Crop-analysis.R

SOWMIYA S

2023-03-21

```
X=file.choose()
```

```
df=read.csv(X)
View(df)

#STRUCTURE OF THE DATA
str(df)
```

```
## 'data.frame':    2200 obs. of  8 variables:
##  $ N          : int  90 85 60 74 78 69 69 94 89 68 ...
##  $ P          : int  42 58 55 35 42 37 55 53 54 58 ...
##  $ K          : int  43 41 44 40 42 42 38 40 38 38 ...
##  $ temperature: num  20.9 21.8 23 26.5 20.1 ...
##  $ humidity   : num  82 80.3 82.3 80.2 81.6 ...
##  $ ph         : num  6.5 7.04 7.84 6.98 7.63 ...
##  $ rainfall   : num  203 227 264 243 263 ...
##  $ label      : chr  "rice" "rice" "rice" "rice" ...
```

```
#Checking the null values in the dataset
```

```
sum(is.na(df)=="TRUE")
```

```
## [1] 0
#SUMMARY OF THE DATA
```

```
summary(df)
```

```
##        N                P                K           temperature
humidity
##  Min.   :  0.00   Min.   :  5.00   Min.   :  5.00   Min.   : 8.826
Min.   :14.26
##  1st Qu.: 21.00   1st Qu.: 28.00   1st Qu.: 20.00   1st Qu.:22.769
1st Qu.:60.26
##  Median : 37.00   Median : 51.00   Median : 32.00   Median :25.599
Median :80.47
##  Mean   : 50.55   Mean   : 53.36   Mean   : 48.15   Mean   :25.616
Mean   :71.48
```

```
##  3rd Qu.: 84.25   3rd Qu.: 68.00   3rd Qu.: 49.00   3rd Qu.:28.562
3rd Qu.:89.95
##  Max.   :140.00   Max.   :145.00   Max.   :205.00   Max.   :43.675
Max.   :99.98
##        ph           rainfall         label
##  Min.   :3.505   Min.   : 20.21   Length:2200
##  1st Qu.:5.972   1st Qu.: 64.55   Class :character
##  Median :6.425   Median : 94.87   Mode  :character
##  Mean   :6.469   Mean   :103.46
##  3rd Qu.:6.924   3rd Qu.:124.27
##  Max.   :9.935   Max.   :298.56
```

```
#View data
```

head(df)

```
##    N  P  K temperature humidity       ph rainfall label
## 1 90 42 43    20.87974 82.00274 6.502985 202.9355  rice
## 2 85 58 41    21.77046 80.31964 7.038096 226.6555  rice
## 3 60 55 44    23.00446 82.32076 7.840207 263.9642  rice
## 4 74 35 40    26.49110 80.15836 6.980401 242.8640  rice
## 5 78 42 42    20.13017 81.60487 7.628473 262.7173  rice
## 6 69 37 42    23.05805 83.37012 7.073454 251.0550  rice
```

```
tail(df)
##        N  P  K temperature humidity       ph rainfall  label
## 2195  97 35 26    24.91461 53.74145 6.334610 166.2549 coffee
## 2196 107 34 32    26.77464 66.41327 6.780064 177.7745 coffee
## 2197  99 15 27    27.41711 56.63636 6.086922 127.9246 coffee
## 2198 118 33 30    24.13180 67.22512 6.362608 173.3228 coffee
## 2199 117 32 34    26.27242 52.12739 6.758793 127.1753 coffee
## 2200 104 18 30    23.60302 60.39647 6.779833 140.9370 coffee
```

```
dim(df)
## [1] 2200    8
names(df)
## [1] "N"           "P"           "K"           "temperature"
"humidity"    "ph"
```

```
## [7] "rainfall"    "label"
```

```
length(df)
## [1] 8
#Subsetting the dataset
```

library(dplyr)

```
#Arrange  function
df1=arrange(df,ph)
head(df1)
```

```
##    N  P  K temperature humidity       ph rainfall     label
## 1  4 46 15    31.01275 62.40393 3.504752 63.77192 mothbeans
## 2 36 43 24    27.09401 43.65305 3.510404 41.53750 mothbeans
## 3 19 51 25    26.80474 48.23991 3.525366 43.87802 mothbeans
## 4 16 51 21    31.01964 49.97675 3.532009 32.81297 mothbeans
## 5 35 51 17    28.79929 49.84213 3.558823 40.85535 mothbeans
## 6  3 49 18    27.91095 64.70931 3.692864 32.67892 mothbeans
```

```
df2=arrange(df,P,desc(K))
```

```
head(df2)
```

```
##     N P  K temperature humidity       ph rainfall      label
## 1 111 5 55    26.28344 84.42479 6.520663 50.78670 watermelon
## 2 119 5 55    29.68847 94.30112 6.168758 26.83925  muskmelon
## 3  94 5 55    28.58546 91.89217 6.085682 26.88373  muskmelon
## 4 116 5 54    25.37601 80.99314 6.653987 57.23028 watermelon
## 5 107 5 52    26.66346 89.98405 6.881426 57.40847 watermelon
## 6 111 5 52    29.88431 94.03711 6.135996 21.00010  muskmelon
```

```
df3=arrange(df,desc(rainfall))
```

```
head(df3)
```

```
##    N  P  K temperature humidity       ph rainfall label
## 1 82 40 40    23.83067 84.81360 6.271479 298.5601  rice
## 2 60 43 44    21.01945 82.95222 7.416245 298.4018  rice
## 3 93 47 37    21.53346 82.14004 6.500343 295.9249  rice
## 4 79 42 37    24.87301 82.84023 6.587919 295.6094  rice
## 5 88 35 40    23.57944 83.58760 5.853932 291.2987  rice
## 6 93 53 38    26.92995 81.91411 7.069172 290.6794  rice
```

```
#Select function
```

```
df4=select(df,N,P,K)
head(df4)
```

```
##    N  P  K
## 1 90 42 43
## 2 85 58 41
```

```
## 3 60 55 44
## 4 74 35 40
## 5 78 42 42
## 6 69 37 42
```

```
df5=select(df,N:ph)
```

head(df5)

```
##    N  P  K temperature humidity       ph
## 1 90 42 43    20.87974 82.00274 6.502985
## 2 85 58 41    21.77046 80.31964 7.038096
## 3 60 55 44    23.00446 82.32076 7.840207
## 4 74 35 40    26.49110 80.15836 6.980401
## 5 78 42 42    20.13017 81.60487 7.628473
## 6 69 37 42    23.05805 83.37012 7.073454
```

```
df6=select(df,-humidity)
```

head(df6)

```
##    N  P  K temperature       ph rainfall label
## 1 90 42 43    20.87974 6.502985 202.9355  rice
## 2 85 58 41    21.77046 7.038096 226.6555  rice
## 3 60 55 44    23.00446 7.840207 263.9642  rice
## 4 74 35 40    26.49110 6.980401 242.8640  rice
## 5 78 42 42    20.13017 7.628473 262.7173  rice
## 6 69 37 42    23.05805 7.073454 251.0550  rice
```

```
df7=select(df,-(ph:label))
```

head(df7)

```
##    N  P  K temperature humidity
## 1 90 42 43    20.87974 82.00274
## 2 85 58 41    21.77046 80.31964
## 3 60 55 44    23.00446 82.32076
## 4 74 35 40    26.49110 80.15836
## 5 78 42 42    20.13017 81.60487
## 6 69 37 42    23.05805 83.37012
```

```
#Filter function
```

df8=filter(df,label=="rice")
head(df8)

```
##    N  P  K temperature humidity       ph rainfall label
```

```
## 1 90 42 43    20.87974 82.00274 6.502985 202.9355  rice
## 2 85 58 41    21.77046 80.31964 7.038096 226.6555  rice
## 3 60 55 44    23.00446 82.32076 7.840207 263.9642  rice
## 4 74 35 40    26.49110 80.15836 6.980401 242.8640  rice
## 5 78 42 42    20.13017 81.60487 7.628473 262.7173  rice
## 6 69 37 42    23.05805 83.37012 7.073454 251.0550  rice
```

```
df9=filter(df,label=="rice",K==42)
```

head(df9)

```
##    N  P  K temperature humidity       ph rainfall label
```

```
## 1 78 42 42    20.13017 81.60487 7.628473 262.7173  rice
## 2 69 37 42    23.05805 83.37012 7.073454 251.0550  rice
## 3 90 46 42    23.97898 81.45062 7.502834 250.0832  rice
## 4 95 55 42    26.79534 82.14809 5.950661 193.3474  rice
## 5 76 49 42    24.95878 84.47963 5.206373 196.9560  rice
## 6 93 56 42    23.85724 82.22573 7.382763 195.0948  rice
```

```
df10=filter(df,label=="maize",ph<8)
```

head(df10)

```
##    N  P  K temperature humidity       ph  rainfall label
```

```
## 1 71 54 16    22.61360 63.69071 5.749914  87.75954 maize
## 2 61 44 17    26.10018 71.57477 6.931757 102.26624 maize
## 3 80 43 16    23.55882 71.59351 6.657965  66.71995 maize
## 4 73 58 21    19.97216 57.68273 6.596061  60.65171 maize
## 5 61 38 20    18.47891 62.69504 5.970458  65.43835 maize
## 6 68 41 16    21.77689 57.80841 6.158831 102.08617 maize
```

```
df11=filter(df,N==70)
```

head(df11)

```
##    N  P  K temperature humidity       ph  rainfall  label
```

```
## 1 70 36 42    21.84107 80.72886 6.946210 202.38383   rice
## 2 70 44 19    23.31689 73.45415 5.852607  94.29713  maize
## 3 70 47 17    24.61291 70.41624 6.600827 104.16261  maize
## 4 70 68 45    33.83509 92.85470 6.991626 203.40440 papaya
## 5 70 54 46    39.73149 91.12221 6.919342 122.76287 papaya
## 6 70 68 55    42.84609 94.63548 6.691202  78.80996 papaya
```

```
df12=filter(df,label=="chickpea",rainfall<85)
```

head(df12)

```
##    N  P  K temperature humidity       ph rainfall    label
## 1 23 72 84    19.02061 17.13159 6.920251 79.92698 chickpea
## 2 39 58 85    17.88776 15.40590 5.996932 68.54933 chickpea
## 3 36 67 77    18.36953 19.56381 7.152811 79.26358 chickpea
## 4 58 70 84    20.65432 16.60821 6.231049 74.66311 chickpea
## 5 59 70 84    17.33487 18.74927 7.550808 82.61735 chickpea
## 6 42 62 75    18.17912 18.90427 7.010571 81.84998 chickpea
```

```
#Rename function
```

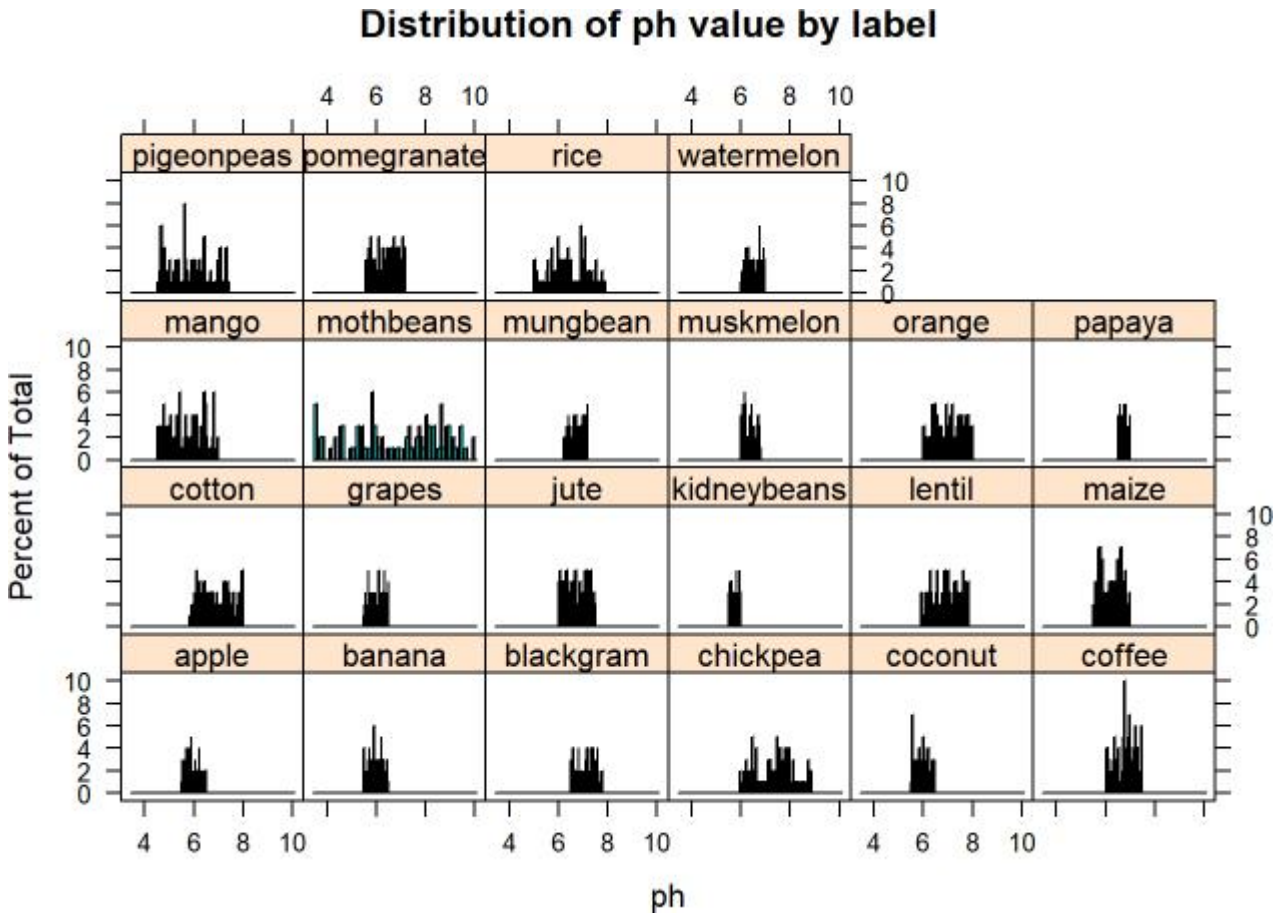
head(rename(df,temp=temperature),5)

```
##    N  P  K     temp humidity       ph rainfall label
## 1 90 42 43 20.87974 82.00274 6.502985 202.9355  rice
## 2 85 58 41 21.77046 80.31964 7.038096 226.6555  rice
## 3 60 55 44 23.00446 82.32076 7.840207 263.9642  rice
## 4 74 35 40 26.49110 80.15836 6.980401 242.8640  rice
## 5 78 42 42 20.13017 81.60487 7.628473 262.7173  rice
```

```
#Visualizing the dataset
```

#Histogram

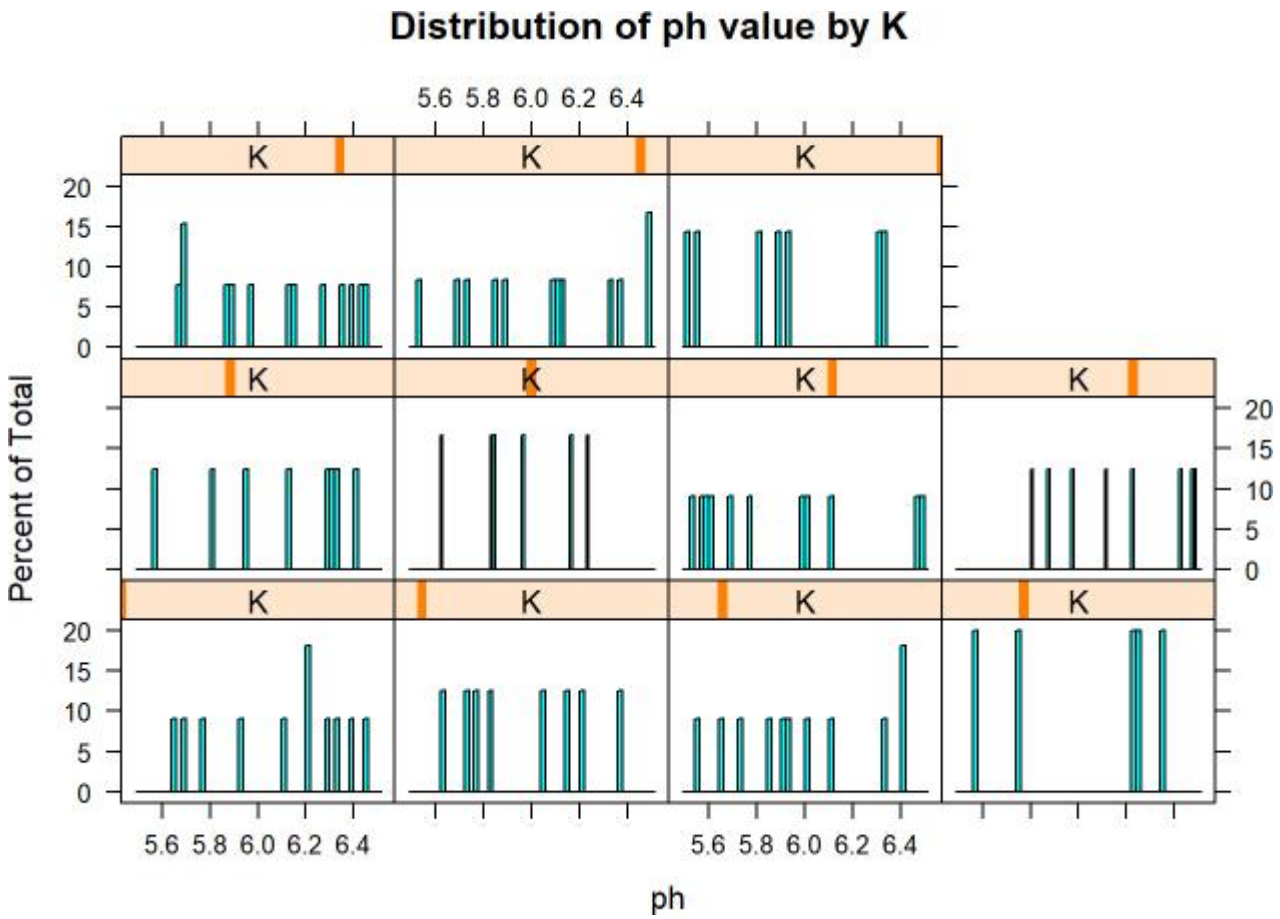
#1.Which label got the maximum number of ph value
#Using histogram to find out the maximum number of ph value in label wise analysis
library(lattice)
histogram(~ph|label,data=df,breaks=50,main="Distribution of ph value by label")




```
#Subset the label grapes by using dplyr package
```

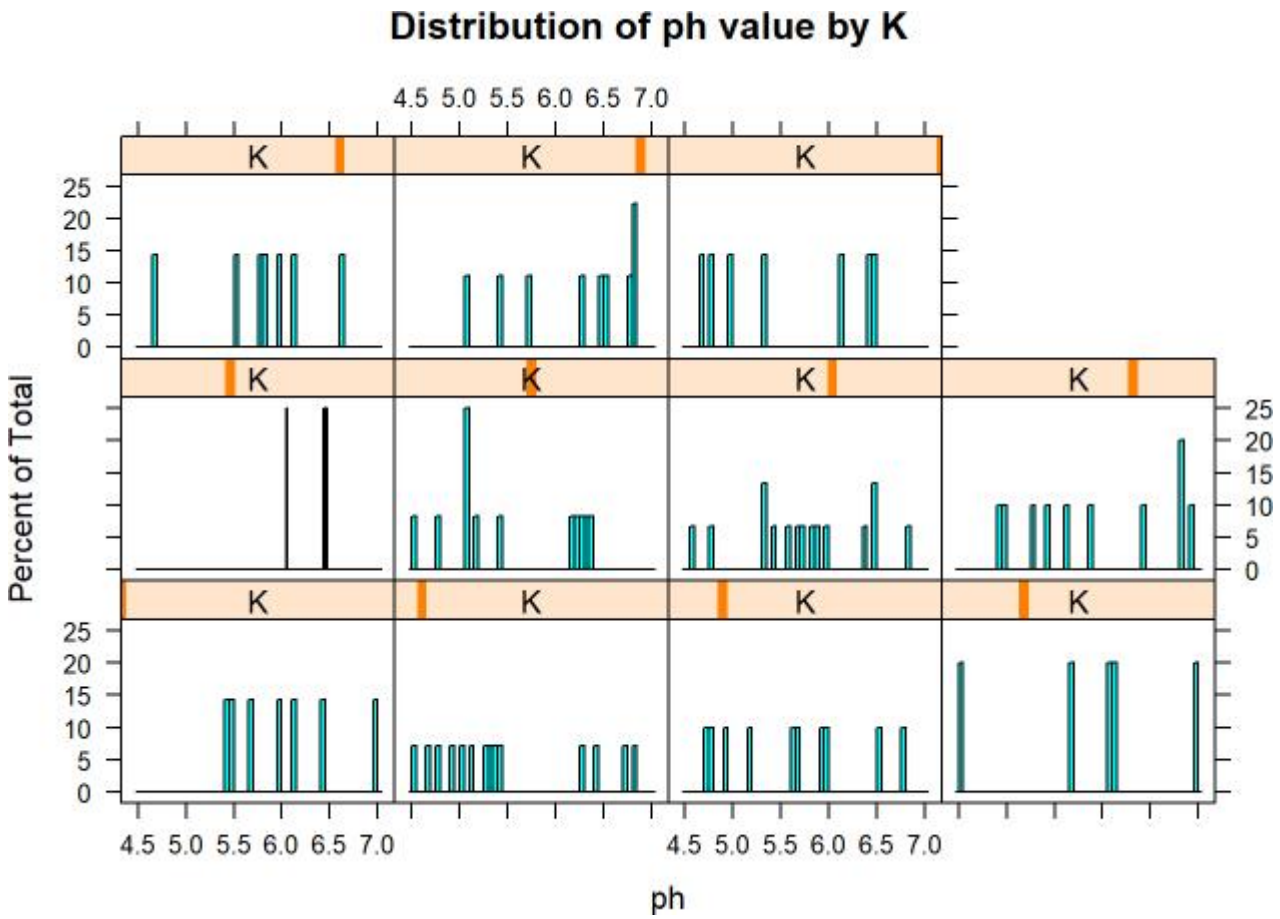
```
library(dplyr)
grapes=filter(df,df$label=="grapes")

#2 Find out which K has maximum number of ph value
library(lattice)
histogram(~ph|K,data=grapes,breaks=50,main="Distribution of ph value by K")
```



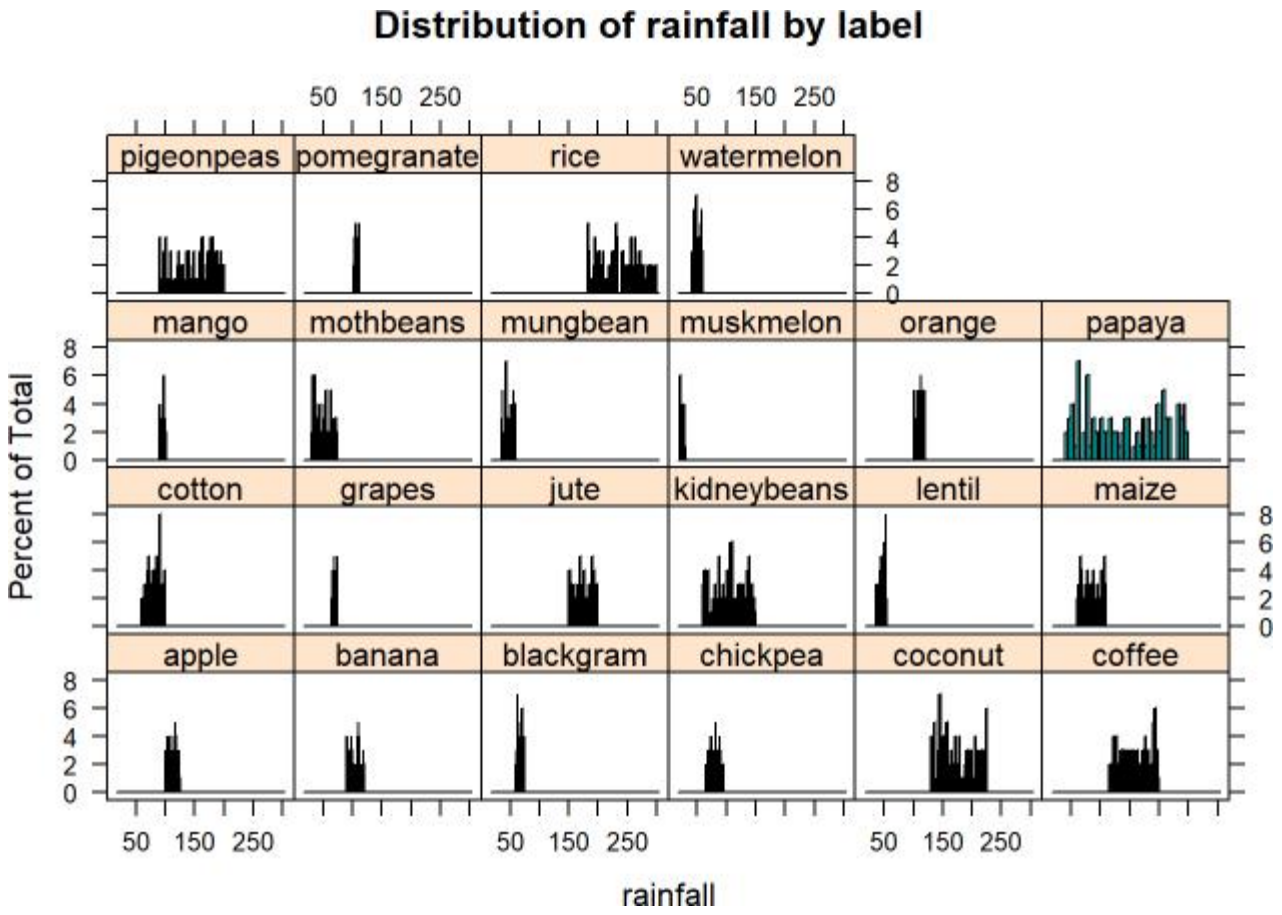

```
#Subset thelabel mango by using dplyr package
```

```
mango=filter(df,df$label=="mango")
#Find out which K has maximum number of ph value
histogram(~ph|K,data=mango,breaks=50,main="Distribution of ph value by K")
```



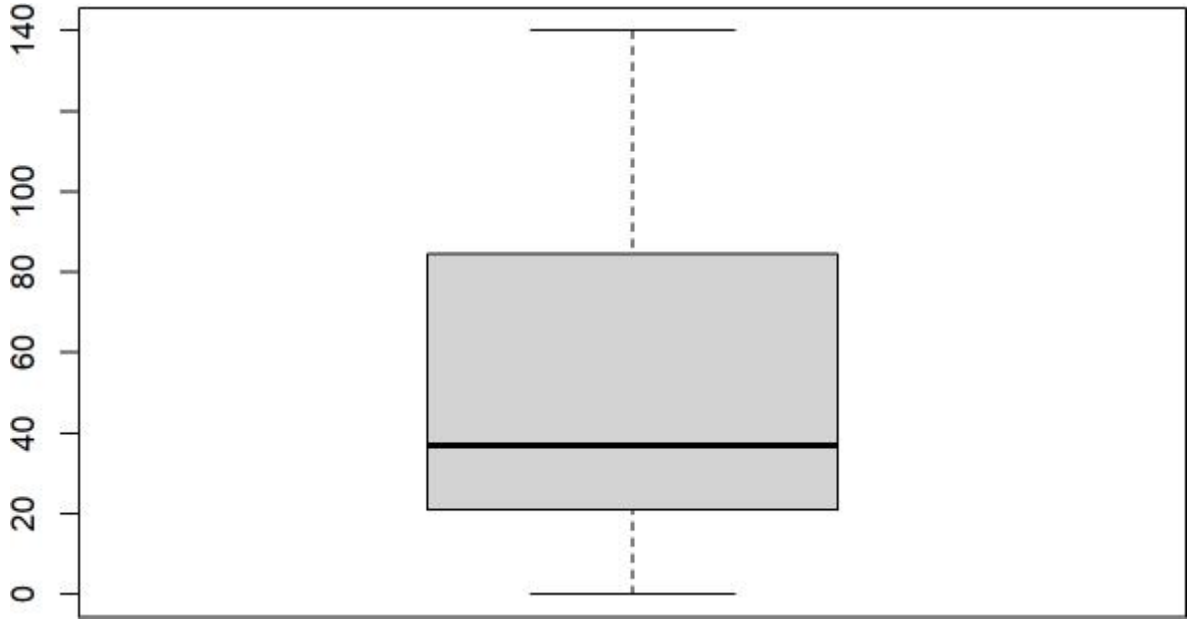

```
#3 Which label got most rainfall
```

#Using histogram to find out the maximum number of rainfall in label wise analysis
library(lattice)
histogram(~rainfall|label,data=df,breaks=50,main="Distribution of rainfall by label")



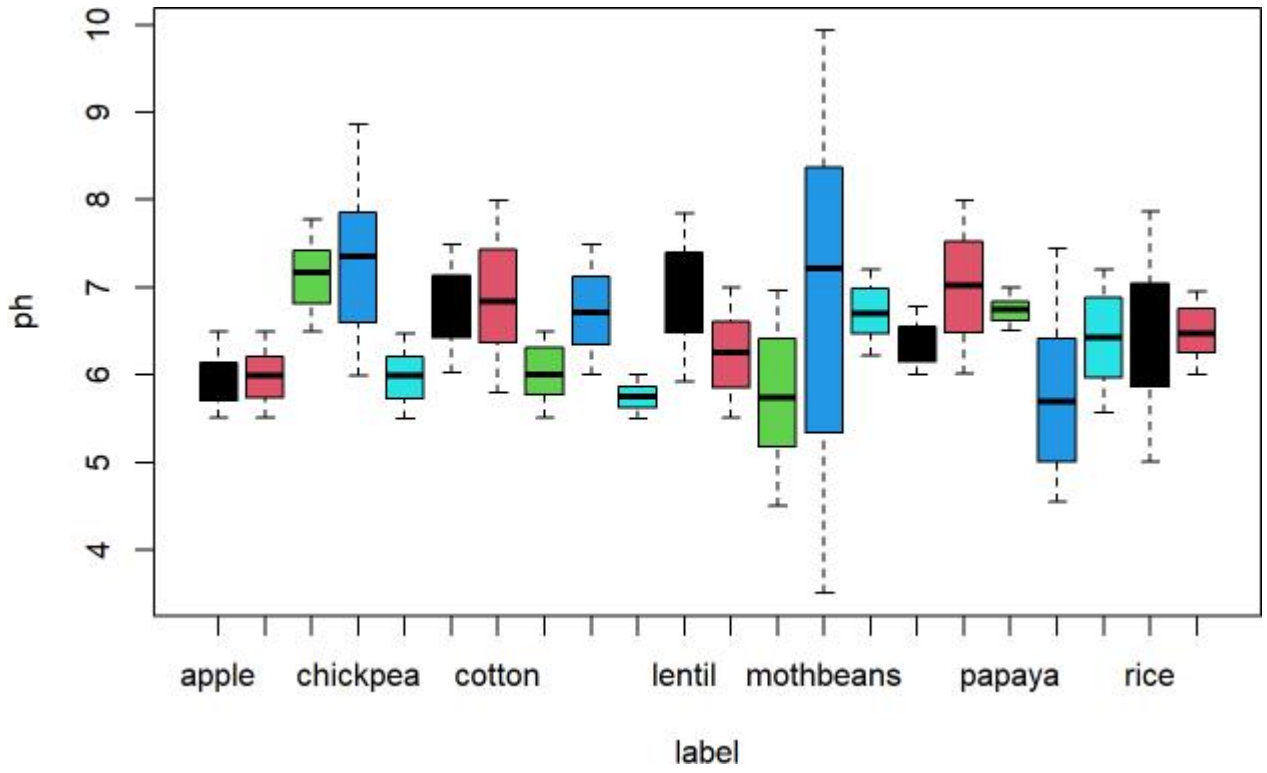

```
#Boxplot
```

#1
with(df,boxplot(N))



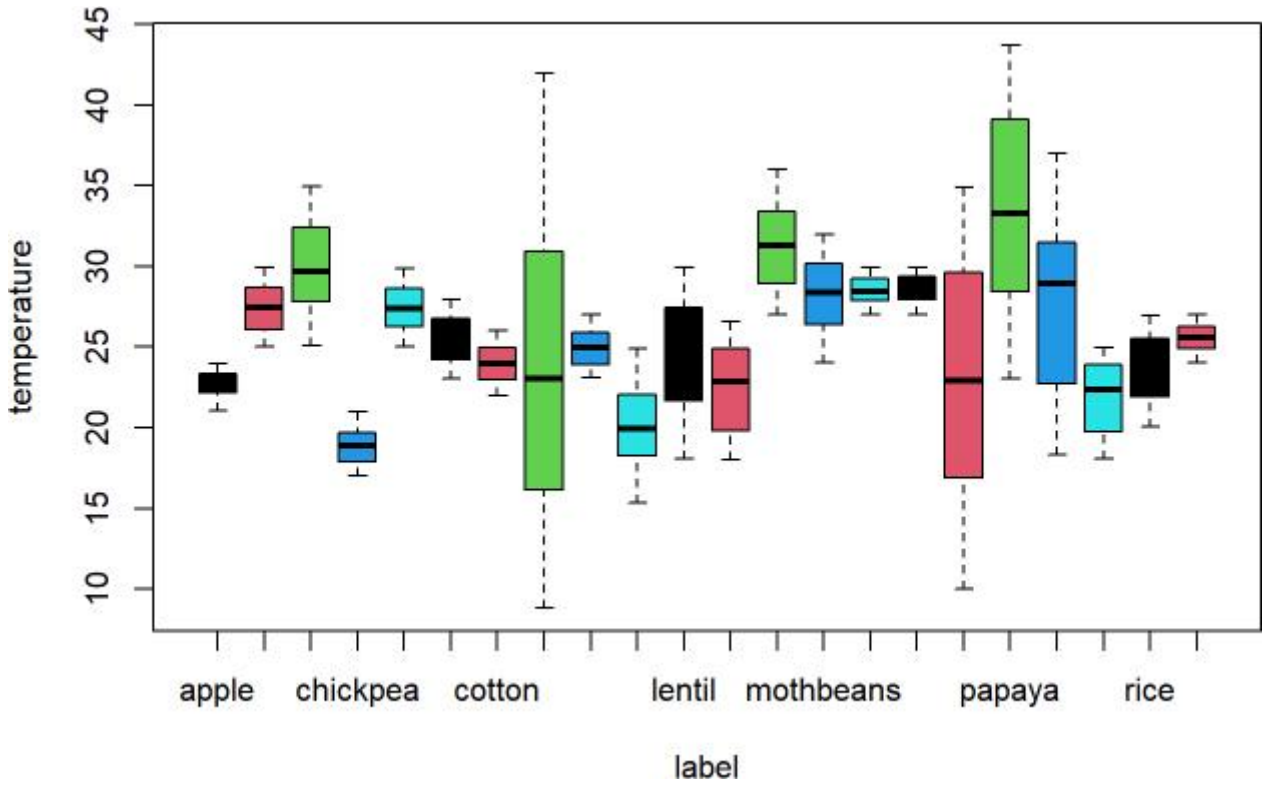

```
#2
```

```
with(df,boxplot(ph~label,col=c(1,2,3,4,5)))
```




```
#3
```

```
with(df,boxplot(temperature~label,col=c(1,2,3,4,5)))
```




```
#Scatterplot
```

#1 Which label has maximum temperature

```
library(plotly)
fig=plot_ly(data=df,x=~label,y=~temperature,type="scatter")%>%layout(title="Scatterplot between label and temperature")
fig
```

```
## No scatter mode specifed:
##   Setting the mode to markers
##   Read more about this attribute ->
https://plotly.com/r/reference/#scatter-mode
```