

Utilização da biblioteca GeoPandas para representação de dados geoespaciais no Python

Marcos Dantas Guimarães Filho

Universidade Federal da Paraíba
Cadernos de Ciência de Dados e Inteligência Artificial
cdia.ci.ufpb.br

20 de setembro de 2021

Este trabalho tem a finalidade de apresentar a biblioteca GeoPandas, de forma resumida, demonstrando apenas a sua principal qualidade: visualização para dados geoespaciais.

Assunto

GeoPandas é um módulo para o Python desenvolvido para trabalhar com dados geoespaciais. O GeoPandas utiliza como *backend* diferentes outras bibliotecas sendo as mais influentes:

- *Pandas*: análise de dados
- *Fiona*: ler e escrever arquivos geoespaciais
- *Shapely*: trata das formas geométricas
- *Pyproj*: trata dos diferentes modelos de projeção cartográfica
- *Descartes*: trata da visualização de dados, é construído em cima do *Matplotlib*
- *Geopy*: manipula geocoding, transforma endereços em tuplas de latitude e longitude
- *Rtree*: para análises espaciais simples
- *Pysal*: para análises espaciais mais complexas, e coloração de mapas
- *NumPy*: muita da parte matemática do GeoPandas é tratado com o *NumPy*



Figura 1 - Logo do GeoPandas

Como é?

Para utilizar o GeoPandas é necessário, primeiramente, um arquivo geoespacial. Nesse artigo irei utilizar como exemplo ilustrativo o arquivo `bcim_2016_21_11_2018.gpkg` obtido diretamente do [site do IBGE \(info ufs\)](#), e também irei utilizar uma [tabela csv \(wikipedia df\)](#), gerada por mim, contendo informações de páginas da [Wikipédia em português](#) relevantes a cada estado brasileiro, mais o Distrito Federal, o que será utilizado dessa tabela serão as colunas de número de caracteres, número de figuras e o estado que está sendo representado.

Irei explorar e ilustrar quais os estados brasileiros possuem maior relevância na Wikipédia através da extensão do conteúdo presente sobre cada estado na Wikipédia.

State name	Number of characters	Number of Images
Goiás	132013	85
Mato Grosso do Sul	199281	463
Paraná	341620	296
Minas Gerais	226786	421
Sergipe	46539	92

Figura 2 - Colunas do DataFrame `wikipedia_df` que serão utilizadas

O `.gpkg` contém as informações necessárias para efetuar o desenho do mapa do Brasil pelo GeoPandas. Ele contém também as seguintes informações: `nome`, `nomeabrev`, `geometriaaproximada`, `sigla`, `geocodigo`, `id_produto`, `id_elementoprodutor`, `cd_insumo_orgao`, `nr_insumo_mes`, `nr_insumo_ano`, `tx_insumo_documento`, `geometry` porém só irei utilizar neste artigo as colunas `name` e `geometry`, que é a coluna necessária para desenhar o mapa.

Então, para inicializar a análise exploratória de dados basta carregar o arquivo geoespacial para o GeoPandas da seguinte maneira:

```
import geopandas as gpd
info_ufs = gpd.read_file('bcim_2016_21_11_2018.gpkg',
                        layer='lim_unidade_federacao_a')
```

Figura 3 – Criação do DataFrame do GeoPandas através do arquivo gpkg

Com o dataframe do GeoPandas criado, contendo nosso mapa, e o nosso dataframe Pandas com os dados para a análise exploratória também criado, basta realizar uma junção (*merge*) entre os dois dataframes utilizando o método `.merge` do GeoPandas:

```
brasil = info_ufs.merge(wikipedia_df,
                        on = 'State name', how = 'left')
```

Figura 4 – Demonstração do método `.merge` do GeoPandas para unir dois DataFrames

E, finalmente, para gerar com o GeoPandas, o mapa do Brasil com as informações relevantes basta gerar um plot, de uma maneira muito similar de como seria feito com o Matplotlib:

```
brasil_com_info.plot(column = 'Number of characters',
                    cmap = 'YlOrRd',
                    figsize = (15,10),
                    legend = True,
                    edgecolor = 'black')

plt.title("Em relação ao número de Caracteres",
         fontdict={'fontsize':20})
```

Figura 5 – Código necessário para gerar um plot

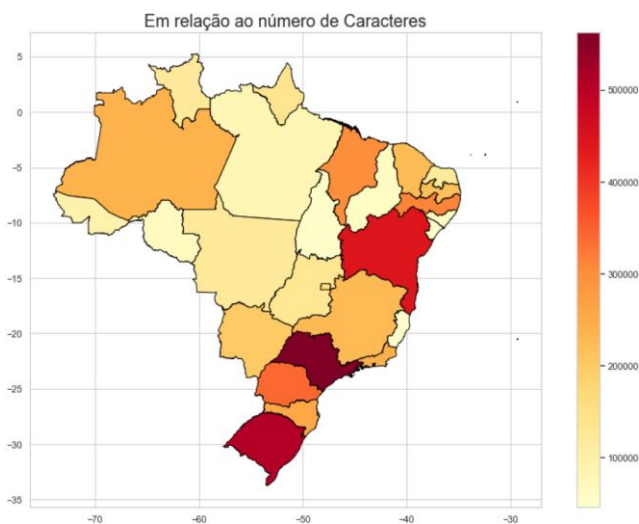


Figura 6 – Mapa gerado com o código acima, os valores que estão marcados na escala de cor são, do amarelo claro para o vermelho escuro: 10.000, 20.000, 30.000, 40.000 e 50.000

O mapa gerado acima está ilustrando a soma de número de caracteres das páginas relevantes a aquele determinado estado, a coluna de contraste ao lado representa o número correspondente para cada coloração, e os eixos das abcissas e ordenadas no gráfico estão representando, respectivamente, as linhas de latitudes e longitudes do globo terrestre.

O mesmo código pode ser utilizado para gerar outro mapa, porém dessa vez ao invés de ser visualizado o mapa em relação ao número de caracteres podemos visualizar em relação ao número de figuras:

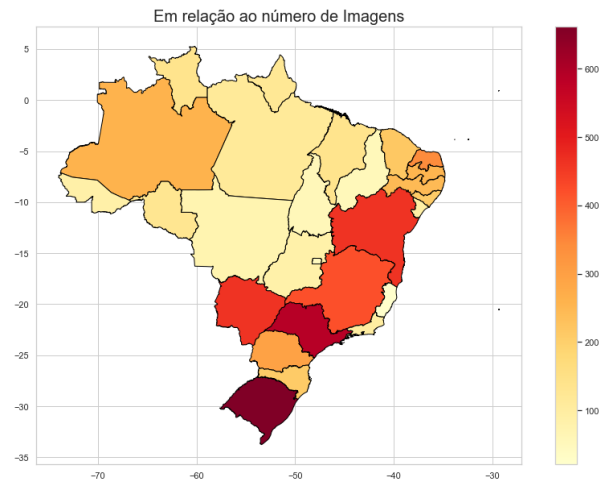


Figura 7 – A coluna de contraste marca os valores, do mais claro para o mais escuro: 100. 200. 300. 400. 500 e 600

Considerações finais

Obviamente, por ser uma ferramenta criada com base no Matplotlib há uma extensiva possibilidade de customização dos gráficos gerados, porém até os gráficos mais simples gerados pelo GeoPandas são capazes de transmitir as informações com clareza.

Apesar de breve, acredito que foi possível demonstrar com clareza a principal característica do GeoPandas: a representação geoespacial de maneira rápida e simples.