

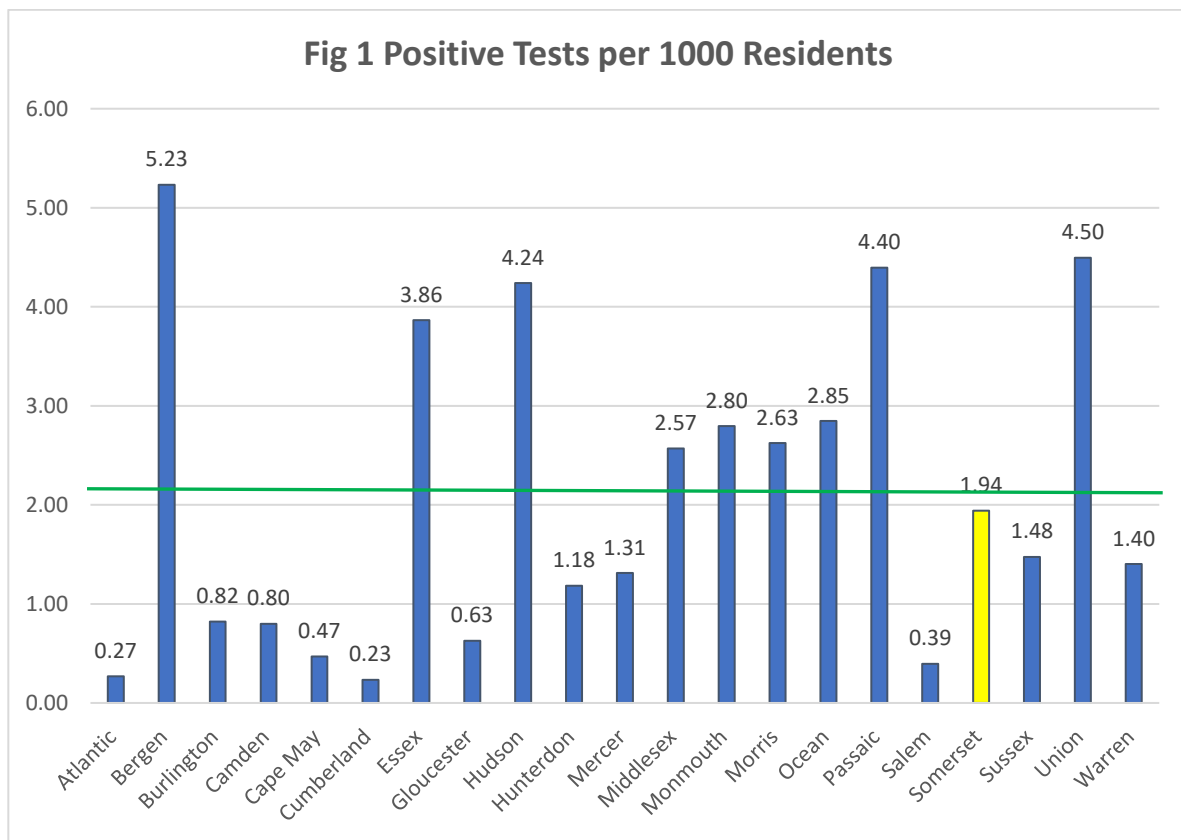
The Influence of population size, distance from NYC and population density on the number of positive tests in a county in NJ on 4/3/2020.

This is a technical post. If you are not into science the bottom line is the higher the population of a county and the closer the county is to NYC the higher the number of positive tests. Surprisingly, population density does not have a significant influence on the number of positive cases.

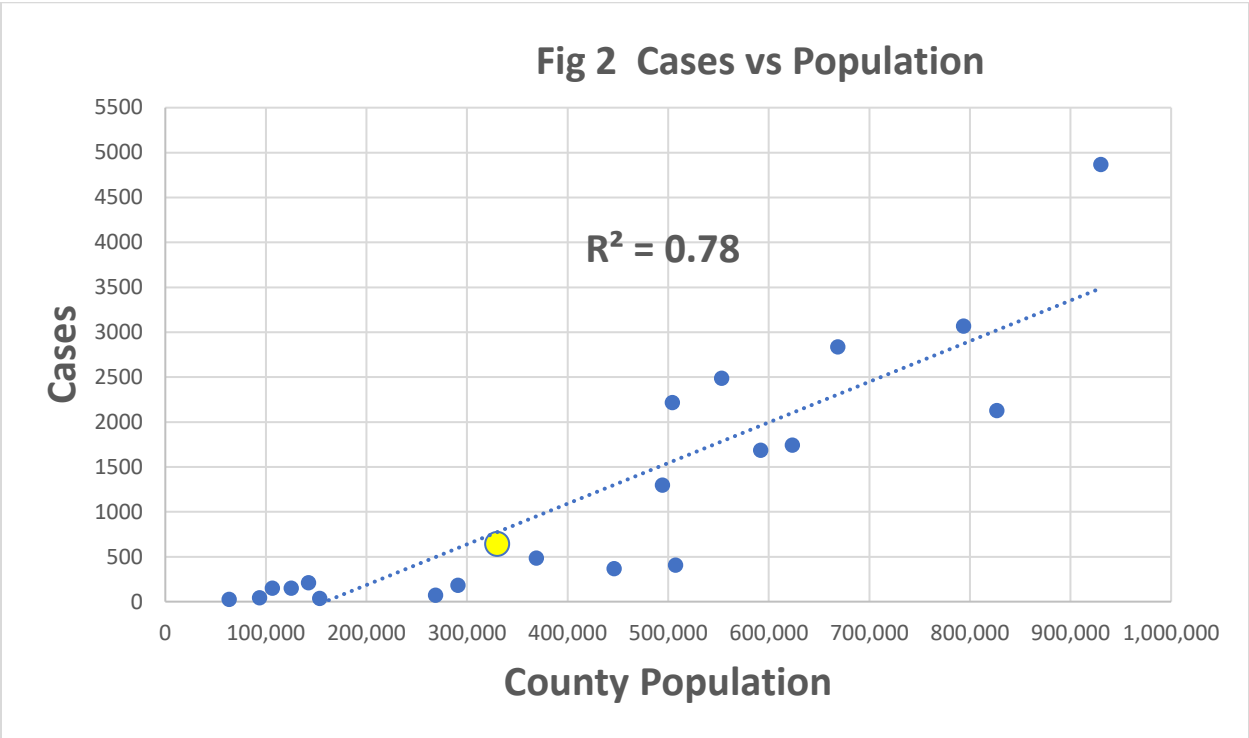
My original post on this was on [3/28/2020](#). I thought I'd look at it again with the current data. My conclusions are the same.

For the rest of this post I will use the word “cases” as shorthand for “positive tests”.

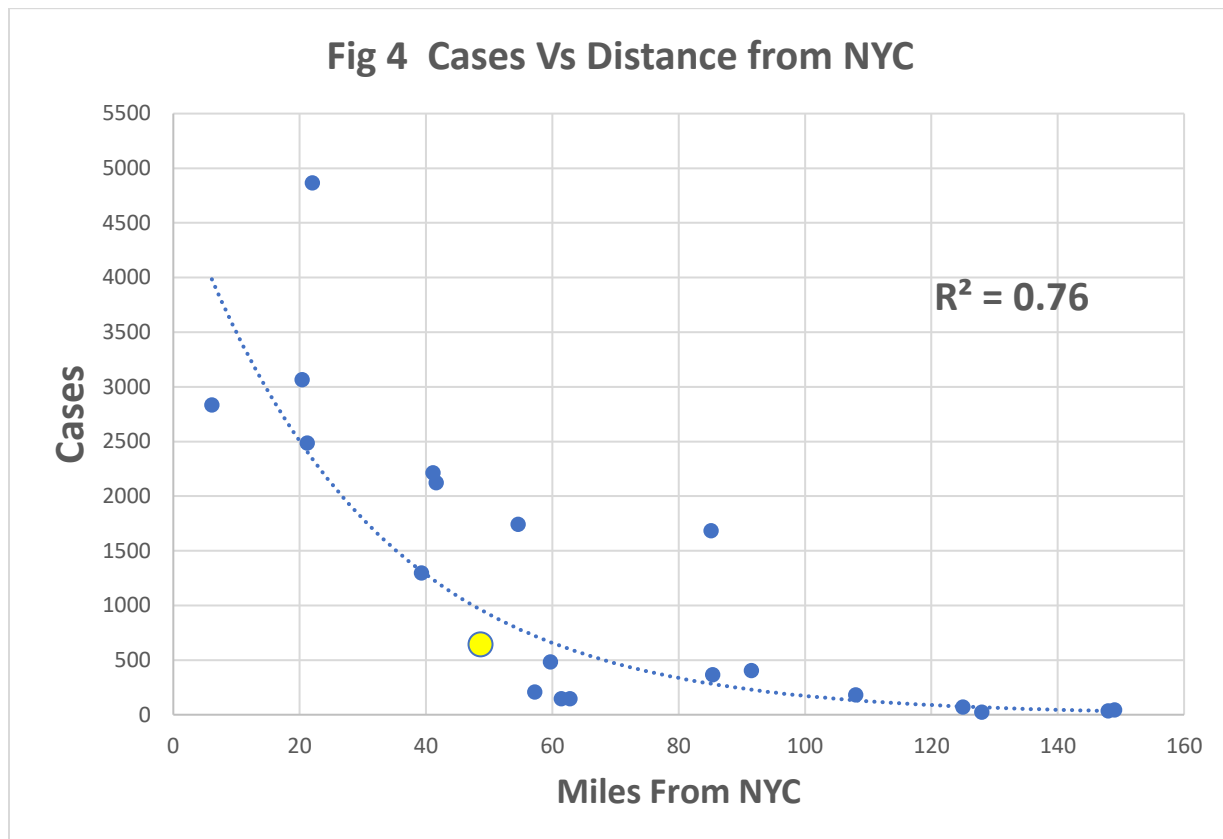
Fig 1 compares the numbers of cases county by county. I hand drew in the average value of 2.1. Somerset County is 1.9.



Does the population of a county influence the result (Fig 2)? On the graph you will note a dotted line. That is the line that best fits the data. You also notice $R^2 = 0.78$. R^2 is a measure of the “goodness” of the fit. Zero is no correlation whatsoever and 1 is perfect correlation. For medical issues anything around 0.8 is pretty good so this isn't bad. So, no surprise larger populations have more cases. The yellow dot is Somerset County and we are right where you'd expect.

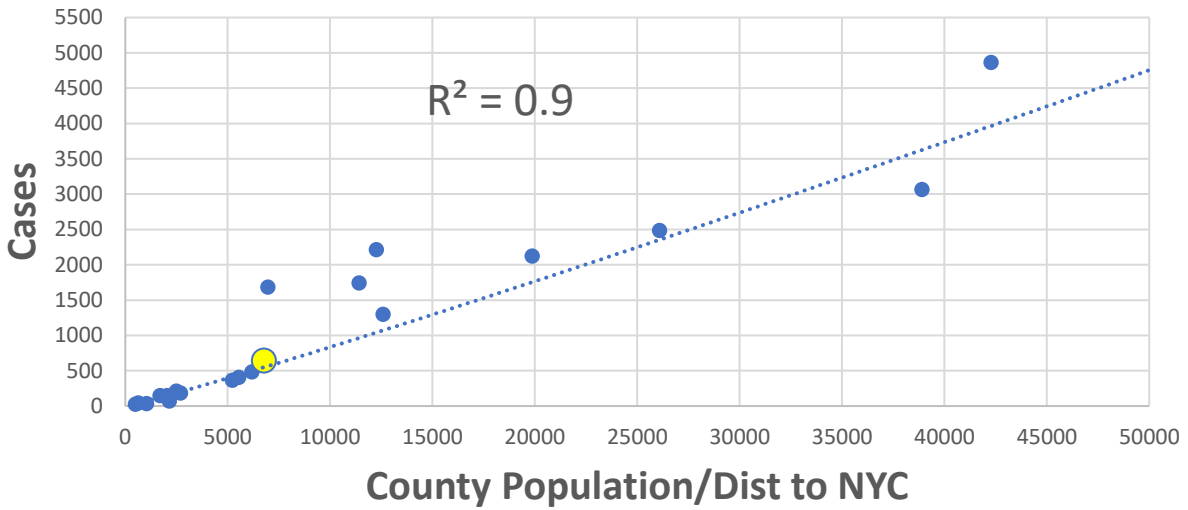


Does proximity to NYC influence the number of cases? To get the distances I used Google Maps to find directions from each county to the Empire State Building. That also provided distance. The result is in Figure 4. The curve fit uses a logarithmic fit. (For those technically minded I'm aware of the caveats of using log fits.) $R^2 = 0.76$ so proximity seems to matter. The closer to NYC the more cases. Somerset County is where you'd expect.



As a final exercise I examined how the population of a county together with its proximity to NYC influence the number of cases. I plotted the number of cases vs population divided by the distance to NYC for each county. The result is in Fig 5. This has a very good correlation. BTW Hudson County had an X axis value of almost 110,000 so for clarity I excluded it from the graph, but it was included in the R^2 calculation. The fit was a power fit, not a line.

Fig 5 How Pop and Distance From NYC affects the Number of Positive Results
 (Hudson County in calculation but not shown)



For the really technically minded a confounding factor is these calculations assume the population of a county is independent of its proximity to NYC. We all know that isn't true.

So, what we have learned is the greater the population of a county and the closer it is to NYC the higher the number of people who have tested positive. Population density is not a factor.

For completeness here is the data.

County	Population	Cases	Area sq mi	Dist NYC	cases / 1000	cases / sq mi	Pop Density	Pop/Dist
1 Atlantic	268,539	72	555.7	125	0.27	0.1	483	2148
2 Bergen	929,999	4,866	233.0	22	5.23	20.9	3991	42273
3 Burlington	446,367	367	798.6	85.4	0.82	0.5	559	5227
4 Camden	507,367	406	221.3	91.5	0.80	1.8	2293	5545
5 Cape May	93,705	44	251.4	149	0.47	0.2	373	629
6 Cumberland	153,400	36	483.7	148	0.23	0.1	317	1036
7 Essex	793,555	3,067	126.2	20.4	3.86	24.3	6288	38900
8 Gloucester	290,852	183	322.0	108	0.63	0.6	903	2693
9 Hudson	668,631	2835	46.2	6.1	4.24	61.4	14476	109612
10 Hunterdon	125,051	148	427.8	61.4	1.18	0.3	292	2037
11 Mercer	368,762	484	224.6	59.7	1.31	2.2	1642	6177

12	Middlesex	826,698	2125	308.9	41.6	2.57	6.9	2676	19873
13	Monmouth	623,387	1743	468.8	54.6	2.80	3.7	1330	11417
14	Morris	494,383	1298	460.2	39.3	2.63	2.8	1074	12580
15	Ocean	591,939	1685	628.8	85.1	2.85	2.7	941	6956
16	Passaic	504,041	2216	184.6	41.1	4.40	12.0	2731	12264
17	Salem	63,336	25	331.9	128	0.39	0.1	191	495
18	Somerset	330,176	641	301.8	48.7	1.94	2.1	1094	6780
19	Sussex	142,298	210	519.0	57.2	1.48	0.4	274	2488
20	Union	553,066	2487	102.9	21.2	4.50	24.2	5377	26088
21	Warren	106,293	149	356.9	62.8	1.40	0.4	298	1693

Ave 2.10