

1º a)

lê o arquivo csv e armazena na variável enem

```
enem <- read.csv("../data/ENEM-AL-2019.csv", header = TRUE, dec = ".")
```

filtra os dados de maceió e armazena na variável maceio

```
maceio <- subset(enem, enem$NO_MUNICIPIO_RESIDENCIA == "Maceió")
```

cria um vetor com as notas de cada matéria, e a nota geral

```
a_cn <- c(maceio$NU_NOTA_CN)
```

```
a_ch <- c(maceio$NU_NOTA_CH)
```

```
a_lc <- c(maceio$NU_NOTA_LC)
```

```
a_mt <- c(maceio$NU_NOTA_MT)
```

```
a_red <- c(maceio$NU_NOTA_REDACAO)
```

```
a_geral <- c(maceio$NU_NOTA_ENEM)
```

cria um boxplot com as notas de cada matéria e a nota geral

```
boxplot(a_cn, a_ch, a_lc, a_mt, a_red, a_geral,  
        main = "Notas do ENEM - Maceió",  
        names = c("CN", "CH", "LC", "MT", "RED", "GERAL"),  
        col = c(  
            "#fdc5f5",  
            "#f7aef8",  
            "#b388eb",  
            "#8093f1",  
            "#79B8F4",  
            "#72ddf7"  
        )  
)
```

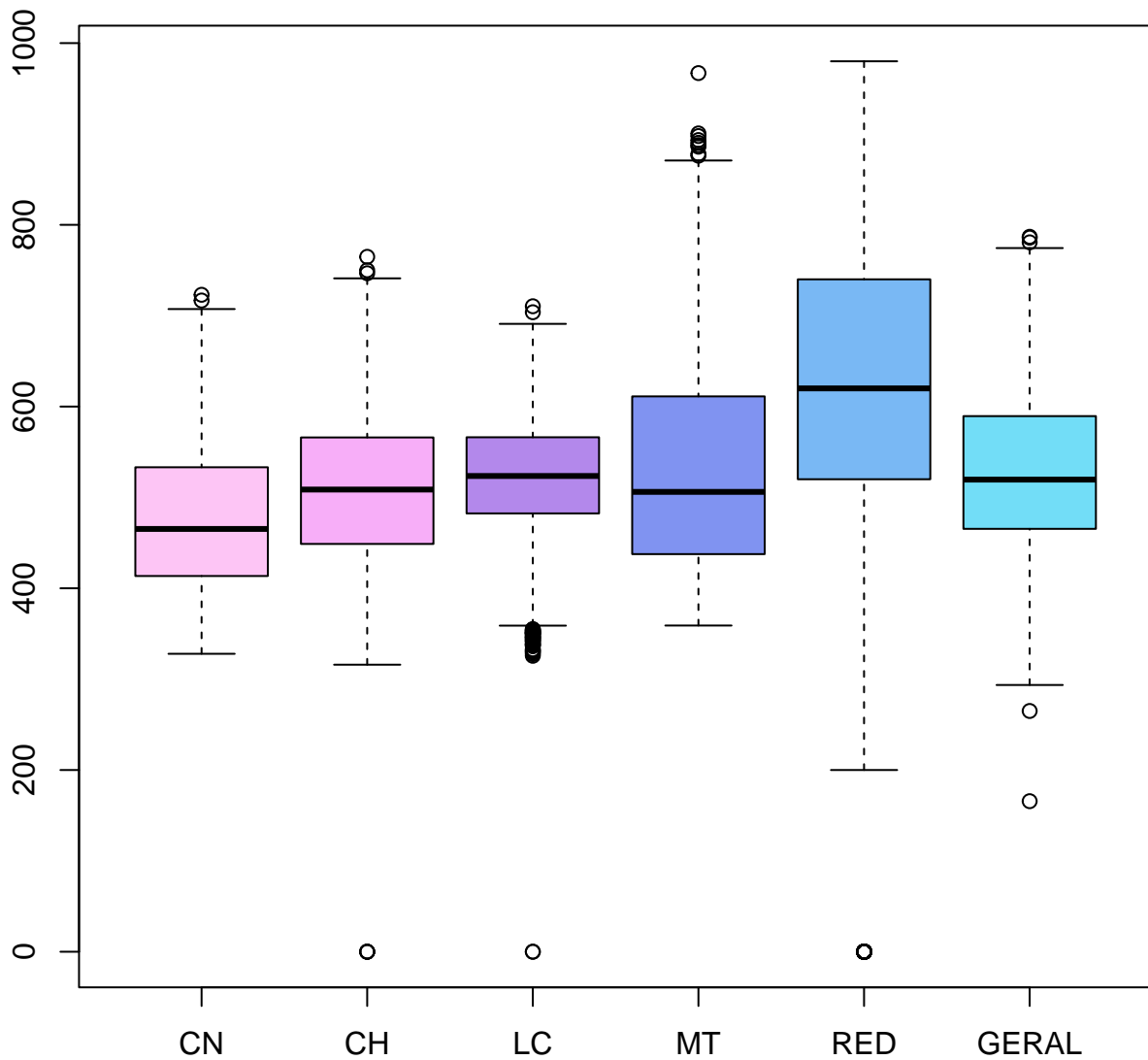
Análise:

Pela forma de como é mostrada pelo boxplot, as notas de matemática e redação

aparentam ser mais fáceis de se alcançar uma nota alta, em comparação com as

notas de linguagens e ch e cn.

Notas do ENEM – Maceió




```

# 1º b)

# lê o arquivo csv e armazena na variável enem
enem <- read.csv("./data/ENEM-AL-2019.csv", header = TRUE, dec = ".")

# filtra os dados de arapiraca e armazena na variável arapiraca
arapiraca <- subset(enem, enem$NO_MUNICIPIO_RESIDENCIA == "Arapiraca")

# cria um vetor com a nota geral do enem
nota_enem <- c(arapiraca$NU_NOTA_ENEM)

# cria um vetor com os nomes dos quartis
classes <- c(
  "297 - 460",
  "460 - 515",
  "515 - 583",
  "583 - 769"
)

# cria uma tabela de frequência
freq <- table(cut(nota_enem, breaks = 4, labels = classes))

# cria uma tabela de frequência relativa
freq_acum <- cumsum(freq)

# calcula a frequência relativa e acumulada
freq_rel <- prop.table(freq)
freq_rel_acum <- cumsum(freq_rel)

# arredonda os valores para 2 casas decimais
freq_rel <- round(freq_rel * 100, 2)
freq_rel_acum <- round(freq_rel_acum * 100, 2)

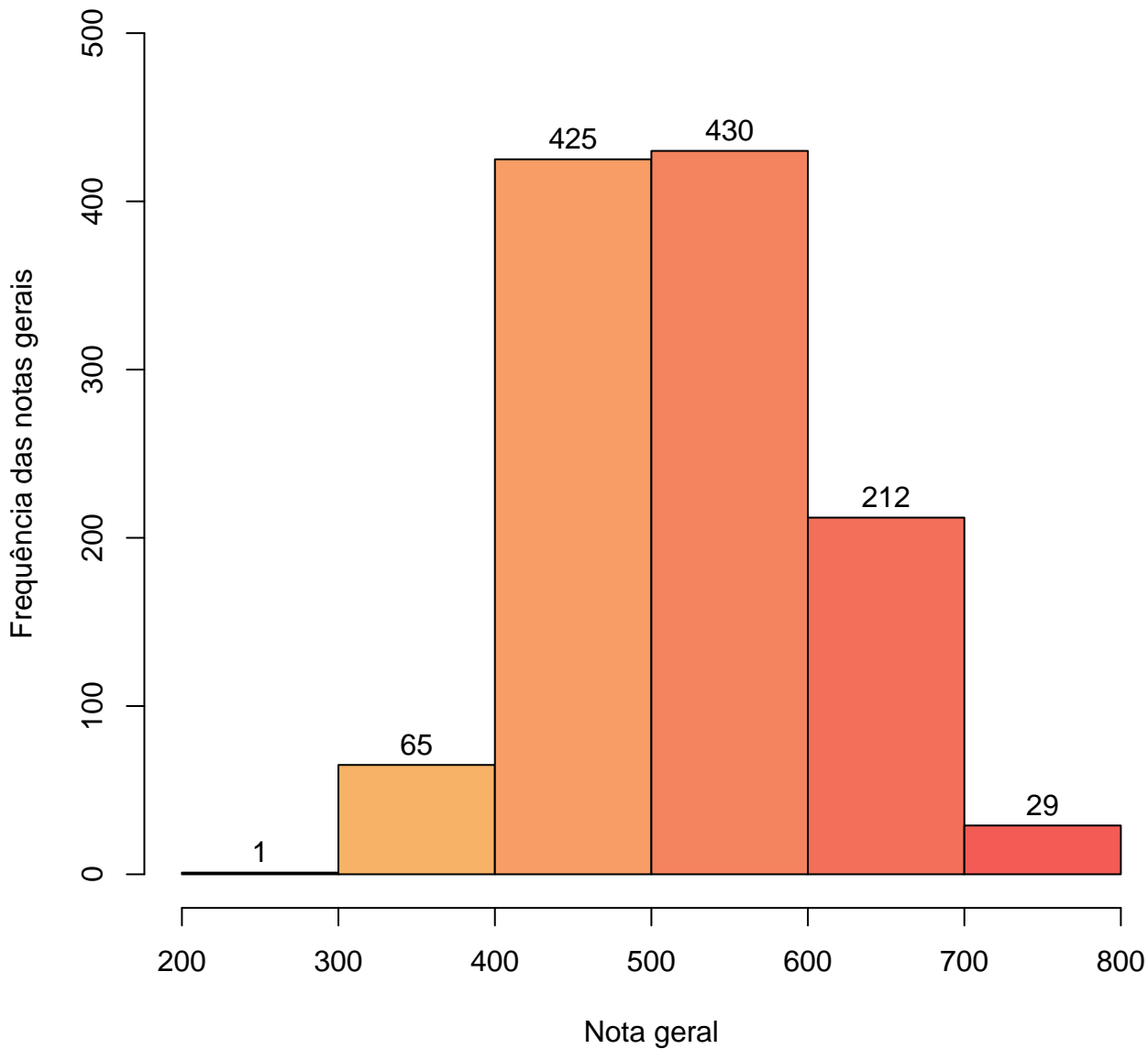
# cria uma tabela com as frequências e frequências relativas
tabela <- cbind(freq, freq_acum, freq_rel, freq_rel_acum)

# cria um histograma com as notas de arapiraca
hist(nota_enem,
  breaks = 4,
  xlim = c(200, 800),
  ylim = c(0, 500),
  main = "Notas do ENEM - Arapiraca",
  xlab = "Nota geral",
  ylab = "Frequência das notas gerais",
  col = c(
    "#ffffff",
    "#f7b267",
    "#f79d65",
    "#f4845f",
    "#f27059",
    "#f25c54"
  ),
  labels = TRUE,
)

# Análise:
# A partir da tabela, é possível perceber que a maior parte dos alunos de
# arapiraca tiraram notas entre 460 e 515, e a menor parte entre as
# extremidades (297 e 769).

```

Notas do ENEM – Arapiraca




```

# 1º c)

# lê o arquivo csv e armazena na variável enem
enem <- read.csv("./data/ENEM-AL-2019.csv", header = TRUE, dec = ".")

# filtra os dados de rio largo e armazena na variável rio_largo
rio_largo <- subset(enem, enem$NO_MUNICIPIO_RESIDENCIA == "Rio Largo")

# cria um vetor com a cor/raça e a escolaridade
cor_raca <- rio_largo$TP_COR_RACA
escolaridade <- rio_largo$TP_ESCOLA

# cria uma tabela de frequência cruzada
dados <- with(rio_largo, table(cor_raca, escolaridade))

# cria um vetor com os nomes das cores
nomes_cor_raca <- c(
  "Não consta",
  "Branca",
  "Preta",
  "Parda",
  "Amarela",
  "Indígena"
)

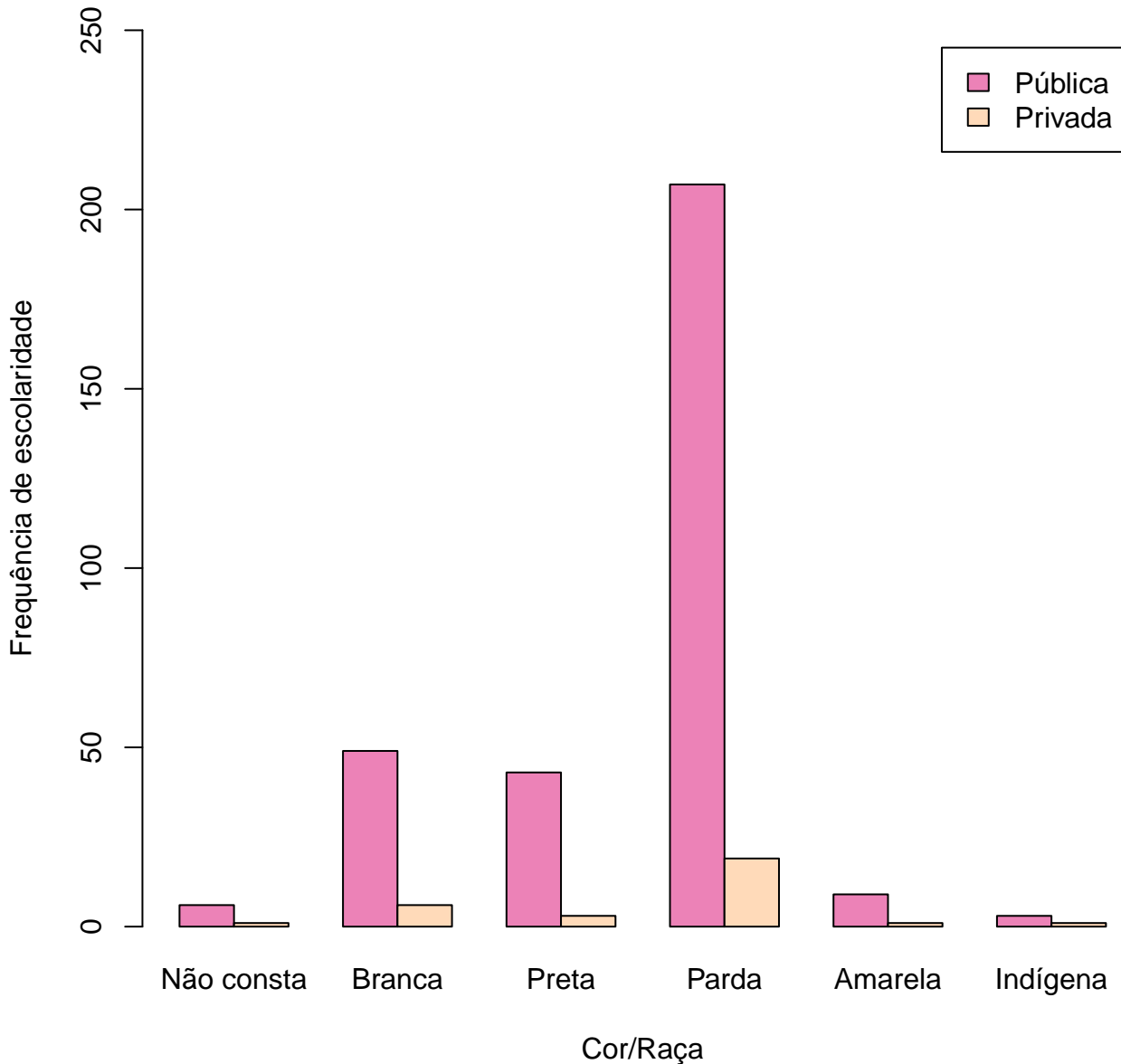
# cria um vetor com os nomes das escolaridades
nomes_escolaridade <- c("Pública", "Privada")

# cria um gráfico de barras com as cores e escolaridades de rio largo
barplot(t(dados),
  main = "Escolaridade por cor/raça - Rio Largo",
  xlab = "Cor/Raça",
  ylab = "Frequência de escolaridade",
  ylim = c(0, 250),
  col = c("#ec82b7", "#FFDAB9"),
  legend = nomes_escolaridade,
  names = nomes_cor_raca,
  beside = TRUE
)

# Análise:
# A partir do gráfico, é possível perceber que a maior parte dos alunos de
# Rio Largo são pardos e que a maior parte são de escola pública.

```

Escolaridade por cor/raça – Rio Largo




```

# 1º d)

# lê o arquivo csv e armazena na variável enem
enem <- read.csv("./data/ENEM-AL-2019.csv", header = TRUE, dec = ".")

# filtra os dados de palmeira e armazena na variável palmeira
palmeira <- subset(enem, enem$NO_MUNICIPIO_RESIDENCIA == "Palmeira dos Índios")

# cria um vetor com a nota de redação e o sexo
nota_redacao <- palmeira$NU_NOTA_REDACAO
sexo <- palmeira$TP_SEX0

# define os intervalos de notas para agrupar as notas em 5 grupos
intervalos <- cut(nota_redacao, breaks = c(0, 500, 600, 700, 800, 1000))

# define os nomes dos grupos de notas
grupos <- c(
  "(0, 500]",
  "(500, 600]",
  "(600, 700]",
  "(700, 800]",
  "(800, 980]"
)

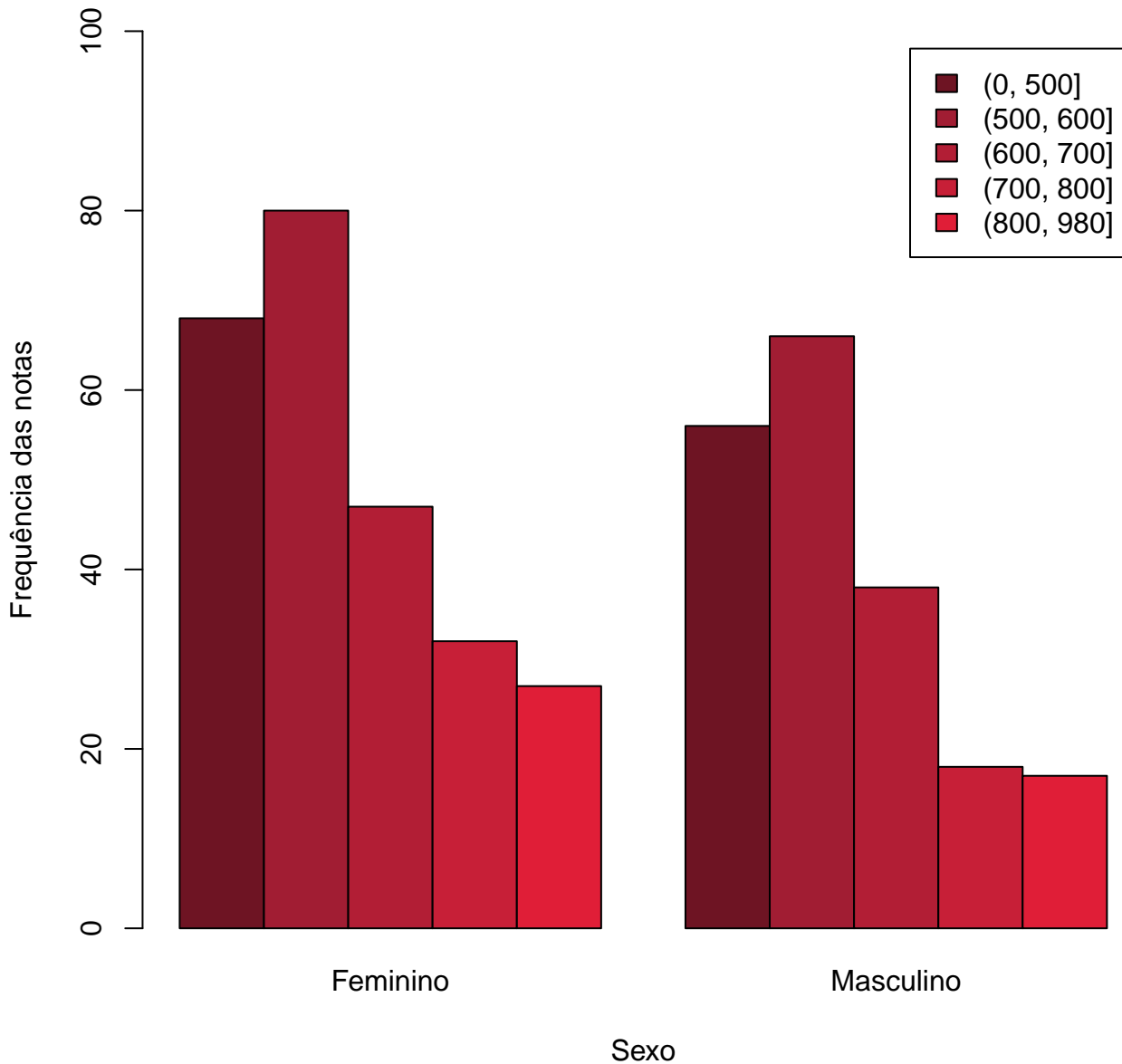
# cria uma tabela que relaciona o grupo de notas com o sexo dos participantes
dados <- table(intervalos, sexo)

# cria um gráfico de barras com os dados da tabela
barplot(dados,
  main = "Notas de redação por sexo - Palmeira dos Índios",
  ylab = "Frequência das notas",
  xlab = "Sexo",
  ylim = c(0, 100),
  col = c(
    "#6e1423",
    "#a11d33",
    "#b21e35",
    "#c71f37",
    "#e01e37"
  ),
  legend = grupos,
  names = c("Feminino", "Masculino"),
  beside = TRUE
)

# Análise:
# A partir do gráfico, é possível perceber que a maior parte das maiores
# notas de redação são de alunas do sexo feminino. Porém a maior parte das
# menores notas também são de alunas do sexo feminino. Portanto, é possível
# perceber que há uma maior concentração de alunas
# em Palmeira dos Índios em geral.
# Entretanto, não tira o fato de que as alunas de Palmeira dos Índios tem uma
# média maior de notas de redação que os alunos, em geral.

```


Notas de redação por sexo – Palmeira dos Índios




```

# 1º e)

# lê o arquivo csv e armazena na variável enem
enem <- read.csv("./data/ENEM-AL-2019.csv", header = TRUE, dec = ".")

# filtra os dados de maceio e armazena na variável maceio
maceio <- subset(enem, enem$NO_MUNICIPIO_RESIDENCIA == "Maceió")

# escolhe as variáveis quantitativas
nota_geral <- maceio$NU_NOTA_ENEM
nota_red <- maceio$NU_NOTA_REDACAO

# agrupa as notas em 4 grupos
nota_geral.cut <- cut(nota_geral, breaks = 4)
nota_red.cut <- cut(nota_red, breaks = 4)

# cria uma tabela de frequência cruzada
table(nota_geral.cut, nota_red.cut)

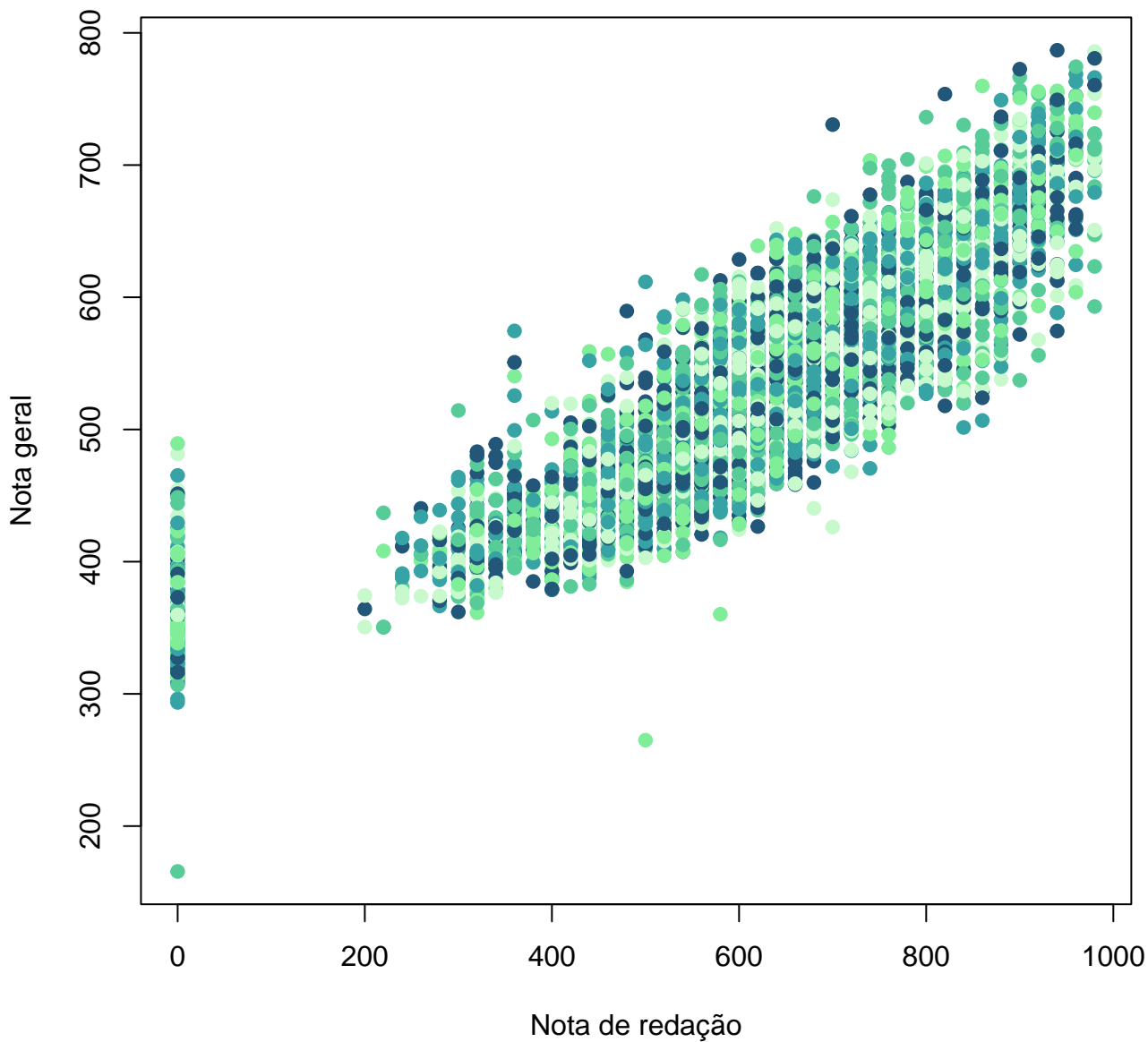
# calcula o coeficiente de correlação
cor(nota_geral, nota_red)
# 0.863, ou seja, há uma correlação positiva forte

# cria um gráfico de dispersão com as notas e as idades
plot(nota_red, nota_geral,
     main = "Notas de redação por nota geral - Maceió",
     xlab = "Nota de redação",
     ylab = "Nota geral",
     col = c(
       "#22577a",
       "#38a3a5",
       "#57cc99",
       "#80ed99",
       "#c7f9cc"
     ),
     pch = 19
)

# Análise:
# A partir do gráfico, é possível perceber que a nota geral e a nota de redação
# são positivamente correlacionadas, ou seja, quanto maior a nota geral,
# maior a nota de redação.

```

Notas de redação por nota geral – Maceió



2º

Agrupar variáveis a sua escolha, como por exemplo notas e idade, ou notas e
números de computadores, ou notas por estado civil ou por sexo, gere os
gráficos que achar adequados e faça uma síntese dos resultados obtidos.

lê o arquivo csv e armazena na variável enem
enem <- read.csv("./data/ENEM-AL-2019.csv", header = TRUE, dec = ".")

filtra os dados de maceio e armazena na variável maceio
maceio <- subset(enem, enem\$NO_MUNICIPIO_RESIDENCIA == "Maceió")

escolhendo as variáveis: nota geral e língua estrangeira
nota_geral <- maceio\$NU_NOTA_ENEM
lingua_estrangeira <- maceio\$TP_LINGUA

table(lingua_estrangeira)

agrupa as notas em 4 grupos
nota_geral.cut <- cut(nota_geral, breaks = 4)

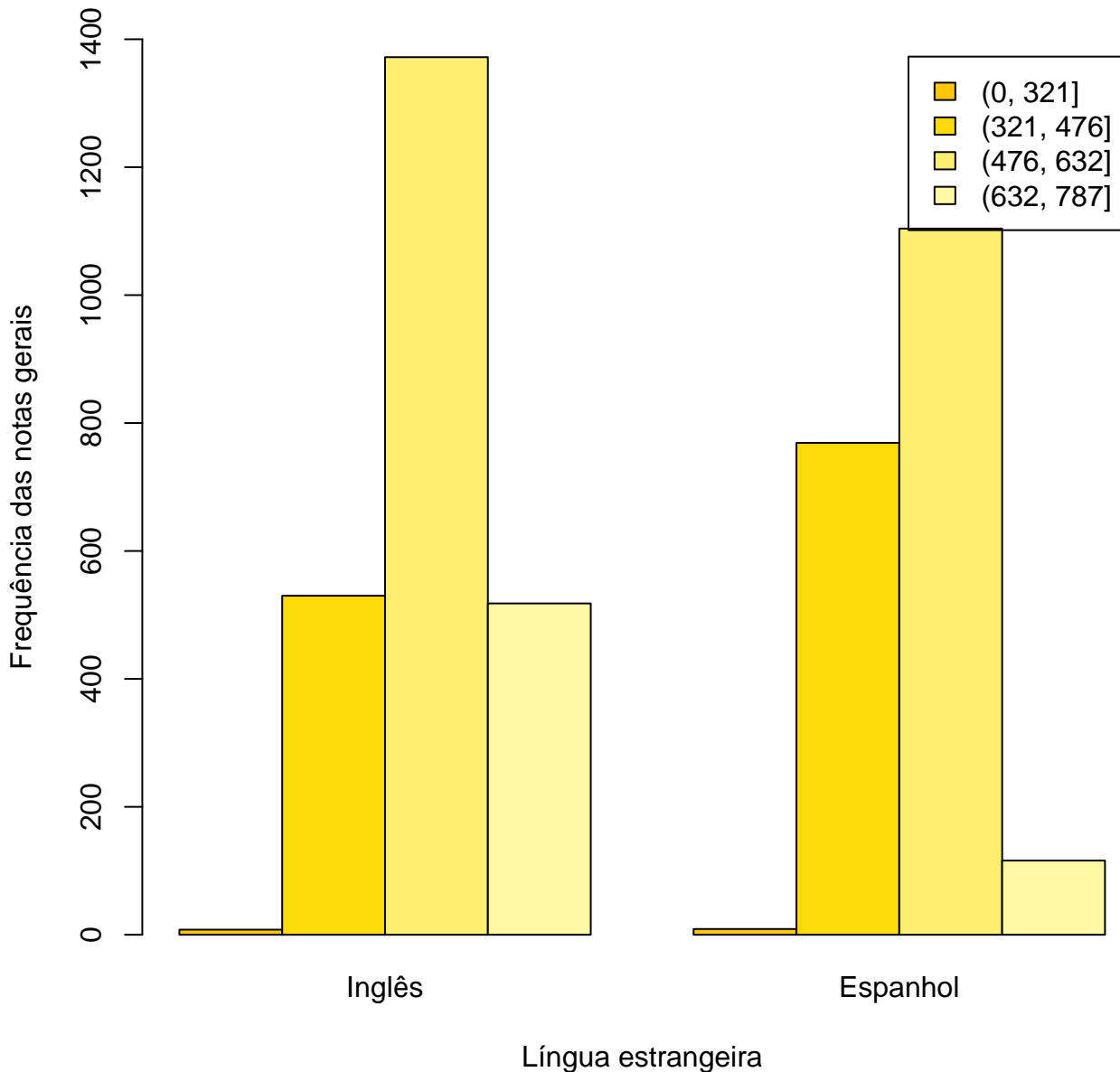
cria uma tabela de frequência cruzada
table(nota_geral.cut, lingua_estrangeira)

cria um gráfico de barras com os dados da tabela
barplot(table(nota_geral.cut, lingua_estrangeira),
 main = "Nota geral por língua estrangeira - Maceió",
 ylab = "Frequência das notas gerais",
 xlab = "Língua estrangeira",
 ylim = c(0, 1400),
 col = c(
 "#ffc60a",
 "#FFDA0A",
 "#FFEE70",
 "#FFF8A5"
),
 names = c("Inglês", "Espanhol"),
 legend = c(
 "(0, 321]",
 "(321, 476]",
 "(476, 632]",
 "(632, 787]"
),
 beside = TRUE
)

Análise:

Apesar do número de alunos que escolheram a língua estrangeira inglesa ser
maior, a nota geral média dos alunos que escolheram o inglês é maior por
uma grande margem. Isso pode ser explicado pelo fato de que o inglês é
uma língua mais utilizada no mundo, e por isso, é mais fácil para os alunos
aprenderem. Além do motivo de que o espanhol, apesar de se parecer com o
português, é uma língua muito diferente e com diversos falso cognatos.

Nota geral por língua estrangeira – Maceió



2º

Verifique se existe alguma correlação entre variáveis quantitativas, como por exemplo, notas de Redação e Matemática, ou Redação com Média final obtida.

lê o arquivo csv e armazena na variável enem

```
enem <- read.csv("./data/ENEM-AL-2019.csv", header = TRUE, dec = ".")
```

filtra os dados de maceio e armazena na variável maceio

```
maceio <- subset(enem, enem$NO_MUNICIPIO_RESIDENCIA == "Maceió")
```

```
nota_mat <- maceio$NU_NOTA_MT
```

```
nota_red <- maceio$NU_NOTA_REDACA0
```

```
nota_mat.cut <- cut(nota_mat, breaks = 4)
```

```
nota_red.cut <- cut(nota_red, breaks = 4)
```

cria uma tabela de frequência cruzada

```
table(nota_mat.cut, nota_red.cut)
```

calcula o coeficiente de correlação

```
cor(nota_mat, nota_red)
```

0.545, ou seja, há uma correlação positiva moderada

Análise:

A partir do coeficiente de correlação, é possível perceber que há uma correlação positiva moderada entre as notas de matemática e de redação.

A correlação acaba não sendo mais forte por conta do fato de que a nota de redação e a nota de matemática não são diretamente relacionadas.

As duas notas são consideradas as mais fáceis de conseguir uma boa nota, e por isso, alguns podem focar mais em uma e deixar a outra de lado.

Como pode ser visto na tabela de frequência cruzada, a nota de redação é maior que a nota de matemática em todos os grupos, e a maioria dos alunos se encontram no grupo de notas onde, em redação, a nota é maior que a nota de matemática.


```

# 2º

# Elabore gráficos que possam ilustrar as Nota Final com o número de
# computadores que o estudante tem em casa (atributo Q024).

# lê o arquivo csv e armazena na variável enem
enem <- read.csv("./data/ENEM-AL-2019.csv", header = TRUE, dec = ".")

# filtra os dados de maceio e armazena na variável maceio
maceio <- subset(enem, enem$NO_MUNICIPIO_RESIDENCIA == "Maceió")

nota_geral <- maceio$NU_NOTA_ENEM
computadores <- maceio$Q024

# agrupa as notas em 4 grupos
nota_geral.cut <- cut(nota_geral, breaks = 4)

# cria um vetor com os nomes das classes de computadores
classes <- c(
  "Zero",
  "Um",
  "Dois",
  "Três",
  "Quatro ou mais"
)

# cria uma tabela de frequência cruzada
table(nota_geral.cut, computadores)

# cria um gráfico de barras com os dados da tabela
barplot(table(nota_geral.cut, computadores),
  main = "Nota geral por número de computadores - Maceió",
  ylab = "Frequência das notas gerais",
  xlab = "Número de computadores",
  ylim = c(0, 1200),
  col = c(
    "#2c2a4a",
    "#4f518c",
    "#907ad6",
    "#dabfff"
  ),
  names = classes,
  legend = c(
    "(0, 321]",
    "(321, 476]",
    "(476, 632]",
    "(632, 787]"
  ),
  beside = TRUE
)

# Análise:
# Apesar do número de alunos que não possuem computadores
# em casa ser superior ao número de alunos
# que possuem mais de um computador, é possível perceber
# que a nota geral média dos alunos que possuem
# computadores em casa é maior por uma grande margem.
# Além disso, a medida em que o número de computadores
# aumenta, o grupo de notas baixas diminui significativamente,
# e o grupo de notas altas sempre aumenta.

```


Nota geral por número de computadores – Maceió

