SWEETMOMENTSINR

KERAMA JAMES THE DATA ANALYST

2023-04-25

#DEALING WITH THE DATASET AIRQUALITY IN R  
#Data cleaning  
# Load the airquality dataset  
data(airquality)  
  
# Check for missing values  
sum(is.na(airquality))

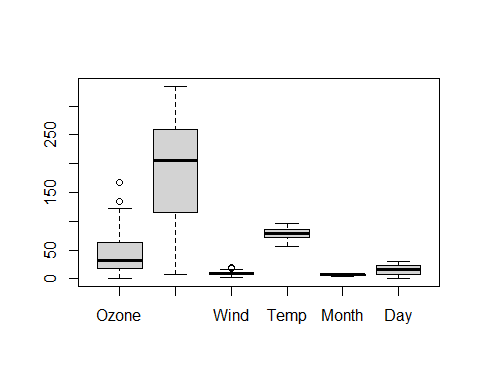
## [1] 44

# Check for outliers using boxplots  
boxplot(airquality)  
  
# Check data types  
str(airquality)

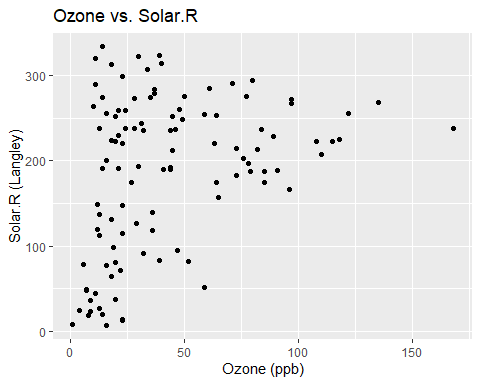
## 'data.frame': 153 obs. of 6 variables:  
## $ Ozone : int 41 36 12 18 NA 28 23 19 8 NA ...  
## $ Solar.R: int 190 118 149 313 NA NA 299 99 19 194 ...  
## $ Wind : num 7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...  
## $ Temp : int 67 72 74 62 56 66 65 59 61 69 ...  
## $ Month : int 5 5 5 5 5 5 5 5 5 5 ...  
## $ Day : int 1 2 3 4 5 6 7 8 9 10 ...

# Delete rows with missing values  
airquality <- na.omit(airquality)  
airquality <- na.omit(airquality)  
View(airquality)  
#Data visualization using the scatter plots  
# Load ggplot2 library  
library(ggplot2)

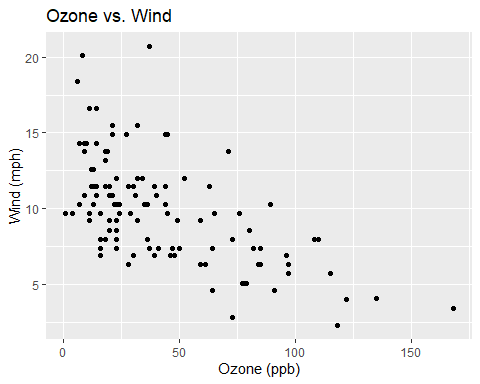
## Warning: package 'ggplot2' was built under R version 4.2.3



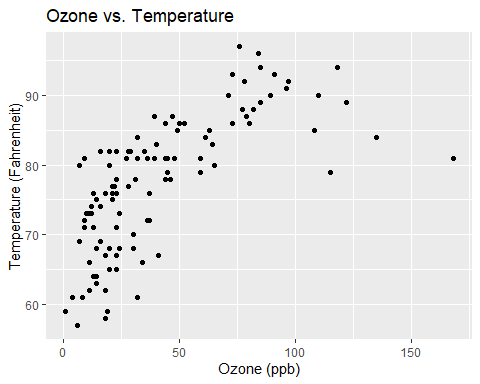
# Scatterplot of Ozone and Solar.R  
  
  
ggplot(airquality, aes(x = Ozone, y = Solar.R)) +   
 geom\_point() +   
 labs(title = "Ozone vs. Solar.R", x = "Ozone (ppb)", y = "Solar.R (Langley)")



# Scatterplot of Ozone and Wind  
  
ggplot(airquality, aes(x = Ozone, y = Wind)) +   
 geom\_point() +   
 labs(title = "Ozone vs. Wind", x = "Ozone (ppb)", y = "Wind (mph)")



# Scatterplot of Ozone and Temperature  
ggplot(airquality, aes(x = Ozone, y = Temp)) +   
 geom\_point() +   
 labs(title = "Ozone vs. Temperature", x = "Ozone (ppb)", y = "Temperature (Fahrenheit)")



#Data modeling using the linear regression models  
# Model Ozone and Solar.R  
model1 <- lm(Ozone ~ Solar.R, data = airquality)  
summary(model1)

##   
## Call:  
## lm(formula = Ozone ~ Solar.R, data = airquality)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -48.292 -21.361 -8.864 16.373 119.136   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 18.59873 6.74790 2.756 0.006856 \*\*   
## Solar.R 0.12717 0.03278 3.880 0.000179 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 31.33 on 109 degrees of freedom  
## Multiple R-squared: 0.1213, Adjusted R-squared: 0.1133   
## F-statistic: 15.05 on 1 and 109 DF, p-value: 0.0001793

# Model Ozone and Wind  
model2 <- lm(Ozone ~ Wind, data = airquality)  
summary(model2)

##   
## Call:  
## lm(formula = Ozone ~ Wind, data = airquality)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -43.513 -18.597 -5.035 15.814 88.437   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 99.0413 7.4724 13.25 < 2e-16 \*\*\*  
## Wind -5.7288 0.7082 -8.09 9.09e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 26.42 on 109 degrees of freedom  
## Multiple R-squared: 0.3752, Adjusted R-squared: 0.3694   
## F-statistic: 65.44 on 1 and 109 DF, p-value: 9.089e-13

# Model Ozone and Temperature  
model3 <- lm(Ozone ~ Temp, data = airquality)  
summary(model3)

##   
## Call:  
## lm(formula = Ozone ~ Temp, data = airquality)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -40.922 -17.459 -0.874 10.444 118.078   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -147.6461 18.7553 -7.872 2.76e-12 \*\*\*  
## Temp 2.4391 0.2393 10.192 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 23.92 on 109 degrees of freedom  
## Multiple R-squared: 0.488, Adjusted R-squared: 0.4833   
## F-statistic: 103.9 on 1 and 109 DF, p-value: < 2.2e-16

#Analysis  
# Summary of model1  
summary(model1)

##   
## Call:  
## lm(formula = Ozone ~ Solar.R, data = airquality)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -48.292 -21.361 -8.864 16.373 119.136   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 18.59873 6.74790 2.756 0.006856 \*\*   
## Solar.R 0.12717 0.03278 3.880 0.000179 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 31.33 on 109 degrees of freedom  
## Multiple R-squared: 0.1213, Adjusted R-squared: 0.1133   
## F-statistic: 15.05 on 1 and 109 DF, p-value: 0.0001793

# Summary of model2  
summary(model2)

##   
## Call:  
## lm(formula = Ozone ~ Wind, data = airquality)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -43.513 -18.597 -5.035 15.814 88.437   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 99.0413 7.4724 13.25 < 2e-16 \*\*\*  
## Wind -5.7288 0.7082 -8.09 9.09e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 26.42 on 109 degrees of freedom  
## Multiple R-squared: 0.3752, Adjusted R-squared: 0.3694   
## F-statistic: 65.44 on 1 and 109 DF, p-value: 9.089e-13

# Summary of model3  
summary(model3)

##   
## Call:  
## lm(formula = Ozone ~ Temp, data = airquality)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -40.922 -17.459 -0.874 10.444 118.078   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -147.6461 18.7553 -7.872 2.76e-12 \*\*\*  
## Temp 2.4391 0.2393 10.192 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 23.92 on 109 degrees of freedom  
## Multiple R-squared: 0.488, Adjusted R-squared: 0.4833   
## F-statistic: 103.9 on 1 and 109 DF, p-value: < 2.2e-16

# Load the airquality dataset  
data(airquality)  
  
# Fit linear regression models  
model1 <- lm(Ozone ~ Solar.R, data = airquality)  
model2 <- lm(Ozone ~ Wind, data = airquality)  
model3 <- lm(Ozone ~ Temp, data = airquality)  
  
# Display model summaries  
summary(model1)

##   
## Call:  
## lm(formula = Ozone ~ Solar.R, data = airquality)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -48.292 -21.361 -8.864 16.373 119.136   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 18.59873 6.74790 2.756 0.006856 \*\*   
## Solar.R 0.12717 0.03278 3.880 0.000179 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 31.33 on 109 degrees of freedom  
## (42 observations deleted due to missingness)  
## Multiple R-squared: 0.1213, Adjusted R-squared: 0.1133   
## F-statistic: 15.05 on 1 and 109 DF, p-value: 0.0001793

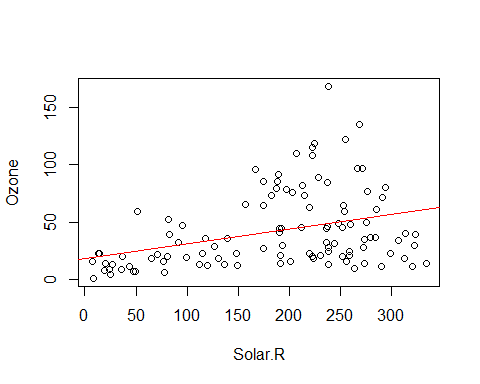
summary(model2)

##   
## Call:  
## lm(formula = Ozone ~ Wind, data = airquality)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -51.572 -18.854 -4.868 15.234 90.000   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 96.8729 7.2387 13.38 < 2e-16 \*\*\*  
## Wind -5.5509 0.6904 -8.04 9.27e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 26.47 on 114 degrees of freedom  
## (37 observations deleted due to missingness)  
## Multiple R-squared: 0.3619, Adjusted R-squared: 0.3563   
## F-statistic: 64.64 on 1 and 114 DF, p-value: 9.272e-13

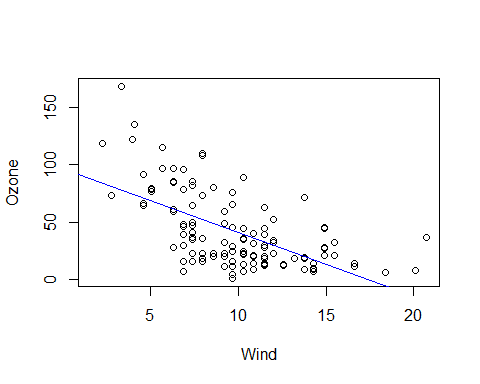
summary(model3)

##   
## Call:  
## lm(formula = Ozone ~ Temp, data = airquality)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -40.729 -17.409 -0.587 11.306 118.271   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -146.9955 18.2872 -8.038 9.37e-13 \*\*\*  
## Temp 2.4287 0.2331 10.418 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 23.71 on 114 degrees of freedom  
## (37 observations deleted due to missingness)  
## Multiple R-squared: 0.4877, Adjusted R-squared: 0.4832   
## F-statistic: 108.5 on 1 and 114 DF, p-value: < 2.2e-16

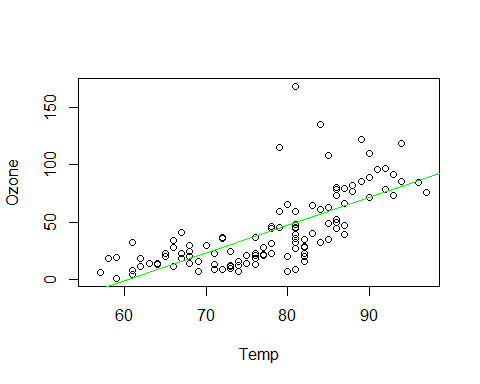
# Load the airquality dataset  
data(airquality)  
  
# Fit linear regression models  
model1 <- lm(Ozone ~ Solar.R, data = airquality)  
model2 <- lm(Ozone ~ Wind, data = airquality)  
model3 <- lm(Ozone ~ Temp, data = airquality)  
  
# Visualize the models  
plot(Ozone ~ Solar.R, data = airquality)  
abline(model1, col = "red")



plot(Ozone ~ Wind, data = airquality)  
abline(model2, col = "blue")



plot(Ozone ~ Temp, data = airquality)  
abline(model3, col = "green")



#Another study in R concerning with the dataset cancer  
#First of everything load the library that contains the dataset to enhance the smooth running of the workouts  
library(survival)  
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.2.3

##   
## 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

str(lung)

## 'data.frame': 228 obs. of 10 variables:  
## $ inst : num 3 3 3 5 1 12 7 11 1 7 ...  
## $ time : num 306 455 1010 210 883 ...  
## $ status : num 2 2 1 2 2 1 2 2 2 2 ...  
## $ age : num 74 68 56 57 60 74 68 71 53 61 ...  
## $ sex : num 1 1 1 1 1 1 2 2 1 1 ...  
## $ ph.ecog : num 1 0 0 1 0 1 2 2 1 2 ...  
## $ ph.karno : num 90 90 90 90 100 50 70 60 70 70 ...  
## $ pat.karno: num 100 90 90 60 90 80 60 80 80 70 ...  
## $ meal.cal : num 1175 1225 NA 1150 NA ...  
## $ wt.loss : num NA 15 15 11 0 0 10 1 16 34 ...

#Secondly load the dataset itself to portray the dataset content  
data("cancer")  
#Type the view command to see the dataset composition  
View(cancer)  
#Get the summary of the dataset to better comprehend what the dataset entails  
summary(cancer)

## inst time status age   
## Min. : 1.00 Min. : 5.0 Min. :1.000 Min. :39.00   
## 1st Qu.: 3.00 1st Qu.: 166.8 1st Qu.:1.000 1st Qu.:56.00   
## Median :11.00 Median : 255.5 Median :2.000 Median :63.00   
## Mean :11.09 Mean : 305.2 Mean :1.724 Mean :62.45   
## 3rd Qu.:16.00 3rd Qu.: 396.5 3rd Qu.:2.000 3rd Qu.:69.00   
## Max. :33.00 Max. :1022.0 Max. :2.000 Max. :82.00   
## NA's :1   
## sex ph.ecog ph.karno pat.karno   
## Min. :1.000 Min. :0.0000 Min. : 50.00 Min. : 30.00   
## 1st Qu.:1.000 1st Qu.:0.0000 1st Qu.: 75.00 1st Qu.: 70.00   
## Median :1.000 Median :1.0000 Median : 80.00 Median : 80.00   
## Mean :1.395 Mean :0.9515 Mean : 81.94 Mean : 79.96   
## 3rd Qu.:2.000 3rd Qu.:1.0000 3rd Qu.: 90.00 3rd Qu.: 90.00   
## Max. :2.000 Max. :3.0000 Max. :100.00 Max. :100.00   
## NA's :1 NA's :1 NA's :3   
## meal.cal wt.loss   
## Min. : 96.0 Min. :-24.000   
## 1st Qu.: 635.0 1st Qu.: 0.000   
## Median : 975.0 Median : 7.000   
## Mean : 928.8 Mean : 9.832   
## 3rd Qu.:1150.0 3rd Qu.: 15.750   
## Max. :2600.0 Max. : 68.000   
## NA's :47 NA's :14

#Get the structure of the dataset to better comprehend what kind of data types the dataset contains  
str(cancer)

## 'data.frame': 228 obs. of 10 variables:  
## $ inst : num 3 3 3 5 1 12 7 11 1 7 ...  
## $ time : num 306 455 1010 210 883 ...  
## $ status : num 2 2 1 2 2 1 2 2 2 2 ...  
## $ age : num 74 68 56 57 60 74 68 71 53 61 ...  
## $ sex : num 1 1 1 1 1 1 2 2 1 1 ...  
## $ ph.ecog : num 1 0 0 1 0 1 2 2 1 2 ...  
## $ ph.karno : num 90 90 90 90 100 50 70 60 70 70 ...  
## $ pat.karno: num 100 90 90 60 90 80 60 80 80 70 ...  
## $ meal.cal : num 1175 1225 NA 1150 NA ...  
## $ wt.loss : num NA 15 15 11 0 0 10 1 16 34 ...

#Getting the headings of the dataset in concern  
names(cancer)

## [1] "inst" "time" "status" "age" "sex" "ph.ecog"   
## [7] "ph.karno" "pat.karno" "meal.cal" "wt.loss"

#Getting the summation of the certain columns in concern  
sum(cancer$age)

## [1] 14238

sum(cancer$time)

## [1] 69593

sum(cancer$status)

## [1] 393

sum(cancer$inst)

## [1] NA

sum(cancer$sex)

## [1] 318

sum(cancer$ph.ecog)

## [1] NA

sum(cancer$ph.karno)

## [1] NA

#Attaching the dataset  
attach(cancer)  
detach(cancer)  
#Doing some calculations on the dataset of concern  
mean(cancer$time)

## [1] 305.2325

median(cancer$time)

## [1] 255.5

mode(cancer$time)

## [1] "numeric"

mode(cancer$age)

## [1] "numeric"

#Getting the number of columns of the old cancer dataset  
ncol(cancer)

## [1] 10

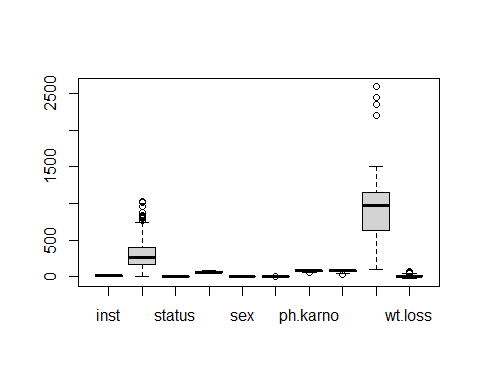
#Getting the number of rows of the old cancer dataset  
nrow(cancer)

## [1] 228

#Get the total number of missing values in the dataset of concern  
sum(is.na(cancer))

## [1] 67

#Workout the outliers concept using the boxplot idea  
boxplot(cancer)  
summary(boxplot(cancer))

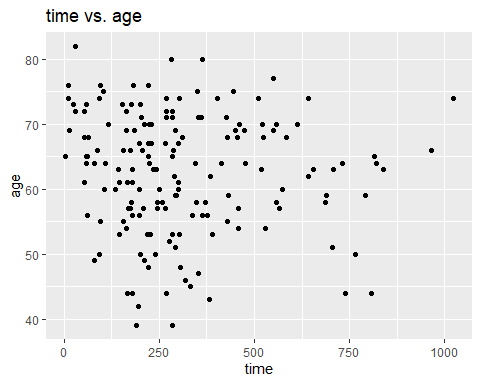


## Length Class Mode   
## stats 50 -none- numeric   
## n 10 -none- numeric   
## conf 20 -none- numeric   
## out 28 -none- numeric   
## group 28 -none- numeric   
## names 10 -none- character

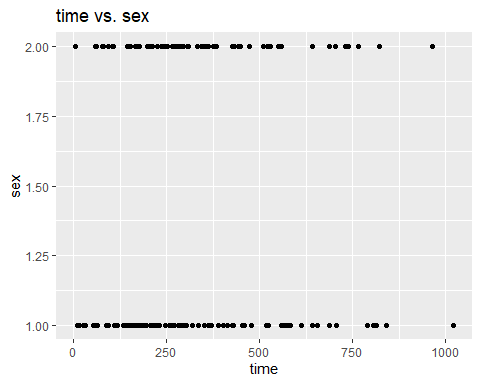
#Do away with the missing values by deletion of the NAS by omit concept  
cancer <- na.omit(cancer)  
#Do away with the missing values with the use of the tidyr library and the drop concept  
library(tidyr)

## Warning: package 'tidyr' was built under R version 4.2.3

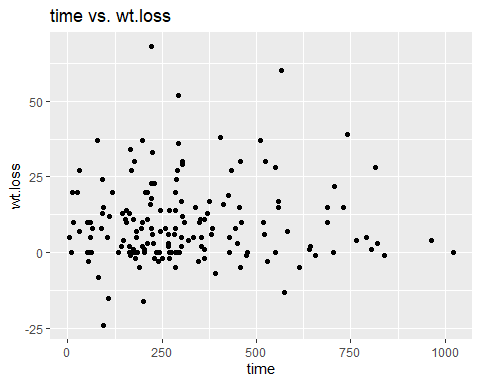
#Creation of the new dataset from the old dataset that has no missing values  
#Data visualization using the scatter plots  
# Load ggplot2 library  
library(ggplot2)  
# Scatterplot of time and age  
ggplot(cancer, aes(x = time, y = age)) +   
 geom\_point() +   
 labs(title = "time vs. age", x = "time", y = "age")



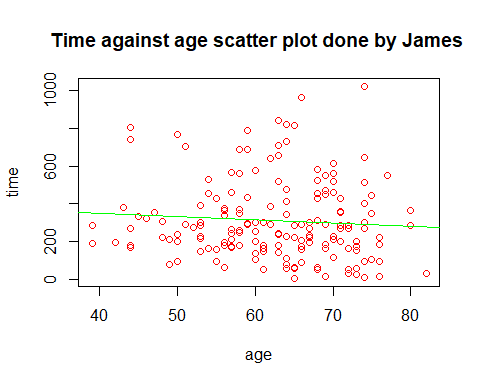
# Scatterplot of time and sex  
ggplot(cancer, aes(x = time, y = sex)) +   
 geom\_point() +   
 labs(title = "time vs. sex", x = "time", y = "sex")



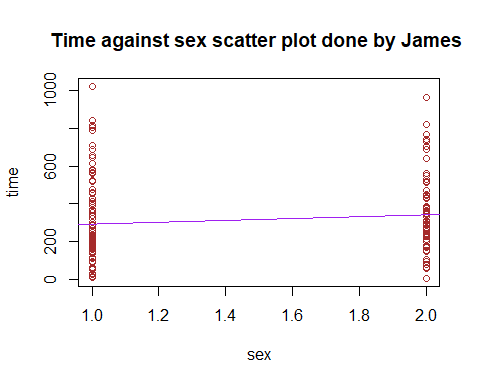
# Scatterplot of time and wt.loss  
ggplot(cancer, aes(x = time, y = wt.loss)) +   
 geom\_point() +   
 labs(title = "time vs. wt.loss", x = "time", y = "wt.loss")



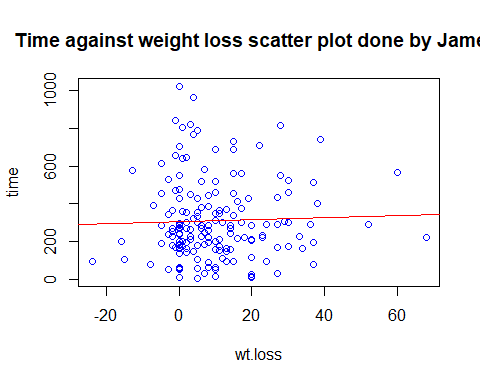
#Data modeling using the linear regression models  
# Fit linear regression models  
model1 <- lm(time ~ age, data = cancer)  
model2 <- lm(time ~ sex, data = cancer)  
model3 <- lm(time ~ wt.loss, data = cancer)  
# Visualize the models  
plot(time ~ age, data = cancer , main = "Time against age scatter plot done by James ", col = "red")  
abline(model1, col = "green")



plot(time ~ sex, data = cancer , main = "Time against sex scatter plot done by James ", col = "brown")  
abline(model2, col = "purple")



plot(time ~ wt.loss, data = cancer , main = "Time against weight loss scatter plot done by James ", col = "blue")  
abline(model3, col = "red")



# Display model summaries  
summary(model1)

##   
## Call:  
## lm(formula = time ~ age, data = cancer)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -300.59 -143.27 -53.88 110.05 732.48   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 421.676 111.598 3.779 0.00022 \*\*\*  
## age -1.786 1.765 -1.012 0.31301   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 209.4 on 165 degrees of freedom  
## Multiple R-squared: 0.006169, Adjusted R-squared: 0.0001456   
## F-statistic: 1.024 on 1 and 165 DF, p-value: 0.313

summary(model2)

##   
## Call:  
## lm(formula = time ~ sex, data = cancer)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -335.17 -132.16 -55.17 101.34 730.85   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 242.12 48.70 4.971 1.65e-06 \*\*\*  
## sex 49.03 33.22 1.476 0.142   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 208.7 on 165 degrees of freedom  
## Multiple R-squared: 0.01303, Adjusted R-squared: 0.007048   
## F-statistic: 2.178 on 1 and 165 DF, p-value: 0.1419

summary(model3)

##   
## Call:  
## lm(formula = time ~ wt.loss, data = cancer)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -303.31 -136.47 -40.42 105.49 717.15   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 304.8488 20.1016 15.16 <2e-16 \*\*\*  
## wt.loss 0.5233 1.2180 0.43 0.668   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 210 on 165 degrees of freedom  
## Multiple R-squared: 0.001117, Adjusted R-squared: -0.004937   
## F-statistic: 0.1846 on 1 and 165 DF, p-value: 0.6681

#Observation of the first and last six rows and columns of the new cancer dataset  
#Observation of the firs and last six rows and columns of the old cancer dataset  
head(cancer)

## inst time status age sex ph.ecog ph.karno pat.karno meal.cal wt.loss  
## 2 3 455 2 68 1 0 90 90 1225 15  
## 4 5 210 2 57 1 1 90 60 1150 11  
## 6 12 1022 1 74 1 1 50 80 513 0  
## 7 7 310 2 68 2 2 70 60 384 10  
## 8 11 361 2 71 2 2 60 80 538 1  
## 9 1 218 2 53 1 1 70 80 825 16

tail(cancer)

## inst time status age sex ph.ecog ph.karno pat.karno meal.cal wt.loss  
## 221 22 197 1 67 1 1 80 90 1500 2  
## 222 11 203 1 71 2 1 80 90 1025 0  
## 225 13 191 1 39 1 0 90 90 2350 -5  
## 226 32 105 1 75 2 2 60 70 1025 5  
## 227 6 174 1 66 1 1 90 100 1075 1  
## 228 22 177 1 58 2 1 80 90 1060 0

#Another study in R  
data()  
library(survival)  
data("cancer")  
names(cancer)

## [1] "inst" "time" "status" "age" "sex" "ph.ecog"   
## [7] "ph.karno" "pat.karno" "meal.cal" "wt.loss"

sum(is.na(cancer))

## [1] 67

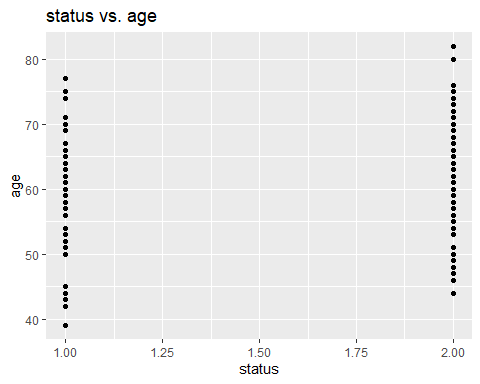
library(tidyr)  
cancer <- na.omit(cancer)  
nrow(cancer)

## [1] 167

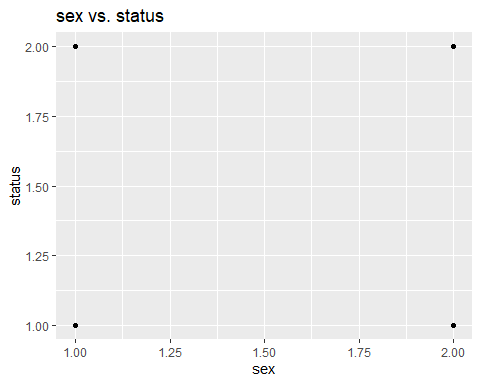
ncol(cancer)

## [1] 10

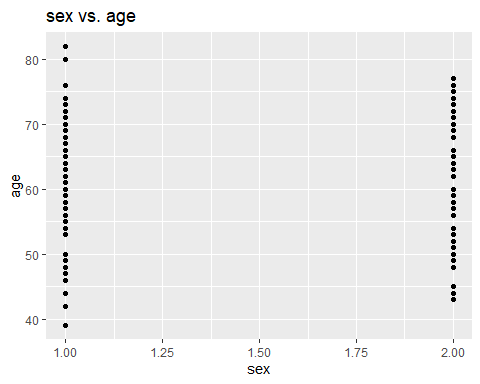
# Scatter plot of age and circumference  
ggplot(cancer, aes(x = status, y = age)) +   
 geom\_point() +   
 labs(title = "status vs. age", x = "status", y = "age")



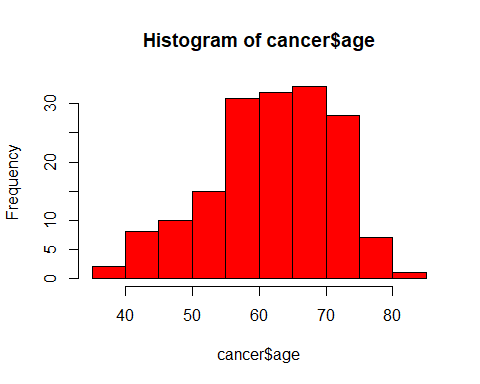
# Scatter plot of Tree and age  
ggplot(cancer, aes(x = sex, y = status)) +   
 geom\_point() +   
 labs(title = "sex vs. status", x = "sex", y = "status")



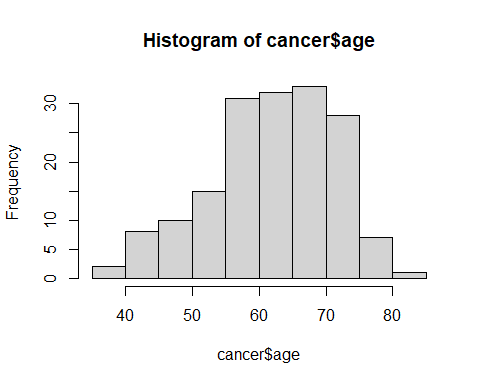
# Scatter plot of sex and age  
ggplot(cancer, aes(x = sex, y = age)) +   
 geom\_point() +   
 labs(title = "sex vs. age", x = "sex", y = "age")



#Drawing of the histogram cancer and age  
hist(cancer$age , col = "red")

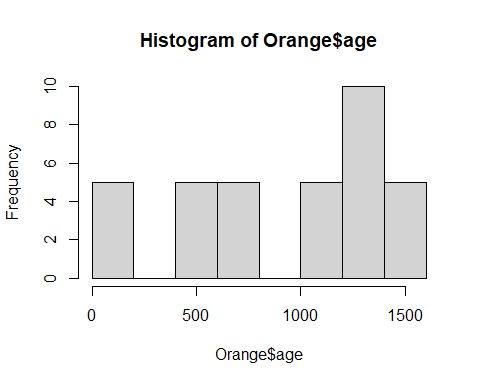


#Getting the summary of the output  
summary(hist(cancer$age))



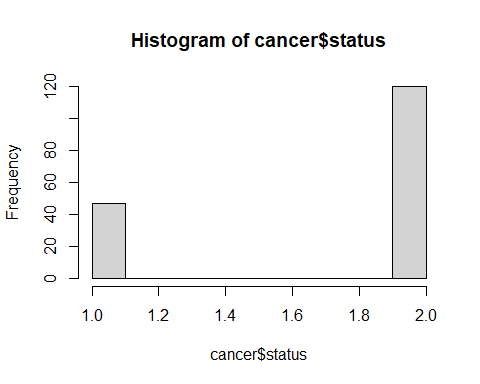
## Length Class Mode   
## breaks 11 -none- numeric   
## counts 10 -none- numeric   
## density 10 -none- numeric   
## mids 10 -none- numeric   
## xname 1 -none- character  
## equidist 1 -none- logical

#Getting the structure of the summary  
str(summary(hist(Orange$age)))



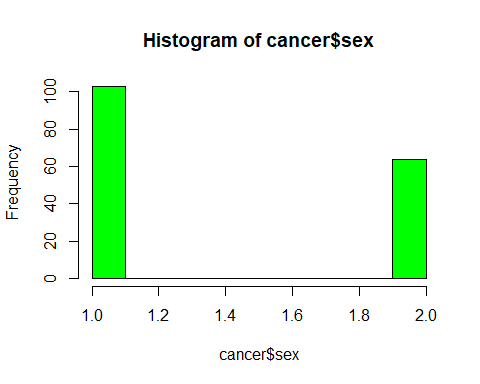
## 'summaryDefault' chr [1:6, 1:3] "9" "8" "8" "8" "1" "1" "-none-" "-none-" ...  
## - attr(\*, "dimnames")=List of 2  
## ..$ : chr [1:6] "breaks" "counts" "density" "mids" ...  
## ..$ : chr [1:3] "Length" "Class" "Mode"

str(hist(cancer$status))



## List of 6  
## $ breaks : num [1:11] 1 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 ...  
## $ counts : int [1:10] 47 0 0 0 0 0 0 0 0 120  
## $ density : num [1:10] 2.81 0 0 0 0 ...  
## $ mids : num [1:10] 1.05 1.15 1.25 1.35 1.45 1.55 1.65 1.75 1.85 1.95  
## $ xname : chr "cancer$status"  
## $ equidist: logi TRUE  
## - attr(\*, "class")= chr "histogram"

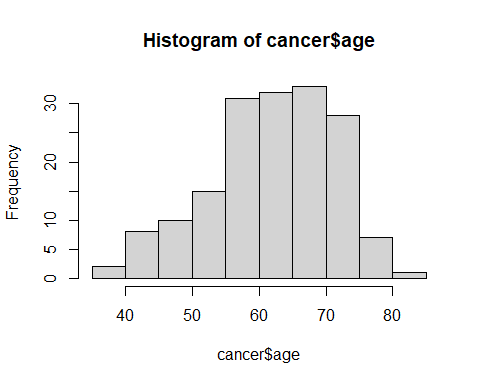
#Drawing of the histogram of the cancer and sex  
hist(cancer$sex , col = "green")



#Getting the summary of the sex n relation to cancer  
summary(hist(cancer$age))

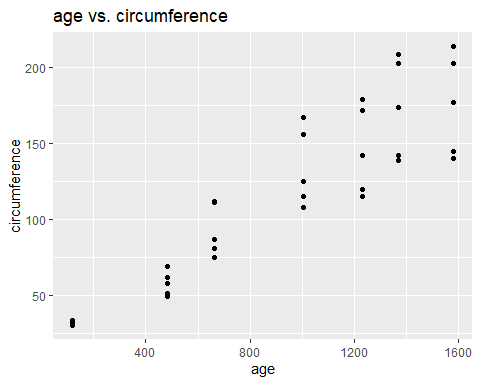
## Length Class Mode   
## breaks 11 -none- numeric   
## counts 10 -none- numeric   
## density 10 -none- numeric   
## mids 10 -none- numeric   
## xname 1 -none- character  
## equidist 1 -none- logical

#Getting the structure of the summary  
str(hist(cancer$age))



## List of 6  
## $ breaks : int [1:11] 35 40 45 50 55 60 65 70 75 80 ...  
## $ counts : int [1:10] 2 8 10 15 31 32 33 28 7 1  
## $ density : num [1:10] 0.0024 0.00958 0.01198 0.01796 0.03713 ...  
## $ mids : num [1:10] 37.5 42.5 47.5 52.5 57.5 62.5 67.5 72.5 77.5 82.5  
## $ xname : chr "cancer$age"  
## $ equidist: logi TRUE  
## - attr(\*, "class")= chr "histogram"

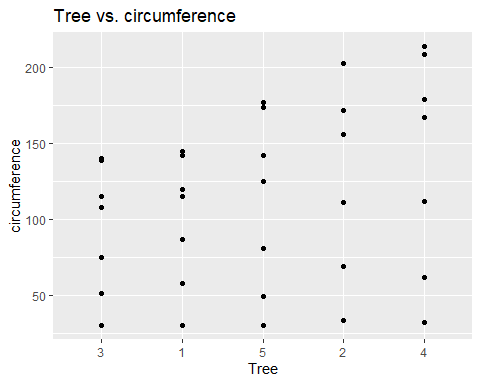
#Dealing with the oranges dataset  
data()  
data("Orange")  
# Scatter plot of age and circumference  
ggplot(Orange, aes(x = age, y = circumference)) +   
 geom\_point() +   
 labs(title = "age vs. circumference", x = "age", y = "circumference")



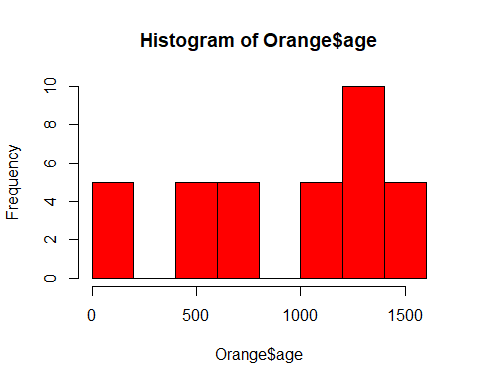
# Scatter plot of Tree and age  
ggplot(Orange, aes(x = Tree, y = age)) +   
 geom\_point() +   
 labs(title = "Tree vs. age", x = "Tree", y = "age")



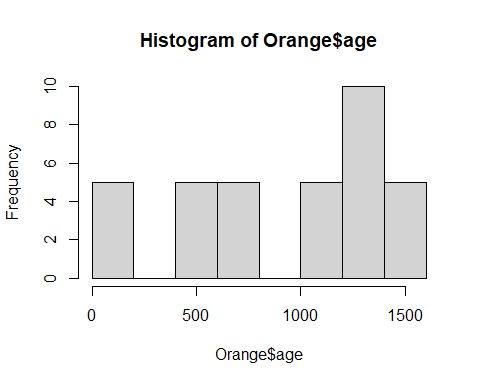
# Scatter plot of Tree and circumference  
ggplot(Orange, aes(x = Tree, y = circumference)) +   
 geom\_point() +   
 labs(title = "Tree vs. circumference", x = "Tree", y = "circumference")



#Drawing of the histogram orange and age  
hist(Orange$age , col = "red")



#Getting the summary of the histogram  
summary(hist(Orange$age))



## Length Class Mode   
## breaks 9 -none- numeric   
## counts 8 -none- numeric   
## density 8 -none- numeric   
## mids 8 -none- numeric   
## xname 1 -none- character  
## equidist 1 -none- logical

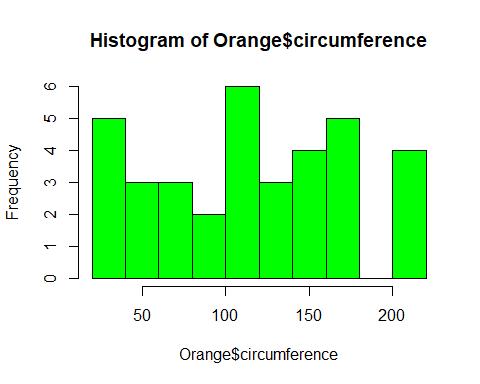
#Getting the structure of the summary  
str(summary(hist(Orange$age)))

## 'summaryDefault' chr [1:6, 1:3] "9" "8" "8" "8" "1" "1" "-none-" "-none-" ...  
## - attr(\*, "dimnames")=List of 2  
## ..$ : chr [1:6] "breaks" "counts" "density" "mids" ...  
## ..$ : chr [1:3] "Length" "Class" "Mode"

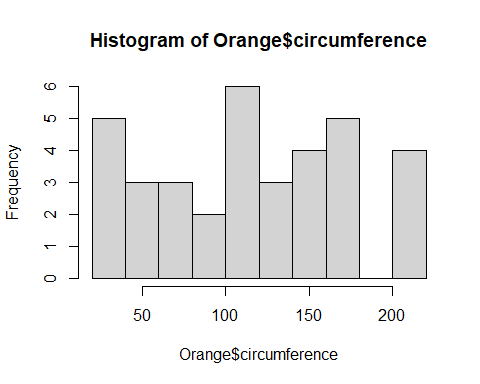
str(hist(Orange$age))

## List of 6  
## $ breaks : num [1:9] 0 200 400 600 800 1000 1200 1400 1600  
## $ counts : int [1:8] 5 0 5 5 0 5 10 5  
## $ density : num [1:8] 0.000714 0 0.000714 0.000714 0 ...  
## $ mids : num [1:8] 100 300 500 700 900 1100 1300 1500  
## $ xname : chr "Orange$age"  
## $ equidist: logi TRUE  
## - attr(\*, "class")= chr "histogram"

#Drawing of the histogram of oranges and circumference  
hist(Orange$circumference , col = "green")



#Getting the summary of the histogram  
summary(hist(Orange$circumference))



## Length Class Mode   
## breaks 11 -none- numeric   
## counts 10 -none- numeric   
## density 10 -none- numeric   
## mids 10 -none- numeric   
## xname 1 -none- character  
## equidist 1 -none- logical

#Getting the structure of the summary  
str(hist(Orange$circumference))

## List of 6  
## $ breaks : int [1:11] 20 40 60 80 100 120 140 160 180 200 ...  
## $ counts : int [1:10] 5 3 3 2 6 3 4 5 0 4  
## $ density : num [1:10] 0.00714 0.00429 0.00429 0.00286 0.00857 ...  
## $ mids : num [1:10] 30 50 70 90 110 130 150 170 190 210  
## $ xname : chr "Orange$circumference"  
## $ equidist: logi TRUE  
## - attr(\*, "class")= chr "histogram"

#This marks the end of today's study