

```
In [42]: import numpy as np
import pandas as pd
import seaborn as sns
import plotly
import matplotlib as plt
import plotly.graph_objects as go
import plotly.express as px
import plotly.offline as pyo
from datetime import date, timedelta
import folium
from scipy.interpolate import interp1d
pyo.init_notebook_mode()
```

```
In [43]: covid = pd.read_csv("https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv", error_bad_lines = False)
```

```
In [44]: covid.head()
```

Out[44]:

	date	county	state	fips	cases	deaths
0	2020-01-21	Snohomish	Washington	53061.0	1	0
1	2020-01-22	Snohomish	Washington	53061.0	1	0
2	2020-01-23	Snohomish	Washington	53061.0	1	0
3	2020-01-24	Cook	Illinois	17031.0	1	0
4	2020-01-24	Snohomish	Washington	53061.0	1	0

```
In [45]: covid.shape
```

Out[45]: (411727, 6)

```
In [46]: del covid["fips"]
```

```
In [47]: covid.head()
```

Out[47]:

	date	county	state	cases	deaths
0	2020-01-21	Snohomish	Washington	1	0
1	2020-01-22	Snohomish	Washington	1	0
2	2020-01-23	Snohomish	Washington	1	0
3	2020-01-24	Cook	Illinois	1	0
4	2020-01-24	Snohomish	Washington	1	0

```
In [48]: covid.columns
```

Out[48]: Index(['date', 'county', 'state', 'cases', 'deaths'], dtype='object')

```
In [49]: today = date.today()
three_days_back = (today - timedelta(days=3)).strftime('%Y-%m-%d')
covid_latest = covid[covid.date.eq(three_days_back)]
covid_latest.head()
```

Out[49]:

	date	county	state	cases	deaths
402068	2020-08-05	Autauga	Alabama	1073	22
402069	2020-08-05	Baldwin	Alabama	3380	24
402070	2020-08-05	Barbour	Alabama	615	5
402071	2020-08-05	Bibb	Alabama	392	4
402072	2020-08-05	Blount	Alabama	839	3

```
In [55]: covid_states_lst = np.array(['Alabama', 'Alaska', 'Arizona', 'Arkansas', 'California', 'Colorado', 'Connecticut', 'Delaware',
'District of Columbia', 'Florida', 'Georgia', 'Guam', 'Hawaii', 'Idaho', 'Illinois', 'Indiana', 'Iowa', 'Kansas', 'Kentucky', 'Louisiana', 'Maine', 'Maryland', 'Massachusetts', 'Michigan', 'Minnesota', 'Mississippi', 'Missouri', 'Montana', 'Nebraska', 'Nevada', 'New Hampshire', 'New Jersey', 'New Mexico', 'New York', 'North Carolina', 'North Dakota', 'Northern Mariana Islands', 'Ohio', 'Oklahoma', 'Oregon', 'Pennsylvania', 'Puerto Rico', 'Rhode Island', 'South Carolina', 'South Dakota', 'Tennessee', 'Texas', 'Utah', 'Vermont', 'Virgin Islands', 'Virginia', 'Washington', 'West Virginia', 'Wisconsin', 'Wyoming'])
print(covid_states_lst)
```

```
['Alabama' 'Alaska' 'Arizona' 'Arkansas' 'California' 'Colorado'
'Connecticut' 'Delaware' 'District of Columbia' 'Florida' 'Georgia'
'Guam' 'Hawaii' 'Idaho' 'Illinois' 'Indiana' 'Iowa' 'Kansas' 'Kentucky'
'Louisiana' 'Maine' 'Maryland' 'Massachusetts' 'Michigan' 'Minnesota'
'Mississippi' 'Missouri' 'Montana' 'Nebraska' 'Nevada' 'New Hampshire'
'New Jersey' 'New Mexico' 'New York' 'North Carolina' 'North Dakota'
'Northern Mariana Islands' 'Ohio' 'Oklahoma' 'Oregon' 'Pennsylvania'
'Puerto Rico' 'Rhode Island' 'South Carolina' 'South Dakota' 'Tennessee'
'Texas' 'Utah' 'Vermont' 'Virgin Islands' 'Virginia' 'Washington'
'West Virginia' 'Wisconsin' 'Wyoming']
```

```
In [83]: #scd by state df.groupby('Outlet_Location_Type')
covid_county_total = covid_latest.groupby('county', as_index = False).sum()
print(covid_county_total)
```

	county	cases	deaths
0	Abbeville	304	8
1	Acadia	2536	77
2	Accomack	1085	15
3	Ada	8415	72
4	Adair	698	25
...	...	...	...
1910	Yukon-Koyukuk Census Area	49	1
1911	Yuma	11382	276
1912	Zapata	165	2
1913	Zavala	249	5
1914	Ziebach	9	0

[1915 rows x 3 columns]

```
In [84]: covid_state_total = covid_latest.groupby('state', as_index = False).sum()
#covid_state_total.drop('state')
print(covid_state_total)
```

	state	cases	deaths
0	Alabama	94654	1695
1	Alaska	4181	23
2	Arizona	182230	3933
3	Arkansas	46293	508
4	California	532776	9866
5	Colorado	49143	1857
6	Connecticut	50225	4437
7	Delaware	15296	587
8	District of Columbia	12443	587
9	Florida	502731	7626
10	Georgia	186395	3899
11	Guam	1358	6
12	Hawaii	2740	26
13	Idaho	22854	218
14	Illinois	188474	7776
15	Indiana	71685	3007
16	Iowa	46836	900
17	Kansas	30047	371
18	Kentucky	33823	771
19	Louisiana	126061	4096
20	Maine	3992	124
21	Maryland	93000	3536
22	Massachusetts	119643	8659
23	Michigan	94045	6480
24	Minnesota	57820	1670
25	Mississippi	63444	1804
26	Missouri	55919	1337
27	Montana	4481	65
28	Nebraska	27497	342
29	Nevada	52919	891
30	New Hampshire	6719	418
31	New Jersey	185180	15842
32	New Mexico	21566	667
33	New York	422935	32431
34	North Carolina	129511	2079
35	North Dakota	7061	112
36	Northern Mariana Islands	46	2
37	Ohio	96305	3596
38	Oklahoma	40555	583
39	Oregon	19979	342
40	Pennsylvania	120492	7310
41	Puerto Rico	19651	0
42	Rhode Island	19481	1012
43	South Carolina	95472	1894
44	South Dakota	9168	137
45	Tennessee	111305	1130
46	Texas	480789	8149
47	Utah	42478	330
48	Vermont	1436	57
49	Virgin Islands	501	9
50	Virginia	95049	2274
51	Washington	62390	1705
52	West Virginia	7159	124
53	Wisconsin	61247	979
54	Wyoming	2923	27

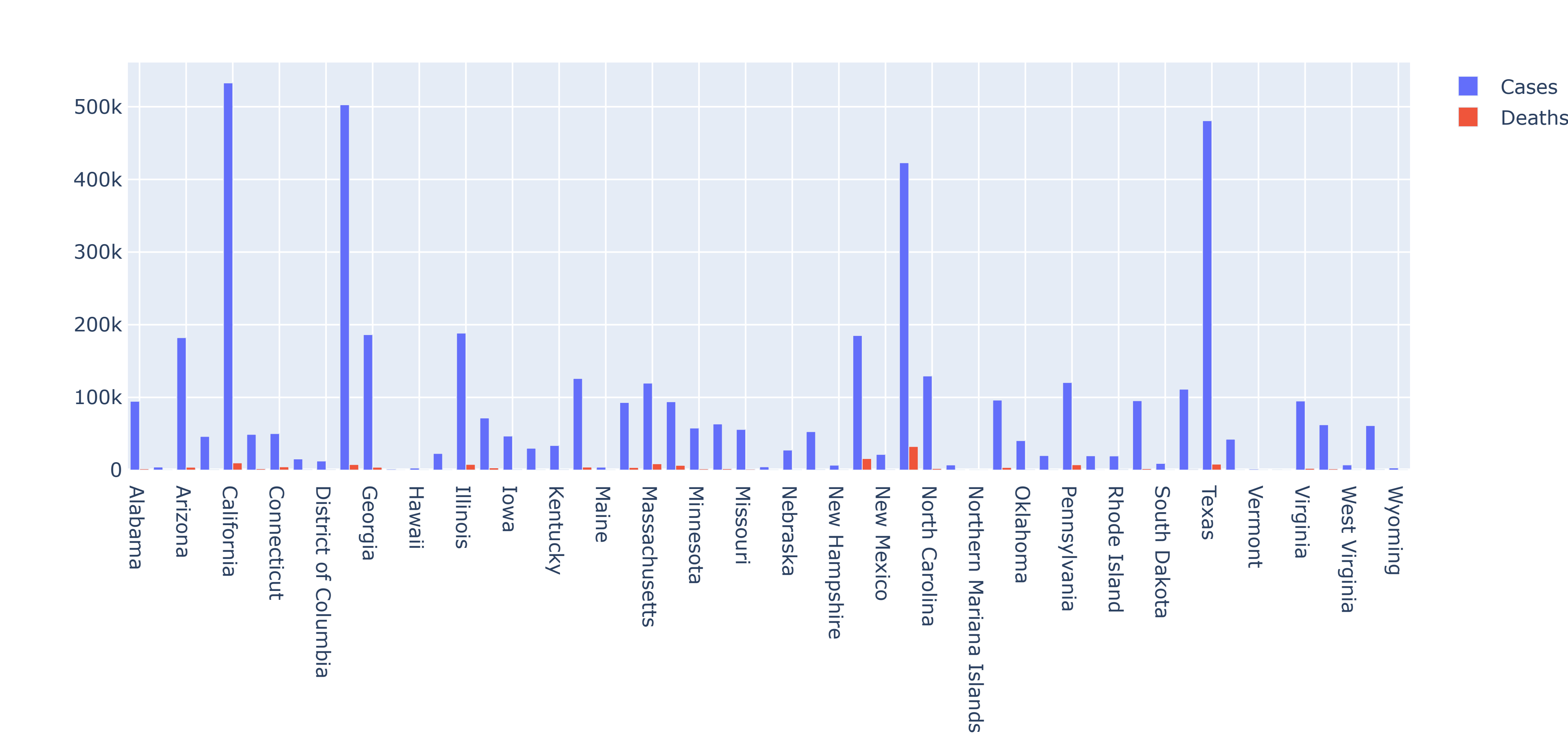
```
In [85]: scd[scd.state.eq("New York")]
```

Out[85]:

state	total_cases	total_deaths
-------	-------------	--------------

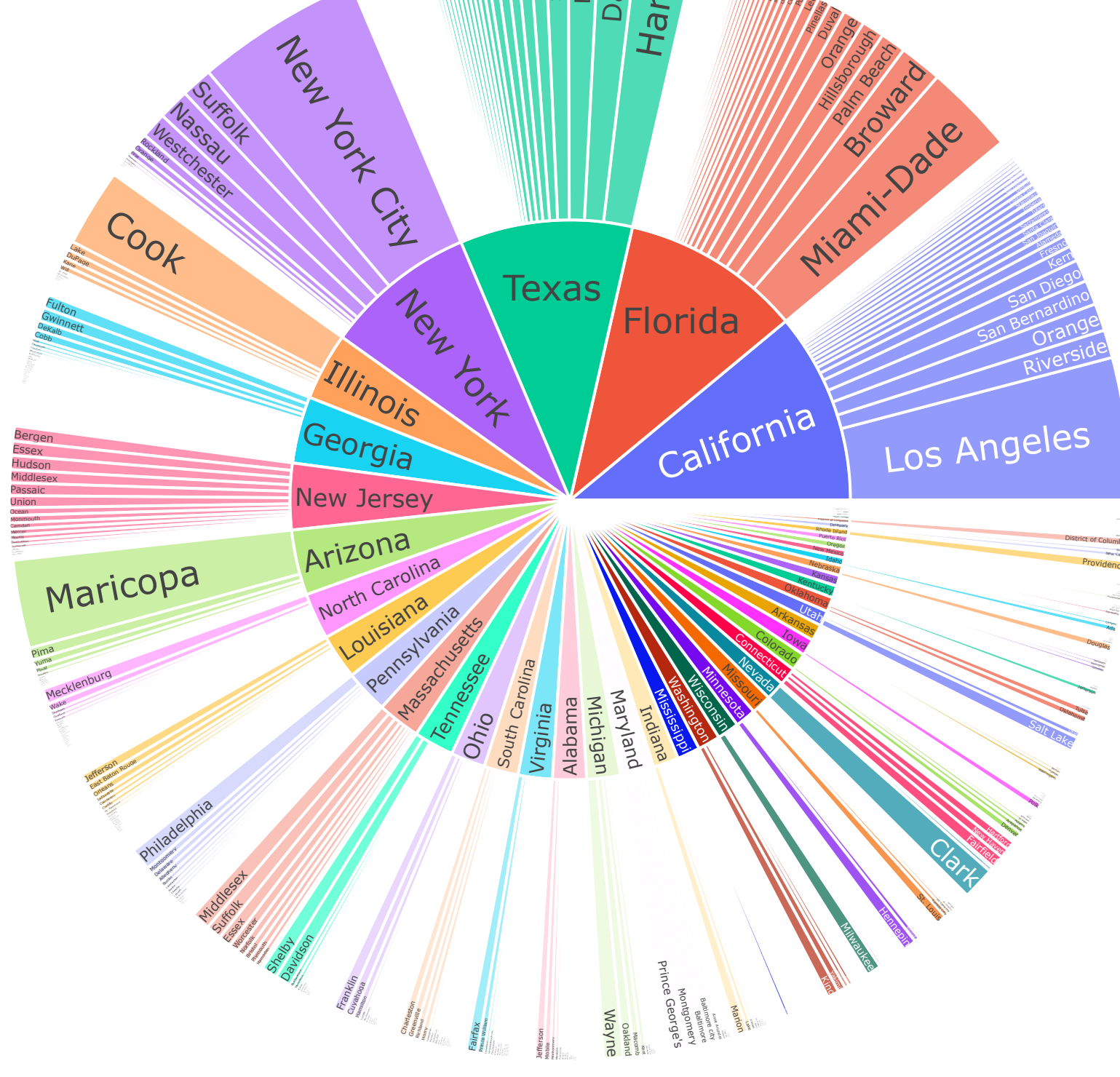
```
In [86]: fig = go.Figure(data=[
    go.Bar(name='Cases', x=covid_state_total['state'], y= covid_state_total['cases']),
    go.Bar(name='Deaths', x=covid_state_total['state'], y= covid_state_total['deaths'])
])
fig.update_layout(title="COVID-19 Cases in the United States as of " + three_days_back, barmode='group')
fig.show()
```

COVID-19 Cases in the United States as of 2020-08-05



```
In [87]: fig = px.sunburst(covid_latest, path=['state', 'county'], values = 'cases')
fig.update_layout(title = "COVID-19 Cases by State, County as of " + three_days_back)
fig.show()
```

COVID-19 Cases by State, County as of 2020-08-05



```
In [88]: STATES = pd.read_csv("STATES.csv")
```

```
In [89]: STATES.head()
```

Out[89]:

	state	latitude	longitude	name
0	AK	63.588753	-154.493062	Alaska
1	AL	32.318231	-86.902298	Alabama
2	AR	35.201050	-91.831833	Arkansas
3	AZ	34.048928	-111.093731	Arizona
4	CA	36.778261	-119.417932	California

```
In [90]: STATES_LST = list(STATES["name"])
print(STATES_LST)
```

```
['Alaska', 'Alabama', 'Arkansas', 'Arizona', 'California', 'Colorado', 'Connecticut', 'District of Columbia', 'Delaware', 'Florida', 'Georgia', 'Guam', 'Hawaii', 'Iowa', 'Idaho', 'Illinois', 'Indiana', 'Kansas', 'Kentucky', 'Louisiana', 'Massachusetts', 'Maryland', 'Maine', 'Michigan', 'Minnesota', 'Missouri', 'Montana', 'Nebraska', 'Nevada', 'New Hampshire', 'New Jersey', 'New Mexico', 'New York', 'North Carolina', 'North Dakota', 'Northern Mariana Islands', 'Ohio', 'Oklahoma', 'Oregon', 'Pennsylvania', 'Puerto Rico', 'Rhode Island', 'South Carolina', 'South Dakota', 'Tennessee', 'Texas', 'Utah', 'Virginia', 'Virgin Islands', 'Vermont', 'Washington', 'Wisconsin', 'West Virginia', 'Wyoming']
```

```
In [91]: STATES.columns = ["ISO", "Lat", "Lon", "state"]
STATES.head()
```

Out[91]:

	ISO	Lat	Lon	state
0	AK	63.588753	-154.493062	Alaska
1	AL	32.318231	-86.902298	Alabama
2	AR	35.201050	-91.831833	Arkansas
3	AZ	34.048928	-111.093731	Arizona
4	CA	36.778261	-119.417932	California

```
In [92]: covid_LL = covid_latest
covid_LL.head()
```

Out[92]:

	date	county	state	cases	deaths
402068	2020-08-05	Autauga	Alabama	1073	22
402069	2020-08-05	Baldwin	Alabama	3380	24
402070	2020-08-05	Barbour	Alabama	615	5
402071	2020-08-05	Bibb	Alabama	392	4
402072	2020-08-05	Blount	Alabama	839	3

```
In [93]: merged_covid = pd.merge(scd, STATES, on = "state")
merged_covid.head()
```

Out[93]:

	total_cases	total_deaths	ISO	Lat	Lon	state
--	-------------	--------------	-----	-----	-----	-------

```
In [94]: total_cases_lst = merged_covid.total_cases.values.tolist()
m = interp1d([1, max(total_cases_lst)], [5, 50])
circle_radius = m(total_cases_lst)
fig = px.density_mapbox(merged_covid, lat = "Lat", lon = "Lon", radius = circle_radius, zoom = 3, mapbox_style = "open-street-map")
fig.show()
```

```
ValueError                                Traceback (most recent call last)
<ipython-input-94-372c31c932c1> in <module>
      1 total_cases_lst = merged_covid.total_cases.values.tolist()
----> 2 m = interp1d([1, max(total_cases_lst)], [5, 50])
      3 circle_radius = m(total_cases_lst)
      4 fig = px.density_mapbox(merged_covid, lat = "Lat", lon = "Lon", radius = circle_radius, zoom = 3, mapbox_style = "open-street-map")
      5 fig.show()
```

ValueError: max() arg is an empty sequence

**For the months of Mar - July, do a scatter plot for 5 US States (e.g. CA, NY ...) to show the trend lines for total number of cases and total number of deaths over the 5 months**

```
In [ ]:
```