

```
In [1]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv("New_York_State_Statewide_COVID-19_Testing.csv")
```

```
In [2]: df
```

Out[2]:	Test Date	County	New Positives	Cumulative Number of Positives	Total Number of Tests Performed	Cumulative Number of Tests Performed
	0	03/01/2020	Albany	0	0	0
	1	03/02/2020	Albany	0	0	0
	2	03/03/2020	Albany	0	0	0
	3	03/04/2020	Albany	0	0	0
	4	03/05/2020	Albany	0	0	3

	25663	04/14/2021	Yates	3	1101	210
	25664	04/15/2021	Yates	1	1102	218
	25665	04/16/2021	Yates	6	1108	214
	25666	04/17/2021	Yates	2	1110	228
	25667	04/18/2021	Yates	0	1110	186

25668 rows × 6 columns

```
In [3]: df.describe()
```

Out[3]:	New Positives	Cumulative Number of Positives	Total Number of Tests Performed	Cumulative Number of Tests Performed
count	25668.000000	25668.000000	25668.000000	2.566800e+04
mean	77.499922	11534.093112	1913.191990	2.517956e+05
std	216.239943	29594.954950	4264.492906	6.404424e+05
min	0.000000	0.000000	0.000000	0.000000e+00
25%	1.000000	159.000000	132.000000	9.130750e+03
50%	9.000000	1054.000000	377.000000	4.377900e+04
75%	40.000000	5355.000000	1308.250000	1.453280e+05
max	2722.000000	261599.000000	47155.000000	6.011397e+06

```
In [4]: df[df['Cumulative Number of Positives'] > 0]
```

Out[4]:	Test Date	County	New Positives	Cumulative Number of Positives	Total Number of Tests Performed	Cumulative Number of Tests Performed
	11	03/12/2020	Albany	1	1	11
	12	03/13/2020	Albany	2	3	26
	13	03/14/2020	Albany	3	6	108
	14	03/15/2020	Albany	6	12	108
	15	03/16/2020	Albany	11	23	140

	25663	04/14/2021	Yates	3	1101	210
	25664	04/15/2021	Yates	1	1102	218
	25665	04/16/2021	Yates	6	1108	214
	25666	04/17/2021	Yates	2	1110	228
	25667	04/18/2021	Yates	0	1110	186

24739 rows × 6 columns

```
In [5]: #Any_missing_values?
df.isnull().sum()
```

Out[5]:	Test Date	0
	County	0
	New Positives	0
	Cumulative Number of Positives	0
	Total Number of Tests Performed	0
	Cumulative Number of Tests Performed	0
	dtype: int64	

```
In [23]: #Standard_Deviation
df['New Positives'].std()
```

Out[23]:	216.2399425267739
----------	-------------------

```
In [24]: #Variance
df['New Positives'].var()
```

Out[24]:	46759.712743982476
----------	--------------------

```
In [27]: #Inter_Quartile_Range
IQR1= df['Cumulative Number of Positives'].describe()['75%']- df['Cumulative Number of Positives'].describe()['25%']
print("Cumulative Number of Positives IQR: ", IQR1)
IQR2= df['Cumulative Number of Tests Performed'].describe()['75%']- df['Cumulative Number of Tests Performed'].describe()['25%']
print("Cumulative Number of Tests Performed IQR: ", IQR2)
```

Cumulative Number of Positives IQR: 5196.0
Cumulative Number of Tests Performed IQR: 136197.25

```
In [6]: df.corr()
```

Out[6]:		New Positives	Cumulative Number of Positives	Total Number of Tests Performed	Cumulative Number of Tests Performed
	New Positives	1.000000	0.754922	0.776774	0.727878
	Cumulative Number of Positives	0.754922	1.000000	0.874101	0.941141
	Total Number of Tests Performed	0.776774	0.874101	1.000000	0.918066
	Cumulative Number of Tests Performed	0.727878	0.941141	0.918066	1.000000

```
In [7]: df[df['New Positives']>0].mean()
```

Out[7]:	New Positives	96.034952
	Cumulative Number of Positives	14261.405378
	Total Number of Tests Performed	2345.712320
	Cumulative Number of Tests Performed	309944.034614
	dtype: float64	

```
In [8]: df[df['New Positives']>0]['Total Number of Tests Performed'].mean()
```

Out[8]:	2345.7123201699333
---------	--------------------

```
In [14]: df.groupby(['County', 'New Positives']).mean()
```

Out[14]:	County	New Positives	Cumulative Number of Positives	Total Number of Tests Performed	Cumulative Number of Tests Performed
	Albany	0	0.000000	2.545455	7.272727
		1	1027.500000	397.500000	20940.000000
		2	2073.222222	573.777778	59482.777778
		3	1955.333333	552.777778	47917.666667
		4	2195.900000	757.700000	67080.400000

	Yates	19	515.000000	202.000000	20126.500000
		21	522.000000	151.000000	20787.000000
		24	673.000000	306.000000	22406.000000
		25	695.500000	263.500000	22712.500000
		26	649.000000	201.000000	22100.000000

6608 rows × 3 columns

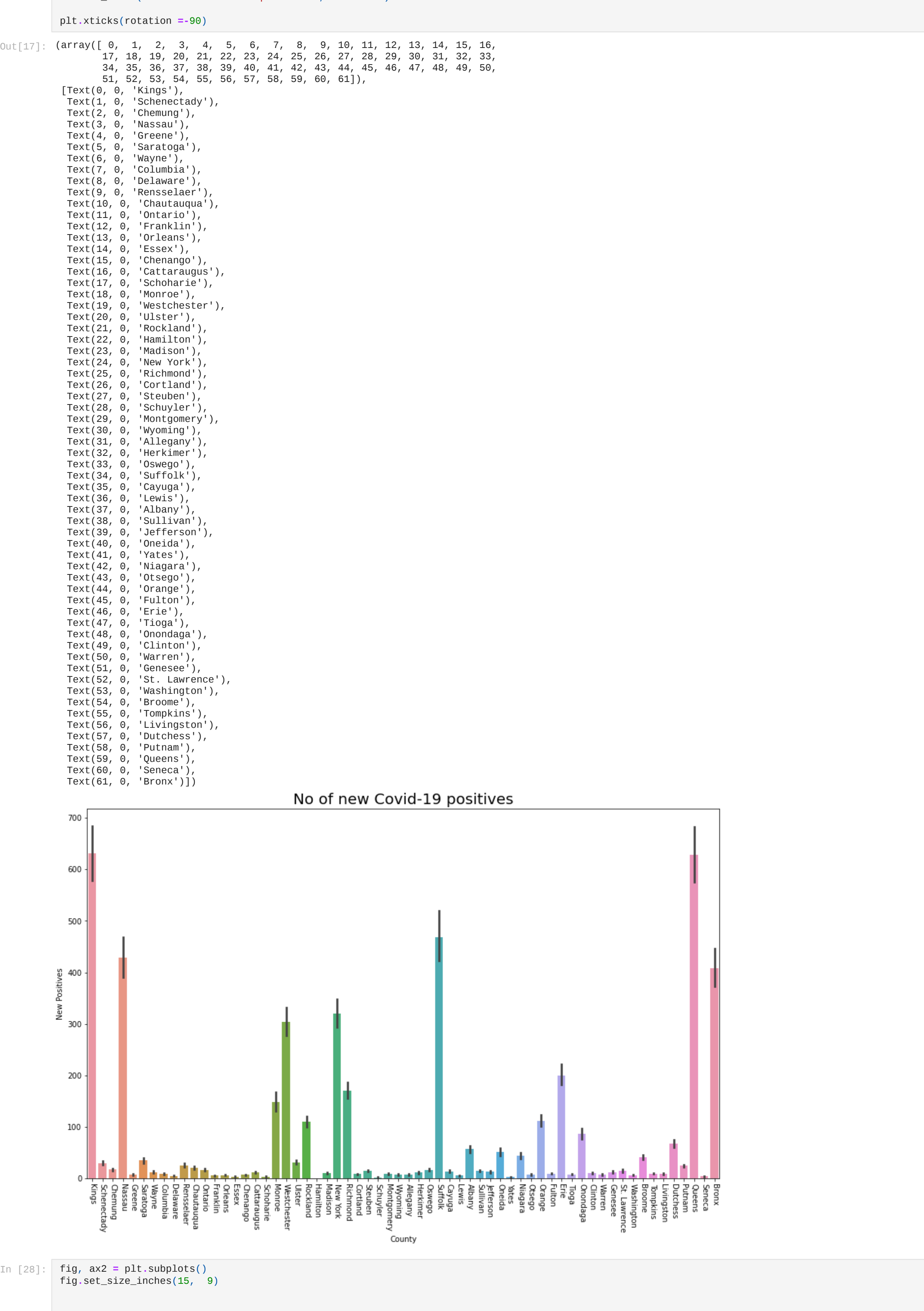
```
In [17]: fig, ax1 = plt.subplots()
fig.set_size_inches(15, 9)

ax1 = sns.barplot(x ="County", y ="New Positives",
                  data = df.sort_values('Test Date'),
                  )

ax1.set(xlabel ='County', ylabel ='New Positives')
ax1.set_title('No of new Covid-19 positives', size = 20)

plt.xticks(rotation =-90)
```

```
Out[17]: (array([ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50,
51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61]),
[Text(0, 0, 'Kings'),
Text(1, 0, 'Schenectady'),
Text(2, 0, 'Chemung'),
Text(3, 0, 'Nassau'),
Text(4, 0, 'Greene'),
Text(5, 0, 'Saratoga'),
Text(6, 0, 'Wayne'),
Text(7, 0, 'Columbia'),
Text(8, 0, 'Delaware'),
Text(9, 0, 'Rensselaer'),
Text(10, 0, 'Chautauqua'),
Text(11, 0, 'Ontario'),
Text(12, 0, 'Franklin'),
Text(13, 0, 'Orleans'),
Text(14, 0, 'Essex'),
Text(15, 0, 'Chenango'),
Text(16, 0, 'Cattaraugus'),
Text(17, 0, 'Schoharie'),
Text(18, 0, 'Monroe'),
Text(19, 0, 'Westchester'),
Text(20, 0, 'Ulster'),
Text(21, 0, 'Rockland'),
Text(22, 0, 'Hamilton'),
Text(23, 0, 'Madison'),
Text(24, 0, 'New York'),
Text(25, 0, 'Richmond'),
Text(26, 0, 'Cortland'),
Text(27, 0, 'Steuben'),
Text(28, 0, 'Schuyler'),
Text(29, 0, 'Montgomery'),
Text(30, 0, 'Wyoming'),
Text(31, 0, 'Allegany'),
Text(32, 0, 'Herkimer'),
Text(33, 0, 'Oswego'),
Text(34, 0, 'Suffolk'),
Text(35, 0, 'Cayuga'),
Text(36, 0, 'Lewis'),
Text(37, 0, 'Albany'),
Text(38, 0, 'Sullivan'),
Text(39, 0, 'Jefferson'),
Text(40, 0, 'Oneida'),
Text(41, 0, 'Yates'),
Text(42, 0, 'Niagara'),
Text(43, 0, 'Otsego'),
Text(44, 0, 'Orange'),
Text(45, 0, 'Fulton'),
Text(46, 0, 'Erie'),
Text(47, 0, 'Tioga'),
Text(48, 0, 'Onondaga'),
Text(49, 0, 'Clinton'),
Text(50, 0, 'Warren'),
Text(51, 0, 'Genesee'),
Text(52, 0, 'St. Lawrence'),
Text(53, 0, 'Washington'),
Text(54, 0, 'Broome'),
Text(55, 0, 'Tompkins'),
Text(56, 0, 'Livingston'),
Text(57, 0, 'Dutchess'),
Text(58, 0, 'Putnam'),
Text(59, 0, 'Queens'),
Text(60, 0, 'Seneca'),
Text(61, 0, 'Bronx')])
```



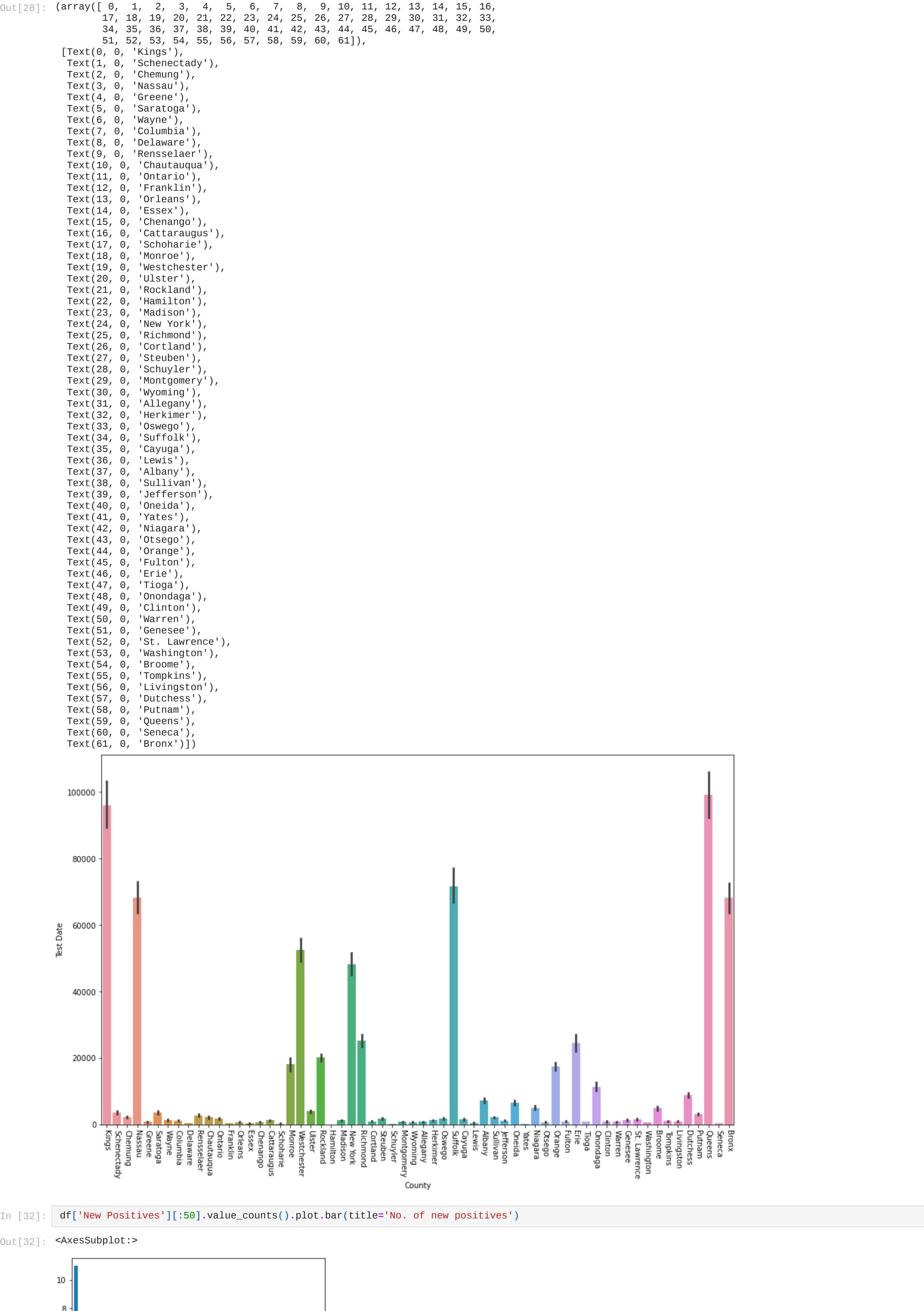
```
In [28]: fig, ax2 = plt.subplots()
fig.set_size_inches(15, 9)

ax2 = sns.barplot(
    x ="County", y ="Cumulative Number of Positives",
    data = df.sort_values('Test Date', ascending = 1),
    )

ax2.set(xlabel ='County', ylabel ='Test Date')
ax2.set_title('', size = 20)

plt.xticks(rotation =-90)
```

```
Out[28]: (array([ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50,
51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61]),
[Text(0, 0, 'Kings'),
Text(1, 0, 'Schenectady'),
Text(2, 0, 'Chemung'),
Text(3, 0, 'Nassau'),
Text(4, 0, 'Greene'),
Text(5, 0, 'Saratoga'),
Text(6, 0, 'Wayne'),
Text(7, 0, 'Columbia'),
Text(8, 0, 'Delaware'),
Text(9, 0, 'Rensselaer'),
Text(10, 0, 'Chautauqua'),
Text(11, 0, 'Ontario'),
Text(12, 0, 'Franklin'),
Text(13, 0, 'Orleans'),
Text(14, 0, 'Essex'),
Text(15, 0, 'Chenango'),
Text(16, 0, 'Cattaraugus'),
Text(17, 0, 'Schoharie'),
Text(18, 0, 'Monroe'),
Text(19, 0, 'Westchester'),
Text(20, 0, 'Ulster'),
Text(21, 0, 'Rockland'),
Text(22, 0, 'Hamilton'),
Text(23, 0, 'Madison'),
Text(24, 0, 'New York'),
Text(25, 0, 'Richmond'),
Text(26, 0, 'Cortland'),
Text(27, 0, 'Steuben'),
Text(28, 0, 'Schuyler'),
Text(29, 0, 'Montgomery'),
Text(30, 0, 'Wyoming'),
Text(31, 0, 'Allegany'),
Text(32, 0, 'Herkimer'),
Text(33, 0, 'Oswego'),
Text(34, 0, 'Suffolk'),
Text(35, 0, 'Cayuga'),
Text(36, 0, 'Lewis'),
Text(37, 0, 'Albany'),
Text(38, 0, 'Sullivan'),
Text(39, 0, 'Jefferson'),
Text(40, 0, 'Oneida'),
Text(41, 0, 'Yates'),
Text(42, 0, 'Niagara'),
Text(43, 0, 'Otsego'),
Text(44, 0, 'Orange'),
Text(45, 0, 'Fulton'),
Text(46, 0, 'Erie'),
Text(47, 0, 'Tioga'),
Text(48, 0, 'Onondaga'),
Text(49, 0, 'Clinton'),
Text(50, 0, 'Warren'),
Text(51, 0, 'Genesee'),
Text(52, 0, 'St. Lawrence'),
Text(53, 0, 'Washington'),
Text(54, 0, 'Broome'),
Text(55, 0, 'Tompkins'),
Text(56, 0, 'Livingston'),
Text(57, 0, 'Dutchess'),
Text(58, 0, 'Putnam'),
Text(59, 0, 'Queens'),
Text(60, 0, 'Seneca'),
Text(61, 0, 'Bronx')])
```



```
In [32]: df['New Positives'][ :50].value_counts().plot.bar(title='No. of new positives')
```

```
Out[32]: <AxesSubplot:~>
```

