In [ ]:

```
import pandas as pd
import numpy as np
import seaborn as sns
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
df = pd.read_csv(r"https://Applications/covid_deaths_usafacts.csv")
```

In [ ]:

```
df
```

In [112]:

```
df.columns
```

Out[112]:

```
Index(['County Name', 'State', '1/22/20', '1/23/20', '1/24/20', '1/25/20',
       '1/26/20', '1/27/20', '1/28/20', '1/29/20',
       ...
       '7/20/20', '7/21/20', '7/22/20', '7/23/20', '7/24/20', '7/25/20',
       '7/26/20', '7/27/20', '7/28/20', '7/29/20'],
      dtype='object', length=192)
```

In [111]:

```
df.head()
```

Out[111]:

	County Name	State	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	...	7/20/20	7/21/20	7/22/20	7/23/20	7/24/20
0	Statewide Unallocated	AL	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	Autauga County	AL	0	0	0	0	0	0	0	0	...	857	865	886	905	921
2	Baldwin County	AL	0	0	0	0	0	0	0	0	...	2013	2102	2196	2461	2513
3	Barbour County	AL	0	0	0	0	0	0	0	0	...	503	514	518	534	539
4	Bibb County	AL	0	0	0	0	0	0	0	0	...	279	283	287	289	303

5 rows × 192 columns

In [110]:

```
df.describe()
```

Out[110]:

	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	1/3
count	3195.000000	3195.000000	3195.000000	3195.000000	3195.000000	3195.000000	3195.000000	3195.000000	3195.000000	3195.00
mean	0.000313	0.000313	0.000626	0.000626	0.001565	0.001878	0.001878	0.002191	0.002191	0.00
std	0.017691	0.017691	0.025016	0.025016	0.039535	0.043301	0.043301	0.053038	0.053038	0.06
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00

	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	1/3
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	2.000000	2.000000	2.00

8 rows × 190 columns

In [14]:

```
df.tail()
```

Out[14]:

	countyFIPS	County Name	State	stateFIPS	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	7/20/20	7/21/20	7/22/20	7/23/20
3190	56037	Sweetwater County	WY	56	0	0	0	0	0	0	...	181	194	201	206
3191	56039	Teton County	WY	56	0	0	0	0	0	0	...	223	226	234	254
3192	56041	Uinta County	WY	56	0	0	0	0	0	0	...	222	223	224	227
3193	56043	Washakie County	WY	56	0	0	0	0	0	0	...	44	44	45	45
3194	56045	Weston County	WY	56	0	0	0	0	0	0	...	4	4	4	4

5 rows × 194 columns

In [17]:

```
df.isnull().sum()
```

Out[17]:

```
countyFIPS     0
County Name    0
State          0
stateFIPS      0
1/22/20        0
              ..
7/25/20        0
7/26/20        0
7/27/20        0
7/28/20        0
7/29/20        0
Length: 194, dtype: int64
```

In [19]:

```
sorted_df = df.sort_values(['States'].reset_index(drop=True))
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-19-bbfd8efeb7e2> in <module>
----> 1 sorted_df = df.sort_values(['States'].reset_index(drop=True))

NameError: name 'df' is not defined
```

In [20]:

```
sns.relplot(x = "State", y = "stateFIPS", data=df)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-20-ed471d5c72ed> in <module>
----> 1 sns.relplot(x = "State", y = "stateFIPS", data=df)

NameError: name 'df' is not defined
```

In [ ]:

In [ ]:

In [ ]:

In [21]:

```
sns.pairplot(df)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-21-1a6fe1782b2f> in <module>
----> 1 sns.pairplot(df)

NameError: name 'df' is not defined
```

In [ ]:

In [ ]:

Subselect New Jersey

In [89]:

```
is_state = df["State"] == "NJ"
data = df[is_state]
```

In [64]:

```
data
```

Out[64]:

	County Name	State	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	...	7/20/20	7/21/20	7/22/20	7/23/20	7/24
1805	Statewide Unallocated	NJ	0	0	0	0	0	0	0	0	...	650	688	682	671	6
1806	Atlantic County	NJ	0	0	0	0	0	0	0	0	...	3126	3148	3161	3169	31
1807	Bergen County	NJ	0	0	0	0	0	0	0	0	...	20034	20052	20099	20133	201
1808	Burlington County	NJ	0	0	0	0	0	0	0	0	...	5450	5478	5490	5502	55
1809	Camden County	NJ	0	0	0	0	0	0	0	0	...	7845	7867	7888	7910	79
1810	Cape May County	NJ	0	0	0	0	0	0	0	0	...	757	760	762	767	7
1811	Cumberland County	NJ	0	0	0	0	0	0	0	0	...	3079	3088	3095	3100	31
1812	Essex County	NJ	0	0	0	0	0	0	0	0	...	19121	19153	19173	19198	192
1813	Gloucester County	NJ	0	0	0	0	0	0	0	0	...	2829	2855	2870	2894	29
1814	Hudson County	NJ	0	0	0	0	0	0	0	0	...	19263	19267	19281	19289	193
1815	Hunterdon County	NJ	0	0	0	0	0	0	0	0	...	1100	1101	1102	1102	11
	Mercer															

	County Name	State	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	...	7/20/20	7/21/20	7/22/20	7/23/20	7/24
1816	Mercer County	NJ	0	0	0	0	0	0	0	0	...	7828	7839	7858	7869	78
1817	Middlesex County	NJ	0	0	0	0	0	0	0	0	...	17168	17155	17379	17442	174
1818	Monmouth County	NJ	0	0	0	0	0	0	0	0	...	9672	9683	9720	9748	97
1819	Morris County	NJ	0	0	0	0	0	0	0	0	...	6967	6990	7010	7016	70
1820	Ocean County	NJ	0	0	0	0	0	0	0	0	...	9956	9988	10054	10074	101
1821	Passaic County	NJ	0	0	0	0	0	0	0	0	...	17172	17181	17196	17214	172
1822	Salem County	NJ	0	0	0	0	0	0	0	0	...	831	837	841	843	8
1823	Somerset County	NJ	0	0	0	0	0	0	0	0	...	5068	5072	5082	5094	51
1824	Sussex County	NJ	0	0	0	0	0	0	0	0	...	1250	1253	1256	1259	12
1825	Union County	NJ	0	0	0	0	0	0	0	0	...	16297	16297	16297	16297	163
1826	Warren County	NJ	0	0	0	0	0	0	0	0	...	1280	1286	1295	1296	13

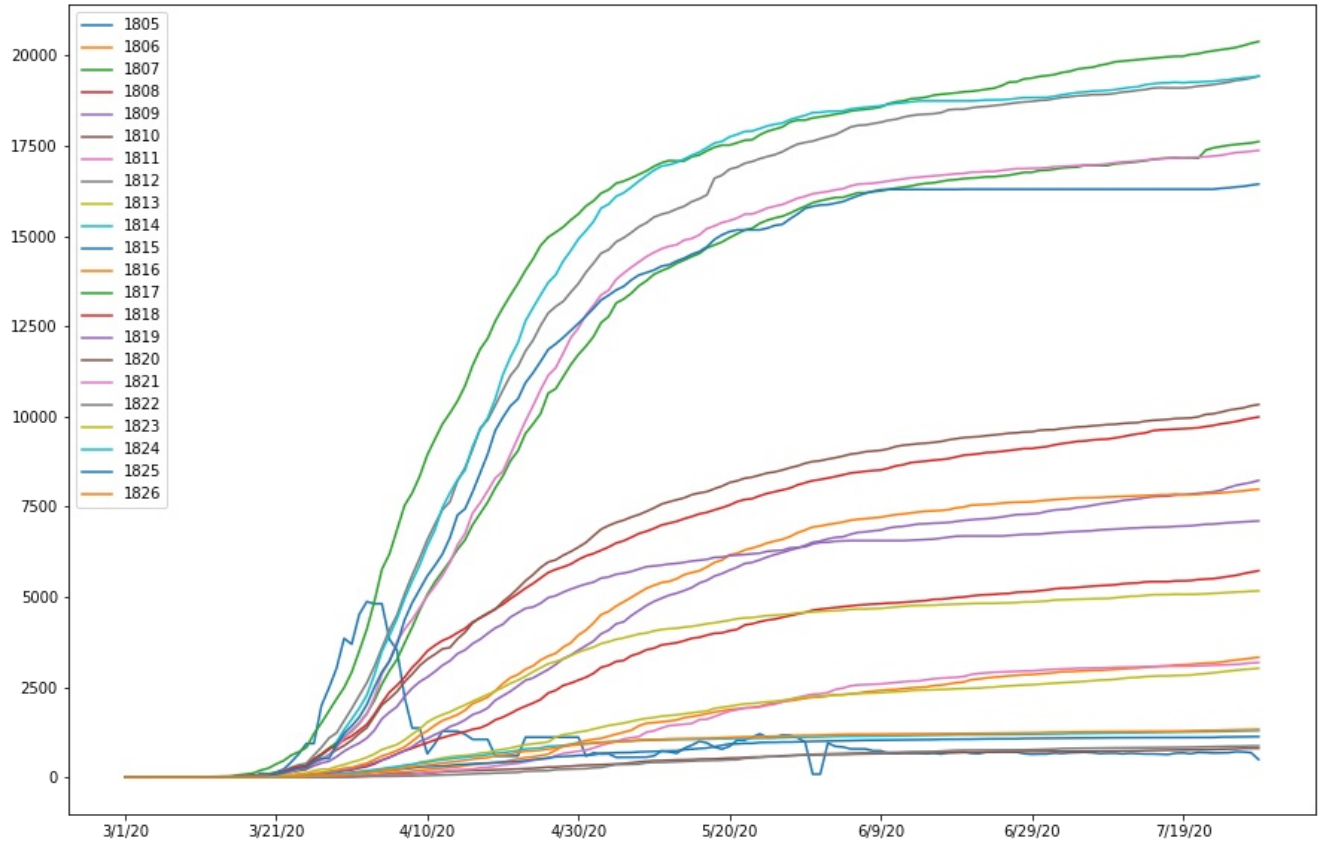
22 rows × 192 columns

In [58]:

```
data.T.loc['3/1/20':,:].plot(figsize=(15,10))
```

Out[58]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fefd6999190>
```



In [ ]:

```
#plot data for the specified counties in the NJ state
data.T.loc['3/1/20':,'Middlesex county'].plot(figsize=(15,10))
plt.show()
```

In [ ]:

```
data.head()
```

In [96]:

```
#sum the values for each county(i.e. by row)
data_totals = data.sum(axis=1)

#add a column called Totals with the county total
data['Totals'] = data_totals

#display the first 5 rows of the updated dataframe
data_NJ.head()
```

```
/Users/anishaagrawal/opt/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:5:
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
  """
```

Out[96]:

	County Name	State	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	...	7/20/20	7/21/20	7/22/20	7/23/20	7/24/
1805	Statewide Unallocated	NJ	0	0	0	0	0	0	0	0	...	650	688	682	671	6
1806	Atlantic County	NJ	0	0	0	0	0	0	0	0	...	3126	3148	3161	3169	31
1807	Bergen County	NJ	0	0	0	0	0	0	0	0	...	20034	20052	20099	20133	201
1808	Burlington County	NJ	0	0	0	0	0	0	0	0	...	5450	5478	5490	5502	55
1809	Camden County	NJ	0	0	0	0	0	0	0	0	...	7845	7867	7888	7910	79

5 rows × 192 columns

Plot top Eight Counties

In [98]:

```
#filter the top 8 counties totals
top = data.nlargest(8,'Totals')
#display the first eight rows of the new data frame
top.head(n=8)
```

Out[98]:

	County Name	State	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	...	7/21/20	7/22/20	7/23/20	7/24/20	7/25/2
1807	Bergen County	NJ	0	0	0	0	0	0	0	0	...	20052	20099	20133	20162	2019
1814	Hudson County	NJ	0	0	0	0	0	0	0	0	...	19267	19281	19289	19313	1932
1812	Essex County	NJ	0	0	0	0	0	0	0	0	...	19153	19173	19198	19228	1926
1821	Passaic County	NJ	0	0	0	0	0	0	0	0	...	17181	17196	17214	17237	1727
1817	Middlesex County	NJ	0	0	0	0	0	0	0	0	...	17155	17379	17442	17476	1751
1825	Union	NJ	0	0	0	0	0	0	0	0		16297	16297	16297	16321	1633

	County Name	State	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	...	7/21/20	7/22/20	7/23/20	7/24/20	7/25/2
1820	Ocean County	NJ	0	0	0	0	0	0	0	0	...	9988	10054	10074	10116	1017
1818	Monmouth County	NJ	0	0	0	0	0	0	0	0	...	9683	9720	9748	9791	982

8 rows × 193 columns

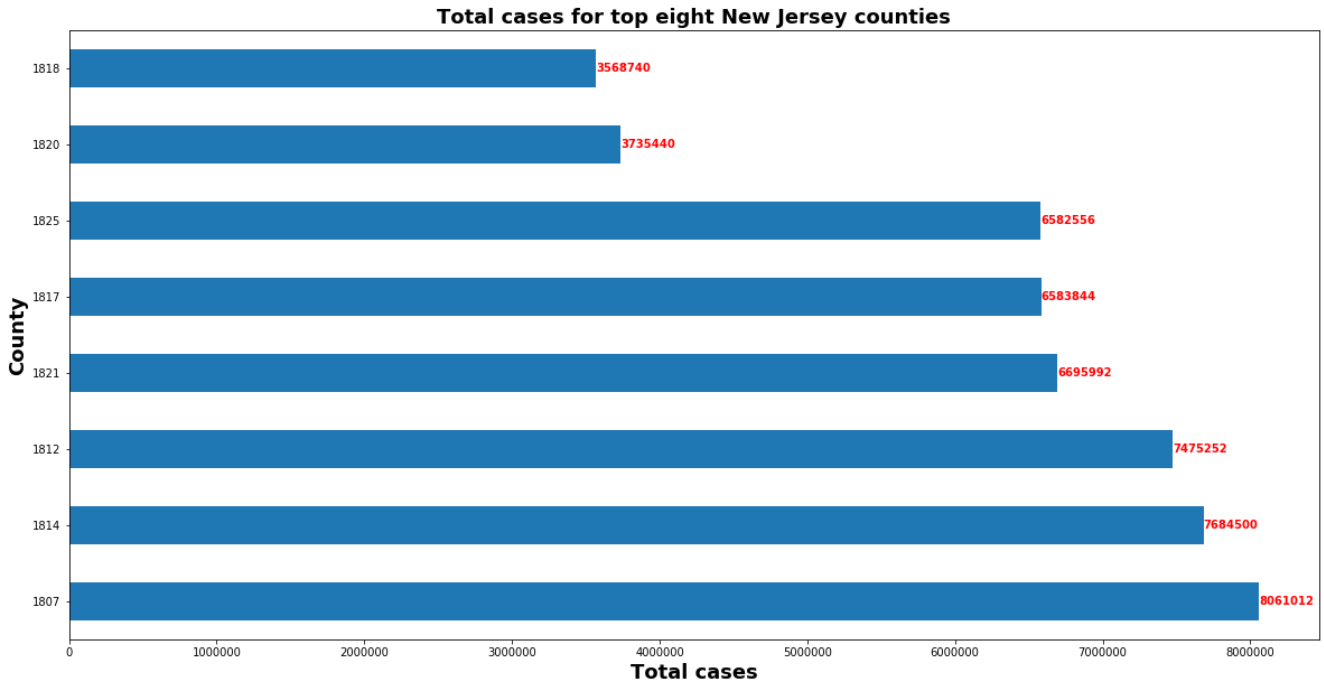
In [106]:

```
#display a bar chart

barplot = top.loc[:,'Totals'].plot(kind='barh', figsize=(20,10))
for i, v in enumerate(top.Totals):
    barplot.text(v + 3, i, ""+str(v), va= 'center', color='Red', fontweight='bold')
plt.title('Total cases for top eight New Jersey counties', fontsize= '18', fontweight = 'bold')
plt.xlabel('Total cases',fontsize= '18', fontweight = 'bold')
plt.ylabel('County',fontsize= '18', fontweight = 'bold')
```

Out[106]:

```
Text(0, 0.5, 'County')
```



In [ ]: