

Parte 1: deve ser resolvida sem uso do R.

1. Considere os comandos abaixo

```
> x <- 0:5; y <- seq(from=4, by=0.5, length.out=length(x))  
> z <- matrix(x, ncol=2); w <- 2*matrix(y, ncol=ncol(z))
```

Escreva por extenso o resultado que aparece no console para cada um desses quatro objetos.

> x

```
[1] 0 1 2 3 4 5
```

> y

```
[1] 4.0 4.5 5.0 5.5 6.0 6.5
```

> z

```
      [,1] [,2]  
[1,]    0    3  
[2,]    1    4  
[3,]    2    5
```

> w

```
      [,1] [,2]  
[1,]    8   11  
[2,]    9   12  
[3,]   10   13
```

2. Considere os comandos abaixo

```
> x <- (0:7)+0.5; y <- seq(from=3, to=6, length.out=7)  
> z <- matrix(x, ncol=2); w <- 2*matrix(y, 3, 3)
```

Escreva por extenso o resultado que aparece no console para cada um dos comandos abaixo.

a) `x[1:3]`

e) `y[x[4:5]-0.5]`

b) `y[c(2,4,6)]`

f) `z[z[,1]<2,]`

c) `z[,1]`

g) `w<=8`

d) `w[3,]`

h) `sum(z<4)`

> `x[1:3]`

```
[1] 0.5 1.5 2.5
```

> `y[c(2,4,6)]`

```
[1] 3.5 4.5 5.5
```

> `z[,1]`

```
[1] 0.5 1.5 2.5 3.5
```

> `w[3,]`

```

[1] 8 11 7

> y[x[4:5]-0.5]

[1] 4.0 4.5

> z[z[,1]<2,]

      [,1] [,2]
[1,] 0.5  4.5
[2,] 1.5  5.5

> w<=8

      [,1] [,2] [,3]
[1,] TRUE FALSE FALSE
[2,] TRUE FALSE TRUE
[3,] TRUE FALSE TRUE

> sum(z<4)

[1] 4

```

Parte 2: deve ser resolvida com uso do R. As respostas são os códigos empregados para resolver os problemas propostos e o resultados da avaliação dos códigos.

3. Considere o conjunto de dados `USArrests` do R.

- Informe a estrutura do objeto (dimensões, nomes, tipos de valores).
- Obtenha a tabela de distribuição de frequências absolutas para a primeira variável do objeto usando os limites de classe dessa sequência regular 0, 4, ..., 20.
- Baseado no número de dados e na amplitude total sugira outros limites de classe para a variável `Murder` e obtenha a respectiva tabela de distribuição de frequências.
- Por meio de gráficos descreva a relação entre `Murder` e `Assault` e entre `Murder` e `UrbanPop` em termos de forma, direção e intensidade.

```

> # a)
> data(USArrests)
> str(USArrests)

'data.frame':      50 obs. of  4 variables:
 $ Murder   : num  13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...
 $ Assault  : int  236 263 294 190 276 204 110 238 335 211 ...
 $ UrbanPop : int  58 48 80 50 91 78 77 72 80 60 ...
 $ Rape     : num  21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...

> # b)
> class <- seq(0,20,4)
> tb <- table(cut(USArrests[,1], class))
> tb

 (0,4]  (4,8]  (8,12] (12,16] (16,20]
      13      15      11       9       2

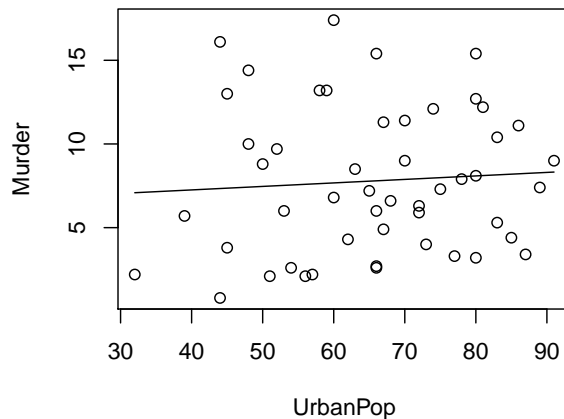
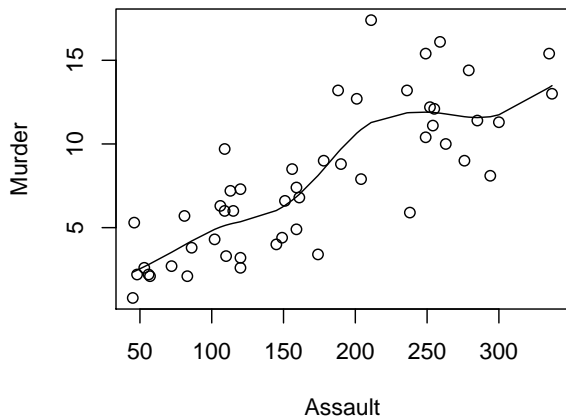
> # c)
> # 50 dados, usar a regra da raiz(n)
> sqrt(nrow(USArrests))

[1] 7.071068

> range(USArrests$Murder)

[1] 0.8 17.4

```



```

> seq(0, 18, l=7+1)      # 7 classes, limites irracionais
[1] 0.000000 2.571429 5.142857 7.714286 10.285714 12.857143 15.428571
[8] 18.000000

> seq(0, 18, l=6+1)      # 6 classes, limites inteiros
[1] 0 3 6 9 12 15 18

> seq(0.5, l=7+1, by=2.5) # 7 classes, limites decimais
[1] 0.5 3.0 5.5 8.0 10.5 13.0 15.5 18.0

> seq(0, l=7+1, by=2.5) # 7 classes, limites decimais
[1] 0.0 2.5 5.0 7.5 10.0 12.5 15.0 17.5

> tb <- table(cut(USArrests$Murder, seq(0, 18, l=6+1)))
> tb

 (0,3] (3,6] (6,9] (9,12] (12,15] (15,18]
      8     13     12      6      7      4

> # d)
> plot(Murder~Assault, data=USArrests)
> with(USArrests, lines(smooth.spline(Assault, Murder)))
> # relação linear, crescente de intensidade moderada
> plot(Murder~UrbanPop, data=USArrests)
> with(USArrests, lines(smooth.spline(UrbanPop, Murder)))
> # sem relação
> # as linhas de tendência são opcionais

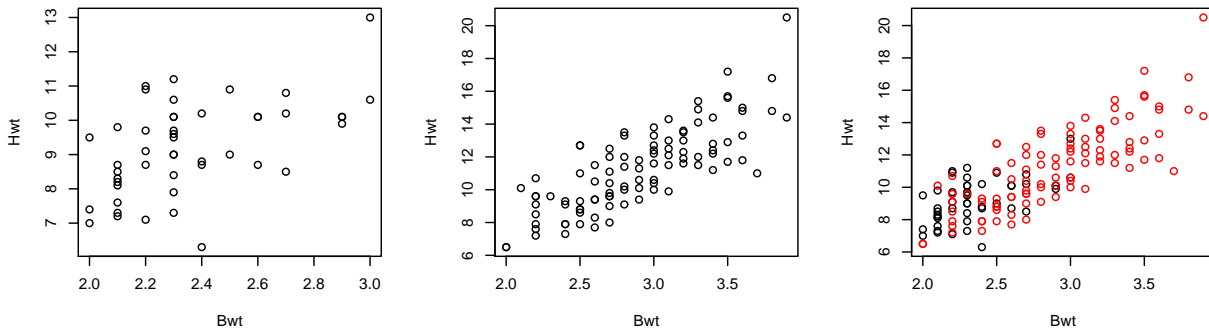
```

-
4. Considere o conjunto de dados `cats` do R (`data(cats, package="MASS")`).
- Informe a estrutura do objeto (dimensões, nomes, tipos de valores).
 - Obtenha para cada nível de `Sex` o número observações, o mínimo, o máximo e a média dos valores de `Hwt`.
 - Faça um gráfico de dispersão de `Hwt` em função de `Bwt` para cada nível de `Sex`. Descreva-o.
 - Faça um gráfico de dispersão de `Hwt` em função de `Bwt` e represente os níveis de `Sex` com símbolos ou cores diferentes. Descreva-o.

```

> # a)
> data(cats, package="MASS")
> str(cats)

```



```
'data.frame':      144 obs. of  3 variables:
 $ Sex: Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 1 1 1 ...
 $ Bwt: num  2 2 2 2.1 2.1 2.1 2.1 2.1 2.1 2.1 ...
 $ Hwt: num  7 7.4 9.5 7.2 7.3 7.6 8.1 8.2 8.3 8.5 ...

> # b)
> myfun <- function(x){c(nobs=length(x), media=mean(x), minimo=min(x), maximo=max(x))}
> with(cats, tapply(Hwt, Sex, myfun))

$F
  nobs  media  minimo  maximo
47.000000 9.202128 6.300000 13.000000

$M
  nobs  media  minimo  maximo
97.000000 11.32268 6.50000 20.50000

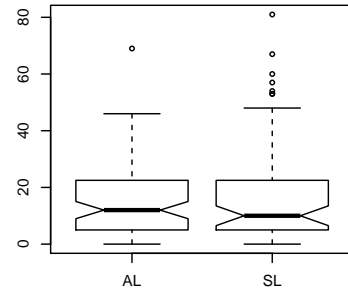
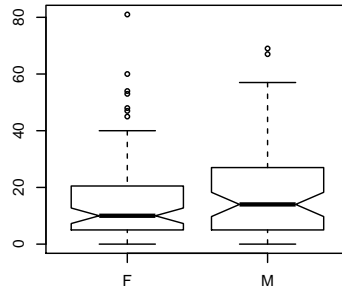
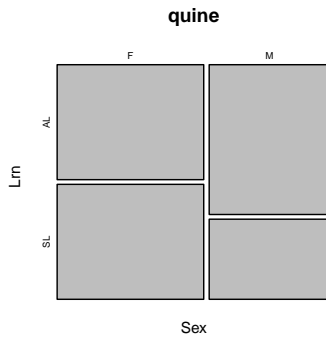
> # c)
> plot(Hwt~Bwt, subset(cats, Sex=="F"))
> plot(Hwt~Bwt, subset(cats, Sex=="M"))
> #par(mfrow=c(1,2)) # modo ninja
> #by(cats, INDICES=cats$Sex, FUN=function(i) plot(Hwt~Bwt, data=i))
> #layout(1)
> # d)
> plot(Hwt~Bwt, cats, col=as.numeric(cats$Sex))
> plot(Hwt~Bwt, cats, col=ifelse(cats$Sex=="M", 3, 4))
```

5. O conjunto de dados quine é sobre as crianças de Walgett em New South Wales, Austrália. Elas foram classificadas pela cultura (Eth), idade (Age), sexo (Sex) e estado de aprendizado (Lrn) e o número de dias de ausência da escola (Days) em um ano letivo especial (`data(quine, package="MASS")`). Use gráficos adequados para investigar a relação entre as variáveis Sex, Lrn e Days (duas à duas). Descreva as relações observadas. Para dar suporte à sua descrição, apresente medidas de frequência, posição e dispersão.

```
> data(quine, package="MASS")
> str(quine)

'data.frame':      146 obs. of  5 variables:
 $ Eth : Factor w/ 2 levels "A","N": 1 1 1 1 1 1 1 1 1 1 ...
 $ Sex : Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 2 2 ...
 $ Age : Factor w/ 4 levels "F0","F1","F2",...: 1 1 1 1 1 1 1 1 2 2 ...
 $ Lrn : Factor w/ 2 levels "AL","SL": 2 2 2 1 1 1 1 1 2 2 ...
 $ Days: int  2 11 14 5 5 13 20 22 6 6 ...

> # Sex vs Lrn (categórica vs categórica)
> mosaicplot(~Sex+Lrn, data=quine)
> xt <- xtabs(~Sex+Lrn, data=quine)/nrow(quine)
> xt
```



```
Lrn
Sex      AL      SL
F 0.2739726 0.2739726
M 0.2945205 0.1575342
```

```
> colSums(xt)
```

```
      AL      SL
0.5684932 0.4315068
```

```
> rowSums(xt)
```

```
      F      M
0.5479452 0.4520548
```

```
> # a amostra contém mais mulheres e para elas metade são AL.
> # Para os tem mais AL. Deve haver uma associação poritiva entre Homens e AL.
>
> # Sex vs Days (categórica vs contínua)
> boxplot(Days~Sex, data=quine, notch=TRUE)
> with(quine, tapply(Days, Sex, summary))
```

```
$F
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00   5.00   10.00   15.22  20.25   81.00
```

```
$M
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00   5.25   14.00   17.95  27.00   69.00
```

```
> # valores centrais dos homens são maiores. A AIQ dos homens é maior mas a AT é menor.
>
> # Lrn vs Days (categórica vs contínua)
> boxplot(Days~Lrn, data=quine)
> with(quine, tapply(Days, Lrn, summary))
```

```
$AL
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00   5.00   12.00   15.82  22.50   69.00
```

```
$SL
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.0    5.0   10.0   17.3   22.5   81.0
```

```
> # todas as estatísticas de descritivas e o gráfico mostram comportamentos
> # bem próximos (não existe efeito) entre os níveis de Lrn em relação à Days.
```