

Especialização em Ciências de Dados - Lista 2

15 de outubro de 2018

1. Carregue o arquivo "poluicao.csv", e chame o dataframe de `poluicao`. Ele contem as medições atmosféricas diárias de partículas sólidas (PM10), monóxido de carbono (co) e dióxido de nitrogênio (no2), durante o ano de 1997. Não esqueça de verificar o campo DATA, além de corrigir o limitador de decimal, codificação, etc.
2. Faça o histograma das medições de PM10, co e no2. Utilize os comandos opcionais para deixar o gráfico mais amigável. Exporte os gráficos com os nomes *hist1.png*, *hist2.png* e *hist3.png*.
3. Construa tabelas de frequência para as variáveis, com os mesmos intervalos e quantidades de categorias utilizados no histograma. Salve as tabelas de frequência em uma planilha, com o nome *tabela1*.
4. Utilizando o comando `plot()`, construa um gráfico de linhas para a evolução temporal de cada uma das variáveis PM10, co e no2. Exporte os gráficos com os nomes *PM10.png*, *co.png* e *no2.png*. Quais os problemas ao tentar analisar os dados com esses gráficos?
5. Cria duas novas variáveis em *poluicao*, uma com o número da semana à qual aquela data corresponde (semana 1, semana 2, etc), e outra com o mês. Dica: instale o pacote `lubridate`, que facilita a manipulação de datas. Utilize o comando `week()` para extrair a semana e `month()` para extrair o mês.
6. Utilizando o comando `group_by()`, em conjunto com `sumarize()`, crie dois novos dataframes, um chamado *poluicao_semana*, contendo as medições com as médias semanais, e outro chamado *poluicao_mes*, com as medições em médias mensais.
7. Construa o gráfico da evolução temporal de cada uma das variáveis de *poluicao_semana* e *poluicao_mes*, utilizando os comandos opcionais para tornar os gráficos o mais amigável possível. Exporte-os, com os nomes *PM10_semana.png*, *co_semana.png*, *no2_semana.png*, para os gráficos semanais, e o equivalente para os dados mensais.
8. Para analisar uma possível relação entre as três medições (PM10, co e no2) ao longo do tempo, tente plotar as três medições no mesmo gráfico, nos casos semanais e mensais. Quais foram as dificuldades encontradas? Esse tipo de análise gráfica é uma boa ferramenta para avaliar relação entre as diferentes medições?
9. Você deve entregar
 - (a) Os gráficos salvos.
 - (b) A planilha salva.
 - (c) Um relatório, em formato pdf, contendo todos os gráficos e planilhas elaboradas, com a resposta dos questionamentos feitos ao longo da tarefa.
 - (d) Os comandos utilizados.