**Soil Moisture Tamilnadu -2018**

**Problem:**

Farmers need to analysis their soil moisture level to [helping farmers make optimum cultivation decisions](https://www.myfarminfo.com/blog/soil-success-helping-farmers-make-optimum-cultivation-decisions/) for **agricultural**

**Assumption:**

The basic idea of analyzing the soil moisture dataset is to get a fair idea about the soil moisture level in Tamilnadu.

We need to find which of district has the best soil for **agriculture.**

Hence build a model to predict the Soil Moistureof each district.

**Features on data set:**

* **Date:** It show date of recording
* **State\_Name:** It show the name of the state date of recording
* **District\_Name:** It show the name of the District date of recording
* **Average soil moisture level:** The water in your soil is stored on the surface of the soil particles, as well as in the pores, which are holes or gaps between individual soil particles. Pores will contain both water and air/oxygen.
* **Aggregate soil moisture percentage:** It is important to note that the majority of flowers, trees, and shrubs require moisture levels between 21% - 40%,
* **Volume soil moisture percentage:** It's calculated by weighing the wet soil sampled from the field, drying it in an oven, and then weighing the dry soil.

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

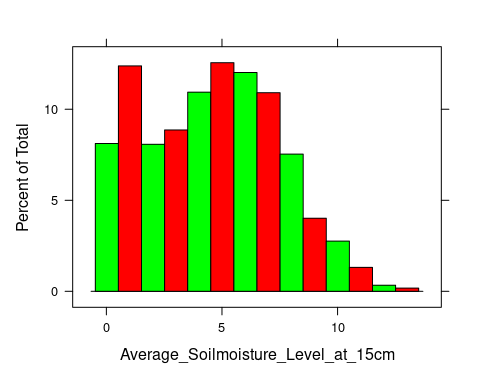
library(lattice)  
library(ggplot2)  
df=read.csv("soil.csv")  
colnames(df) <- c('Date','State\_Name','District\_Name','Average\_Soilmoisture\_Level\_at\_15cm','Aggregate\_Soilmoisture\_Percentage\_at\_15cm','Volume\_Soilmoisture\_percentage\_at\_15cm')  
df

## Date State\_Name District\_Name Average\_Soilmoisture\_Level\_at\_15cm  
## 1 2018/06/30 TAMILNADU THENI 5.42  
## 2 2018/06/30 TAMILNADU COIMBATORE 8.11  
## 3 2018/06/30 TAMILNADU KARUR 3.55  
## 4 2018/06/30 TAMILNADU VIRUDUNAGAR 3.00  
## 5 2018/06/30 TAMILNADU NAGAPPATTINAM 0.50  
## 6 2018/06/30 TAMILNADU NAMAKKAL 2.71  
## 7 2018/06/30 TAMILNADU CUDDALORE 0.30  
## 8 2018/06/30 TAMILNADU TIRUPPUR 8.74  
## 9 2018/06/30 TAMILNADU THE NILGIRIS 3.53  
## 10 2018/06/30 TAMILNADU THANJAVUR 1.64  
## 11 2018/06/30 TAMILNADU THOOTHUKKUDI 1.27  
## 12 2018/06/30 TAMILNADU DHARMAPURI 3.34  
## 13 2018/06/30 TAMILNADU ERODE 5.71  
## 14 2018/06/30 TAMILNADU KANNIYAKUMARI 0.61  
## 15 2018/06/30 TAMILNADU DINDIGUL 4.39  
## 16 2018/06/30 TAMILNADU VILUPPURAM 0.45  
## 17 2018/06/30 TAMILNADU TIRUVANNAMALAI 0.08  
## 18 2018/06/30 TAMILNADU RAMANATHAPURAM 0.82  
## 19 2018/06/30 TAMILNADU SIVAGANGA 0.95  
## 20 2018/06/30 TAMILNADU ARIYALUR 0.62  
## 21 2018/06/30 TAMILNADU KANCHEEPURAM 0.57  
## 22 2018/06/30 TAMILNADU VELLORE 1.07  
## 23 2018/06/30 TAMILNADU KRISHNAGIRI 4.34  
## 24 2018/06/30 TAMILNADU TIRUCHIRAPPALLI 2.51  
## 25 2018/06/30 TAMILNADU PERAMBALUR 0.52  
## 26 2018/06/30 TAMILNADU THIRUVARUR 0.15  
## 27 2018/06/30 TAMILNADU TIRUNELVELI 3.41  
## 28 2018/06/30 TAMILNADU CHENNAI 3.45  
## 29 2018/06/30 TAMILNADU THIRUVALLUR 0.41  
## 30 2018/06/30 TAMILNADU SALEM 3.55  
## 31 2018/06/30 TAMILNADU MADURAI 4.58  
## 32 2018/06/30 TAMILNADU PUDUKKOTTAI 3.69  
## 33 2018/06/29 TAMILNADU THENI 5.33  
## 34 2018/06/29 TAMILNADU COIMBATORE 8.14  
## 35 2018/06/29 TAMILNADU KARUR 3.49  
## 36 2018/06/29 TAMILNADU VIRUDUNAGAR 3.02  
## 37 2018/06/29 TAMILNADU NAGAPPATTINAM 0.45  
## 38 2018/06/29 TAMILNADU NAMAKKAL 2.70  
## 39 2018/06/29 TAMILNADU CUDDALORE 0.31  
## 40 2018/06/29 TAMILNADU TIRUPPUR 8.71  
## 41 2018/06/29 TAMILNADU THE NILGIRIS 3.48  
## 42 2018/06/29 TAMILNADU THANJAVUR 1.29  
## 43 2018/06/29 TAMILNADU THOOTHUKKUDI 1.37  
## 44 2018/06/29 TAMILNADU DHARMAPURI 3.42  
## 45 2018/06/29 TAMILNADU ERODE 5.78  
## 46 2018/06/29 TAMILNADU KANNIYAKUMARI 0.62  
## 47 2018/06/29 TAMILNADU DINDIGUL 4.05  
## 48 2018/06/29 TAMILNADU VILUPPURAM 0.46  
## 49 2018/06/29 TAMILNADU TIRUVANNAMALAI 0.09  
## 50 2018/06/29 TAMILNADU RAMANATHAPURAM 0.82  
## Aggregate\_Soilmoisture\_Percentage\_at\_15cm  
## 1 42.36  
## 2 61.20  
## 3 23.58  
## 4 23.13  
## 5 6.22  
## 6 21.50  
## 7 3.56  
## 8 54.66  
## 9 28.59  
## 10 12.54  
## 11 12.68  
## 12 22.21  
## 13 36.64  
## 14 4.93  
## 15 27.03  
## 16 4.07  
## 17 0.77  
## 18 9.54  
## 19 9.13  
## 20 5.45  
## 21 5.51  
## 22 6.74  
## 23 28.59  
## 24 18.60  
## 25 3.15  
## 26 1.61  
## 27 29.53  
## 28 39.03  
## 29 3.41  
## 30 25.56  
## 31 36.62  
## 32 26.91  
## 33 41.61  
## 34 61.43  
## 35 23.07  
## 36 23.29  
## 37 5.43  
## 38 21.40  
## 39 3.68  
## 40 54.47  
## 41 28.23  
## 42 10.14  
## 43 13.44  
## 44 22.68  
## 45 37.14  
## 46 5.07  
## 47 25.03  
## 48 4.14  
## 49 0.84  
## 50 9.59  
  
Volume\_Soilmoisture\_percentage\_at\_15cm NA  
## 1 38.176163 NA  
## 2 37.690124 NA  
## 3 30.816165 NA  
## 4 30.894787 NA  
## 5 22.360402 NA  
## 6 29.782532 NA  
## 7 28.088107 NA  
## 8 30.062344 NA  
## 9 41.136121 NA  
## 10 29.872339 NA  
## 11 26.849510 NA  
## 12 30.678367 NA  
## 13 30.263841 NA  
## 14 33.670358 NA  
## 15 35.297029 NA  
## 16 28.922069 NA  
## 17 29.228539 NA  
## 18 23.320870 NA  
## 19 33.850359 NA  
## 20 30.936665 NA  
## 21 25.257855 NA  
## 22 28.379394 NA  
## 23 30.669593 NA  
## 24 31.872773 NA  
## 25 34.450674 NA  
## 26 28.357739 NA  
## 27 27.974361 NA  
## 28 14.791859 NA  
## 29 25.064694 NA  
## 30 31.352249 NA  
## 31 41.716951 NA  
## 32 32.318523 NA  
## 33 36.529461 NA  
## 34 38.233076 NA  
## 35 26.632956 NA  
## 36 27.969064 NA  
## 37 19.942142 NA  
## 38 28.587858 NA  
## 39 27.235769 NA  
## 40 29.920783 NA  
## 41 42.345105 NA  
## 42 24.680588 NA  
## 43 25.557790 NA  
## 44 31.319230 NA  
## 45 30.940608 NA  
## 46 34.582933 NA  
## 47 32.546210 NA  
## 48 29.554178 NA  
## 49 29.901736 NA  
## 50 22.316881 NA

summary(df)

## Date State\_Name District\_Name   
## Length:6848 Length:6848 Length:6848   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## Average\_Soilmoisture\_Level\_at\_15cm Aggregate\_Soilmoisture\_Percentage\_at\_15cm  
## Min. : 0.04 Min. : 0.43   
## 1st Qu.: 2.11 1st Qu.:16.88   
## Median : 4.67 Median :34.79   
## Mean : 4.57 Mean :35.12   
## 3rd Qu.: 6.72 3rd Qu.:51.08   
## Max. :12.99 Max. :91.02   
## Volume\_Soilmoisture\_percentage\_at\_15cm NA   
## Min. : 7.874 Mode:logical   
## 1st Qu.:24.295 NA's:6848   
## Median :29.954   
## Mean :29.812   
## 3rd Qu.:35.989   
## Max. :45.970

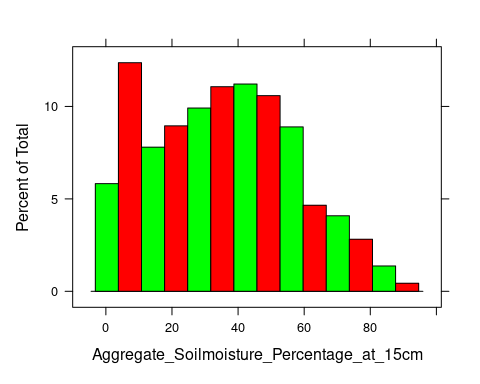
histogram(~Average\_Soilmoisture\_Level\_at\_15cm,data=df,col=c("green","red"))



**INFERENCE**

* As displacement plotted in histogram the mean is high than median so it right skewed
* The **maximum number of** Average Soil moisture Level **is in between 0 and 150**

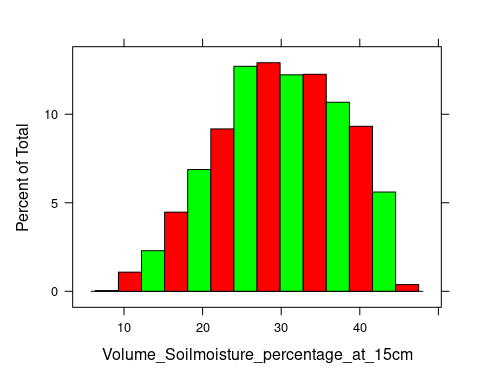
histogram(~Aggregate\_Soilmoisture\_Percentage\_at\_15cm,data=df,col=c("green","red"))



**INFERENCE**

* As Aggregate Soil moisture Percentage plotted in histogram the mean is high than median so it right skewed
* The Aggregate Soil moisture Percentage falls on **Poisson distribution and maximum number** Aggregate Soil moisture **of is in between 5 and 60**

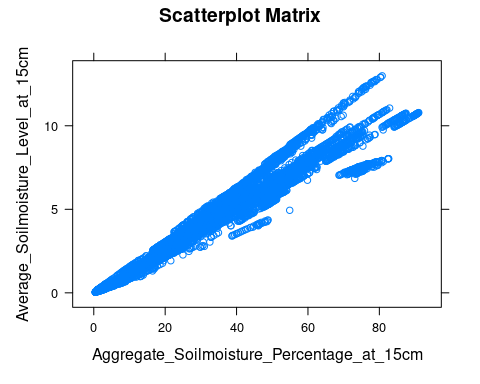
histogram(~Volume\_Soilmoisture\_percentage\_at\_15cm,data=df,col=c("green","red))



**INFERENCE**

* As Volume Soil moisture percentage plotted in histogram the mean is less than median so it left-skewed
* The **maximum number of** Volume Soil moisture percentage **is in between 25 and 35**

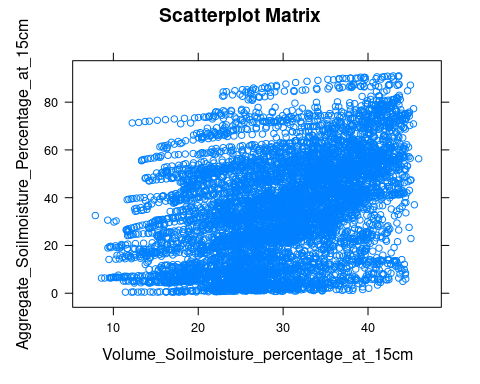
xyplot(Average\_Soilmoisture\_Level\_at\_15cm~Aggregate\_Soilmoisture\_Percentage\_at\_15cm,data = df,main = "Scatterplot Matrix")



**INFERENCE**

* It is high correlation
* High the Average Soil moisture Level of the soil. higher the Aggregate Soil moisture Percentage
* At the Average Soil moisture Level 10 the Aggregate Soil moisture Percentage is high as 86

xyplot(Aggregate\_Soilmoisture\_Percentage\_at\_15cm~Volume\_Soilmoisture\_percentage\_at\_15cm,data = df,main = "Scatterplot Matrix")



**INFERENCE**

* there is no relationship between the two variables
* This can be indicated by a correlation coefficient of zero, which means that the variables are completely independent of each other.

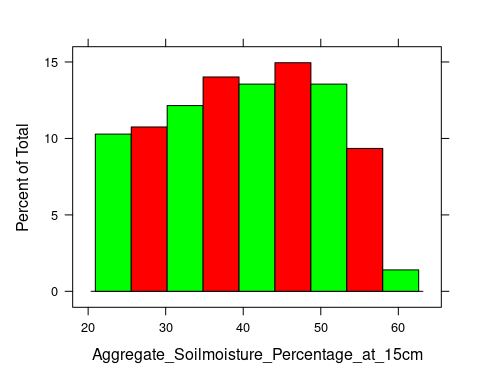
test <- cor.test(df$Aggregate\_Soilmoisture\_Percentage\_at\_15cm, df$Volume\_Soilmoisture\_percentage\_at\_15cm)  
test

##   
## Pearson's product-moment correlation  
##   
## data: df$Aggregate\_Soilmoisture\_Percentage\_at\_15cm and df$Volume\_Soilmoisture\_percentage\_at\_15cm  
## t = 37.279, df = 6846, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3909032 0.4302841  
## sample estimates:  
## cor   
## 0.4107852

df1=read.csv("soil.csv")  
colnames(df1) <- c('Date','State\_Name','District\_Name','Average\_Soilmoisture\_Level\_at\_15cm','Aggregate\_Soilmoisture\_Percentage\_at\_15cm','Volume\_Soilmoisture\_percentage\_at\_15cm')  
  
df1

p1=subset(df1,District\_Name=="DINDIGUL")  
p1

histogram(~Aggregate\_Soilmoisture\_Percentage\_at\_15cm,data=p1,col=c("green","red"))

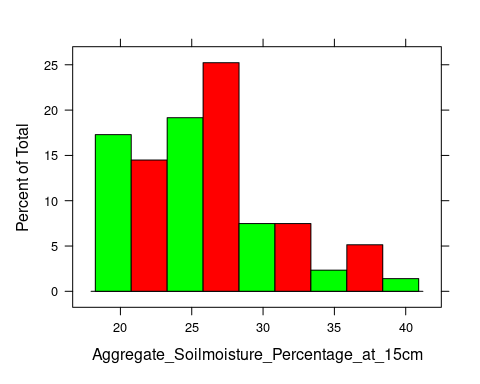


**INFERENCE**

* As Soil moisture percentage plotted of DINDIGUL in histogram the mean is less than median so it left-skewed
* The **maximum number of** Aggregate Soil moisture percentage **is in between 35 and 55**

p2=subset(df1,District\_Name=="NAMAKKAL")  
p2

histogram(~Aggregate\_Soilmoisture\_Percentage\_at\_15cm,data=p2,col=c("green","red"))

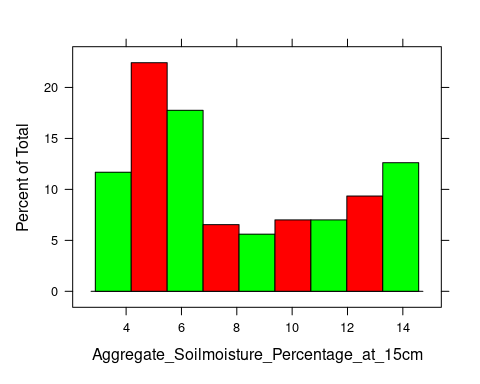


**INFERENCE**

* As Soil moisture percentage plotted of NAMAKKAL in histogram the mean is less than median so it left-skewed
* The **maximum number of** Aggregate Soil moisture percentage **is in between 0 and 28**

p3=subset(df1,District\_Name=="KANNIYAKUMARI")  
p3

histogram(~Aggregate\_Soilmoisture\_Percentage\_at\_15cm,data=p3,col=c("green","red"))

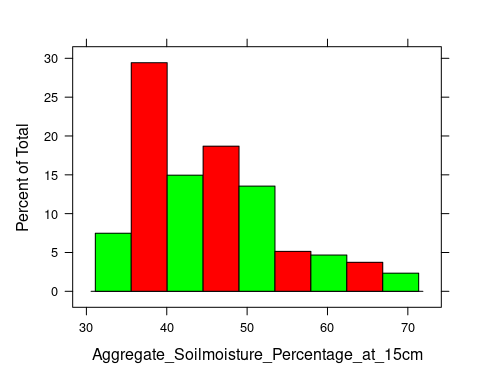


**INFERENCE**

* As Soil moisture percentage plotted of KANNIYAKUMARI in histogram the mean is less than median so it left-skewed
* The **maximum number of** Aggregate Soil moisture percentage **is in between 0 and 7**

p4=subset(df1,District\_Name=="ERODE")  
p4

histogram(~Aggregate\_Soilmoisture\_Percentage\_at\_15cm,data=p4,col=c("green","red"))

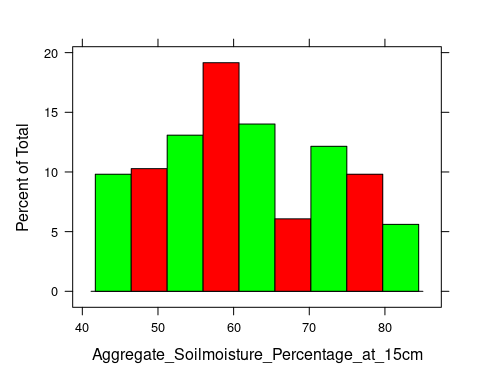


**INFERENCE**

* As Soil moisture percentage plotted of ERODE in histogram the mean is less than median so it left-skewed
* The **maximum number of** Aggregate Soil moisture percentage **is in between 25 and 45**

p5=subset(df1,District\_Name=="COIMBATORE")  
p5

histogram(~Aggregate\_Soilmoisture\_Percentage\_at\_15cm,data=p5,col=c("green","red"))



**INFERENCE**

* As Soil moisture percentage plotted of COIMBATORE in histogram the mean is less than median so it left-skewed
* The **maximum number of** Aggregate Soil moisture percentage **is in between 50 and 65**

**Inferences:**

* Dindigul has the good soil moisture out of all other district.
* High the Average Soil moisture Level of the soil. higher the Aggregate Soil moisture Percentage
* If the soil moisture content is too low, it can lead to drought stress in plants, which can also lead to reduced crop yields.
* Average Soil moisture affect the Aggregate Soil moisture Percentage