In [1]:
```
import pandas as pd
```

In [2]:
```
covidData = pd.read_csv("Datasets/covid.csv")
```

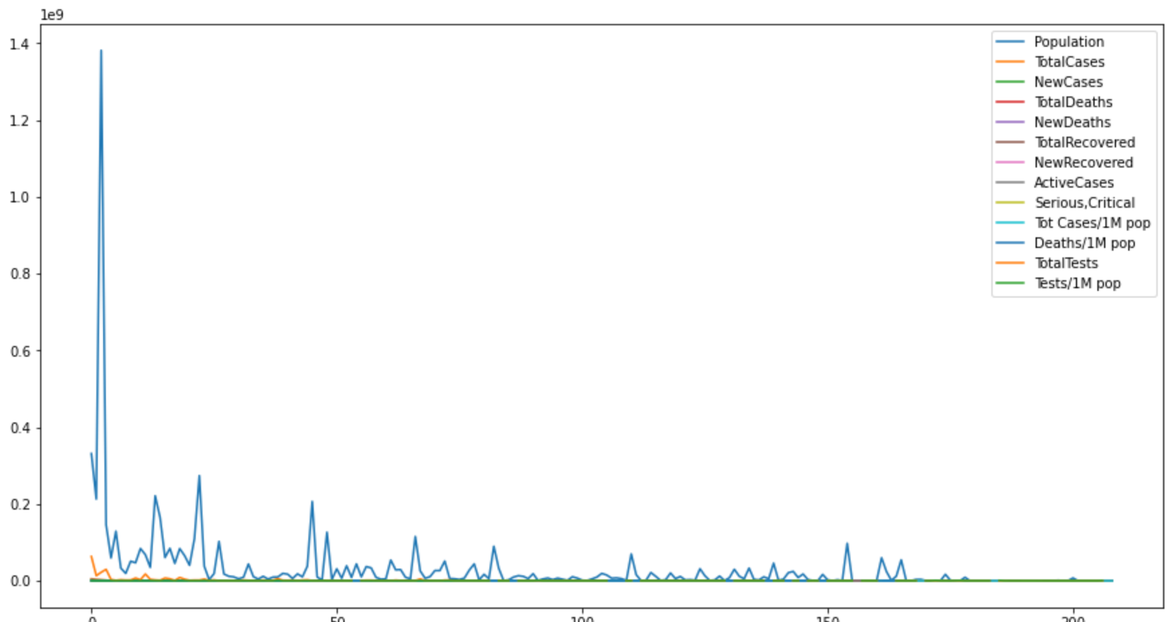
In [3]:
```
covidData.head()
```

Out[3]:

	Country/Region	Continent	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewRecovered	ActiveCases	Serious,Critical	Tot Cases/1M pop	Deaths/1M pop	TotalTests	Tests/1M pop	WHO Region
0	USA	North America	3.311981e+08	5032179	NaN	162804.0	NaN	2576668.0	NaN	2292707.0	18296.0	15194.0	492.0	63139605.0	190640.0	Americas
1	Brazil	South America	2.127107e+08	2917562	NaN	98644.0	NaN	2047660.0	NaN	771258.0	8318.0	13716.0	464.0	13206188.0	62085.0	Americas
2	India	Asia	1.381345e+09	2025409	NaN	41638.0	NaN	1377384.0	NaN	606387.0	8944.0	1466.0	30.0	22149351.0	16035.0	South-EastAsia
3	Russia	Europe	1.459409e+08	871894	NaN	14606.0	NaN	676357.0	NaN	180931.0	2300.0	5974.0	100.0	29716907.0	203623.0	Europe
4	South Africa	Africa	5.938157e+07	538184	NaN	9604.0	NaN	387316.0	NaN	141264.0	539.0	9063.0	162.0	3149807.0	53044.0	Africa

In [4]:
```
covidData.plot(figsize=(15,8))
```

Out[4]:
```
<AxesSubplot:>
```



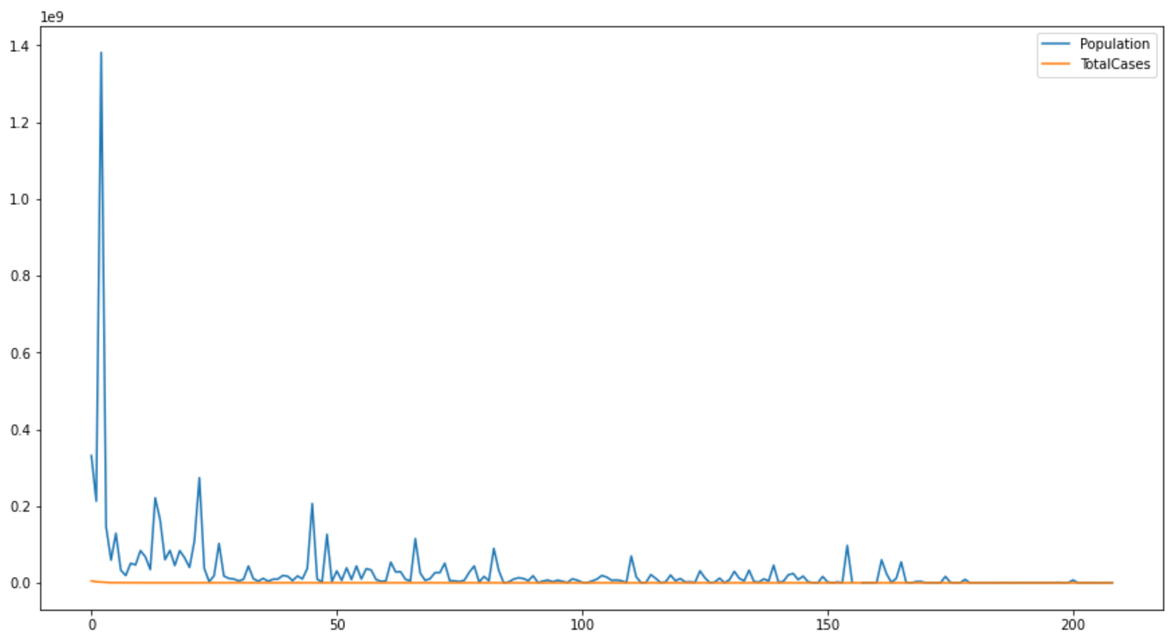
In [10]:
```
covidData.corr()
```

Out[10]:

	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewRecovered	ActiveCases	Serious,Critical	Tot Cases/1M pop	Deaths/1M pop	TotalTests	Tests/1M pop
Population	1.000000	0.546158	0.889610	0.431072	0.910836	0.590239	0.854194	0.459124	0.595379	-0.009337	0.025686	0.497937	-0.075129
TotalCases	0.546158	1.000000	0.999459	0.938622	0.998062	0.985764	0.998258	0.969423	0.967270	0.252627	0.280476	0.891001	0.029141
NewCases	0.889610	0.999459	1.000000	0.991894	0.995554	0.991416	0.999636	0.935500	0.985526	0.321203	0.842468	0.214572	-0.669339
TotalDeaths	0.431072	0.938622	0.991894	1.000000	0.999507	0.935410	0.988362	0.927625	0.906627	0.237206	0.425186	0.850304	0.053870
NewDeaths	0.910836	0.998062	0.995554	0.999507	1.000000	0.998935	0.992651	0.888717	0.997112	0.046499	0.751519	0.039460	-0.803470
TotalRecovered	0.590239	0.985764	0.991416	0.935410	0.998935	1.000000	0.986012	0.914566	0.949628	0.268483	0.296443	0.870113	0.022027
NewRecovered	0.854194	0.998258	0.999636	0.988362	0.992651	0.986012	1.000000	0.937663	0.980595	0.167036	0.825828	-0.081745	-0.869608
ActiveCases	0.459124	0.969423	0.935500	0.927625	0.888717	0.914566	0.937663	1.000000	0.945943	0.212899	0.255001	0.911488	0.028963
Serious,Critical	0.595379	0.967270	0.985526	0.906627	0.997112	0.949628	0.980595	0.945943	1.000000	0.209668	0.271842	0.859526	0.020649
Tot Cases/1M pop	-0.009337	0.252627	0.321203	0.237206	0.046499	0.268483	0.167036	0.212899	0.209668	1.000000	0.502134	0.194120	0.302728
Deaths/1M pop	0.025686	0.280476	0.842468	0.425186	0.751519	0.296443	0.825828	0.255001	0.271842	0.502134	1.000000	0.264218	0.156827
TotalTests	0.497937	0.891001	0.214572	0.850304	0.039460	0.870113	-0.081745	0.911488	0.859526	0.194120	0.264218	1.000000	0.096177
Tests/1M pop	-0.075129	0.029141	-0.669339	0.053870	-0.803470	0.022027	-0.869608	0.028963	0.020649	0.302728	0.156827	0.096177	1.000000

In [16]:
```
covidData[['Population','TotalCases']].plot(figsize=(15,8))
```

Out[16]:
```
<AxesSubplot:>
```



In [18]:
```
covidData[['Population','TotalCases']].corr()
```

Out[18]:

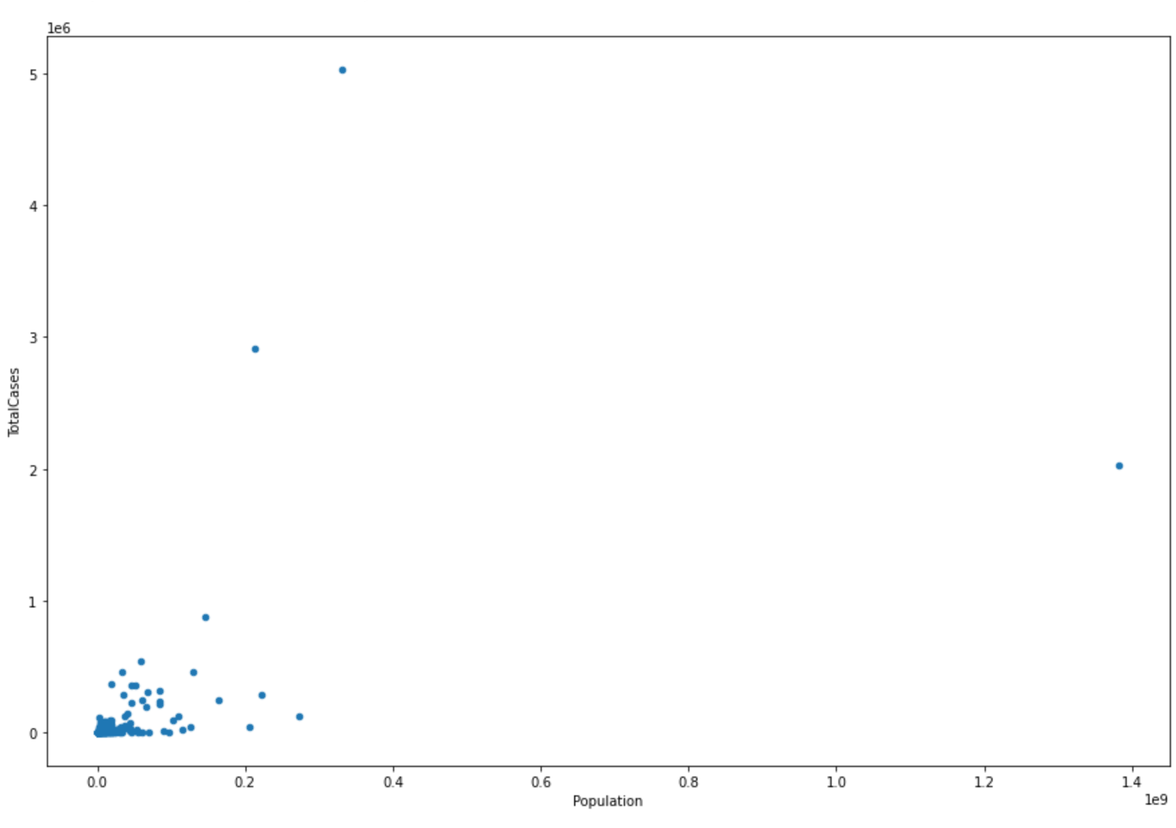
	Population	TotalCases
Population	1.000000	0.546158
TotalCases	0.546158	1.000000

In the scatter plot below I will analyse the population total with the total cases to see if they correlate.

Correlation between TotalCases and Population is 0.546158

In [14]:
```
covidData.plot(kind='scatter', x='Population', y='TotalCases', figsize=(15,10))
```

Out[14]:
```
<AxesSubplot:xlabel='Population', ylabel='TotalCases'>
```



It seems like there is some correlation, but not much. Most countries with a low population seem to have a less total cases, but its still quite scattered.

Since the correlation index is above 0 and even above 0.5 shows that a correlation is present.