



Prova 01

1. Para esse exercício você deverá utilizar as duas matrizes fornecidas. [🔗](#)

- Considere a matriz A dada por:

A	B	C	D	E
28	32	8	9	49
7	21	35	28	10
47	43	15	34	2
48	42	19	32	26
45	44	39	50	26

- e a matriz B dada por:

A	B	C	D	E
0	26	3	8	30
35	12	19	27	27
27	24	12	17	29
31	36	40	35	8
24	43	31	21	39

Considere:

$$C = (A * B^T)^{-1}$$

e

$$P = B * (B^T * B) * B^T$$

a) Considere a matriz de projeção P . A soma de seus autovetores é dada por?

```
# Defina as matrizes A e B
A <- matrix(c(28, 32, 8, 9, 49,
              7, 21, 35, 28, 10,
              47, 43, 15, 34, 2,
              48, 42, 19, 32, 26,
              45, 44, 39, 50, 26),
            nrow = 5,
            byrow = TRUE)

B <- matrix(c(0, 26, 3, 8, 30,
              35, 12, 19, 27, 27,
              27, 24, 12, 17, 29,
              31, 36, 40, 35, 8,
              24, 43, 31, 21, 39),
            nrow = 5,
            byrow = TRUE)

# Calcule os autovetores para cada matriz
eigen_A <- eigen(A)$vectors
eigen_B <- eigen(B)$vectors

# Some os autovetores
soma_autovetores <- eigen_A + eigen_B

# Calculando P
P <- B %*% (t(B) %*% B) %*% t(B)

# Calculando os autovetores de P
eigen_P <- eigen(P)$vectors

# Calculando a soma dos autovetores de P
soma_autovetores_P <- colSums(eigen_P)

# Exibindo a soma dos autovetores de P
print(soma_autovetores_P)
```

```
[1] -2.1633934 -0.3386760 0.1557935 0.4236154 -0.0361385
```

```
# Arredondando para 4 casas decimais para facilitar a comparação
soma_autovetores_P_arredondada <- round(sum(soma_autovetores_P), 4)

soma_autovetores_P_arredondada
```

```
[1] -1.9588
```

b. A soma dos valores absolutos da diagonal da matriz C é 0,0722.

```
# Calculando C
C <- solve(A %*% t(B))

# Calculando a soma dos valores absolutos da diagonal de C
soma_diagonal_C <- sum(abs(diag(C)))
```

```
# Exibindo a soma  
soma_diagonal_C
```

```
[1] 0.07224898
```

c. A soma de uma matriz triangular inferior para a matriz A é 233.

```
# Definindo a matriz A  
A <- matrix(c(28, 32, 8, 9, 49,  
              7, 21, 35, 28, 10,  
              47, 43, 15, 34, 2,  
              48, 42, 19, 32, 26,  
              45, 44, 39, 50, 26),  
            nrow = 5,  
            byrow = TRUE)  
  
# Calculando a soma dos elementos da matriz triangular inferior  
soma_triangular_inferior <- sum(A[row(A) > col(A)])  
  
# Exibindo a soma  
soma_triangular_inferior
```

```
[1] 384
```

d. O \log_{10} do valor absoluto do determinante de A é 6,335. O \log_{10} do valor absoluto do determinante de B é 6,7168. O \log_{10} do valor absoluto do determinante da matriz resultante do produto matricial entre A e B é 13,0518.

```
# Calculando o log10 do valor absoluto do determinante de A  
log_det_A <- log10(abs(det(A)))  
  
# Exibindo o resultado  
print(log_det_A)
```

```
[1] 6.335031
```

```
# Calculando o log10 do valor absoluto do determinante de B  
log_det_B <- log10(abs(det(B)))  
  
# Exibindo o resultado  
print(log_det_B)
```

```
[1] 6.716753
```

```
# Calculando o determinante da matriz resultante do produto matricial entre A e B  
det_AB <- det(A %*% B)  
  
# Calculando o log10 do valor absoluto do determinante de AB  
log_det_AB <- log10(abs(det_AB))
```

```
# Exibindo o resultado  
print(log_det_AB)
```

[1] 13.05178

e. O maior elemento da diagonal do inverso da matriz resultante do produto matricial entre A e o transposto de B é 0,026.

```
# Calculando o produto matricial entre A e o transposto de B  
ABt <- A %*% t(B)  
  
# Calculando o inverso da matriz resultante  
inv_ABt <- solve(ABt)  
  
# Obtendo a diagonal da matriz inversa  
diagonal_inv_ABt <- diag(inv_ABt)  
  
# Encontrando o maior elemento da diagonal  
maior_elemento_diagonal <- max(diagonal_inv_ABt)  
  
# Exibindo o resultado  
maior_elemento_diagonal
```

[1] 0.01596027

2. Para esse exercício você deverá utilizar o banco de dados **chocolate.csv.gz**.

O dicionário das variáveis encontra-se disponível abaixo. |Variável| Descrição| |---| |local_compania| Região do Fabricante| |ano| Ano da Revisão| |origem_cacau| País de Origem dos Grãos de Cacau| |cocoa_percent| Percentagem de Cacau (% chocolate)| |ingredientes| Ingredientes| |caracteristicas| Características mais memoráveis daquele chocolate|

Lista de ingredientes: "#": representa o número de ingredientes no chocolate; B: Grãos, S: Açúcar, S*: Adoçante diferente de açúcar de cana branco ou beterraba, C: Manteiga de Cacau, V: Baunilha, L: Lecitina, Sa: Sal

Assinale todas as alternativas incorreta(s).

```
require(data.table)
```

Carregando pacotes exigidos: data.table

```
require(dplyr)
```

Carregando pacotes exigidos: dplyr

Attaching package: 'dplyr'

The following objects are masked from 'package:data.table':

between, first, last

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
require(tidyr)
```

Carregando pacotes exigidos: tidyr

```
chocolate = fread("./avaliação_01-datasets/chocolate.csv.gz")

glimpse(chocolate)
```

Rows: 2,443

Columns: 7

```
$ local_compania    <chr> "U.S.A.", "U.S.A.", "U.S.A.", "U.S.A.", "U.S.A.", "U...
$ ano              <int> 2019, 2019, 2019, 2021, 2021, 2021, 2021, 2012, 2012...
$ origem_cacau     <chr> "Tanzania", "Dominican Republic", "Madagascar", "Fij...
$ ingredientes     <chr> "3- B,S,C", "3- B,S,C", "3- B,S,C", "3- B,S,C", "3- ...
$ caracteristicas  <chr> "rich cocoa, fatty, bready", "cocoa, vegetal, savory...
$ percentagem_cacau <chr> "76%", "76%", "76%", "68%", "72%", "80%", "68%", "70...
$ ID              <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
```

a. Existem 2443 países que produzem chocolate.

```
num_paises <- chocolate %>%
  select(origem_cacau) %>%
  distinct() %>%
  nrow()

# Exibir o número de países
print(num_paises)
```

[1] 62

b. Existem 104 chocolates com 4 ingredientes que são descritos por 2 características.

```
chocolate$qtidade_caracteristicas <- lengths(strsplit(chocolate$caracteristicas, ","))

q2 <- chocolate %>%
  filter(substr(ingredientes, 1, 1) == "4",
         qtidade_caracteristicas == "2")

# Contar o número de chocolates resultantes
num_chocolates <- nrow(q2)
```

```
# Exibir o número de chocolates
print(num_chocolates)
```

[1] 104

c. A frequência absoluta para chocolates que contenham 5 ingredientes é 750.

```
# Filtrar chocolates que contenham 5 ingredientes
chocolates_5_ingredientes <- chocolate %>%
  filter(substr(ingredientes, 1, 1) == "5")

# Calcular a frequência absoluta
frequencia_absoluta <- nrow(chocolates_5_ingredientes)

# Exibir a frequência absoluta
print(frequencia_absoluta)
```

[1] 191

d. As 8 características mais marcantes dos chocolates são sweet, nutty, cocoa, roasty, creamy, earthy, sandy e fatty e juntas correspondem a 1663 descrições dos chocolates.

```
# Criar um novo dataframe com as características separadas em linhas
caracteristicas_chocolate <- chocolate %>%
  separate_rows(caracteristicas, sep = ",") %>%
  group_by(caracteristicas) %>%
  summarise(frequencia = n()) %>%
  arrange(desc(frequencia))

caracteristicas_chocolate
```

A tibble: 1,202 × 2

	caracteristicas	frequencia
	<chr>	<int>
1	"cocoa"	206
2	"sweet"	175
3	"nutty"	173
4	"creamy"	161
5	"roasty"	160
6	"earthy"	141
7	"sandy"	139
8	"sour"	122
9	"floral"	91
10	"intense"	88

i 1,192 more rows

```
# Selecionar apenas as características desejadas
palavras_chave <- c("sweet", "nutty", "cocoa", "roasty", "creamy", "earthy", "sandy",
  "fatty")

# Criar uma expressão regular para corresponder a variações nas palavras-chave
padrao_regex <- paste0("\\b", paste(palavras_chave, collapse = "|"), "\\b", ignore.case =
  TRUE)
```

```
# Filtrar as características desejadas e somar as frequências
q2b <- caracteristicas_chocolate %>%
  filter(grepl(padrao_regex, caracteristicas)) %>%
  group_by(caracteristicas)

# Exibir a soma das frequências
print(q2b)
```

```
# A tibble: 120 × 2
# Groups:   caracteristicas [120]
  caracteristicas frequencia
  <chr>           <int>
1 " cocoa"        206
2 " sweet"        175
3 " nutty"        173
4 "creamy"        161
5 " roasty"       160
6 " earthy"       141
7 "sandy"        139
8 "sweet"         85
9 "nutty"         83
10 "roasty"       52
# i 110 more rows
```

```
sum(q2b$frequencia)
```

```
[1] 1748
```

e. Existem 81 chocolates que incluem o ingrediente *Adoçante* em sua composição.

```
# Criar um novo dataframe com as características separadas em linhas
ingredientes_chocolate <- chocolate %>%
  separate_rows(ingredientes, sep = ",") %>%
  separate_rows(ingredientes, sep = "-") %>%
  group_by(ingredientes) %>%
  summarise(frequencia = n()) %>%
  arrange(desc(frequencia))

sum(ingredientes_chocolate$frequencia[ingredientes_chocolate$ingredientes == "S*"])
```

```
[1] 76
```

3. Para esse exercício você deverá utilizar os banco de dados *Art.csv.gz* e *Art_Moma.csv.gz*. Desconsidere artistas sem nacionalidade e/ou sem nome.

O dicionário das variáveis encontra-se disponível abaixo.

variável	descrição
artist_name	O nome de cada artista
edition_number	O número da edição do livro.
year	O ano de publicação de uma determinada edição do livro
artist_nationality	A nacionalidade de um artista.
artist_nationality_other	A nacionalidade do artista
artist_gender	O gênero do artista
artist_race	A raça do artista
artist_ethnicity	A etnia do artista
book	Qual livro, "Janson" ou "Gardner"
space_ratio_per_page_total	A área em centímetros quadrados do texto e da figura de um determinado artista
artist_unique_id	O número de identificação exclusivo atribuído aos artistas
moma_count_to_year	O número total de exposições já realizadas pelo Museu de Arte Moderna (MoMA)
whitney_count_to_year	O número de exposições realizadas pelo The Whitney
artist_race_nwi	O indicador de raça não branca para a raça do artista

```
require(data.table)
require(dplyr)
require(tidyr)

art_moma = fread("./avaliação_01-datasets/Art_Moma.csv.gz")
art = fread("./avaliação_01-datasets/Art.csv.gz")

glimpse(art_moma)
```

Rows: 3,162

Columns: 7

```
$ edition_number    <dbl> 9, 10, 11, 12, 13, 14, 15, 16, 14, 15, 16, ...
$ year              <int> 1991, 1996, 2001, 2005, 2009, 2013, 2016, 2...
$ book              <chr> "Gardner", "Gardner", "Gardner", "Gardner",...
$ space_ratio_per_page_total <dbl> 0.3533658, 0.3739470, 0.3032593, 0.3770489,...
```



```
$ moma_count_to_year      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
$ whitney_count_to_year   <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
$ artist_unique_id        <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 4, 4, 4, 6, 6, 6...
```

```
glimpse(art)
```

Rows: 413

Columns: 8

```
$ artist_name      <chr> "Aaron Douglas", "Adélaïde Labille-Guiard", "...
$ artist_nationality <chr> "American", "French", "French", "German-Ameri...
$ artist_nationality_other <chr> "American", "French", "French", "Other", "Ame...
$ artist_gender     <chr> "Male", "Female", "Male", "Male", "Male", "Ma...
$ artist_race       <chr> "Black or African American", "White", "White"...
$ artist_ethnicity  <chr> "Not Hispanic or Latino origin", "Not Hispani...
$ artist_unique_id  <int> 2, 4, 6, 7, 8, 12, 14, 15, 16, 17, 18, 19, 20...
$ artist_race_nwi   <chr> "Non-White", "White", "White", "White", "Whit..."
```

Assinale todas as alternativas **correta(s)**. a) Os 3 artista(s) com mais exposições no The Whitney classificados em ordem decrescente de exposições são: Edward Hopper, Georgia O'Keeffe e Stuart Davis.

```
q3a <- inner_join(art, art_moma,
                  by = "artist_unique_id") %>%
  group_by(artist_name) %>%
  summarise(sum_whitney_count_to_year = sum(whitney_count_to_year)) %>%
  arrange(desc(sum_whitney_count_to_year))
```

```
q3a
```

A tibble: 413 × 2

	artist_name	sum_whitney_count_to_year
	<chr>	<int>
1	Edward Hopper	529
2	Georgia O'Keeffe	352
3	Stuart Davis	339
4	Willem de Kooning	333
5	George Grosz	332
6	Ellsworth Kelly	276
7	Jacob Lawrence	248
8	Jasper Johns	235
9	Jackson Pollock	203
10	Roy Lichtenstein	200

i 403 more rows

b. Do total de artistas, 152 são Swiss, Mexican ou Japanese.

```
q3b <- inner_join(art, art_moma,
                  by = "artist_unique_id") %>%
  group_by(artist_nationality) %>%
  summarise(frequencia = n())
```

```
q3b
```

```
# A tibble: 52 × 2
  artist_nationality frequencia
  <chr>              <int>
1 American          908
2 Argentine           1
3 Armenian-American  10
4 Australian           7
5 Austrian           36
6 Austrian-American   5
7 Belgian            30
8 Brazilian            1
9 British            317
10 Canadian           14
# i 42 more rows
```

```
palavras_chave <- c("Swiss", "Mexican", "Japanese")

# Filtrar as características desejadas e somar as frequências
q3b <- q3b %>%
  filter(artist_nationality %in% palavras_chave)

# Exibir a soma das frequências
print(q3b)
```

```
# A tibble: 3 × 2
  artist_nationality frequencia
  <chr>              <int>
1 Japanese           56
2 Mexican            52
3 Swiss              44
```

```
sum(q3b$frequencia)
```

```
[1] 152
```

c. Apenas 6 artista(s) com a nacionalidade Swiss tiveram entre 0 e 1 exposições no The Whitney.

```
q3c <- inner_join(art, art_moma,
  by = "artist_unique_id") %>%
  group_by(artist_name, artist_nationality) %>%
  summarise(sum_whitney_count_to_year = sum(whitney_count_to_year))
```

`summarise()` has grouped output by 'artist_name'. You can override using the
`.groups` argument.

```
q3c
```

```
# A tibble: 413 × 3
# Groups:   artist_name [413]
  artist_name          artist_nationality sum_whitney_count_to_year
  <chr>              <chr>              <int>
```

	artist_name	artist_nationality	sum_whitney_count_to_year
1	A. R. Penck	German	0
2	Aaron Douglas	American	0
3	Aaron Siskind	American	0
4	Adolph Gottlieb	American	46
5	Adolphe William Bouguereau	French	0
6	Adélaïde Labille-Guiard	French	0
7	Albert Bierstadt	German-American	0
8	Albert Pinkham Ryder	American	0
9	Albert Renger-Patzsch	German	0
10	Albert Sands Southworth	American	0

i 403 more rows

```
palavras_chave <- c("Swiss")

# Filtrar as características desejadas e somar as frequências
q3c <- q3c %>%
  filter(artist_nationality %in% palavras_chave &
         sum_whitney_count_to_year <= 1)

# Exibir a soma das frequências
print(q3c)
```

```
# A tibble: 5 × 3
# Groups:   artist_name [5]
  artist_name      artist_nationality sum_whitney_count_to_year
  <chr>            <chr>                                <int>
1 Angelica Kauffmann Swiss                                0
2 Arnold Böcklin   Swiss                                0
3 Friedrich von Martens Swiss                            0
4 Henry Fuseli     Swiss                                0
5 Jean Tinguely    Swiss                                0
```

```
nrow(q3c)
```

```
[1] 5
```

d. A diferença entre a média de páginas para artistas Brancos e Não Brancos no ano de 2007 é -0,24.

```
q3d <- inner_join(art, art_moma,
                  by = "artist_unique_id")

# Filtrar por artist_race == "white" e calcular a média
media_white <- q3d %>%
  filter(artist_race_nwi == "White") %>%
  summarise(media_space_ratio = mean(space_ratio_per_page_total))

# Filtrar por artist_race diferente de "white" e calcular a média
media_nao_white <- q3d %>%
  filter(artist_race_nwi != "White") %>%
  summarise(media_space_ratio = mean(space_ratio_per_page_total))

# Calcular a diferença
```

```
diferenca <- media_white$media_space_ratio - media_nao_white$media_space_ratio

# Exibir o resultado
print(diferenca)
```

[1] 0.1395619

e. Dos artista(s) que expuseram no The Whitney, apenas 164 aparecem nos livros 'Gardner' e 'Janson'.

```
q3e <- inner_join(art, art_moma,
                  by = "artist_unique_id")

unique(q3e$book)
```

[1] "Gardner" "Janson"

```
q3e <- q3e %>%
  filter(whitney_count_to_year > 0) %>%
  group_by(artist_name) %>%
  summarise(frequencia = n())

nrow(q3e)
```

[1] 101

4. Para esse exercício você deverá utilizar os banco de dados refugiados_pais.csv.gz e refugiados.csv.gz. Considere apenas observações completas.

Assinale todas as alternativas **correta(s)**.

```
require(data.table)
require(dplyr)
require(tidyr)

refugiados_pais= fread("./avaliação_01-datasets/refugiados_pais.csv.gz")
refugiados = fread("./avaliação_01-datasets/refugiados.csv.gz")

glimpse(refugiados_pais)
```

Rows: 230

Columns: 4

```
$ id      <chr> "AFG", "ALB", "DZA", "AND", "AGO", "AIA", "ATG", "ARG", "ARM..."
$ nome    <chr> "Afghanistan", "Albania", "Algeria", "Andorra", "Angola", "A..."
$ regioao <chr> "Asia", "Europe", "Africa", "Europe", "Africa", "Americas", ...
$ subregiao <chr> "Southern Asia", "Southern Europe", "Northern Africa", "Sout..."
```

```
glimpse(refugiados)
```

Rows: 120,338

Columns: 4

```
$ ano          <int> 1951, 1951, 1951, 1951, 1951, 1951, 1951, 1951, 1951, 1951, 1951,...
$ id_origem    <chr> "UNK", "UNK", "UNK", "UNK", "UNK", "UNK", "UNK", "UNK", "UNK", "UN...
$ id_destino    <chr> "AUS", "AUT", "BEL", "CAN", "DNK", "FRA", "GBR", "DEU", "GR...
$ refugiados   <int> 180000, 282000, 55000, 168511, 2000, 290000, 208000, 265000...
```

```
# Realizar o left join
tb4 <- refugiados %>%
  left_join(refugiados_pais, by = c("id_origem" = "id")) %>%
  left_join(refugiados_pais, by = c("id_destino" = "id"), suffix = c("_origem", "_destino"))

# Exibir a planilha
print(tb4)
```

	ano	id_origem	id_destino	refugiados	nome_origem
1:	1951	UNK	AUS	180000	<NA>
2:	1951	UNK	AUT	282000	<NA>
3:	1951	UNK	BEL	55000	<NA>
4:	1951	UNK	CAN	168511	<NA>
5:	1951	UNK	DNK	2000	<NA>

120334:	2022	VEN	ABW	0	Venezuela (Bolivarian Republic of)
120335:	2022	ABW	ABW	0	Aruba
120336:	2022	VEN	CUW	0	Venezuela (Bolivarian Republic of)
120337:	2022	CUW	CUW	0	Curacao
120338:	2022	VEN	SXM	0	Venezuela (Bolivarian Republic of)

	regiao_origem	subregiao_origem	nome_destino
1:	<NA>	<NA>	Australia
2:	<NA>	<NA>	Austria
3:	<NA>	<NA>	Belgium
4:	<NA>	<NA>	Canada
5:	<NA>	<NA>	Denmark

120334:	Americas Latin America and the Caribbean	Aruba
120335:	Americas Latin America and the Caribbean	Aruba
120336:	Americas Latin America and the Caribbean	Curacao
120337:	Americas Latin America and the Caribbean	Curacao
120338:	Americas Latin America and the Caribbean Sint Maarten (Dutch part)	

	regiao_destino	subregiao_destino
1:	Oceania	Australia and New Zealand
2:	Europe	Western Europe
3:	Europe	Western Europe
4:	Americas	Northern America
5:	Europe	Northern Europe

120334:	Americas Latin America and the Caribbean
120335:	Americas Latin America and the Caribbean
120336:	Americas Latin America and the Caribbean
120337:	Americas Latin America and the Caribbean
120338:	Americas Latin America and the Caribbean

a. A matriz de migração [origem, destino] intercontinental do ano 2006 é dada por: |Região |Africa
|Americas|Asia |Europe|Oceania| |——|——|——|——|——|——| |Africa |2507581|262745

```
|98175 |250070|37124 | |Americas|0 |150149 |0 |14850 |174 | |Asia |76780 |308706
|4411284|664075|42704 | |Europe |94 |306672 |7816 |454237|3423 | |Oceania |0 |1679 |0 |92 |59 |
```

```
# Filtrar para o ano de 2006
tb4_2006 <- tb4 %>% filter(ano == 2006)

# Criar a matriz de migração [origem, destino]
matriz_migracao_2006 <- tb4_2006 %>%
  group_by(regiao_origem, regiao_destino) %>%
  summarise(total_migrantes = sum(refugiados)) %>%
  pivot_wider(names_from = regiao_destino, values_from = total_migrantes, values_fill = 0)
```

`summarise()` has grouped output by 'regiao_origem'. You can override using the
`.groups` argument.

```
# Exibir a matriz de migração no formato desejado
print(matriz_migracao_2006)
```

```
# A tibble: 6 × 7
# Groups:   regiao_origem [6]
  regiao_origem Africa Americas Asia Europe Oceania `NA`
  <chr>          <int>    <int> <int> <int> <int> <int>
1 Africa        2507581  262745  98175 250070  37124    0
2 Americas            0  150149    0  14850   174    0
3 Asia           76780  308706 4411284 664075  42704    0
4 Europe            94  306672   7816 454237  3423    0
5 Oceania            0   1679    0    92    59    0
6 <NA>          22980   5748  20381 228763   518    0
```

b. A partir de 1972 houveram 172075 refugiados partindo do país: Afghanistan para o país: Canada, e 219920 refugiados partindo do país: Pakistan para o país: Canada.

```
# Especificar o nome_origem, nome_destino e ano desejados
ano_desejado <- 1972

q4b <- tb4 %>%
  group_by(nome_origem, nome_destino, ano) %>%
  summarise(refugiados = sum(refugiados), .groups = "drop")

resultado <- q4b %>%
  filter(ano >= ano_desejado) %>%
  group_by(nome_origem, nome_destino) %>%
  summarise(total_refugiados = sum(refugiados))
```

`summarise()` has grouped output by 'nome_origem'. You can override using the
`.groups` argument.

```
refugiados_af_can <- resultado %>%
  filter(nome_origem == "Afghanistan", nome_destino == "Canada")
print(refugiados_af_can)
```

```
# A tibble: 1 × 3
# Groups:   nome_origem [1]
  nome_origem nome_destino total_refugiados
  <chr>       <chr>          <int>
1 Afghanistan Canada              172022
```

```
refugiados_paq_can <- resultado %>%
  filter(nome_origem == "Pakistan", nome_destino == "Canada")
print(refugiados_paq_can)
```

```
# A tibble: 1 × 3
# Groups:   nome_origem [1]
  nome_origem nome_destino total_refugiados
  <chr>       <chr>          <int>
1 Pakistan    Canada              219988
```

c. Os 5 países que mais enviaram refugiados no ano de 1965 pertencem às subregiões Sub-Saharan Africa e Southern Europe.

```
ano_c <- 1965

q4c <- tb4 %>%
  group_by(nome_origem, subregiao_origem, ano) %>%
  summarise(refugiados = sum(refugiados), .groups = "drop")

q4c <- q4c %>%
  filter(ano == ano_c) %>%
  group_by(nome_origem, subregiao_origem) %>%
  summarise(total_refugiados = sum(refugiados)) %>%
  arrange(desc(total_refugiados))
```

`summarise()` has grouped output by 'nome_origem'. You can override using the `.groups` argument.

```
q4c
```

```
# A tibble: 15 × 3
# Groups:   nome_origem [15]
  nome_origem      subregiao_origem total_refugiados
  <chr>          <chr>          <int>
1 <NA>          <NA>          1581335
2 China         Eastern Asia    1330000
3 Angola        Sub-Saharan Africa 220000
4 Rwanda        Sub-Saharan Africa 155900
5 Sudan         Northern Africa  81560
6 Guinea-Bissau Sub-Saharan Africa 50000
7 Dem. Rep. of the Congo Sub-Saharan Africa 47000
8 Greece        Southern Europe  17690
9 Mozambique    Sub-Saharan Africa 17000
10 Viet Nam     South-eastern Asia 15000
11 Ghana        Sub-Saharan Africa  5000
12 Albania      Southern Europe   4860
13 Hungary      Eastern Europe    3270
```

14	Burundi	Sub-Saharan Africa	3000
15	Congo	Sub-Saharan Africa	0

d. Os 6 países que mais receberam refugiados a partir de 1982 receberam juntos 19523 refugiados.

```
ano_d <- 1982

q4d <- tb4 %>%
  group_by(nome_destino, ano) %>%
  summarise(refugiados = sum(refugiados), .groups = "drop") %>%
  na.omit()

q4d <- q4d %>%
  filter(ano >= ano_d) %>%
  group_by(nome_destino) %>%
  summarise(total_refugiados = sum(refugiados)) %>%
  arrange(desc(total_refugiados))

q4d
```

```
# A tibble: 191 × 2
  nome_destino      total_refugiados
  <chr>              <int>
1 Pakistan          75814675
2 Iran (Islamic Rep. of) 72881894
3 Germany           32908138
4 Türkiye           29883329
5 Sudan             23165858
6 United States of America 17905925
7 Dem. Rep. of the Congo 16212832
8 Uganda            16166236
9 Ethiopia          15899126
10 United Rep. of Tanzania 14793351
# i 181 more rows
```

e. Existem 27 países que receberam pelo menos 5382652 refugiados.

```
n_refugiados <- 5382652

q4e <- tb4 %>%
  group_by(nome_destino) %>%
  summarise(refugiados = sum(refugiados), .groups = "drop") %>%
  na.omit()

q4e <- q4e %>%
  filter(refugiados >= n_refugiados) %>%
  arrange(desc(refugiados))

cat('Existem', nrow(q4e), 'países que receberam pelo menos', n_refugiados, 'refugiados.\n')
```

Existem 30 países que receberam pelo menos 5382652 refugiados.

