grupo2-trabalho1

Unidade Curricular : Métodos Estatísticos

# moda  
moda <- function(x)  
{  
 ux <- unique(x)  
 tab <- tabulate(match(x, ux))  
 ux[tab == max(tab)]  
}  
  
# formula sturges  
sturges <- function(x)  
{  
 floor((1 + log(length(x)) / log(2)))  
}  
  
# amplitude sturges  
amplitude <- function(x)  
{  
 max <- max(x)  
 min <- min(x)  
 nclass <- sturges(x)  
 (max - min) / nclass  
}  
  
# coeficiente de variacao  
coeficiente\_variacao <- function(x)  
{  
 s <- sd(x)  
 m <- mean(x)  
 s/m \* 100  
}  
  
# funcao de assimetria  
assimetria <- function(x)  
{  
 n <-length(x)  
 s <- sd(x)  
 m <- mean(x)  
 1/n \* sum((x - m)^3/s^3)  
}  
  
# funcao de achatamento ou curtose  
achatamento <- function(x)  
{  
 n <- length(x)  
 s <- sd(x)  
 m <- mean(x)  
 1/n \* sum((x - m)^4/s^4) - 3  
}  
  
# funcao de oulier quantiles  
outliers\_quantiles <- function(x, o)  
{  
 sx <- sort(x)  
 n <- length(x)  
 nps <- c(round(n \* .25), round(n \* .50), round(n \* .75))  
 qs <- c(sx[nps[1]],sx[nps[2]], sx[nps[3]])  
}  
  
# funcao de outlier moderado  
outlier\_moderado <- function(x, o)  
{  
 result <- FALSE  
 oqs <- outliers\_quantiles(x, o)  
 om <- c(oqs[1] - 1.5 \* (oqs[3] - oqs[1]), oqs[1] - 3 \* (oqs[3] - oqs[1]), oqs[3] + 1.5 \* (oqs[3] - oqs[1]) , oqs[3] + 3 \* (oqs[3] - oqs[1]))  
 result <- ((o >= om[1] && o <= om[2]) || (o >= om[3] && o <= om[4]))  
 result  
}  
  
# funcao de oulier severo  
outlier\_severo <- function(x, o)  
{  
 result <- FALSE  
 oqs <- outliers\_quantiles(x, o)  
 os <- c(oqs[1] - 3 \* (oqs[3] - oqs[1]), oqs[3] + 3 \* (oqs[3] - oqs[1]))  
 result <- (o < os[1]) || (o > os[2])  
 result  
}  
  
# funcao de outliers tipo moderado  
outliers\_type\_moderado <- function(x, os)  
{  
 result <- c()  
 if(!is.na(os[1]))  
 {  
 for(i in 1:length(os))  
 {  
 m <- outlier\_moderado(x, os[i])  
 if (m)  
 {  
 result <- c(result, os[i])  
 }  
 }  
 }  
 result  
}  
  
# funcao de outliers tipo severo  
outliers\_type\_severo <- function(x, os)  
{  
 result <- c()  
 if(!is.na(os[1]))  
 {  
 for(i in 1:length(os))  
 {  
 s <- outlier\_severo(x, os[i])  
 if (s)  
 {  
 result <- c(result, os[i])  
 }  
 }  
 }  
 result  
}

# importar dataset  
# alterar caminho para o vosso  
# importar dataset  
hour.dataset <- "D:/LEI/matematica/Metodos\_Estatisticos/2122/ME\_Trabalhos\_2122/hour.csv"  
  
# Todos os alugueres todas as horas  
hour <- read.csv(file = hour.dataset)  
  
# Todos os alugueres entre as 7 às 18  
hour.day <- subset(hour, 6<hr & hr<19)  
  
# Todos os alugueres entre as 19 às 6  
hour.night <- subset(hour, hr>18 | hr<7)  
  
# Ver tabelas importadas no RStudio  
View(hour)

# tabela frequencia  
local({  
 .Table <- with(hour, table(hr))  
 cat("\ncounts:\n")  
 print(.Table)  
 cat("\npercentages:\n")  
 print(round(100\*.Table/sum(.Table), 2))  
})

##   
## counts:  
## hr  
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19   
## 726 724 715 697 697 717 725 727 727 727 727 727 728 729 729 729 730 730 728 728   
## 20 21 22 23   
## 728 728 728 728   
##   
## percentages:  
## hr  
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15   
## 4.18 4.17 4.11 4.01 4.01 4.13 4.17 4.18 4.18 4.18 4.18 4.18 4.19 4.19 4.19 4.19   
## 16 17 18 19 20 21 22 23   
## 4.20 4.20 4.19 4.19 4.19 4.19 4.19 4.19

# quartis  
summary(hour)

## instant dteday season yr   
## Min. : 1 Length:17379 Min. :1.000 Min. :0.0000   
## 1st Qu.: 4346 Class :character 1st Qu.:2.000 1st Qu.:0.0000   
## Median : 8690 Mode :character Median :3.000 Median :1.0000   
## Mean : 8690 Mean :2.502 Mean :0.5026   
## 3rd Qu.:13034 3rd Qu.:3.000 3rd Qu.:1.0000   
## Max. :17379 Max. :4.000 Max. :1.0000   
## mnth hr holiday weekday   
## Min. : 1.000 Min. : 0.00 Min. :0.00000 Min. :0.000   
## 1st Qu.: 4.000 1st Qu.: 6.00 1st Qu.:0.00000 1st Qu.:1.000   
## Median : 7.000 Median :12.00 Median :0.00000 Median :3.000   
## Mean : 6.538 Mean :11.55 Mean :0.02877 Mean :3.004   
## 3rd Qu.:10.000 3rd Qu.:18.00 3rd Qu.:0.00000 3rd Qu.:5.000   
## Max. :12.000 Max. :23.00 Max. :1.00000 Max. :6.000   
## workingday weathersit temp atemp   
## Min. :0.0000 Min. :1.000 Min. :0.020 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:0.340 1st Qu.:0.3333   
## Median :1.0000 Median :1.000 Median :0.500 Median :0.4848   
## Mean :0.6827 Mean :1.425 Mean :0.497 Mean :0.4758   
## 3rd Qu.:1.0000 3rd Qu.:2.000 3rd Qu.:0.660 3rd Qu.:0.6212   
## Max. :1.0000 Max. :4.000 Max. :1.000 Max. :1.0000   
## hum windspeed casual registered   
## Min. :0.0000 Min. :0.0000 Min. : 0.00 Min. : 0.0   
## 1st Qu.:0.4800 1st Qu.:0.1045 1st Qu.: 4.00 1st Qu.: 34.0   
## Median :0.6300 Median :0.1940 Median : 17.00 Median :115.0   
## Mean :0.6272 Mean :0.1901 Mean : 35.68 Mean :153.8   
## 3rd Qu.:0.7800 3rd Qu.:0.2537 3rd Qu.: 48.00 3rd Qu.:220.0   
## Max. :1.0000 Max. :0.8507 Max. :367.00 Max. :886.0   
## cnt   
## Min. : 1.0   
## 1st Qu.: 40.0   
## Median :142.0   
## Mean :189.5   
## 3rd Qu.:281.0   
## Max. :977.0

## 1. temp

### Classificação da variável

Variável Qualitativa nominal, e é tratada como Qualitativa nominal normal, visto que, tem grande variedade de variaveis distintas.

### Medidas de Tendência Central

# Contagem de valores  
table(hour$temp)

##   
## 0.02 0.04 0.06 0.08 0.1 0.12 0.14 0.16 0.18 0.2 0.22 0.24 0.26 0.28 0.3 0.32   
## 17 16 16 17 51 76 138 230 155 354 421 522 559 301 641 611   
## 0.34 0.36 0.38 0.4 0.42 0.44 0.46 0.48 0.5 0.52 0.54 0.56 0.58 0.6 0.62 0.64   
## 645 671 372 614 548 507 559 288 531 556 569 579 305 675 726 692   
## 0.66 0.68 0.7 0.72 0.74 0.76 0.78 0.8 0.82 0.84 0.86 0.88 0.9 0.92 0.94 0.96   
## 693 349 690 570 516 396 169 325 213 138 131 53 90 49 17 16   
## 0.98 1   
## 1 1

#### Moda

hour.temp.sort <- sort(hour$temp)  
  
hour.temp.mode <- moda(hour.temp.sort)  
hour.temp.mode

## [1] 0.62

#### Média

hour.temperatura.media <- mean(hour$temp)  
hour.temperatura.media

## [1] 0.4969872

#### Mediana

hour.temperatura.mediana <- median(hour$temp)  
hour.temperatura.mediana

## [1] 0.5

#### Quartil

hour.temperatura.quantile <- quantile(hour$temp)  
hour.temperatura.quantile

## 0% 25% 50% 75% 100%   
## 0.02 0.34 0.50 0.66 1.00

### Medidas de Dispersão Central

#### Variância

hour.temperatura.variancia <- var(hour$temp)  
hour.temperatura.variancia

## [1] 0.03707786

#### Desvio Padrão

hour.temperatura.desvio\_padrao <- sd(hour$temp)  
hour.temperatura.desvio\_padrao

## [1] 0.1925561

#### Coeficiente de Variação

hour.temperatura.coeficiente\_variacao <- coeficiente\_variacao(hour$temp)  
hour.temperatura.coeficiente\_variacao

## [1] 38.74469

### Valores Extremos

#### Máximo

hour.temperatura.maxima <- max(hour$temp)  
hour.temperatura.maxima

## [1] 1

#### Mínimo

hour.temperatura.minima <- min(hour$temp)  
hour.temperatura.minima

## [1] 0.02

### Medidas de Assimetria

hour.temperatura.assimetria <- assimetria(hour$temp)  
hour.temperatura.assimetria

## [1] -0.006019844

De acordo com o resultado, é assimétrica negativa fraca (ou enviesada para a esquerda).

### Medidas de Curtose ou Achatamento

hour.temperatura.achatamento <- achatamento(hour$temp)  
hour.temperatura.achatamento

## [1] -0.9421553

Curva platicúrtica, achatada, o que significa que os valores estão muito concentrados em torno da média e, consequentemente, existe uma variação baixa.

### Histograma

# formula de sturges  
cat("Número de classes\n")

## Número de classes

hour.temp.sturges.nclass <- sturges(hour$temp)  
hour.temp.sturges.nclass

## [1] 15

# amplitude da classe  
cat("\nAmplitude da classe\n")

##   
## Amplitude da classe

hour.temp.sturges.amplitude <- amplitude(hour$temp)  
hour.temp.sturges.amplitude

## [1] 0.06533333

# tabela de frequencias  
cat("\nClasses\n")

##   
## Classes

hour.temp.sturges.breaks <- seq(hour.temperatura.minima, hour.temperatura.maxima, hour.temp.sturges.amplitude)  
hour.temp.sturges.breaks

## [1] 0.02000000 0.08533333 0.15066667 0.21600000 0.28133333 0.34666667  
## [7] 0.41200000 0.47733333 0.54266667 0.60800000 0.67333333 0.73866667  
## [13] 0.80400000 0.86933333 0.93466667 1.00000000

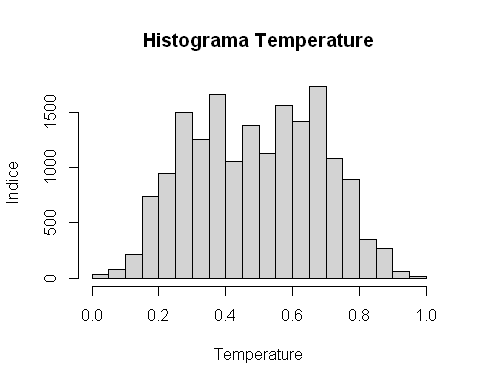
cat("\nTabela de frequências\n")

##   
## Tabela de frequências

table(cut(hour$temp, breaks=hour.temp.sturges.breaks))

##   
## (0.02,0.0853] (0.0853,0.151] (0.151,0.216] (0.216,0.281] (0.281,0.347]   
## 49 265 739 1803 1897   
## (0.347,0.412] (0.412,0.477] (0.477,0.543] (0.543,0.608] (0.608,0.673]   
## 1657 1614 1944 1559 2111   
## (0.673,0.739] (0.739,0.804] (0.804,0.869] (0.869,0.935] (0.935,1]   
## 1609 1406 482 192 34

# histograma  
hist(hour$temp, main="Histograma Temperature", xlab="Temperature", ylab="Indice", nclass=hour.temp.sturges.nclass)



A partir do histograma, pode-se observar que o valor do índice aumenta no intervalo 0.02, a 1, tendo um comportamento crescente até atingir o maior índice no intervalo de 0.608 a 0.673.

### Diagrama de Extremos e Quartis

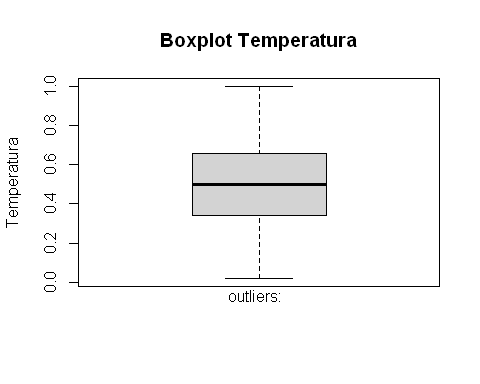
# Outliers  
cat("\nOutliers\n")

##   
## Outliers

hour.temperatura.outliers <- boxplot.stats(hour$temp)$out  
hour.temperatura.outliers

## numeric(0)

# Diagrama  
boxplot(hour$temp, ylab="Temperatura", main="Boxplot Temperatura")  
  
mtext(text = paste("outliers: ", paste(hour.temperatura.outliers, collapse = ", ")), side = 1)

 Verificou-se que não existem outliers presentes no gráfico.

## 2.. weekday

### Classificação da variável

Variável qualitativa ordinal, e é tratada como quantitativa discreta normal, visto que, não tem grande variedade de variaveis distintas.

### Medidas de Tendência Central

#### Moda

hour.weekday.moda <- moda(hour$weekday)  
hour.weekday.moda

## [1] 6

#### Média

hour.weekday.media <- mean(hour$weekday)  
hour.weekday.media

## [1] 3.003683

#### Mediana

hour.weekday.mediana <- median(hour$weekday)  
hour.weekday.mediana

## [1] 3

#### Quartil

hour.weekday.quantile <- quantile(hour$weekday)  
hour.weekday.quantile

## 0% 25% 50% 75% 100%   
## 0 1 3 5 6

### Medidas de Dispersão Central

#### Variância

hour.weekday.variancia <- var(hour$weekday)  
hour.weekday.variancia

## [1] 4.023119

#### Desvio Padrão

hour.weekday.desvio\_padrao <- sd(hour$weekday)  
hour.weekday.desvio\_padrao

## [1] 2.005771

#### Coeficiente de Variação

hour.weekday.coeficiente\_variacao <- coeficiente\_variacao(hour$weekday)  
hour.weekday.coeficiente\_variacao

## [1] 66.77708

### Valores Extremos

#### Máximo

hour.weekday.maxima <- max(hour$weekday)  
hour.weekday.maxima

## [1] 6

#### Mínimo

hour.weekday.minima <- min(hour$weekday)  
hour.weekday.minima

## [1] 0

### Medidas de Assimetria

hour.weekday.assimetria <- assimetria(hour$weekday)  
hour.weekday.assimetria

## [1] -0.002997704

De acordo com o resultado, é assimétrica negativa (ou enviesada para a esquerda).

### Medidas de Curtose ou Achatamento

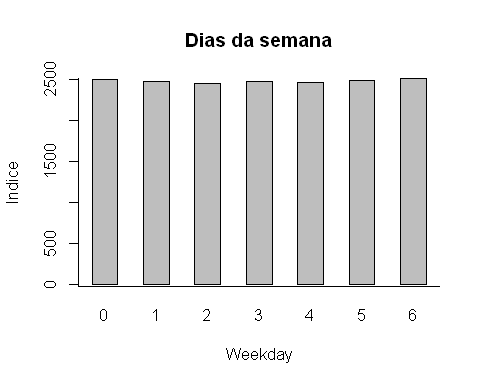
hour.weekday.achatamento <- achatamento(hour$weekday)  
hour.weekday.achatamento

## [1] -1.256181

Curva platicúrtica, achatada, o que significa que os valores estão pouco concentrados em torno da média e, consequentemente, existe uma variação elevada.

### Grafico de barras

count <- table(hour$weekday)  
  
barplot(count, main="Dias da semana", xlab="Weekday", ylab="Indice", space =1.0)  
box(bty="L")

 A partir do Gráfico de barras, pode-se observar que o valor do índice mantém se no intervalo 0 a 6, tendo um comportamento crescente até atingir o maior índice que é 6.

### Diagrama de Extremos e Quartis

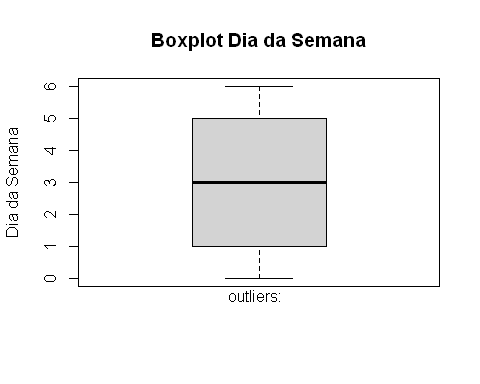
# Outliers  
cat("\nOutliers\n")

##   
## Outliers

hour.weekday.outliers <- boxplot.stats(hour$weekday)$out  
hour.weekday.outliers

## integer(0)

# Diagrama  
boxplot(hour$weekday, ylab="Dia da Semana", main="Boxplot Dia da Semana")  
  
mtext(text = paste("outliers: ", paste(hour.weekday.outliers, collapse = ", ")), side = 1)



Verificou-se que não existem outliers presentes no gráfico.

## 3.. casual

### Classificação da variável

Variável qualitativa ordinal, e é tratada como quantitativa discreta normal, visto que, não tem grande variedade de variaveis distintas.

### Medidas de Tendência Central

#### Moda

hour.casual.moda <- moda(hour$casual)  
hour.casual.moda

## [1] 0

#### Média

hour.casual.media <- mean(hour$casual)  
hour.casual.media

## [1] 35.67622

#### Mediana

hour.casual.mediana <- median(hour$casual)  
hour.casual.mediana

## [1] 17

#### Quartil

hour.casual.quantile <- quantile(hour$casual)  
hour.casual.quantile

## 0% 25% 50% 75% 100%   
## 0 4 17 48 367

### Medidas de Dispersão Central

#### Variância

hour.casual.variancia <- var(hour$casual)  
hour.casual.variancia

## [1] 2430.986

#### Desvio Padrão

hour.casual.desvio\_padrao <- sd(hour$casual)  
hour.casual.desvio\_padrao

## [1] 49.30503

#### Coeficiente de Variação

hour.casual.coeficiente\_variacao <- coeficiente\_variacao(hour$casual)  
hour.casual.coeficiente\_variacao

## [1] 138.2014

### Valores Extremos

#### Máximo

hour.casual.maxima <- max(hour$casual)  
hour.casual.maxima

## [1] 367

#### Mínimo

hour.casual.minima <- min(hour$casual)  
hour.casual.minima

## [1] 0

### Medidas de Assimetria

hour.casual.assimetria <- assimetria(hour$casual)  
hour.casual.assimetria

## [1] 2.498805

De acordo com o resultado, é assimétrica negativa (ou enviesada para a esquerda).

### Medidas de Curtose ou Achatamento

hour.casual.achatamento <- achatamento(hour$casual)  
hour.casual.achatamento

## [1] 7.567262

Curva platicúrtica, achatada, o que significa que os valores estão pouco concentrados em torno da média e, consequentemente, existe uma variação elevada.

### Histograma

# formula de sturges  
cat("Número de classes\n")

## Número de classes

hour.casual.sturges.nclass <- sturges(hour$casual)  
hour.casual.sturges.nclass

## [1] 15

# amplitude da classe  
cat("\nAmplitude da classe\n")

##   
## Amplitude da classe

hour.casual.sturges.amplitude <- amplitude(hour$casual)  
hour.casual.sturges.amplitude

## [1] 24.46667

# tabela de frequencias  
cat("\nClasses\n")

##   
## Classes

hour.casual.sturges.breaks <- seq(hour.casual.minima, hour.casual.maxima, hour.casual.sturges.amplitude)  
hour.casual.sturges.breaks

## [1] 0.00000 24.46667 48.93333 73.40000 97.86667 122.33333 146.80000  
## [8] 171.26667 195.73333 220.20000 244.66667 269.13333 293.60000 318.06667  
## [15] 342.53333 367.00000

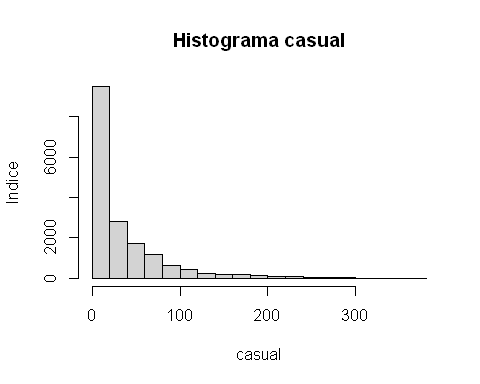
cat("\nTabela de frequências\n")

##   
## Tabela de frequências

table(cut(hour$casual, breaks=hour.casual.sturges.breaks))

##   
## (0,24.5] (24.5,48.9] (48.9,73.4] (73.4,97.9] (97.9,122] (122,147]   
## 8547 2918 1795 947 533 278   
## (147,171] (171,196] (196,220] (220,245] (245,269] (269,294]   
## 212 186 109 112 82 38   
## (294,318] (318,343] (343,367]   
## 21 10 10

# histograma  
hist(hour$casual, main="Histograma casual", xlab="casual", ylab="Indice", nclass=hour.casual.sturges.nclass)



A partir do histograma, pode-se observar que o valor do índice diminui no intervalo 0, a 367, tendo um comportamento decrescente até atingir o maior índice no intervalo de 0 a 24.5.

### Diagrama de Extremos e Quartis

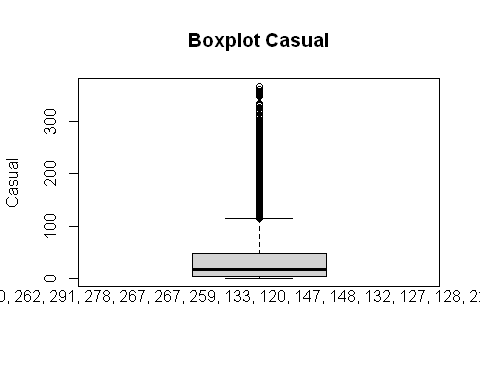
# Outliers  
cat("\nOutliers\n")

##   
## Outliers

hour.casual.outliers <- boxplot.stats(hour$casual)$out  
hour.casual.outliers

## [1] 144 149 124 126 174 168 170 175 138 120 145 172 118 139 166 219 240 174  
## [19] 147 148 128 121 148 156 135 119 167 181 170 179 161 143 126 182 171 180  
## [37] 168 149 118 124 168 205 197 167 162 142 120 178 185 184 217 191 162 134  
## [55] 150 123 132 129 196 143 148 119 138 142 139 166 126 128 122 120 120 195  
## [73] 183 206 158 137 122 173 222 191 187 232 204 191 191 117 135 164 146 180  
## [91] 125 120 171 145 168 172 187 201 180 173 130 139 216 237 183 221 194 214  
## [109] 151 141 116 153 179 133 142 132 115 120 150 188 193 180 168 142 127 121  
## [127] 123 154 161 161 138 126 124 150 148 142 169 147 131 120 131 176 142 115  
## [145] 134 162 135 138 153 193 210 118 141 141 149 124 146 121 159 122 140 140  
## [163] 151 157 175 152 121 136 127 175 197 197 174 177 178 117 144 142 178 177  
## [181] 168 206 179 193 156 136 131 215 198 248 225 194 195 181 142 166 177 237  
## [199] 242 235 224 236 240 222 170 195 195 115 145 167 158 182 171 160 132 128  
## [217] 129 133 139 122 142 128 146 141 120 171 187 221 201 205 183 234 185 164  
## [235] 127 147 177 200 168 129 129 133 147 145 125 126 146 142 130 138 130 176  
## [253] 176 130 155 161 118 130 118 122 175 150 148 123 144 126 128 116 133 125  
## [271] 115 156 158 147 175 151 180 169 132 135 149 178 133 133 129 157 155 196  
## [289] 145 145 140 153 204 187 186 170 160 147 150 174 214 245 205 218 196 204  
## [307] 187 178 141 154 145 144 131 129 217 170 185 119 120 144 119 120 134 180  
## [325] 152 119 115 124 123 148 116 141 116 135 162 127 132 134 118 141 170 164  
## [343] 168 167 180 157 131 143 122 116 119 144 145 125 120 176 191 256 251 262  
## [361] 221 174 135 155 189 212 225 272 198 223 208 186 116 139 164 146 147 130  
## [379] 154 125 115 153 195 171 242 166 179 177 121 121 159 167 169 204 182 177  
## [397] 152 132 136 133 159 169 143 131 164 160 192 171 188 120 137 148 159 141  
## [415] 128 120 142 128 191 165 137 118 121 120 126 143 122 165 122 137 177 141  
## [433] 154 137 115 120 149 156 132 133 123 131 128 165 183 229 186 142 151 145  
## [451] 167 117 132 115 146 176 212 201 208 199 133 140 140 226 286 286 352 357  
## [469] 367 291 221 155 139 129 222 198 218 240 229 233 165 118 120 136 152 125  
## [487] 154 147 117 125 131 199 172 208 191 219 264 237 213 117 148 118 122 123  
## [505] 139 153 146 202 184 263 265 240 275 243 238 132 180 187 190 283 295 236  
## [523] 232 153 127 132 128 120 120 145 123 139 143 145 173 240 218 184 172 115  
## [541] 187 251 320 355 326 321 354 299 227 170 195 198 229 254 260 232 185 151  
## [559] 125 115 117 139 207 274 308 288 311 253 251 197 163 186 227 275 298 282  
## [577] 266 286 262 184 115 156 145 151 122 115 183 260 262 291 278 267 267 259  
## [595] 133 120 147 148 132 127 128 223 226 281 279 240 188 164 128 148 221 220  
## [613] 217 187 241 230 204 156 151 144 207 197 226 229 181 195 131 117 116 140  
## [631] 128 132 157 206 293 257 269 254 233 164 166 128 150 189 255 228 190 225  
## [649] 191 129 133 134 131 151 124 177 235 276 332 361 356 331 279 254 203 118  
## [667] 129 174 258 247 244 236 246 238 209 146 142 120 118 137 124 131 181 208  
## [685] 236 265 307 274 261 235 199 159 130 209 268 301 270 317 290 258 275 266  
## [703] 230 168 121 189 254 233 272 238 181 223 158 132 148 145 179 250 279 259  
## [721] 297 275 248 171 185 139 127 199 276 265 267 236 226 192 153 116 129 128  
## [739] 147 117 168 218 196 236 239 218 219 238 175 128 125 148 184 173 204 186  
## [757] 159 173 184 164 115 123 117 121 121 132 125 121 139 213 254 293 264 295  
## [775] 238 244 256 184 123 148 196 260 267 255 203 240 232 203 146 177 167 203  
## [793] 226 238 217 196 214 194 185 158 161 215 227 237 253 197 181 203 186 176  
## [811] 143 117 131 134 121 143 164 164 177 190 155 163 161 159 177 237 222 175  
## [829] 129 128 122 124 116 120 118 125 120 138 123 143 181 219 269 243 226 199  
## [847] 156 141 170 153 171 162 182 152 122 156 121 148 130 191 178 238 256 254  
## [865] 245 188 173 180 124 130 116 129 162 157 173 186 206 208 199 151 115 123  
## [883] 183 154 181 231 195 191 164 117 143 155 222 195 183 205 197 179 169 133  
## [901] 117 155 161 208 161 164 143 122 125 117 119 125 203 214 228 248 246 227  
## [919] 220 167 163 178 213 235 213 186 164 148 116 124 130 119 128 127 116 144  
## [937] 166 180 266 289 242 250 287 256 225 164 138 121 130 117 148 143 217 186  
## [955] 150 174 132 122 127 150 225 166 125 132 119 120 120 129 128 219 187 237  
## [973] 240 257 209 205 184 157 158 226 248 263 249 179 222 219 227 140 135 187  
## [991] 191 209 163 168 165 181 145 173 220 233 260 175 172 194 247 238 232 284  
## [1009] 228 213 156 134 115 127 168 119 137 116 119 228 287 327 325 312 350 295  
## [1027] 232 169 160 183 188 240 225 215 194 177 120 129 119 124 189 228 273 250  
## [1045] 307 253 195 171 138 186 250 257 266 265 254 227 135 134 217 191 233 302  
## [1063] 271 275 256 174 119 160 222 206 180 208 230 202 195 161 117 150 153 158  
## [1081] 130 198 258 362 310 269 279 317 268 183 121 138 131 192 202 235 243 251  
## [1099] 294 193 174 150 156 200 218 249 213 203 193 134 122 131 204 267 273 335  
## [1117] 308 325 347 165 167 191 236 243 235 240 182 153 142 137 125 165 197 264  
## [1135] 310 325 310 257 248 138 124 120 118 136 117 134 150 122 149 149 119 143  
## [1153] 258 268 280 216 227 183 133 180 195 262 292 304 260 151 134 117 179 175  
## [1171] 175 144 125 117 125 125 132 121 186 224 240 233 158 117 119 123 142 122  
## [1189] 148 164 167 139

# Diagrama  
boxplot(hour$casual, ylab="Casual", main="Boxplot Casual")  
  
mtext(text = paste("outliers: ", paste(hour.casual.outliers, collapse = ", ")), side = 1)



Verificou-se que existem outliers presentes no gráfico.

## 4.. atemp

### Classificação da variável

Variável qualitativa ordinal, e é tratada como quantitativa discreta normal, visto que, não tem grande variedade de variaveis distintas.

### Medidas de Tendência Central

#### Moda

hour.atemp.moda <- moda(hour$atemp)  
hour.atemp.moda

## [1] 0.6212

#### Média

hour.atemp.media <- mean(hour$atemp)  
hour.atemp.media

## [1] 0.4757751

#### Mediana

hour.atemp.mediana <- median(hour$atemp)  
hour.atemp.mediana

## [1] 0.4848

#### Quartil

hour.atemp.quantile <- quantile(hour$atemp)  
hour.atemp.quantile

## 0% 25% 50% 75% 100%   
## 0.0000 0.3333 0.4848 0.6212 1.0000

### Medidas de Dispersão Central

#### Variância

hour.atemp.variancia <- var(hour$atemp)  
hour.atemp.variancia

## [1] 0.0295325

#### Desvio Padrão

hour.atemp.desvio\_padrao <- sd(hour$atemp)  
hour.atemp.desvio\_padrao

## [1] 0.1718502

#### Coeficiente de Variação

hour.atemp.coeficiente\_variacao <- coeficiente\_variacao(hour$atemp)  
hour.atemp.coeficiente\_variacao

## [1] 36.12005

### Valores Extremos

#### Máximo

hour.atemp.maxima <- max(hour$atemp)  
hour.atemp.maxima

## [1] 1

#### Mínimo

hour.atemp.minima <- min(hour$atemp)  
hour.atemp.minima

## [1] 0

### Medidas de Assimetria

hour.atemp.assimetria <- assimetria(hour$atemp)  
hour.atemp.assimetria

## [1] -0.09041325

De acordo com o resultado, é assimétrica negativa (ou enviesada para a esquerda).

### Medidas de Curtose ou Achatamento

hour.atemp.achatamento <- achatamento(hour$atemp)  
hour.atemp.achatamento

## [1] -0.8457618

Curva platicúrtica, achatada, o que significa que os valores estão pouco concentrados em torno da média e, consequentemente, existe uma variação elevada.

### Histograma

# formula de sturges  
cat("Número de classes\n")

## Número de classes

hour.atemp.sturges.nclass <- sturges(hour$atemp)  
hour.atemp.sturges.nclass

## [1] 15

# amplitude da classe  
cat("\nAmplitude da classe\n")

##   
## Amplitude da classe

hour.atemp.sturges.amplitude <- amplitude(hour$atemp)  
hour.atemp.sturges.amplitude

## [1] 0.06666667

# tabela de frequencias  
cat("\nClasses\n")

##   
## Classes

hour.atemp.sturges.breaks <- seq(hour.atemp.minima, hour.atemp.maxima, hour.atemp.sturges.amplitude)  
hour.atemp.sturges.breaks

## [1] 0.00000000 0.06666667 0.13333333 0.20000000 0.26666667 0.33333333  
## [7] 0.40000000 0.46666667 0.53333333 0.60000000 0.66666667 0.73333333  
## [13] 0.80000000 0.86666667 0.93333333 1.00000000

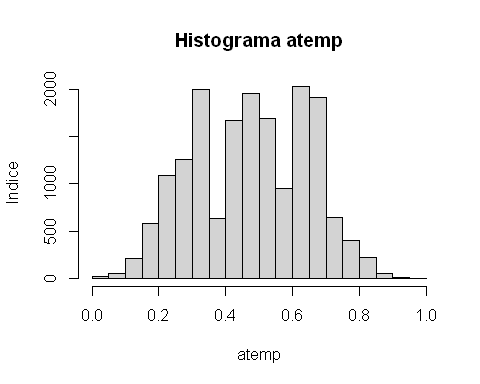
cat("\nTabela de frequências\n")

##   
## Tabela de frequências

table(cut(hour$atemp, breaks=hour.atemp.sturges.breaks))

##   
## (0,0.0667] (0.0667,0.133] (0.133,0.2] (0.2,0.267] (0.267,0.333]   
## 35 162 665 1510 2388   
## (0.333,0.4] (0.4,0.467] (0.467,0.533] (0.533,0.6] (0.6,0.667]   
## 1075 2228 2591 1444 2525   
## (0.667,0.733] (0.733,0.8] (0.8,0.867] (0.867,0.933] (0.933,1]   
## 1862 599 245 44 4

# histograma  
hist(hour$atemp, main="Histograma atemp", xlab="atemp", ylab="Indice", nclass=hour.atemp.sturges.nclass)



A partir do histograma, pode-se observar que o valor do índice aumenta no intervalo 0 a 0.5337 e diminui no intervalo 0.5337 a 1, tendo um comportamento crescente e decrescente até atingir o maior índice no intervalo de 0.467 a 0.533.

### Diagrama de Extremos e Quartis

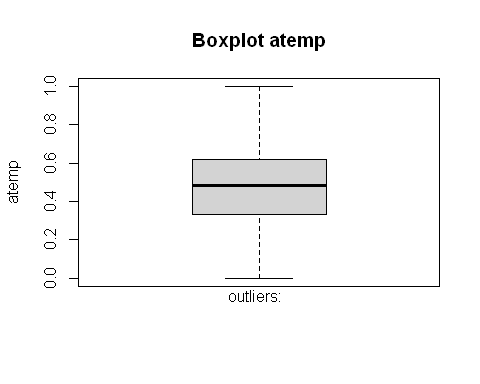
# Outliers  
cat("\nOutliers\n")

##   
## Outliers

hour.atemp.outliers <- boxplot.stats(hour$atemp)$out  
hour.atemp.outliers

## numeric(0)

# Diagrama  
boxplot(hour$atemp, ylab="atemp", main="Boxplot atemp")  
  
mtext(text = paste("outliers: ", paste(hour.atemp.outliers, collapse = ", ")), side = 1)



Verificou-se que não existem outliers presentes no gráfico.

## 5.. hum

### Classificação da variável

Variável qualitativa ordinal, e é tratada como quantitativa discreta normal, visto que, não tem grande variedade de variaveis distintas.

### Medidas de Tendência Central

#### Moda

hour.hum.moda <- moda(hour$hum)  
hour.hum.moda

## [1] 0.88

#### Média

hour.hum.media <- mean(hour$hum)  
hour.hum.media

## [1] 0.6272288

#### Mediana

hour.hum.mediana <- median(hour$hum)  
hour.hum.mediana

## [1] 0.63

#### Quartil

hour.hum.quantile <- quantile(hour$hum)  
hour.hum.quantile

## 0% 25% 50% 75% 100%   
## 0.00 0.48 0.63 0.78 1.00

### Medidas de Dispersão Central

#### Variância

hour.hum.variancia <- var(hour$hum)  
hour.hum.variancia

## [1] 0.03722192

#### Desvio Padrão

hour.hum.desvio\_padrao <- sd(hour$hum)  
hour.hum.desvio\_padrao

## [1] 0.1929298

#### Coeficiente de Variação

hour.hum.coeficiente\_variacao <- coeficiente\_variacao(hour$hum)  
hour.hum.coeficiente\_variacao

## [1] 30.75908

### Valores Extremos

#### Máximo

hour.hum.maxima <- max(hour$hum)  
hour.hum.maxima

## [1] 1

#### Mínimo

hour.hum.minima <- min(hour$hum)  
hour.hum.minima

## [1] 0

### Medidas de Assimetria

hour.hum.assimetria <- assimetria(hour$hum)  
hour.hum.assimetria

## [1] -0.1112679

De acordo com o resultado, é assimétrica negativa (ou enviesada para a esquerda).

### Medidas de Curtose ou Achatamento

hour.hum.achatamento <- achatamento(hour$hum)  
hour.hum.achatamento

## [1] -0.8264745

Curva platicúrtica, achatada, o que significa que os valores estão pouco concentrados em torno da média e, consequentemente, existe uma variação elevada.

### Histograma

# formula de sturges  
cat("Número de classes\n")

## Número de classes

hour.hum.sturges.nclass <- sturges(hour$hum)  
hour.hum.sturges.nclass

## [1] 15

# amplitude da classe  
cat("\nAmplitude da classe\n")

##   
## Amplitude da classe

hour.hum.sturges.amplitude <- amplitude(hour$hum)  
hour.hum.sturges.amplitude

## [1] 0.06666667

# tabela de frequencias  
cat("\nClasses\n")

##   
## Classes

hour.hum.sturges.breaks <- seq(hour.hum.minima, hour.hum.maxima, hour.hum.sturges.amplitude)  
hour.hum.sturges.breaks

## [1] 0.00000000 0.06666667 0.13333333 0.20000000 0.26666667 0.33333333  
## [7] 0.40000000 0.46666667 0.53333333 0.60000000 0.66666667 0.73333333  
## [13] 0.80000000 0.86666667 0.93333333 1.00000000

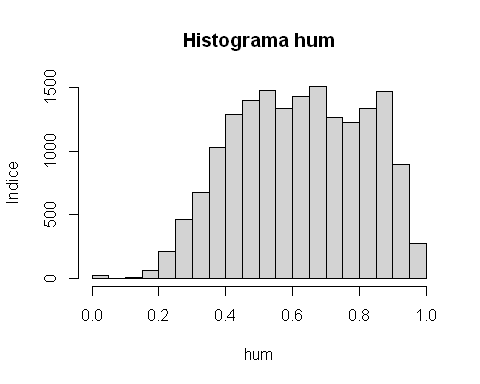
cat("\nTabela de frequências\n")

##   
## Tabela de frequências

table(cut(hour$hum, breaks=hour.hum.sturges.breaks))

##   
## (0,0.0667] (0.0667,0.133] (0.133,0.2] (0.2,0.267] (0.267,0.333]   
## 0 4 69 292 766   
## (0.333,0.4] (0.4,0.467] (0.467,0.533] (0.533,0.6] (0.6,0.667]   
## 1326 1603 1921 1977 1818   
## (0.667,0.733] (0.733,0.8] (0.8,0.867] (0.867,0.933] (0.933,1]   
## 1823 1790 1409 1725 834

# histograma  
hist(hour$hum, main="Histograma hum", xlab="hum", ylab="Indice", nclass=hour.hum.sturges.nclass)



A partir do histograma, pode-se observar que o valor do índice aumenta no intervalo 0 a 0.6 e diminui no intervalo 0.6 a 1, tendo um comportamento crescente e decrescente até atingir o maior índice no intervalo de 0.533 a 0.6.

### Diagrama de Extremos e Quartis

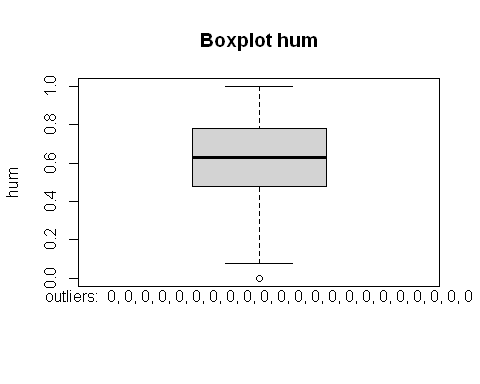
# Outliers  
cat("\nOutliers\n")

##   
## Outliers

hour.hum.outliers <- boxplot.stats(hour$hum)$out  
hour.hum.outliers

## [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

# Diagrama  
boxplot(hour$hum, ylab="hum", main="Boxplot hum")  
  
mtext(text = paste("outliers: ", paste(hour.hum.outliers, collapse = ", ")), side = 1)



Verificou-se que existem outliers presentes no gráfico no ponto 0.

## 6.. cnt

### Classificação da variável

Variável qualitativa ordinal, e é tratada como quantitativa discreta normal, visto que, não tem grande variedade de variaveis distintas.

### Medidas de Tendência Central

#### Moda

hour.cnt.moda <- moda(hour$cnt)  
hour.cnt.moda

## [1] 5

#### Média

hour.cnt.media <- mean(hour$cnt)  
hour.cnt.media

## [1] 189.4631

#### Mediana

hour.cnt.mediana <- median(hour$cnt)  
hour.cnt.mediana

## [1] 142

#### Quartil

hour.cnt.quantile <- quantile(hour$cnt)  
hour.cnt.quantile

## 0% 25% 50% 75% 100%   
## 1 40 142 281 977

### Medidas de Dispersão Central

#### Variância

hour.cnt.variancia <- var(hour$cnt)  
hour.cnt.variancia

## [1] 32901.46

#### Desvio Padrão

hour.cnt.desvio\_padrao <- sd(hour$cnt)  
hour.cnt.desvio\_padrao

## [1] 181.3876

#### Coeficiente de Variação

hour.cnt.coeficiente\_variacao <- coeficiente\_variacao(hour$cnt)  
hour.cnt.coeficiente\_variacao

## [1] 95.7377

### Valores Extremos

#### Máximo

hour.cnt.maxima <- max(hour$cnt)  
hour.cnt.maxima

## [1] 977

#### Mínimo

hour.cnt.minima <- min(hour$cnt)  
hour.cnt.minima

## [1] 1

### Medidas de Assimetria

hour.cnt.assimetria <- assimetria(hour$cnt)  
hour.cnt.assimetria

## [1] 1.277191

De acordo com o resultado, é assimétrica negativa (ou enviesada para a esquerda).

### Medidas de Curtose ou Achatamento

hour.cnt.achatamento <- achatamento(hour$cnt)  
hour.cnt.achatamento

## [1] 1.415942

Curva platicúrtica, achatada, o que significa que os valores estão pouco concentrados em torno da média e, consequentemente, existe uma variação elevada.

### Histograma

# formula de sturges  
cat("Número de classes\n")

## Número de classes

hour.cnt.sturges.nclass <- sturges(hour$cnt)  
hour.cnt.sturges.nclass

## [1] 15

# amplitude da classe  
cat("\nAmplitude da classe\n")

##   
## Amplitude da classe

hour.cnt.sturges.amplitude <- amplitude(hour$cnt)  
hour.cnt.sturges.amplitude

## [1] 65.06667

# tabela de frequencias  
cat("\nClasses\n")

##   
## Classes

hour.cnt.sturges.breaks <- seq(hour.cnt.minima, hour.cnt.maxima, hour.cnt.sturges.amplitude)  
hour.cnt.sturges.breaks

## [1] 1.00000 66.06667 131.13333 196.20000 261.26667 326.33333 391.40000  
## [8] 456.46667 521.53333 586.60000 651.66667 716.73333 781.80000 846.86667  
## [15] 911.93333 977.00000

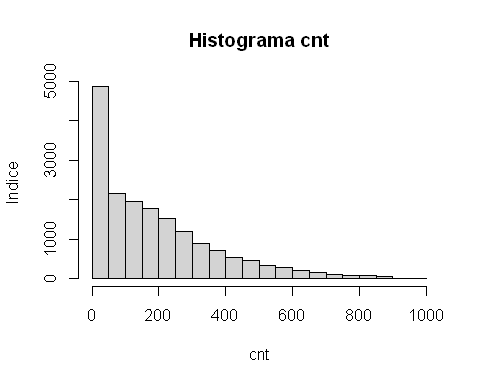
cat("\nTabela de frequências\n")

##   
## Tabela de frequências

table(cut(hour$cnt, breaks=hour.cnt.sturges.breaks))

##   
## (1,66.1] (66.1,131] (131,196] (196,261] (261,326] (326,391] (391,456]   
## 5458 2669 2392 1900 1413 1006 704   
## (456,522] (522,587] (587,652] (652,717] (717,782] (782,847] (847,912]   
## 552 394 263 190 104 105 55   
## (912,977]   
## 16

# histograma  
hist(hour$cnt, main="Histograma cnt", xlab="cnt", ylab="Indice", nclass=hour.cnt.sturges.nclass)



A partir do histograma, pode-se observar que o valor do índice diminui no intervalo 0, a 977, tendo um comportamento decrescente até atingir o maior índice no intervalo de 1 a 66.1.

### Diagrama de Extremos e Quartis

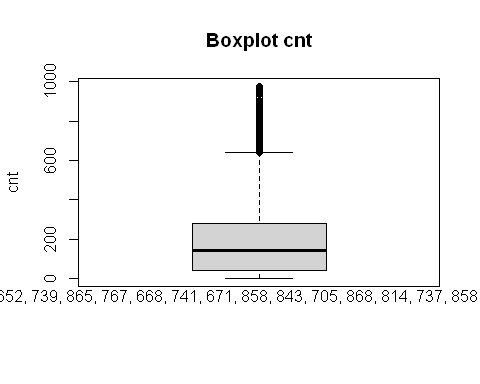
# Outliers  
cat("\nOutliers\n")

##   
## Outliers

hour.cnt.outliers <- boxplot.stats(hour$cnt)$out  
hour.cnt.outliers

## [1] 651 644 712 676 734 662 782 749 713 746 651 686 690 679 685 648 721 801  
## [19] 750 801 729 779 649 810 801 662 957 830 686 657 664 684 644 658 654 703  
## [37] 681 729 757 800 684 744 759 822 698 655 643 646 692 744 704 656 738 671  
## [55] 678 678 660 658 681 712 676 673 781 775 677 748 776 700 819 668 649 691  
## [73] 654 732 709 664 702 654 653 683 664 681 743 666 729 813 704 706 643 769  
## [91] 680 717 710 705 732 770 779 659 678 733 650 873 846 852 868 745 812 669  
## [109] 704 730 672 645 785 785 719 692 798 752 653 781 710 671 839 796 693 827  
## [127] 785 694 668 679 647 702 644 684 686 678 644 662 665 834 822 645 710 850  
## [145] 790 668 724 782 681 869 813 700 793 723 651 800 831 681 653 713 857 744  
## [163] 671 719 867 823 653 823 693 723 673 811 795 643 691 672 646 833 791 656  
## [181] 900 824 687 843 804 643 705 697 747 730 646 722 689 849 872 649 872 819  
## [199] 674 830 814 702 795 825 713 835 667 755 794 661 770 772 679 643 657 771  
## [217] 777 691 715 738 847 741 743 869 877 788 913 891 698 729 738 688 676 699  
## [235] 751 827 812 665 760 820 857 681 837 891 652 739 865 767 668 741 671 858  
## [253] 843 705 868 814 737 858 862 686 698 810 811 730 673 818 812 812 854 682  
## [271] 851 848 649 682 897 832 677 668 791 669 654 644 655 893 815 723 878 740  
## [289] 783 683 707 820 941 744 812 736 699 845 834 693 864 818 808 870 812 643  
## [307] 754 844 853 690 856 839 725 863 839 662 646 808 835 719 772 792 694 668  
## [325] 757 729 647 696 701 671 730 871 968 750 970 877 770 925 977 758 884 852  
## [343] 674 766 894 808 706 704 715 654 783 729 656 694 683 842 774 672 797 886  
## [361] 892 790 976 900 757 846 805 688 707 654 750 680 646 776 702 666 691 723  
## [379] 750 898 853 805 967 822 838 953 884 794 905 899 705 808 667 663 682 686  
## [397] 750 727 722 712 680 649 648 784 856 715 687 809 917 810 738 901 887 785  
## [415] 900 761 647 743 710 659 713 806 784 839 948 844 798 827 692 743 745 837  
## [433] 670 737 766 835 943 838 817 888 884 834 890 788 703 714 711 711 691 731  
## [451] 675 728 922 786 761 938 826 801 963 858 779 886 809 734 817 665 668 760  
## [469] 750 711 723 680 689 678 687 648 668 693 654 651 686 654 724 653 691 646  
## [487] 711 663 665 692 679 729 731 708 692 721 743 731 759 659 724 688 679 662  
## [505] 678

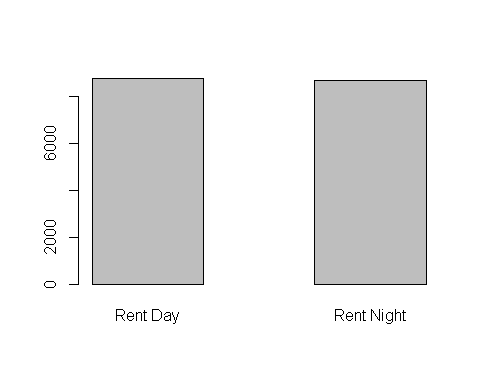
# Diagrama  
boxplot(hour$cnt, ylab="cnt", main="Boxplot cnt")  
  
mtext(text = paste("outliers: ", paste(hour.cnt.outliers, collapse = ", ")), side = 1)



Verificou-se que existem bastantes outliers presentes no gráfico.

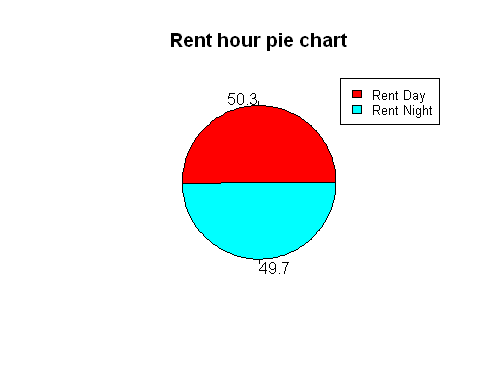
### Gráfico de Barras

values <- c(nrow(hour.day), nrow(hour.night))   
labels <- c("Rent Day", "Rent Night")  
  
barplot(values, names.arg=labels, space =1.0)



### Gráfico Circular

x <- c(nrow(hour.day), nrow(hour.night))  
labels <- c("Rent Day", "Rent Night")  
  
piepercent<- round(100\*x/sum(x), 1)  
  
pie(x, labels = piepercent, main = "Rent hour pie chart",col = rainbow(length(x)))  
legend("topright", c("Rent Day", "Rent Night"), cex = 0.8,  
 fill = rainbow(length(x)))



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.