

PROBABILIDADE E ESTATÍSTICA

Professor Petrúcio Barros

Aluna: **Karla Sophia Santana da Cruz**

Matrícula: **19111113**

Aluna: **Maria Eduarda Cardoso Aciole**

Matrícula: **19111117**

Curso: **Engenharia de Computação**

Período: **2020.2**

Questão 1

Dada a lista a seguir,

```
1 lista <- c(48, 58, 56, 63, 52, 50, 59, 51, 59, 38, 57, 56, 73, 61, 41, 55, 49, 61,  
2           49, 49, 52, 55, 60, 52, 54, 57, 47, 66, 60, 53, 59, 50, 45, 57, 64, 56,  
3           57, 60, 47, 58, 53, 58, 66, 47, 40)  
4  
5 grid <- grid(nx = NA, ny = NULL, lty = 2, col = "gray", lwd = 1)
```

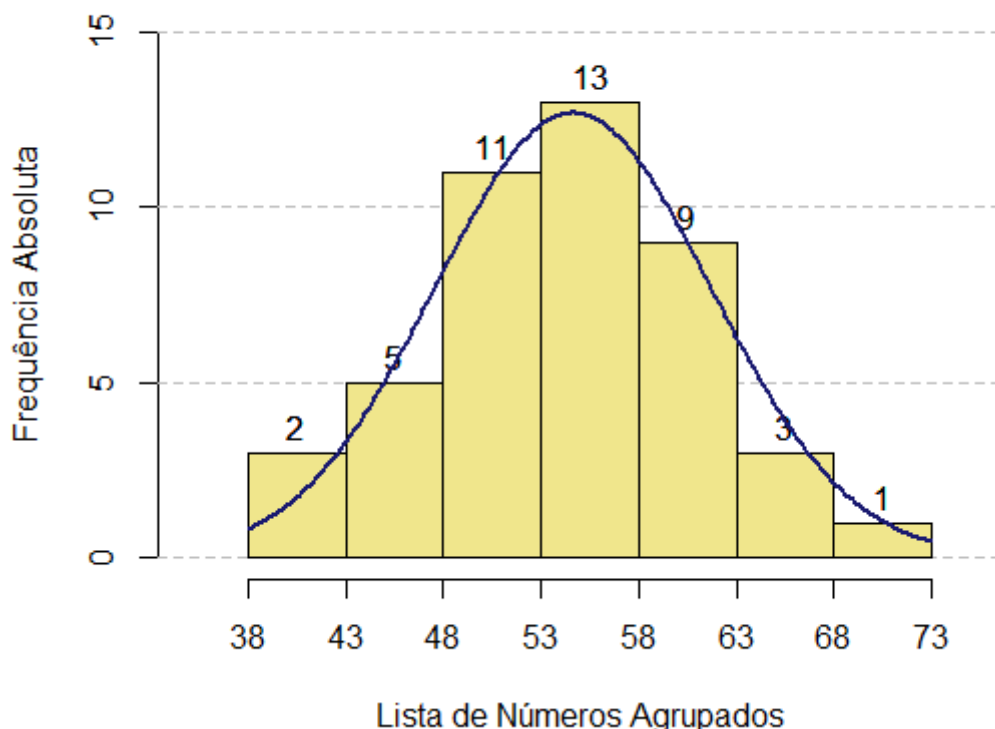
A) Histograma agrupado com a frequência absoluta.

IMPLEMENTAÇÃO

```
7 #####  
8  
9 #LETRA A: HISTOGRAMA DA FREQUÊNCIA DA LISTA DE NÚMEROS AGRUPADOS  
10  
11 lim_intervalo <- c(38, 43, 48, 53, 58, 63, 68, 73)  
12  
13 lista_df <- as.data.frame(table(cut(lista, breaks = lim_intervalo)))  
14  
15 h_lista = hist(lista, col = "khaki",  
16                main = "HISTOGRAMA DA LISTA DE NÚMEROS AGRUPADOS",  
17                xlab = "Lista de Números Agrupados", ylab = "Frequência Absoluta",  
18                xlim = c(35, 75), ylim = c(0, 15),  
19                xaxp = c(38, 73, 7), add = TRUE,  
20                labels = as.character(lista_df$Freq), breaks = lim_intervalo)  
21  
22 xfit <- seq(min(lista), max(lista), length = 200)  
23 yfit <- dnorm(xfit, mean(lista), sd(lista))  
24 yfit <- yfit*diff(h_lista$mids[1:2])*length(lista)  
25 lines(xfit, yfit, col = "midnightblue", lwd = 2)  
26  
27 #####
```

GRÁFICO

HISTOGRAMA DA LISTA DE NÚMEROS AGRUPADOS



B) Em cálculos diferentes, com os dados agrupados, encontre:

1. Média

IMPLEMENTAÇÃO

```
51 #MÉDIA
52 media <- mean(lista)
```

RESULTADO

```
> media
[1] 54.62222
```

2. Moda de Czuber

IMPLEMENTAÇÃO

```
31 ##### FUNÇÃO MODA CZUBER #####
32
33 moda.czuber <- function(dados){
34   amp <- max(dados) - min(dados)
35   nk <- round(1 + 3.222 * log10(length(dados)))
36   amp_classe <- amp / nk
37
38   lim_classe <- seq(min(dados), max(dados), amp_classe)
39   tab_classe <- table(cut(dados, breaks = lim_classe))
40
41   lim_inferior <- lim_classe[which.max(tab_classe)]
42
43   d1 <- max(tab_classe) - as.numeric(tab_classe[which.max(tab_classe) - 1])
44   d2 <- max(tab_classe) - as.numeric(tab_classe[which.max(tab_classe) + 1])
45
46   return(lim_inferior + (d1 / (d1 + d2)) * amp_classe)
47 }
48
49 #####
57 #MODA DE CZUBER
58 moda_czuber <- moda.czuber(lista)
```

RESULTADO

```
> moda_czuber
[1] 57.44444
```

3. Mediana

IMPLEMENTAÇÃO

```
54 #MEDIANA
55 mediana <- median(lista)
```

RESULTADO

```
> mediana
[1] 56
```

4. Terceiro Quartil

IMPLEMENTAÇÃO

```
60 #TERCEIRO QUARTIL
61 terceiro_quartil <- quantile(lista, 0.75)
```

RESULTADO

```
> terceiro_quartil
75%
59
```

5. Percentis 8, 50 e 80

IMPLEMENTAÇÃO

```
63 #PERCENTIL 8
64 percentil_8 <- quantile(lista, 0.08)
65
66 #PERCENTIL 50
67 percentil_50 <- quantile(lista, 0.5)
68
69 #PERCENTIL 80
70 percentil_80 <- quantile(lista, 0.8)
```

RESULTADO

```
> percentil_8
8%
46.04
> percentil_50
50%
56
> percentil_80
80%
60
```

C) Boxplot e comentários referentes aos resultados apresentados nos gráficos.

IMPLEMENTAÇÃO

```
76 #BOXPLOT VERTICAL
77
78 boxplot_lista_v <- boxplot(lista, add = TRUE,
79                             main = "Boxplot da Lista de Números",
80                             ylab = "Lista de Números",
81                             col = "khaki")
82
83 #BOXPLOT HORIZONTAL
84
85 boxplot_lista_h <- boxplot(lista, horizontal = TRUE, notch = TRUE,
86                             main = "Boxplot da Lista de Números",
87                             ylab = "Lista de Números",
88                             col = "khaki")
89
```

GRÁFICO VERTICAL

Boxplot da Lista de Números

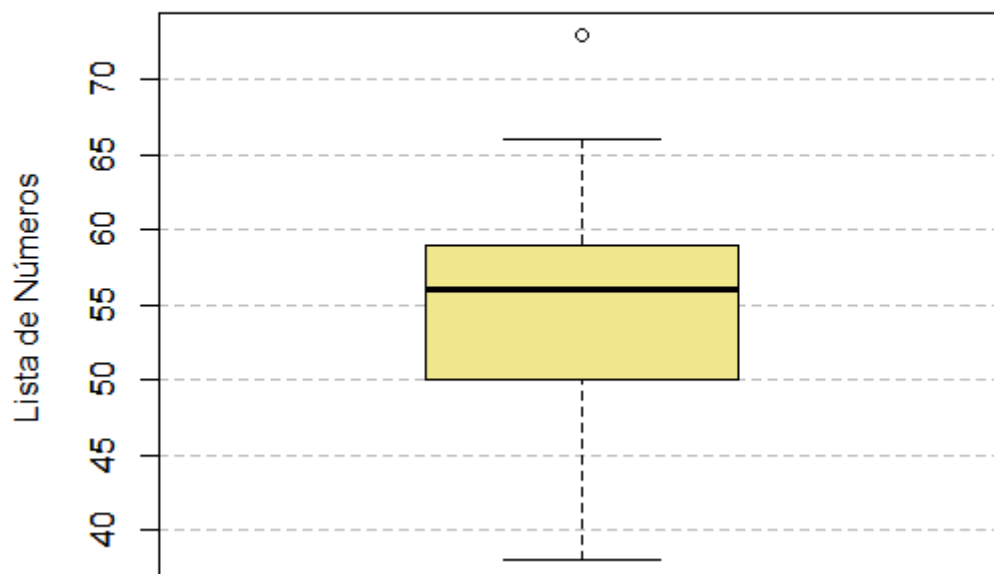
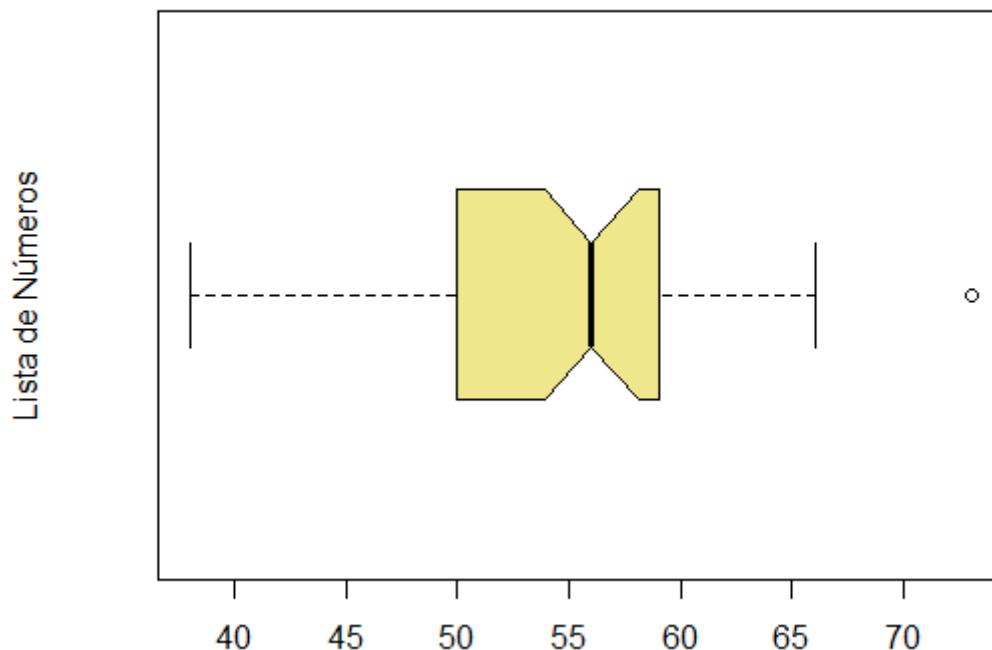


GRÁFICO HORIZONTAL

Boxplot da Lista de Números



A partir da análise de dados da Lista de Números, com mínimo 38 e máximo 73, pode-se afirmar que a maior frequência está no intervalo 53 - 58, como também a média, a mediana e a Moda de Czuber. Exceto pelo percentil 50, o terceiro quartil e os percentis 8 e 80 fogem um pouco desse intervalo, estando entre 46 - 47 e 59 - 60.

Questão 2

Dada a base de dados a seguir,

```
1 dados <- read.csv("dadosPacientes2021.csv")
2
3 dados[dados == "Masculino"] <- "Masculino"
4
5 check_20 <- grepl("2020", dados$data_resultado_exame, fixed = TRUE)
6 check_21 <- grepl("2021", dados$data_resultado_exame, fixed = TRUE)
7
8 dados <- subset(dados, check_20 == TRUE | check_21 == TRUE)
9 dados <- subset(dados, idade > 0 & idade < 120)
10
11 grid <- grid(nx = NA, ny = NULL, lty = 2, col = "gray", lwd = 1)
12
```

Com a base de dados *dadosPacientes2021.csv* com mais de 190 mil pacientes, foi necessário realizar um tratamento sobre alguns erros encontrados, tais como erros de digitação do sexo do paciente, datas e idades inválidas. Após o tratamento, a base retirou quase mil casos.

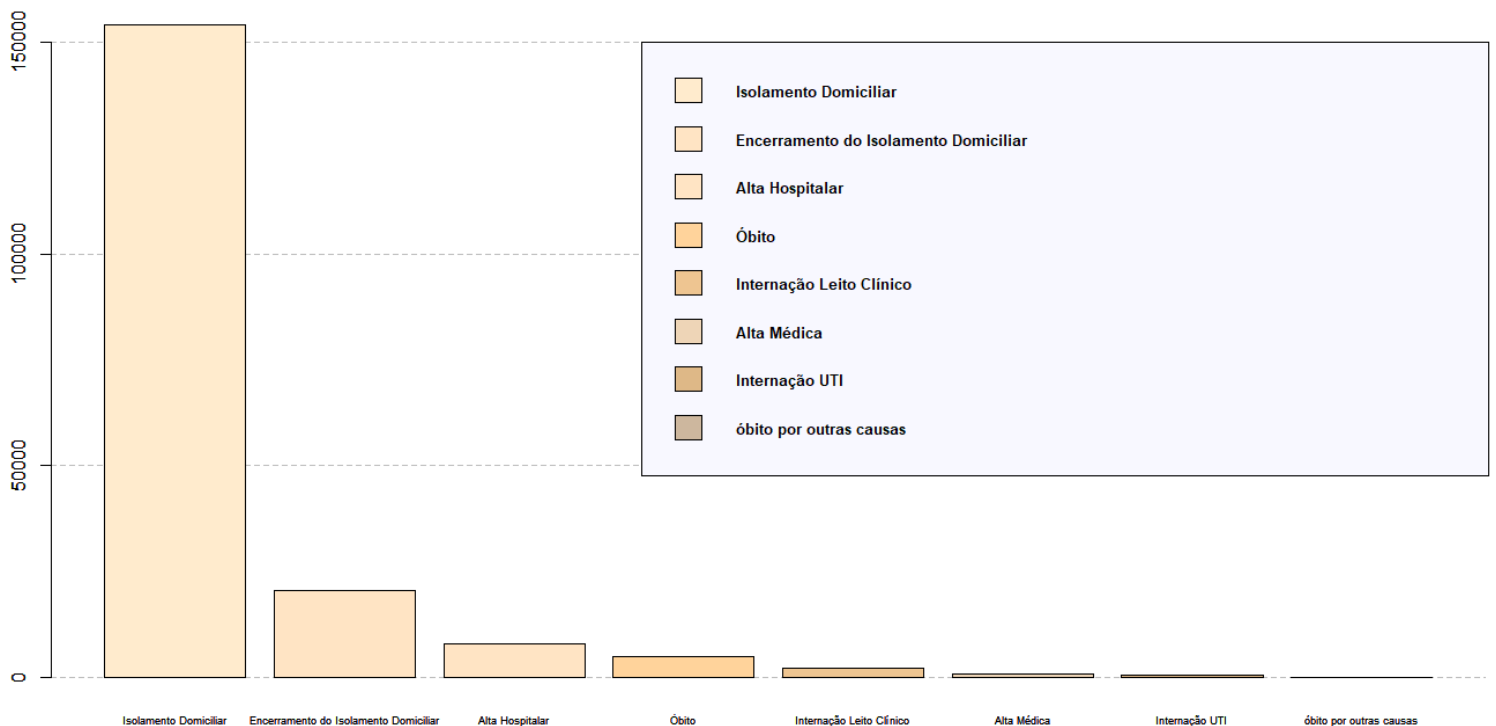
A) Gráfico de barras da situação atual dos pacientes.

IMPLEMENTAÇÃO

```
15 #LETRA A: GRÁFICO DE BARRAS DA SITUAÇÃO ATUAL DOS PACIENTES
16
17 gb_situacao = barplot(sort(table(dados$situacao_atual)), decreasing = TRUE),
18                       cex.names = 0.59, srt = 45, ylim = c(0, 200000, 10000),
19                       legend = TRUE, args.legend = list(x = 10, y = 150000,
20                                                         cex = 0.8, text.font = 2, bg = "ghostwhite"),
21                       col = c("blanchedalmond", "bisque", "bisque1", "burlywood1",
22                               "burlywood2", "bisque2", "burlywood", "bisque3"),
23                       main = "GRÁFICO DE BARRAS DA SITUAÇÃO ATUAL DOS PACIENTES")
24
25 #####
```

GRÁFICO

GRÁFICO DE BARRAS DA SITUAÇÃO ATUAL DOS PACIENTES



Pode-se analisar a partir do Gráfico da Situação Atual dos Pacientes que quase 80% dos pacientes estão em isolamento domiciliar, enquanto aproximadamente 5% dos pacientes vieram a óbito. Seguido do isolamento estão Encerramento do Isolamento Domiciliar e Alta Hospitalar. Na situação com menos pacientes, encontra-se os óbitos por outras causas.

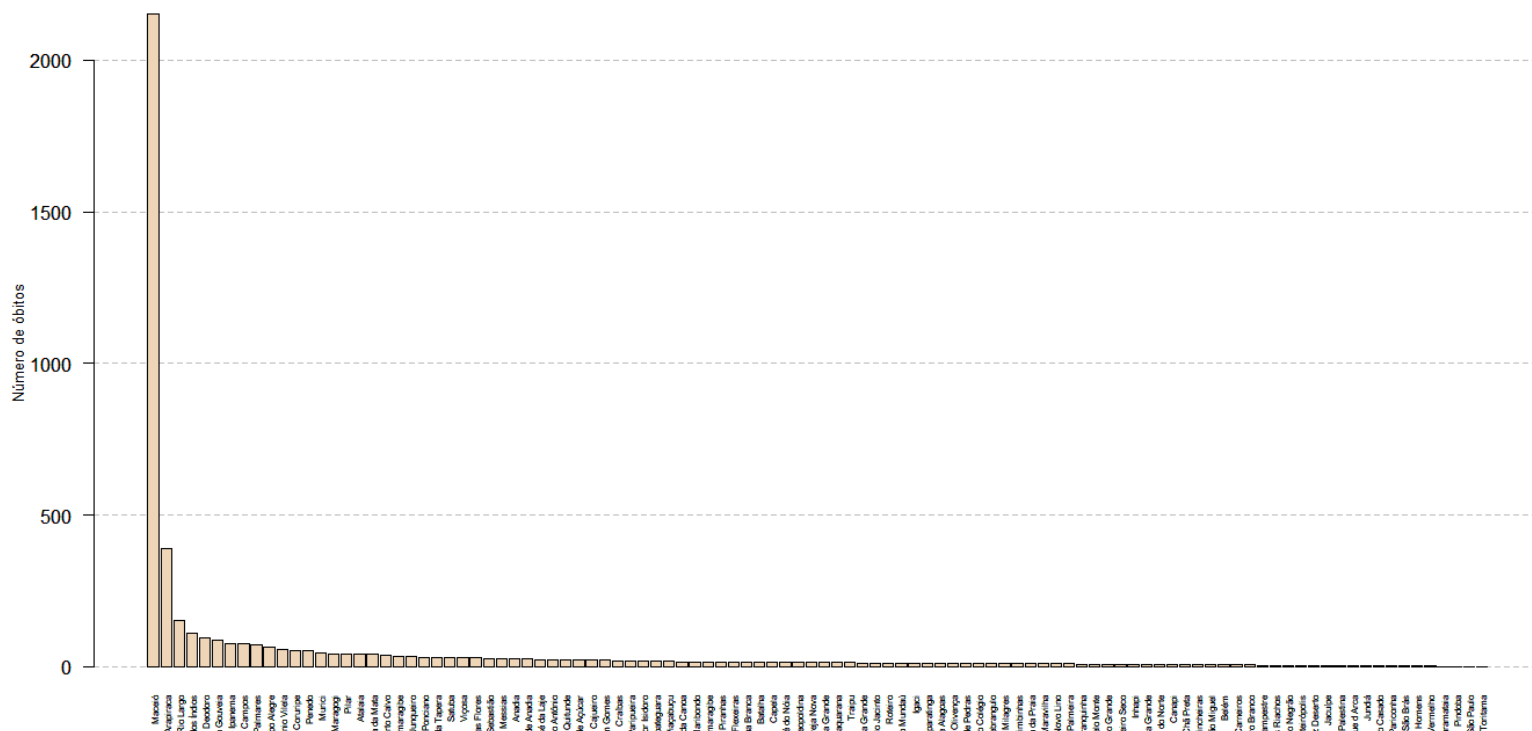
B) Gráfico de barras, representando o número de óbitos por município.

IMPLEMENTAÇÃO

```
27 #LETRA B: GRÁFICO DE BARRAS DOS ÓBITOS POR MUNICÍPIO
28
29 obitos <- subset(dados, situacao_atual == "óbito")
30
31 obitos_muni <- table(obitos$municipio_residencia)
32
33 gb_obitos_muni <- barplot(sort(obitos_muni, decreasing = TRUE), col = "bisque2",
34                           main = "GRÁFICO DE BARRAS DOS ÓBITOS POR MUNICÍPIO",
35                           ylab = "Número de óbitos", add = TRUE,
36                           cex.axis = 1, cex.names = 0.5, las = 2, cex.lab = 0.8, srt = 45)
37
38 #####
```

GRÁFICO

GRÁFICO DE BARRAS DOS ÓBITOS POR MUNICÍPIO

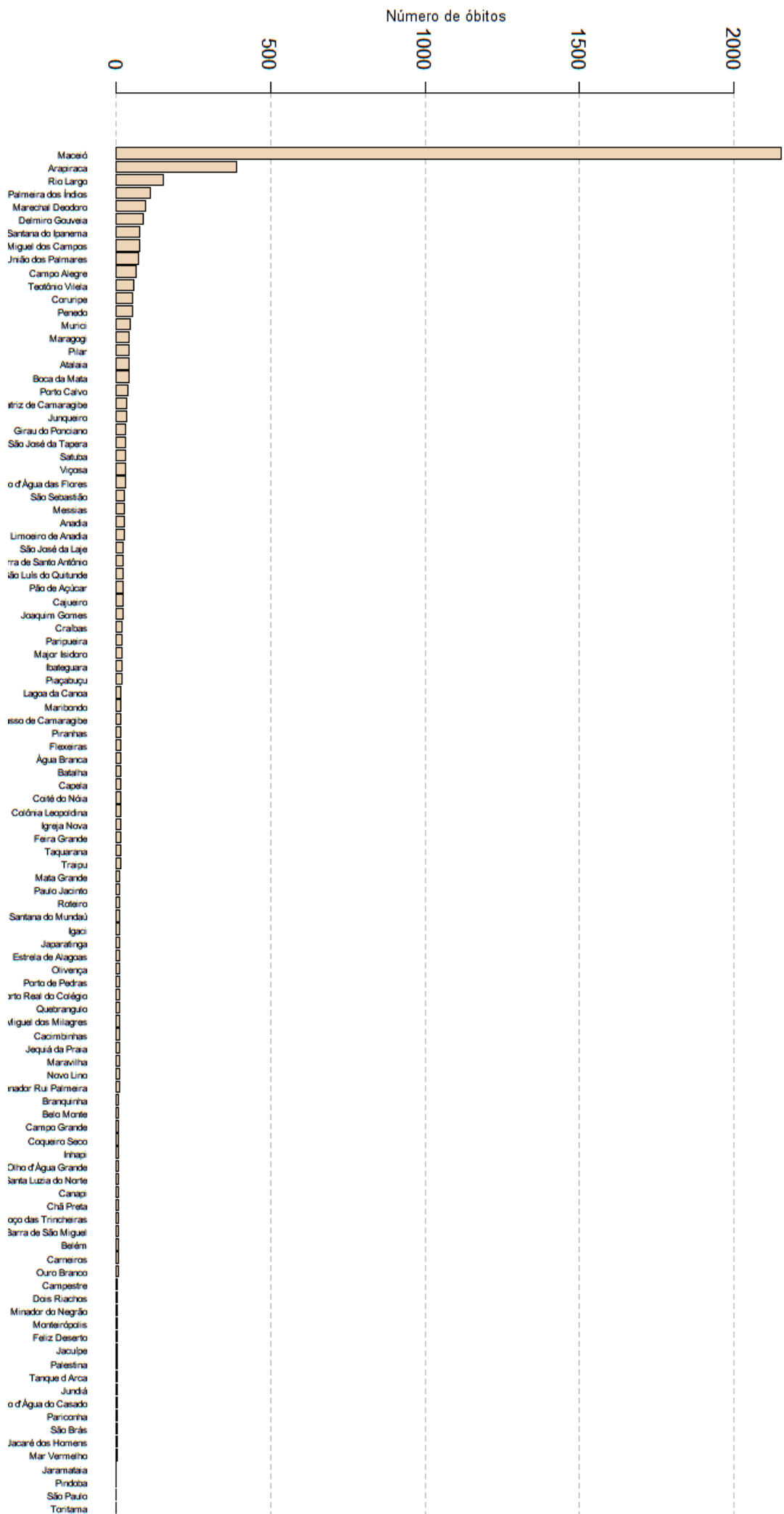


Com o gráfico da quantidade de óbitos por município, é notório perceber que a capital de Alagoas, Maceió, obtém a maior parte das mortes de COVID, com mais de 2000 óbitos, aproximadamente 25% dos óbitos totais no estado. Seguida dela está Arapiraca, segunda maior cidade de Alagoas em população e território, também ocupa essa posição no número de mortes, com quase 500 casos.

Os outros casos estão distribuídos nas mais de 100 cidades restantes do estado. Pode-se analisar que a quantidade de habitantes de cada local possivelmente é diretamente proporcional à quantidade de óbitos da cidade.

Para uma melhor análise do gráfico, segue este na página seguinte na orientação paisagem.

GRÁFICO DE BARRAS DOS ÓBITOS POR MUNICÍPIO

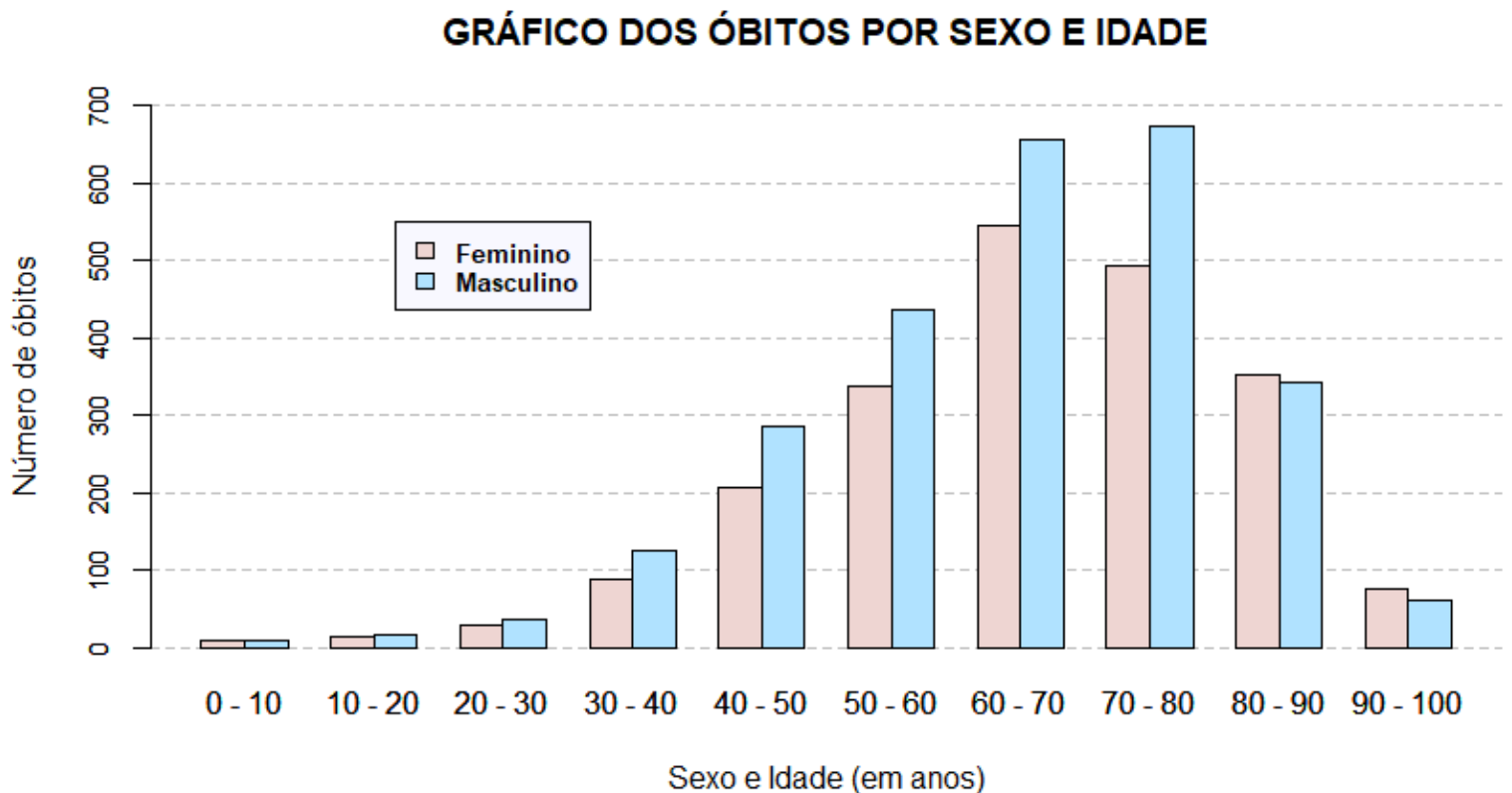


C) Gráfico representando o número de óbitos por sexo e idade. Considere para a idade um intervalo de classe de 10 anos.

IMPLEMENTAÇÃO

```
40 #LETRA C: GRÁFICO DOS ÓBITOS POR SEXO E IDADE
41
42 lim_idade <- c("0 - 10", "10 - 20", "20 - 30", "30 - 40", "40 - 50", "50 - 60",
43               "60 - 70", "70 - 80", "80 - 90", "90 - 100")
44
45 obitos_idade <- cut(obitos$idade, seq(0, 100, by = 10))
46
47 obitos_sexo_idade <- table(obitos$sexo, obitos_idade)
48
49 gb_obitos_sexidade <- barplot(obitos_sexo_idade, beside = TRUE,
50                               col = c("mistyrose2", "lightskyblue1"),
51                               main = "GRÁFICO DOS ÓBITOS POR SEXO E IDADE",
52                               ylim = c(0, 700), add = TRUE, cex.axis = 0.8,
53                               names.arg = lim_idade, legend = TRUE,
54                               args.legend = list(x = 10, y = 550, cex = 0.8, text.font = 2,
55                                                  bg = "ghostwhite"),
56                               ylab = "Número de óbitos", xlab = "Sexo e Idade (em anos)")
57
58 #####
```

GRÁFICO



Dado o gráfico dos óbitos por sexo e idade, é interessante analisar que a quantidade de óbitos entre 0 e 10 anos se manteve igual entre os sexos, com um aumento de óbitos do sexo masculino entre 10 e 80 anos, com seu pico entre os 70 e 80 anos. Porém, entre 80 e 100 anos, a quantidade de óbitos do sexo feminino é maior que o masculino.

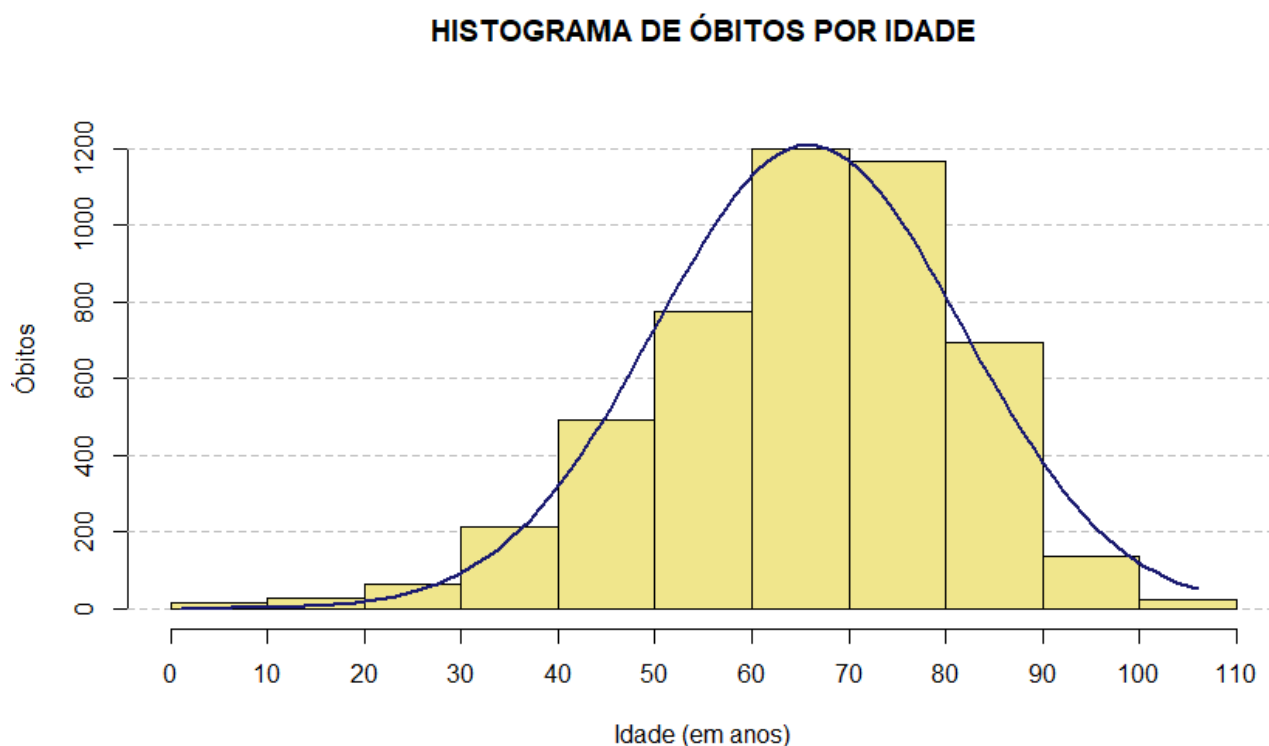
Nota-se que o pico de óbitos em ambos os sexos se dá entre 60 e 80 anos, porém, há uma diminuição de óbitos do sexo feminino a partir dos 70 anos, enquanto no masculino, essa diminuição ocorre somente a partir dos 80 anos.

D) Gráfico de óbitos por idade (histograma).

IMPLEMENTAÇÃO

```
60 #LETRA D: HISTOGRAMA DE ÓBITOS POR IDADE
61
62 h_obitos_idade <- hist(obitos$idade, col = "khaki",
63                        main = "HISTOGRAMA DE ÓBITOS POR IDADE",
64                        xlab = "Idade (em anos)", ylab = "Óbitos", add= TRUE,
65                        xlim = c(0, 110), ylim = c(0, 1300), xaxp = c(0, 110, 11))
66
67 xfit <- seq(min(obitos$idade), max(obitos$idade), length = 100)
68 yfit <- dnorm(xfit, mean(obitos$idade), sd(obitos$idade))
69 yfit <- yfit*diff(h_obitos_idade$mids[1:2])*length(obitos$idade)
70 lines(xfit, yfit, col = "midnightblue", lwd = 2)
71
72 #####
```

GRÁFICO



Como na questão anterior, percebe-se ao analisar o histograma de óbitos por idade que a quantidade de mortes segue crescente até o seu pico entre 60 e 70 anos, com aproximadamente 1200 mortes e decrescendo a partir daí.

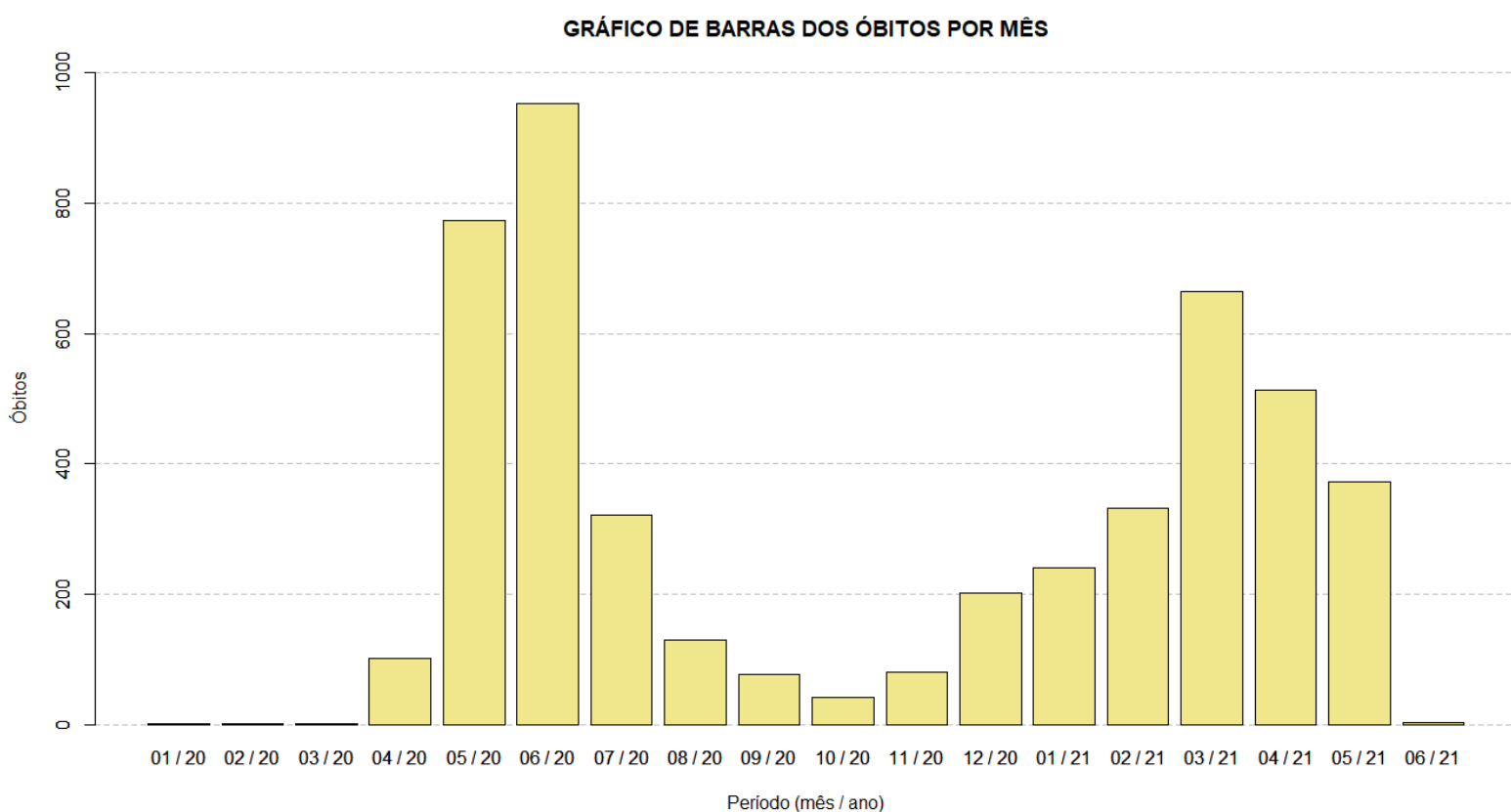
Pode-se ligar o número de óbitos a sua idade, pois, mesmo aparentemente havendo uma diminuição entre as idades mais avançadas, vale ressaltar que em relação ao número de contaminações pelas idades, a porcentagem de óbitos a partir de 60 anos é significativamente maior do que nos outros casos.

E) Gráfico de barras, representando o número de óbitos por mês.

IMPLEMENTAÇÃO

```
74 #LETRA E: GRÁFICO DE BARRAS DOS ÓBITOS POR MÊS
75
76 lim_datas <- c("01 / 20", "02 / 20", "03 / 20", "04 / 20", "05 / 20", "06 / 20",
77               "07 / 20", "08 / 20", "09 / 20", "10 / 20", "11 / 20", "12 / 20",
78               "01 / 21", "02 / 21", "03 / 21", "04 / 21", "05 / 21", "06 / 21")
79
80 obito_data <- as.Date(obitos$data_resultado_exame)
81 obito_data <- table(format(obito_data, '%y %m'))
82
83 gb_obitos_mes <- barplot(date, col = "khaki", add = TRUE,
84                           main = "GRÁFICO DE BARRAS DOS ÓBITOS POR MÊS",
85                           xlab = "Período (mês / ano)", ylab = "Óbitos",
86                           ylim = c(0, 1000), names.arg = lim_datas)
87
88 #####
```

GRÁFICO



Ao analisar o gráfico dos óbitos por mês, percebe-se dois picos com intervalo de 9 meses, no período de 06/20, com quase mil mortes, e 03/21, com quase 700 casos. Percebe-se uma diminuição nesse intervalo até chegar ao menor número no mês de outubro de 2020, crescendo a partir daí.

Vale ressaltar que, entre janeiro e março de 2020, quase não houve óbitos por COVID, porém há um aumento de quase 100 casos em 04/20 e principalmente em 05/20, quando em comparação ao mês anterior, o número de óbitos aumentou em quase 8 vezes.

Em 06/21, a diminuição se deve à desatualização da base de dados no número de confirmações de óbitos por COVID.

Questão 3

Dada a base de dados a seguir,

```
1 dados <- read.csv("dadosPacientes2021.csv")
2
3 dados[dados == "Masculino"] <- "Masculino"
4
5 check_20 <- grepl(("2020"), dados$data_resultado_exame, fixed = TRUE)
6 check_21 <- grepl(("2021"), dados$data_resultado_exame, fixed = TRUE)
7
8 dados <- subset(dados, check_20 == TRUE | check_21 == TRUE)
9 dados <- subset(dados, idade > 0 & idade < 120)
10
11 grid <- grid(nx = NA, ny = NULL, lty = 2, col = "gray", lwd = 1)
12
13 dados.craibas <- subset(dados, municipio_residencia == "Craíbas")
14
15 #####
```

Como dito na questão anterior, foi realizado um tratamento sobre alguns erros. Após o tratamento, houve uma restrição para pegar os dados do município de Craíbas.

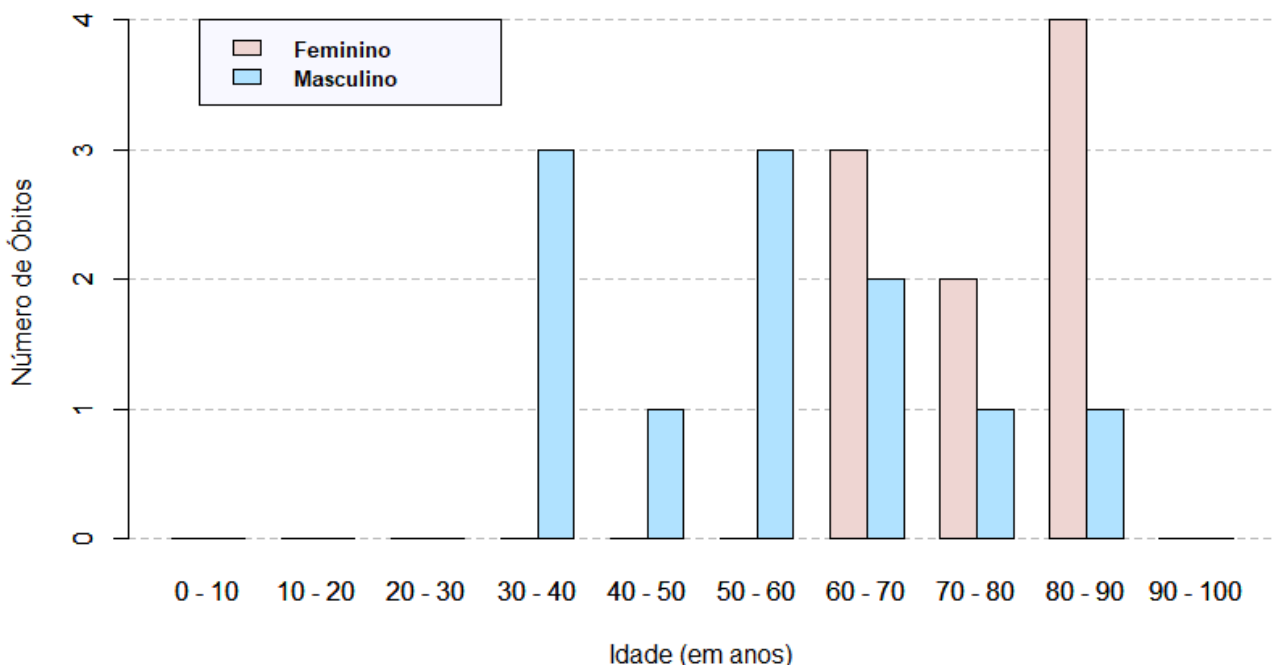
1. Gráfico dos óbitos em Craíbas por sexo e idade

IMPLEMENTAÇÃO

```
17 #GRÁFICO DOS ÓBITOS DE CRAÍBAS POR SEXO E IDADE
18
19 obitos.craibas <- subset(dados.craibas, situacao_atual == "Óbito")
20 obitos.craibas.idade <- cut(obitos.craibas$idade, seq(0, 100, by = 10))
21
22 plot.obitos.sexo <- table(obitos.craibas$sexo, obitos.craibas.idade)
23
24 lim_idade <- c("0 - 10", "10 - 20", "20 - 30", "30 - 40", "40 - 50", "50 - 60",
25               "60 - 70", "70 - 80", "80 - 90", "90 - 100")
26
27 gb_craibas <- barplot(plot.obitos.sexo, beside = TRUE, add = TRUE,
28                       col = c("mistyrose2", "lightskyblue1"), names.arg = lim_idade,
29                       main = "GRÁFICO DE ÓBITOS POR SEXO E IDADE EM CRAÍBAS", legend = TRUE,
30                       args.legend = list(x = 10, y = 4, cex = 0.8, text.font = 2,
31                                           bg = "ghostwhite"),
32                       xlab = "Idade (em anos)", ylab = "Número de Óbitos")
33
```

GRÁFICO

GRÁFICO DE ÓBITOS POR SEXO E IDADE EM CRAÍBAS

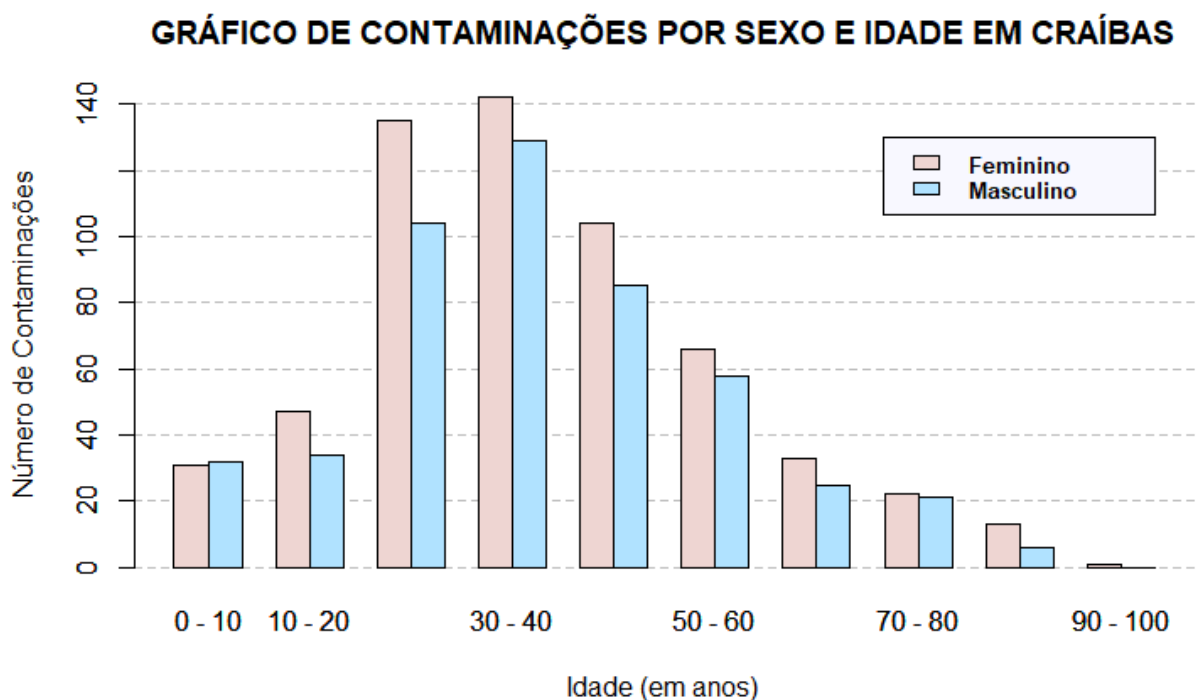


2. Gráfico dos Contaminação em Craíbas por sexo e idade

IMPLEMENTAÇÃO

```
36 #GRÁFICO DE CONTAMINAÇÕES EM CRAÍBAS POR SEXO E IDADE
37
38 dados.craibas.idade <- table(dados.craibas$idade)
39
40 craibas_idade <- cut(dados.craibas$idade, seq(0, 100, by = 10))
41
42 casos.confirmados <- subset(dados.craibas, classificacao == "Confirmado")
43
44 casos.craibas <- table(casos.confirmados$sexo, craibas_idade)
45
46 gb_craibas2 <- barplot(casos.craibas, beside = TRUE, add = TRUE,
47   col = c("mistyrose2", "lightskyblue1"), names.arg = lim_idade,
48   xlab = "Idade (em anos)", ylab = "Número de Contaminações",
49   main = "GRÁFICO DE CONTAMINAÇÕES POR SEXO E IDADE EM CRAÍBAS",
50   legend = TRUE, args.legend = list(x = 30, y = 130, cex = 0.8, text.font = 2,
51   bg = "ghostwhite"))
52
```

GRÁFICO



Ao observarmos os Gráficos 1 e 2, é nítido que a maior taxa de contaminação ocorre em mulheres e homens entre os 30 a 40 anos, já, em comparação, com número de óbitos é relativamente mais baixa.

É possível observar e destacar que na faixa etária de 80 a 90 anos as mulheres possuem a maior taxa de mortalidade e para os homens a taxa mais alta foi na faixa de 30 a 40 anos. Com isso, podemos concluir que mulheres mais velhas possuem a maior taxa de mortalidade.

Todos os códigos mostrados acima estão no seguinte repositório:

[Repositório da Atividade de Probabilidade e Estatística](#)