Assignment 1

Moushreeta Debroy(22122132)

2023-01-15

library(readxl)  
crop\_data <- read\_excel("C:/Users/PRASANTA/Downloads/crop.data.xlsx")  
View(crop\_data)

View(crop\_data)

dim(crop\_data)

## [1] 96 4

str(crop\_data)

## tibble [96 × 4] (S3: tbl\_df/tbl/data.frame)  
## $ density : num [1:96] 1 2 1 2 1 2 1 2 1 2 ...  
## $ soil\_type : num [1:96] 1 2 3 4 1 2 3 4 1 2 ...  
## $ fertilizer: num [1:96] 1 1 1 1 1 1 1 1 1 1 ...  
## $ yield : num [1:96] 177 178 176 178 177 ...

attributes(crop\_data)

## $class  
## [1] "tbl\_df" "tbl" "data.frame"  
##   
## $row.names  
## [1] 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] 26 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] 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75  
## [76] 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96  
##   
## $names  
## [1] "density" "soil\_type" "fertilizer" "yield"

sum(is.na(crop\_data))

## [1] 0

dim(crop\_data) str(crop\_data)

names(crop\_data)

## [1] "density" "soil\_type" "fertilizer" "yield"

head(crop\_data)

## # A tibble: 6 × 4  
## density soil\_type fertilizer yield  
## <dbl> <dbl> <dbl> <dbl>  
## 1 1 1 1 177.  
## 2 2 2 1 178.  
## 3 1 3 1 176.  
## 4 2 4 1 178.  
## 5 1 1 1 177.  
## 6 2 2 1 177.

tail(crop\_data)

## # A tibble: 6 × 4  
## density soil\_type fertilizer yield  
## <dbl> <dbl> <dbl> <dbl>  
## 1 1 3 3 177.  
## 2 2 4 3 177.  
## 3 1 1 3 178.  
## 4 2 2 3 178.  
## 5 1 3 3 178.  
## 6 2 4 3 177.

crop\_data[5:10,]

## # A tibble: 6 × 4  
## density soil\_type fertilizer yield  
## <dbl> <dbl> <dbl> <dbl>  
## 1 1 1 1 177.  
## 2 2 2 1 177.  
## 3 1 3 1 177.  
## 4 2 4 1 177.  
## 5 1 1 1 176.  
## 6 2 2 1 178.

crop\_data[5:10,"yield"]

## # A tibble: 6 × 1  
## yield  
## <dbl>  
## 1 177.  
## 2 177.  
## 3 177.  
## 4 177.  
## 5 176.  
## 6 178.

summary(crop\_data)

## density soil\_type fertilizer yield   
## Min. :1.0 Min. :1.00 Min. :1 Min. :175.4   
## 1st Qu.:1.0 1st Qu.:1.75 1st Qu.:1 1st Qu.:176.5   
## Median :1.5 Median :2.50 Median :2 Median :177.1   
## Mean :1.5 Mean :2.50 Mean :2 Mean :177.0   
## 3rd Qu.:2.0 3rd Qu.:3.25 3rd Qu.:3 3rd Qu.:177.4   
## Max. :2.0 Max. :4.00 Max. :3 Max. :179.1

#mean: Mean is the average of the given numbers and is calculated by dividing the sum of given numbers by the total number of numbers.

#median:Median, in statistics, is the middle value of the given list of data when arranged in an order. The arrangement of data or observations can be made either in ascending order or descending order.

#IQR:The interquartile range defines the difference between the third and the first quartile. Quartiles are the partitioned values that divide the whole series into 4 equal parts. So, there are 3 quartiles. First Quartile is denoted by Q1 known as the lower quartile, the second Quartile is denoted by Q2 and the third Quartile is denoted by Q3 known as the upper quartile. Therefore, the interquartile range is equal to the upper quartile minus lower quartile.

#correlation:Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate in relation to each other. A positive correlation indicates the extent to which those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases.

#OUTLIERS:An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. In a sense, this definition leaves it up to the analyst (or a consensus process) to decide what will be considered abnormal. Before abnormal observations can be singled out, it is necessary to characterize normal observations.

mean(crop\_data$yield)

## [1] 177.0155

median(crop\_data$yield)

## [1] 177.0581

IQR(crop\_data$yield)

## [1] 0.9298747

range(crop\_data$yield)

## [1] 175.3608 179.0609

quantile(crop\_data$yield,c(0.10,0.60,0.70))

## 10% 60% 70%   
## 176.0880 177.1649 177.3484

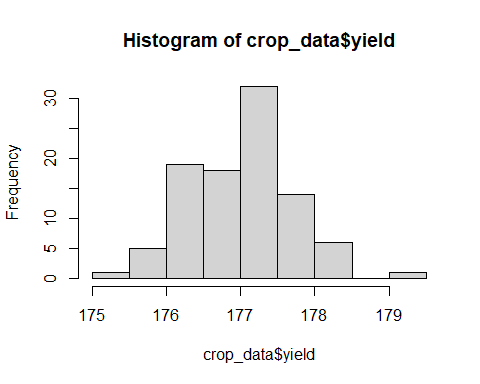
var(crop\_data$yield)

## [1] 0.4416235

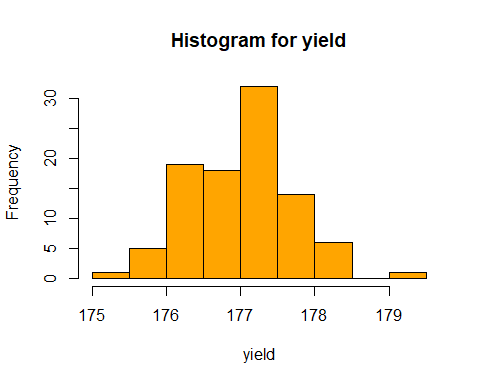
sd(crop\_data$yield)

## [1] 0.6645476

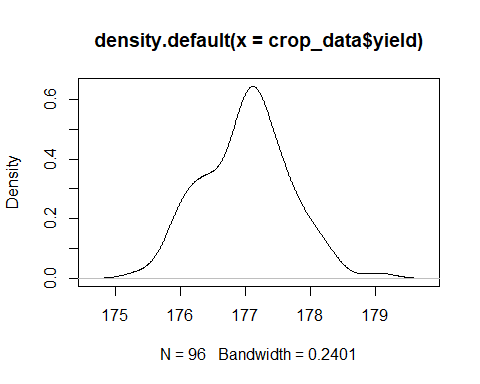
hist(crop\_data$yield)



hist(crop\_data$yield,col="orange",  
 xlab="yield",  
 main="Histogram for yield")

 #density plot

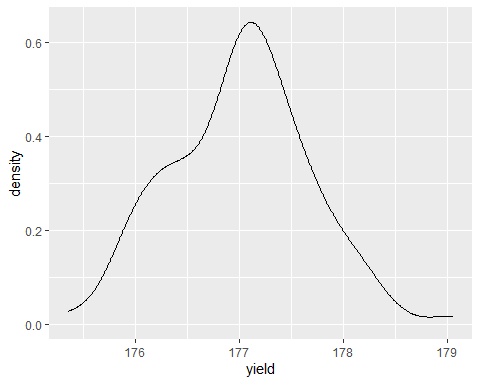
plot(density(crop\_data$yield))



library(ggplot2)

ggplot(data=crop\_data,aes(yield,fill=density))+  
 geom\_density()

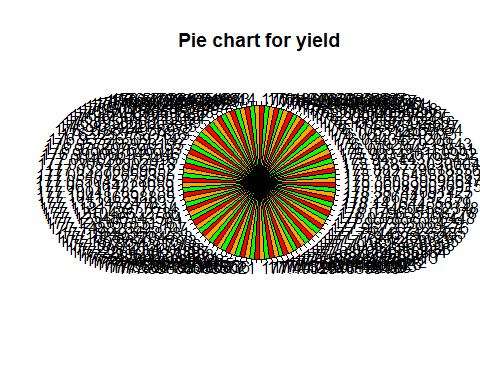
## Warning: The following aesthetics were dropped during statistical transformation: fill  
## ℹ This can happen when ggplot fails to infer the correct grouping structure in  
## the data.  
## ℹ Did you forget to specify a `group` aesthetic or to convert a numerical  
## variable into a factor?



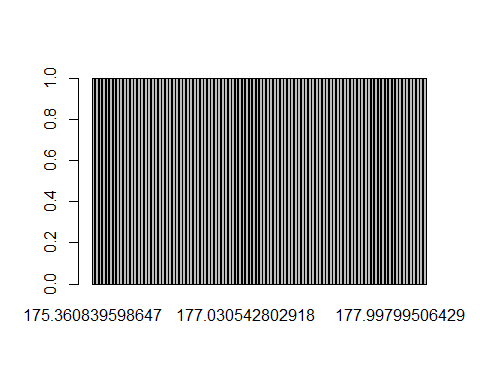
table(crop\_data$yield)

##   
## 175.360839598647 175.747545582883 175.882779618556 175.938533030004   
## 1 1 1 1   
## 175.945443795425 175.983480169992 176.008394511531 176.039297939843   
## 1 1 1 1   
## 176.0443421205 176.06774497035 176.108312589654 176.114388315927   
## 1 1 1 1   
## 176.160586502816 176.239015775225 176.274949297277 176.300517050366   
## 1 1 1 1   
## 176.353940642858 176.405410230471 176.408461852371 176.414495002246   
## 1 1 1 1   
## 176.429156002757 176.430830126686 176.434862565867 176.436762368181   
## 1 1 1 1   
## 176.479340919814 176.573069754873 176.601299834494 176.60372408049   
## 1 1 1 1   
## 176.668322935082 176.746301895632 176.778342481435 176.817922165499   
## 1 1 1 1   
## 176.819076695268 176.874129801452 176.878897743394 176.895866859859   
## 1 1 1 1   
## 176.903422145535 176.918844938972 176.92557577835 176.933265092142   
## 1 1 1 1   
## 176.957268920193 176.975807696393 176.999066949946 177.004038102495   
## 1 1 1 1   
## 177.030542802918 177.03270969652 177.034092659355 177.055045778695   
## 1 1 1 1   
## 177.061164221059 177.100417857738 177.104186398865 177.115245241837   
## 1 1 1 1   
## 177.118175977614 177.121048632786 177.125486343584 177.128367531761   
## 1 1 1 1   
## 177.148828595107 177.16493668281 177.168302204533 177.187967031073   
## 1 1 1 1   
## 177.197721367386 177.208171431058 177.226424131397 177.228692278688   
## 1 1 1 1   
## 177.262445084271 177.276995681026 177.344163946335 177.352595079015   
## 1 1 1 1   
## 177.360818239535 177.379778689577 177.385499180836 177.396330635021   
## 1 1 1 1   
## 177.405291855943 177.412461749009 177.479507160588 177.540284161493   
## 1 1 1 1   
## 177.550041265426 177.580683078112 177.59335237605 177.632808263613   
## 1 1 1 1   
## 177.652274608017 177.687264356834 177.701366283038 177.703625478918   
## 1 1 1 1   
## 177.710612540335 177.779492860266 177.794457435485 177.967202929576   
## 1 1 1 1   
## 177.99799506429 178.036858366448 178.079635168278 178.134604582348   
## 1 1 1 1   
## 178.14164435679 178.28604192157 178.35744091425 179.060899036915   
## 1 1 1 1

pie(table(crop\_data$yield),  
 main="Pie chart for yield",  
 col=c("orange","green","red"))



barplot(table(crop\_data$yield))

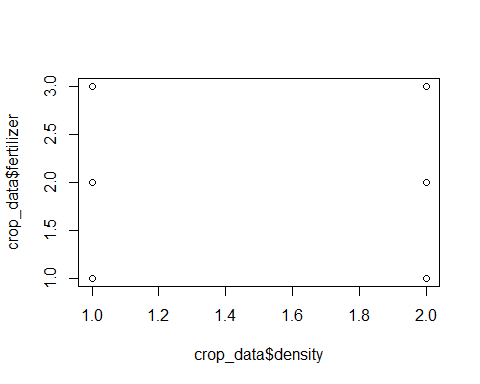


#—Correlation

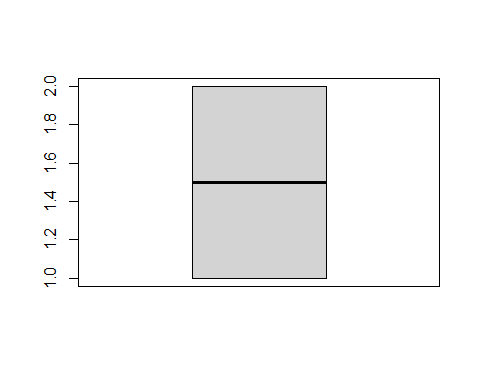
cor(crop\_data$density,crop\_data$soil\_type)

## [1] 0.4472136

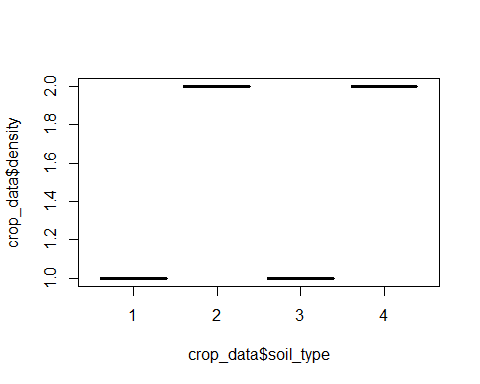
plot(crop\_data$density,crop\_data$fertilizer)



boxplot(crop\_data$density)



boxplot(data = crop\_data,crop\_data$density~crop\_data$soil\_type)

 # A matrix of scatter plot

pairs(crop\_data[,1:4])

