

1. Considere a tabela de dados guardada num objeto da classe `data frame`, `C02` no R. Estes dados referem-se ao consumo de CO_2 em plantas de duas regiões distintas (Mississippi e Quebec) quando expostas a diferentes concentrações de CO_2 ou quando sujeitas, ou não, a arrefecimento.
 - (a) Verifique a informação contida na tabela de dados em causa (recorra ao `help`), a dimensão, nomes das variáveis e classe de cada um dos objetos.
 - (b) Quantas plantas foram sujeitas a tratamento (`Treatment`) de arrefecimento (`chilled`)?
 - (c) Defina um subconjunto de dados apenas com as plantas da região do Quebec. Exporte-o para um documento `txt`.
 - (d) Defina uma matriz com os dados da concentração de CO_2 e de `uptake`. Atribua nomes às linhas (identificação da planta) e às colunas (strings `concentra` e `consumo`).
 - (e) Usando a matriz definida na alínea anterior, selecione os valores de concentração e consumo da planta identificada com `Mc3`. Crie um vetor com os valores de consumo a partir desta matriz.
 - (f) Num campo com 1000 plantas do tipo `QC3`, sujeitas a uma concentração constante de 675 mL/L, qual o consumo de CO_2 ?
 - (g) Em que condições o consumo de CO_2 é o mais elevado? Obtenha a mesma informação para cada uma das regiões.
 - (h) Qual a média de consumo de CO_2 realizado pelas plantas da região do Mississippi, sujeitas a tratamento?
2. Considere o `data frame` `airquality` que contém dados sobre a qualidade do ar em Nova York entre maio e setembro de 1973.
 - (a) Visualize a estrutura do `data frame` (use a função `str`).
 - (b) Verifique quantos valores dos campos `Ozone` e `Solar.R` não estão disponíveis (missing values – `Not Available`; função de teste: `is.na`).
 - (c) Quais os índices das linhas que têm missing values na variável `Ozone`?
 - (d) Qual o valor médio da radiação solar durante o mês de julho?
 - (e) Defina um novo `data frame` a partir desta em que não existam observações NA.

- (f) Defina uma função que, para cada um dos meses de maio a setembro, calcule o máximo, mínimo, média, mediana e desvio padrão de radiação solar, usando o novo data frame definido na alínea anterior.
- (g) Com a função `boxplot` construa diagramas de extremos e quartis para a concentração de Ozono por mês e para a temperatura por mês.
- (h) Use a função `split` para obter as temperaturas por mês. Tendo em conta a classe do resultado obtido, obtenha a média das temperaturas por mês usando uma função adequada da família `apply`.
- (i) Obtenha a média das concentrações de Ozono e a média das temperaturas, por mês, usando agora a função `aggregate`.

3. A tabela de dados em Excel `bnames` contém a percentagem de bebés com cada um de 6872 nomes próprios, por ano, do género masculino ou feminino, registados nos EUA desde 1880 até 2008.

- (a) Importe a tabela usando a função `read.csv`, guardando-o no objeto `nomEUA`, verifique a classe de `nomEUA` e a sua estrutura.
- (b) Classifique as variáveis constantes no `data.frame` e passe a variável `year` para a classe `factor`.
- (c) Quantos bebés foram avaliados em cada ano? Quantas meninas? Qual o nome mais popular entre as meninas no ano de 1880? E no ano 2008? E entre os meninos?
- (d) Faça uma ordenação por ano e por género dos nomes (do mais popular ao menos popular). Encontre uma forma automática de ir buscar o nome mais popular de menina e de menino em cada ano para todos os anos.

4. Em 1880 Christian Zeller, matemático alemão, notou que se percorrermos um ano de março a fevereiro, o número acumulado de dias em cada mês forma *quase* uma linha reta que passa *quase* pela origem (por se ter passado fevereiro para último lugar).

A fórmula que aqui se apresenta pode ser usada para determinar o dia da semana em função do dia do mês, do mês - 2, e do ano. De facto, março é 1, abril 2, e assim sucessivamente até dezembro 10. O mês de janeiro é 11 e fevereiro 12 (último mês) ambos do ano precedente. O ano (com 4 dígitos, por exemplo 1965), por sua vez, encontra-se dividido em século (os primeiros dois dígitos +1) e ano (os últimos 2 dígitos). Por exemplo 29 de mar de 1965, corresponde ao dia 29, do mês 1 do ano 65 do século 20, enquanto que 29 de janeiro de 1965, corresponde ao dia 29, do mês 11, do ano 64 (ano precedente!) do século 20.

Considerando d o dia, m o mês menos 2, a os últimos dois dígitos do ano e s os primeiros dois dígitos de ano, o dia da semana pode ser calculado através do resto da divisão de $[2.6m - 0.2] + d + a + [a/4] + [s/4] - 2s$ por 7 (sete dias da semana) onde 0 é sábado, 1 é domingo e assim sucessivamente até 6 sexta-feira.

A notação $[]$ é usada para designar a parte inteira de um número. Por exemplo $[6.3]$ é 6, assim como

também é 6, [6.95]. No R usa-se a função `floor` com este objetivo.

Implemente em R o algoritmo de Zeller através de uma função com argumentos dia, mês (1-janeiro a 12-dezembro), ano com 4 dígitos (exemplo 2002). Esta função deve ser vetorizável, ou seja, deve ser possível obter simultaneamente os dias da semana de 22 de nov de 1965, 19 de ago de 1992 e 8 de jan de 1997 (`f(c(22,19,8),c(11,8,1), c(1965,1992,1997))`). Caso necessite de usar, numa primeira abordagem, a instrução `if else`, (`else` não é obrigatório), a sintaxe é `if(condicao verdadeira) ... else ...`. No entanto esta instrução não é uma função vetorizável. Existe uma opção vetorizável desta instrução – função `ifelse` (consulte a ajuda do programa).

Repare que se o mês for janeiro ou fevereiro, há necessidade de somar 12 e, simultaneamente, subtrair 1 ao ano.

5. Considere a sucessão definida recursivamente por $x_0 = 1$, $x_1 = 2$; $x_j = x_{j-1} + \frac{2}{x_{j-1}}$ para $j = 1, 2, \dots$. Escreva uma função com um único argumento `n` que devolva os `n-1` termos da sucessão $\{x_j\}_{j \geq 0}$. A sintaxe para os comandos associados a instruções condicionais, *loops* (iterações), ou seja usados para *control flow* (fluxo de execução) pode ser vista na ajuda do R fazendo, por exemplo, `?for` (tal como para operadores aritméticos são usadas aspas no acesso ao *help*).
6. Considere a experiência que consiste em contar o número de vezes que sai `coroa` antes que saia pela primeira vez `cara` no lançamento de uma moeda equilibrada. Simule esta experiência usando a função `while`. Repita a experiência 4000 vezes e obtenha uma tabela onde figurem os resultados obtidos. Elabore um gráfico adequado.
7. Pretende-se estudar o comportamento do valor médio do número de pontos obtido em 10 lançamentos (nas mesmas condições) de um dado equilibrado. Comece por simular um resultado para 10 lançamentos de um dado equilibrado e calcule a sua média. Repita agora o procedimento anterior 400 vezes (simule 400 vezes o lançamento de um dado equilibrado 10 vezes). Calcule as médias observadas das 400 repitções da experiência e elabore um gráfico adequado. Que intervalo de valores lhe parece razoável para o valor médio do número de pontos obtido em 10 lançamentos?