

Programação em R - Exercícios - Lista 3

Carlos Cinelli

Julho, 2016

1. Considere os dados em `roubo2.rds`. Esses dados contêm registros online de crimes por todo o Brasil, e foram coletados do site *Onde Fui Roubado?* (<http://www.ondedefuiroubado.com.br>).
 - Utilize `sapply` para calcular o número de NA's por colunas.
 - Calcule o número de crimes por cidade e ordene de forma decrescente. Onde houve mais crimes registrados? Quantos crimes foram registrados em Brasília?
 - Calcule o número de crimes por tipo de crime. Qual o tipo de crime mais frequente? Calcule, por cidade, a frequência relativa de tipo de crime. Das cidades que tem mais de 100 registros, qual cidade tem a maior proporção de assalto à mão armada?
 - Crie uma coluna que indique o dia da semana em que o crime foi cometido. Em que dia da semana foram registrados mais crimes? E considerando somente Brasília? Quais os tipos de crimes mais cometidos para cada dia da semana?
2. Considere os dados `dados.rds`.
 - O índice de concentração de Herfindahl–Hirschman (HHI) de um mercado é dado pelo somatório do quadrado do market share das empresas. Considere como empresa a corretora. Calcule o HHI para a oferta de imóveis de Brasília, separado por aluguel e venda (elimine os imóveis duplicados). Qual mercado parece ser mais concentrado? Agora calcule o HHI separado por bairro. Qual bairro parece ter um mercado mais concentrado?
 - Crie uma série temporal com a mediana dos preços de imóveis, de aluguel e venda, no plano piloto (a coluna `coleta` contém a data em que o preço do imóvel foi coletado, utilize esta coluna). Crie a mesma série separada por tipo de imóvel. Calcule a média e desvio-padrão móvel das séries (veja a ajuda da função `rollapply()` do pacote `zoo`), com uma janela de 15 dias.
 - Retire os outliers de `dados` e elimine os duplicados: quais os bairros com o maior aluguel mediano para cada tipo de imóvel? Quais os bairros com a maior mediana de preço por m2 de venda, por tipo de imóvel?
 - Usando a base completa, tente encontrar imóveis com as características que você deseja. Utilize os dados da coleta em 2014-08-31 e use a função `grep1()` em algum momento do filtro.
3. Considere os dados de salários de servidores públicos federais `20140131_Remuneracao.csv` e os dados de cadastro `20140131_Cadastro.csv` dentro do arquivo zip `201401_servidores.zip` (descompacte o zip se necessário):
 - Leia os arquivos no R usando os comandos ao final da questão (pode ser necessário modificar alguns argumentos do comando).

- Selecione da base de salários apenas as colunas ID do servidor e a remuneração básica bruta.
- Selecione da base de cadastro apenas as colunas de ID do servidor, nome do servidor e órgão de exercício.
- Renomeie as colunas para nomes mais fáceis de usar (e deixe as duas colunas de ID do servidor com o mesmo nome).
- Crie uma base de dados chamada **salcad** com o resultado do merge entre as duas bases acima.
- Com a base criada, responda: qual(is) o(s) órgão(s) com maiores remunerações média, mediana, máxima e mínima? Qual o órgão com maior variabilidade de salários (escolha uma medida de dispersão)?

```
salarios <- read.csv("Dados/20140131_Remuneracao.csv",
                    dec = ",",
                    header = TRUE,
                    fileEncoding = "latin1",
                    sep = "\t",
                    stringsAsFactors = FALSE)

cadastro <- read.csv("Dados/20140131_Cadastro.csv",
                    dec = ",",
                    header = TRUE,
                    fileEncoding = "ASCII",
                    sep = "\t",
                    stringsAsFactors = FALSE)
```