

Faculdade Cotemig
Pós-Graduação em Data Science Big Data
Fundamentos Matemáticos para Ciência dos Dados
Lista de Exercícios 1

Professora: Melissa Pinho

Data de entrega: 19/11/2018

- O arquivo "dados_rais_2014.txt" contém informações de número de empregos e remuneração dos empregados formais em 2014, por Unidade Federativa (sg_estado) e por Ocupação (id_cbo).
 - O arquivo "CBO2002_Familia.txt" contém a relação de códigos de ocupação (codigo) e sua respectiva descrição (descricao).
 - O arquivo "CBO2002_Grande_Grupo.txt" contém a relação de códigos de grande grupo de ocupação (Codigo) e sua respectiva descrição (Titulo). Cada ocupação pertence a um grande grupo, que é definido pelo primeiro caracter de seu código; por exemplo, a ocupação 'Engenheiros ambientais e afins' possui o código CBO igual a "2140", portanto pertence ao Grande Grupo 2: Profissionais das Ciências e das Artes.
1. Carregue os arquivos no R. **Dica:** verifique os argumentos *encoding* e *colClasses* da função *read.table*.
 2. Insira em seu data frame que contém os dados da RAIS as descrições de Família e Grandes Grupos. **Dica:** Para criar o Grande Grupo, utilize a função *substr*.

3. Calcule a Renda Mensal Total e o número total de empregos por estado, por ocupação e por grande grupo.
 4. Ordene os estados por ordem decrescente de número de empregos
 5. Qual ocupação possui o maior número de empregos no Brasil?
 6. Qual ocupação possui maior remuneração média? E a menor?
 7. Quantas ocupações possuem remuneração média menor que R\$ 2.000,00 reais?
 8. Dentre os técnicos de nível médio de Minas Gerais, qual ocupação possui maior remuneração média?
 9. Faça um gráfico de barras representando o número de empregos por Grande Grupo.
 10. Calcule o % de empregos por Região e construa um gráfico de pizza representando tais valores.
 11. Construa uma tabela de referência cruzada com o número de empregos por grande grupo (linhas) e por região (colunas). **Dica:** Utilize a função *wtd.table* do pacote *questionr*.
-

- Utilizando o conjunto de dados **infert** (`data(infert)`), faça:

1. Calcule o número total de abortos (induzidos e espontâneos)
2. Construa uma tabela de frequência cruzada entre o número total de abortos e a escolaridade das mulheres.

3. Utilizando o teste qui-quadrado, verifique se há relação entre o número de abortos e a escolaridade. *Obs.:* Para corrigir o problema de aproximação do teste, tente unir categorias da escolaridade.
4. Construa um boxplot da variável idade para cada escolaridade.