1.3-exploratory_analysis_fancy_plot

October 20, 2020

1 Imports

[1]:
```
import sys
import os
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
```

[2]:
```
project_path = os.path.abspath(os.path.join('..'))

if project_path not in sys.path:
    sys.path.append(f'{project_path}/src/visualizations/')

from covid_data_viz import CovidDataViz
```

2 Setup

[3]:
```
mpl.rcParams['figure.figsize'] = (9, 5)
```

3 Goal

My goal is to visualize various aspect of the `COVID-19` pandemic.

4 Data sources

In this project I use data from the following sources: - https://github.com/CSSEGISandData/COVID-19 - JHU CSSE COVID-19 Data. - https://datahub.io/JohnSnowLabs/country-and-continent-codes-list - country codes and continents.

5 Data loading

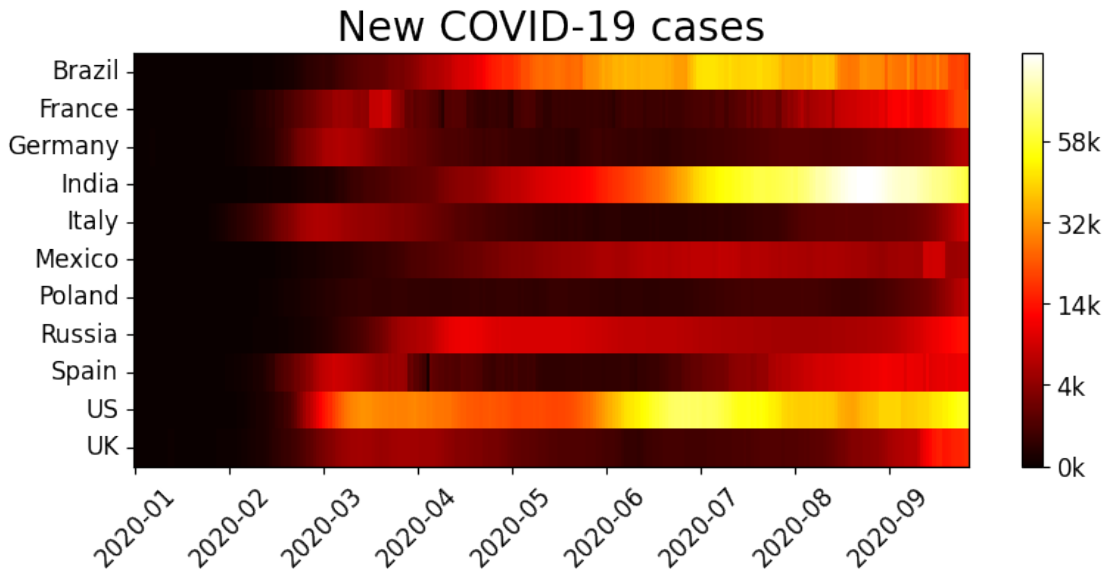
[4]:
```
cdv = CovidDataViz()
```

6 Fancy plot

Visual for repo readme.

[5]:
```
countries = ['Germany',
             'France',
             'Italy',
             'Spain',
             'United Kingdom',
             'Russia',
             'India',
             'Brazil',
             'US',
             'Poland',
             'Mexico']

width = 825
height = width / 2
dpi = 100
period = 7
step = 30
label_size = 12
n_clabels = 6

countries = sorted(countries)

plot_df = cdv.data['Confirmed chg'][countries]
plot_df = plot_df.rename(columns={'United Kingdom': 'UK'})
countries = plot_df.columns.to_list()
plot_df = plot_df.rolling(period)
plot_df = plot_df.mean()
plot_df = plot_df.dropna()
plot_df = plot_df.to_numpy()
plot_df = plot_df.astype(float)
plot_df = plot_df.transpose()
plot_df = np.sqrt(plot_df)

xticks = range(plot_df.shape[1])[::step]
xlabels = list(cdv.data['Confirmed chg']['Date'])[period:]
xlabels = [x.strftime(format='%Y-%m') for x in xlabels]
# xlabels = [x.date() for x in xlabels]
xlabels = xlabels[::step]

yticks = range(len(countries))
ylabels = countries

cticks = np.round(np.linspace(0, np.max(plot_df), 6), -1)
cticks = cticks.astype(np.int)
```

```
clabels = np.power(cticks, 2)
cticks = sorted(set(cticks))
clabels = np.power(cticks, 2)
clabels = [int((round(x, -3))/1000) for x in clabels]
clabels = [str(x)+'k' for x in clabels]
# clabels = list(map(str, clabels))

plt.figure(figsize=(width / dpi, height / dpi), dpi=dpi)
plt.imshow(plot_df, aspect='auto', interpolation='nearest')
plt.set_cmap('hot')

plt.yticks(ticks=yticks,
           labels=ylabels,
           fontsize=label_size,
           verticalalignment='center')

plt.xticks(ticks=xticks,
           labels=xlabels,
           rotation=45,
           fontsize=label_size,
           horizontalalignment='center')

cbar = plt.colorbar()
cbar.set_ticks(cticks)
cbar.set_ticklabels(clabels)
cbar.ax.tick_params(labelsize=label_size)

plt.title('New COVID-19 cases', fontsize=20)

plt.tight_layout()
plt.savefig('../img/covid_tiles.png', bbox_inches='tight')

plt.show()
```


New COVID-19 cases
Brazil
France
Germany
India
Italy
Mexico
Poland
Russia
Spain
US
UK
2020-01
2020-02
2020-03
2020-04
2020-05
2020-06
2020-07
2020-08
2020-09
58k
32k
14k
4k
0k