

```
dados = read.csv("tempo.csv", sep = ";", na.strings = "", stringsAsFactors = T)
summary(dados)
```

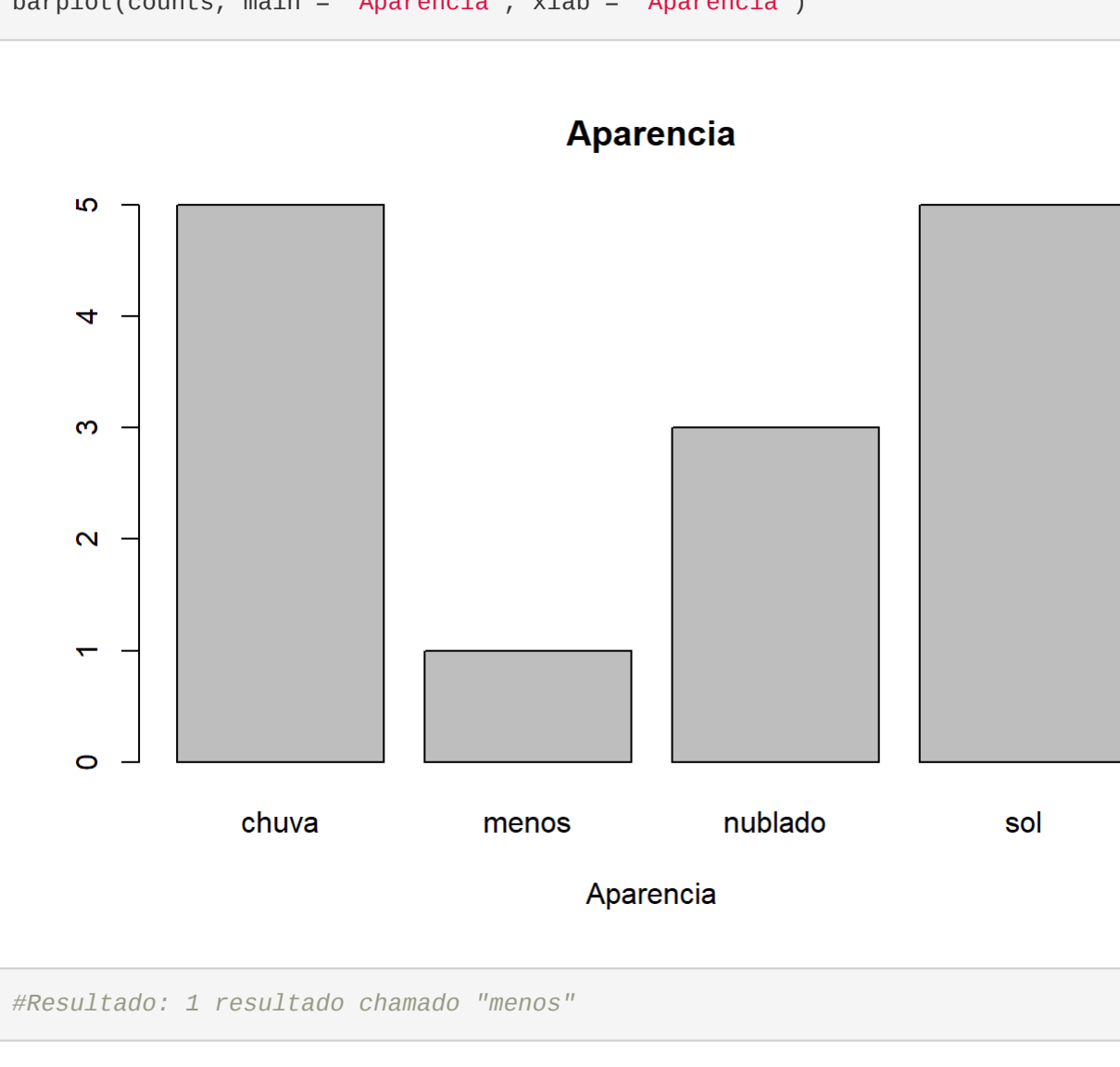
```
##      Aparencia  Temperatura      Umidade      Vento  Jogar
## chuva  :5      Min.   : 64.00      Min.   : 65.00  FALSO   :7      nao:5
## menos  :1      1st Qu.: 69.25      1st Qu.: 70.00  VERDADEIRO:6      sim:9
## nublado:3      Median : 73.50      Median : 85.00  NA's     :1
## sol    :5      Mean   : 155.57      Mean   : 89.77
##      3rd Qu.: 80.75      3rd Qu.: 90.00
##      Max.   :1220.00      Max.   :200.00
##
##      NA's      :1
```

```
head(dados)
```

```
##      Aparencia  Temperatura  Umidade      Vento  Jogar
## 1      sol          85          85      FALSO   nao
## 2      sol          80          90  VERDADEIRO  nao
## 3  nublado          83          86      FALSO   sim
## 4      chuva          70          NA      FALSO   sim
## 5      chuva          68          80      FALSO   sim
## 6      chuva          65          70  VERDADEIRO  nao
```

Explorar dados categóricos - Aparencia

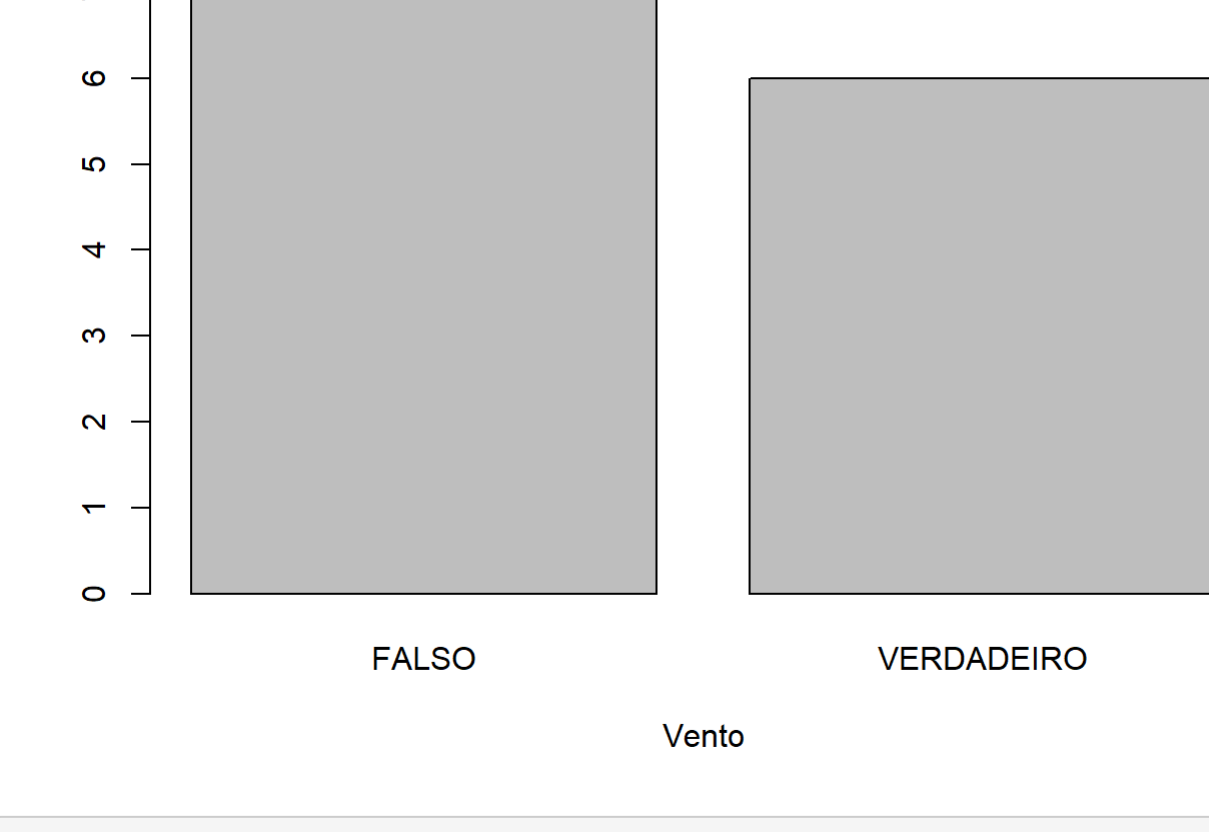
```
counts = table(dados$Aparencia)
barplot(counts, main = "Aparencia", xlab = "Aparencia")
```



```
#Resultado: 1 resultado chamado "menos"
```

Explorar dados categóricos - vento

```
counts = table(dados$Vento)
barplot(counts, main = "Vento", xlab = "Vento")
```



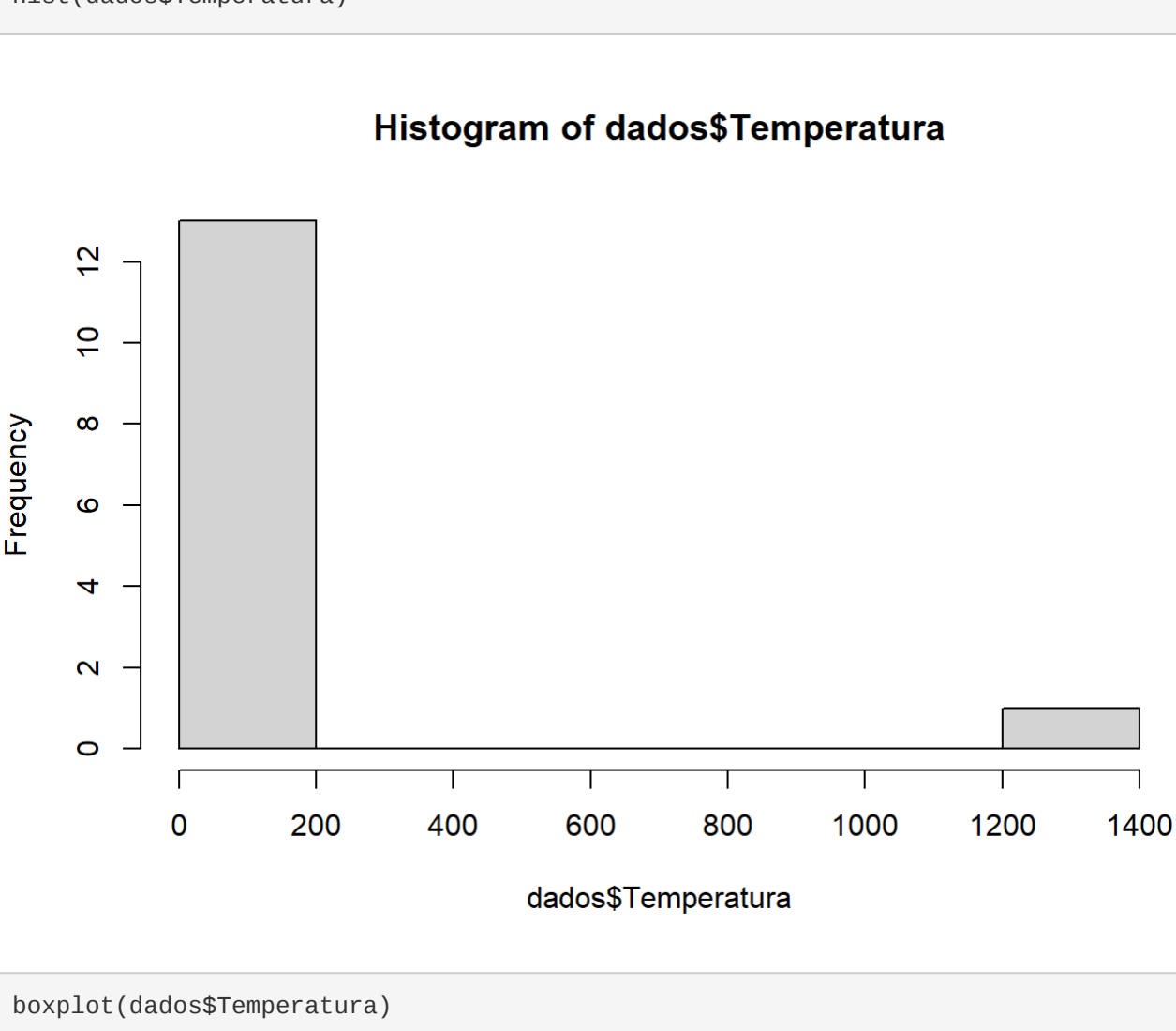
```
#Resultado: Tudo ok
```

Explorar dados numéricos - Temperatura

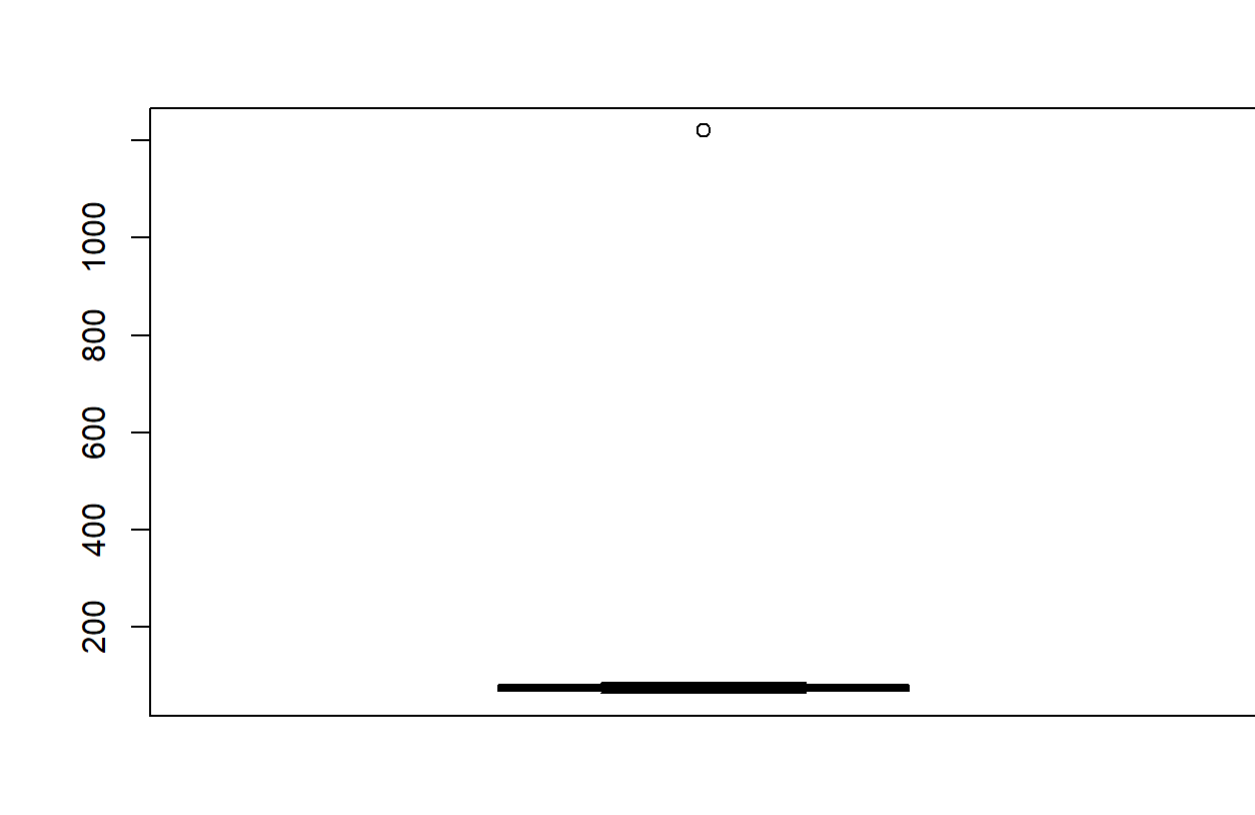
```
summary(dados$Temperatura)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 64.00   69.25   73.50   155.57   80.75  1220.00
```

```
hist(dados$Temperatura)
```



```
boxplot(dados$Temperatura)
```



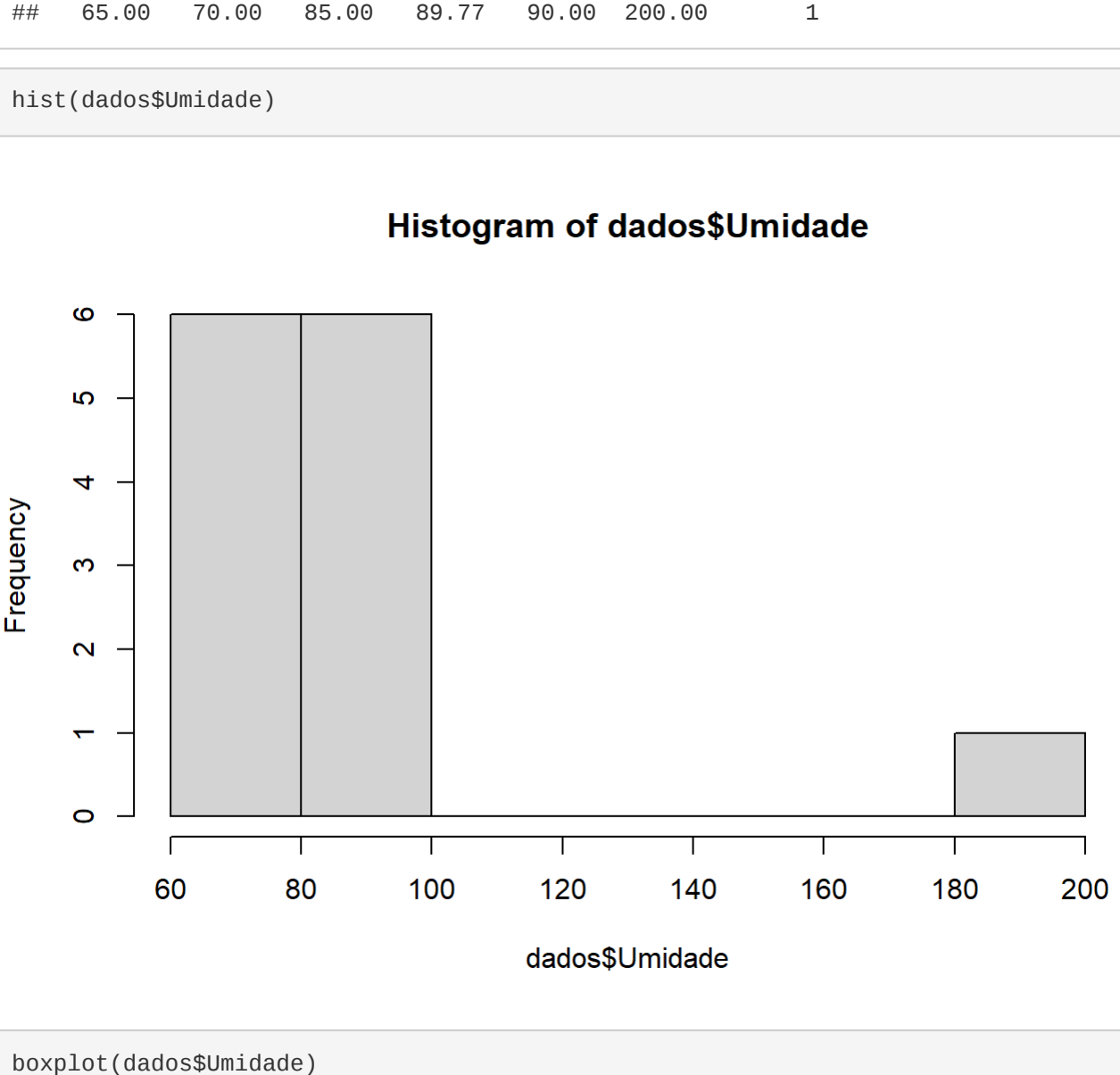
```
#Resultado: 1 dado outlier
```

Explorar dados numéricos - Umidade

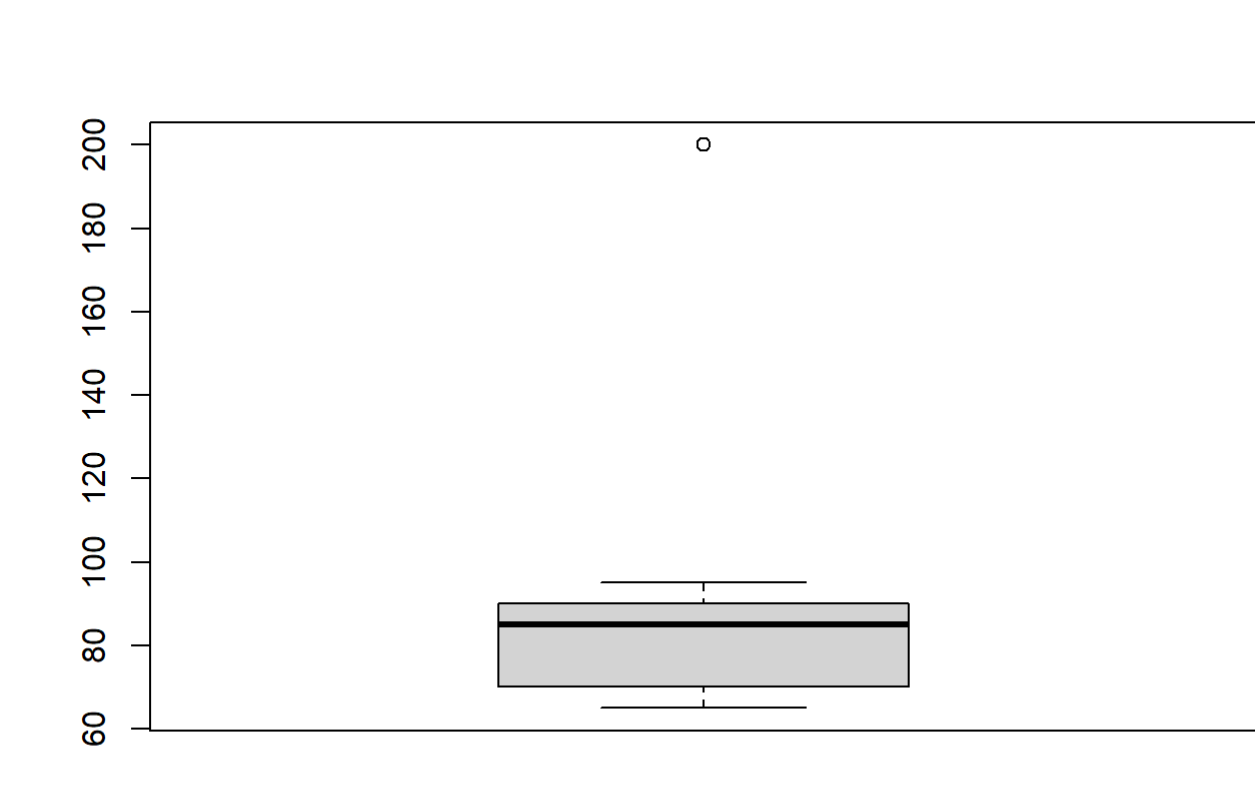
```
summary(dados$Umidade)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 65.00   70.00   85.00   89.77   90.00   200.00      1
```

```
hist(dados$Umidade)
```



```
boxplot(dados$Umidade)
```



```
#Resultado: 1 dado NA, 1 outlier
```

Verificar valores faltantes

```
dados[!complete.cases(dados),]
```

```
##      Aparencia  Temperatura  Umidade  Vento  Jogar
## 4      chuva          70          NA  FALSO   sim
## 10     chuva          75          200 <NA>   sim
```

```
#Resultado: Falta 1 dado em Umidade e 1 em vento
```

substituir valores faltantes e retirar outliers

```
#Umidade - valores NA
summary(dados$Umidade)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 65.00   70.00   85.00   89.77   90.00   200.00      1
```

```
median(dados$Umidade, na.rm = T)
```

```
## [1] 85
```

```
dados[is.na(dados$Umidade),]$Umidade = median(dados$Umidade, na.rm = T)
dados[!complete.cases(dados),]
```

```
##      Aparencia  Temperatura  Umidade  Vento  Jogar
## 10     chuva          75          200 <NA>   sim
```

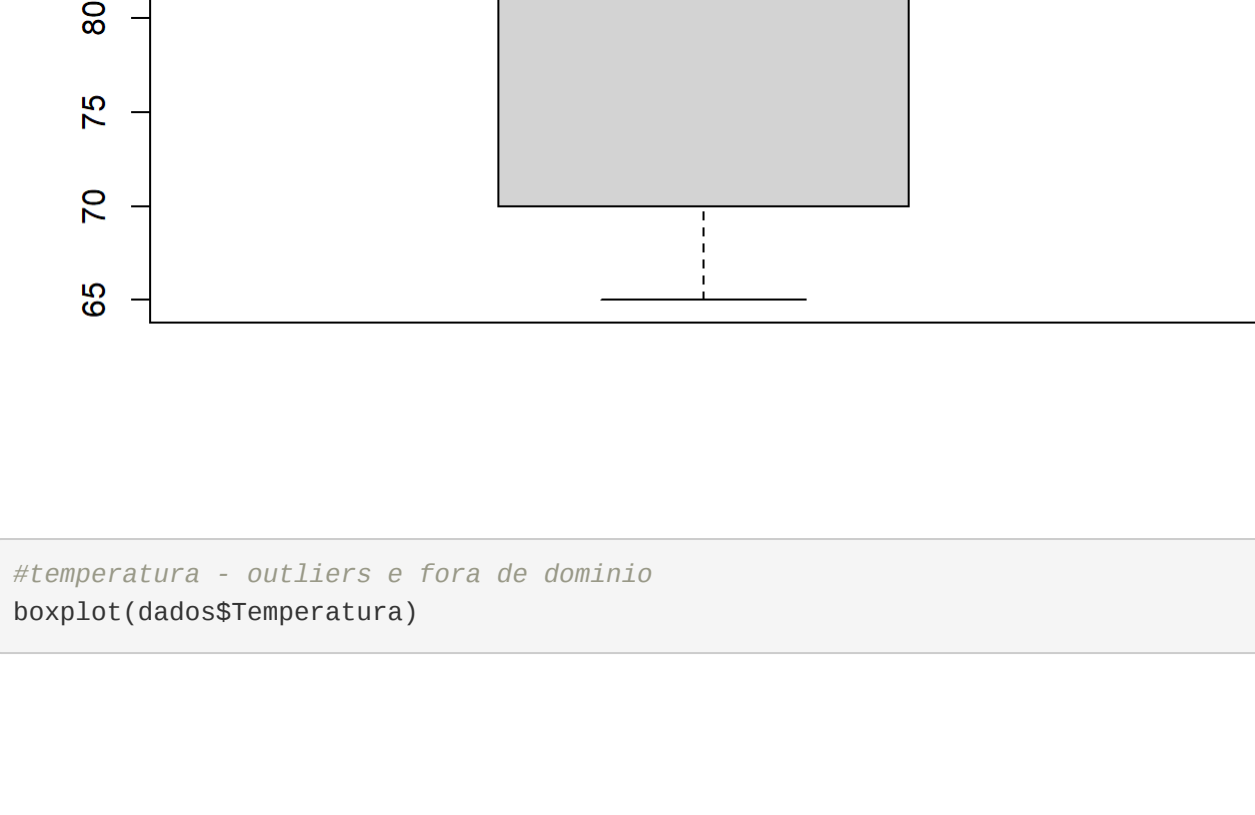
```
#Umidade - outliers e fora de dominio
boxplot(dados$Umidade)
```



```
dados[dados$Umidade > 100,]$Umidade = median(dados$Umidade, na.rm = T)
boxplot(dados$Umidade)
```



```
#Temperatura - outliers e fora de dominio
boxplot(dados$Temperatura)
```



```
dados[dados$Temperatura > 100, ]
```

```
##      Aparencia  Temperatura  Umidade  Vento  Jogar
## 8      sol          1220          95  FALSO   nao
```

```
median(dados$Temperatura)
```

```
## [1] 73.5
```

```
dados[dados$Temperatura > 100,]$Temperatura = median(dados$Temperatura)
```

substituir valores faltantes e escritos errados nas variaveis categoricas

```
#Vento - valores NA
unique(dados$Vento)
```

```
## [1] FALSO  VERDADEIRO <NA>
## Levels: FALSO VERDADEIRO
```

```
summary(dados$Vento)
```

```
##      FALSO  VERDADEIRO    NA's
##         7           6         1
```

```
dados[is.na(dados$Vento),]$Vento = "FALSO"
#Aparencia - dados errados
unique(dados$Aparencia)
```

```
## [1] sol    nublado  chuva  menos
## Levels: chuva menos nublado sol
```

```
summary(dados$Aparencia)
```

```
##      chuva  menos  nublado    sol
##         5         1         3         5
```

```
dados[dados$Aparencia == "menos",]$Aparencia = "sol"
dados$Aparencia = factor(dados$Aparencia)
```

Conferir tabela final

```
dados
```

```
##      Aparencia  Temperatura  Umidade      Vento  Jogar
## 1      sol          85.0       85      FALSO   nao
## 2      sol          80.0       90  VERDADEIRO  nao
## 3  nublado          83.0       86      FALSO   sim
## 4      chuva          70.0       85      FALSO   sim
## 5      chuva          68.0       80      FALSO   sim
## 6      chuva          65.0       70  VERDADEIRO  nao
## 7      sol          64.0       65  VERDADEIRO   sim
## 8      sol          73.5       95      FALSO   nao
## 9      sol          60.0       70      FALSO   sim
## 10     chuva          75.0       85      FALSO   sim
## 11     sol          75.0       70  VERDADEIRO   sim
## 12  nublado          72.0       90  VERDADEIRO   sim
## 13  nublado          81.0       75      FALSO   sim
## 14     chuva          71.0       91  VERDADEIRO  nao
```