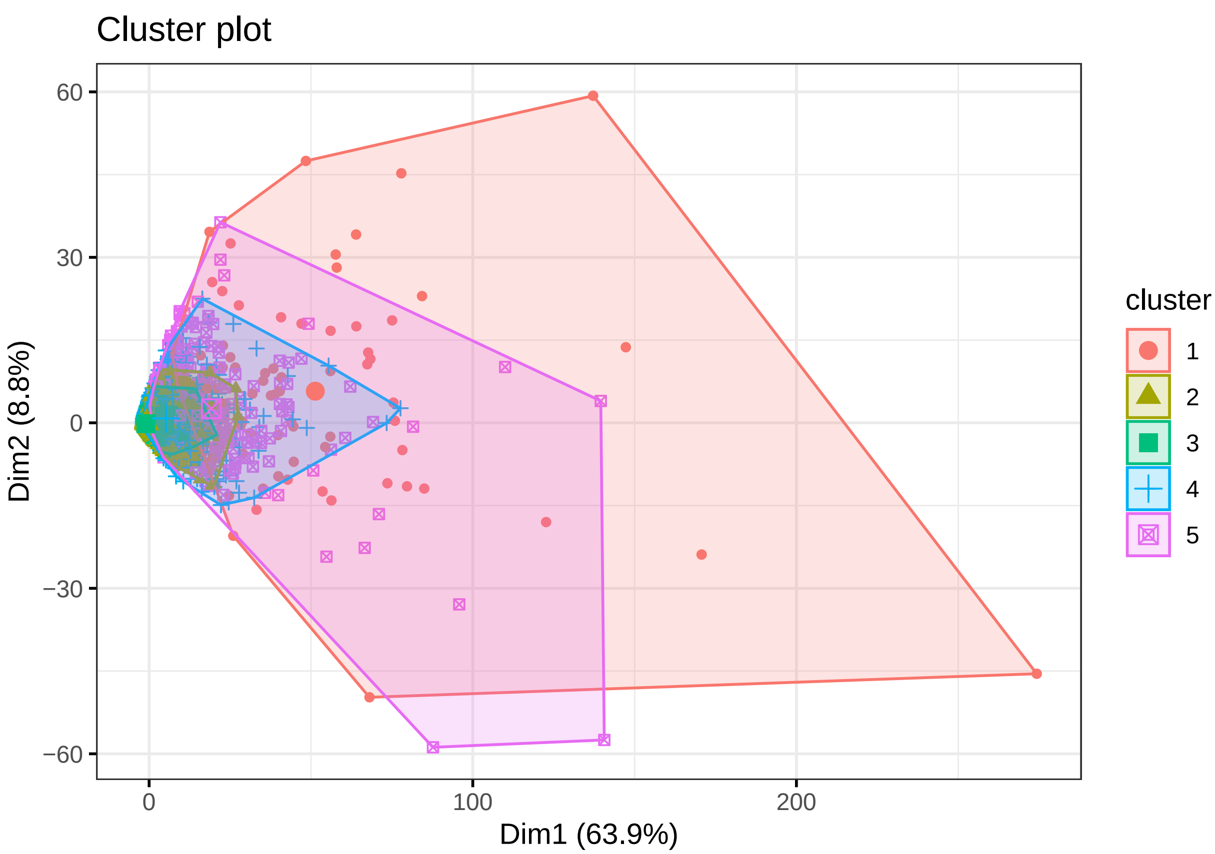
```
## 2  0.5584196 0.2415804 0.1659454 0.1123236 0.1328316 0.01429915 0.3555974
## 3  0.3098773 0.1601403 0.1873758 0.1329047 0.1497370 0.02291058 0.2046756
## 4  2.0113636 1.1568182 0.7875000 0.5534091 0.6920455 0.08636364 1.3113636
## 5  5.1250000 3.3437500 2.3593750 1.6914062 1.9804688 0.20703125 3.2968750
##       fraspm        ga         gl        hi        ia       id        il
## 1 1.70833333 8.2638889 0.75000000 2.1944444 5.2638889 6.902778 7.2222222
## 2 0.04252117 0.2692380 0.02728128 0.2015052 0.1243650 0.225588 0.2048918
## 3 0.05341905 0.1931034 0.01542957 0.1380479 0.1539451 0.155114 0.2029223
## 4 0.20909091 1.2840909 0.11250000 0.5386364 0.6818182 1.060227 0.9943182
## 5 0.60156250 3.3632812 0.34765625 1.0976562 1.9609375 3.160156 2.7343750
##          in        ks        ky        la         lb        ma        mb
## 1 5.9583333 5.5694444 5.9305556 7.2916667 2.15277778 6.3194444 4.6388889
## 2 0.1666980 0.1527752 0.1834431 0.2466604 0.05832549 0.1998119 0.1158984
## 3 0.1829340 0.1402688 0.1791935 0.1706604 0.05599065 0.2051432 0.1175921
## 4 0.8306818 0.7659091 0.8613636 1.0170455 0.25795455 0.9147727 0.5545455
## 5 2.3828125 2.1601562 2.5468750 2.8203125 0.66406250 2.6328125 1.6289062
##          md        me        mi        mn        mo        ms        mt
## 1 7.0416667 5.4166667 6.4722222 5.6388889 6.6805556 6.9027778 6.6250000
## 2 0.2101599 0.1465663 0.1749765 0.1429915 0.1870179 0.2142992 0.2062088
## 3 0.1945061 0.1703098 0.2007013 0.1628288 0.1795441 0.1594389 0.1495032
## 4 0.9613636 0.6977273 0.8590909 0.7363636 0.8647727 0.9988636 1.0079545
## 5 2.8867188 2.1523438 2.4804688 2.0898438 2.4921875 2.7578125 2.9726562
##          nb        nc        nd        ne         nf        nh        nj
## 1 3.9305556 8.3194444 4.0000000 4.8055556 2.90277778 5.1250000 6.3611111
## 2 0.1000941 0.2626529 0.0957667 0.1222954 0.08447789 0.1275635 0.1900282
## 3 0.1281122 0.2093513 0.1077732 0.1250731 0.09316189 0.1565167 0.1969608
## 4 0.4954545 1.1852273 0.5011364 0.6181818 0.35795455 0.6522727 0.9125000
## 5 1.4257812 3.2890625 1.4492188 1.8945312 1.16406250 1.9648438 2.5742188
##          nm        ns         nt         nu       nv        ny        oh
## 1 8.8888889 3.8611111 2.88888889 1.45833333 7.125000 7.5694444 6.2361111
## 2 0.3010348 0.1017874 0.08523048 0.04402634 0.274318 0.2408278 0.1843838
## 3 0.1756867 0.1227352 0.07165400 0.03366452 0.150789 0.2338983 0.1978960
## 4 1.2863636 0.5125000 0.39659091 0.18522727 1.152273 1.0465909 0.8784091
## 5 3.5000000 1.4921875 1.03515625 0.46093750 3.214844 3.0742188 2.5703125
##          ok        on        or        pa        pe        pr        qc
## 1 6.8333333 7.1111111 8.5555556 7.2222222 2.6527778 3.1388889 5.7083333
## 2 0.2013170 0.1964252 0.3288805 0.2306679 0.0607714 0.3243650 0.1633114
## 3 0.1552309 0.2060783 0.2002338 0.2191701 0.0842782 0.1430742 0.1777908
## 4 0.9443182 0.9193182 1.4159091 0.9943182 0.2943182 0.7454545 0.7738636
## 5 2.5507812 2.7265625 3.9609375 2.9023438 0.9882812 1.7968750 2.2460938
##          ri        sc        sd        sk        tn         tx        ut
## 1 4.4305556 7.6805556 4.6111111 3.8472222 6.8055556 11.4444444 7.8472222
## 2 0.1140169 0.2415804 0.1194732 0.1119473 0.2043274  0.4124177 0.2871119
## 3 0.1436587 0.1830508 0.1200468 0.1124489 0.1828171  0.2437171 0.1729982
## 4 0.6022727 1.1318182 0.6193182 0.5295455 0.9454545  1.6727273 1.2022727
## 5 1.7539062 3.0625000 1.7890625 1.5468750 2.7148438  4.2734375 3.5625000
##          va        vi        vt        wa        wi        wv        wy
## 1 7.4861111 1.6805556 4.9027778 7.5277778 6.0138889 5.5277778 6.7777778
## 2 0.2453434 0.1634995 0.1354657 0.2524929 0.1586077 0.1505174 0.2045155
## 3 0.2082992 0.0603156 0.1602572 0.1742840 0.1855056 0.1668030 0.1416715
## 4 1.0988636 0.3352273 0.6602273 1.1693182 0.7602273 0.7500000 0.9750000
## 5 3.1210938 0.9179688 2.0234375 3.2968750 2.2695312 2.2890625 2.9414062
##           yt
## 1 2.91666667
```

```
## 2 0.08955786
## 3 0.07293980
## 4 0.41477273
## 5 1.08593750
```

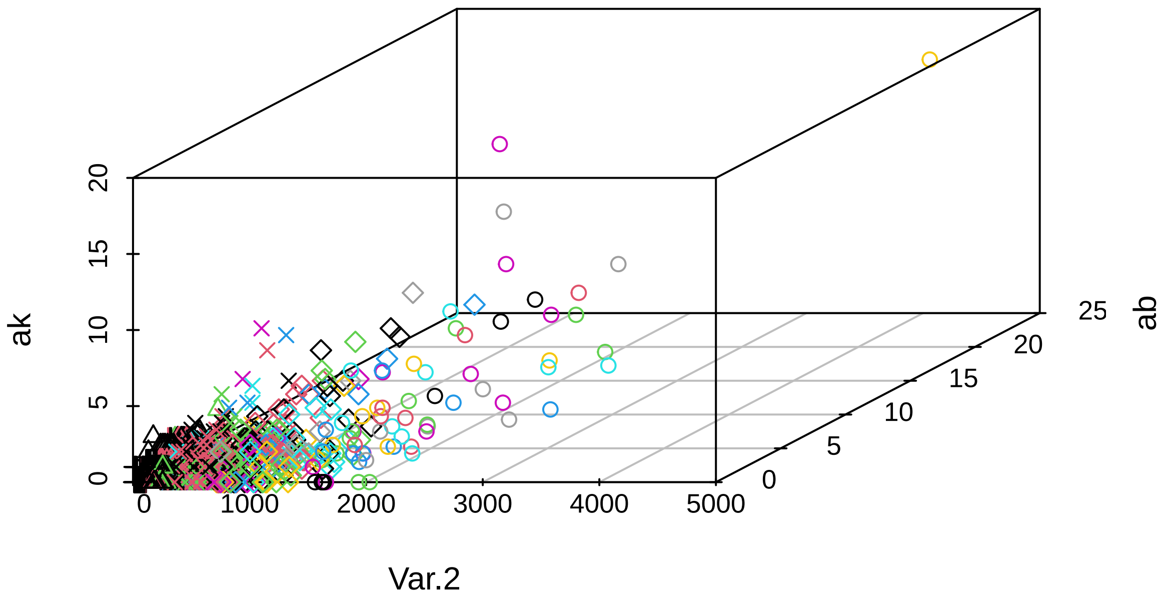
5. Visualize the clusters.

```r
#Importing libraries
library(factoextra)
library(scatterplot3d)

#Plotting 2D graph to represent the clusters
fviz_cluster(plant.cluster, data = plant.new[,-1],
             geom = "point",
             ellipse.type = "convex",
             ggtheme = theme_bw()
             )
```




```r
#Plotting scatter plot for first 3 values in 3D
scatterplot3d(plant.new[,-1], pch = plant.cluster$cluster, color = plant.cluster$cluster)
```


ak
0
5
10
15
20
0
1000
2000
3000
4000
5000
Var.2
0
5
10
15
20
25
ab