

EXERCÍCIOS

Baixe do SIGAA/Email o arquivo “baleia.xls” e salve-o em seu computador. Você vai utilizá-los nas questões seguintes. Os dados desse arquivo foram coletados durante dois cruzeiros de pesquisa no Atlântico Norte em maio e outubro de 2003. Durante esses dois meses, o navio de pesquisa visitou várias estações (áreas) e pesquisadores de mamíferos marinhos registraram o número de baleias em cada uma dessas estações. O tempo que a embarcação passou em cada estação também foi registrado junto com outras variáveis específicas do local, como latitude e longitude, profundidade da água e gradiente do fundo do mar. Os pesquisadores também registraram o nível ambiente de ruído subterrâneo com um hidrofone e categorizaram essa variável em “low”, “medium” ou “high”. Cada linha é uma observação individual e cada coluna uma variável separada. Os nomes das variáveis estão contidos na primeira linha do conjunto de dados (também conhecido como cabeçalho).

1 - Agora, vamos ler no R o arquivo ‘baleia.xlsx’ para R e atribuir ele ao data.frame **baleia**. Quantas observações este conjunto de dados tem? Quantas variáveis existem neste conjunto de dados? Que tipo de variáveis são mês e ruído da água?

2 - A função `summary()` fornece algumas estatísticas de resumo úteis para cada variável. Observe como o tipo de saída depende se a variável é um fator ou um número. Outro recurso útil da função `summary()` é que ela também contará o número de valores ausentes em cada variável. Quais variáveis têm valores ausentes e quantos?

3 - Acesse todos os elementos das primeiras 10 linhas e das primeiras 4 colunas do data.frame **baleia** e atribua a um novo objeto chamado **baleia.sub**. Em seguida, acesse todas as observações de **baleia** e das colunas mês, ruído da água e Numero_baleias e atribua a um objeto chamada **baleia.num**. Agora, acesse as primeiras 50 linhas e todas as colunas e atribua a um objeto **baleia.maio**. Finalmente, acesse todas as linhas, exceto as primeiras 10 linhas e todas as colunas, exceto a última coluna. Lembre-se de que, para algumas dessas perguntas, você pode especificar as colunas que deseja por posição ou por nome. Pratique dos dois lados. Você tem alguma preferência? Se sim, por quê?

4 - Além de acessar linhas e colunas de seu data.frame por posição, você também pode usar instruções condicionais para selecionar linhas específicas com base em alguns critérios lógicos. Acesse linhas de **baleia** (todas as colunas por padrão) com base nos seguintes critérios (observação: você deverá atribuir os resultados dessas declarações objetos nomeados apropriadamente:

em profundidades superiores a 1200 m

gradiente mais íngreme do que 200 graus

nível de ruído da água “low”

nível de ruído da água de “high” no mês de “May”

mês de “October”, nível de ruído da água de “low” e gradiente maior do que o valor médio do gradiente

todas as observações entre as latitudes 60,0 e 61,0 e longitudes -6,0 e -4,0

todas as linhas que não têm um nível de ruído de água “medium”

5 - Escreva um código para acessar todas as linhas do data.frame **baleia** com profundidades maiores que 1500m e com um número de baleias avistadas maior do que a média.

6 - Outra maneira útil de manipular seus data.frames é ordenar as linhas com base no valor de uma variável (ou combinações de variáveis). Use a função `order()` para classificar todas as linhas de

baleia em ordem crescente de profundidade. Armazene esse data.frame ordenado em um objeto chamado **baleia.ord.prof**.

7 - Classifique todas as linhas de **baleia** por ordem crescente de profundidade dentro de cada nível de ruído da água. O truque aqui é lembrar que você pode ordenar por mais de uma variável ao usar a função `order()`. Atribua seu data.frame ordenado a um novo objeto com um nome adequado.

8 - Use a função `aggregate()` para calcular a média de tempo na estação, o número de baleias, a profundidade e o gradiente para cada nível de ruído da água (não se esqueça daquele valor NA). Em seguida, calcule a média de tempo na estação, número de baleias, profundidade e gradiente para cada nível de ruído da água em cada mês.

9 - Saber quantas observações estão presentes para cada nível de fator (ou combinações de níveis de fator) é útil para determinar se você tem um tamanho de amostra adequado. Use a função `table()` para determinar o número de observações para cada nível de ruído da água. Em seguida, use a mesma função para exibir o número de observações para cada combinação de ruído da água e mês.

10 - Exporte o data.frame **baleia.num** criado anteriormente para um arquivo chamado "baleia_num.txt". Para fazer isso, você precisará usar o `write.table()` função. Você deseja incluir os nomes das variáveis na primeira linha do arquivo, mas não deseja incluir os nomes das linhas. Além disso, certifique-se de que o arquivo seja um arquivo delimitado por tabulação.