



HOMEWORK I

NOME COMPLETO: Erick de Oliveira Eugênio

NUMERO DE MATRICULA: 569610

QUESTÃO 1

As emissões diárias de um gás poluente de uma planta industrial foram registradas 80 vezes, em uma determinada unidade de medida. Os dados obtidos estão apresentados na Tabela 1.

15.8	22.7	26.8	19.1	18.5	14.4	8.3	25.9	26.4	9.8	21.9	10.5
17.3	6.2	18.0	22.9	24.6	19.4	12.3	15.9	20.1	17.0	22.3	27.5
23.9	17.5	11.0	20.4	16.2	20.8	20.9	21.4	18.0	24.3	11.8	17.9
18.7	12.8	15.5	19.2	13.9	28.6	19.4	21.6	13.5	24.6	20.0	24.1
9.0	17.6	25.7	20.1	13.2	23.7	10.7	19.0	14.5	18.1	31.8	28.5
22.7	15.2	23.0	29.6	11.2	14.7	20.5	26.6	13.3	18.1	24.8	26.1
7.7	22.5	19.3	19.4	16.7	16.9	23.5	18.4				

Tabela 1: Emissões diárias de gas poluente (questão 1).

1. Calcule as medidas de tendência central (média, mediana e moda) e as medidas de dispersão (amplitude, variância, desvio padrão e coeficiente de variação) para o conjunto de dados da Tabela 1. Interprete os resultados.
2. Construa um histograma e um boxplot para os dados de emissões. Os dados parecem estar simetricamente distribuídos? Existem valores atípicos?
3. Determine os quartis (Q1, Q2, Q3) e o intervalo interquartil (IQR). Utilize esses valores para reforçar sua análise sobre a presença de valores atípicos.
4. Suponha que o limite máximo aceitável diário para as emissões seja de 25 unidades. Qual a proporção de dias em que a planta excedeu esse limite? O comportamento geral das emissões estaria em conformidade com esse padrão regulatório?

SOLUÇÃO DA QUESTÃO 1

1. Para a questão 1.1, utilizei o R para os cálculos. Para as medidas de tendência central, utilizei as seguintes funções oferecidas pelo próprio R: `mean()` e `median()`. No entanto, como o R não oferece uma função própria para a moda, pesquisei um método para resolver. Acessando o site [TutorialsPoint](#), que oferece vários métodos e funções em R para aprendizado, encontrei um método `getmode` que calcula a moda baseada em uma função criada. As respostas para as tendências de medida central foram:

Para as medidas de dispersão, o próprio R fornece o cálculo da variância (`var()`) e do desvio padrão (`sd()`). Para a amplitude, apenas subtraí o maior valor do vetor pelo menor valor utilizando `max()` e `min()`. O cálculo do coeficiente foi feito matematicamente no código.

Os resultados foram dispostos em uma tabela apresentada abaixo:

Medidas - Questão 1.1	
Medida	Valor
Média	19,02
Mediana	19,15
Moda	10.27
Amplitude	25.6
Variância	19.4
Desvio Padrão	5.55
Coeficiente de Variação	29.18

De acordo com a tabela, a média e a mediana tem valores bem parecidos. Isso significa que os dados estão bem simétricos. A moda ser menor que a média e a mediana indica uma leve assimetria à esquerda. A amplitude de 25.6 indica que a variação entre maior valor e o menor é considerável. O valor da variância (e, consequentemente, do desvio padrão), indica que a variação dos dados em relação à média também é considerável, ou seja, os dados estão um pouco dispersos. Por fim, um coeficiente de variação de praticamente 30% indica que os dados estão um pouco mais variados em relação à média, mas nada fora do comum.

2. Com o auxílio do ChatGPT e do [TutorialsPoint](#), consegui achar uma função em R que produz um histograma e um boxplot baseado nos dados fornecidos pela questão. Os gráficos estão na página 3. O boxplot indica que a mediana está levemente abaixo do centro, ou seja, os dados estão assimétricos (não muito) em relação à esquerda. De resto, os dados estão melhor distribuídos. Outliers não são extremos. O histograma apresenta leve assimetria entre os dados de 20 a 25, ou seja, a frequência se encontra levemente positiva.

3. O próprio R fornece uma função para calcular os quartis (`quantile(dados, probs = c(0.25, 0.5, 0.75))`) e o IQR (`IQR()`). Assim, os resultados obtidos foram indicados

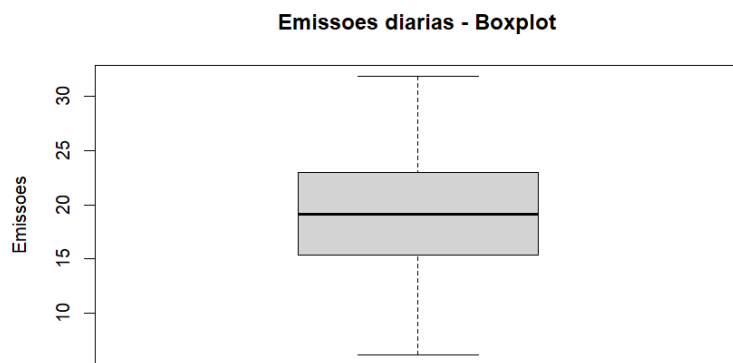


Figura 1: Solução da Questão 1.2 - Boxplot

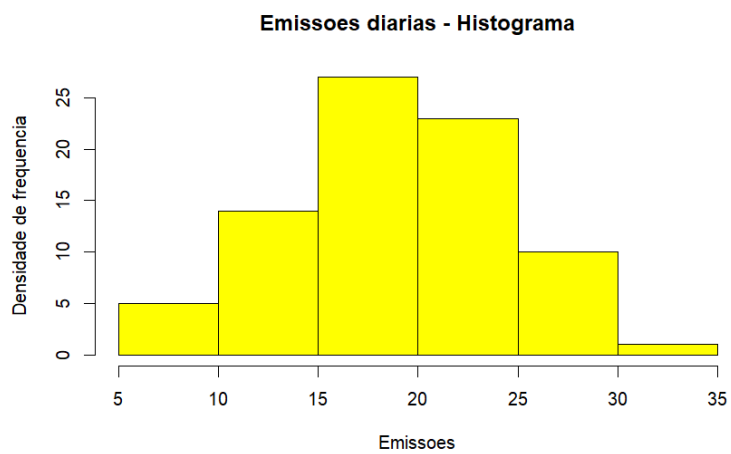


Figura 2: Solução da Questão 1.2 - Boxplot

na tabela. De acordo com os resultados, não existem valores atípicos, pois a distância entre os quartis é muito baixa.

Medidas - Questão 1.3

Quartis	Valor
1°Q	15.43
2°Q	19.15
3°Q	22.92
IQR	7.5

4. Para solucionar essa questão, fiz um código simples em R que conta quantos dias a planta excedeu o limite usando a função `sum()` e depois fazendo a proporção. Segundo a resposta, o excedente de dias em que a planta superou o limite diário foi 11. Calculando $11/80$, a resposta da proporção nos dá 13,75.

Listado 1: Solution of exercise 1

```
# Vetor formado pelos dados fornecidos na questao 1
dados <- c(15.8, 22.7, 26.8, 19.1, 18.5, 14.4, 8.3, 25.9, 26.4, 9.8,
          21.9, 10.5,
          17.3, 6.2, 18.0, 22.9, 24.6, 19.4, 12.3, 15.9, 20.1, 17.0,
          22.3, 27.5,
          23.9, 17.5, 11.0, 20.4, 16.2, 20.8, 20.9, 21.4, 18.0, 24.3,
          11.8, 17.9,
          18.7, 12.8, 15.5, 19.2, 13.9, 28.6, 19.4, 21.6, 13.5, 24.6,
          20.0, 24.1,
          9.0, 17.6, 25.7, 20.1, 13.2, 23.7, 10.7, 19.0, 14.5, 18.1,
          31.8, 28.5,
          22.7, 15.2, 23.0, 29.6, 11.2, 14.7, 20.5, 26.6, 13.3, 18.1,
          24.8, 26.1,
          7.7, 22.5, 19.3, 19.4, 16.7, 16.9, 23.5, 18.4)

# Funcao feita para calcular a moda
getmode <- function(v) {
  uniqv <- unique(v) # Acha os valores unicos do vetor
  uniqv[which.max(tabulate(match(v, uniqv)))] # Acha o valor que mais se
    repete
}

# Calculos (arredondados para duas casas decimais)
media <- round(mean(dados), 2)
mediana <- round(median(dados), 2)
moda <- round(getmode(dados), 2)
amplitude <- round(max(dados) - min(dados), 2)
variancia <- round(var(dados), 2)
dp <- round(sd(dados), 2)
coeficiente <- round((dp / media) * 100, 2)

# Calculo dos quartis
quartis <- quantile(dados, probs = c(0.25, 0.5, 0.75))
```

```

iqr <- IQR(dados)

# Calculo da Questao 4
excedente <- sum(dados > 25) # Calculo do numero de dias que passaram de
    25 unidades
proporcao <- (excedente / 80) * 100 # Calculo da proporcao

# Print dos resultados
cat("Media:", media, "\n")
cat("Mediana:", mediana, "\n")
cat("Moda:", moda, "\n")
cat("Amplitude:", amplitude, "\n")
cat("Variancia:", variancia, "\n")
cat("Desvio padrao:", dp, "\n")
cat("Coeficiente de variacao (%):", coeficiente, "\n")
cat("Calculo dos quartis:", round(quartis,2), "\n")
cat("Calculo da proporcao:", round(proporcao,2), "\n")

# Plotagem do histograma e do boxplot
hist(dados, main="Emissoes diarias - Histograma", xlab= "Emissoes", ylab=
    "Densidade de frequencia", col="yellow", border="black")
boxplot(dados, main="Emissoes diarias - Boxplot", ylab= "Emissoes")

```

QUESTÃO 2

Uma empresa italiana recebeu 20 currículos de cidadãos italianos e estrangeiros na seleção de pessoal qualificado para o cargo de gerente de relações exteriores. A tabela 2 reporta as informações consideradas relevantes na seleção: a idade, a nacionalidade, o nível mínimo de renda desejada (em milhares de euros), os anos de experiência no trabalho.

1. Calcule a média, mediana e desvio padrão para as variáveis idade, renda desejada e anos de experiência. O que você pode inferir a partir desses valores sobre o perfil típico dos candidatos?
2. Agrupe os candidatos por nacionalidade e calcule a renda média desejada e os anos médios de experiência para cada grupo. Qual nacionalidade apresenta a maior renda média desejada? Qual grupo aparenta ser o mais experiente?
3. Existe correlação entre anos de experiência e renda desejada? Utilize ferramentas visuais apropriadas (por exemplo, gráfico de dispersão) e calcule o coeficiente de correlação de Pearson. Interprete o resultado.
4. Suponha que a empresa queira priorizar candidatos com pelo menos 10 anos de experiência e renda desejada inferior a 2,0 (mil euros). Quantos candidatos atendem a ambos os critérios? Liste suas nacionalidades e idades.
5. Construa gráficos que permitam visualizar a distribuição da idade e da renda desejada, separados por nacionalidade. Utilize histogramas, box-plots ou gráficos de barras, e comente as principais diferenças observadas entre os grupos.

	Idade	Nacionalidade	Renda	Experiência
1	28	Italiana	2.3	2
2	34	Inglesa	1.6	8
3	46	Belga	1.2	21
4	26	Espanhola	0.9	1
5	37	Italiana	2.1	15
6	29	Espanhola	1.6	3
7	51	Francesa	1.8	28
8	31	Belga	1.4	5
9	39	Italiana	1.2	13
10	43	Italiana	2.8	20
11	58	Italiana	3.4	32
12	44	Inglesa	2.7	23
13	25	Francesa	1.6	1
14	23	Espanhola	1.2	0
15	52	Italiana	1.1	29
16	42	Alemana	2.5	18
17	48	Francesa	2.0	19
18	33	Italiana	1.7	7
19	38	Alemana	2.1	12
20	46	Italiana	3.2	23

Tabela 2: Informações na seleção da empresa italiana (questão 2).

SOLUÇÃO DA QUESTÃO 2

1. O cálculo da média, da mediana e do desvio padrão foram feitos de acordo com o código no R. O R tem funções próprias para o cálculo dessas medidas de tendência central: a média é calculada por meio da função `mean()`; a mediana é calculada por meio da função `median()`; o desvio padrão é calculado por meio da função `sd()`.

Com esses dados, nota-se que, para as três variáveis, a média é bem semelhante à mediana. Isso significa que a distribuição dos dados de cada variável é simétrica. No entanto, ao analisar o desvio padrão da idade e da experiência, nota-se que o valor é alto (distante de zero). Isso evidencia que, apesar dos valores serem simétricos, eles estão bem dispersos em relação à média, ou seja, temos idades e anos de experiência bem variados.

Tabela – Solução da questão 2.1

Variável	Média	Mediana	Desvio Padrão
Idade	39.65	40.5	10.12
Renda	1.92	1.75	0.71
Experiência	14.00	14.0	10.27

2. Esse cálculo também foi feito utilizando os códigos no R. Ao inserir os dados da tabela original, utilizei o método `data.frame` para agrupar os dados em uma tabela dentro do R. Após isso, utilizei a função `aggregate` para associar apenas os dados das colunas de nacionalidade e de renda e, depois, os dados das colunas de nacionalidade e de idade. Os resultados foram:

- Nacionalidade com maior renda desejada: alemã \rightarrow 2,3 mil euros
- Nacionalidade com mais anos de experiência: alemã \rightarrow 15 anos

Isso mostra que a nacionalidade alemã se destacou nos dois espectros observados. Existem apenas dois alemães presentes, e eles são bem experientes e desejam uma renda alta.

3. Utilizei um gráfico de dispersão produzido no R. Como não sabia fazer, pesquisei no site TutorialsPoint novamente e, com a ajuda de uma IA, adaptei o código para as especificações devidas.

De acordo com o gráfico, a linha de regressão linear indica que, com o aumento dos anos de experiência, a renda desejada também aumenta. Isso significa que pessoas com mais anos de atuação tendem a desejar receber mais. No entanto, nota-se que existem pontos que não se aproximam tanto da linha de regressão. Isso indica que, apesar da tendência de crescimento na análise dos dois dados, não é uma regra que pessoas mais experientes desejem receber uma renda maior (e vice-versa). Logo, os dados têm uma tendência, mas se dispersam em certos pontos. O cálculo do coeficiente de Pearson foi feito usando uma função própria do R para esse cálculo e retornou o valor 0,5. Isso mostra que os valores tem relação, mas não é uma relação tão forte (o que é mostrado pelos gráficos e pelos dados apresentados).

4. Essa questão também foi resolvida com o código em R. Usando a função `subset`, que seleciona subconjuntos da tabela criada pelo `data.frame` para filtrar dados específicos, manuseei o código para procurar as pessoas com anos de experiência maior ou igual a 10 e renda desejada inferior a 2,0 mil euros. Assim, a resposta obtida foi inserida na tabela:

Tabela – Solução da questão 2.4

Candidato	Idade	Nacionalidade	Renda	Experiência
3	46	Belga	1.2	21
7	51	Francesa	1.8	28
9	39	Italiana	1.2	13
15	52	Italiana	1.1	29

5. Com a ajuda do TutorialPoint e do ChatGPT, pude formular gráficos por meio de boxplots e do pacote `ggplot2`. Os gráficos obtidos estão dispostos neste documento. Dito isso, o gráfico de idade por nacionalidade mostra que os belgas são mais velhos (na média) e que os espanhóis são mais novos. O outro gráfico mostra que os alemães

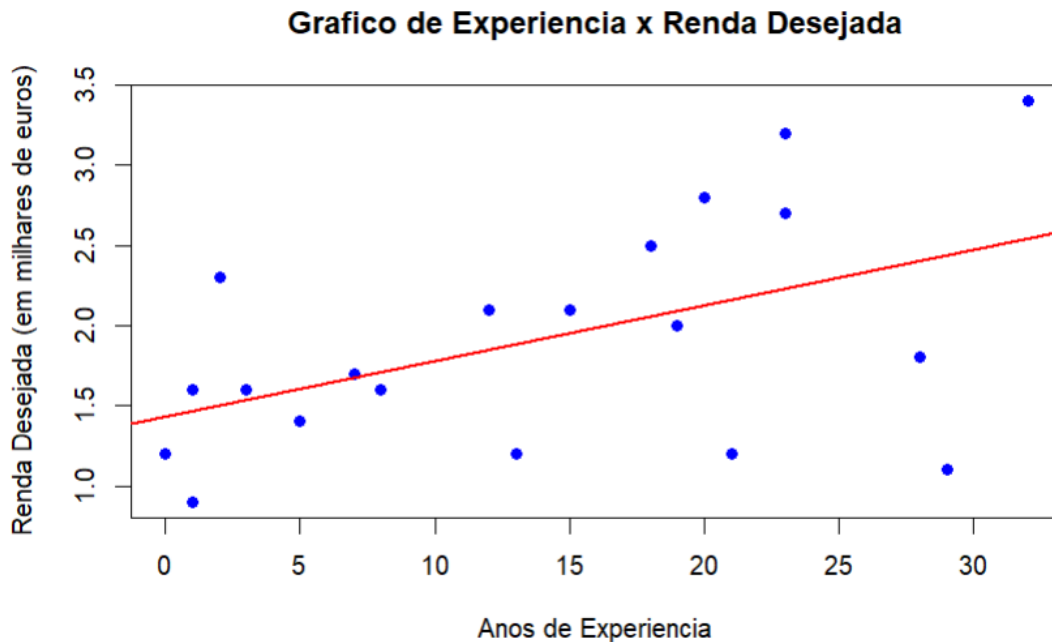


Figura 3: Gráfico - Solução da Questão 2.3

esperam receber mais e os espanhóis esperam receber menos. Isso está em concordância com os dados apresentados nas respostas anteriores. Quem é mais velho espera receber mais pelos fatores de idade e de experiência, mas isso não é uma regra (olhe a relação entre os belgas e o que eles esperam receber, por exemplo).

Listado 2: Solution of exercise 2

```
library(ggplot2)
library(dplyr)

# Criação da tabela utilizando o método data.frame
Idade <- c(28, 34, 46, 26, 37,
          29, 51, 51, 39, 43,
          58, 44, 25, 23, 52,
          42, 48, 33, 38, 46)

Nacionalidade <- c("Italiana", "Inglesa", "Belga", "Espanhola", "Italiana",
                  "Espanhola", "Francesa", "Belga", "Italiana", "Italiana",
                  "Italiana", "Inglesa", "Francesa", "Espanhola", "Italiana",
                  "Alemana", "Francesa", "Italiana", "Alemana", "Italiana")

Renda <- c(2.3, 1.6, 1.2, 0.9, 2.1,
```

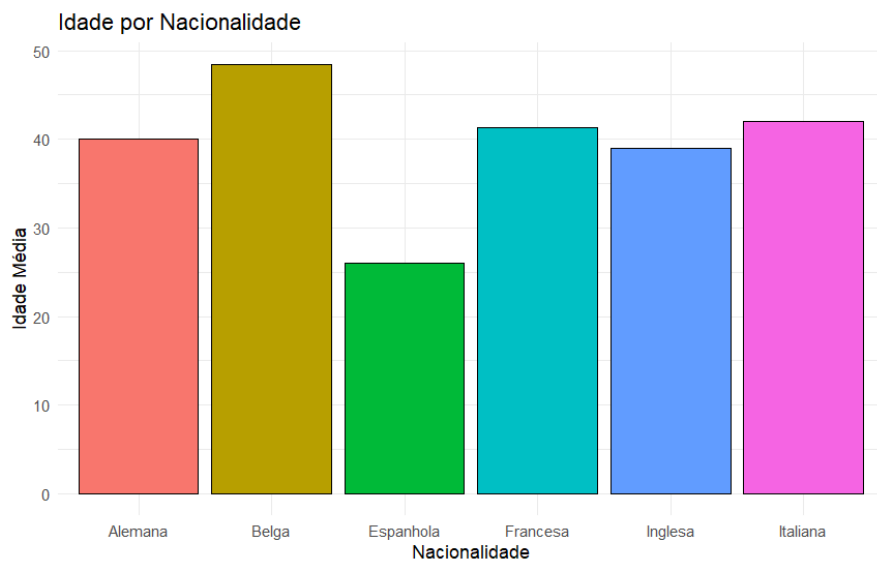



Figura 4: Gráfico de idade média por nacionalidade - Solução da questão 2.5

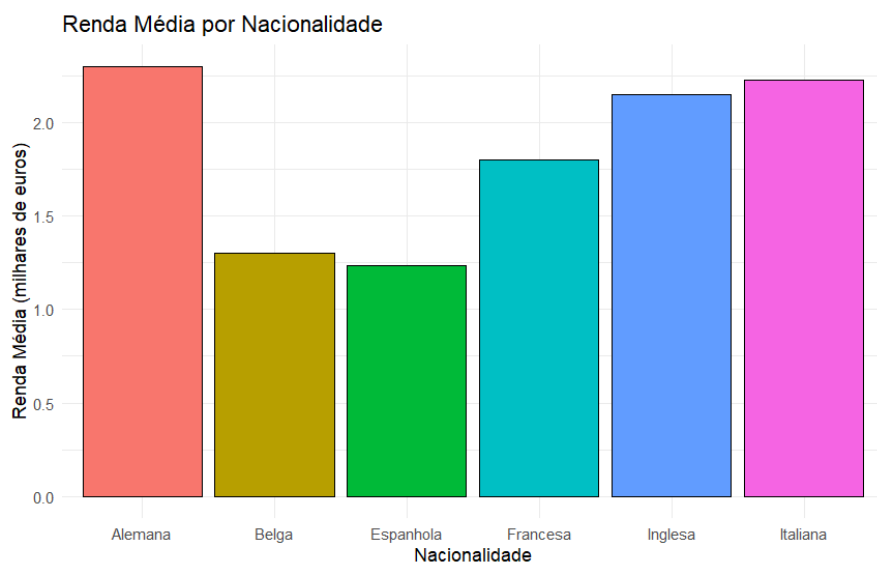


Figura 5: Gráfico de renda média por nacionalidade - Solução da questão 2.5

```

1.6, 1.8, 1.4, 1.2, 2.8,
3.4, 2.7, 1.6, 1.2, 1.1,
2.5, 2.0, 1.7, 2.1, 3.2)

Experiencia <- c(2, 8, 21, 1, 15,
                 3, 28, 5, 13, 20,
                 32, 23, 1, 0, 29,

```

```

18, 19, 7, 12, 23)

Dados <- data.frame(Idade, Nacionalidade, Renda, Experiencia)

# Calculo das medidas de tendencia central da idade
media_idade <- mean(Idade)
mediana_idade <- median(Idade)
dp_idade <- sd(Idade)

# Calculo das medidas de tendencia central da renda
media_renda <- mean(Renda)
mediana_renda <- median(Renda)
dp_renda <- sd(Renda)

# Calculo das medidas de tendencia central dos anos de experiencia
media_exp <- mean(Experiencia)
mediana_exp <- median(Experiencia)
dp_exp <- sd(Experiencia)

# Uso da funcao aggregate para calcular e dispor a maior renda media
# desejada por nacionalidade
renda_por_nacionalidade <- aggregate (Renda ~ Nacionalidade, data = Dados
, FUN = mean)
maior_renda <- renda_por_nacionalidade[which.max(renda_por_nacionalidade$
Renda), ]

# Uso da funcao aggregate para calcular e dispor a maior media de anos de
# experiencia desejada por nacionalidade
exp_por_nacionalidade <- aggregate (Experiencia ~ Nacionalidade, data =
Dados, FUN = mean)
maior_exp <- exp_por_nacionalidade[which.max(renda_por_nacionalidade$
Renda), ]

# Uso da funcao subset para encontrar os candidatos com pelo menos 10
# anos de experiencia e renda desejada menor do que 2,0
candidatos <- subset(Dados, Experiencia>=10 & Renda < 2.0)

# Calculo do coeficiente de Pearson
pearson <- cor(Dados$Renda, Dados$Experiencia, method = "pearson")

# Print dos resultados
cat("Media das idades:", round(media_idade,2), "\n")
cat("Mediana das idades:", round(mediana_idade,2), "\n")
cat("Desvio padrao das idades:", round(dp_idade,2), "\n")
cat("Media das rendas desejadas:", round(media_renda,2), "\n")
cat("Mediana das rendas desejadas:", round(mediana_renda,2), "\n")
cat("Desvio padrao das rendas desejadas:", round(dp_renda,2), "\n")
cat("Media dos anos de experiencias:", round(media_exp,2), "\n")
cat("Mediana dos anos de experiencias:", round(mediana_exp,2), "\n")
cat("Desvio padrao dos anos de experiencias:", round(dp_exp,2), "\n")

maior_renda
maior_exp

```

```

cat("Coeficiente de Pearson: ", round(pearson,2), "\n")

candidatos

# Plotagem do grafico de dispersao utilizando o metodo plot - Questao 2.3
plot(Dados$Experiencia, Dados$Renda,
      xlab = "Anos de Experiencia",
      ylab = "Renda Desejada (em milhares de euros)",
      main = "Grafico de Experiencia x Renda Desejada",
      pch = 19, col = "blue"
)

# Adicao da linha de regressao linear ao grafico utilizando o metodo
  proprio lm
modelo <- lm(Renda ~ Experiencia, data = Dados)
abline(modelo, col = "red", lwd = 2)

# Agrupando por nacionalidade e calculando a media de idade
idade_por_nacionalidade <- Dados %>%
  group_by(Nacionalidade) %>%
  summarise(IdadeMedia = mean(Idade))

# Grafico de idade por nacionalidade - Q5
ggplot(idade_por_nacionalidade, aes(x = Nacionalidade, y = IdadeMedia,
  fill = Nacionalidade)) +
  geom_bar(stat = "identity", color = "black") +
  labs(title = "Idade Media por Nacionalidade",
    x = "Nacionalidade",
    y = "Idade Media") +
  theme_minimal() +
  theme(legend.position = "none")

# Agrupando por nacionalidade e calculando a media de renda
renda_por_nacionalidade <- Dados %>%
  group_by(Nacionalidade) %>%
  summarise(RendaMedia = mean(Renda))

# Grafico de renda por nacionalidade - Q5
ggplot(renda_por_nacionalidade, aes(x = Nacionalidade, y = RendaMedia,
  fill = Nacionalidade)) +
  geom_bar(stat = "identity", color = "black") +
  labs(title = "Renda Media por Nacionalidade",
    x = "Nacionalidade",
    y = "Renda Media (milhares de euros)") +
  theme_minimal() +
  theme(legend.position = "none")

```

QUESTÃO 3

O conjunto de dados em anexo, `HW1_bike_sharing.csv`¹, refere-se ao processo de compartilhamento de bicicletas em uma cidade dos Estados Unidos. O conjunto contém as colunas descritas na Tabela 3. A variável `season` inclui as quatro estações do hemisfério norte: primavera, verão, outono e inverno. A variável `weathersit` representa quatro condições meteorológicas: ‘Céu limpo’, ‘Nublado’, ‘Chuva fraca’, ‘Chuva forte’. A variável `temp` é a temperatura normalizada em graus Celsius, ou seja, os valores foram divididos por 41 (valor máximo).

TAG	DESCRIÇÃO
<code>instant</code>	Índice de registro
<code>dteday</code>	Data da observação
<code>season</code>	Estação do ano
<code>weathersit</code>	Condições meteorológicas
<code>temp</code>	Temperatura em °C (normalizada)
<code>casual</code>	Número de usuários casuais
<code>registered</code>	Número de usuários registrados

Tabela 3: Variáveis do conjunto `HW1_bike_sharing` (questão 3).

1. Carregue o conjunto de dados `HW1_bike_sharing.csv` no R. Classifique as variáveis quanto ao tipo (categórica ou numérica), identifique o número total de observações e as datas de início e fim da amostra.
2. Calcule medidas de tendência central (média, mediana) e os quartis para cada característica numérica relevante. Apresente os resultados em uma tabela com título apropriado. Comente os principais pontos.
3. Atribua os níveis correspondentes às variáveis `season` e `weathersit`. Construa gráficos de barras para ambas. Qual estação do ano apresenta maior número de usuários? O uso de bicicletas depende da estação? Qual é a condição climática mais favorável para o uso do sistema?
4. Calcule o número total de usuários por dia, somando `casual` e `registered`. Converta a variável `temp` para temperatura real (multiplicando por 41). Em seguida, construa os gráficos de séries temporais para temperatura e número total de usuários. Essas séries apresentam tendência semelhante?

SOLUÇÃO DA QUESTÃO 3

1. Com ajuda do site TutorialsPoint, do ChatGPT e dos códigos fornecidos pela professora Michela, consegui carregar o conjunto de dados para realizar a questão. Após isso,

¹ Os dados estão disponíveis no material do homework.

elaborei um código para classificar as variáveis, contar o número total de observações e as datas de início e fim da amostra. Os resultados estão nas tabelas abaixo:

Tabela – Solução da questão 3.1 - Tipo de variável

Variável	Tipo
instant	Numérica
dteday	Data
season	Numérica
weathersit	Numérica
temp	Numérica
casual	Numérica
registered	Numérica

Tabela – Solução da questão 3.1 - Número de observações e datas

Medida	Valor
Número de observações	731
Data de início	01/01/2011
Data de término	31/12/2012

2. As estatísticas numéricas mais relevantes são **temp**, **casual** e **registered**. O cálculo da média, da mediana e dos quartis está disposto em uma tabela deste documento. De início, analisando as medidas de **temp**, a média e a mediana estão bem parecidas, ou seja, os dados estão simetricamente distribuídos (como deve ser a tabela de uma média de temperatura anual). Os quartis também se encontram igualmente espaçados em relação à mediana. Em **casual**, a média é bem maior do que a mediana, ou seja, os dados estão bem assimétricos. Em **registered**, acontece o mesmo que em **temp**.

Tabela – Solução da questão 3.1 - Média, mediana e quartis para temp, casual, registered

Variável	Média	Mediana	1° Quartil	3° Quartil
temp	20.31	20.4	13.8	26.9
casual	848.18	713.0	315.5	1096.0
registered	3656.17	3662.0	2497.0	4776.5

3. Fiz essa questão utilizando gráficos para o R. Coloquei níveis para as variáveis correspondentes. O código está apresentado ao final. A estação que apresenta maior número de usuários é o verão. E sim, o uso de bicicletas depende da estação. No verão e primavera (climas melhores), o uso de bicicletas é maior. A condição climática mais favorável foi a que eu classifiquei como "tempo bom".

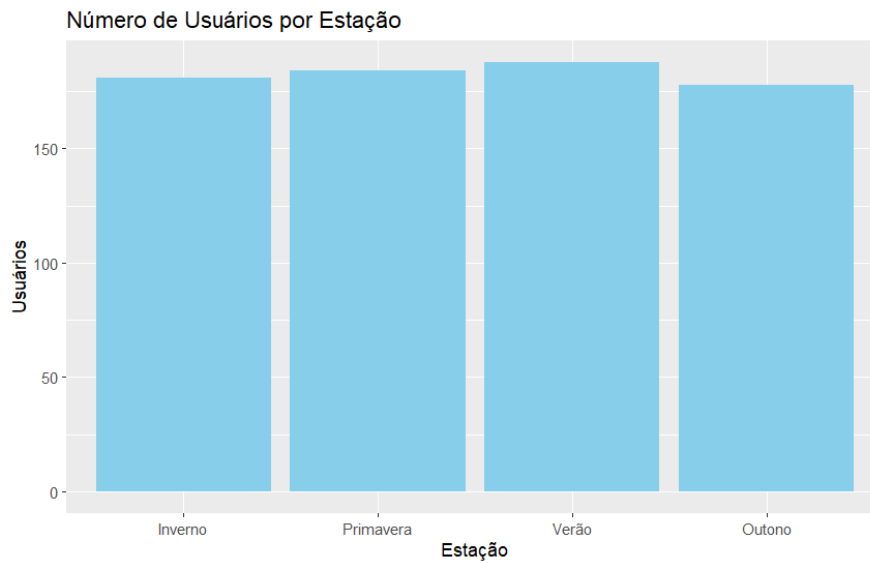


Figura 6: Usuários por estação - Questão 3.3

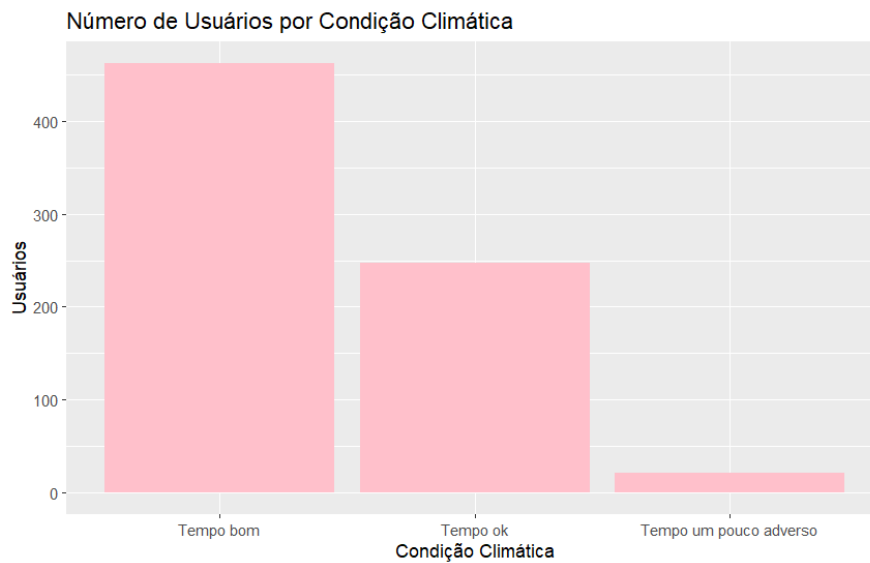


Figura 7: Usuários por condição - Questão 3.3

4. Com o código em R, calculei o número total de usuários e converti a variável `temp`. Depois disso, construí os gráficos. De acordo com o código, o número total de usuários foi de 3.292,679. E, de acordo com os gráficos construídos, há, sim, uma tendência de que as pessoas utilizem mais o sistema de bicicletas com um aumento da temperatura, ou seja, gráficos bem parecidos.

Listado 3: Solution of exercise 3

```
library(readr)
library(dplyr)
library(ggplot2)

rm(list = ls())
graphics.off()
setwd("C:/Users/Erick/Documents") # parte do codigo tirada do documento
enviado
getwd()
list.files()

dados <- read_csv("HW1_bike_sharing.csv", show_col_types = FALSE)
head(dados)
View(dados)

str(dados) # Mostra o tipo da variavel (ao lado do nome da variavel)
numero_obs <- nrow(dados) # Mostra o numero de observacoes totais
primeirodia <- min(dados$dteday) # Mostra a data do primeiro dia
observado
ultimodia <- max(dados$dteday) # Mostra a data do ultimo dia observado

# Selecao das estatisticas numericas relevantes
numeric_vars <- dados %>% select(temp, casual, registered)

# Coloquei as variaveis em uma tabela com as medidas, com seus devidos
calculos
tabelatendencias <- data.frame(
  Variavel = names(numeric_vars),
  Media = round(sapply(numeric_vars, mean), 2),
  Mediana = round(sapply(numeric_vars, median), 2),
  Q1 = round(sapply(numeric_vars, function(x) quantile(x, 0.25)), 2),
  Q3 = round(sapply(numeric_vars, function(x) quantile(x, 0.75)), 2)
)

# Atribuicao de niveis as variaveis season e weathersit
dados$season <- factor(dados$season, levels=c(1,2,3,4),
  labels=c("Inverno", "Primavera", "Verao", "Outono")
)
dados$weathersit <- factor(dados$weathersit, levels=c(1,2,3,4),
  labels=c("Tempo_bom", "Tempo_ok", "Tempo_um",
    pouco_adverso", "Tempo_ruim"))

# Construcão do grafico de barras para season
ggplot(dados, aes(x=season)) +
  geom_bar(fill="skyblue") + labs(title="Numero_de_Usuarios_por_Estacao",
  x="Estacao", y="Usuarios")

# Construcão do grafico de barras para weathersit
ggplot(dados, aes(x=weathersit)) +
  geom_bar(fill="pink") +
  labs(title="Numero_de_Usuarios_por_Condicao_Climatica", x="Condicao_
    Climatica", y="Usuarios")
```

```

# Calculo do total de usuarios e da temperatura real de acordo com o
  proprio enunciado
dados <- dados %>%
  mutate(total_usuarios = casual + registered,
         temp_real = temp * 41)

total_usuarios = sum(dados$casual + dados$registered)

# Print dos resultados
cat("Numero de observacoes:", round(numero_obs,2), "\n")
print(primeirodia)
print(ultimodia)
cat("Numero total de usuarios:", total_usuarios, "\n")
print(tabelatendencias)

# Grafico de series temporais: temperatura
ggplot(dados, aes(x=dteday, y=temp_real)) +
  geom_line(color="red") +
  labs(title="Temperatura Real ao Longo do Tempo", x="Data", y="
    Temperatura")

# Grafico de series temporais: total de usuarios
ggplot(dados, aes(x=dteday, y=total_usuarios)) +
  geom_line(color="blue") +
  labs(title="Numero Total de Usuarios ao Longo do Tempo", x="Data", y="
    Total de Usuarios")

```