Iris\_Dataset.R

Innova

Tue Jan 17 11:37:04 2017

library(ggplot2)  
library(reshape2)  
library(plotly)

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:plotly':  
##   
## select

library(caret)

## Loading required package: lattice

#Iris Dataset  
head(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 2 4.9 3.0 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 4 4.6 3.1 1.5 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 6 5.4 3.9 1.7 0.4 setosa

ls.str(iris)

## Petal.Length : num [1:150] 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
## Petal.Width : num [1:150] 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
## Sepal.Length : num [1:150] 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
## Sepal.Width : num [1:150] 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
## Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...

levels(iris$Species)

## [1] "setosa" "versicolor" "virginica"

summary(iris)

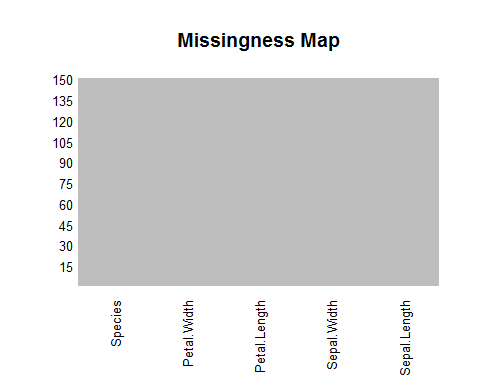
## Sepal.Length Sepal.Width Petal.Length Petal.Width   
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100   
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300   
## Median :5.800 Median :3.000 Median :4.350 Median :1.300   
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199   
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800   
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500   
## Species   
## setosa :50   
## versicolor:50   
## virginica :50   
##   
##   
##

library(Amelia)

## Loading required package: Rcpp

## ##   
## ## Amelia II: Multiple Imputation  
## ## (Version 1.7.4, built: 2015-12-05)  
## ## Copyright (C) 2005-2017 James Honaker, Gary King and Matthew Blackwell  
## ## Refer to http://gking.harvard.edu/amelia/ for more information  
## ##

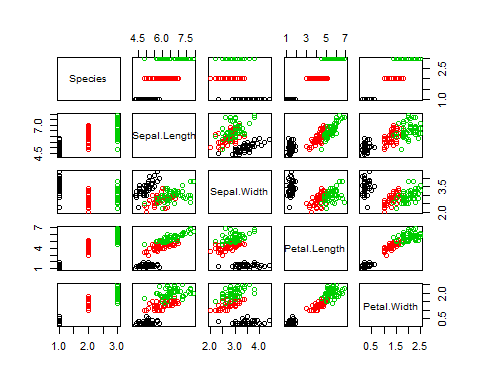
library(mlbench)  
missmap(iris, col=c("black", "grey"), legend=FALSE)



# pair-wise scatterplots of all 5 attributes  
pairs(iris)



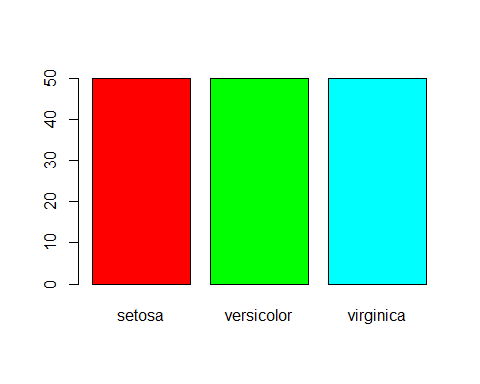
pairs(Species~., data=iris, col=iris$Species)



Species =iris$Species  
Species.freq = table(Species) # apply the table function   
Species.freq

## Species  
## setosa versicolor virginica   
## 50 50 50

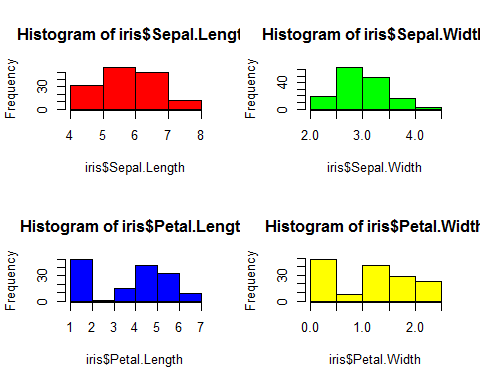
barplot(Species.freq, col = c("red","green","cyan"))



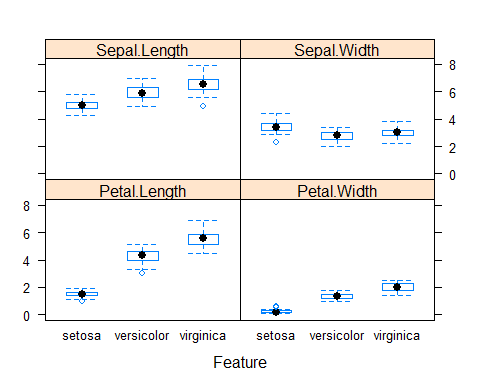
Species.relfreq = Species.freq / nrow(iris)  
Species.relfreq

## Species  
## setosa versicolor virginica   
## 0.3333333 0.3333333 0.3333333

par(mfrow=c(2,2))  
hist(iris$Sepal.Length,col = "red",breaks = 5)  
hist(iris$Sepal.Width,col = "green", breaks = 5)  
hist(iris$Petal.Length,col = "blue",breaks = 6)  
hist(iris$Petal.Width,col = "yellow", breaks = 5)



# box and whisker plots for each attribute by class value  
x <- iris[,1:4]  
y <- iris[,5]  
featurePlot(x=x, y=y, plot="box")



Sepal.Length = iris$Sepal.Length # the sepal length interval   
Sepal.Width = iris$Sepal.Width # the sepal width interval   
Petal.Length = iris$Petal.Length # the Petal length interval   
Petal.Width = iris$Petal.Width # the Petal width interval   
  
plot(Sepal.Length, Sepal.Width) #apply the scatter function  
  
abline(lm(Sepal.Width ~ Sepal.Length)) #generate a linear regression model of the two variables  
  
plot(Petal.Length, Petal.Width )  
abline(lm( Petal.Width ~ Petal.Length)) #generate a linear regression model of the two variables  
  
  
library(survival)

##   
## Attaching package: 'survival'

## The following object is masked from 'package:caret':  
##   
## cluster

data("iris")  
model1 <- glm(Species ~ Petal.Length + Petal.Width +Sepal.Length+Sepal.Width,  
 data = iris, family = binomial())

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(model1)

##   
## Call:  
## glm(formula = Species ~ Petal.Length + Petal.Width + Sepal.Length +   
## Sepal.Width, family = binomial(), data = iris)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.173e-05 -2.100e-08 2.100e-08 2.100e-08 3.185e-05   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 16.946 457457.097 0 1  
## Petal.Length 20.088 107724.589 0 1  
## Petal.Width 21.608 154350.604 0 1  
## Sepal.Length -11.759 130504.037 0 1  
## Sepal.Width -7.842 59415.373 0 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1.9095e+02 on 149 degrees of freedom  
## Residual deviance: 3.2940e-09 on 145 degrees of freedom  
## AIC: 10  
##   
## Number of Fisher Scoring iterations: 25

sp <- ggplot(iris, aes(x=Sepal.Length, y=Sepal.Width)) + geom\_point(shape=2)  
print(sp)  
  
sp1 <- ggplot(iris, aes(x=Petal.Length, y=Petal.Width)) + geom\_point(shape=3)  
print(sp1)  
sp + facet\_grid(Species ~.) #Divide by levels of "Species", in the vertical direction  
  
sp1 + facet\_grid(.~ Species) #Divide by levels of "Species", in the horizontal direction  
  
  
#Modifying facet label appearance  
sp + facet\_grid(.~Species) +  
 theme(strip.text.x = element\_text(size=12, angle=30, face="bold"),  
 strip.text.y = element\_text(size=12, angle=60, face="bold"),  
 strip.background = element\_rect(colour="red", fill="yellow"))

