D02\_teacher.R

B0113

2020-01-08

data(iris)  
data(state)  
class(iris)

## [1] "data.frame"

class(state.x77)

## [1] "matrix"

str(state.x77)

## num [1:50, 1:8] 3615 365 2212 2110 21198 ...  
## - attr(\*, "dimnames")=List of 2  
## ..$ : chr [1:50] "Alabama" "Alaska" "Arizona" "Arkansas" ...  
## ..$ : chr [1:8] "Population" "Income" "Illiteracy" "Life Exp" ...

dim(state.x77)

## [1] 50 8

colnames(state.x77)

## [1] "Population" "Income" "Illiteracy" "Life Exp" "Murder"   
## [6] "HS Grad" "Frost" "Area"

rownames(state.x77)

## [1] "Alabama" "Alaska" "Arizona" "Arkansas"   
## [5] "California" "Colorado" "Connecticut" "Delaware"   
## [9] "Florida" "Georgia" "Hawaii" "Idaho"   
## [13] "Illinois" "Indiana" "Iowa" "Kansas"   
## [17] "Kentucky" "Louisiana" "Maine" "Maryland"   
## [21] "Massachusetts" "Michigan" "Minnesota" "Mississippi"   
## [25] "Missouri" "Montana" "Nebraska" "Nevada"   
## [29] "New Hampshire" "New Jersey" "New Mexico" "New York"   
## [33] "North Carolina" "North Dakota" "Ohio" "Oklahoma"   
## [37] "Oregon" "Pennsylvania" "Rhode Island" "South Carolina"  
## [41] "South Dakota" "Tennessee" "Texas" "Utah"   
## [45] "Vermont" "Virginia" "Washington" "West Virginia"   
## [49] "Wisconsin" "Wyoming"

df <- data.frame(state.x77)  
str(df)

## 'data.frame': 50 obs. of 8 variables:  
## $ Population: num 3615 365 2212 2110 21198 ...  
## $ Income : num 3624 6315 4530 3378 5114 ...  
## $ Illiteracy: num 2.1 1.5 1.8 1.9 1.1 0.7 1.1 0.9 1.3 2 ...  
## $ Life.Exp : num 69 69.3 70.5 70.7 71.7 ...  
## $ Murder : num 15.1 11.3 7.8 10.1 10.3 6.8 3.1 6.2 10.7 13.9 ...  
## $ HS.Grad : num 41.3 66.7 58.1 39.9 62.6 63.9 56 54.6 52.6 40.6 ...  
## $ Frost : num 20 152 15 65 20 166 139 103 11 60 ...  
## $ Area : num 50708 566432 113417 51945 156361 ...

str(iris)

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

mtx <- as.matrix(iris[, -5])  
mtx

## Sepal.Length Sepal.Width Petal.Length Petal.Width  
## [1,] 5.1 3.5 1.4 0.2  
## [2,] 4.9 3.0 1.4 0.2  
## [3,] 4.7 3.2 1.3 0.2  
## [4,] 4.6 3.1 1.5 0.2  
## [5,] 5.0 3.6 1.4 0.2  
## [6,] 5.4 3.9 1.7 0.4  
## [7,] 4.6 3.4 1.4 0.3  
## [8,] 5.0 3.4 1.5 0.2  
## [9,] 4.4 2.9 1.4 0.2  
## [10,] 4.9 3.1 1.5 0.1  
## [11,] 5.4 3.7 1.5 0.2  
## [12,] 4.8 3.4 1.6 0.2  
## [13,] 4.8 3.0 1.4 0.1  
## [14,] 4.3 3.0 1.1 0.1  
## [15,] 5.8 4.0 1.2 0.2  
## [16,] 5.7 4.4 1.5 0.4  
## [17,] 5.4 3.9 1.3 0.4  
## [18,] 5.1 3.5 1.4 0.3  
## [19,] 5.7 3.8 1.7 0.3  
## [20,] 5.1 3.8 1.5 0.3  
## [21,] 5.4 3.4 1.7 0.2  
## [22,] 5.1 3.7 1.5 0.4  
## [23,] 4.6 3.6 1.0 0.2  
## [24,] 5.1 3.3 1.7 0.5  
## [25,] 4.8 3.4 1.9 0.2  
## [26,] 5.0 3.0 1.6 0.2  
## [27,] 5.0 3.4 1.6 0.4  
## [28,] 5.2 3.5 1.5 0.2  
## [29,] 5.2 3.4 1.4 0.2  
## [30,] 4.7 3.2 1.6 0.2  
## [31,] 4.8 3.1 1.6 0.2  
## [32,] 5.4 3.4 1.5 0.4  
## [33,] 5.2 4.1 1.5 0.1  
## [34,] 5.5 4.2 1.4 0.2  
## [35,] 4.9 3.1 1.5 0.2  
## [36,] 5.0 3.2 1.2 0.2  
## [37,] 5.5 3.5 1.3 0.2  
## [38,] 4.9 3.6 1.4 0.1  
## [39,] 4.4 3.0 1.3 0.2  
## [40,] 5.1 3.4 1.5 0.2  
## [41,] 5.0 3.5 1.3 0.3  
## [42,] 4.5 2.3 1.3 0.3  
## [43,] 4.4 3.2 1.3 0.2  
## [44,] 5.0 3.5 1.6 0.6  
## [45,] 5.1 3.8 1.9 0.4  
## [46,] 4.8 3.0 1.4 0.3  
## [47,] 5.1 3.8 1.6 0.2  
## [48,] 4.6 3.2 1.4 0.2  
## [49,] 5.3 3.7 1.5 0.2  
## [50,] 5.0 3.3 1.4 0.2  
## [51,] 7.0 3.2 4.7 1.4  
## [52,] 6.4 3.2 4.5 1.5  
## [53,] 6.9 3.1 4.9 1.5  
## [54,] 5.5 2.3 4.0 1.3  
## [55,] 6.5 2.8 4.6 1.5  
## [56,] 5.7 2.8 4.5 1.3  
## [57,] 6.3 3.3 4.7 1.6  
## [58,] 4.9 2.4 3.3 1.0  
## [59,] 6.6 2.9 4.6 1.3  
## [60,] 5.2 2.7 3.9 1.4  
## [61,] 5.0 2.0 3.5 1.0  
## [62,] 5.9 3.0 4.2 1.5  
## [63,] 6.0 2.2 4.0 1.0  
## [64,] 6.1 2.9 4.7 1.4  
## [65,] 5.6 2.9 3.6 1.3  
## [66,] 6.7 3.1 4.4 1.4  
## [67,] 5.6 3.0 4.5 1.5  
## [68,] 5.8 2.7 4.1 1.0  
## [69,] 6.2 2.2 4.5 1.5  
## [70,] 5.6 2.5 3.9 1.1  
## [71,] 5.9 3.2 4.8 1.8  
## [72,] 6.1 2.8 4.0 1.3  
## [73,] 6.3 2.5 4.9 1.5  
## [74,] 6.1 2.8 4.7 1.2  
## [75,] 6.4 2.9 4.3 1.3  
## [76,] 6.6 3.0 4.4 1.4  
## [77,] 6.8 2.8 4.8 1.4  
## [78,] 6.7 3.0 5.0 1.7  
## [79,] 6.0 2.9 4.5 1.5  
## [80,] 5.7 2.6 3.5 1.0  
## [81,] 5.5 2.4 3.8 1.1  
## [82,] 5.5 2.4 3.7 1.0  
## [83,] 5.8 2.7 3.9 1.2  
## [84,] 6.0 2.7 5.1 1.6  
## [85,] 5.4 3.0 4.5 1.5  
## [86,] 6.0 3.4 4.5 1.6  
## [87,] 6.7 3.1 4.7 1.5  
## [88,] 6.3 2.3 4.4 1.3  
## [89,] 5.6 3.0 4.1 1.3  
## [90,] 5.5 2.5 4.0 1.3  
## [91,] 5.5 2.6 4.4 1.2  
## [92,] 6.1 3.0 4.6 1.4  
## [93,] 5.8 2.6 4.0 1.2  
## [94,] 5.0 2.3 3.3 1.0  
## [95,] 5.6 2.7 4.2 1.3  
## [96,] 5.7 3.0 4.2 1.2  
## [97,] 5.7 2.9 4.2 1.3  
## [98,] 6.2 2.9 4.3 1.3  
## [99,] 5.1 2.5 3.0 1.1  
## [100,] 5.7 2.8 4.1 1.3  
## [101,] 6.3 3.3 6.0 2.5  
## [102,] 5.8 2.7 5.1 1.9  
## [103,] 7.1 3.0 5.9 2.1  
## [104,] 6.3 2.9 5.6 1.8  
## [105,] 6.5 3.0 5.8 2.2  
## [106,] 7.6 3.0 6.6 2.1  
## [107,] 4.9 2.5 4.5 1.7  
## [108,] 7.3 2.9 6.3 1.8  
## [109,] 6.7 2.5 5.8 1.8  
## [110,] 7.2 3.6 6.1 2.5  
## [111,] 6.5 3.2 5.1 2.0  
## [112,] 6.4 2.7 5.3 1.9  
## [113,] 6.8 3.0 5.5 2.1  
## [114,] 5.7 2.5 5.0 2.0  
## [115,] 5.8 2.8 5.1 2.4  
## [116,] 6.4 3.2 5.3 2.3  
## [117,] 6.5 3.0 5.5 1.8  
## [118,] 7.7 3.8 6.7 2.2  
## [119,] 7.7 2.6 6.9 2.3  
## [120,] 6.0 2.2 5.0 1.5  
## [121,] 6.9 3.2 5.7 2.3  
## [122,] 5.6 2.8 4.9 2.0  
## [123,] 7.7 2.8 6.7 2.0  
## [124,] 6.3 2.7 4.9 1.8  
## [125,] 6.7 3.3 5.7 2.1  
## [126,] 7.2 3.2 6.0 1.8  
## [127,] 6.2 2.8 4.8 1.8  
## [128,] 6.1 3.0 4.9 1.8  
## [129,] 6.4 2.8 5.6 2.1  
## [130,] 7.2 3.0 5.8 1.6  
## [131,] 7.4 2.8 6.1 1.9  
## [132,] 7.9 3.8 6.4 2.0  
## [133,] 6.4 2.8 5.6 2.2  
## [134,] 6.3 2.8 5.1 1.5  
## [135,] 6.1 2.6 5.6 1.4  
## [136,] 7.7 3.0 6.1 2.3  
## [137,] 6.3 3.4 5.6 2.4  
## [138,] 6.4 3.1 5.5 1.8  
## [139,] 6.0 3.0 4.8 1.8  
## [140,] 6.9 3.1 5.4 2.1  
## [141,] 6.7 3.1 5.6 2.4  
## [142,] 6.9 3.1 5.1 2.3  
## [143,] 5.8 2.7 5.1 1.9  
## [144,] 6.8 3.2 5.9 2.3  
## [145,] 6.7 3.3 5.7 2.5  
## [146,] 6.7 3.0 5.2 2.3  
## [147,] 6.3 2.5 5.0 1.9  
## [148,] 6.5 3.0 5.2 2.0  
## [149,] 6.2 3.4 5.4 2.3  
## [150,] 5.9 3.0 5.1 1.8

colnames(mtx)

## [1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width"

aq <- read.csv("airquality.csv")  
str(aq)

## '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 ...

set.seed(3)  
col <- sample(letters, 153, replace = T)  
aq$NewCol <- col  
str(aq)

## 'data.frame': 153 obs. of 7 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 ...  
## $ NewCol : chr "e" "z" "l" "g" ...

write.csv(aq, "airquality2.csv", row.names = F)  
  
aq2 <- read.csv("airquality2.csv", stringsAsFactors = F)  
str(aq2)

## 'data.frame': 153 obs. of 7 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 ...  
## $ NewCol : chr "e" "z" "l" "g" ...

aq3 <- read.csv("airquality3.csv",   
 stringsAsFactors = T, header = F)  
str(aq3)

## 'data.frame': 153 obs. of 7 variables:  
## $ V1: int 41 36 12 18 NA 28 23 19 8 NA ...  
## $ V2: int 190 118 149 313 NA NA 299 99 19 194 ...  
## $ V3: num 7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...  
## $ V4: int 67 72 74 62 56 66 65 59 61 69 ...  
## $ V5: int 5 5 5 5 5 5 5 5 5 5 ...  
## $ V6: int 1 2 3 4 5 6 7 8 9 10 ...  
## $ V7: Factor w/ 26 levels "a","b","c","d",..: 5 26 12 7 4 26 8 11 8 20 ...

colnames(aq3) <- c("C1", "C2", "C3", "C4", "C5", "C6", "C7")  
str(aq3)

## 'data.frame': 153 obs. of 7 variables:  
## $ C1: int 41 36 12 18 NA 28 23 19 8 NA ...  
## $ C2: int 190 118 149 313 NA NA 299 99 19 194 ...  
## $ C3: num 7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...  
## $ C4: int 67 72 74 62 56 66 65 59 61 69 ...  
## $ C5: int 5 5 5 5 5 5 5 5 5 5 ...  
## $ C6: int 1 2 3 4 5 6 7 8 9 10 ...  
## $ C7: Factor w/ 26 levels "a","b","c","d",..: 5 26 12 7 4 26 8 11 8 20 ...

iris$Label <- ifelse(iris$Petal.Length >= 5.1, "H",   
 ifelse(iris$Petal.Length > 1.6, "M", "L"))  
str(iris)

## 'data.frame': 150 obs. of 6 variables:  
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
## $ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Label : chr "L" "L" "L" "L" ...

unique(iris$Label)

## [1] "L" "M" "H"

iris$Label <- factor(iris$Label, levels = c("H", "M", "L"))  
str(iris)

## 'data.frame': 150 obs. of 6 variables:  
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
## $ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Label : Factor w/ 3 levels "H","M","L": 3 3 3 3 3 2 3 3 3 3 ...

iris$Label <- factor(iris$Label,   
 levels = c("L", "M", "H"),   
 ordered = T)  
str(iris)

## 'data.frame': 150 obs. of 6 variables:  
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
## $ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Label : Ord.factor w/ 3 levels "L"<"M"<"H": 1 1 1 1 1 2 1 1 1 1 ...

apply(iris[, 1:4], 2, mean)

## Sepal.Length Sepal.Width Petal.Length Petal.Width   
## 5.843333 3.057333 3.758000 1.199333

rowIndex <- which(iris$Petal.Length > 5)  
iris[rowIndex, ]

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species Label  
## 84 6.0 2.7 5.1 1.6 versicolor H  
## 101 6.3 3.3 6.0 2.5 virginica H  
## 102 5.8 2.7 5.1 1.9 virginica H  
## 103 7.1 3.0 5.9 2.1 virginica H  
## 104 6.3 2.9 5.6 1.8 virginica H  
## 105 6.5 3.0 5.8 2.2 virginica H  
## 106 7.6 3.0 6.6 2.1 virginica H  
## 108 7.3 2.9 6.3 1.8 virginica H  
## 109 6.7 2.5 5.8 1.8 virginica H  
## 110 7.2 3.6 6.1 2.5 virginica H  
## 111 6.5 3.2 5.1 2.0 virginica H  
## 112 6.4 2.7 5.3 1.9 virginica H  
## 113 6.8 3.0 5.5 2.1 virginica H  
## 115 5.8 2.8 5.1 2.4 virginica H  
## 116 6.4 3.2 5.3 2.3 virginica H  
## 117 6.5 3.0 5.5 1.8 virginica H  
## 118 7.7 3.8 6.7 2.2 virginica H  
## 119 7.7 2.6 6.9 2.3 virginica H  
## 121 6.9 3.2 5.7 2.3 virginica H  
## 123 7.7 2.8 6.7 2.0 virginica H  
## 125 6.7 3.3 5.7 2.1 virginica H  
## 126 7.2 3.2 6.0 1.8 virginica H  
## 129 6.4 2.8 5.6 2.1 virginica H  
## 130 7.2 3.0 5.8 1.6 virginica H  
## 131 7.4 2.8 6.1 1.9 virginica H  
## 132 7.9 3.8 6.4 2.0 virginica H  
## 133 6.4 2.8 5.6 2.2 virginica H  
## 134 6.3 2.8 5.1 1.5 virginica H  
## 135 6.1 2.6 5.6 1.4 virginica H  
## 136 7.7 3.0 6.1 2.3 virginica H  
## 137 6.3 3.4 5.6 2.4 virginica H  
## 138 6.4 3.1 5.5 1.8 virginica H  
## 140 6.9 3.1 5.4 2.1 virginica H  
## 141 6.7 3.1 5.6 2.4 virginica H  
## 142 6.9 3.1 5.1 2.3 virginica H  
## 143 5.8 2.7 5.1 1.9 virginica H  
## 144 6.8 3.2 5.9 2.3 virginica H  
## 145 6.7 3.3 5.7 2.5 virginica H  
## 146 6.7 3.0 5.2 2.3 virginica H  
## 148 6.5 3.0 5.2 2.0 virginica H  
## 149 6.2 3.4 5.4 2.3 virginica H  
## 150 5.9 3.0 5.1 1.8 virginica H

data("mtcars")  
str(mtcars)

## 'data.frame': 32 obs. of 11 variables:  
## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...  
## $ cyl : num 6 6 4 6 8 6 8 4 4 6 ...  
## $ disp: num 160 160 108 258 360 ...  
## $ hp : num 110 110 93 110 175 105 245 62 95 123 ...  
## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...  
## $ wt : num 2.62 2.88 2.32 3.21 3.44 ...  
## $ qsec: num 16.5 17 18.6 19.4 17 ...  
## $ vs : num 0 0 1 1 0 1 0 1 1 1 ...  
## $ am : num 1 1 1 0 0 0 0 0 0 0 ...  
## $ gear: num 4 4 4 3 3 3 3 4 4 4 ...  
## $ carb: num 4 4 1 1 2 1 4 2 2 4 ...

head(mtcars)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

rownames(mtcars)

## [1] "Mazda RX4" "Mazda RX4 Wag" "Datsun 710"   
## [4] "Hornet 4 Drive" "Hornet Sportabout" "Valiant"   
## [7] "Duster 360" "Merc 240D" "Merc 230"   
## [10] "Merc 280" "Merc 280C" "Merc 450SE"   
## [13] "Merc 450SL" "Merc 450SLC" "Cadillac Fleetwood"   
## [16] "Lincoln Continental" "Chrysler Imperial" "Fiat 128"   
## [19] "Honda Civic" "Toyota Corolla" "Toyota Corona"   
## [22] "Dodge Challenger" "AMC Javelin" "Camaro Z28"   
## [25] "Pontiac Firebird" "Fiat X1-9" "Porsche 914-2"   
## [28] "Lotus Europa" "Ford Pantera L" "Ferrari Dino"   
## [31] "Maserati Bora" "Volvo 142E"

colnames(mtcars)

## [1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am" "gear"  
## [11] "carb"

apply(mtcars[,c("mpg", "hp", "wt")], 2, mean)

## mpg hp wt   
## 20.09062 146.68750 3.21725

vec <- sample(letters[1:3], 100, replace = T)  
table(vec)

## vec  
## a b c   
## 31 38 31

length(vec)

## [1] 100

vec[100]

## [1] "c"

vec[101] # NA

## [1] NA

vec[101] <- "a"  
vec

## [1] "b" "b" "a" "c" "b" "c" "b" "c" "a" "a" "a" "b" "b" "c" "a" "b" "b" "a"  
## [19] "c" "a" "c" "a" "b" "a" "c" "b" "c" "b" "c" "a" "a" "a" "c" "a" "b" "c"  
## [37] "b" "a" "b" "b" "a" "b" "b" "c" "b" "a" "a" "c" "b" "b" "b" "b" "c" "a"  
## [55] "b" "b" "b" "a" "b" "c" "c" "b" "a" "a" "b" "a" "b" "c" "c" "b" "c" "a"  
## [73] "c" "a" "c" "c" "a" "b" "b" "b" "b" "c" "c" "c" "a" "b" "b" "b" "c" "a"  
## [91] "c" "c" "b" "a" "a" "c" "a" "a" "c" "c" "a"

vec[50:52] <- NA  
table(vec)

## vec  
## a b c   
## 32 35 31

table(vec, useNA = "ifany")

## vec  
## a b c <NA>   
## 32 35 31 3

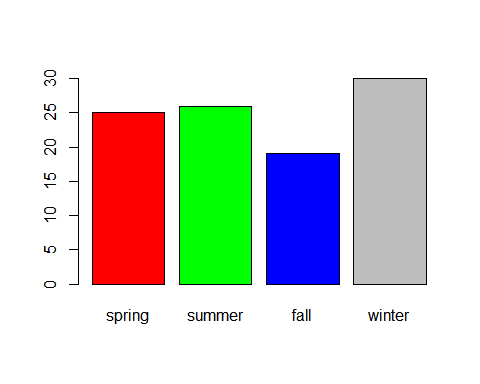
prop.table(table(vec, useNA = "ifany"))

## vec  
## a b c <NA>   
## 0.31683168 0.34653465 0.30693069 0.02970297

season <- c("spring", "summer", "fall", "winter")  
set.seed(3)  
favorite <- sample(season, 100, replace = T)   
favorite2 <- factor(favorite, levels = season)  
table(favorite2)

## favorite2  
## spring summer fall winter   
## 25 26 19 30

barplot(table(favorite2),   
 col = c("red", "green", "blue", "gray"))  
  
library(ggplot2)



library(dplyr)

##   
## Attaching package: 'dplyr'

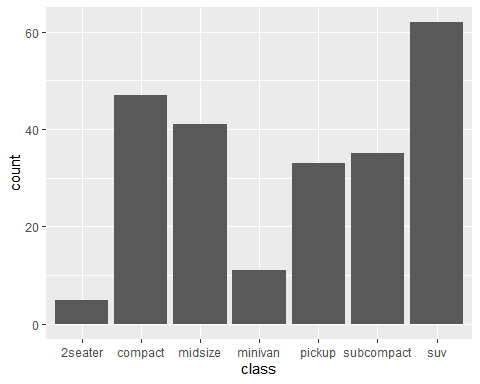
## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

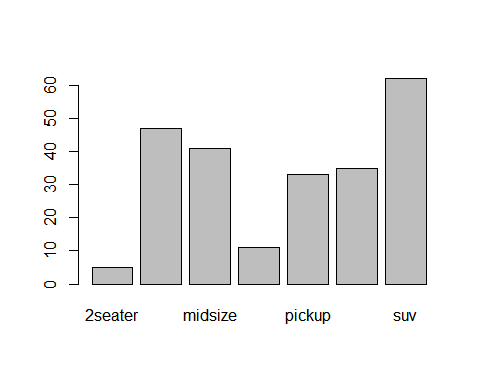
data(mpg)  
str(mpg)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 234 obs. of 11 variables:  
## $ manufacturer: chr "audi" "audi" "audi" "audi" ...  
## $ model : chr "a4" "a4" "a4" "a4" ...  
## $ displ : num 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...  
## $ year : int 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...  
## $ cyl : int 4 4 4 4 6 6 6 4 4 4 ...  
## $ trans : chr "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...  
## $ drv : chr "f" "f" "f" "f" ...  
## $ cty : int 18 21 20 21 16 18 18 18 16 20 ...  
## $ hwy : int 29 29 31 30 26 26 27 26 25 28 ...  
## $ fl : chr "p" "p" "p" "p" ...  
## $ class : chr "compact" "compact" "compact" "compact" ...

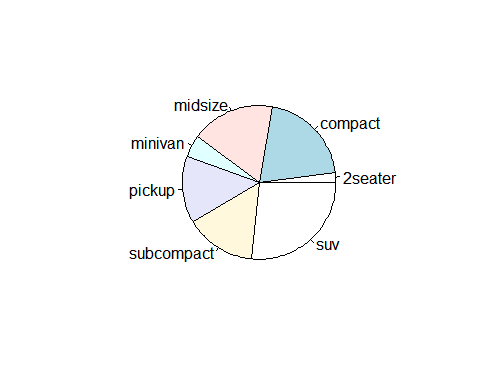
mpg %>%  
 ggplot(aes(x = class)) +   
 geom\_bar()



barplot(table(mpg$class))



pie(table(mpg$class))



mean(mpg$cty)

## [1] 16.85897

mean(mpg$hwy)

## [1] 23.44017

apply(mpg[c("cty", "hwy")], 2, mean)

## cty hwy   
## 16.85897 23.44017

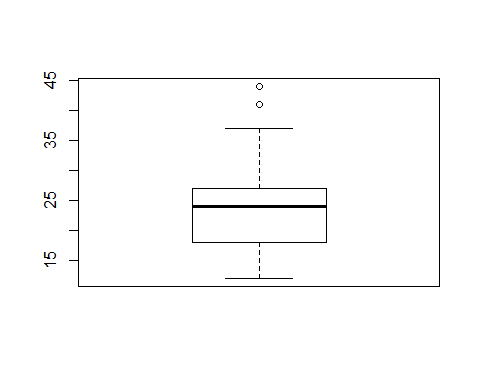
quantile(mpg$hwy, probs = 0:5/5)

## 0% 20% 40% 60% 80% 100%   
## 12 17 22 26 29 44

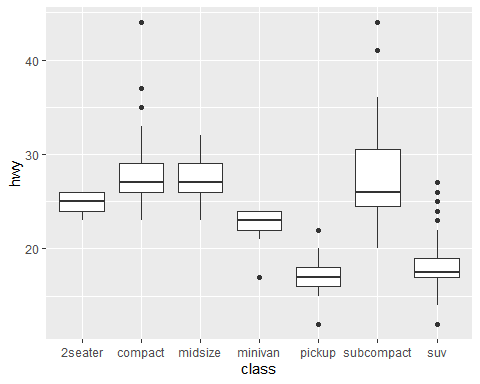
0:5/5

## [1] 0.0 0.2 0.4 0.6 0.8 1.0

boxplot(mpg$hwy)



mpg %>%  
 ggplot(aes(x = class, y = hwy)) +   
 geom\_boxplot()



(q <- quantile(mpg$hwy))

## 0% 25% 50% 75% 100%   
## 12 18 24 27 44

q1 <- q["25%"]  
q3 <- q["75%"]  
iqr <- q3 - q1  
mpg$hwy

## [1] 29 29 31 30 26 26 27 26 25 28 27 25 25 25 25 24 25 23 20 15 20 17 17 26 23  
## [26] 26 25 24 19 14 15 17 27 30 26 29 26 24 24 22 22 24 24 17 22 21 23 23 19 18  
## [51] 17 17 19 19 12 17 15 17 17 12 17 16 18 15 16 12 17 17 16 12 15 16 17 15 17  
## [76] 17 18 17 19 17 19 19 17 17 17 16 16 17 15 17 26 25 26 24 21 22 23 22 20 33  
## [101] 32 32 29 32 34 36 36 29 26 27 30 31 26 26 28 26 29 28 27 24 24 24 22 19 20  
## [126] 17 12 19 18 14 15 18 18 15 17 16 18 17 19 19 17 29 27 31 32 27 26 26 25 25  
## [151] 17 17 20 18 26 26 27 28 25 25 24 27 25 26 23 26 26 26 26 25 27 25 27 20 20  
## [176] 19 17 20 17 29 27 31 31 26 26 28 27 29 31 31 26 26 27 30 33 35 37 35 15 18  
## [201] 20 20 22 17 19 18 20 29 26 29 29 24 44 29 26 29 29 29 29 23 24 44 41 29 26  
## [226] 28 29 29 29 28 29 26 26 26

mpg$hwy > q3 + 1.5 \* iqr

## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [193] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [205] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE  
## [217] FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE  
## [229] FALSE FALSE FALSE FALSE FALSE FALSE

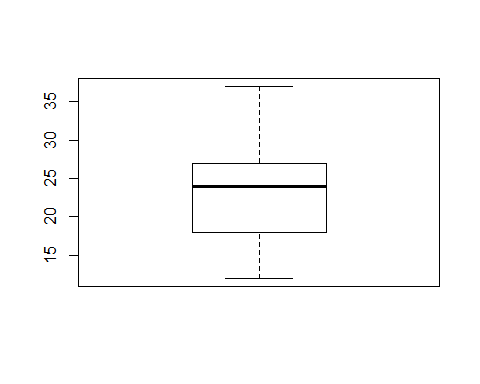
(outlier\_idx1 <- which(mpg$hwy > q3 + 1.5 \* iqr))

## [1] 213 222 223

(outlier\_idx2 <- which(mpg$hwy < q1 - 1.5 \* iqr))

## integer(0)

boxplot(mpg$hwy[-outlier\_idx1])



quantile(mpg$hwy[-outlier\_idx1])

## 0% 25% 50% 75% 100%   
## 12 18 24 27 37

quantile(mpg$hwy)

## 0% 25% 50% 75% 100%   
## 12 18 24 27 44

set.seed(123)  
scores <- sample(seq(35, 100, 5),   
 1000, replace = T)  
cut(scores,   
 breaks = c(0, 60, 70, 80, 90, 100),   
 include.lowest = T,  
 right = F,  
 labels = c("F", "D", "C", "B", "A"))

## [1] F A F B F D B F F A D C B B F F B C A C C A F C B D B C A F F A F B D F A  
## [38] B A D C C B D B A F D F B D C F A F C A A F F C B C D F C B A F A D B C D  
## [75] D D F D F F F F A F D F C F D C A C D F C A A C A A F A A D F D D B F F C  
## [112] F B F B A F B D A F A F A A D F C A F A C F D B D A D B D A D B B F D C F  
## [149] D B F A A A A F C D D B C D F A F C F D F F C A A F C C D B F C F A F F A  
## [186] D F B B F B C D C A F A F B A A D C C A B F A A A F D A F C C B C C F A F  
## [223] C D D B A A B C D D B B F C A A F B F D A B B C F B D C D B F A C C C C F  
## [260] D A A D B F B B F B F A B B F D A F B C F D F A A F F C F F C A F F F B F  
## [297] F F C C D C F A F B D D A C F F A A F C F D F F A F F F F F A C F F F F F  
## [334] A C A B A D B B D A F F B B F A D F F D F C A D A A C F F D B A C A A A B  
## [371] D A F B F F A F F F A A F A C F D F F C F F D B C B B A F D F A F A C C D  
## [408] B F F C F B D B F A F F A A A F D A B F F F F A F A F A F F A F F D B F B  
## [445] F A F C F B F C A A B B C A D F F F B C F B A C F F A F D B B C B D D A F  
## [482] B C B A D F C A F A C D F D B B C A F B C D F F C C F F D F F C D A F A D  
## [519] F C B F D F A A C B B F A A B D A F C A D F A A D F F C C B F C D F D C B  
## [556] B F F D D C F D A F F A F F F B F F A A D F B F B F D A D A F F D F F C F  
## [593] D A B D B B D F D C F F B F F A F F C B F A C A D C A F D B D A C F A F F  
## [630] A C C A A D F F B B C C B D B F A C B F A F D F D A B A B F D F D B A D C  
## [667] C F D F C A A B C C C C C F F F C F F C D D A A F A A D F A A C B A F B A  
## [704] C A C D C F F B D F D B F D D F B F C F F A F A F A A D C F F A F F F F A  
## [741] F A F B B B B F D F F F F D F A D C C C F F F C A F D F F B D D C F C B F  
## [778] B C C B F A C A C F F D F C C F F C D F F B B A D F C C B B F A B A B C F  
## [815] F B C A B D D D F F F F B A B F D C F F D A B F C D C D C A C F D C F C C  
## [852] B C B F D C F A B D F B A B D F B C F F F F B C A A C F B D D F A D A C D  
## [889] B B F D D A A F F C A A B B F F F F F B B B F C F D D F F A A A F A C F F  
## [926] D B F F F A A A B F D F B C F A D F C A F F A B A A F F F B D F B D B D F  
## [963] F F F C A C F D A C F C F C F B F C F B A F D A F A D F A A A D D D D C C  
## [1000] F  
## Levels: F D C B A

iris[1:4]

## Sepal.Length Sepal.Width Petal.Length Petal.Width  
## 1 5.1 3.5 1.4 0.2  
## 2 4.9 3.0 1.4 0.2  
## 3 4.7 3.2 1.3 0.2  
## 4 4.6 3.1 1.5 0.2  
## 5 5.0 3.6 1.4 0.2  
## 6 5.4 3.9 1.7 0.4  
## 7 4.6 3.4 1.4 0.3  
## 8 5.0 3.4 1.5 0.2  
## 9 4.4 2.9 1.4 0.2  
## 10 4.9 3.1 1.5 0.1  
## 11 5.4 3.7 1.5 0.2  
## 12 4.8 3.4 1.6 0.2  
## 13 4.8 3.0 1.4 0.1  
## 14 4.3 3.0 1.1 0.1  
## 15 5.8 4.0 1.2 0.2  
## 16 5.7 4.4 1.5 0.4  
## 17 5.4 3.9 1.3 0.4  
## 18 5.1 3.5 1.4 0.3  
## 19 5.7 3.8 1.7 0.3  
## 20 5.1 3.8 1.5 0.3  
## 21 5.4 3.4 1.7 0.2  
## 22 5.1 3.7 1.5 0.4  
## 23 4.6 3.6 1.0 0.2  
## 24 5.1 3.3 1.7 0.5  
## 25 4.8 3.4 1.9 0.2  
## 26 5.0 3.0 1.6 0.2  
## 27 5.0 3.4 1.6 0.4  
## 28 5.2 3.5 1.5 0.2  
## 29 5.2 3.4 1.4 0.2  
## 30 4.7 3.2 1.6 0.2  
## 31 4.8 3.1 1.6 0.2  
## 32 5.4 3.4 1.5 0.4  
## 33 5.2 4.1 1.5 0.1  
## 34 5.5 4.2 1.4 0.2  
## 35 4.9 3.1 1.5 0.2  
## 36 5.0 3.2 1.2 0.2  
## 37 5.5 3.5 1.3 0.2  
## 38 4.9 3.6 1.4 0.1  
## 39 4.4 3.0 1.3 0.2  
## 40 5.1 3.4 1.5 0.2  
## 41 5.0 3.5 1.3 0.3  
## 42 4.5 2.3 1.3 0.3  
## 43 4.4 3.2 1.3 0.2  
## 44 5.0 3.5 1.6 0.6  
## 45 5.1 3.8 1.9 0.4  
## 46 4.8 3.0 1.4 0.3  
## 47 5.1 3.8 1.6 0.2  
## 48 4.6 3.2 1.4 0.2  
## 49 5.3 3.7 1.5 0.2  
## 50 5.0 3.3 1.4 0.2  
## 51 7.0 3.2 4.7 1.4  
## 52 6.4 3.2 4.5 1.5  
## 53 6.9 3.1 4.9 1.5  
## 54 5.5 2.3 4.0 1.3  
## 55 6.5 2.8 4.6 1.5  
## 56 5.7 2.8 4.5 1.3  
## 57 6.3 3.3 4.7 1.6  
## 58 4.9 2.4 3.3 1.0  
## 59 6.6 2.9 4.6 1.3  
## 60 5.2 2.7 3.9 1.4  
## 61 5.0 2.0 3.5 1.0  
## 62 5.9 3.0 4.2 1.5  
## 63 6.0 2.2 4.0 1.0  
## 64 6.1 2.9 4.7 1.4  
## 65 5.6 2.9 3.6 1.3  
## 66 6.7 3.1 4.4 1.4  
## 67 5.6 3.0 4.5 1.5  
## 68 5.8 2.7 4.1 1.0  
## 69 6.2 2.2 4.5 1.5  
## 70 5.6 2.5 3.9 1.1  
## 71 5.9 3.2 4.8 1.8  
## 72 6.1 2.8 4.0 1.3  
## 73 6.3 2.5 4.9 1.5  
## 74 6.1 2.8 4.7 1.2  
## 75 6.4 2.9 4.3 1.3  
## 76 6.6 3.0 4.4 1.4  
## 77 6.8 2.8 4.8 1.4  
## 78 6.7 3.0 5.0 1.7  
## 79 6.0 2.9 4.5 1.5  
## 80 5.7 2.6 3.5 1.0  
## 81 5.5 2.4 3.8 1.1  
## 82 5.5 2.4 3.7 1.0  
## 83 5.8 2.7 3.9 1.2  
## 84 6.0 2.7 5.1 1.6  
## 85 5.4 3.0 4.5 1.5  
## 86 6.0 3.4 4.5 1.6  
## 87 6.7 3.1 4.7 1.5  
## 88 6.3 2.3 4.4 1.3  
## 89 5.6 3.0 4.1 1.3  
## 90 5.5 2.5 4.0 1.3  
## 91 5.5 2.6 4.4 1.2  
## 92 6.1 3.0 4.6 1.4  
## 93 5.8 2.6 4.0 1.2  
## 94 5.0 2.3 3.3 1.0  
## 95 5.6 2.7 4.2 1.3  
## 96 5.7 3.0 4.2 1.2  
## 97 5.7 2.9 4.2 1.3  
## 98 6.2 2.9 4.3 1.3  
## 99 5.1 2.5 3.0 1.1  
## 100 5.7 2.8 4.1 1.3  
## 101 6.3 3.3 6.0 2.5  
## 102 5.8 2.7 5.1 1.9  
## 103 7.1 3.0 5.9 2.1  
## 104 6.3 2.9 5.6 1.8  
## 105 6.5 3.0 5.8 2.2  
## 106 7.6 3.0 6.6 2.1  
## 107 4.9 2.5 4.5 1.7  
## 108 7.3 2.9 6.3 1.8  
## 109 6.7 2.5 5.8 1.8  
## 110 7.2 3.6 6.1 2.5  
## 111 6.5 3.2 5.1 2.0  
## 112 6.4 2.7 5.3 1.9  
## 113 6.8 3.0 5.5 2.1  
## 114 5.7 2.5 5.0 2.0  
## 115 5.8 2.8 5.1 2.4  
## 116 6.4 3.2 5.3 2.3  
## 117 6.5 3.0 5.5 1.8  
## 118 7.7 3.8 6.7 2.2  
## 119 7.7 2.6 6.9 2.3  
## 120 6.0 2.2 5.0 1.5  
## 121 6.9 3.2 5.7 2.3  
## 122 5.6 2.8 4.9 2.0  
## 123 7.7 2.8 6.7 2.0  
## 124 6.3 2.7 4.9 1.8  
## 125 6.7 3.3 5.7 2.1  
## 126 7.2 3.2 6.0 1.8  
## 127 6.2 2.8 4.8 1.8  
## 128 6.1 3.0 4.9 1.8  
## 129 6.4 2.8 5.6 2.1  
## 130 7.2 3.0 5.8 1.6  
## 131 7.4 2.8 6.1 1.9  
## 132 7.9 3.8 6.4 2.0  
## 133 6.4 2.8 5.6 2.2  
## 134 6.3 2.8 5.1 1.5  
## 135 6.1 2.6 5.6 1.4  
## 136 7.7 3.0 6.1 2.3  
## 137 6.3 3.4 5.6 2.4  
## 138 6.4 3.1 5.5 1.8  
## 139 6.0 3.0 4.8 1.8  
## 140 6.9 3.1 5.4 2.1  
## 141 6.7 3.1 5.6 2.4  
## 142 6.9 3.1 5.1 2.3  
## 143 5.8 2.7 5.1 1.9  
## 144 6.8 3.2 5.9 2.3  
## 145 6.7 3.3 5.7 2.5  
## 146 6.7 3.0 5.2 2.3  
## 147 6.3 2.5 5.0 1.9  
## 148 6.5 3.0 5.2 2.0  
## 149 6.2 3.4 5.4 2.3  
## 150 5.9 3.0 5.1 1.8

cor(iris$Sepal.Length, iris$Sepal.Width)

## [1] -0.1175698

cor(iris$Sepal.Length, iris$Petal.Length)

## [1] 0.8717538

cor(iris$Sepal.Length, iris$Petal.Width)

## [1] 0.8179411

cor(iris[-(5:6)])[1, -1]

## Sepal.Width Petal.Length Petal.Width   
## -0.1175698 0.8717538 0.8179411

sort(abs(cor(iris[-(5:6)])[1, -1]),   
 decreasing = T)[1:2]

## Petal.Length Petal.Width   
## 0.8717538 0.8179411

library(MASS)

##   
## Attaching package: 'MASS'

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

data("birthwt")  
str(birthwt)

## 'data.frame': 189 obs. of 10 variables:  
## $ low : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ age : int 19 33 20 21 18 21 22 17 29 26 ...  
## $ lwt : int 182 155 105 108 107 124 118 103 123 113 ...  
## $ race : int 2 3 1 1 1 3 1 3 1 1 ...  
## $ smoke: int 0 0 1 1 1 0 0 0 1 1 ...  
## $ ptl : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ ht : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ ui : int 1 0 0 1 1 0 0 0 0 0 ...  
## $ ftv : int 0 3 1 2 0 0 1 1 1 0 ...  
## $ bwt : int 2523 2551 2557 2594 2600 2622 2637 2637 2663 2665 ...

(m <- birthwt[c("bwt", "age", "lwt",   
 "ptl", "ftv")])

## bwt age lwt ptl ftv  
## 85 2523 19 182 0 0  
## 86 2551 33 155 0 3  
## 87 2557 20 105 0 1  
## 88 2594 21 108 0 2  
## 89 2600 18 107 0 0  
## 91 2622 21 124 0 0  
## 92 2637 22 118 0 1  
## 93 2637 17 103 0 1  
## 94 2663 29 123 0 1  
## 95 2665 26 113 0 0  
## 96 2722 19 95 0 0  
## 97 2733 19 150 0 1  
## 98 2751 22 95 0 0  
## 99 2750 30 107 1 2  
## 100 2769 18 100 0 0  
## 101 2769 18 100 0 0  
## 102 2778 15 98 0 0  
## 103 2782 25 118 0 3  
## 104 2807 20 120 0 0  
## 105 2821 28 120 0 1  
## 106 2835 32 121 0 2  
## 107 2835 31 100 0 3  
## 108 2836 36 202 0 1  
## 109 2863 28 120 0 0  
## 111 2877 25 120 0 2  
## 112 2877 28 167 0 0  
## 113 2906 17 122 0 0  
## 114 2920 29 150 0 2  
## 115 2920 26 168 0 0  
## 116 2920 17 113 0 1  
## 117 2920 17 113 0 1  
## 118 2948 24 90 1 1  
## 119 2948 35 121 1 1  
## 120 2977 25 155 0 1  
## 121 2977 25 125 0 0  
## 123 2977 29 140 0 2  
## 124 2977 19 138 0 2  
## 125 2922 27 124 0 0  
## 126 3005 31 215 0 2  
## 127 3033 33 109 0 1  
## 128 3042 21 185 0 2  
## 129 3062 19 189 0 2  
## 130 3062 23 130 0 1  
## 131 3062 21 160 0 0  
## 132 3062 18 90 0 0  
## 133 3062 18 90 0 0  
## 134 3080 32 132 0 4  
## 135 3090 19 132 0 0  
## 136 3090 24 115 0 2  
## 137 3090 22 85 0 0  
## 138 3100 22 120 0 1  
## 139 3104 23 128 0 0  
## 140 3132 22 130 0 0  
## 141 3147 30 95 0 2  
## 142 3175 19 115 0 0  
## 143 3175 16 110 0 0  
## 144 3203 21 110 0 0  
## 145 3203 30 153 0 0  
## 146 3203 20 103 0 0  
## 147 3225 17 119 0 0  
## 148 3225 17 119 0 0  
## 149 3232 23 119 0 2  
## 150 3232 24 110 0 0  
## 151 3234 28 140 0 0  
## 154 3260 26 133 2 0  
## 155 3274 20 169 1 1  
## 156 3274 24 115 0 2  
## 159 3303 28 250 0 6  
## 160 3317 20 141 2 1  
## 161 3317 22 158 1 2  
## 162 3317 22 112 2 0  
## 163 3321 31 150 0 2  
## 164 3331 23 115 0 1  
## 166 3374 16 112 0 0  
## 167 3374 16 135 0 0  
## 168 3402 18 229 0 0  
## 169 3416 25 140 0 1  
## 170 3430 32 134 1 4  
## 172 3444 20 121 0 0  
## 173 3459 23 190 0 0  
## 174 3460 22 131 0 1  
## 175 3473 32 170 0 0  
## 176 3544 30 110 0 0  
## 177 3487 20 127 0 0  
## 179 3544 23 123 0 0  
## 180 3572 17 120 0 0  
## 181 3572 19 105 0 0  
## 182 3586 23 130 0 0  
## 183 3600 36 175 0 0  
## 184 3614 22 125 0 1  
## 185 3614 24 133 0 0  
## 186 3629 21 134 0 2  
## 187 3629 19 235 0 0  
## 188 3637 25 95 3 0  
## 189 3643 16 135 0 0  
## 190 3651 29 135 0 1  
## 191 3651 29 154 0 1  
## 192 3651 19 147 0 0  
## 193 3651 19 147 0 0  
## 195 3699 30 137 0 1  
## 196 3728 24 110 0 1  
## 197 3756 19 184 0 0  
## 199 3770 24 110 1 0  
## 200 3770 23 110 0 1  
## 201 3770 20 120 0 0  
## 202 3790 25 241 0 0  
## 203 3799 30 112 0 1  
## 204 3827 22 169 0 0  
## 205 3856 18 120 0 2  
## 206 3860 16 170 0 4  
## 207 3860 32 186 0 2  
## 208 3884 18 120 0 1  
## 209 3884 29 130 0 2  
## 210 3912 33 117 0 1  
## 211 3940 20 170 0 0  
## 212 3941 28 134 0 1  
## 213 3941 14 135 0 0  
## 214 3969 28 130 0 0  
## 215 3983 25 120 0 2  
## 216 3997 16 95 0 1  
## 217 3997 20 158 0 1  
## 218 4054 26 160 0 0  
## 219 4054 21 115 0 1  
## 220 4111 22 129 0 0  
## 221 4153 25 130 0 2  
## 222 4167 31 120 0 2  
## 223 4174 35 170 1 1  
## 224 4238 19 120 0 0  
## 225 4593 24 116 0 1  
## 226 4990 45 123 0 1  
## 4 709 28 120 1 0  
## 10 1021 29 130 0 2  
## 11 1135 34 187 0 0  
## 13 1330 25 105 1 0  
## 15 1474 25 85 0 0  
## 16 1588 27 150 0 0  
## 17 1588 23 97 0 1  
## 18 1701 24 128 1 1  
## 19 1729 24 132 0 0  
## 20 1790 21 165 0 1  
## 22 1818 32 105 0 0  
## 23 1885 19 91 2 0  
## 24 1893 25 115 0 0  
## 25 1899 16 130 0 1  
## 26 1928 25 92 0 0  
## 27 1928 20 150 0 2  
## 28 1928 21 200 0 2  
## 29 1936 24 155 1 0  
## 30 1970 21 103 0 0  
## 31 2055 20 125 0 0  
## 32 2055 25 89 2 1  
## 33 2082 19 102 0 2  
## 34 2084 19 112 0 0  
## 35 2084 26 117 1 0  
## 36 2100 24 138 0 0  
## 37 2125 17 130 1 0  
## 40 2126 20 120 0 3  
## 42 2187 22 130 1 1  
## 43 2187 27 130 0 0  
## 44 2211 20 80 0 0  
## 45 2225 17 110 0 0  
## 46 2240 25 105 1 1  
## 47 2240 20 109 0 0  
## 49 2282 18 148 0 0  
## 50 2296 18 110 1 0  
## 51 2296 20 121 1 0  
## 52 2301 21 100 1 4  
## 54 2325 26 96 0 0  
## 56 2353 31 102 1 1  
## 57 2353 15 110 0 0  
## 59 2367 23 187 0 1  
## 60 2381 20 122 0 0  
## 61 2381 24 105 0 0  
## 62 2381 15 115 0 0  
## 63 2410 23 120 0 0  
## 65 2410 30 142 1 0  
## 67 2410 22 130 0 1  
## 68 2414 17 120 0 3  
## 69 2424 23 110 1 0  
## 71 2438 17 120 0 2  
## 75 2442 26 154 1 1  
## 76 2450 20 105 0 3  
## 77 2466 26 190 0 0  
## 78 2466 14 101 1 0  
## 79 2466 28 95 0 2  
## 81 2495 14 100 0 2  
## 82 2495 23 94 0 0  
## 83 2495 17 142 0 0  
## 84 2495 21 130 0 3

(c <- cor(m)[1, -1])

## age lwt ptl ftv   
## 0.09031781 0.18573328 -0.15465339 0.05831777

sort(abs(c), decreasing = T)[1:2]

## lwt ptl   
## 0.1857333 0.1546534

housing\_df <- read.table("housing.data", sep = "", stringsAsFactors = F)  
str(housing\_df)

## 'data.frame': 506 obs. of 14 variables:  
## $ V1 : num 0.00632 0.02731 0.02729 0.03237 0.06905 ...  
## $ V2 : num 18 0 0 0 0 0 12.5 12.5 12.5 12.5 ...  
## $ V3 : num 2.31 7.07 7.07 2.18 2.18 2.18 7.87 7.87 7.87 7.87 ...  
## $ V4 : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ V5 : num 0.538 0.469 0.469 0.458 0.458 0.458 0.524 0.524 0.524 0.524 ...  
## $ V6 : num 6.58 6.42 7.18 7 7.15 ...  
## $ V7 : num 65.2 78.9 61.1 45.8 54.2 58.7 66.6 96.1 100 85.9 ...  
## $ V8 : num 4.09 4.97 4.97 6.06 6.06 ...  
## $ V9 : int 1 2 2 3 3 3 5 5 5 5 ...  
## $ V10: num 296 242 242 222 222 222 311 311 311 311 ...  
## $ V11: num 15.3 17.8 17.8 18.7 18.7 18.7 15.2 15.2 15.2 15.2 ...  
## $ V12: num 397 397 393 395 397 ...  
## $ V13: num 4.98 9.14 4.03 2.94 5.33 ...  
## $ V14: num 24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...

colnames(housing\_df) <- c("CRIM","ZN","INDUS",  
 "CHAS","NOX","RM","AGE",  
 "DIS","RAD","TAX",  
 "PTRATIO","B","LSTAT",  
 "MEDV")  
str(housing\_df)

## 'data.frame': 506 obs. of 14 variables:  
## $ CRIM : num 0.00632 0.02731 0.02729 0.03237 0.06905 ...  
## $ ZN : num 18 0 0 0 0 0 12.5 12.5 12.5 12.5 ...  
## $ INDUS : num 2.31 7.07 7.07 2.18 2.18 2.18 7.87 7.87 7.87 7.87 ...  
## $ CHAS : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ NOX : num 0.538 0.469 0.469 0.458 0.458 0.458 0.524 0.524 0.524 0.524 ...  
## $ RM : num 6.58 6.42 7.18 7 7.15 ...  
## $ AGE : num 65.2 78.9 61.1 45.8 54.2 58.7 66.6 96.1 100 85.9 ...  
## $ DIS : num 4.09 4.97 4.97 6.06 6.06 ...  
## $ RAD : int 1 2 2 3 3 3 5 5 5 5 ...  
## $ TAX : num 296 242 242 222 222 222 311 311 311 311 ...  
## $ PTRATIO: num 15.3 17.8 17.8 18.7 18.7 18.7 15.2 15.2 15.2 15.2 ...  
## $ B : num 397 397 393 395 397 ...  
## $ LSTAT : num 4.98 9.14 4.03 2.94 5.33 ...  
## $ MEDV : num 24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...

str(housing\_df[-4])

## 'data.frame': 506 obs. of 13 variables:  
## $ CRIM : num 0.00632 0.02731 0.02729 0.03237 0.06905 ...  
## $ ZN : num 18 0 0 0 0 0 12.5 12.5 12.5 12.5 ...  
## $ INDUS : num 2.31 7.07 7.07 2.18 2.18 2.18 7.87 7.87 7.87 7.87 ...  
## $ NOX : num 0.538 0.469 0.469 0.458 0.458 0.458 0.524 0.524 0.524 0.524 ...  
## $ RM : num 6.58 6.42 7.18 7 7.15 ...  
## $ AGE : num 65.2 78.9 61.1 45.8 54.2 58.7 66.6 96.1 100 85.9 ...  
## $ DIS : num 4.09 4.97 4.97 6.06 6.06 ...  
## $ RAD : int 1 2 2 3 3 3 5 5 5 5 ...  
## $ TAX : num 296 242 242 222 222 222 311 311 311 311 ...  
## $ PTRATIO: num 15.3 17.8 17.8 18.7 18.7 18.7 15.2 15.2 15.2 15.2 ...  
## $ B : num 397 397 393 395 397 ...  
## $ LSTAT : num 4.98 9.14 4.03 2.94 5.33 ...  
## $ MEDV : num 24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...

cor(housing\_df[-4])

## CRIM ZN INDUS NOX RM AGE  
## CRIM 1.0000000 -0.2004692 0.4065834 0.4209717 -0.2192467 0.3527343  
## ZN -0.2004692 1.0000000 -0.5338282 -0.5166037 0.3119906 -0.5695373  
## INDUS 0.4065834 -0.5338282 1.0000000 0.7636514 -0.3916759 0.6447785  
## NOX 0.4209717 -0.5166037 0.7636514 1.0000000 -0.3021882 0.7314701  
## RM -0.2192467 0.3119906 -0.3916759 -0.3021882 1.0000000 -0.2402649  
## AGE 0.3527343 -0.5695373 0.6447785 0.7314701 -0.2402649 1.0000000  
## DIS -0.3796701 0.6644082 -0.7080270 -0.7692301 0.2052462 -0.7478805  
## RAD 0.6255051 -0.3119478 0.5951293 0.6114406 -0.2098467 0.4560225  
## TAX 0.5827643 -0.3145633 0.7207602 0.6680232 -0.2920478 0.5064556  
## PTRATIO 0.2899456 -0.3916785 0.3832476 0.1889327 -0.3555015 0.2615150  
## B -0.3850639 0.1755203 -0.3569765 -0.3800506 0.1280686 -0.2735340  
## LSTAT 0.4556215 -0.4129946 0.6037997 0.5908789 -0.6138083 0.6023385  
## MEDV -0.3883046 0.3604453 -0.4837252 -0.4273208 0.6953599 -0.3769546  
## DIS RAD TAX PTRATIO B LSTAT  
## CRIM -0.3796701 0.6255051 0.5827643 0.2899456 -0.3850639 0.4556215  
## ZN 0.6644082 -0.3119478 -0.3145633 -0.3916785 0.1755203 -0.4129946  
## INDUS -0.7080270 0.5951293 0.7207602 0.3832476 -0.3569765 0.6037997  
## NOX -0.7692301 0.6114406 0.6680232 0.1889327 -0.3800506 0.5908789  
## RM 0.2052462 -0.2098467 -0.2920478 -0.3555015 0.1280686 -0.6138083  
## AGE -0.7478805 0.4560225 0.5064556 0.2615150 -0.2735340 0.6023385  
## DIS 1.0000000 -0.4945879 -0.5344316 -0.2324705 0.2915117 -0.4969958  
## RAD -0.4945879 1.0000000 0.9102282 0.4647412 -0.4444128 0.4886763  
## TAX -0.5344316 0.9102282 1.0000000 0.4608530 -0.4418080 0.5439934  
## PTRATIO -0.2324705 0.4647412 0.4608530 1.0000000 -0.1773833 0.3740443  
## B 0.2915117 -0.4444128 -0.4418080 -0.1773833 1.0000000 -0.3660869  
## LSTAT -0.4969958 0.4886763 0.5439934 0.3740443 -0.3660869 1.0000000  
## MEDV 0.2499287 -0.3816262 -0.4685359 -0.5077867 0.3334608 -0.7376627  
## MEDV  
## CRIM -0.3883046  
## ZN 0.3604453  
## INDUS -0.4837252  
## NOX -0.4273208  
## RM 0.6953599  
## AGE -0.3769546  
## DIS 0.2499287  
## RAD -0.3816262  
## TAX -0.4685359  
## PTRATIO -0.5077867  
## B 0.3334608  
## LSTAT -0.7376627  
## MEDV 1.0000000

cor(housing\_df[-4])[, "MEDV"]

## CRIM ZN INDUS NOX RM AGE DIS   
## -0.3883046 0.3604453 -0.4837252 -0.4273208 0.6953599 -0.3769546 0.2499287   
## RAD TAX PTRATIO B LSTAT MEDV   
## -0.3816262 -0.4685359 -0.5077867 0.3334608 -0.7376627 1.0000000

(c <- cor(housing\_df[-4])[, "MEDV"])

## CRIM ZN INDUS NOX RM AGE DIS   
## -0.3883046 0.3604453 -0.4837252 -0.4273208 0.6953599 -0.3769546 0.2499287   
## RAD TAX PTRATIO B LSTAT MEDV   
## -0.3816262 -0.4685359 -0.5077867 0.3334608 -0.7376627 1.0000000

c[-length(c)]

## CRIM ZN INDUS NOX RM AGE DIS   
## -0.3883046 0.3604453 -0.4837252 -0.4273208 0.6953599 -0.3769546 0.2499287   
## RAD TAX PTRATIO B LSTAT   
## -0.3816262 -0.4685359 -0.5077867 0.3334608 -0.7376627

sort(abs(c[-length(c)]), decreasing = T)[1:5]

## LSTAT RM PTRATIO INDUS TAX   
## 0.7376627 0.6953599 0.5077867 0.4837252 0.4685359

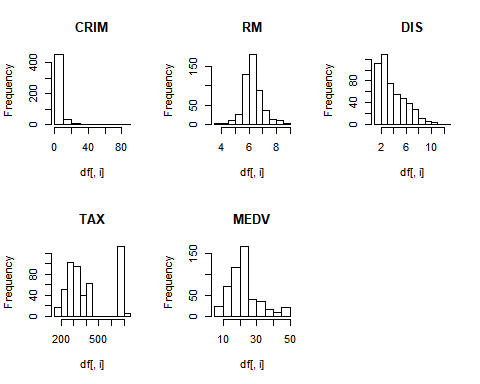
housing\_df$GRP <- ifelse(housing\_df$MEDV >= 25.0, "H",  
 ifelse(housing\_df$MEDV >= 17.0, "M", "L"))  
housing\_df$GRP <- factor(housing\_df$GRP,  
 levels = c("L", "M", "H"))  
table(housing\_df$GRP)

##   
## L M H   
## 126 248 132

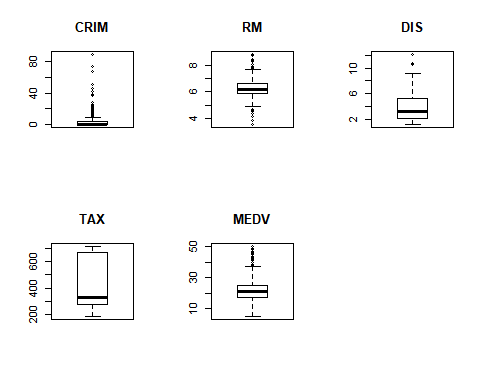
df <- housing\_df[c("CRIM", "RM", "DIS",   
 "TAX", "MEDV")]  
str(df)

## 'data.frame': 506 obs. of 5 variables:  
## $ CRIM: num 0.00632 0.02731 0.02729 0.03237 0.06905 ...  
## $ RM : num 6.58 6.42 7.18 7 7.15 ...  
## $ DIS : num 4.09 4.97 4.97 6.06 6.06 ...  
## $ TAX : num 296 242 242 222 222 222 311 311 311 311 ...  
## $ MEDV: num 24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...

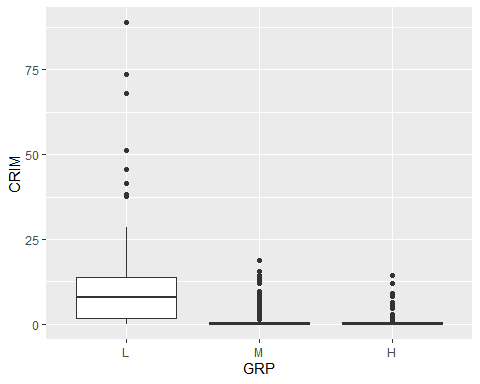
opar <- par(mfrow = c(2, 3))  
for (i in 1:5) {  
 hist(df[,i], main=colnames(df)[i])  
}  
par(opar)



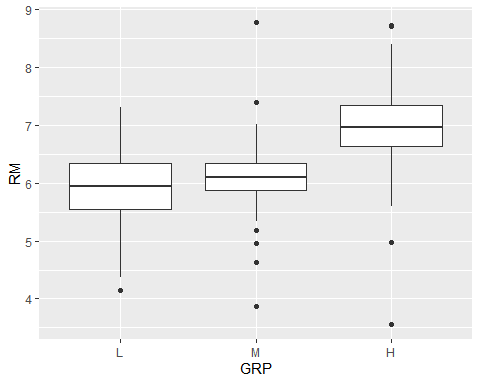
opar <- par(mfrow = c(2, 3))  
for (i in 1:5) {  
 boxplot(df[,i], main=colnames(df)[i])  
}  
par(opar)



housing\_df %>%  
 ggplot(aes(x = GRP, y = CRIM)) +  
 geom\_boxplot()



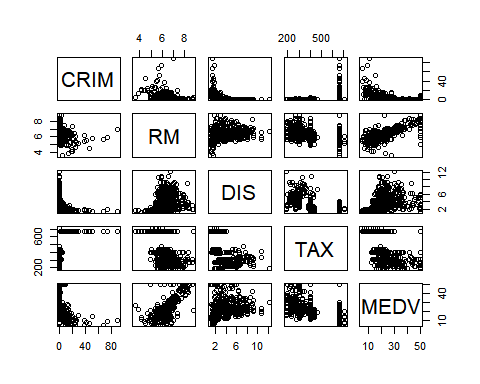
housing\_df %>%  
 ggplot(aes(x = GRP, y = RM)) +  
 geom\_boxplot()



str(df)

## 'data.frame': 506 obs. of 5 variables:  
## $ CRIM: num 0.00632 0.02731 0.02729 0.03237 0.06905 ...  
## $ RM : num 6.58 6.42 7.18 7 7.15 ...  
## $ DIS : num 4.09 4.97 4.97 6.06 6.06 ...  
## $ TAX : num 296 242 242 222 222 222 311 311 311 311 ...  
## $ MEDV: num 24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...

pairs(df)



data <- c(1, 2, 3, NA, 5, NA, NA, 8)  
is.na(data)

## [1] FALSE FALSE FALSE TRUE FALSE TRUE TRUE FALSE

(na\_idx <- which(is.na(data)))

## [1] 4 6 7

data[-na\_idx]

## [1] 1 2 3 5 8

as.vector(na.omit(data))

## [1] 1 2 3 5 8

x <- iris  
x[1, 2] <- NA  
x[1, 3] <- NA  
x[2, 3] <- NA  
x[3, 4] <- NA  
idx <- which(apply(x, 1, function(row) {  
 return(sum(is.na(row)))  
}) >= 2)  
x2 <- x[-idx, ]  
x2

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species Label  
## 2 4.9 3.0 NA 0.2 setosa L  
## 3 4.7 3.2 1.3 NA setosa L  
## 4 4.6 3.1 1.5 0.2 setosa L  
## 5 5.0 3.6 1.4 0.2 setosa L  
## 6 5.4 3.9 1.7 0.4 setosa M  
## 7 4.6 3.4 1.4 0.3 setosa L  
## 8 5.0 3.4 1.5 0.2 setosa L  
## 9 4.4 2.9 1.4 0.2 setosa L  
## 10 4.9 3.1 1.5 0.1 setosa L  
## 11 5.4 3.7 1.5 0.2 setosa L  
## 12 4.8 3.4 1.6 0.2 setosa L  
## 13 4.8 3.0 1.4 0.1 setosa L  
## 14 4.3 3.0 1.1 0.1 setosa L  
## 15 5.8 4.0 1.2 0.2 setosa L  
## 16 5.7 4.4 1.5 0.4 setosa L  
## 17 5.4 3.9 1.3 0.4 setosa L  
## 18 5.1 3.5 1.4 0.3 setosa L  
## 19 5.7 3.8 1.7 0.3 setosa M  
## 20 5.1 3.8 1.5 0.3 setosa L  
## 21 5.4 3.4 1.7 0.2 setosa M  
## 22 5.1 3.7 1.5 0.4 setosa L  
## 23 4.6 3.6 1.0 0.2 setosa L  
## 24 5.1 3.3 1.7 0.5 setosa M  
## 25 4.8 3.4 1.9 0.2 setosa M  
## 26 5.0 3.0 1.6 0.2 setosa L  
## 27 5.0 3.4 1.6 0.4 setosa L  
## 28 5.2 3.5 1.5 0.2 setosa L  
## 29 5.2 3.4 1.4 0.2 setosa L  
## 30 4.7 3.2 1.6 0.2 setosa L  
## 31 4.8 3.1 1.6 0.2 setosa L  
## 32 5.4 3.4 1.5 0.4 setosa L  
## 33 5.2 4.1 1.5 0.1 setosa L  
## 34 5.5 4.2 1.4 0.2 setosa L  
## 35 4.9 3.1 1.5 0.2 setosa L  
## 36 5.0 3.2 1.2 0.2 setosa L  
## 37 5.5 3.5 1.3 0.2 setosa L  
## 38 4.9 3.6 1.4 0.1 setosa L  
## 39 4.4 3.0 1.3 0.2 setosa L  
## 40 5.1 3.4 1.5 0.2 setosa L  
## 41 5.0 3.5 1.3 0.3 setosa L  
## 42 4.5 2.3 1.3 0.3 setosa L  
## 43 4.4 3.2 1.3 0.2 setosa L  
## 44 5.0 3.5 1.6 0.6 setosa L  
## 45 5.1 3.8 1.9 0.4 setosa M  
## 46 4.8 3.0 1.4 0.3 setosa L  
## 47 5.1 3.8 1.6 0.2 setosa L  
## 48 4.6 3.2 1.4 0.2 setosa L  
## 49 5.3 3.7 1.5 0.2 setosa L  
## 50 5.0 3.3 1.4 0.2 setosa L  
## 51 7.0 3.2 4.7 1.4 versicolor M  
## 52 6.4 3.2 4.5 1.5 versicolor M  
## 53 6.9 3.1 4.9 1.5 versicolor M  
## 54 5.5 2.3 4.0 1.3 versicolor M  
## 55 6.5 2.8 4.6 1.5 versicolor M  
## 56 5.7 2.8 4.5 1.3 versicolor M  
## 57 6.3 3.3 4.7 1.6 versicolor M  
## 58 4.9 2.4 3.3 1.0 versicolor M  
## 59 6.6 2.9 4.6 1.3 versicolor M  
## 60 5.2 2.7 3.9 1.4 versicolor M  
## 61 5.0 2.0 3.5 1.0 versicolor M  
## 62 5.9 3.0 4.2 1.5 versicolor M  
## 63 6.0 2.2 4.0 1.0 versicolor M  
## 64 6.1 2.9 4.7 1.4 versicolor M  
## 65 5.6 2.9 3.6 1.3 versicolor M  
## 66 6.7 3.1 4.4 1.4 versicolor M  
## 67 5.6 3.0 4.5 1.5 versicolor M  
## 68 5.8 2.7 4.1 1.0 versicolor M  
## 69 6.2 2.2 4.5 1.5 versicolor M  
## 70 5.6 2.5 3.9 1.1 versicolor M  
## 71 5.9 3.2 4.8 1.8 versicolor M  
## 72 6.1 2.8 4.0 1.3 versicolor M  
## 73 6.3 2.5 4.9 1.5 versicolor M  
## 74 6.1 2.8 4.7 1.2 versicolor M  
## 75 6.4 2.9 4.3 1.3 versicolor M  
## 76 6.6 3.0 4.4 1.4 versicolor M  
## 77 6.8 2.8 4.8 1.4 versicolor M  
## 78 6.7 3.0 5.0 1.7 versicolor M  
## 79 6.0 2.9 4.5 1.5 versicolor M  
## 80 5.7 2.6 3.5 1.0 versicolor M  
## 81 5.5 2.4 3.8 1.1 versicolor M  
## 82 5.5 2.4 3.7 1.0 versicolor M  
## 83 5.8 2.7 3.9 1.2 versicolor M  
## 84 6.0 2.7 5.1 1.6 versicolor H  
## 85 5.4 3.0 4.5 1.5 versicolor M  
## 86 6.0 3.4 4.5 1.6 versicolor M  
## 87 6.7 3.1 4.7 1.5 versicolor M  
## 88 6.3 2.3 4.4 1.3 versicolor M  
## 89 5.6 3.0 4.1 1.3 versicolor M  
## 90 5.5 2.5 4.0 1.3 versicolor M  
## 91 5.5 2.6 4.4 1.2 versicolor M  
## 92 6.1 3.0 4.6 1.4 versicolor M  
## 93 5.8 2.6 4.0 1.2 versicolor M  
## 94 5.0 2.3 3.3 1.0 versicolor M  
## 95 5.6 2.7 4.2 1.3 versicolor M  
## 96 5.7 3.0 4.2 1.2 versicolor M  
## 97 5.7 2.9 4.2 1.3 versicolor M  
## 98 6.2 2.9 4.3 1.3 versicolor M  
## 99 5.1 2.5 3.0 1.1 versicolor M  
## 100 5.7 2.8 4.1 1.3 versicolor M  
## 101 6.3 3.3 6.0 2.5 virginica H  
## 102 5.8 2.7 5.1 1.9 virginica H  
## 103 7.1 3.0 5.9 2.1 virginica H  
## 104 6.3 2.9 5.6 1.8 virginica H  
## 105 6.5 3.0 5.8 2.2 virginica H  
## 106 7.6 3.0 6.6 2.1 virginica H  
## 107 4.9 2.5 4.5 1.7 virginica M  
## 108 7.3 2.9 6.3 1.8 virginica H  
## 109 6.7 2.5 5.8 1.8 virginica H  
## 110 7.2 3.6 6.1 2.5 virginica H  
## 111 6.5 3.2 5.1 2.0 virginica H  
## 112 6.4 2.7 5.3 1.9 virginica H  
## 113 6.8 3.0 5.5 2.1 virginica H  
## 114 5.7 2.5 5.0 2.0 virginica M  
## 115 5.8 2.8 5.1 2.4 virginica H  
## 116 6.4 3.2 5.3 2.3 virginica H  
## 117 6.5 3.0 5.5 1.8 virginica H  
## 118 7.7 3.8 6.7 2.2 virginica H  
## 119 7.7 2.6 6.9 2.3 virginica H  
## 120 6.0 2.2 5.0 1.5 virginica M  
## 121 6.9 3.2 5.7 2.3 virginica H  
## 122 5.6 2.8 4.9 2.0 virginica M  
## 123 7.7 2.8 6.7 2.0 virginica H  
## 124 6.3 2.7 4.9 1.8 virginica M  
## 125 6.7 3.3 5.7 2.1 virginica H  
## 126 7.2 3.2 6.0 1.8 virginica H  
## 127 6.2 2.8 4.8 1.8 virginica M  
## 128 6.1 3.0 4.9 1.8 virginica M  
## 129 6.4 2.8 5.6 2.1 virginica H  
## 130 7.2 3.0 5.8 1.6 virginica H  
## 131 7.4 2.8 6.1 1.9 virginica H  
## 132 7.9 3.8 6.4 2.0 virginica H  
## 133 6.4 2.8 5.6 2.2 virginica H  
## 134 6.3 2.8 5.1 1.5 virginica H  
## 135 6.1 2.6 5.6 1.4 virginica H  
## 136 7.7 3.0 6.1 2.3 virginica H  
## 137 6.3 3.4 5.6 2.4 virginica H  
## 138 6.4 3.1 5.5 1.8 virginica H  
## 139 6.0 3.0 4.8 1.8 virginica M  
## 140 6.9 3.1 5.4 2.1 virginica H  
## 141 6.7 3.1 5.6 2.4 virginica H  
## 142 6.9 3.1 5.1 2.3 virginica H  
## 143 5.8 2.7 5.1 1.9 virginica H  
## 144 6.8 3.2 5.9 2.3 virginica H  
## 145 6.7 3.3 5.7 2.5 virginica H  
## 146 6.7 3.0 5.2 2.3 virginica H  
## 147 6.3 2.5 5.0 1.9 virginica M  
## 148 6.5 3.0 5.2 2.0 virginica H  
## 149 6.2 3.4 5.4 2.3 virginica H  
## 150 5.9 3.0 5.1 1.8 virginica H