NIA_IRP_COVID19_dataviz_workshop

March 9, 2021

1 Interactive COVID-19 data visualizations using Python & Bokeh

1.0.1 Chris Coletta

1.0.2 Friday, January 29, 2021

2 Goals

- Reproduce many figures from the New York Times U.S. tracking page and other sources
  - "Epidemic curve": Cumulative cases and deaths plotted by day
  - "Chloropleth Map": New cases and deaths plotted by state
- Answer some important questions:
  - Maryland: What is the county-by-county situation?
  - US: Is the increase in cases coming from an increase in testing, or from actual virus spread?

3 Analysis Strategy

- Use one or more of the widely available APIs that provide COVID-19 data
- Download them programmatically so figures are up to date every time you run this notebook

3.1 NIA IRP Data Viz workshop

- Close all anaconda windows
- Open up anaconda prompt
- Run the first few cells to make sure you have everything installed
  - `!pip install pandas-bokeh`
  - `!pip install geopandas`
    - Requires GDAL (On Mac: `brew install gdal`)
- RAISE YOUR HAND IN ZOOM IF YOU HAVE DONE ALL OF THE ABOVE
- The notebook we will use today is in the Zoom chat box
- Two options for running:
  - Download and run locally.
  - OR run this notebook in Google CoLab
- Workshop participants are expected to run the code along with me
- Select `Kernel->Restart & Clear Output` to start

- Have my presentation side-by-side with your Jupyter Notebook open
- Unmute yourself if you have questions

- Can also use Jupyterlab, but you must run additional command and restart JupyterLab
  - `jupyter labextension install @jupyter-widgets/jupyterlab-manager`
  - `jupyter labextension install @bokeh/jupyter_bokeh`

4 What we won't cover

- Effective reproduction number
  - rt.live - state-by-state COVID-19 transmission rate R_t modelling from the makers of Instagram

- More sophisticated epidemic modelling in general
  - Please refer to John Holmes' brilliant talk AI in the Age of COVID-19: Computational Tools for a Pandemic) from yesterday: "AI in the Age of COVID-19: Computational Tools for a Pandemic"

- Information retrieval from corpus of published and pre-published scientific manuscripts adressing coronavirus

5 Available Data Sources

- NYTimes COVID-19 data on GitHub
  - Cases and deaths by US County
  - Excess deaths by time series data country

- The COVID Tracking Project at The Atlantic
  - Data broken down by race and other demographic factors/social determinants

- CSSEGISandData COVID-19 data on GitHub
  - COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University
  - Includes normalized case, testing, and mortality rate per 100k

- Humanitarian Data Exchange
  - UN website
  - COVID *testing* data by country

6 Load the necessary libraries

- We use the Pandas DataFrame as a container for the data, and also to load the data from internet URLs
- We use pandas-bokeh to create the interactive plots.
- We use GeoPandas to make the maps
- Optional: we use requests to download state and county shapes from the US census website

[1]:
```
import pandas as pd

pd.set_option( 'display.max_rows', 100 )
```

```
pd.set_option( 'display.max_columns', 100 )
```

[2]:
```
!pip install pandas_bokeh
```

```
Requirement already satisfied: pandas_bokeh in /usr/local/lib/python3.9/site-
packages (0.5.3)
Requirement already satisfied: pandas>=0.22.0 in /usr/local/lib/python3.9/site-
packages (from pandas_bokeh) (1.2.0)
Requirement already satisfied: bokeh>=2.0 in /usr/local/lib/python3.9/site-
packages (from pandas_bokeh) (2.2.3)
Requirement already satisfied: typing-extensions>=3.7.4 in
/usr/local/lib/python3.9/site-packages (from bokeh>=2.0->pandas_bokeh) (3.7.4.3)
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.9/site-
packages (from bokeh>=2.0->pandas_bokeh) (5.3.1)
Requirement already satisfied: packaging>=16.8 in /usr/local/lib/python3.9/site-
packages (from bokeh>=2.0->pandas_bokeh) (20.8)
Requirement already satisfied: Jinja2>=2.7 in /usr/local/lib/python3.9/site-
packages (from bokeh>=2.0->pandas_bokeh) (2.11.2)
Requirement already satisfied: python-dateutil>=2.1 in
/usr/local/lib/python3.9/site-packages (from bokeh>=2.0->pandas_bokeh) (2.8.1)
Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.9/site-
packages (from bokeh>=2.0->pandas_bokeh) (6.1)
Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.9/site-
packages (from bokeh>=2.0->pandas_bokeh) (8.1.0)
Requirement already satisfied: numpy>=1.11.3 in /usr/local/lib/python3.9/site-
packages (from bokeh>=2.0->pandas_bokeh) (1.19.5)
Requirement already satisfied: MarkupSafe>=0.23 in
/usr/local/lib/python3.9/site-packages (from
Jinja2>=2.7->bokeh>=2.0->pandas_bokeh) (1.1.1)
Requirement already satisfied: pyparsing>=2.0.2 in
/usr/local/lib/python3.9/site-packages (from
packaging>=16.8->bokeh>=2.0->pandas_bokeh) (2.4.7)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.9/site-
packages (from pandas>=0.22.0->pandas_bokeh) (2020.5)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.9/site-
packages (from python-dateutil>=2.1->bokeh>=2.0->pandas_bokeh) (1.15.0)
```

[3]:
```
import pandas_bokeh

#from bokeh.layouts import column, row
#from bokeh.models import CustomJS, Slider, Dropdown
#from bokeh.plotting import ColumnDataSource, figure, show
#from bokeh.io import output_notebook, output_file
```

[4]:
```
pandas_bokeh.output_notebook()
```

[ ]:
```
!pip install geopandas
```

[ ]:
```
import geopandas as gpd
```

[ ]:
```
import requests
```

[5]:
```
import numpy as np
```

7 Part 1: US Epidemic curves

- Rate of transmission is the slope of the epidemic curve
- Examples
  - New York Times - U.S. tracking page
  - AP news - COVID-19 tracker with state pulldown
  - Visual Capitalist - Flattening the Curve by Country

8 Download state-wide COVID time series

[6]:
```
state_url = "https://covidtracking.com/api/v1/states/daily.csv"
```

[7]:
```
state_data = pd.read_csv( state_url )
```

[8]:
```
type( state_data )
```

[8]:
```
pandas.core.frame.DataFrame
```

8.1 Quick Review of PANDAS DataFrame capabilities

- Use period and tab key to get pop up list of methods and attributes
- Can always read the manual for each function by using the question mark

[ ]:
```
state_data.
```

[9]:
```
state_data.shape
```

[9]:
```
(18642, 55)
```

[10]:
```
state_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18642 entries, 0 to 18641
Data columns (total 55 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   date                        18642 non-null  int64
 1   state                       18642 non-null  object
 2   positive                    18456 non-null  float64
 3   probableCases               8032 non-null   float64
 4   negative                    14849 non-null  float64
 5   pending                     1949 non-null   float64
```

```
6   totalTestResultsSource       18642 non-null  object
7   totalTestResults             18540 non-null  float64
8   hospitalizedCurrently        15324 non-null  float64
9   hospitalizedCumulative       11447 non-null  float64
10  inIcuCurrently               9938 non-null   float64
11  inIcuCumulative              3336 non-null   float64
12  onVentilatorCurrently        7906 non-null   float64
13  onVentilatorCumulative       1141 non-null   float64
14  recovered                    13332 non-null  float64
15  dataQualityGrade             17323 non-null  object
16  lastUpdateEt                 18074 non-null  object
17  dateModified                 18074 non-null  object
18  checkTimeEt                  18074 non-null  object
19  death                        17774 non-null  float64
20  hospitalized                 11447 non-null  float64
21  dateChecked                  18074 non-null  object
22  totalTestsViral              12428 non-null  float64
23  positiveTestsViral           7122 non-null   float64
24  negativeTestsViral           4253 non-null   float64
25  positiveCasesViral           13271 non-null  float64
26  deathConfirmed               8324 non-null   float64
27  deathProbable                6547 non-null   float64
28  totalTestEncountersViral     4367 non-null   float64
29  totalTestsPeopleViral        8202 non-null   float64
30  totalTestsAntibody           4162 non-null   float64
31  positiveTestsAntibody        3188 non-null   float64
32  negativeTestsAntibody        1313 non-null   float64
33  totalTestsPeopleAntibody     1629 non-null   float64
34  positiveTestsPeopleAntibody  942 non-null    float64
35  negativeTestsPeopleAntibody  858 non-null    float64
36  totalTestsPeopleAntigen      809 non-null    float64
37  positiveTestsPeopleAntigen   606 non-null    float64
38  totalTestsAntigen            2640 non-null   float64
39  positiveTestsAntigen         1658 non-null   float64
40  fips                         18642 non-null  int64
41  positiveIncrease             18642 non-null  int64
42  negativeIncrease             18642 non-null  int64
43  total                        18642 non-null  int64
44  totalTestResultsIncrease     18642 non-null  int64
45  posNeg                       18642 non-null  int64
46  deathIncrease                18642 non-null  int64
47  hospitalizedIncrease         18642 non-null  int64
48  hash                         18642 non-null  object
49  commercialScore              18642 non-null  int64
50  negativeRegularScore         18642 non-null  int64
51  negativeScore                18642 non-null  int64
52  positiveScore                18642 non-null  int64
53  score                        18642 non-null  int64
```

```
 54  grade                         0 non-null      float64
dtypes: float64(33), int64(14), object(8)
memory usage: 7.8+ MB
```

[11]:
```
state_data.head()
```

[11]:

	date	state	positive	probableCases	negative	pending	totalTestResultsSource	totalTestResults	hospitalizedCurrently	hospitalizedCumulative	inIcuCurrently	inIcuCumulative	onVentilatorCurrently	onVentilatorCumulative	recovered	dataQualityGrade	lastUpdateEt	dateModified	checkTimeEt	death	hospitalized	dateChecked	totalTestsViral	positiveTestsViral
0	20210128	AK	52150.0	NaN	NaN	NaN	totalTestsViral	1479849.0	42.0	1201.0	NaN	NaN	5.0	NaN	NaN	A	1/28/2021 03:59	2021-01-28T03:59:00Z	01/27 22:59	262.0	1201.0	2021-01-28T03:59:00Z	1479849.0	62705.0
1	20210128	AL	452734.0	95263.0	1759171.0	NaN	totalTestsPeopleViral	2116642.0	2052.0	41483.0	NaN	2533.0	NaN	1449.0	242143.0	A	1/28/2021 11:00	2021-01-28T11:00:00Z	01/28 06:00	7340.0	41483.0	2021-01-28T11:00:00Z	NaN	NaN
2	20210128	AR	290856.0	58518.0	2161833.0	NaN	totalTestsViral	2394171.0	996.0	13505.0	307.0	NaN	143.0	1414.0	268495.0	A+	1/28/2021 00:00	2021-01-28T00:00:00Z	01/27 19:00	4784.0	13505.0	2021-01-28T00:00:00Z	2394171.0	NaN
3	20210128	AS	0.0	NaN	2140.0	NaN	totalTestsViral	2140.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	12/1/2020 00:00	2020-12-01T00:00:00Z	11/30 19:00	0.0	NaN	2020-12-01T00:00:00Z	2140.0	NaN
4	20210128	AZ	743232.0	46684.0	2712847.0	NaN	totalTestsViral	6549568.0	4087.0	51349.0	1023.0	NaN	692.0	NaN	101104.0	A+	1/28/2021 00:00	2021-01-28T00:00:00Z	01/27 19:00	12819.0	51349.0	2021-01-28T00:00:00Z	6549568.0	NaN

```
   negativeTestsViral  positiveCasesViral  deathConfirmed  deathProbable  \
0           1415473.0                 NaN             NaN            NaN
1                 NaN            357471.0          5928.0         1412.0
2           2161833.0            232338.0          3864.0          920.0
3                 NaN                 0.0             NaN            NaN
4                 NaN            696548.0         11440.0         1379.0

   totalTestEncountersViral  totalTestsPeopleViral  totalTestsAntibody  \
0                       NaN                    NaN                 NaN
1                       NaN              2116642.0                 NaN
2                       NaN                    NaN                 NaN
3                       NaN                    NaN                 NaN
4                       NaN              3409395.0            415234.0

   positiveTestsAntibody  negativeTestsAntibody  totalTestsPeopleAntibody  \
0                    NaN                    NaN                       NaN
1                    NaN                    NaN                  103530.0
2                    NaN                    NaN                       NaN
3                    NaN                    NaN                       NaN
4                    NaN                    NaN                       NaN

   positiveTestsPeopleAntibody  negativeTestsPeopleAntibody  \
0                          NaN                          NaN
1                          NaN                          NaN
2                          NaN                          NaN
3                          NaN                          NaN
4                          NaN                          NaN

   totalTestsPeopleAntigen  positiveTestsPeopleAntigen  totalTestsAntigen  \
0                      NaN                         NaN                NaN
1                      NaN                         NaN                NaN
2                 354207.0                     68660.0                NaN
3                      NaN                         NaN                NaN
4                      NaN                         NaN                NaN

   positiveTestsAntigen  fips  positiveIncrease  negativeIncrease    total  \
0                   NaN     2               199                 0    52150
1                   NaN     1              3648              8103  2211905
2                   NaN     5              1892             12762  2452689
3                   NaN    60                 0                 0     2140
4                   NaN     4              4671             10704  3456079

   totalTestResultsIncrease   posNeg  deathIncrease  hospitalizedIncrease  \
0                     12449    52150              1                     1
1                     10607  2211905            168                     0
2                     14233  2452689             42                    58
```

```
3                         0     2140              0                     0
4                     45161  3456079            176                   312

                                     hash  commercialScore  \
0  6932258339f816d5d92e6907200e546fb5342eb3                0
1  29405d617beb57f1ee9b5f00a86aac1008f78acd                0
2  7f491ba52aea5b93016e6ca41059b491f11fc5bf                0
3  0138eb6fe94624353fb7066d0113058280086d56                0
4  ef4f84f80eb115d6b32a4c6136a6ed1f4a5cf71f                0

   negativeRegularScore  negativeScore  positiveScore  score  grade
0                     0              0              0      0    NaN
1                     0              0              0      0    NaN
2                     0              0              0      0    NaN
3                     0              0              0      0    NaN
4                     0              0              0      0    NaN
```

[12]:
```
state_data.tail(10)
```

[12]:
```
           date state  positive  probableCases  negative  pending  \
18632  20200122    WA       2.0            NaN       NaN      NaN
18633  20200121    WA       2.0            NaN       NaN      NaN
18634  20200120    WA       1.0            NaN       NaN      NaN
18635  20200119    WA       1.0            NaN       NaN      NaN
18636  20200118    WA       0.0            NaN       NaN      NaN
18637  20200117    WA       0.0            NaN       NaN      NaN
18638  20200116    WA       0.0            NaN       NaN      NaN
18639  20200115    WA       0.0            NaN       NaN      NaN
18640  20200114    WA       0.0            NaN       NaN      NaN
18641  20200113    WA       NaN            NaN       NaN      NaN

         totalTestResultsSource  totalTestResults  hospitalizedCurrently  \
18632  totalTestEncountersViral               0.0                    NaN
18633  totalTestEncountersViral               NaN                    NaN
18634  totalTestEncountersViral               NaN                    NaN
18635  totalTestEncountersViral               NaN                    NaN
18636  totalTestEncountersViral               NaN                    NaN
18637  totalTestEncountersViral               NaN                    NaN
18638  totalTestEncountersViral               NaN                    NaN
18639  totalTestEncountersViral               NaN                    NaN
18640  totalTestEncountersViral               NaN                    NaN
18641  totalTestEncountersViral               NaN                    NaN

       hospitalizedCumulative  inIcuCurrently  inIcuCumulative  \
18632                     NaN             NaN              NaN
18633                     NaN             NaN              NaN
18634                     NaN             NaN              NaN
```

```
18635                    NaN             NaN              NaN
18636                    NaN             NaN              NaN
18637                    NaN             NaN              NaN
18638                    NaN             NaN              NaN
18639                    NaN             NaN              NaN
18640                    NaN             NaN              NaN
18641                    NaN             NaN              NaN

       onVentilatorCurrently  onVentilatorCumulative  recovered  \
18632                    NaN                     NaN        NaN
18633                    NaN                     NaN        NaN
18634                    NaN                     NaN        NaN
18635                    NaN                     NaN        NaN
18636                    NaN                     NaN        NaN
18637                    NaN                     NaN        NaN
18638                    NaN                     NaN        NaN
18639                    NaN                     NaN        NaN
18640                    NaN                     NaN        NaN
18641                    NaN                     NaN        NaN

      dataQualityGrade lastUpdateEt dateModified checkTimeEt  death  \
18632              NaN          NaN          NaN         NaN    NaN
18633              NaN          NaN          NaN         NaN    NaN
18634              NaN          NaN          NaN         NaN    NaN
18635              NaN          NaN          NaN         NaN    NaN
18636              NaN          NaN          NaN         NaN    NaN
18637              NaN          NaN          NaN         NaN    NaN
18638              NaN          NaN          NaN         NaN    NaN
18639              NaN          NaN          NaN         NaN    NaN
18640              NaN          NaN          NaN         NaN    NaN
18641              NaN          NaN          NaN         NaN    NaN

       hospitalized dateChecked  totalTestsViral  positiveTestsViral  \
18632           NaN         NaN              NaN                 NaN
18633           NaN         NaN              NaN                 NaN
18634           NaN         NaN              NaN                 NaN
18635           NaN         NaN              NaN                 NaN
18636           NaN         NaN              NaN                 NaN
18637           NaN         NaN              NaN                 NaN
18638           NaN         NaN              NaN                 NaN
18639           NaN         NaN              NaN                 NaN
18640           NaN         NaN              NaN                 NaN
18641           NaN         NaN              NaN                 NaN

       negativeTestsViral  positiveCasesViral  deathConfirmed  deathProbable  \
18632                 NaN                 2.0             NaN            NaN
18633                 NaN                 2.0             NaN            NaN
```

```
18634                NaN                1.0            NaN            NaN
18635                NaN                1.0            NaN            NaN
18636                NaN                0.0            NaN            NaN
18637                NaN                0.0            NaN            NaN
18638                NaN                0.0            NaN            NaN
18639                NaN                0.0            NaN            NaN
18640                NaN                0.0            NaN            NaN
18641                NaN                NaN            NaN            NaN

       totalTestEncountersViral  totalTestsPeopleViral  totalTestsAntibody  \
18632                       0.0                    NaN                 NaN
18633                       NaN                    NaN                 NaN
18634                       NaN                    NaN                 NaN
18635                       NaN                    NaN                 NaN
18636                       NaN                    NaN                 NaN
18637                       NaN                    NaN                 NaN
18638                       NaN                    NaN                 NaN
18639                       NaN                    NaN                 NaN
18640                       NaN                    NaN                 NaN
18641                       NaN                    NaN                 NaN

       positiveTestsAntibody  negativeTestsAntibody  totalTestsPeopleAntibody  \
18632                    NaN                    NaN                       NaN
18633                    NaN                    NaN                       NaN
18634                    NaN                    NaN                       NaN
18635                    NaN                    NaN                       NaN
18636                    NaN                    NaN                       NaN
18637                    NaN                    NaN                       NaN
18638                    NaN                    NaN                       NaN
18639                    NaN                    NaN                       NaN
18640                    NaN                    NaN                       NaN
18641                    NaN                    NaN                       NaN

       positiveTestsPeopleAntibody  negativeTestsPeopleAntibody  \
18632                          NaN                          NaN
18633                          NaN                          NaN
18634                          NaN                          NaN
18635                          NaN                          NaN
18636                          NaN                          NaN
18637                          NaN                          NaN
18638                          NaN                          NaN
18639                          NaN                          NaN
18640                          NaN                          NaN
18641                          NaN                          NaN

       totalTestsPeopleAntigen  positiveTestsPeopleAntigen  totalTestsAntigen  \
18632                      NaN                         NaN                NaN
```

```
18633                      NaN                          NaN                NaN
18634                      NaN                          NaN                NaN
18635                      NaN                          NaN                NaN
18636                      NaN                          NaN                NaN
18637                      NaN                          NaN                NaN
18638                      NaN                          NaN                NaN
18639                      NaN                          NaN                NaN
18640                      NaN                          NaN                NaN
18641                      NaN                          NaN                NaN

       positiveTestsAntigen  fips  positiveIncrease  negativeIncrease  total  \
18632                   NaN    53                 0                 0      2
18633                   NaN    53                 1                 0      2
18634                   NaN    53                 0                 0      1
18635                   NaN    53                 1                 0      1
18636                   NaN    53                 0                 0      0
18637                   NaN    53                 0                 0      0
18638                   NaN    53                 0                 0      0
18639                   NaN    53                 0                 0      0
18640                   NaN    53                 0                 0      0
18641                   NaN    53                 0                 0      0

       totalTestResultsIncrease  posNeg  deathIncrease  hospitalizedIncrease  \
18632                         0       2              0                     0
18633                         0       2              0                     0
18634                         0       1              0                     0
18635                         0       1              0                     0
18636                         0       0              0                     0
18637                         0       0              0                     0
18638                         0       0              0                     0
18639                         0       0              0                     0
18640                         0       0              0                     0
18641                         0       0              0                     0

                                       hash  commercialScore  \
18632  1c59843d8491d0424238185b5466a7c70b3d829b                0
18633  bc80b5a0a1604d12220439ae023c84dfe8486945                0
18634  527045c87fb9bfea340434654ed88b2dabd96a7e                0
18635  b4bb523b5dfabd6570a6659cc884168b7a7a7411                0
18636  852b2c808c5c506316af0e6f3dbb0a1f6131ede5                0
18637  effcc7b17054638471dec2f2a9cdbd3229d92042                0
18638  d9453d58f00994cff1f81639e95e532d1dd10138                0
18639  bd1eef4afdffd883e37be1d3f321dc3b0e75f591                0
18640  84aeea9c6a284b151fdb42a863db3c0aadb651bf                0
18641  4b278ad9e69da1c2a818f761c25f10c4bb6bf2a7                0

       negativeRegularScore  negativeScore  positiveScore  score  grade
```

```
18632    0    0    0    0    NaN
18633    0    0    0    0    NaN
18634    0    0    0    0    NaN
18635    0    0    0    0    NaN
18636    0    0    0    0    NaN
18637    0    0    0    0    NaN
18638    0    0    0    0    NaN
18639    0    0    0    0    NaN
18640    0    0    0    0    NaN
18641    0    0    0    0    NaN
```

8.1.1 Optional: use `to_csv()`method to save downloaded data to a csv

[13]:
```
state_data.to_csv( 'state_data.csv' )
```

8.1.2 Select one column

Enclose the column label string in brackets `[]`

[14]:
```
state_data[ 'state' ]
```

[14]:
```
0        AK
1        AL
2        AR
3        AS
4        AZ
         ..
18637    WA
18638    WA
18639    WA
18640    WA
18641    WA
Name: state, Length: 18642, dtype: object
```

8.1.3 Select one or more rows

Enclose the selection criteria in brackets `[]`

[15]:
```
len( state_data )
```

[15]:
```
18642
```

[16]:
```
state_data['state'].value_counts()
```

[16]:
```
WA    382
MA    373
VA    368
FL    366
```

NJ	354
NE	349
IN	337
WY	334
MI	334
RI	334
NY	333
TX	332
WI	332
VT	332
IL	331
AZ	331
NC	331
GA	331
HI	331
CA	331
OR	331
SC	331
CO	331
NH	331
DC	330
NV	330
MD	330
OH	330
NM	330
TN	330
KY	329
AK	329
IA	329
DE	329
AR	329
PA	329
WV	329
MN	329
KS	329
AL	328
ME	328
ND	328
UT	328
SD	328
ID	328
LA	328
MO	328
OK	328
CT	328
MS	328
MT	328

```
MP    319
PR    319
GU    319
AS    319
VI    319
Name: state, dtype: int64
```

[18]:
```
maryland_rows = state_data[ 'state' ] == 'MD'
```

[20]:
```
maryland_rows.value_counts()
```

[20]:
```
False    18312
True       330
Name: state, dtype: int64
```

[21]:
```
state_data[ state_data[ 'state' ] == 'MD' ]
```

[21]:

	date	state	positive	probableCases	negative	pending	\
22	20210128	MD	348749.0	NaN	2810526.0	NaN	
78	20210127	MD	346559.0	NaN	2801971.0	NaN	
134	20210126	MD	344620.0	NaN	2793416.0	NaN	
190	20210125	MD	343138.0	NaN	2787084.0	NaN	
246	20210124	MD	341452.0	NaN	2779405.0	NaN	
…	…	…	…	…	…	…	
18190	20200309	MD	5.0	NaN	73.0	NaN	
18241	20200308	MD	3.0	NaN	52.0	NaN	
18292	20200307	MD	3.0	NaN	41.0	NaN	
18339	20200306	MD	3.0	NaN	26.0	12.0	
18372	20200305	MD	0.0	NaN	17.0	14.0	

	totalTestResultsSource	totalTestResults	hospitalizedCurrently	\
22	totalTestsViral	6865525.0	1636.0	
78	totalTestsViral	6819049.0	1647.0	
134	totalTestsViral	6784162.0	1642.0	
190	totalTestsViral	6758236.0	1669.0	
246	totalTestsViral	6723502.0	1668.0	
…	…	…	…	
18190	totalTestsViral	NaN	NaN	
18241	totalTestsViral	NaN	NaN	
18292	totalTestsViral	NaN	NaN	
18339	totalTestsViral	NaN	NaN	
18372	totalTestsViral	NaN	NaN	

	hospitalizedCumulative	inIcuCurrently	inIcuCumulative	\
22	31614.0	376.0	NaN	
78	31468.0	374.0	NaN	
134	31330.0	367.0	NaN	

```
190                   31188.0           395.0              NaN
246                   31034.0           392.0              NaN
…                       …                 …                …
18190                   NaN               NaN              NaN
18241                   NaN               NaN              NaN
18292                   NaN               NaN              NaN
18339                   NaN               NaN              NaN
18372                   NaN               NaN              NaN

       onVentilatorCurrently  onVentilatorCumulative  recovered  \
22                       NaN                     NaN     9482.0
78                       NaN                     NaN     9479.0
134                      NaN                     NaN     9477.0
190                      NaN                     NaN     9476.0
246                      NaN                     NaN     9474.0
…                        …                       …         …
18190                    NaN                     NaN        NaN
18241                    NaN                     NaN        NaN
18292                    NaN                     NaN        NaN
18339                    NaN                     NaN        NaN
18372                    NaN                     NaN        NaN

      dataQualityGrade     lastUpdateEt          dateModified  checkTimeEt  \
22                   A  1/28/2021 10:00  2021-01-28T10:00:00Z  01/28 05:00
78                   A  1/27/2021 10:00  2021-01-27T10:00:00Z  01/27 05:00
134                  A  1/26/2021 10:00  2021-01-26T10:00:00Z  01/26 05:00
190                  A  1/25/2021 10:00  2021-01-25T10:00:00Z  01/25 05:00
246                  A  1/24/2021 10:00  2021-01-24T10:00:00Z  01/24 05:00
…                  …                 …                     …           …
18190              NaN               NaN                   NaN         NaN
18241              NaN               NaN                   NaN         NaN
18292              NaN               NaN                   NaN         NaN
18339              NaN               NaN                   NaN         NaN
18372              NaN               NaN                   NaN         NaN

        death  hospitalized           dateChecked  totalTestsViral  \
22     7037.0       31614.0  2021-01-28T10:00:00Z        6865525.0
78     6996.0       31468.0  2021-01-27T10:00:00Z        6819049.0
134    6963.0       31330.0  2021-01-26T10:00:00Z        6784162.0
190    6900.0       31188.0  2021-01-25T10:00:00Z        6758236.0
246    6865.0       31034.0  2021-01-24T10:00:00Z        6723502.0
…       …             …                     …                …
18190     NaN           NaN                   NaN              NaN
18241     NaN           NaN                   NaN              NaN
18292     NaN           NaN                   NaN              NaN
18339     NaN           NaN                   NaN              NaN
18372     NaN           NaN                   NaN              NaN
```

```
       positiveTestsViral  negativeTestsViral  positiveCasesViral  \
22                424593.0                 NaN            348749.0
78                421831.0                 NaN            346559.0
134               419497.0                 NaN            344620.0
190               417738.0                 NaN            343138.0
246               415608.0                 NaN            341452.0
…                      …                   …                   …
18190                NaN                 NaN                 5.0
18241                NaN                 NaN                 3.0
18292                NaN                 NaN                 3.0
18339                NaN                 NaN                 3.0
18372                NaN                 NaN                 0.0

       deathConfirmed  deathProbable  totalTestEncountersViral  \
22             6861.0          176.0                       NaN
78             6821.0          175.0                       NaN
134            6788.0          175.0                       NaN
190            6726.0          174.0                       NaN
246            6690.0          175.0                       NaN
…                 …               …                         …
18190            NaN             NaN                       NaN
18241            NaN             NaN                       NaN
18292            NaN             NaN                       NaN
18339            NaN             NaN                       NaN
18372            NaN             NaN                       NaN

       totalTestsPeopleViral  totalTestsAntibody  positiveTestsAntibody  \
22                 3159275.0                 NaN                    NaN
78                 3148530.0                 NaN                    NaN
134                3138036.0                 NaN                    NaN
190                3130222.0                 NaN                    NaN
246                3120857.0                 NaN                    NaN
…                        …                   …                      …
18190                  NaN                 NaN                    NaN
18241                  NaN                 NaN                    NaN
18292                  NaN                 NaN                    NaN
18339                  NaN                 NaN                    NaN
18372                  NaN                 NaN                    NaN

       negativeTestsAntibody  totalTestsPeopleAntibody  \
22                       NaN                  186392.0
78                       NaN                  186392.0
134                      NaN                  183103.0
190                      NaN                  183103.0
246                      NaN                  183103.0
…                          …                      …
```

```
18190                    NaN                        NaN
18241                    NaN                        NaN
18292                    NaN                        NaN
18339                    NaN                        NaN
18372                    NaN                        NaN

       positiveTestsPeopleAntibody  negativeTestsPeopleAntibody  \
22                         23333.0                     163059.0
78                         23333.0                     163059.0
134                        22101.0                     161002.0
190                        22101.0                     161002.0
246                        22101.0                     161002.0
…                              …                            …
18190                         NaN                          NaN
18241                         NaN                          NaN
18292                         NaN                          NaN
18339                         NaN                          NaN
18372                         NaN                          NaN

       totalTestsPeopleAntigen  positiveTestsPeopleAntigen  totalTestsAntigen  \
22                         NaN                         NaN                NaN
78                         NaN                         NaN                NaN
134                        NaN                         NaN                NaN
190                        NaN                         NaN                NaN
246                        NaN                         NaN                NaN
…                            …                           …                  …
18190                      NaN                         NaN                NaN
18241                      NaN                         NaN                NaN
18292                      NaN                         NaN                NaN
18339                      NaN                         NaN                NaN
18372                      NaN                         NaN                NaN

       positiveTestsAntigen  fips  positiveIncrease  negativeIncrease  \
22                      NaN    24              2190              8555
78                      NaN    24              1939              8555
134                     NaN    24              1482              6332
190                     NaN    24              1686              7679
246                     NaN    24              2145              7679
…                         …    …                  …                 …
18190                   NaN    24                 2                21
18241                   NaN    24                 0                11
18292                   NaN    24                 0                15
18339                   NaN    24                 3                 9
18372                   NaN    24                 0                 0

         total  totalTestResultsIncrease   posNeg  deathIncrease  \
22     3159275                     46476  3159275             41
```

```
78     3148530                    34887  3148530             33
134    3138036                    25926  3138036             63
190    3130222                    34734  3130222             35
246    3120857                    45247  3120857             28
…        …                          …      …                 …
18190     78                         0      78                0
18241     55                         0      55                0
18292     44                         0      44                0
18339     41                         0      29                0
18372     31                         0      17                0

       hospitalizedIncrease                                 hash  \
22                      146  56c272f234cd160cd29afd1c11b467637bfb5906
78                      138  8f596f67caba35a458ae9002728a5005e5d3a138
134                     142  eb8d7d931baae7c2766507be331584f395542444
190                     154  044ddf54b4cb506df3589f7823a535da644b54ed
246                     168  f3c2ff99635fd79c0b48f598504dc0df87d43778
…                         …                                      …
18190                     0  2440ca6317977560a4a6d4028bf31fe8fd2571f9
18241                     0  7b991089ef44365c653f8acbd7021ef451020e36
18292                     0  35f68a20c2a36c4ad386b3b1f31c9fc971912e83
18339                     0  ef5005974414f326294a040fb638e65cf293addf
18372                     0  6e4bc0e2b5dd241cf7495b2355a4ab7d6600bece

       commercialScore  negativeRegularScore  negativeScore  positiveScore  \
22                   0                     0              0              0
78                   0                     0              0              0
134                  0                     0              0              0
190                  0                     0              0              0
246                  0                     0              0              0
…                  …                     …              …              …
18190                0                     0              0              0
18241                0                     0              0              0
18292                0                     0              0              0
18339                0                     0              0              0
18372                0                     0              0              0

       score  grade
22         0    NaN
78         0    NaN
134        0    NaN
190        0    NaN
246        0    NaN
…        …    …
18190      0    NaN
18241      0    NaN
18292      0    NaN
```

```
18339       0    NaN
18372       0    NaN

[330 rows x 55 columns]
```

8.1.4 `groupby()` operation example

Sum all daily deaths since the beginning of the pandemic on a state-by-state basis, and sort in descending order

8.1.5 Using `groupby()`, `sum()`, and `sort_values()` method in a single line of code

1. Group the rows by the value in the `'state'` column.
2. For each batch of rows:
    1. Take the values in the `'deathIncrease'` column
    2. Sum them, returning a new DataFrame that has one number per state.
3. Sort the values in descending order

[24]:
```
state_data.groupby( 'state' )[ 'deathIncrease' ].sum().sort_values(␣
↪ascending=False )
```

[24]:
```
state
CA    38961
TX    35639
NY    34742
FL    26456
PA    21303
NJ    21301
IL    21074
MI    15402
MA    14348
GA    13827
AZ    12819
OH    11006
IN     9879
TN     9417
NC     9046
LA     8743
AL     7340
MD     7037
CT     7020
SC     6903
MO     6725
WI     6334
VA     6308
MN     6140
MS     5945
CO     5569
```

```
AR     4784
IA     4532
WA     4211
NV     4181
KS     3718
KY     3611
OK     3423
NM     3226
RI     2144
WV     1983
OR     1924
NE     1917
PR     1801
SD     1763
ID     1714
UT     1620
ND     1442
MT     1210
DE     1075
NH     1022
DC      902
WY      596
ME      567
HI      406
AK      262
VT      172
GU      129
VI       24
MP        2
AS        0
Name: deathIncrease, dtype: int64
```

[25]:
```
state_data.groupby( 'date' )[ 'deathIncrease' ].sum().sort_values( )
```

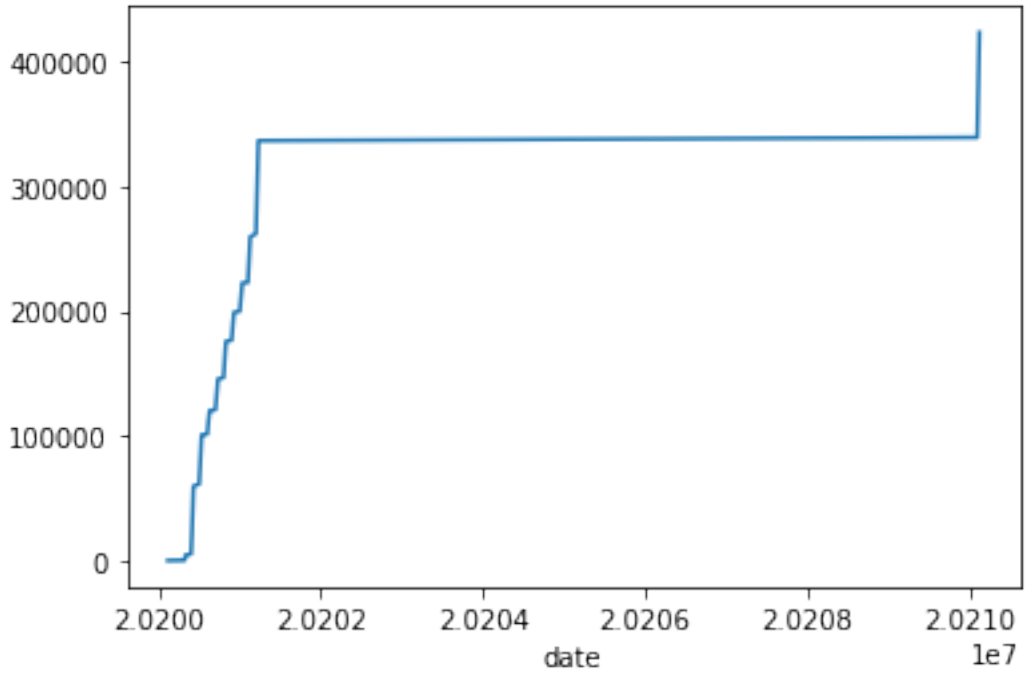
[25]:
```
date
20200113       0
20200208       0
20200209       0
20200210       0
20200211       0
              …
20210112    4064
20210127    4077
20210107    4079
20210113    4087
20210120    4409
Name: deathIncrease, Length: 382, dtype: int64
```

8.1.6 CAVIAT: Understanding the data types of the columns is key

- string, integer, float, datatime, etc

[26]:
```
state_data.groupby( 'date' )[ "death" ].sum().plot()
```

[26]:
```
<AxesSubplot:xlabel='date'>
```



8.1.7 PANDAS is interpreting the date column as an integer

[27]:
```
state_data[ 'date' ]
```

[27]:
```
0        20210128
1        20210128
2        20210128
3        20210128
4        20210128
            ...
18637    20200117
18638    20200116
18639    20200115
18640    20200114
18641    20200113
Name: date, Length: 18642, dtype: int64
```

8.1.8 Use `pd.to_datetime()` function to reinterpret dates

Remember `pd` is a shortcut for `pandas` because we `import pandas as pd`

[28]:
```
state_data[ 'date' ] = pd.to_datetime( state_data[ 'date' ], format="%Y%m%d" )
```

[29]:
```
state_data[ 'date' ]
```

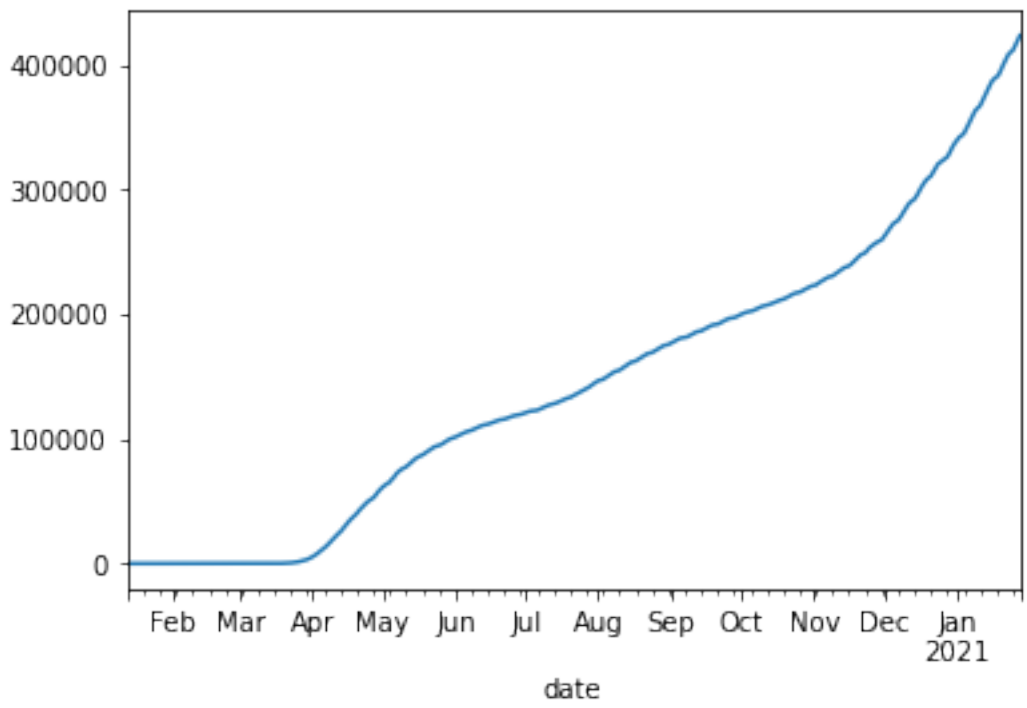
[29]:
```
0       2021-01-28
1       2021-01-28
2       2021-01-28
3       2021-01-28
4       2021-01-28
           ...
18637   2020-01-17
18638   2020-01-16
18639   2020-01-15
18640   2020-01-14
18641   2020-01-13
Name: date, Length: 18642, dtype: datetime64[ns]
```

9 Quick (non-interactive) plot #1

9.1 Cumulative number of COVID-19 *deaths* in the US

[30]:
```
state_data.groupby( 'date' )[ "death" ].sum().plot()
```

[30]:
```
<AxesSubplot:xlabel='date'>
```



9.1.1 In this case, the x and y axes were inferred from the structure of the DataFrame

[31]:
```
state_data.groupby( 'date' )[ "death" ].sum()
```

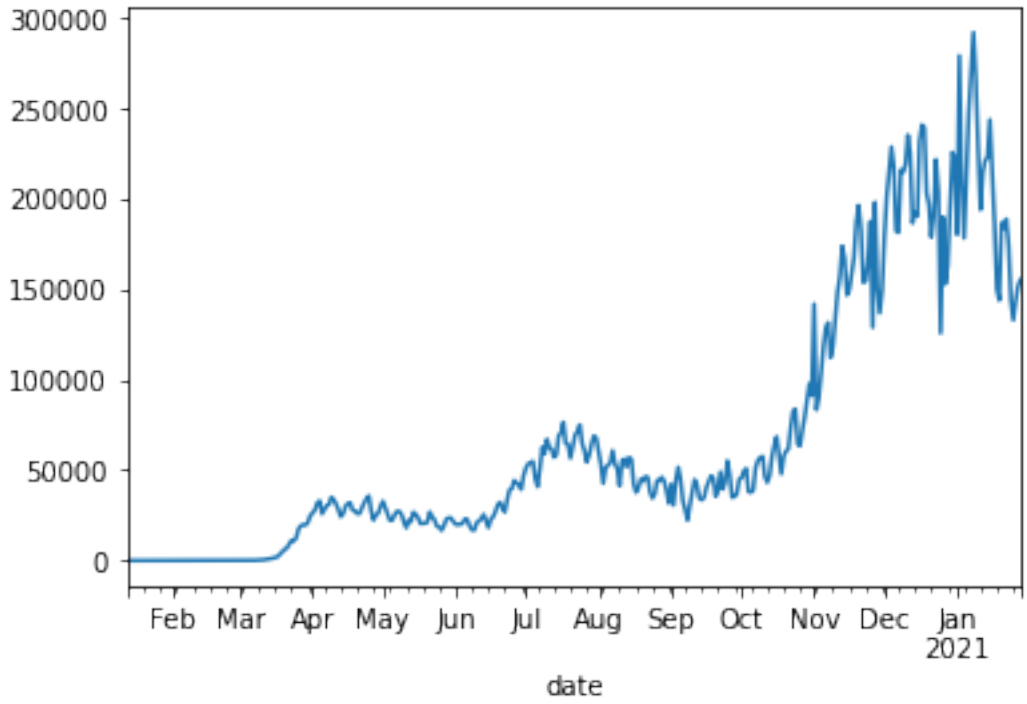
[31]:
```
date
2020-01-13         0.0
2020-01-14         0.0
2020-01-15         0.0
2020-01-16         0.0
2020-01-17         0.0
                  ...
2021-01-24    410230.0
2021-01-25    411823.0
2021-01-26    415557.0
2021-01-27    419634.0
2021-01-28    423645.0
Name: death, Length: 382, dtype: float64
```

10 Quick (non-interactive) plot #2

10.1 Daily US COVID-19 *cases* increase

[32]:
```
state_data.groupby( 'date' )[ "positiveIncrease" ].sum().plot()
```

[32]:
```
<AxesSubplot:xlabel='date'>
```



11 Interactive plot #1

11.1 Daily US COVID-19 *cases* increase with 7-day rolling average

11.1.1 Strategy:

1. Construct a new DataFrame with two columns
    1. Raw daily increase
    2. Rolling daily increase
2. Use the `plot_bokeh()` function to generate the interactive plot

11.1.2 Store daily increase in a variable

[33]:
```
us_new_cases_daily = state_data.groupby( 'date' )[ "positiveIncrease" ].sum()
```

[34]:
```
us_new_cases_daily
```

[34]:
```
date
2020-01-13         0
2020-01-14         0
2020-01-15         0
2020-01-16         0
2020-01-17         0
                 ...
2021-01-24    143691
2021-01-25    133067
2021-01-26    143745
2021-01-27    151675
2021-01-28    155333
Name: positiveIncrease, Length: 382, dtype: int64
```

[35]:
```
len( us_new_cases_daily)
```

[35]:
```
382
```

[36]:
```
us_new_cases_daily.name = "Daily increase"
```

11.1.3 Use the rolling()method to create smooth data series

[37]:
```
# can also pass "center=True" argument to rolling()
us_new_cases_7day_mean = us_new_cases_daily.rolling(7).mean()
```

[38]:
```
us_new_cases_7day_mean
```

[38]:
```
date
2020-01-13              NaN
2020-01-14              NaN
2020-01-15              NaN
2020-01-16              NaN
2020-01-17              NaN
                       ...
2021-01-24    167485.857143
2021-01-25    165035.428571
2021-01-26    164992.285714
2021-01-27    159985.571429
2021-01-28    155885.571429
Name: Daily increase, Length: 382, dtype: float64
```

11.1.4 Round the smoothed data to the nearest person

[39]:
```
us_new_cases_7day_mean = us_new_cases_daily.rolling(7).mean().fillna(0).round().
↪astype( int )
```

[40]:
```
us_new_cases_7day_mean.name = '7-day average'
```

[41]:
```
us_new_cases_7day_mean
```

[41]:
```
date
2020-01-13         0
2020-01-14         0
2020-01-15         0
2020-01-16         0
2020-01-17         0
                 ...
2021-01-24    167486
2021-01-25    165035
2021-01-26    164992
2021-01-27    159986
2021-01-28    155886
Name: 7-day average, Length: 382, dtype: int64
```

11.1.5 Use `pandas.concat()` to bind columns together

[42]:
```
new_cases_df = pd.concat( ( us_new_cases_daily, us_new_cases_7day_mean ),␣
↪axis='columns' )
```

12 Comparison of quick plot and interactive plot

[44]:
```
new_cases_df.head()
```

[44]:

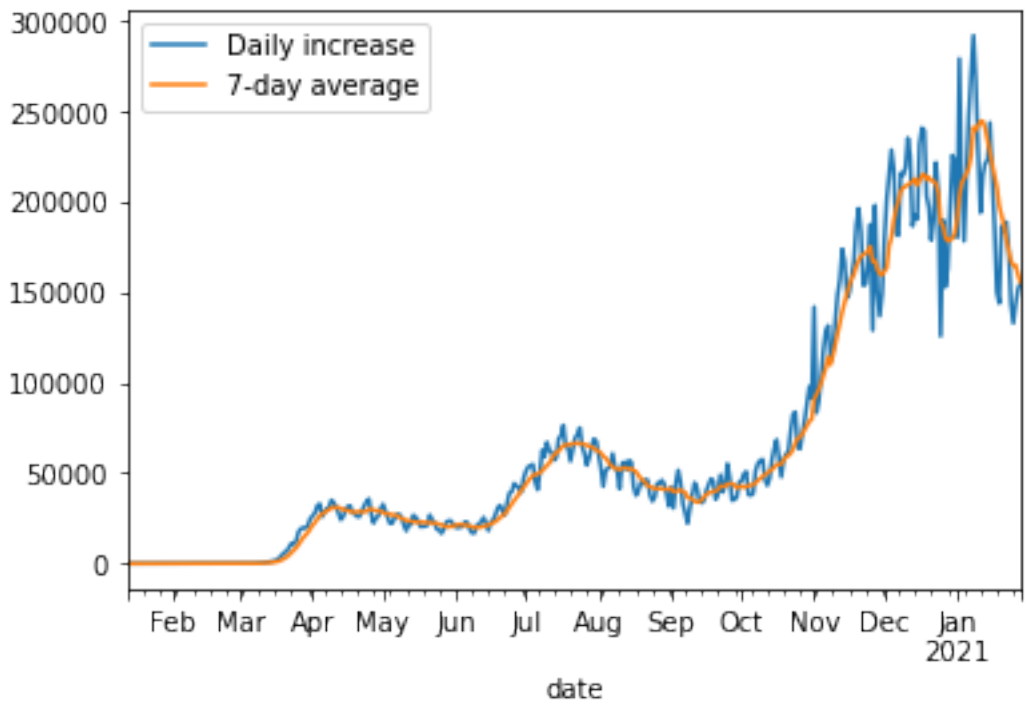
date	Daily increase	7-day average
2020-01-13	0	0
2020-01-14	0	0
2020-01-15	0	0
2020-01-16	0	0
2020-01-17	0	0

[45]:
```
new_cases_df.shape
```

[45]:
```
(382, 2)
```

[43]:
```
new_cases_df.plot()
```

[43]:
```
<AxesSubplot:xlabel='date'>
```



[46]:
```
new_cases_df.plot_bokeh( legend="top_left" )
```

[46]:
```
Figure(id='1002', …)
```

13 Saving plot to HTML file

I have been using a somewhat hacky workaround to save plots. I use the `return_html=True` argument to grab the raw html, and writing that directly to a file.

[ ]:
```
html = new_cases_df.plot_bokeh( legend="top_left", return_html=True )
```

[ ]:
```
with open("/Users/colettace/Desktop/CHRIS_BOKEH_PLOT.html" , "w") as f:
    f.write( '<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"␣
↪"http://www.w3.org/TR/html4/loose.dtd">\n' )
    f.write( '<html lang="en">\n' )
    f.write( '  <head>\n' )
    f.write( '    <meta http-equiv="content-type" content="text/html;␣
↪charset=utf-8">\n' )
    f.write( '    <title>CHRIS\'S BOKEH PLOT</title>\n' )
    f.write( '  </head>\n' )
    f.write( '  <body>\n' )
    f.write( html )
    f.write( '  </body>\n' )
```

```
    f.write( '</html>\n' )
```

14 Part 2: Maryland county-level chloropleth map

[ ]:
```
us_county_covid_death_url = "https://raw.githubusercontent.com/CSSEGISandData/
 ↪COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/
 ↪time_series_covid19_deaths_US.csv"
us_county_covid_url = "https://raw.githubusercontent.com/CSSEGISandData/
 ↪COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/
 ↪time_series_covid19_confirmed_US.csv"
```

[ ]:
```
us_county_covid_data = pd.read_csv( us_county_covid_url )
```

[ ]:
```
type( us_county_covid_data )
```

[ ]:
```
us_county_covid_data.shape
```

[ ]:
```
us_county_covid_data.head()
```

[ ]:
```
us_county_covid_data.sample(10)
```

[ ]:
```
wanted_rows = us_county_covid_data['Province_State'] == 'Maryland'
```

[ ]:
```
wanted_rows
```

[ ]:
```
maryland_county_covid_data = us_county_covid_data[ wanted_rows ].copy()
```

[ ]:
```
maryland_county_covid_data.shape
```

[ ]:
```
maryland_county_covid_data
```

14.1 Download US geographic boundaries from www.census.gov

- US geographic boundaries download page
  - Manually download and unzip 1:500,000 scale shapefiles
  - Use Python to download the file and unzip it programmatically

[ ]:
```
url = 'https://www2.census.gov/geo/tiger/GENZ2019/shp/cb_2019_us_county_500k.
 ↪zip'
r = requests.get(url, allow_redirects=True)
open('cb_2019_us_county_500k.zip', 'wb').write(r.content)
```

[ ]:
```
from zipfile import ZipFile
```

[ ]:
```
with ZipFile('cb_2019_us_county_500k.zip') as myzip:
    myzip.extractall()
```

[ ]:
```
us_county_shapes = gpd.read_file( 'cb_2019_us_county_500k.shp' )
```

[ ]:
```
type( us_county_shapes )
```

[ ]:
```
us_county_shapes.shape
```

[ ]:
```
us_county_shapes.info()
```

[ ]:
```
us_county_shapes.head()
```

15 Join the COVID table and the geography table together

Create a table join key 'FIPS' Federal Information Processing Standard state code

[ ]:
```
us_county_shapes['FIPS'] = us_county_shapes['STATEFP'] +␣
↪us_county_shapes['COUNTYFP']
```

[ ]:
```
us_county_shapes.head()
```

15.1 Make sure the FIPS column has the same data type

[ ]:
```
us_county_shapes['FIPS'].dtype
```

[ ]:
```
maryland_county_covid_data['FIPS'].dtype
```

[ ]:
```
us_county_shapes['FIPS'] = us_county_shapes['FIPS'].astype( int )
```

[ ]:
```
maryland_county_covid_data['FIPS'] = maryland_county_covid_data['FIPS'].astype(␣
↪int )
```

15.2 Perform table merge

- INNER join: resultant output contains only the rows for which the join keys appear in both tables

[ ]:
```
merged_data = us_county_shapes.merge( maryland_county_covid_data, on='FIPS',␣
↪how='inner' )
```

[ ]:
```
merged_data
```

16 Choropleth Map

- Use `plot_bokeh()`
- IMPORTANT:
  - Choropleth maps require the data frame object to be a `GeoDataFrame`
  - A column with label `'geometry'` must be in the `GeoDataFrame`

[ ]:
```
type( merged_data )
```

[ ]:
```
'geometry' in merged_data.columns
```

[ ]:
```
merged_data.plot_bokeh()
```

[ ]:
```
date_columns = [ _ for _ in merged_data.columns if _.endswith( '/20' ) ]
```

[ ]:
```
len( date_columns )
```

[ ]:
```
date_columns[-10:]
```

[ ]:
```
merged_data.plot_bokeh(
    #figsize=(1600, 900),
    #figsize=(600, 400),
    slider=date_columns,
    slider_name="Date",
    hovertool_columns= ['NAME'] + date_columns[-5:] )
```

[ ]:
```
gpd.__version__
```

[ ]:
```

```