PROVA GED-13 1º Bimestre

Rafael Camargo

2022-09-07

Bibliotecas:

```
library(tidyverse)
library(ggplot2)
library(quantmod)
library(data.table)
library(ggpubr)
library(cowplot)
library(ichimoku)
```

Questão 1

Fazendo um a questão para um grupo de dados genérico, a média e a mediana já possuem funções prontas no R, para a moda:

```
amoda <- function(data) {
  res <- which(tabulate(data) == max(tabulate(data)))
  if (length(res) > 1) {
    print('Amostra com várias modas:')
  }
  return(res)
}
```

Variancia, desvio padrão e o grafico de linha já possuem funções no R, para o cálculo do retorno temos:

```
retorno <- function(data) {
  res <- c(0)

  for (dia in 2:length(data[, 2])) {
    retorno <- (data[dia, 2]-data[dia-1, 2])/data[dia-1, 2]
    res <- append(res, retorno)
  }

  return(res)
}
```

Para o gráfico de linha do retorno foi novamente utilizada o ggplot do R, assim como o violin plot, histograma, qqplot e a qqline. Para determinar a assimetria amostral, usamos a expressão:

$$s_3 = \frac{n}{(n-1)(n-2)} \times \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{s} \right)^3$$

Sendo s o desvio padrão anteriormente calculado anteriormente.

```r
assimetria <- function(data) {
  n <- length(data)
  theMean <- mean(data)
  dataSd <- sd(data)
  s <- 0
  for (i in 1:n) {
    s <- s + n/((n-1)*(n-2))*((data[i]-theMean)/dataSd)^3
  }
  return(s)
}
```

Para determinar a curtose amostral usamos a expressão:

$$s_4 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \times \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

```r
curtose <- function(data) {
  n <- length(data)
  theMean <- mean(data)
  dataSd <- sd(data)
  s <- (-3)*(n-1)^2/((n-2)*(n-3))
  for (i in 1:n) {
    s <- s + n*(n+1)/((n-1)*(n-2)*(n-3))*((data[i]-theMean)/dataSd)^4
  }
  return(s)
}
```

Assim, a função principal fica da seguinte forma para todos às análises:

```r
main <- function(data) {
  data <- xts_df(data)
  colnames(data) <- c('index', 'fecha')

  # Média
  print(sprintf("Média: %f", mean(data[,2])))

  # Moda
  print(sprintf("Moda: %f", amoda(data[,2])))

  # Mediana
  print(sprintf("Mediana: %f", median(data[,2])))

  # Variância
  print(sprintf("Variância: %f", var(data[,2])))

  # Desvio padrão
  dataSd <- sd(data[,2])
  print(sprintf("Desvio padrão: %f", dataSd))

  # Gráfico de linha do preço de fechamento
  print(ggplot(data, aes(x = get('index'), y = get('fecha'))) + geom_line() +
          labs(y = 'Fechamento', x = 'Dias', title = 'Gráfico de linha do preço de fechamento'))

  # Retorno
```

```
  ret <- retorno(data)
  retdata <- data
  retdata$new <- ret
  # OBS: Após uma longa análise, cheguei à conclusão que é desnecessario printar
  #o vetor de retornos calculados.

  # Gráfico de linha do Retorno
  print(ggplot(retdata, aes(x = 1:nrow(retdata), y = get('new'))) + geom_line() +
          labs(y = 'Retorno', x = 'Dias', title = 'Gráfico de linha do retorno'))

  # Violin Plot para os dados de Preço de Fechamento e do Retorno.
  viodata <- data
  viodata$labedl <- rep(c('Fechamento'), times=length(data[, 2]))
  colnames(viodata) <- c('index', 'fecha', 'label')
  print(ggplot(viodata, aes(x = get('label'), y = get('fecha'))) + geom_violin(trim=FALSE) +
          labs(y = NULL, x = NULL, title='Violin Plot do preço de fechamento'))

  retdata$label <- rep(c('Retorno'), times=length(data[, 2]))
  print(ggplot(retdata, aes(x = get('label'), y = get('new'))) + geom_violin(trim=FALSE) +
          labs(y = NULL, x = NULL, title='Violin Plot do retorno'))

  # Histograma para os dados de Preço de Fechamento e do Retorno.
  print(ggplot(viodata, aes(x = get('fecha'))) + geom_histogram(color="blue", fill="white") +
          labs(y = NULL, x = NULL, title='Histograma do preço de fechamento'))

  print(ggplot(retdata, aes(x = get('new'))) + geom_histogram(color="red", fill="white") +
          labs(y = NULL, x = NULL, title='Histograma do retorno'))

  # QQPlot e QQLine do retorno
  print(qplot(sample = retdata[, 3]) + geom_qq_line() +
          labs(y = 'Retorno', title='QQPlot e QQline do retorno'))

  # Assimetria OBS: O valor será interpretado conforme solicitado no enunciado na chamada
  #da função para cada variável financeira.
  print(sprintf("Assimetria amostral: %f", assimetria(ret)))

  # Curtose amostral OBS: O valor será interpretado conforme solicitado no enunciado na
  #chamada da função para cada variável financeira.
  print(sprintf("Curtose amostral: %f", curtose(ret)))

}
```

S&P500

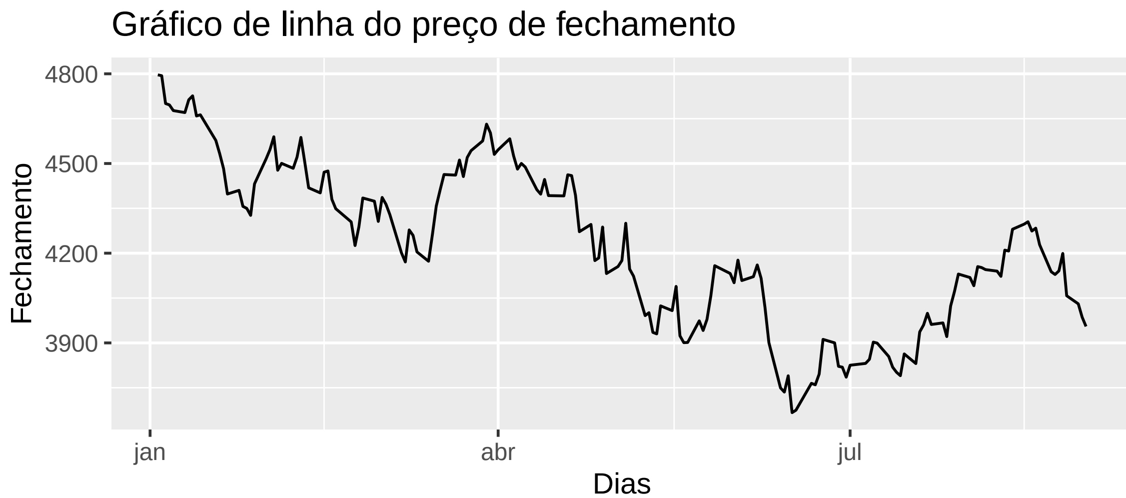
Carregamento das informações do Índice S&P500:

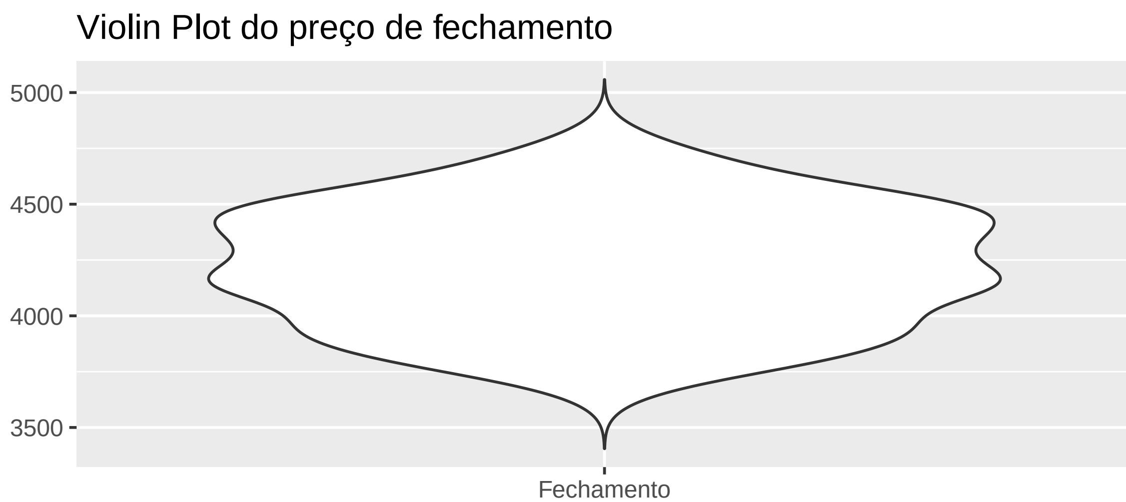
```
start <- as.Date("2022-01-01")
end <- as.Date("2022-09-01")
dados.sp <- quantmod::getSymbols("^GSPC", src = "yahoo", from = start, to = end,auto.assign = FALSE)
stdpoors <- na.omit(dados.sp)

stdpoors <- stdpoors$GSPC.Close
```

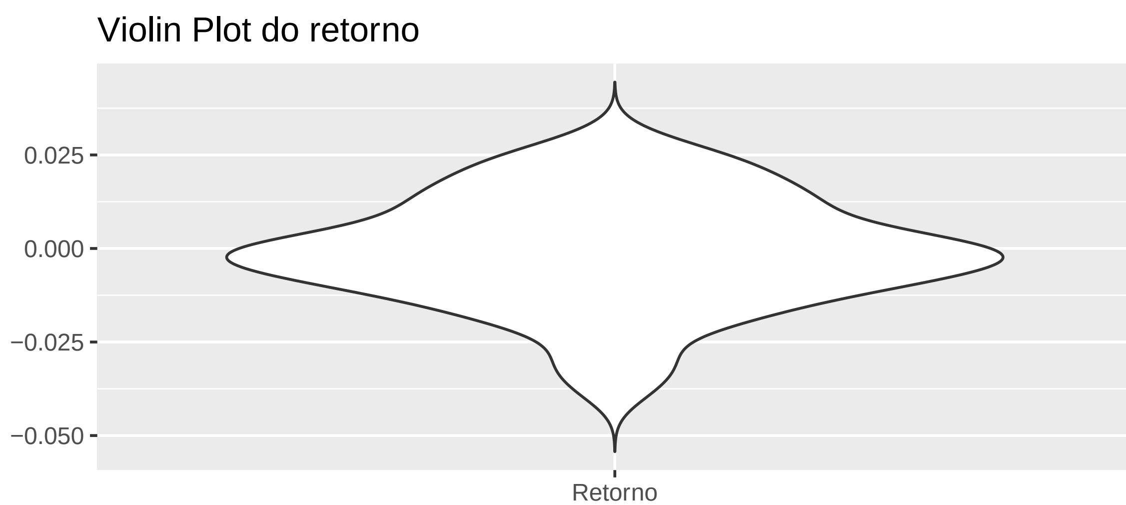
Estatísticas:

```
main(stdpoors)
```

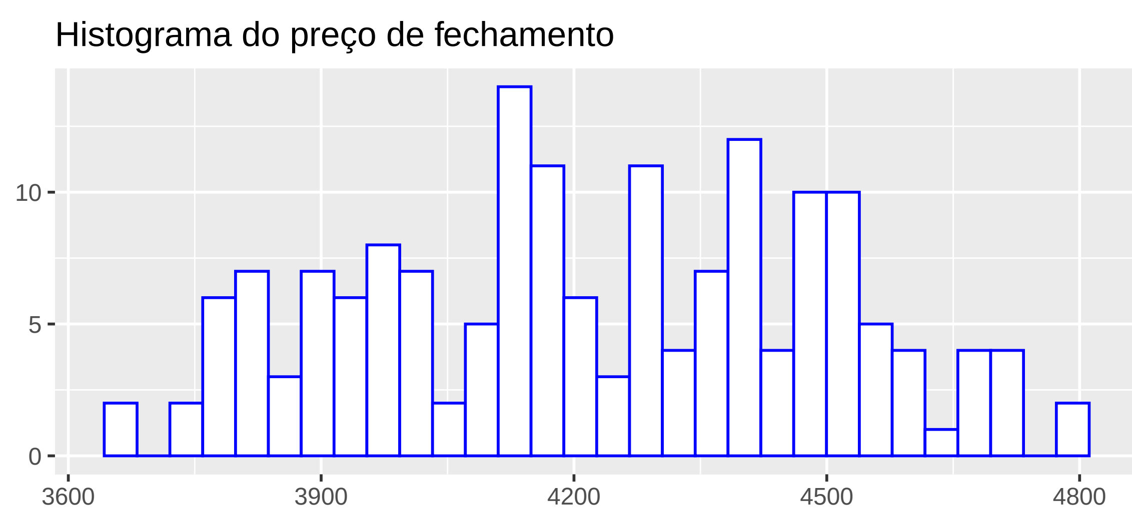
```
## [1] "Média: 4222.706232"
## [1] "Moda: 3900.000000"
## [1] "Mediana: 4207.270020"
## [1] "Variância: 72749.005423"
## [1] "Desvio padrão: 269.720235"
```



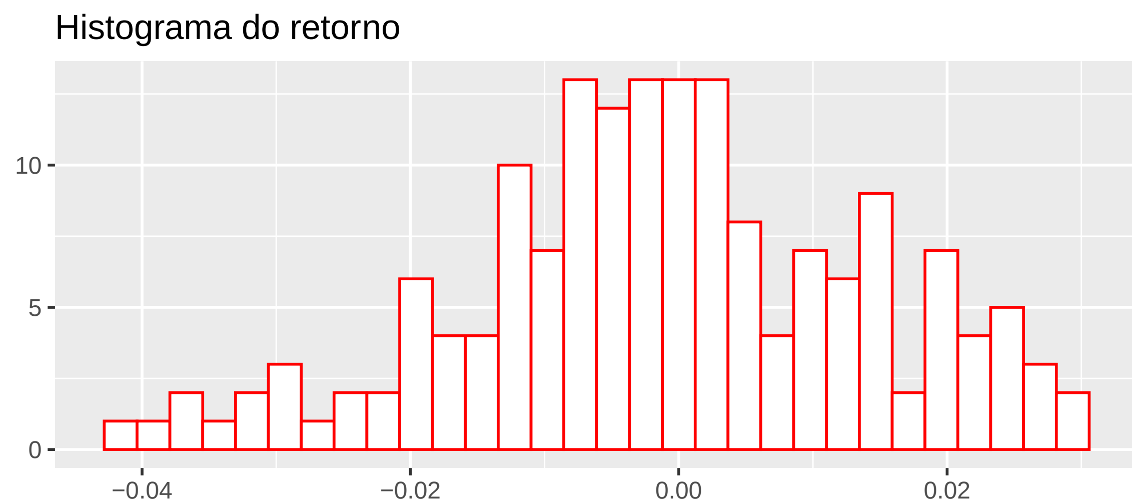


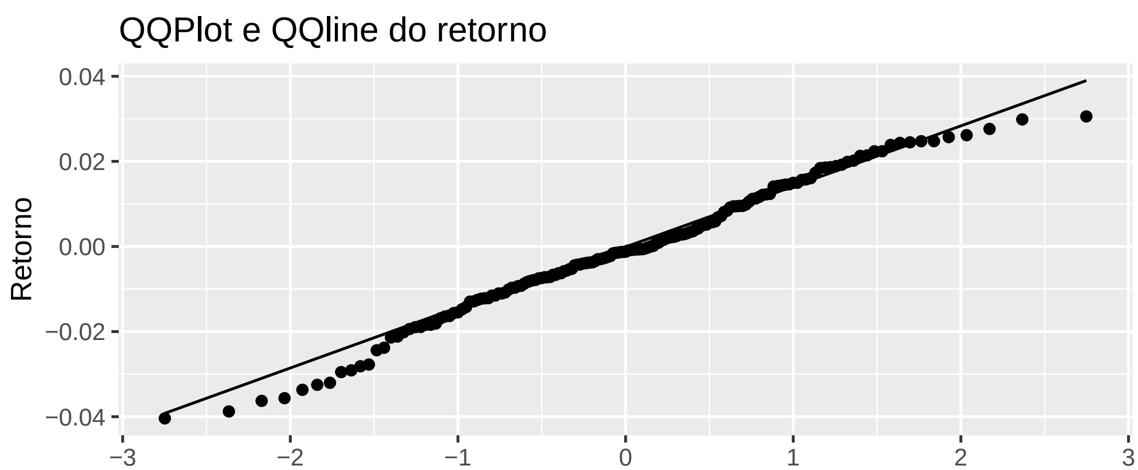



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```






```
## [1] "Assimetria amostral: -0.214683"
## [1] "Curtose amostral: -0.101986"
```

A assimetria amostra resultou -0,214683, ou seja, possui uma assimetria negativa, o que significa que os valores estão mais concentrados à direita. Além disso, como a curtose amostral é de -0,101986, a curtose é platicúrtica, ou seja, a curva é mais "achatada".

Dow Jones

Carregamento das informações do Índice Dow Jones:

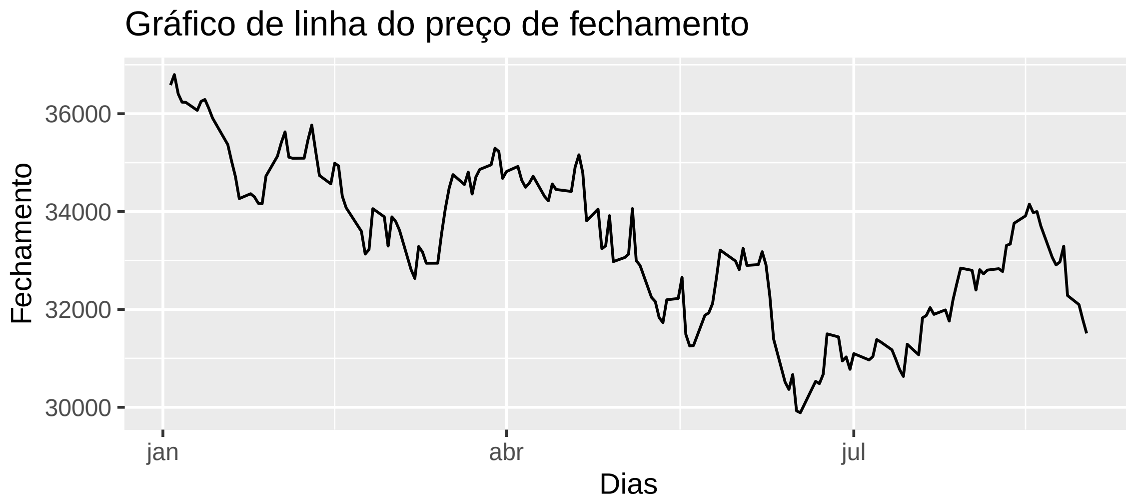
```
dados.dj <- quantmod::getSymbols("^DJI", src = "yahoo", from = start, to = end,auto.assign = FALSE)
dowjones <- na.omit(dados.dj)

dowjones <- dowjones$DJI.Close
```

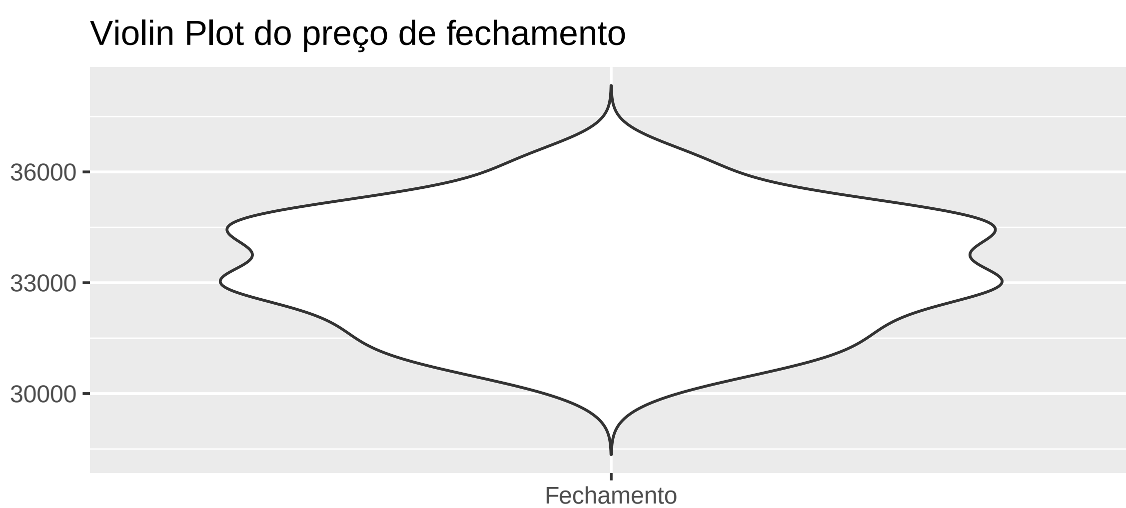
Estatísticas:

```
main(dowjones)
```

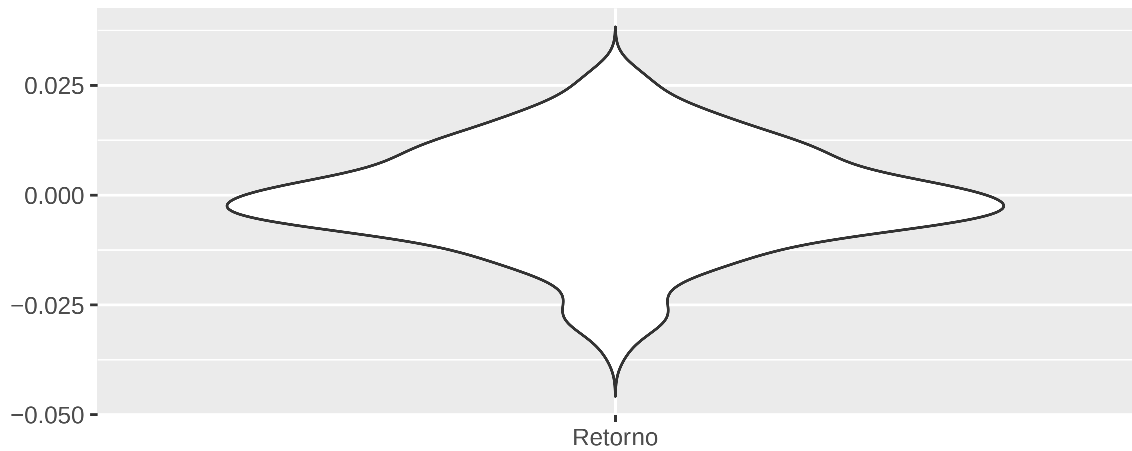
```
## [1] "Média: 33333.556898"
## [1] "Moda: 32899.000000"
## [1] "Mediana: 33248.281250"
```

```
## [1] "Variância: 2520219.322237"
## [1] "Desvio padrão: 1587.519865"
```



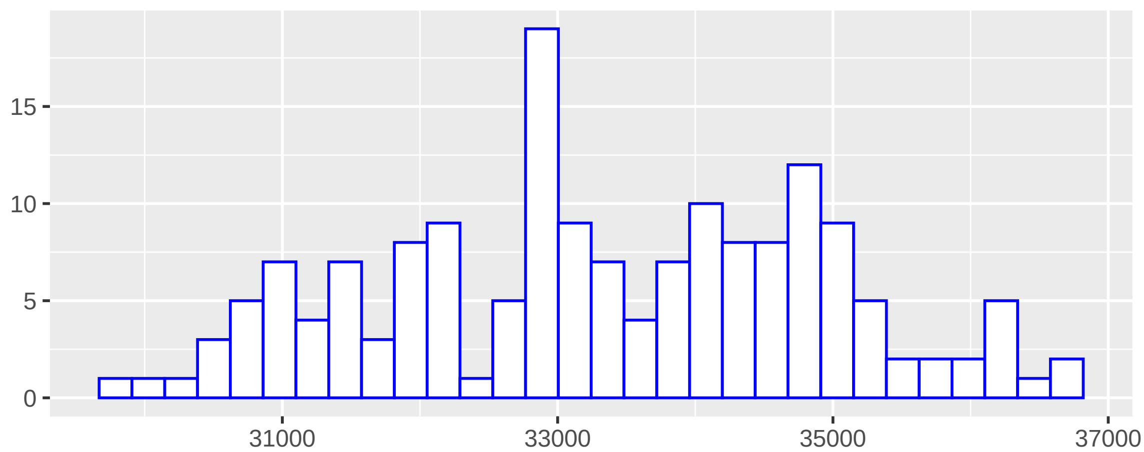


Violin Plot do retorno



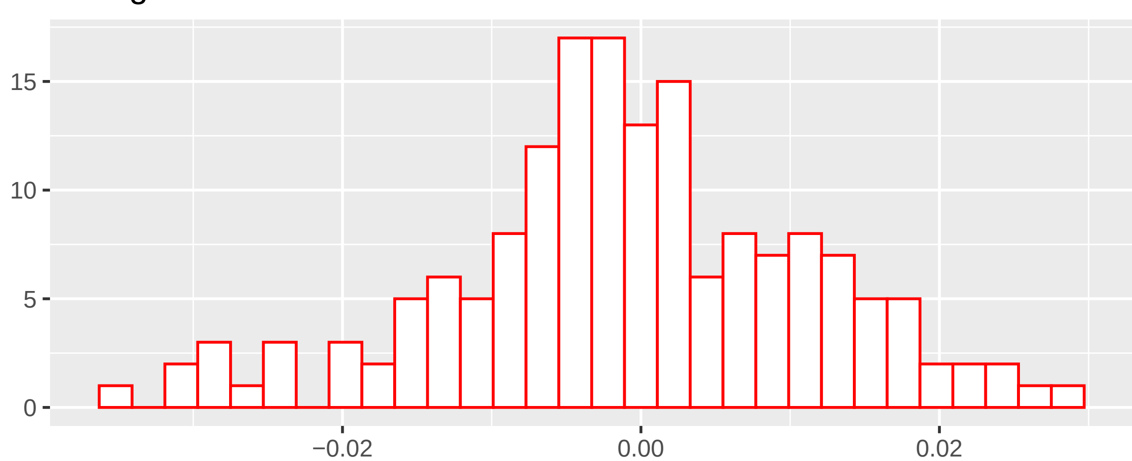

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

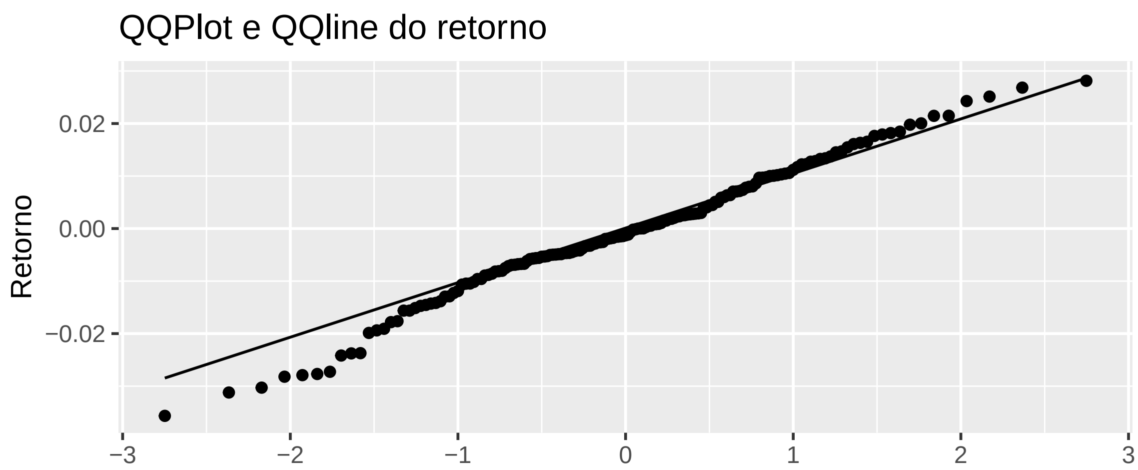
Histograma do preço de fechamento




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Histograma do retorno






```
## [1] "Assimetria amostral: -0.224311"
## [1] "Curtose amostral: 0.262944"
```

A assimetria amostra resultou -0,224311, ou seja, possui uma assimetria negativa, o que significa que os valores estão mais concentrados à direita. Além disso, como a curtose amostral é de 0,262944, a curtose é leptocúrtica, ou seja, a curva é mais "pontuda".

Nasdaq

Carregamento das informações do Índice Nasdaq:

```
dados.nasdaq <- quantmod::getSymbols("^IXIC", src = "yahoo", from = start, to =end, auto.assign = FALSE)
nasdaq <- na.omit(dados.nasdaq)

nasdaq <- nasdaq$IXIC.Close
```

Estatísticas:

```
main(nasdaq)
```

```
## [1] "Média: 12848.574488"
## [1] "Amostra com várias modas:"
## [1] "Moda: 11264.000000" "Moda: 12381.000000"
## [1] "Mediana: 12830.959961"
## [1] "Variância: 1424893.758856"
## [1] "Desvio padrão: 1193.689138"
```

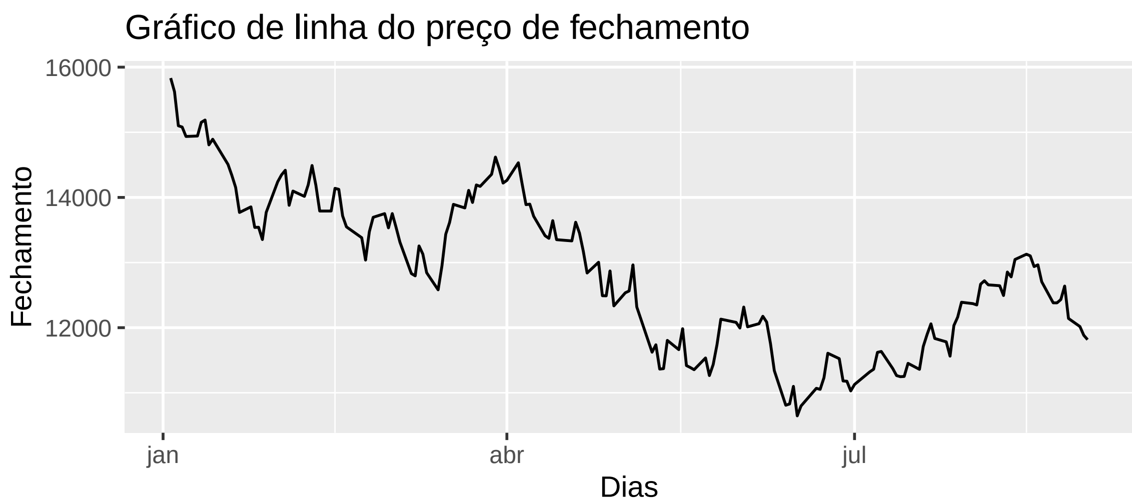
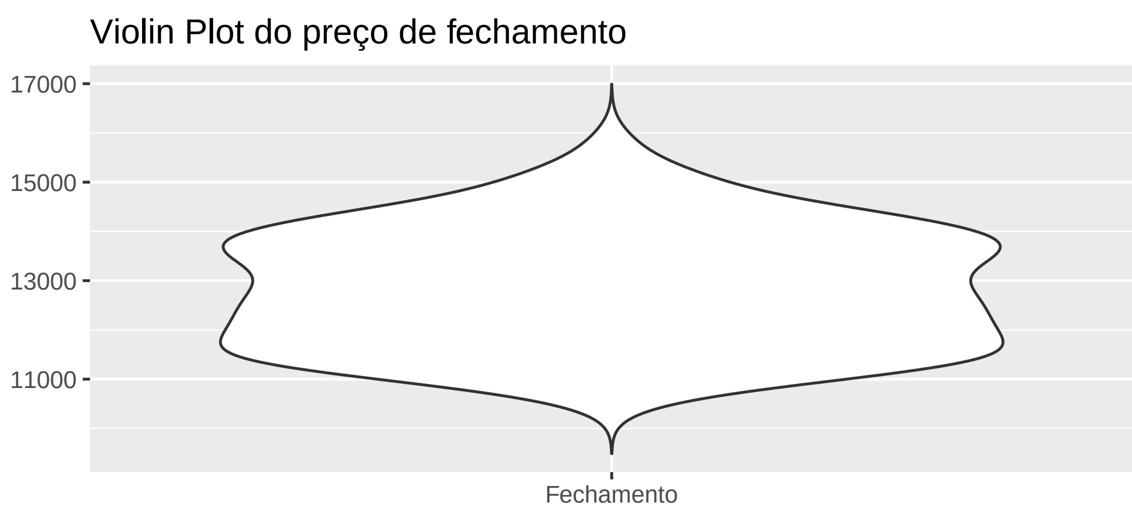
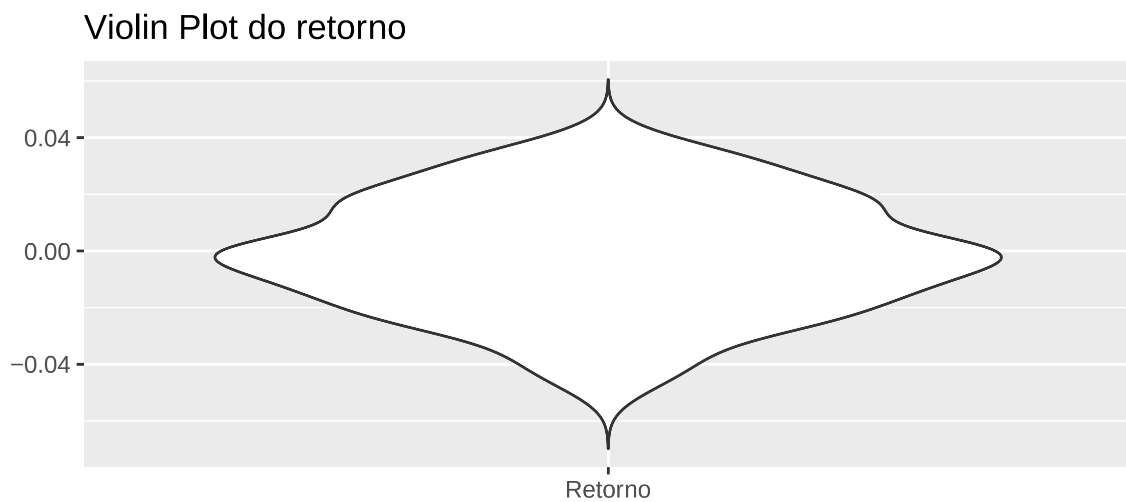


Gráfico de linha do preço de fechamento
16000
14000
12000
Fechamento
jan
abr
jul
Dias

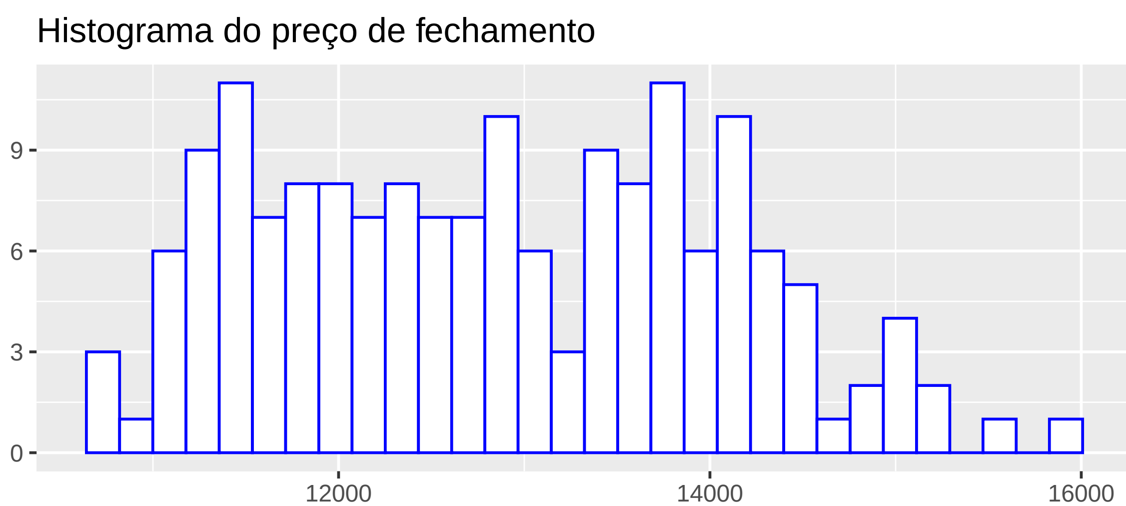
Gráfico de linha do retorno
0.025
0.000
−0.025
−0.050
Retorno
0
50
100
150
Dias



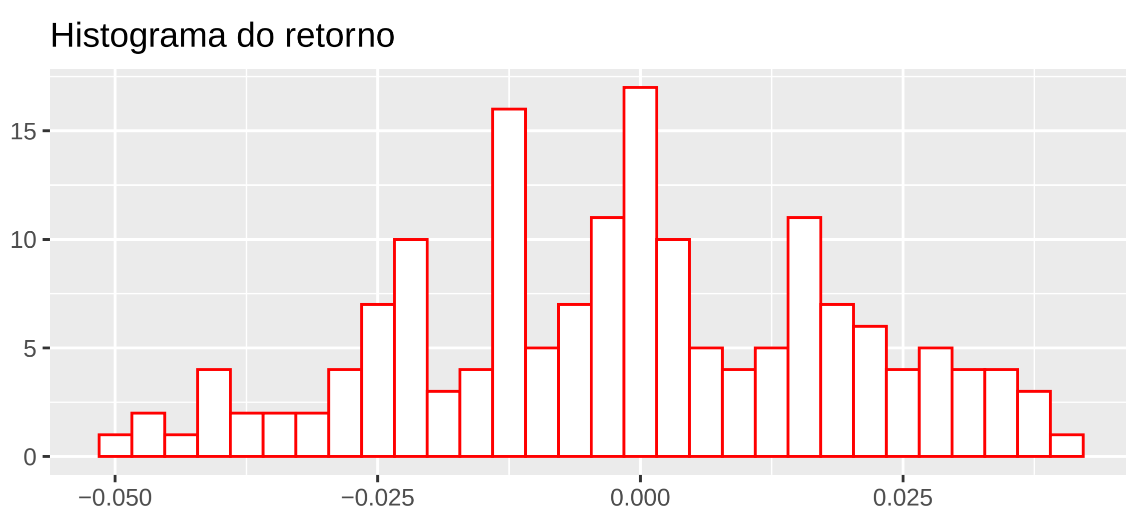
Violin Plot do preço de fechamento
17000
15000
13000
11000
Fechamento

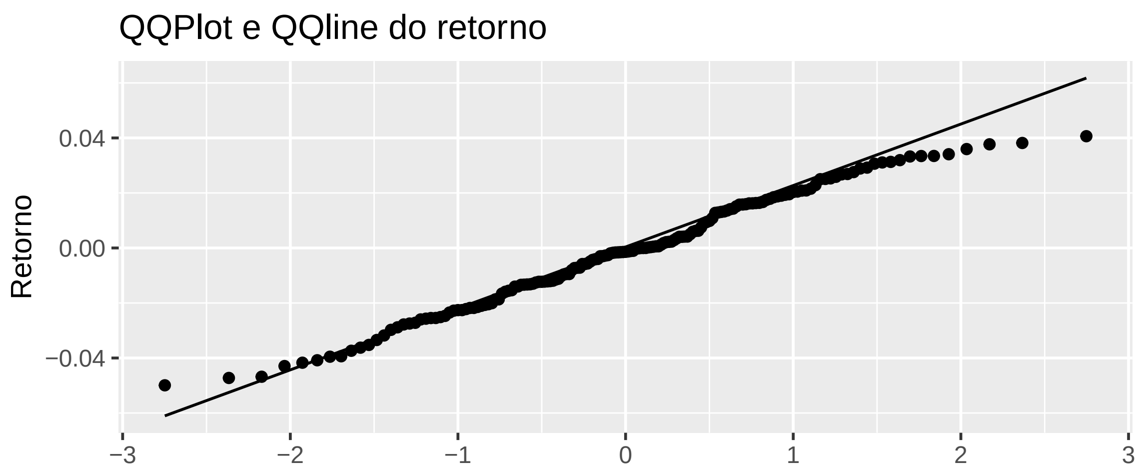



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```






```
## [1] "Assimetria amostral: -0.097247"
## [1] "Curtose amostral: -0.569675"
```

A assimetria amostra resultou -0,097247, ou seja, possui uma assimetria negativa, o que significa que os valores estão mais concentrados à direita. Além disso, como a curtose amostral é de -0,569675, a curtose é platicúrtica, ou seja, a curva é mais "achatada".

IBOVESPA

Carregamento das informações do Índice IBOVESPA:

```
dados.bovespa <- quantmod::getSymbols("^BVSP", src = "yahoo", from = start, to =end, auto.assign = FALSE
bovespa <- na.omit(dados.bovespa)

bovespa <- bovespa$BVSP.Close
```

Estatísticas:

```
main(bovespa)
```

```
## [1] "Média: 108733.863095"
## [1] "Amostra com várias modas:"
##   [1] "Moda: 96121.000000"  "Moda: 96551.000000"  "Moda: 96916.000000"
##   [4] "Moda: 97881.000000"  "Moda: 98080.000000"  "Moda: 98212.000000"
##   [7] "Moda: 98245.000000"  "Moda: 98271.000000"  "Moda: 98287.000000"
##  [10] "Moda: 98295.000000"  "Moda: 98542.000000"  "Moda: 98609.000000"
##  [13] "Moda: 98672.000000"  "Moda: 98719.000000"  "Moda: 98925.000000"
##  [16] "Moda: 98954.000000"  "Moda: 99033.000000"  "Moda: 99522.000000"
##  [19] "Moda: 99622.000000"  "Moda: 99685.000000"  "Moda: 99772.000000"
##  [22] "Moda: 99825.000000"  "Moda: 99853.000000"  "Moda: 100270.000000"
##  [25] "Moda: 100289.000000" "Moda: 100591.000000" "Moda: 100730.000000"
##  [28] "Moda: 100764.000000" "Moda: 101006.000000" "Moda: 101438.000000"
##  [31] "Moda: 101561.000000" "Moda: 101945.000000" "Moda: 102063.000000"
##  [34] "Moda: 102225.000000" "Moda: 102597.000000" "Moda: 102598.000000"
##  [37] "Moda: 102719.000000" "Moda: 102807.000000" "Moda: 103110.000000"
##  [40] "Moda: 103165.000000" "Moda: 103250.000000" "Moda: 103362.000000"
##  [43] "Moda: 103514.000000" "Moda: 103775.000000" "Moda: 103779.000000"
##  [46] "Moda: 103922.000000" "Moda: 104397.000000" "Moda: 105135.000000"
##  [49] "Moda: 105304.000000" "Moda: 105481.000000" "Moda: 105530.000000"
```

```
##  [52] "Moda: 105686.000000" "Moda: 105688.000000" "Moda: 105892.000000"
##  [55] "Moda: 106247.000000" "Moda: 106472.000000" "Moda: 106522.000000"
##  [58] "Moda: 106528.000000" "Moda: 106639.000000" "Moda: 106692.000000"
##  [61] "Moda: 106924.000000" "Moda: 106928.000000" "Moda: 107005.000000"
##  [64] "Moda: 107094.000000" "Moda: 107752.000000" "Moda: 107876.000000"
##  [67] "Moda: 108013.000000" "Moda: 108213.000000" "Moda: 108233.000000"
##  [70] "Moda: 108344.000000" "Moda: 108368.000000" "Moda: 108402.000000"
##  [73] "Moda: 108488.000000" "Moda: 108651.000000" "Moda: 108789.000000"
##  [76] "Moda: 108942.000000" "Moda: 108959.000000" "Moda: 109102.000000"
##  [79] "Moda: 109349.000000" "Moda: 109523.000000" "Moda: 109718.000000"
##  [82] "Moda: 109845.000000" "Moda: 109919.000000" "Moda: 109928.000000"
##  [85] "Moda: 110070.000000" "Moda: 110186.000000" "Moda: 110236.000000"
##  [88] "Moda: 110346.000000" "Moda: 110431.000000" "Moda: 110501.000000"
##  [91] "Moda: 110580.000000" "Moda: 110581.000000" "Moda: 110685.000000"
##  [94] "Moda: 111032.000000" "Moda: 111078.000000" "Moda: 111102.000000"
##  [97] "Moda: 111112.000000" "Moda: 111203.000000" "Moda: 111351.000000"
## [100] "Moda: 111360.000000" "Moda: 111478.000000" "Moda: 111496.000000"
## [103] "Moda: 111573.000000" "Moda: 111592.000000" "Moda: 111593.000000"
## [106] "Moda: 111696.000000" "Moda: 111713.000000" "Moda: 111725.000000"
## [109] "Moda: 111890.000000" "Moda: 111942.000000" "Moda: 111996.000000"
## [112] "Moda: 112008.000000" "Moda: 112161.000000" "Moda: 112234.000000"
## [115] "Moda: 112245.000000" "Moda: 112299.000000" "Moda: 112315.000000"
## [118] "Moda: 112323.000000" "Moda: 112388.000000" "Moda: 112393.000000"
## [121] "Moda: 112461.000000" "Moda: 112764.000000" "Moda: 112768.000000"
## [124] "Moda: 112857.000000" "Moda: 112892.000000" "Moda: 112898.000000"
## [127] "Moda: 113032.000000" "Moda: 113076.000000" "Moda: 113142.000000"
## [130] "Moda: 113147.000000" "Moda: 113359.000000" "Moda: 113512.000000"
## [133] "Moda: 113528.000000" "Moda: 113532.000000" "Moda: 113572.000000"
## [136] "Moda: 113663.000000" "Moda: 113708.000000" "Moda: 113807.000000"
## [139] "Moda: 113813.000000" "Moda: 113900.000000" "Moda: 114344.000000"
## [142] "Moda: 114474.000000" "Moda: 114660.000000" "Moda: 115057.000000"
## [145] "Moda: 115166.000000" "Moda: 115174.000000" "Moda: 115181.000000"
## [148] "Moda: 115311.000000" "Moda: 115687.000000" "Moda: 116147.000000"
## [151] "Moda: 116155.000000" "Moda: 116182.000000" "Moda: 116782.000000"
## [154] "Moda: 116953.000000" "Moda: 117272.000000" "Moda: 117457.000000"
## [157] "Moda: 118228.000000" "Moda: 118322.000000" "Moda: 118738.000000"
## [160] "Moda: 118862.000000" "Moda: 118885.000000" "Moda: 119053.000000"
## [163] "Moda: 119081.000000" "Moda: 119999.000000" "Moda: 120014.000000"
## [166] "Moda: 120260.000000" "Moda: 121280.000000" "Moda: 121570.000000"
## [1] "Mediana: 109999.000000"
## [1] "Variância: 38669231.005097"
## [1] "Desvio padrão: 6218.458893"
```

Gráfico de linha do preço de fechamento

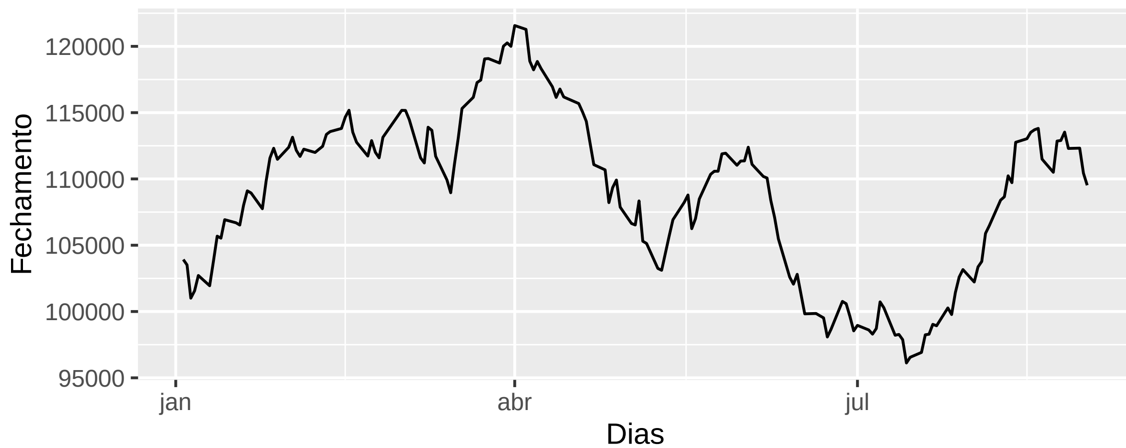
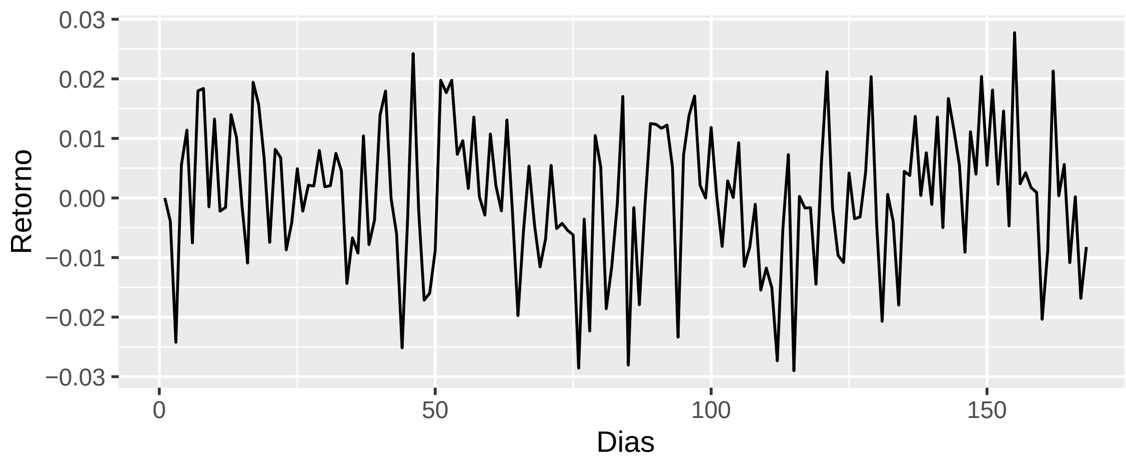
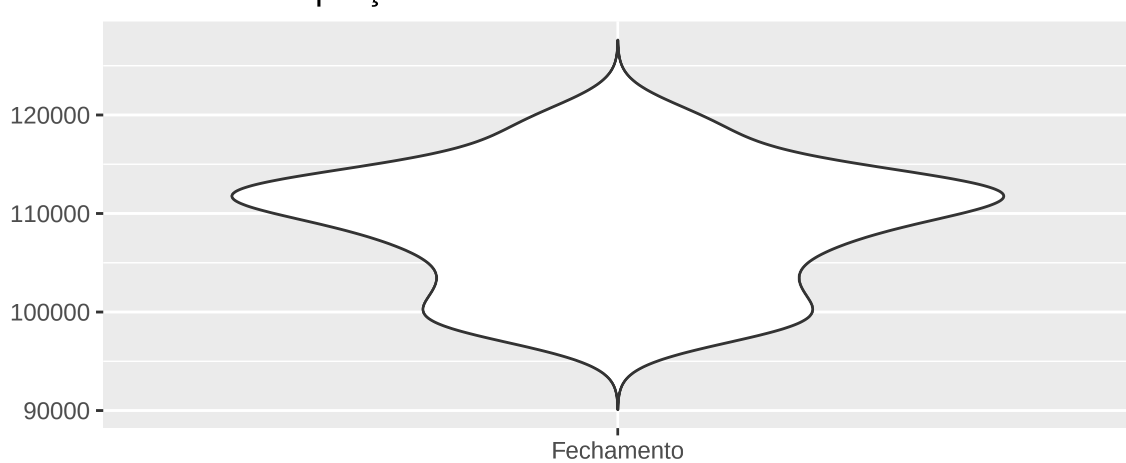
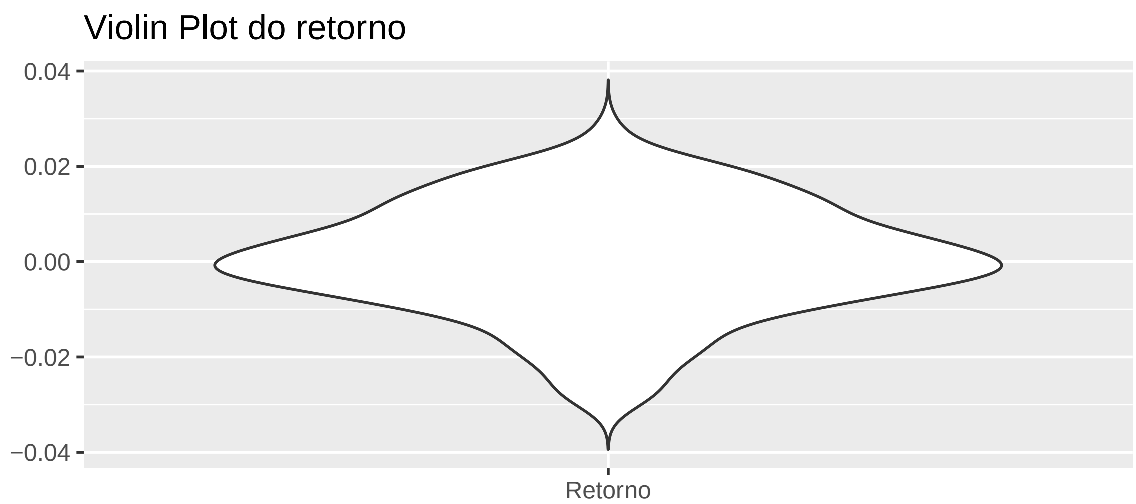


Gráfico de linha do retorno

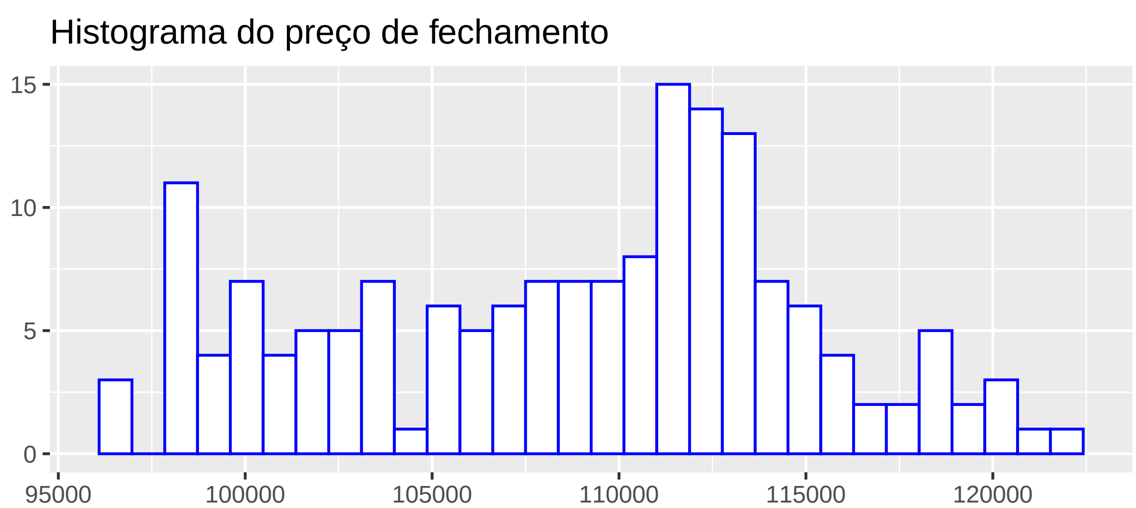


Violin Plot do preço de fechamento

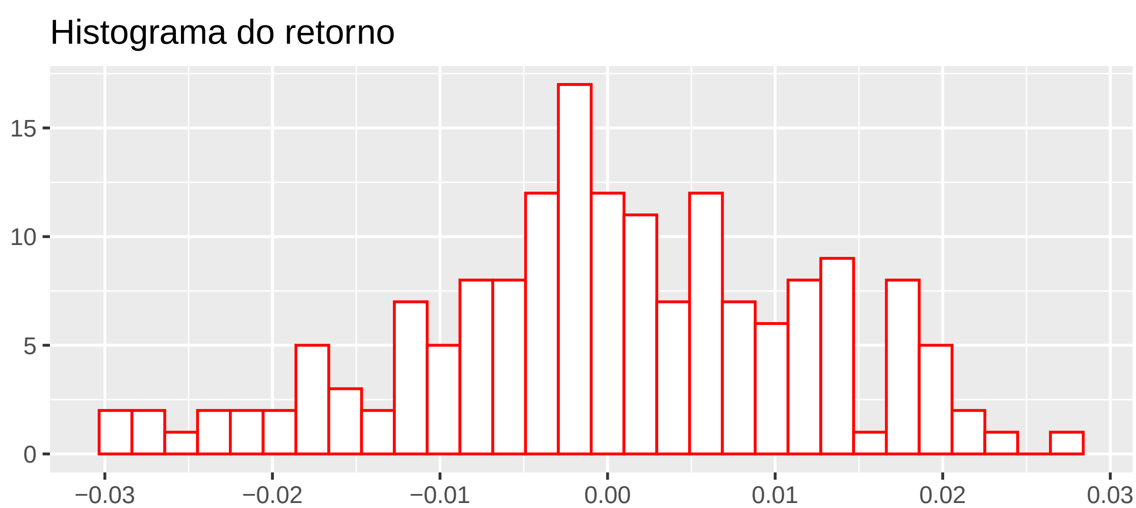


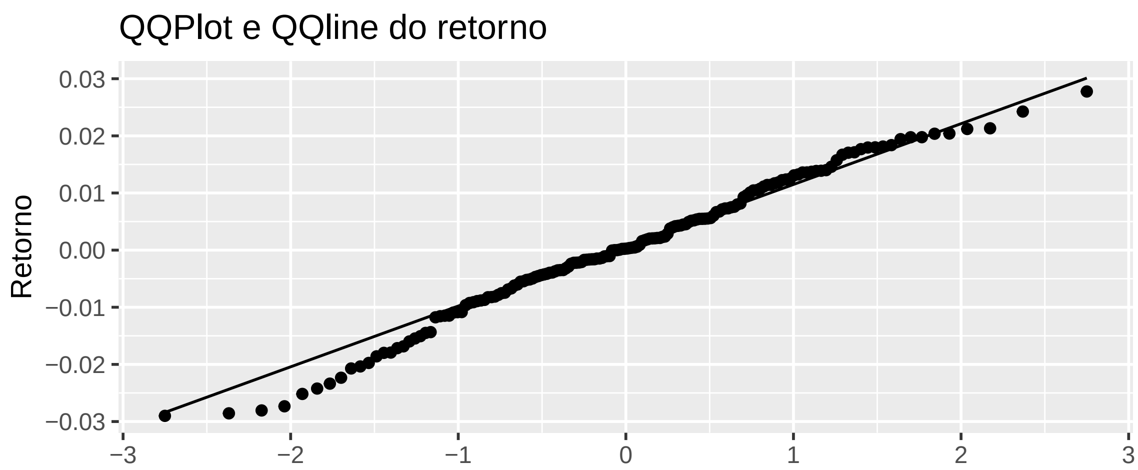



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```






```
## [1] "Assimetria amostral: -0.260256"
## [1] "Curtose amostral: -0.143831"
```

A assimetria amostra resultou -0,260256, ou seja, possui uma assimetria negativa, o que significa que os valores estão mais concentrados à direita. Além disso, como a curtose amostral é de -0,143831, a curtose é platicúrtica, ou seja, a curva é mais "achatada".

Preço do Petróleo Brent

Carregamento das informações do Preço do Petróleo Brent:

```
dados.brent <- quantmod::getSymbols("BZ=F", src = "yahoo", from = start, to = end,auto.assign = FALSE)
```

```
## Warning: BZ=F contains missing values. Some functions will not work if objects
## contain missing values in the middle of the series. Consider using na.omit(),
## na.approx(), na.fill(), etc to remove or replace them.
```

```
brent <- na.omit(dados.brent)

brent <- brent$`BZ=F.Close`
```

Estatísticas:

```
main(brent)
```

```
## [1] "Média: 104.011667"
## [1] "Moda: 107.000000"
## [1] "Mediana: 105.119999"
## [1] "Variância: 117.247090"
## [1] "Desvio padrão: 10.828070"
```

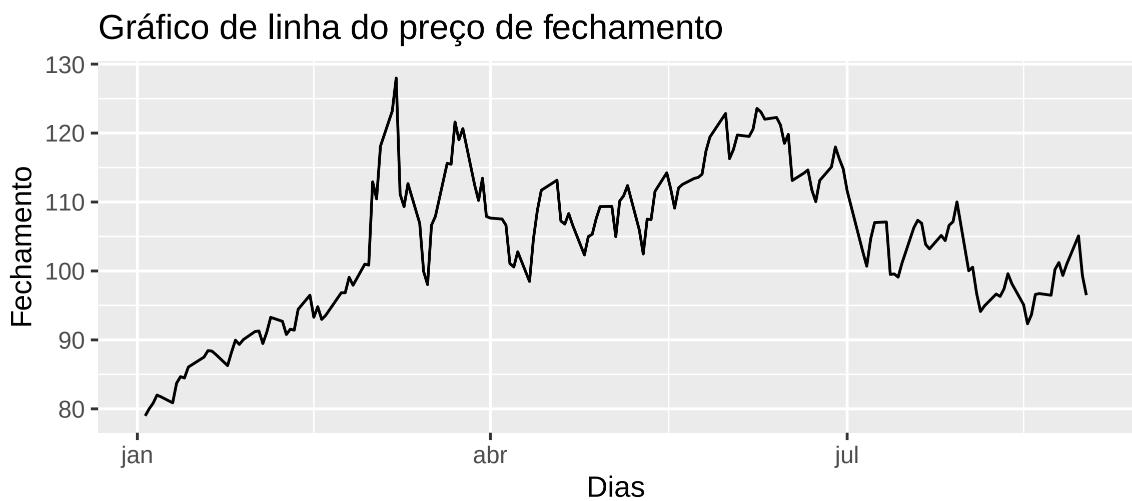
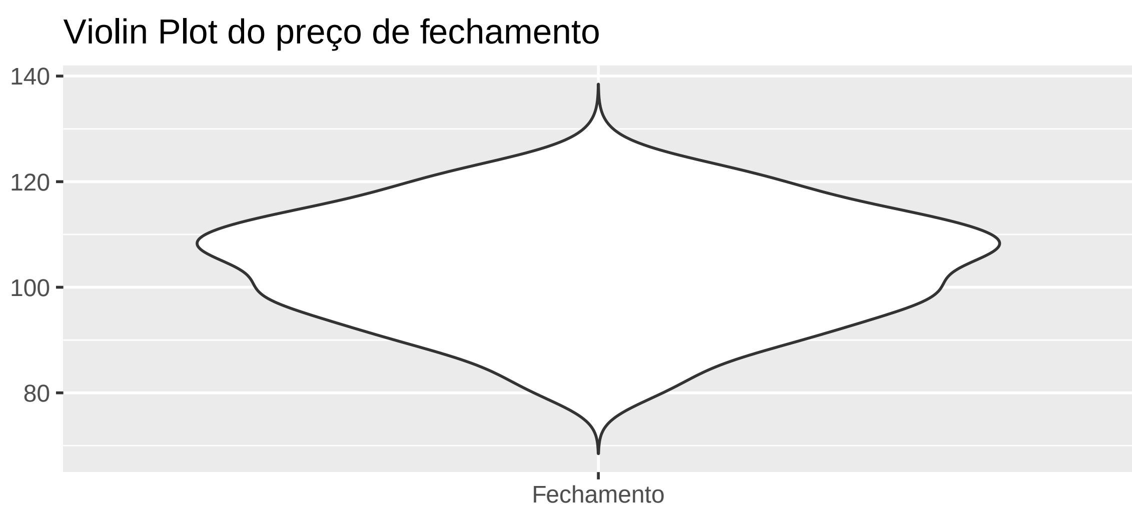
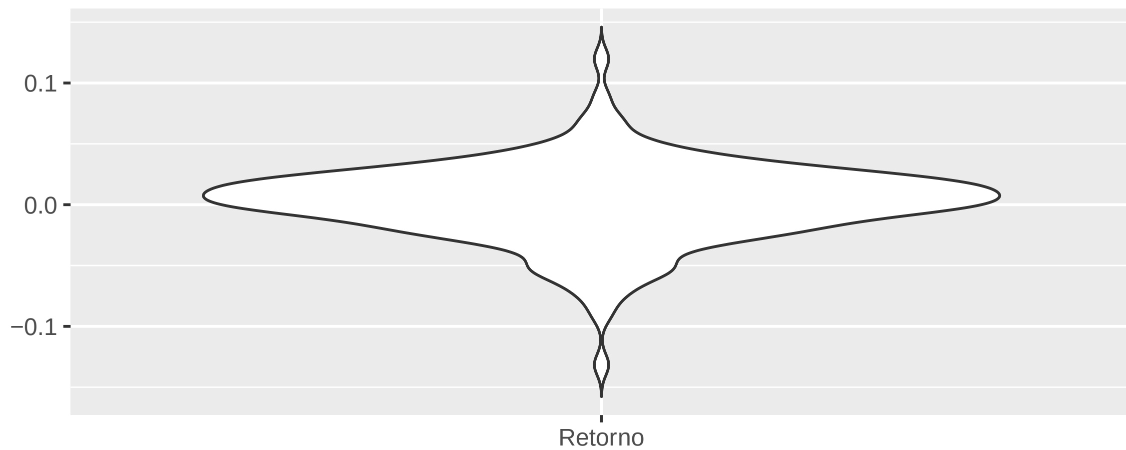


Gráfico de linha do preço de fechamento
Fechamento
130
120
110
100
90
80
jan
abr
jul
Dias


Gráfico de linha do retorno
Retorno
0.10
0.05
0.00
−0.05
−0.10
0
50
100
150
Dias

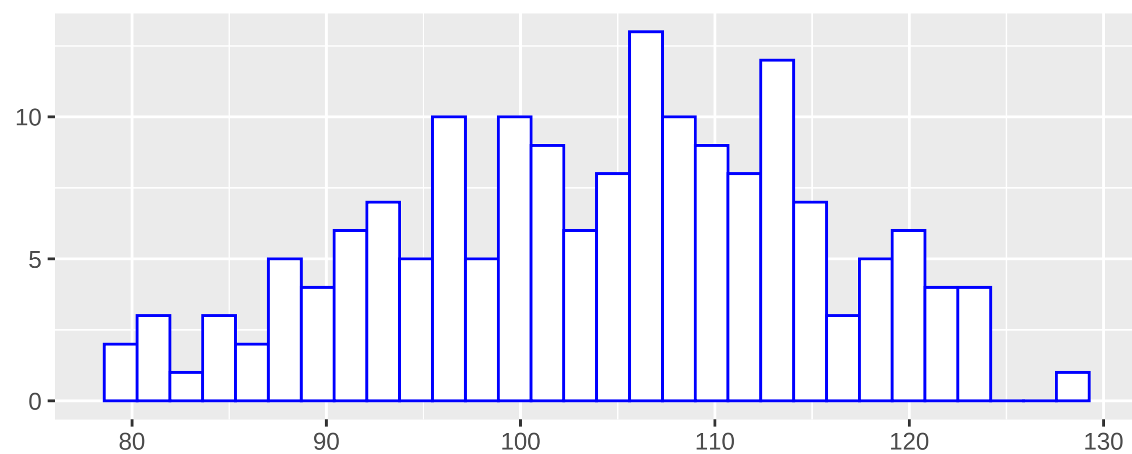

Violin Plot do preço de fechamento
140
120
100
80
Fechamento

Violin Plot do retorno



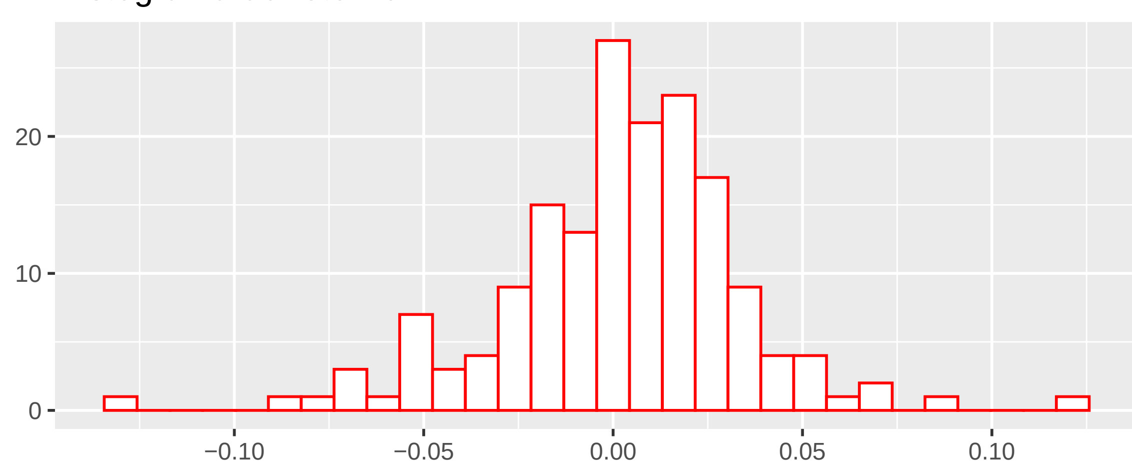

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

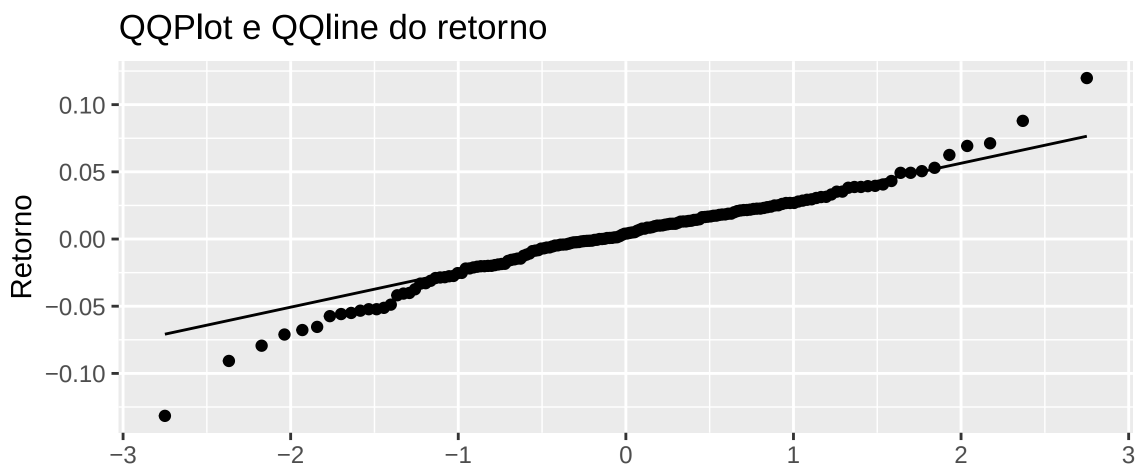
Histograma do preço de fechamento




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Histograma do retorno






```
## [1] "Assimetria amostral: -0.411252"
## [1] "Curtose amostral: 2.475360"
```

A assimetria amostra resultou -0,411252, ou seja, possui uma assimetria negativa, o que significa que os valores estão mais concentrados à direita. Além disso, como a curtose amostral é de 2,475360, a curtose é leptocúrtica, ou seja, a curva é mais "pontuda".

Cotação Dólar/Real

Carregamento das informações da Cotação Dólar/Real:

```
dados.dolar.real <- quantmod::getSymbols("BRL=X", src = "yahoo", from = start, to= end, auto.assign = F
dolar.real <- na.omit(dados.dolar.real)

dolar.real <- dolar.real$`BRL=X.Close`
```

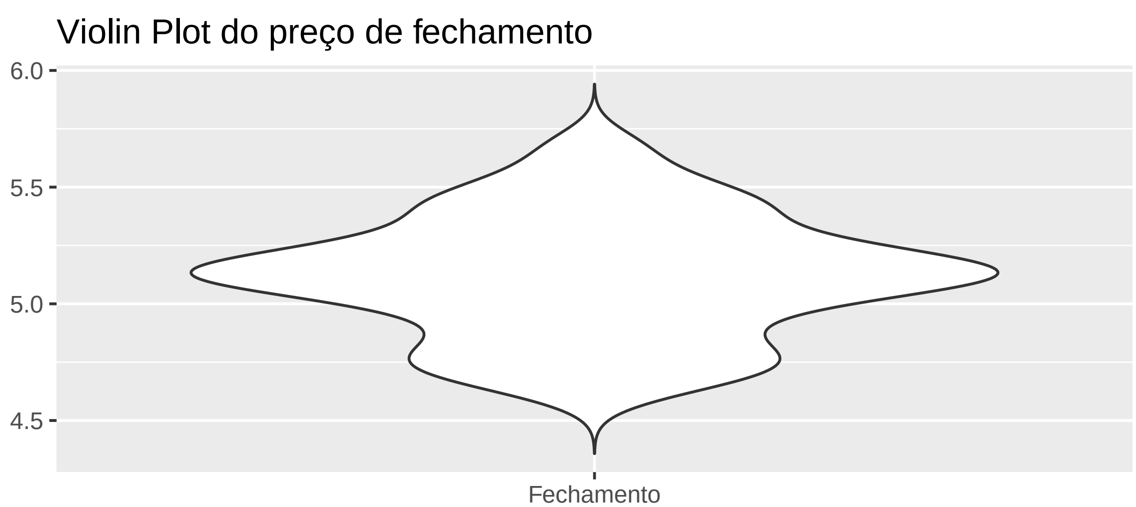
Estatísticas:

```
main(dolar.real)
```

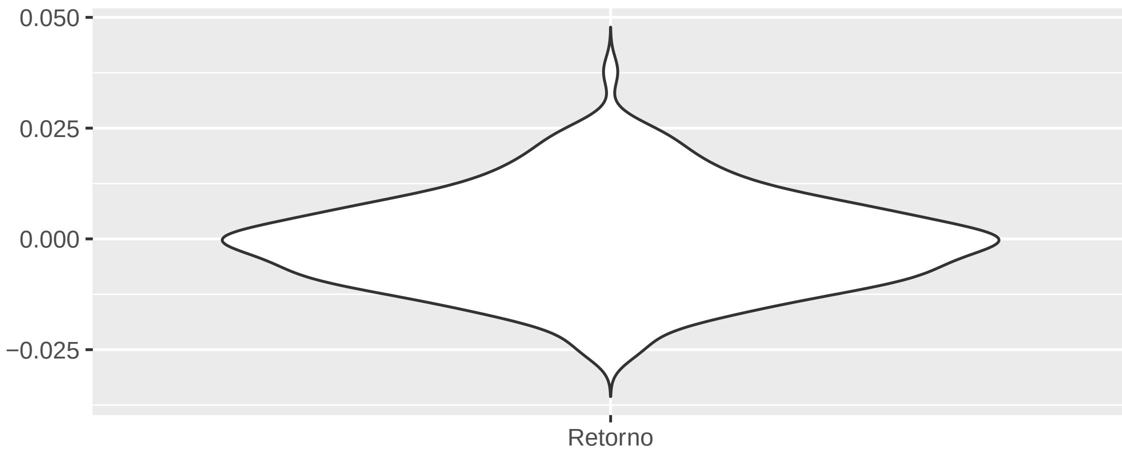
```
## [1] "Média: 5.117594"
## [1] "Moda: 5.000000"
## [1] "Mediana: 5.123250"
## [1] "Variância: 0.067955"
## [1] "Desvio padrão: 0.260682"
```


Gráfico de linha do preço de fechamento
5.75
5.50
5.25
5.00
4.75
Fechamento
jan
abr
jul
Dias


Gráfico de linha do retorno
0.04
0.02
0.00
−0.02
Retorno
0
50
100
150
Dias

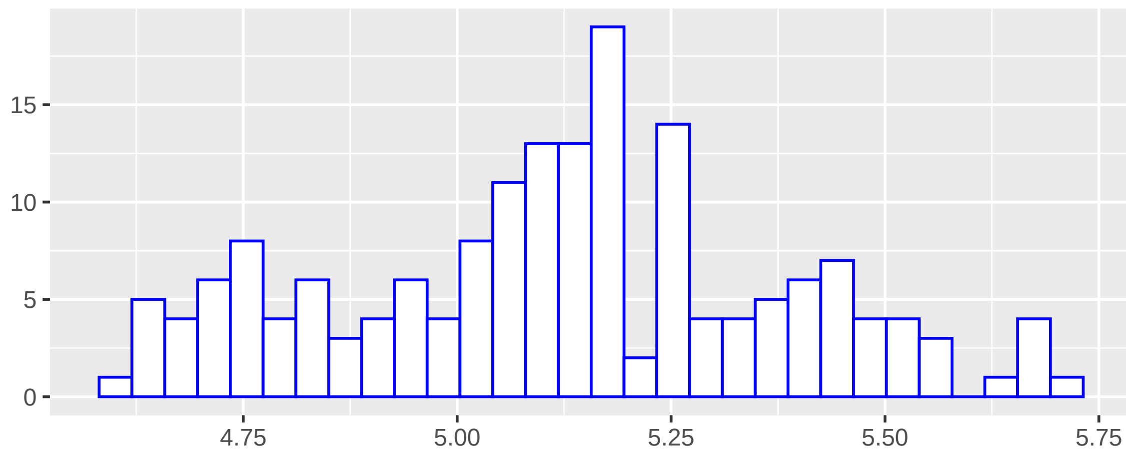

Violin Plot do preço de fechamento
6.0
5.5
5.0
4.5
Fechamento

Violin Plot do retorno



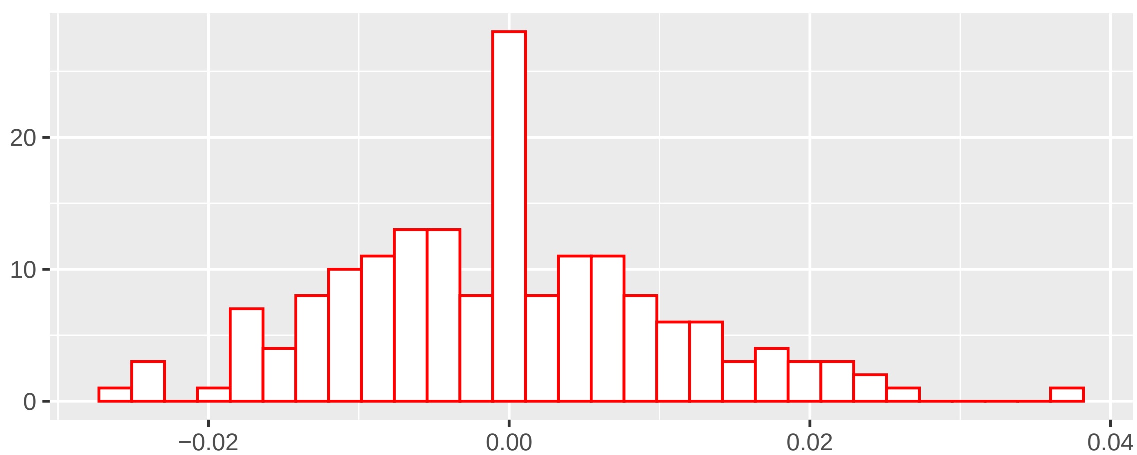

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

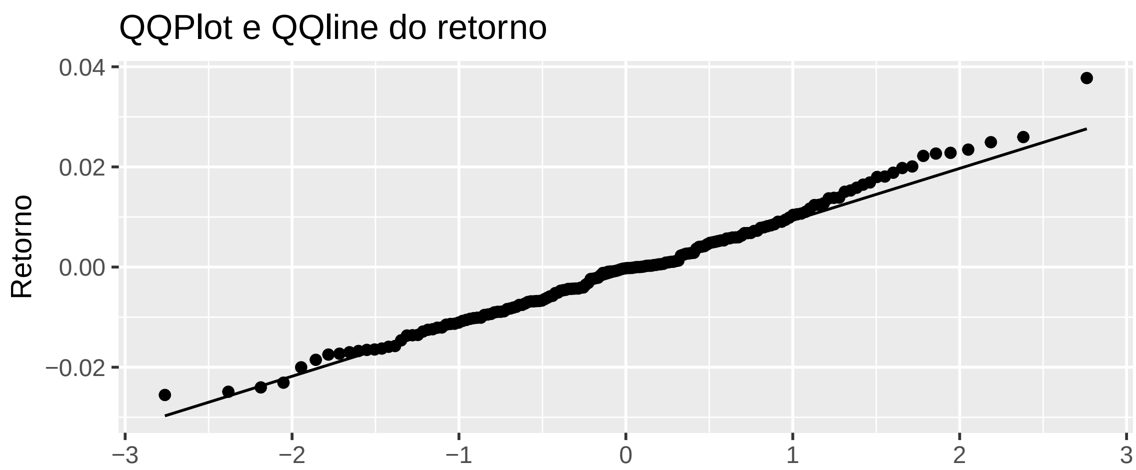
Histograma do preço de fechamento




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Histograma do retorno






```
## [1] "Assimetria amostral: 0.358391"
## [1] "Curtose amostral: 0.314937"
```

A assimetria amostra resultou 0,358391, ou seja, possui uma assimetria positiva, o que significa que os valores estão mais concentrados à esquerda. Além disso, como a curtose amostral é de 0,314937, a curtose é leptocúrtica, ou seja, a curva é mais "pontuda".

Bitcoin

Carregamento das informações do preço do Bitcoin:

```
dados.btc <- quantmod::getSymbols("BTC-USD", src = "yahoo", from = start, to =end, auto.assign = FALSE)
dolar.bitcoin <- na.omit(dados.btc)

dolar.bitcoin <- dolar.bitcoin$`BTC-USD.Close`
```

Estatísticas:

```
main(dolar.bitcoin)
```

```
## [1] "Média: 33014.489394"
## [1] "Moda: 23164.000000"
## [1] "Mediana: 36516.228515"
## [1] "Variância: 84190261.690737"
## [1] "Desvio padrão: 9175.525145"
```

Gráfico de linha do preço de fechamento

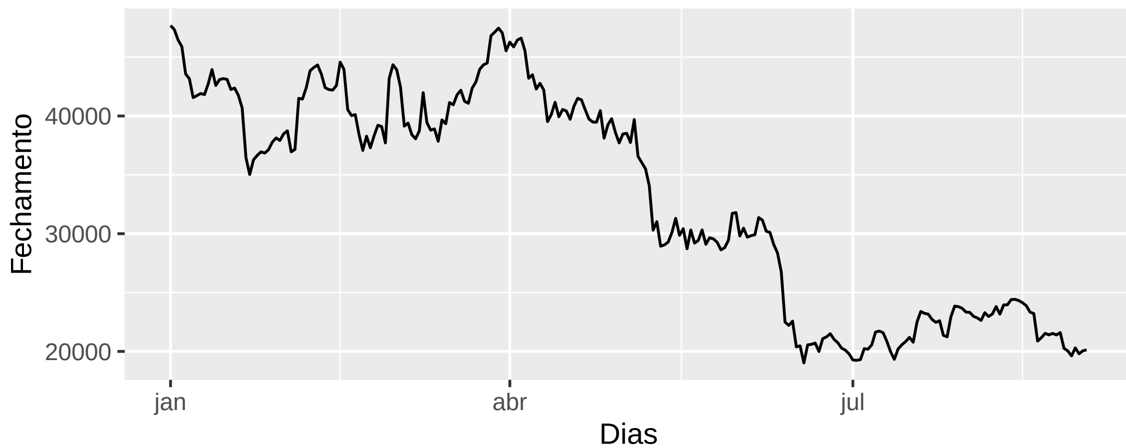
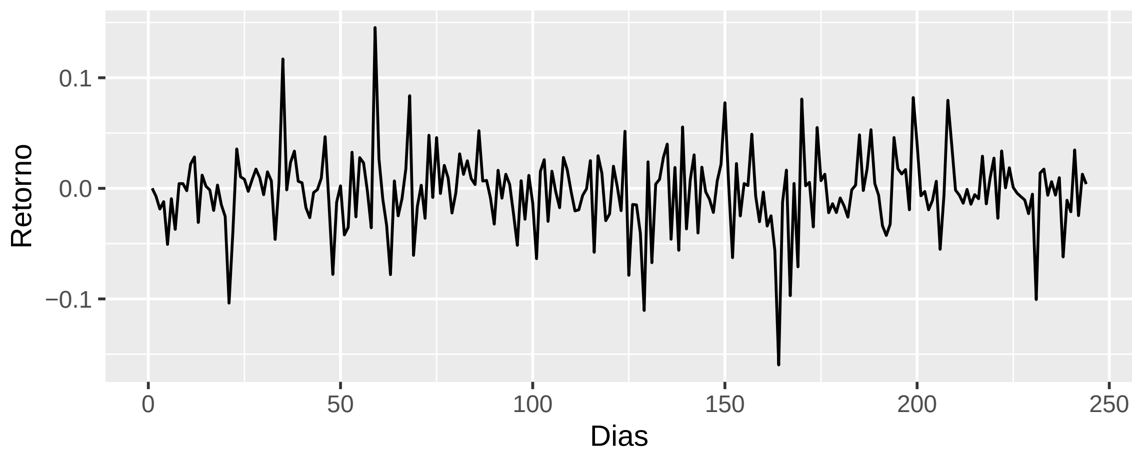
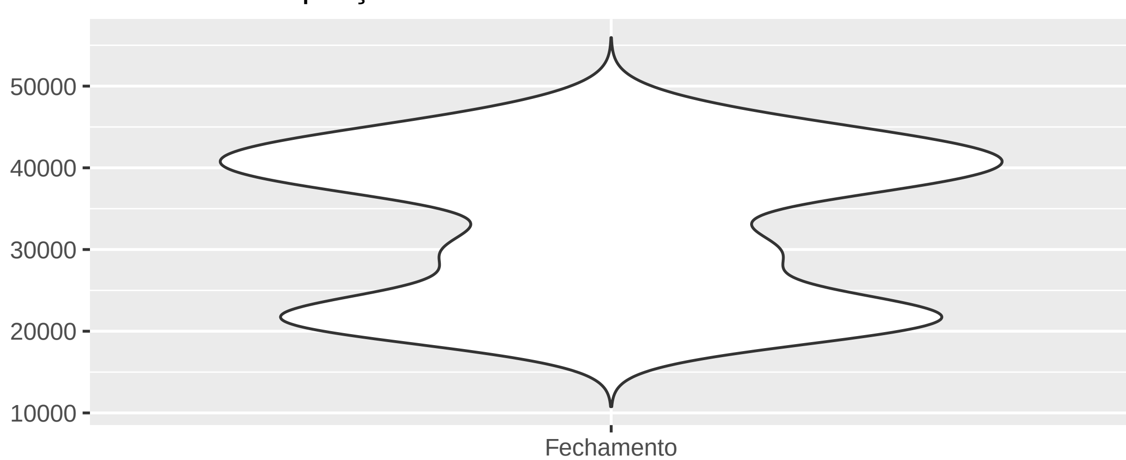
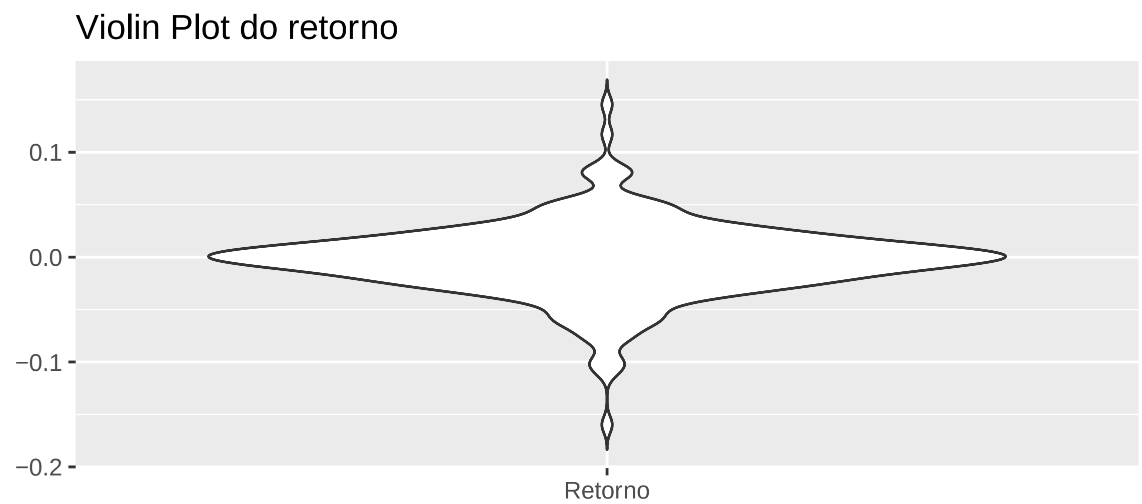


Gráfico de linha do retorno

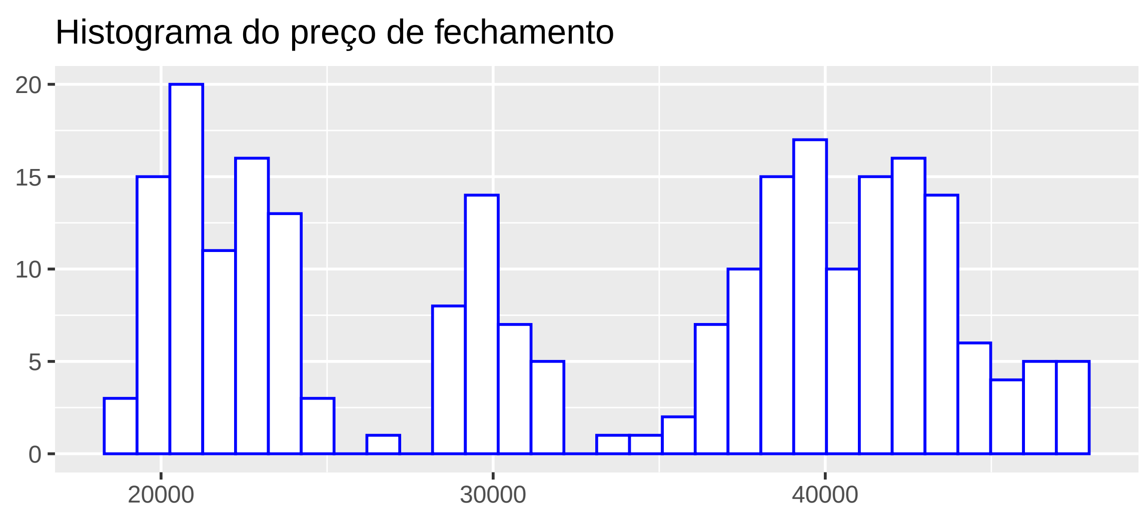


Violin Plot do preço de fechamento

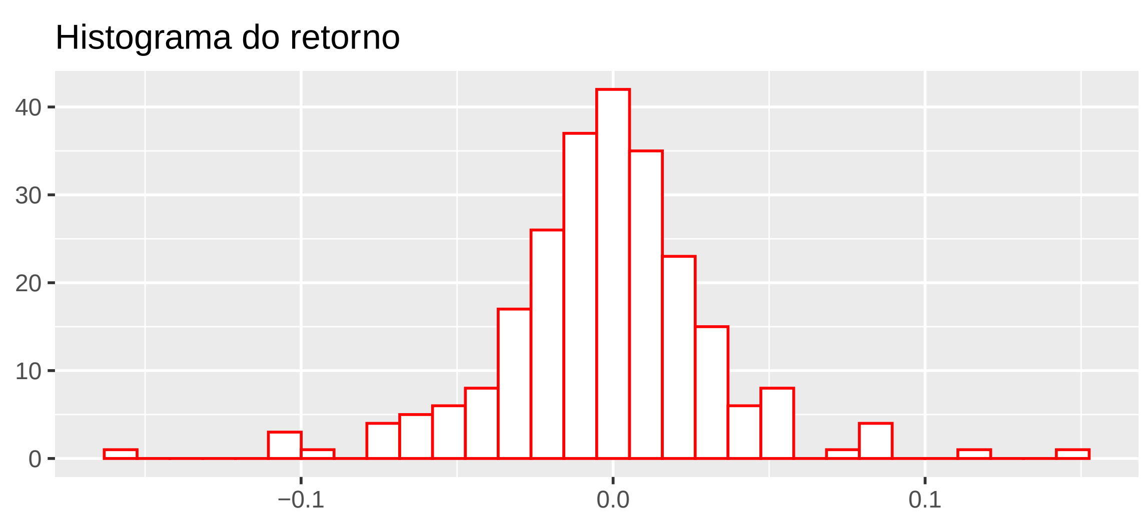


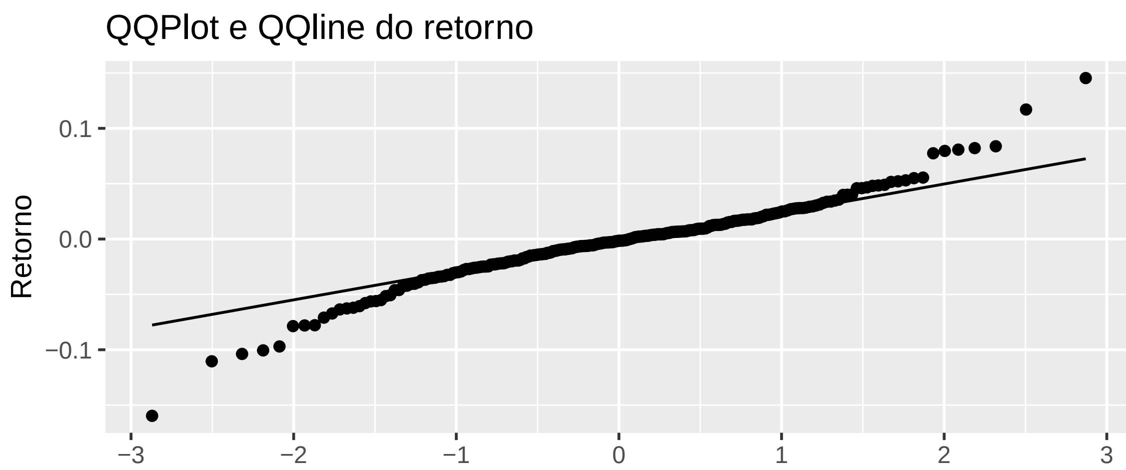



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```






```
## [1] "Assimetria amostral: -0.167268"
## [1] "Curtose amostral: 3.338171"
```

A assimetria amostra resultou -0,167268, ou seja, possui uma assimetria negativa, o que significa que os valores estão mais concentrados à direita. Além disso, como a curtose amostral é de 3,338171, a curtose é leptocúrtica, ou seja, a curva é mais "pontuda".

Questão 2

Faixa 1

O número total de possibilidades de resultados da aposta simples é pode ser calculado observando que há 6 possibilidades de números cuja ordem não importa na matriz normal e 2 na matriz de trevos, ou seja:

$$N = \frac{50!}{44! \times 6!} \times \frac{6!}{4! \times 2!}$$

Como só há um resultado final.

E sua probabilidade:

$$P = \frac{1}{N}$$

```r
N <- factorial(50)/(factorial(44)*factorial(6))*factorial(6)/(factorial(4)*factorial(2))
P1 <- 1/N
N
```

```
## [1] 238360500
```

```r
P1
```

```
## [1] 4.195326e-09
```

Assim, $\mathbf{N = 238360500}$ e $P1 = 4,195326 * 10^{-9}$

Faixa 2

Para a faixa dois ocorre deve-se acertar os 6 números da matriz normal e acertar 1 ou 2 da trevo, ou seja, o número de possibilidades disso ocorre é de:

$$N_2 = \frac{6!}{6!} \times (2 \times 4 + \frac{4!}{2! \times 2!})$$

```r
N2 <- (factorial(4)/(factorial(2)*factorial(2))+2*4)
# Chance em:
N/N2
```

```
## [1] 17025750
```

```r
# Probabilidade:
N2/N
```

```
## [1] 5.873456e-08
```

Assim, a chance é de 1 acerto em 17025750 e a probabilidade é de $5,873456 * 10^{-08}$.

Faixa 3

Como são 5 números corretos e 1 errado na matriz normal e 2 corretos na de trevo, temos que de 6 escolher 5, e um outro número dos 44 restantes:

$$N_3 = \frac{6!}{5! \times 1!} \times 44$$

```r
N3 <- factorial(6)/factorial(5)*44
# Chance em:
N/N3
```

```
## [1] 902880.7
```

```r
# Probabiliade
N3/N
```

```
## [1] 1.107566e-06
```

Portanto, a chance é de 1 acerto em 902881 e a probabilidade é de $1,107566 * 10^{-06}$.

Faixa 4

Repetindo o processo da faixa 2 e 3:

$$N_4 = \frac{6!}{5! \times 1!} \times 44 \times (2 \times 4 + \frac{4!}{2! \times 2!})$$

```r
N4 <- factorial(6)/factorial(5)*44*(2*4+factorial(4)/(factorial(2)*factorial(2)))
# Chance em:
N/N4
```

```
## [1] 64491.48
```

```r
# Probabiliade
N4/N
```

```
## [1] 1.550592e-05
```

Assim, a chance é de 1 acerto em 64491 e a probabilidade é de $1.550592 * 10^{-05}$.

Faixa 5

Como são 2 acertos da matriz do trevo e 4 na normal:

$$N_5 = \frac{6!}{4! \times 2!} \times \frac{44!}{42! \times 2!}$$

```r
N5 <- factorial(6)/(factorial(4)*factorial(2))*factorial(44)/(factorial(42)*factorial(2))
# Chance em:
N/N5
```

```
## [1] 16797.78
```

```r
# Probabiliade
N5/N
```

```
## [1] 5.953168e-05
```

Assim, a chance é de 1 acerto em 16798 e a probabilidade é de $5,953168 * 10^{-05}$.

Faixa 6

Repetindo o processo da faixa 2 e 5

$$N_6 = \frac{6!}{4! \times 2!} \times \frac{44!}{42! \times 2!} \times (2 \times 4 + \frac{4!}{2! \times 2!})$$

```r
N6 <- factorial(6)/(factorial(4)*factorial(2))*factorial(44)/(factorial(42)*factorial(2))*(factorial(4)
# Chance em:
N/N6
```

```
## [1] 1199.841
```

```r
# Probabiliade
N6/N
```

```
## [1] 0.0008334435
```

Assim, a chance é de 1 acerto em 1200 e a probabilidade é de 0,0008334435.

Faixa 7

Como são 3 acertos e 3 erros na matriz normal:

$$N_7 = \frac{6!}{3! \times 3!} \times \frac{44!}{41! \times 3!}$$

```r
N7 <- factorial(6)/(factorial(3)*factorial(3))*factorial(44)/(factorial(41)*factorial(3))
# Chance em:
N/N7
```

```
## [1] 899.8811
```

```r
# Probabiliade
N7/N
```

```
## [1] 0.001111258
```

Assim, a chance é de 1 acerto em 900 e a probabilidade é de 0,001111258.

Faixa 8

Repetindo o processo da faixa 7 e como queremos um unico acerto na matriz do trevo.

$$N_8 = \frac{6!}{3! \times 3!} \times \frac{44!}{41! \times 3!} \times (2 \times 4)$$

```r
N8 <- factorial(6)/(factorial(3)*factorial(3))*factorial(44)/(factorial(41)*factorial(3))*(2*4)
# Chance em:
N/N8
```

```
## [1] 112.4851
```

```r
# Probabiliade
N8/N
```

```
## [1] 0.008890064
```

Assim, a chance é de 1 acerto em 112 e a probabilidade é de 0,008890064.

Faixa 9

Como queremos dois acertos na normal e 4 erros:

$$N_9 = \frac{6!}{2! \times 4!} \times \frac{44!}{40! \times 4!}$$

```r
N9 <- factorial(6)/(factorial(4)*factorial(2))*factorial(44)/(factorial(40)*factorial(4))
# Chance em:
N/N9
```

```
## [1] 117.0577
```

```r
# Probabiliade
N9/N
```

```
## [1] 0.008542795
```

Assim, a chance é de 1 acerto em 117 e a probabilidade é de 0,008542795.

Faixa 10

Como queremos dois acertos na normal e 4 erros na matriz normal e apenas 1 acerto na do trevo:

$$N_{10} = \frac{6!}{2! \times 4!} \times \frac{44!}{40! \times 4!} \times (2 \times 4)$$

```r
N10 <- factorial(6)/(factorial(4)*factorial(2))*factorial(44)/(factorial(40)*factorial(4))*2*4
# Chance em:
N/N10
```

```
## [1] 14.63221
```

```r
# Probabiliade
N10/N
```

```
## [1] 0.06834236
```

Assim, a chance é de 1 acerto em 15 e a probabilidade é de 0,06834236.

Questão 3

Primeiramente, vamos criar uma função que recebe um argumento que é a probabilidade P(CF+) de determinado portador e retorna a probabilidade P(CF+/NEG) dada a fórmula fornecida no enunciado:

$$P(CF\_+\|Neg) = \frac{P(CF_+) \times P(Neg|CF_+)}{\sum_{k=1}^{2} P(CF_k) \times P(Neg|CF_k)}$$

```
ehportador <- function(prob){
  pncposi <- 0.01
  pncneg <- 0.998

  res <- prob*pncposi/(prob*pncposi+(1-prob)*pncneg)
  return(res)
}
```

Assim, respondendo os itens:

Item a)

Temos que P(CF+) = 0.040 para Caucasianos Não Hispânicos, assim:

```
# Probabilidade:
probA <- ehportador(0.04)
probA
```

```
## [1] 0.0004173274
```

```
# Indivíduos por milhão de habitantes:
probA*1e6
```

```
## [1] 417.3274
```

Dessa forma, a probabilidade para esse grupo é *P(CF+/NEG) = 0,0004173274*, tendo aproximadamente *417 pessoas* com a doença por milhão de habitantes.

Item b)

Temos que P(CF+) = 0,017 para Hispano-americanos, assim:

```
# Probabilidade:
probB <- ehportador(0.017)
probB
```

```
## [1] 0.0001732565
```

```
# Indivíduos por milhão de habitantes:
probB*1e6
```

```
## [1] 173.2565
```

Dessa forma, a probabilidade para esse grupo é *P(CF+/NEG) = 0,0001732565*, tendo aproximadamente *173 pessoas* com a doença por milhão de habitantes.

Item c)

Temos que P(CF+) = 0,016 para Afro-americanos, assim:

```
# Probabilidade:
probC <- ehportador(0.016)
probC
```

```
## [1] 0.0001629009
```

```r
# Indivíduos por milhão de habitantes:
probC*1e6
```

```
## [1] 162.9009
```

Dessa forma, a probabilidade para esse grupo é *P(CF+/NEG) = 0,0001629009*, tendo aproximadamente *163 pessoas* com a doença por milhão de habitantes.

Item d)

Temos que P(CF+) = 0,042 para Judeus Asquenazes, assim:

```r
# Probabilidade:
probD <- ehportador(0.042)
probD
```

```
## [1] 0.0004390991
```

```r
# Indivíduos por milhão de habitantes:
probD*1e6
```

```
## [1] 439.0991
```

Dessa forma, a probabilidade para esse grupo é *P(CF+/NEG) = 0,0004390991*, tendo aproximadamente *439 pessoas* com a doença por milhão de habitantes.

Item e)

Temos que P(CF+) = 0,011 para Asiático-americanos, assim:

```r
# Probabilidade:
probE <- ehportador(0.011)
probE
```

```
## [1] 0.0001114339
```

```r
# Indivíduos por milhão de habitantes:
probE*1e6
```

```
## [1] 111.4339
```

Dessa forma, a probabilidade para esse grupo é *P(CF+/NEG) = 0,0001114339*, tendo aproximadamente *111 pessoas* com a doença por milhão de habitantes.

Questão 4:

4.1)

a) Temos que:

$$f_X(x) = \frac{3}{x^4}, x > 1$$

Assim, pra provar que a função é função densidade de probabilidade temos que:

$$\int_1^\infty f_X(x)dx = \int_1^\infty \frac{3}{x^4}dx = \left.\frac{-1}{x^3}\right|_1^\infty = 1$$

b) Como f(x) é função densidade de probabilidade:

$$P(3,4 \leq X < 7,1) = \int_{3,4}^{7,1} f_X(x)dx = \frac{-1}{x^3}\Big|_{3,4}^{7,1} = 0,0226 = 2,26\%$$

c) Temos que o valor esperado é dado por:

$$E[X] = \int_1^{\infty} xf(x)dx = \int_1^{\infty} x\frac{3}{x^4}dx = \int_1^{\infty} \frac{3}{x^3}dx = -\frac{3}{2x^2}\Big|_1^{\infty} = \frac{3}{2}$$

d) Temos que $Var[X] = E[X^2] - E^2[X]$, assim:

$$E[X^2] = \int_1^{\infty} x^2 f(x)dx = \int_1^{\infty} x^2\frac{3}{x^4}dx = \int_1^{\infty} \frac{3}{x^2}dx = -\frac{3}{x}\Big|_1^{\infty} = 3$$

Ou seja:

$$Var[X] = 3 - \left(\frac{3}{2}\right)^2 = 3 - \frac{9}{4} = \frac{3}{4}$$

4.2) Como as variáveis são independentes é válido que $E[XY] = E[X]E[Y]$, dessa forma

$$E[Y] = 2$$
$$E[X^2Y] = 6 \Rightarrow E[X^2] = 3$$
$$E[(XY)^2] = E[X^2Y^2] = 24 \Rightarrow E[Y^2] = 8$$
$$E[XY^2] = 8 \Rightarrow E[X] = 1$$

4.3) Temos que função densidade de probabilidade para uma variável aleatória com distribuição deCauchy é dada por:

$$f_X(x) = \frac{1}{\pi b\left[1 + \left(\frac{x-a}{b}\right)^2\right]}, x \in \mathbb{R}$$

Como a distribuição Cauchy padrão é obtida fazendo a = 0 e b = 1.

$$f_X(x) = \frac{1}{\pi(1+x^2)} E[X] = \int_{-\infty}^{\infty} xf_X(x)dx = \int_{-\infty}^{\infty} x\frac{1}{\pi(1+x^2)}dx$$

Como $x\frac{1}{\pi(1+x^2)}$ é impar $\int_{-\infty}^{\infty} x\frac{1}{\pi(1+x^2)}dx = 0$, ou seja, E[X] = 0.