Faculdade de Tecnologia Bandeirantes - BandTec

 1^o semestre de 2015

Exercícios sobre Análise Exploratória com R

Pós Graduação em Big Data

Prof. Fabrício Jailson Barth

1 Questões gerais sobre R

- O que é data.frame?
- O que é tipo factor?
- Qual é o tipo utilizado em R para representar números contínuos?
- O que a função sapply faz?
- O que a função summary faz?

2 Exercício 1

- Faça a carga do dataset iris.
- Exclua o atributo Species do dataset.
- Calcule a correlação entre os atributos restantes.
- Imprima gráficos que demonstram a correlação positiva alta no caso de alguns atributos.
- Imprima gráficos que demonstram a não correlação entre atributos.
- Calcule o desvio padrão dos atributos. Para aquele atributo com maior desvio padrão e para aquele atributo com menor desvio padrão imprima gráficos de histograma.
- Teste o comando summary(iris). O que ele retorna? Comente o resultado.

3 Exercício 2

- Faça a carga do dataset survey da biblioteca UsingR
- Descreva os atributos do dataset.
- Existem valores NA no dataset? Quantos?
- Qual o percentual do dataset escreve com a mão esquerda?
- Quantos exemplos do dataset são do sexo feminino?
- Faça uma análise exploratória dos dados e descreve os achados que você considera mais "interessantes".

4 Exercício 3

- Faça o download do arquivo que está em http://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/all_month.csv
- Usando o comando read.csv do R¹, importe as informações do arquivo all_month.csv em um data.frame.
- Entenda melhor o significado de cada atributo acessando o site http://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php.
- Usando a função

```
diaHorario <- date()</pre>
```

armazene o dia e horário quando os dados foram coletados.

- Qual a quantidade de exemplos (abalos) foi coletada?
- Quantos exemplos existem no dataset?
- Quais são os tipos de cada atributo?
- O R conseguiu importar corretamente todos os atributos com os tipos corretos?
- Quantos tipos distintos de abalos existem no dataset?
- Existe alguma correlação entre a profundidade do abalo sísmico (depth) e a magnitude (mag) do abalo? Faça uso de algum gráfico para suportar a sua resposta.

5 Desafios

- O dataset com abalos sísmicos possui um atributo chamado time. Como o próprio nome diz, este atributo é
 um atributo do tipo time. No entanto, ele foi importado como factor o que está errado. Converta os valores
 deste atributo para o tipo correto. Depois de convertido, tente utilizar esta informação criando plots com séries
 temporais.
- Cada registro possui uma latitude e uma longitude. Portanto, é possível plotar as informações dos abalos sísmicos em um mapa. Que tal tentar?

6 Sites interessantes para serem navegados

- http://catalog.data.gov/dataset
- https://www.kaggle.com/

¹Se você tiver dúvidas sobre o uso desta função então digite help(read.csv)