covid19-notebook

December 30, 2020

1 Visualisation of COVID-19 datas

The goal of this notebook is to display on a Choropleth map informations related to COVID-19 at different moment in time. We will use the open source dataset maintained by *Our World in Data*.

1.0.1 Table of content

1. Fetching the dataset
2. Understanding and preparation of the data
3. Building the Choropleth map

1.1 1. Fetching the dataset

First of all we need to import different libraries in order to get the datas, work with them and visualize them.

[1]:

```
import numpy as np # Library to handle data in a vectorized manner
import pandas as pd # Library for data analsysis
import branca.colormap as cmp
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
import json # Library to handle JSON files

!pip3 install folium # Install the folium library
import folium # Map rendering library
```

```
Requirement already satisfied: folium in /usr/local/lib/python3.7/site-packages
(0.11.0)
Requirement already satisfied: jinja2>=2.9 in /usr/local/lib/python3.7/site-
packages (from folium) (2.10.3)
Requirement already satisfied: requests in /usr/local/lib/python3.7/site-
packages (from folium) (2.22.0)
Requirement already satisfied: branca>=0.3.0 in /usr/local/lib/python3.7/site-
packages (from folium) (0.4.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.7/site-packages
(from folium) (1.18.1)
Requirement already satisfied: MarkupSafe>=0.23 in
/usr/local/lib/python3.7/site-packages (from jinja2>=2.9->folium) (1.1.1)
Requirement already satisfied: idna<2.9,>=2.5 in /usr/local/lib/python3.7/site-
```

```
packages (from requests->folium) (2.8)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.7/site-packages (from requests->folium) (2019.11.28)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in
/usr/local/lib/python3.7/site-packages (from requests->folium) (1.25.8)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in
/usr/local/lib/python3.7/site-packages (from requests->folium) (3.0.4)
```

Now that all necessary libraries have been imported, we need to put the dataset into a pandas dataframe. To do so we will use the read_csv function offered by pandas to get datas directly from the link of the repository.

[2]:

```
covid_df = pd.read_csv('https://covid.ourworldindata.org/data/owid-covid-data.
 ↪csv')
print('Dataset successfully downloaded !')
```

```
Dataset successfully downloaded !
```

Now that the dataset have been downloaded and transformed into a dataframe, we analyse the content of the dataframe.

[3]:

```
covid_df.head()
```

[3]:

```
  iso_code continent     location        date  total_cases  new_cases  \
0      AFG      Asia  Afghanistan  2020-02-24          1.0        1.0
1      AFG      Asia  Afghanistan  2020-02-25          1.0        0.0
2      AFG      Asia  Afghanistan  2020-02-26          1.0        0.0
3      AFG      Asia  Afghanistan  2020-02-27          1.0        0.0
4      AFG      Asia  Afghanistan  2020-02-28          1.0        0.0

   new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
0                 NaN           NaN         NaN                  NaN
1                 NaN           NaN         NaN                  NaN
2                 NaN           NaN         NaN                  NaN
3                 NaN           NaN         NaN                  NaN
4                 NaN           NaN         NaN                  NaN

   total_cases_per_million  new_cases_per_million  \
0                    0.026                  0.026
1                    0.026                  0.000
2                    0.026                  0.000
3                    0.026                  0.000
4                    0.026                  0.000

   new_cases_smoothed_per_million  total_deaths_per_million  \
0                             NaN                       NaN
1                             NaN                       NaN
2                             NaN                       NaN
3                             NaN                       NaN
```

```
4                             NaN                       NaN

   new_deaths_per_million  new_deaths_smoothed_per_million  reproduction_rate  \
0                     NaN                              NaN                NaN
1                     NaN                              NaN                NaN
2                     NaN                              NaN                NaN
3                     NaN                              NaN                NaN
4                     NaN                              NaN                NaN

   icu_patients  icu_patients_per_million  hosp_patients  \
0           NaN                       NaN            NaN
1           NaN                       NaN            NaN
2           NaN                       NaN            NaN
3           NaN                       NaN            NaN
4           NaN                       NaN            NaN

   hosp_patients_per_million  weekly_icu_admissions  \
0                        NaN                    NaN
1                        NaN                    NaN
2                        NaN                    NaN
3                        NaN                    NaN
4                        NaN                    NaN

   weekly_icu_admissions_per_million  weekly_hosp_admissions  \
0                                NaN                     NaN
1                                NaN                     NaN
2                                NaN                     NaN
3                                NaN                     NaN
4                                NaN                     NaN

   weekly_hosp_admissions_per_million  new_tests  total_tests  \
0                                 NaN        NaN          NaN
1                                 NaN        NaN          NaN
2                                 NaN        NaN          NaN
3                                 NaN        NaN          NaN
4                                 NaN        NaN          NaN

   total_tests_per_thousand  new_tests_per_thousand  new_tests_smoothed  \
0                       NaN                     NaN                 NaN
1                       NaN                     NaN                 NaN
2                       NaN                     NaN                 NaN
3                       NaN                     NaN                 NaN
4                       NaN                     NaN                 NaN

   new_tests_smoothed_per_thousand  positive_rate  tests_per_case tests_units  \
0                              NaN            NaN             NaN         NaN
1                              NaN            NaN             NaN         NaN
```

```
2                              NaN            NaN             NaN         NaN
3                              NaN            NaN             NaN         NaN
4                              NaN            NaN             NaN         NaN

   total_vaccinations  total_vaccinations_per_hundred  stringency_index  \
0                 NaN                             NaN              8.33
1                 NaN                             NaN              8.33
2                 NaN                             NaN              8.33
3                 NaN                             NaN              8.33
4                 NaN                             NaN              8.33

   population  population_density  median_age  aged_65_older  aged_70_older  \
0  38928341.0              54.422        18.6          2.581          1.337
1  38928341.0              54.422        18.6          2.581          1.337
2  38928341.0              54.422        18.6          2.581          1.337
3  38928341.0              54.422        18.6          2.581          1.337
4  38928341.0              54.422        18.6          2.581          1.337

   gdp_per_capita  extreme_poverty  cardiovasc_death_rate  \
0        1803.987              NaN                597.029
1        1803.987              NaN                597.029
2        1803.987              NaN                597.029
3        1803.987              NaN                597.029
4        1803.987              NaN                597.029

   diabetes_prevalence  female_smokers  male_smokers  handwashing_facilities  \
0                 9.59             NaN           NaN                  37.746
1                 9.59             NaN           NaN                  37.746
2                 9.59             NaN           NaN                  37.746
3                 9.59             NaN           NaN                  37.746
4                 9.59             NaN           NaN                  37.746

   hospital_beds_per_thousand  life_expectancy  human_development_index
0                         0.5            64.83                    0.498
1                         0.5            64.83                    0.498
2                         0.5            64.83                    0.498
3                         0.5            64.83                    0.498
4                         0.5            64.83                    0.498
```

1.2 2. Understanding and preparation of the data

Now that we have correctly save the dataset we need to keep only few columns for this representation. We only want to display the total cases in each country over time. To do so we extract 5 columns: - iso_code - continent - location - date - total_cases

The others columns could be very useful in a more advanced context were the cause of contamination could be searched.

[4]:

```
simp_covid_df = covid_df[['iso_code', 'continent', 'location', 'date',␣
 ↪'total_cases']]
simp_covid_df['location'].unique()

# We don't want the general number of case so we remove the 'World' entry
simp_covid_df = simp_covid_df[simp_covid_df.location != 'World']
```

Since we only want to check a specific date we need to create a new DataFrame with only the values of the desired day. Here we choose the christmas day.

[5]:

```
xmas_df = simp_covid_df[simp_covid_df['date'] == '2020-12-25']
xmas_df.head(50)
```

[5]:

```
      iso_code      continent                  location        date  \
305        AFG           Asia               Afghanistan  2020-12-25
601        ALB         Europe                   Albania  2020-12-25
910        DZA         Africa                   Algeria  2020-12-25
1213       AND         Europe                   Andorra  2020-12-25
1498       AGO         Africa                    Angola  2020-12-25
1790       ATG  North America       Antigua and Barbuda  2020-12-25
2122       ARG  South America                 Argentina  2020-12-25
2426       ARM           Asia                   Armenia  2020-12-25
2765       AUS        Oceania                 Australia  2020-12-25
3074       AUT         Europe                   Austria  2020-12-25
3379       AZE           Asia                Azerbaijan  2020-12-25
3668       BHS  North America                   Bahamas  2020-12-25
3978       BHR           Asia                   Bahrain  2020-12-25
4280       BGD           Asia                Bangladesh  2020-12-25
4568       BRB  North America                  Barbados  2020-12-25
4874       BLR         Europe                   Belarus  2020-12-25
5204       BEL         Europe                   Belgium  2020-12-25
5486       BLZ  North America                    Belize  2020-12-25
5775       BEN         Africa                     Benin  2020-12-25
6074       BTN           Asia                    Bhutan  2020-12-25
6368       BOL  South America                   Bolivia  2020-12-25
6668       BIH         Europe    Bosnia and Herzegovina  2020-12-25
6943       BWA         Africa                  Botswana  2020-12-25
7251       BRA  South America                    Brazil  2020-12-25
7547       BRN           Asia                    Brunei  2020-12-25
7844       BGR         Europe                  Bulgaria  2020-12-25
8139       BFA         Africa              Burkina Faso  2020-12-25
8413       BDI         Africa                   Burundi  2020-12-25
8751       KHM           Asia                  Cambodia  2020-12-25
9050       CMR         Africa                  Cameroon  2020-12-25
9389       CAN  North America                    Canada  2020-12-25
9674       CPV         Africa                Cape Verde  2020-12-25
9964       CAF         Africa  Central African Republic  2020-12-25
```

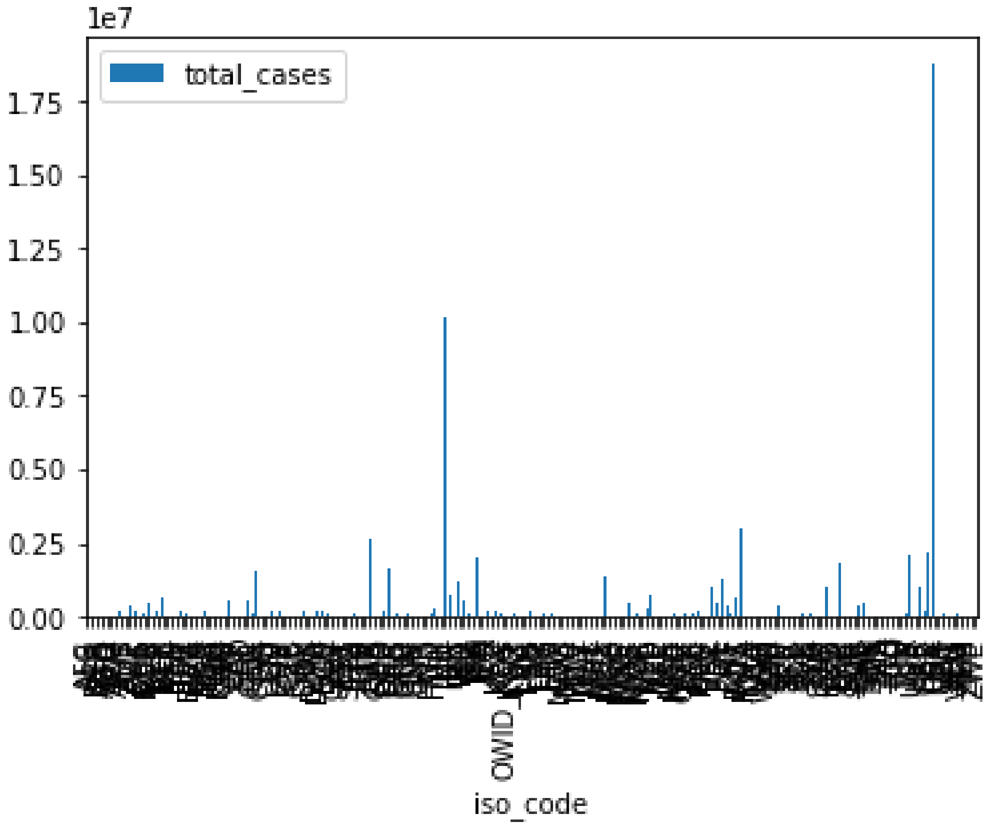
```
10250      TCD         Africa                          Chad  2020-12-25
10561      CHL  South America                         Chile  2020-12-25
10904      CHN           Asia                         China  2020-12-25
11203      COL  South America                      Colombia  2020-12-25
11447      COM         Africa                       Comoros  2020-12-25
11737      COG         Africa                         Congo  2020-12-25
12036      CRI  North America                    Costa Rica  2020-12-25
12330      CIV         Africa                 Cote d'Ivoire  2020-12-25
12639      HRV         Europe                       Croatia  2020-12-25
12932      CUB  North America                          Cuba  2020-12-25
13228      CYP         Europe                        Cyprus  2020-12-25
13566      CZE         Europe                       Czechia  2020-12-25
13860      COD         Africa  Democratic Republic of Congo  2020-12-25
14198      DNK         Europe                       Denmark  2020-12-25
14485      DJI         Africa                      Djibouti  2020-12-25
14768      DMA  North America                      Dominica  2020-12-25
15072      DOM  North America            Dominican Republic  2020-12-25

       total_cases
305        50810.0
601        55380.0
910        97441.0
1213        7756.0
1498       17099.0
1790         155.0
2122     1574554.0
2426      156763.0
2765       28297.0
3074      349055.0
3379      211764.0
3668        7788.0
3978       91304.0
4280      507265.0
4568         347.0
4874      183006.0
5204      637246.0
5486       10490.0
5775        3205.0
6074         576.0
6368      153590.0
6668      108891.0
6943       14025.0
7251     7448560.0
7547         152.0
7844      196915.0
8139        6134.0
8413         786.0
```

```
8751          363.0
9050        26277.0
9389       540939.0
9674        11696.0
9964         4948.0
10250        1971.0
10561      595831.0
10904       95460.0
11203     1574707.0
11447         715.0
11737        6571.0
12036      162990.0
12330       22081.0
12639      203962.0
12932       10900.0
13228       19366.0
13566      664863.0
13860       16472.0
14198      149926.0
14485        5804.0
14768          88.0
15072      165035.0
```

It seems to have large gaps between total cases. This will be an issue in the choropleth map because there will not be enough color shade and thus, the map will not have much interest. Let's check using a plot to make sure it's not only on the 50 first rows.

[6]:
```
xmas_df.plot.bar(x='iso_code', y='total_cases')
```

[6]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x126b9ea50>
```

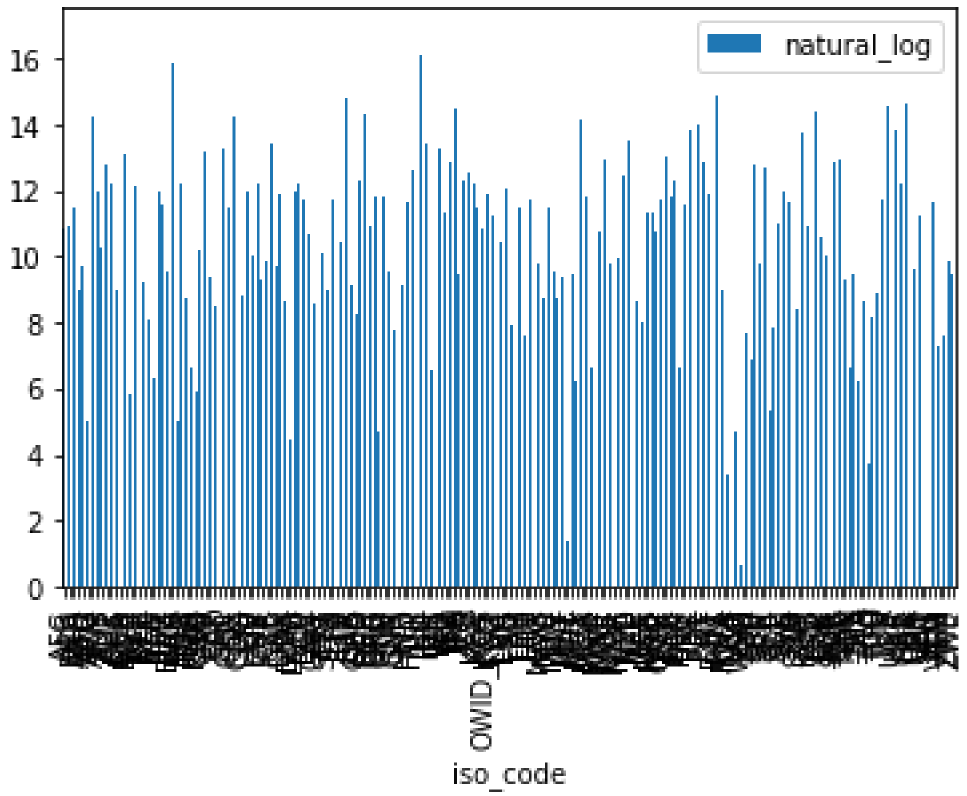


The legend on the plot is not really understandable but what we can see is that the total number of cases have large gaps. In order to solve the problem for the representation we will create a new column with the natural logarithm value of the total_cases in order to normalise the data and see the variations on the map.

```
[7]: xmas_df['natural_log'] = np.log(xmas_df['total_cases'])
     xmas_df.plot.bar(x='iso_code', y='natural_log')
```

```
/usr/local/lib/python3.7/site-packages/ipykernel_launcher.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  """Entry point for launching an IPython kernel.
```

```
[7]: <matplotlib.axes._subplots.AxesSubplot at 0x127719d10>
```



Now that the data have been normalised and that the values seems more close to each other, let's create the choropleth map in order to visualise the evolution of COVID-19 accross the world.

1.3 3. Building the Choropleth map

We isolated the wanted columns but now we need to display it on a Choropleth map. First we will do a fixed map with the datas on the day of christmas and then we will do an interactive map with all the available datas from February 24th 2020 to today.

To do so we need to get a GeoJSON file in order to link data to countries on the map. This open source file is perfect for what we are looking for. An other solution would have to generate the map using this website.

```
[8]: !wget --quiet 'https://raw.githubusercontent.com/johan/world.geo.json/master/
      ↪countries.geo.json'

     print('Geo JSON file downloaded !')
```

```
Geo JSON file downloaded !
```

Now that we have the GeoJSON file, let's create a world map centered on the coordinates [0,0].

```
[9]: world_geo = r'countries.geo.json'
```

And now to create the `Choropleth` map, we will use the *choropleth* method with the following main parameters:

1. world_geo, which is the GeoJSON file.
2. xmas_df, which is the dataframe containing the data.
3. columns, which represents the columns in the dataframe that will be used to create the `Choropleth` map.
4. key_on, which is the key or variable in the GeoJSON file that contains the name of the variable of interest.

[10]:
```
static_map = folium.Map(location=[0, 0], zoom_start=2)

folium.Choropleth(
    geo_data=world_geo,
    data=xmas_df,
    columns=['iso_code', 'natural_log'],
    key_on='feature.id',
    fill_color='OrRd',
    fill_opacity=0.7,
    line_opacity=0.2,
    legend_name='Number of COVID Cases in the world on Christmas 2020'
).add_to(static_map)
```

[10]:
```
<folium.features.Choropleth at 0x127f15a90>
```

[12]:
```
static_map
```

[12]:
```
<folium.folium.Map at 0x127f15050>
```

As we can see from the map, it's mostly the countries in the north of the emispherere that are affected by the virus. It can be explained by different factors. Either it's because of the temperature because she is lower in the north but we can see that in south america there is a large amount of case so it's not only this. Since the virus is more effective in the cold it's a tangible explanation. An other possibility is that the number of test performed and the population amount of the countries are inequal so the number are biased.

Either way we can see that the virus is more present in richer country where the health cares are better. It shows that the life lived by those richer population are not optimal to defeat the virus. It also demonstrate that poorer countries can't be tested enough.

[ ]:
```
```