

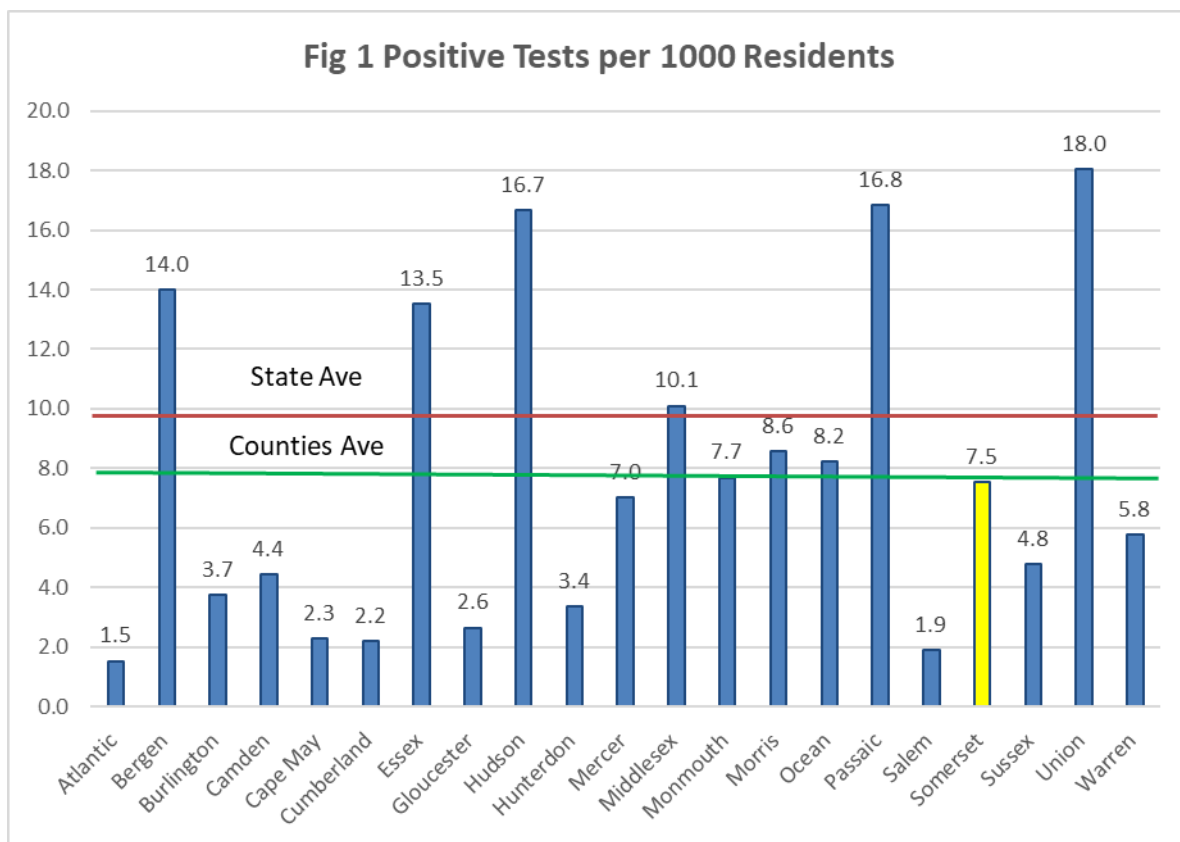
The Influence of population size, distance from NYC and population density on the number of positive tests in a county in NJ on 4/20/2020.

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. (See 2nd to last paragraph)

My original post on this was on [3/28/2020](#) and I updated it on [4/3/2020](#) and [4/11/2020](#). I thought I'd look at it again with the current data. My conclusions are the same.

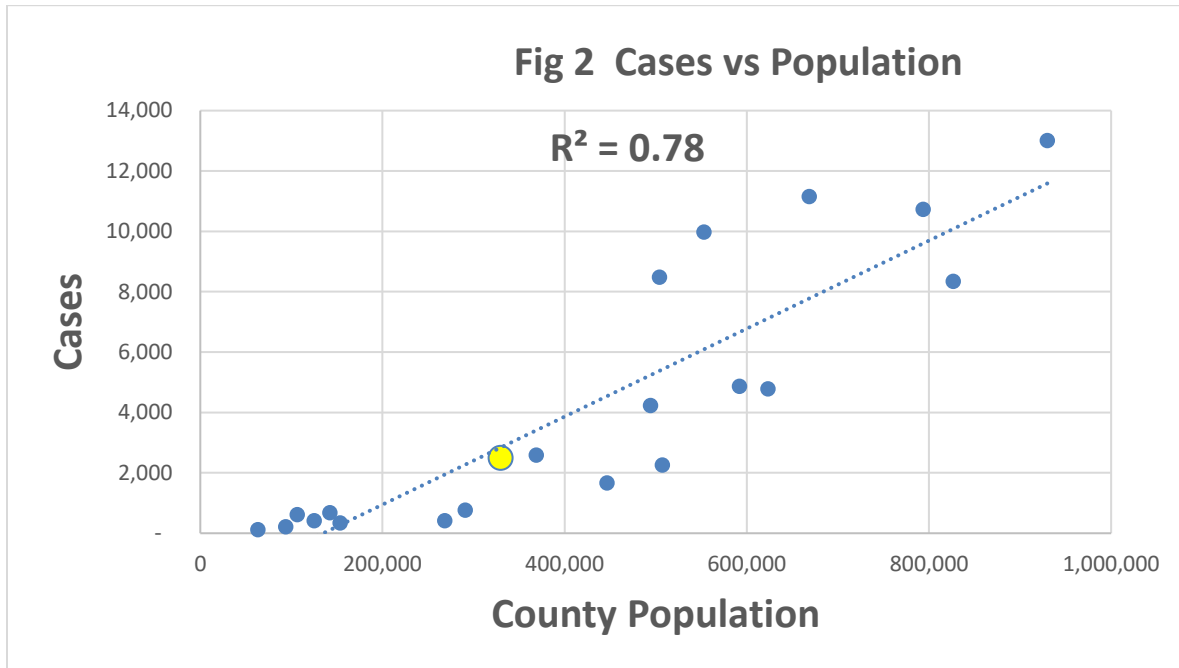
For the rest of this post I will use the word “cases” as shorthand for “positive tests”.

Fig 1 compares the numbers of cases county by county. I drew in the average value of counties 7.7 (green) and the state value of 9.9 (red). Somerset County is 7.5.

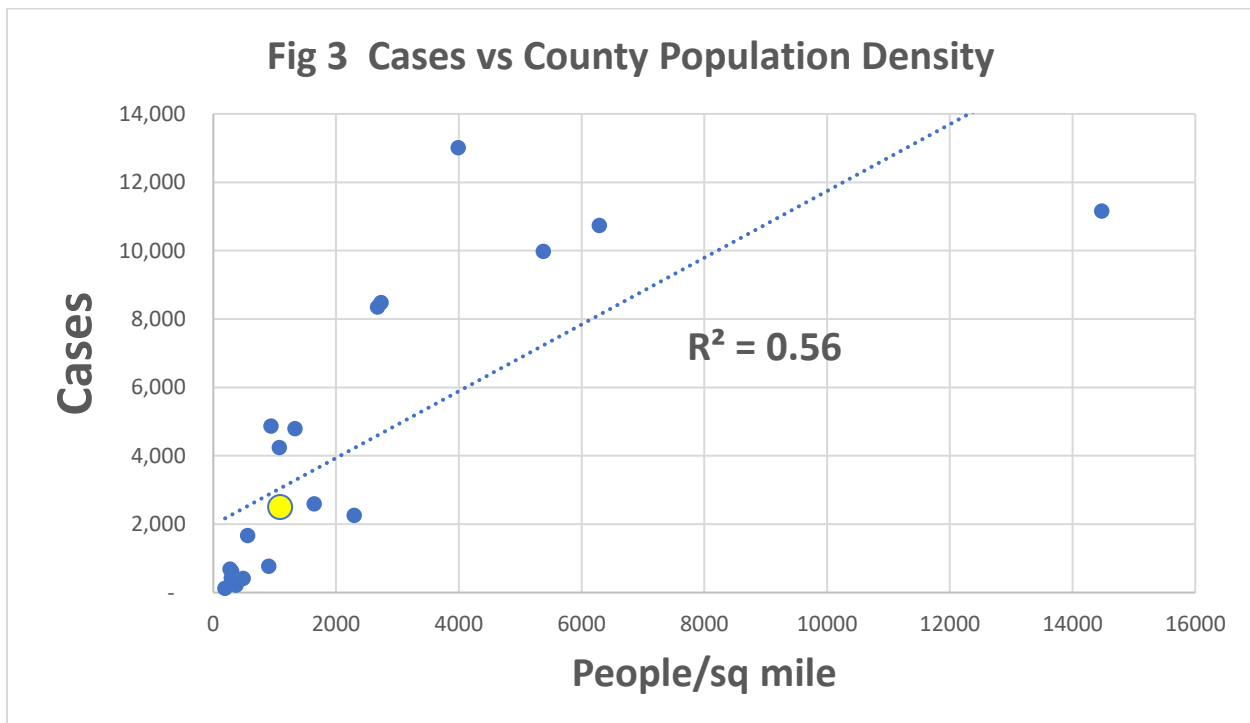


Does the population of a county influence the result (Fig 2)? On the graph you will note a dotted line. That is the line that best fits the data. $R^2 = 0.78$. R^2 is a measure of the “goodness” of the fit. Zero is no correlation and 1 is perfect correlation. For medical issues anything around 0.8 is pretty good so this isn't bad. No surprise

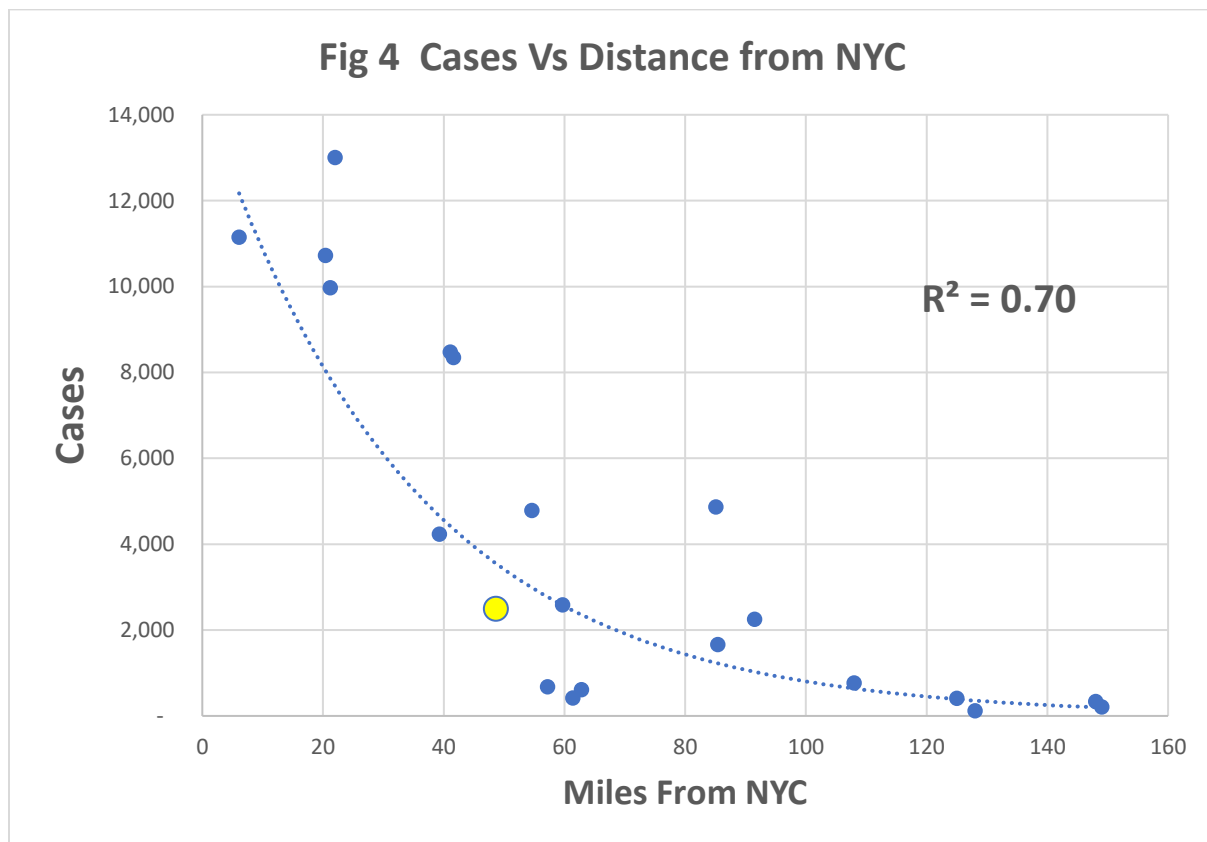
larger populations have more cases. The yellow dot is Somerset County, right where you'd expect.



Does population density influence the number of cases? Fig 3 addresses the question by looking at cases vs the number of people per square mile for each county. The R^2 is 0.56 so no it does not. Interestingly the first time I did the analysis $R^2 = 0.29$ and the next it was 0.42 and then 0.50. Perhaps in the end population density will matter, but not yet.

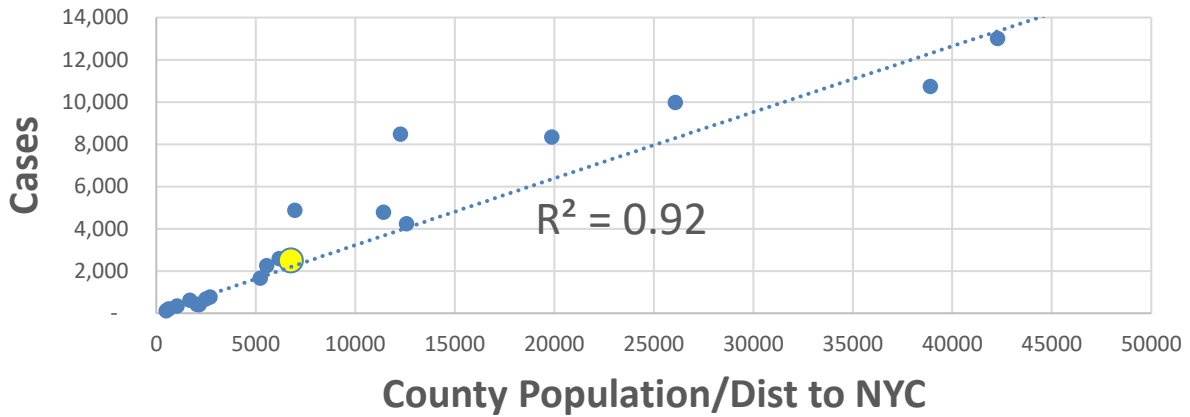


Does proximity to NYC influence the number of cases? To get the distances I used Google Maps to find the distance from each county to the Empire State Building. The result is in Figure 4. The curve fit uses a logarithmic fit. (For those technically minded I'm aware of the caveats of using log fits.) $R^2 = 0.70$ so proximity seems to matter. The closer to NYC the more cases. Somerset County is where you'd expect. The first time the $R^2 = 0.77$ and then 0.76 and 0.70 and now 0.70 again. It may be as the virus spreads distance from NYC is less important.



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. By the way, 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 very 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, with the caveat as time goes on, density may matter and proximity may not.

For completeness here is the data.

County	Population	Cases	Area sq mi	Dist NYC	Cases / 1000	cases/ sqmi	Pop Density	Pop/Dist
Atlantic	268,539	410	555.7	125	1.5	0.7	483	2148
Bergen	929,999	13,011	233.0	22	14.0	55.8	3991	42273
Burlington	446,367	1,663	798.6	85.4	3.7	2.1	559	5227
Camden	507,367	2,255	221.3	91.5	4.4	10.2	2293	5545
Cape May	93,705	212	251.4	149	2.3	0.8	373	629
Cumberland	153,400	338	483.7	148	2.2	0.7	317	1036
Essex	793,555	10,729	126.2	20.4	13.5	85.0	6288	38900

Gloucester	290,852	770	322.0	108	2.6	2.4	903	2693
Hudson	668,631	11,150	46.2	6.1	16.7	241.4	14476	109612
Hunterdon	125,051	419	427.8	61.4	3.4	1.0	292	2037
Mercer	368,762	2,591	224.6	59.7	7.0	11.5	1642	6177
Middlesex	826,698	8,346	308.9	41.6	10.1	27.0	2676	19873
Monmouth	623,387	4,787	468.8	54.6	7.7	10.2	1330	11417
Morris	494,383	4,236	460.2	39.3	8.6	9.2	1074	12580
Ocean	591,939	4,868	628.8	85.1	8.2	7.7	941	6956
Passaic	504,041	8,479	184.6	41.1	16.8	45.9	2731	12264
Salem	63,336	120	331.9	128	1.9	0.4	191	495
Somerset	330,176	2,488	301.8	48.7	7.5	8.2	1094	6780
Sussex	142,298	680	519.0	57.2	4.8	1.3	274	2488
Union	553,066	9,972	102.9	21.2	18.0	97.0	5377	26088
Warren	106,293	613	356.9	62.8	5.8	1.7	298	1693
Counties				Ave	7.7			
State	8,881,845	88,137		Ave	9.9			