

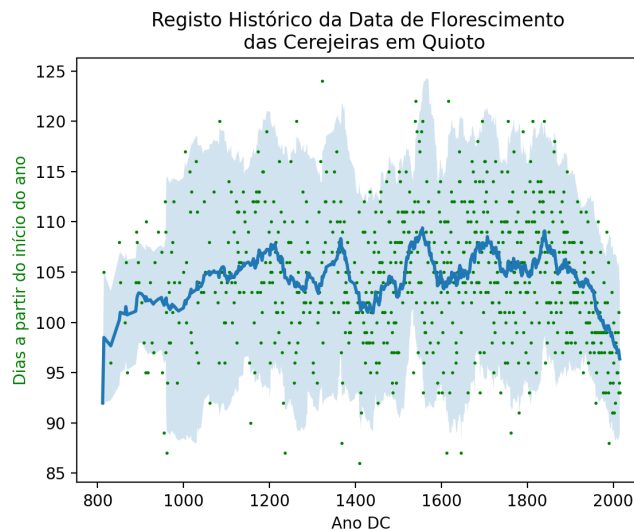
Programação II

Trabalho 4

Gráficos a partir de dados em *csv*

Entrega a 26 de abril de 2021

O objetivo deste trabalho é desenvolver funções que permitam desenhar três gráficos para um dado conjunto de dados: um gráfico de pontos, um gráfico da curva da média móvel, e um gráfico de área sombreada correspondente à área entre os valores 2 desvios-padrão acima e abaixo da média. A figura seguinte exemplifica o resultado para o conjunto de dados do ficheiro `kyoto.csv`.



Ficheiro `kyoto.csv` O ficheiro `kyoto.csv` contém informação sobre a data de florescimento das cerejeiras (*sakura* 🌸) em Quioto, Japão, desde o século IX até ao presente, recolhidos a partir de diários e crónicas de imperadores e monges e, mais recentemente, de jornais.

O ficheiro contém na primeira coluna o ano da observação, e na segunda coluna o dia do ano em que o florescimento ocorreu. Por exemplo, o valor 103 significa 103 dias depois de 1 de janeiro, ou seja, 13 de abril.

As abcissas destes gráficos são os valores dos anos. As ordenadas do gráfico de pontos são os dias passados desde o início do ano.

Ficheiro `all_month.csv` Este ficheiro é utilizado em alguns exemplos das aulas teóricas. Este conjunto de dados descreve a atividade sísmica durante um mês. A data encontra-se na coluna `time`, e a magnitude dos sismos na coluna `mag`.

Para este conjunto de dados, as abcissas são os minutos desde o início do mês. As ordenadas do gráfico de pontos são a média das magnitudes para cada minuto. Para o cálculo do minuto são considerados os dias, horas e minutos decorridos desde o início do ano, pelo que o valor é do tipo de dados `int`.

Gráficos O objetivo é implementar duas funções, uma para cada ficheiro: `sakura` e `sismos`. Cada uma destas funções deve chamar outras funções genéricas para ler o ficheiro, converter tipos de dados, preparar o conjunto de dados, e desenhar os gráficos.

Devem ser implementadas as seguintes funções:

1. `limpa_converte(dados, lista_colunas, pred_filtragem, funs_converter)`
O primeiro parâmetro que a função recebe é o conjunto de dados em forma de lista de dicionários (obtido por leitura de um ficheiro csv). O segundo é a lista de *strings* referente ao cabeçalho das colunas que interessam. O terceiro é um predicado de filtragem e o quarto é uma lista de funções para conversão dos dados. A função deve devolver o conjunto de dados, isto é, uma lista de dicionários, contendo apenas as colunas que interessam para o desenho dos gráficos. Também deve filtrar as linhas sem observações (*strings* vazias), por exemplo, anos sem registo da data de florescimento das cerejeiras. Por último, deve converter as colunas para os valores convenientes—a lista de funções `funs_converter` deve ter uma função para cada coluna escolhida (isto é, as listas `lista_colunas` e `funs_converter` devem ter o mesmo comprimento).
2. `media_movel(yy, janela)`
Esta função recebe a lista dos valores das ordenadas de um conjunto de dados e a dimensão da janela usada para calcular a média móvel. Devolve uma lista com os valores calculados. Por exemplo, para o ficheiro dos sismos, `yy` é a lista das médias de magnitudes para cada minuto. A média móvel é o valor da média calculada numa janela temporal. Por exemplo, para uma janela de tamanho 30, o valor da

média móvel num ponto x_i será a média calculada para o conjunto de 30 valores que antecede x_i , isto é, os valores entre x_{i-29} e x_i , ambos incluídos. Caso não haja valores suficientes para o cálculo, deve usar o número máximo de valores possível. Por exemplo, a média móvel para x_4 , ou seja, a quinta observação, é calculada a partir dos primeiros 5 valores do conjunto de observações, x_0 a x_4 inclusive.

3. `desvio_padrao(yy, janela)` Esta função recebe a lista dos valores das ordenadas de um conjunto de dados e a dimensão da janela usada para calcular o desvio padrão, e devolve uma lista com os valores calculados. Assim como no cálculo da média móvel, o desvio padrão é calculado usando o mesmo conceito de janela. Quando não houver valores suficientes, devem ser usados o maior número de valores possível.
4. `tracar(abcissas, ordenadas, parametros, janela=30)` Esta função recebe uma lista de valores para as abcissas, uma lista de valores para as ordenadas, uma lista de parâmetros, e o valor da janela temporal. Não devolve nada. É responsável por chamar a função que calcula a média móvel e o desvio padrão, e desenhar os três gráficos. A lista de abcissas e ordenadas devem ser as listas usadas para desenhar o gráfico de pontos. A lista de parâmetros é um dicionário que pode incluir o título, as etiquetas, as cores, etc. Fica ao critério de cada um o que incluir neste dicionário. A região a sombreado é a área compreendida entre duas curvas. A linha superior é dada por $\mu + 2 * \sigma$, e a inferior é dada por $\mu - 2 * \sigma$, onde μ é a média móvel e σ o desvio padrão móvel nesse ponto. Este gráfico pode ser feito utilizando a função `plt.fill_between` ↗.
5. `sakura(ficheiro_csv)` Esta função lê dados de um ficheiro CSV, limpa e converte os dados e traça gráficos de modo a obter uma visualização semelhante à figura acima.
6. `sismos(ficheiro_csv)` Mesma coisa, mas desta vez para o ficheiro CSV com os dados dos sismos no mês de março de 2021.

Especial atenção Tome em especial atenção os seguintes pontos.

- No desenvolvimento deste trabalho devem utilizar pelo menos duas das seguintes funções de ordem superior estudadas nas aulas: **map**, **reduce**, e **filter**.
- Não pode utilizar módulos novos além dos estudados nas aulas, uma vez que podem não estar instaladas no sistema de avaliação dos vossos trabalhos.

- As regras de boas práticas de desenvolvimento de software apontam para um número máximo de cerca de 10 linhas por função. Identifique as abstrações relevantes e implemente cada uma numa função separada.
- Pode utilizar e adaptar funções que tenham sido apresentadas nas aulas.
- Cada função que escrever deve estar equipada com uma descrição em formato `docstring`, tal como sugerido nas aulas.
- Pode e deve incluir testes para as funções mais importantes; no entanto, estes não são obrigatórios.
- Não se esqueça de incluir o seu nome e número de estudante no início do ficheiro: `__author__ = Maria Lopes, 45638.`
- O vosso código será testado por um processo automatizado. É indispensável que este contenha as funções com os nomes `limpa_converte`, `media_movel`, `desvio_padrao`, e `tracar`, e que esperem exatamente o número e o tipo dos parâmetros e que devolvam exatamente um objeto do tipo de dados pedido.
- Este é um trabalho de resolução individual. Os trabalhos devem ser entregues no Moodle até às 23:59 do dia 26 de abril de 2021.
- Os trabalhos de todos os alunos serão comparados por uma aplicação de deteção de plágio em programas. Recorde o seguinte texto na secção Integridade Académica da Sinopse:
“Alunos detetados em situação de fraude ou plágio (plagiadores e plagiados) em alguma prova ficam reprovados à disciplina e serão alvo de processo disciplinar, o que levará a um registo dessa incidência no processo de aluno, podendo conduzir à suspensão letiva.”