

DCC001 – Programação de Computadores

Trabalho Prático 2

Professor: Jefersson Alex dos Santos

Monitor: Tiago Moreira Hübner Cançado Santana

Data de Entrega: 24/06/2016 – 23:55 hrs (data provável, sujeita a alterações)

Valor: 25% da nota de prática da disciplina

1. Proposta do Trabalho Prático: Gráficos com matplotlib

Nesse trabalho, você deverá desenvolver três scripts em python (respectivamente chamados de “grafico1.py”, “grafico2.py” e “grafico3.py”) para gerar gráficos semelhantes aos exemplos abaixo:

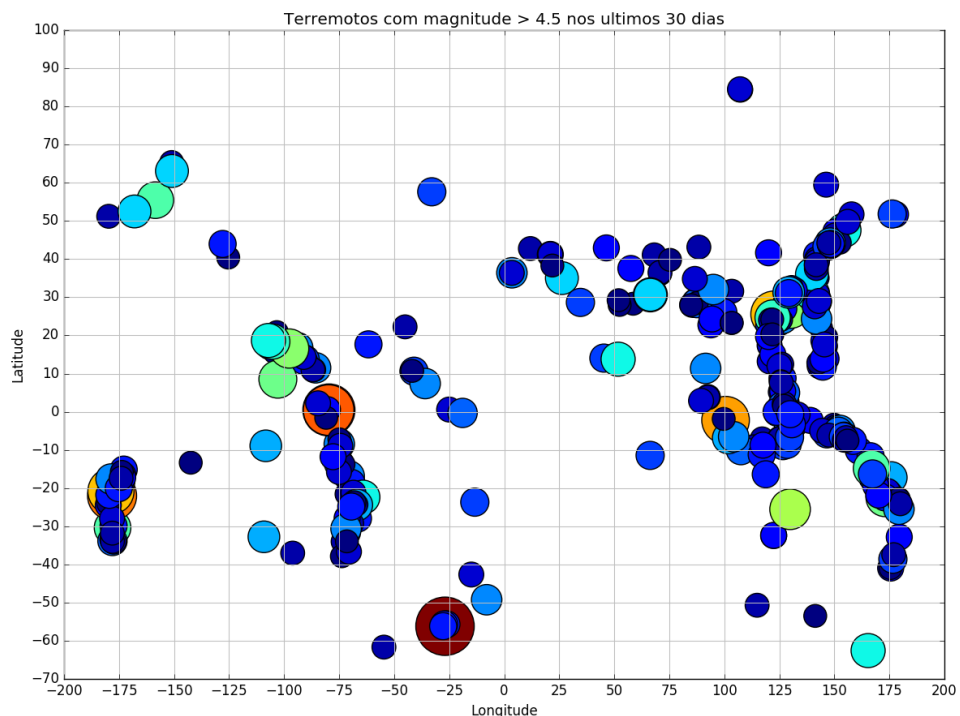


Figura 1 – Scatterplot mostrando as coordenadas dos terremotos.

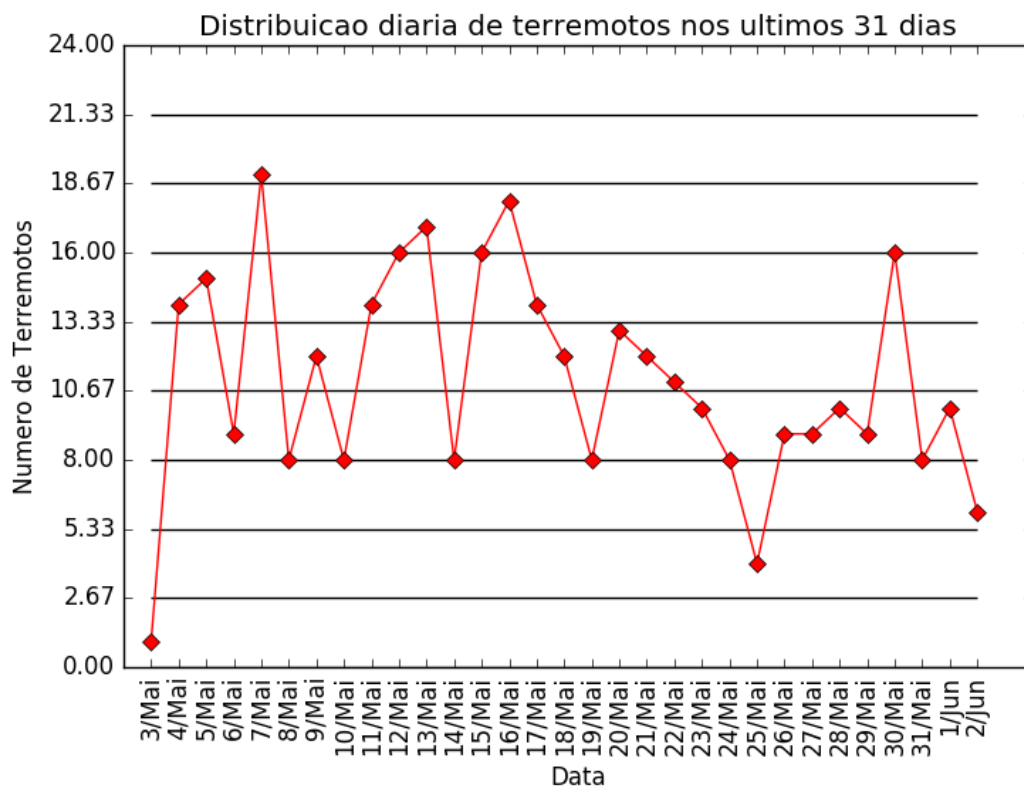


Figura 2 – Frequência diária de terremotos

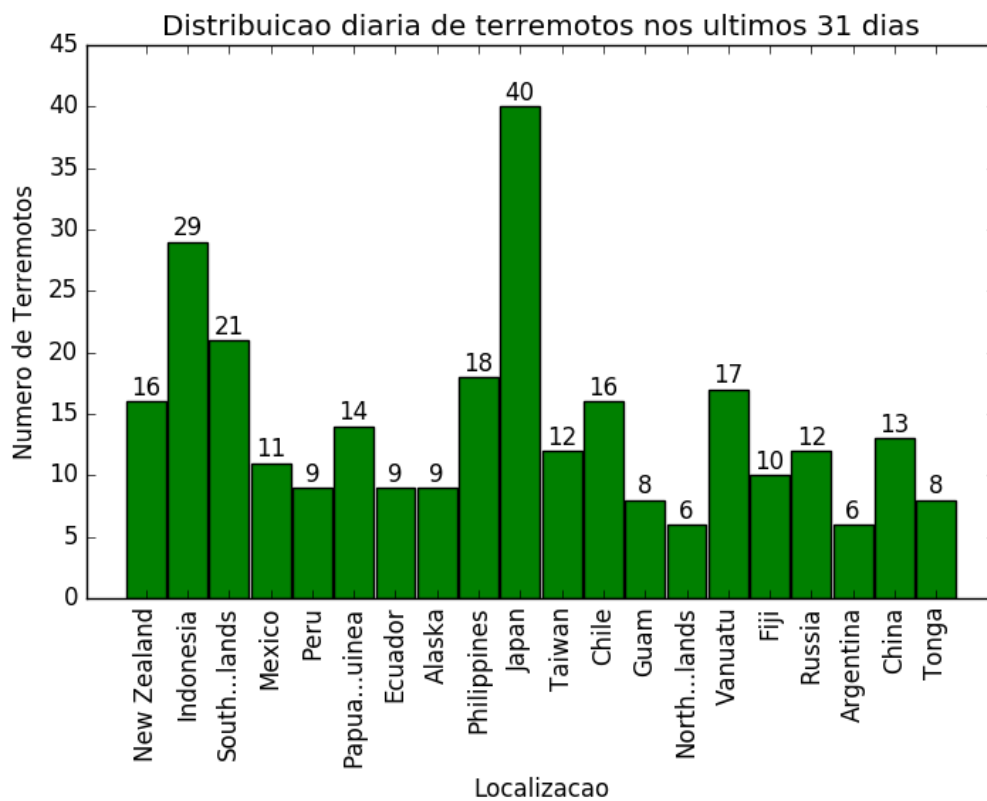


Figura 3 – Locais com maior índice de terremotos (apenas locais com mais de 5 ocorrências).

2. Dataset utilizado

Os dados utilizados para esse trabalho pertencem a um dataset real, distribuído pela USGS (serviço geológico norte-americano), que contém dados sobre terremotos no mundo inteiro. O *USGS Earthquake Hazards Program* é parte de um programa estabelecido pelo congresso americano em 1977. O intuito é monitorar e reportar o acontecimento de terremotos, avaliar seus impactos e ameaças, e pesquisar as causas e efeitos dos terremotos.

Eles disponibilizam diversos dados em seu website (<http://earthquake.usgs.gov/>) sobre os terremotos quase em tempo real. São muitos terremotos pequenos acontecendo a cada hora no planeta e, portanto, vamos trabalhar apenas com terremotos com magnitude maior que 4.5.

Os dados estão disponíveis em formato CSV (Comma Separated Values), que pode ser aberto utilizando um editor de planilhas (como o Microsoft Excel ou o LibreCalc, por exemplo). Entretanto, vamos lê-los diretamente do site utilizando um script em python (que estará disponível para vocês). O seu trabalho será apenas extrair os dados que você precisa. Para isso, é importante você conhecer um pouco o dataset.

Os cabeçalhos de cada uma das 22 colunas são:

time, latitude, longitude, depth, mag, magType, nst, gap, dmin, rms, net, id, updated, place, type, horizontalError, depthError, magError, magNst, status, locationSource, magSource

A maioria desses dados não será utilizada para esse trabalho. Vamos trabalhar apenas com as colunas *time, latitude, longitude, mag, e place*. Assim como o cabeçalho, os dados de cada linha estão separados por vírgulas (é por isso que o formato da tabela é chamado de CSV). Além da vírgula, procure outros separadores que sejam úteis para extrair os dados de interesse.

Importante:

1. Na coluna *place*, geralmente está escrito algo como “22 km ao sul de um lugar, País”. Em outras linhas, há apenas algo como “Sudeste da Ilha X” (sem vírgula). Quando houver a vírgula, vocês deverão pegar apenas a informação que está à direita. Quando não houver, pegue a string inteira.
2. Os dados são atualizados quase em tempo real. Isso significa que toda vez que você executar o seu script, ele vai baixar um dataset diferente. Portanto, não se preocupe em deixar o seu gráfico com os mesmos valores (altura das barras e localização dos círculos, por exemplo) do exemplo mostrado nesse enunciado. O importante é o seu script trabalhar os dados de forma adequada para gerar o gráfico de acordo com o que foi especificado.

3. Scatterplot

Você deverá gerar um scatterplot (Figura 1) dos terremotos onde o centro de cada círculo é dado pelas coordenadas (longitude, latitude) de cada tremor. O tamanho do círculo deverá ser a 4ª potência da magnitude do tremor (mag^4) e a cor é dada pela normalização da magnitude (teoricamente, a escala Richter é ilimitada, pois é logarítmica; mas, como nunca foi registrado um tremor com magnitude maior que 10, utilize esse valor; ou seja, $\text{cor} = \text{mag}/10.0$).

O scatterplot deverá ter um grid cujas linhas possuem distância 25 no eixo X e 10 no eixo Y. Todas as linhas do grid devem ser contínuas, ter largura 0.75 e cor 0.75.

Para esse gráfico, você não deverá mexer nos limites dos eixos X e Y nem em seus marcadores (durante a criação do grid esses aspectos são tratados). Utilize os valores padrão, ou seja, você não precisa especificá-los.

4. Gráfico de Linhas

O segundo gráfico (Figura 2) consiste na distribuição temporal de terremotos nos últimos 31 dias, ou seja, quantos terremotos aconteceram em cada dia. Você deverá plotar uma linha e diamantes grandes vermelhos cuja altura é definida pelo número de terremotos ocorridos naquele dia.

Você não deverá mexer nos limites do eixo Y. Já o eixo X, deverá ir de 0 a (n° de dias + 2). O eixo X deve ter 31 marcadores, que não devem começar em 0, mas sim em 1 e devem ser rotacionados em 90 graus. O eixo Y deve ter 10 marcadores, começando em 0 e indo até (maior n° de terremotos por dia + 5). Além disso, você deverá desenhar linhas horizontais em cada graduação do eixo Y, cobrindo do primeiro ao último marcador do eixo X (ou seja, sem encostar nas laterais).

Não se preocupe com a distância da área do gráfico até a borda da figura (o problema de cortar o texto das marcações dos eixos), pois ela será definida nos scripts de exemplo que estarão disponíveis no Moodle para vocês continuarem.

5. Gráfico de Barras

O último gráfico (Figura 3) é um gráfico de barras mostrando a distribuição da quantidade de terremotos por país nos últimos 31 dias. A altura de cada barra é a quantidade de terremotos ocorridos no país correspondente nos últimos 31 dias. Como muitos países possuem apenas 1 ou 2 terremotos nesse período, você deverá considerar apenas as localidades com mais de 5 terremotos no período.

Novamente, você não deverá mexer nos limites do eixo Y. O eixo X, por sua vez, deverá começar em 0 e ir até (n° de países considerados + 3). As barras

devem começar a ser plotadas a partir de 1 no eixo X, devem ser verdes e possuir largura 1.0. O eixo X deve ter 31 marcações, começando na posição 1.5 com intervalos de 1 em 1 e rotacionadas em 90 graus. Além disso, quando o nome do país for maior que 13, a string deve ser abreviada tomando as 5 primeiras letras e as 5 últimas concatenadas com 3 pontos entre elas (ex: 'abcdefghijklmno' se tornaria 'abcde...klmno'). O eixo Y deve ter 10 marcações, começando em 0 e indo até (maior nº de terremotos + 5).

Além disso, deve ser plotado um texto em cada barra, utilizando as mesmas coordenadas usadas para gerar as barras acrescidas de 0.5 em X e 0.5 em Y. O texto deverá ser um número inteiro informando a quantidade de terremotos que a altura da barra representa. Você deve passar por parâmetro o alinhamento horizontal e vertical do texto, que devem ser definidos como `ha='center'` e `va='bottom'`.

Não se preocupe com a distância da área do gráfico até a borda da figura (o problema de cortar o texto das marcações dos eixos), pois ela será definida nos scripts de exemplo que estarão disponíveis no Moodle para vocês continuarem.

6. Avaliação

O trabalho será avaliado de acordo com os gráficos gerados, considerando-se os seguintes aspectos:

1. Correto dimensionamento dos eixos X e Y
2. Quantidade e texto dos marcadores dos eixos X e Y
3. Cores
4. Título e rótulos dos eixos
5. Dados corretos, ou seja, altura, localização e dimensão das barras, linhas e círculos de acordo com os dados dos terremotos.

Cada gráfico vale 1/3 da nota do trabalho e cada um dos 5 aspectos vale 20% do valor total de cada gráfico.

7. Dicas

Consulte a documentação da biblioteca matplotlib (http://matplotlib.org/api/pyplot_summary.html) para olhar a lista de parâmetros de cada função.

Além disso, pode ser útil olhar o seguinte tutorial : <http://www.labri.fr/perso/nrougier/teaching/matplotlib/>