

Exercícios 4

Manipulação de ficheiros

Em termos computacionais, a informação é habitualmente transmitida em blocos a que denominamos ficheiros. O **R** disponibiliza uma vasta colecção de funções para manipular ficheiros de diferentes formatos. Neste exercício, pretende-se fazer uso de algumas funções de manipulação de ficheiros, e dos objectos resultantes do seu carregamento numa sessão de **R**.

Considere o seguinte contexto. Tornou-se num reconhecido *data analyst*, e uma das maiores empresas da indústria petrolífera está a requisitar os seus serviços. Eles estão a começar a explorar um novo reservatório, tendo já disponível os dados obtidos no primeiro furo (exemplo ilustrativo na Figura 1).



Figura 1: Bomba de produção no oeste do Texas (fonte: wikimedia.org).

O engenheiro de reservatórios tem conhecimento da sua proficiência em **R**, então enumerou as seguintes tarefas a executar.

Nota: Resolva as questões recorrendo, tanto quanto possível, a uma abordagem mais funcional e menos imperativa (i.e., usando funções e evitando usar ciclos, ou, dizendo de outro modo, tirando maior proveito das funcionalidades do **R**).

1. Os dados estão disponíveis para download na seguinte localização: <http://tinyurl.com/welldata1>. Carregue os dados para uma estrutura de dados adequada.
2. Os dados recebidos contêm 7 variáveis:

X: Coordenada X

Y: Coordenada Y

Z: Coordenada Z

facies: diferentes tipos de rocha

density: densidade da rocha

porosity: porosidade da rocha

permeability: permeabilidade da rocha (medida em mD)

Para evitar problemas na análise de dados, verifique se cada coluna ficou com o tipo de dados mais adequado, e ajuste conforme necessário.

3. As fácies correspondem a diferentes zonas no reservatório. Identifique quantas amostras recebeu de cada uma das zonas.
4. Como primeira abordagem para identificar zonas de maior interesse (*high pay zones*), obtenha um resumo com os estatísticos base de cada variável, em cada zona. Deste modo, poderá saber os valores típicos em cada uma das fácies.
5. Outra estratégia consiste em encontrar as amostras com maiores valores de porosidade ($\geq 0,3$). Identifique essas amostras. Qual é a zona com maior número dessas amostras?
6. Entretanto, recebeu dados adicionais, que foram recolhidos mais tarde. Foram enviados através da ligação <http://tinyurl.com/welldata2>. Este novo conjunto de dados contém as seguintes variáveis:

X: Coordenada X

Y: Coordenada Y

Z: Coordenada Z

Vp: Velocidade das ondas P (km/s)

Vs: Velocidade das ondas S (km/s)

Foi-lhe passada a informação de que o processo de aquisição de dados teve alguns problemas, e que é possível que tenha recebido a mesma amostra mais do que uma vez. Tome as medidas necessárias para corrigir esse possível erro.

7. Com esse percalço ultrapassado, pode agora proceder à junção dos dois conjuntos de dados. Verifique as variáveis que têm em comum, e junte-os de acordo com as mesmas.
8. Embora já tenha preparado todos os dados disponíveis, pode ainda ajudar na próxima etapa da modelação do reservatório, providenciado duas variáveis úteis. A impedância acústica é dada pelo produto da densidade com a velocidade sísmica, que varia entre diferentes camadas de rocha. Adicione duas colunas ao conjunto de dados:

Ip: impedância acústica considerando as ondas P

Is: impedância acústica considerando as ondas S

9. O seu trabalho está concluído. Agora, resta prepará-lo para ser entregue. Guarde os dados que preparou num formato de ficheiro apropriado.