Final Project - COVID 19 DATA

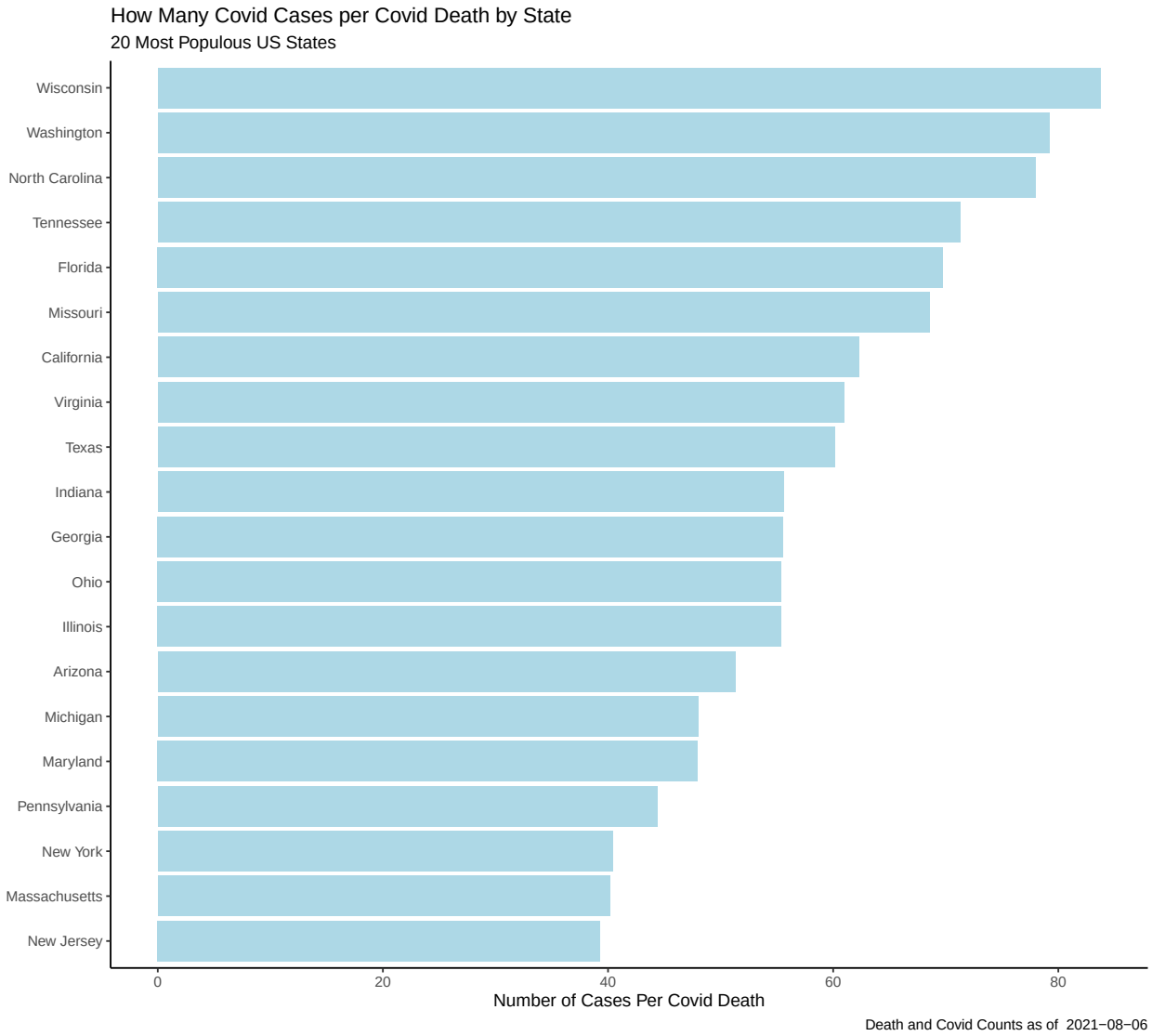
Covid 19 Data Analysis

Today we are using data provided by Johns Hopkins University. The data set takes a look at global and U.S.A. cases of Covid 19.

Today we are mainly going to look at the data for the US and analyze it at a county level. We will finally analyze with a model attempting to answer the question: Given the number of Covid cases in a US county, can we predict the amount of Covid related deaths?

US Analysis by State

How Many Covid Cases Does it Take Until a Death is Likely?



Province_State	Total_Cases	Total_Deaths	Cases_Per_Death
Alaska	77586	398	194.93970
Utah	438479	2494	175.81355
Virgin Islands	4907	40	122.67500
Nebraska	230630	2289	100.75579
Vermont	25320	260	97.38462
Idaho	204091	2217	92.05728
Northern Mariana Islands	183	2	91.50000
Wyoming	66453	786	84.54580
Hawaii	45245	540	83.78704
Wisconsin	695245	8300	83.76446
Colorado	581692	6978	83.36085
Minnesota	617788	7787	79.33582
Washington	488640	6168	79.22179
Maine	71141	900	79.04556
Oregon	226899	2889	78.53894

Province_State	Total_Cases	Total_Deaths	Cases_Per_Death
North Carolina	1071137	13736	77.98027
New Hampshire	101662	1389	73.19078
North Dakota	112169	1573	71.30896
Tennessee	914110	12820	71.30343
Florida	2768985	39695	69.75652
Missouri	703914	10265	68.57418
Montana	117882	1722	68.45645
Kentucky	495132	7372	67.16386
Oklahoma	493983	7531	65.59328
Kansas	340491	5305	64.18303
Arkansas	400275	6269	63.84990
South Carolina	634310	9950	63.74975
California	4031815	64731	62.28569
Iowa	382401	6193	61.74730
Delaware	112298	1833	61.26459
South Dakota	125599	2052	61.20809
Virginia	704664	11558	60.96764
Guam	8673	143	60.65035
Nevada	363574	6005	60.54521
Texas	3224572	53607	60.15207
Puerto Rico	150688	2604	57.86790
West Virginia	169162	2965	57.05295
Rhode Island	155825	2743	56.80824
Indiana	781326	14054	55.59456
Georgia	1211439	21802	55.56550
Ohio	1138600	20556	55.39015
Illinois	1436353	25936	55.38067
Alabama	603318	11600	52.01017
Arizona	940762	18342	51.29004
Louisiana	573903	11210	51.19563
New Mexico	213793	4425	48.31480
Michigan	1018601	21221	47.99967
Maryland	472224	9846	47.96100
Mississippi	358149	7621	46.99501
District of Columbia	51103	1149	44.47607
Pennsylvania	1238552	27898	44.39573
Connecticut	358076	8296	43.16249
New York	2168589	53694	40.38792
Massachusetts	726395	18095	40.14341
New Jersey	1046514	26636	39.28946
Grand Princess	103	3	34.33333

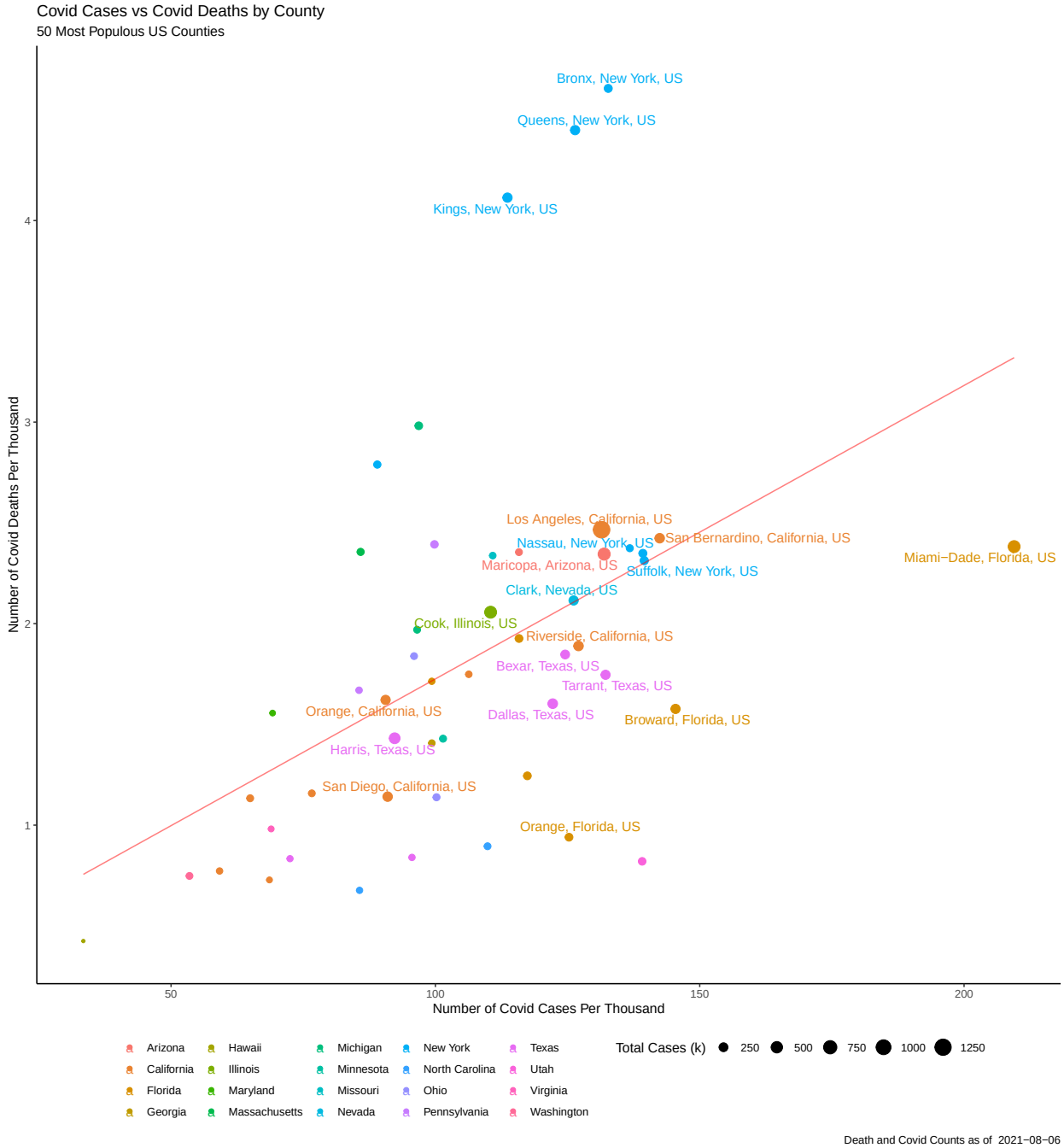
Covid Cases and Deaths by Most Populated Counties

Modeling the Number of Deaths (Per Thousand) as a Function of Cases (Per Thousand)

```
##
## Call:
## lm(formula = Deaths_Per_Thou ~ Cases_Per_Thou, data = county_data)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47477 -0.44776 -0.07542  0.17718  2.45421
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.269273   0.425870   0.632 0.530199
## Cases_Per_Thou 0.014563   0.003866   3.767 0.000452 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8111 on 48 degrees of freedom
## Multiple R-squared:  0.2281, Adjusted R-squared:  0.2121
## F-statistic: 14.19 on 1 and 48 DF,  p-value: 0.0004521
```

Which Counties had more Deaths Per Covid Case?



Final Analysis

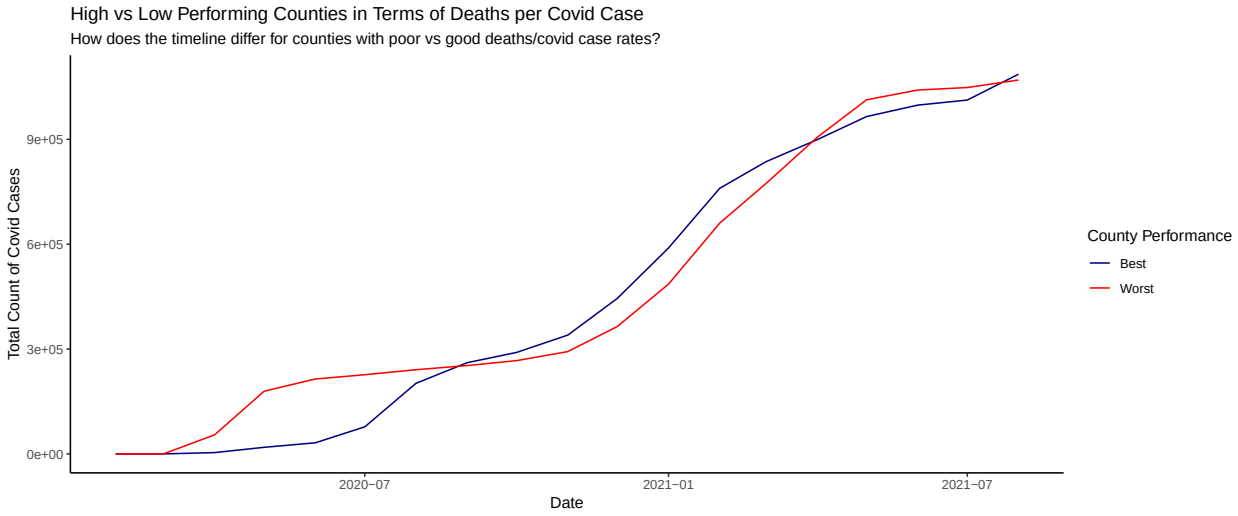
Digging Into County Data

Table 2: Counties Performing Best in Terms of Death Per Covid Case

County	Cases Per Thousand	Deaths Per Thousand	Predicted Deaths
Salt Lake, Utah, US	139.1131	0.8203806	2.295146
Orange, Florida, US	125.2702	0.9401113	2.093555
Mecklenburg, North Carolina, US	109.8639	0.8952084	1.869196
Miami-Dade, Florida, US	209.4687	2.3820916	3.319719
Wake, North Carolina, US	85.6551	0.6764044	1.516649

Table 3: Counties Performing Worst in Terms of Death Per Covid Case

County	Cases Per Thousand	Deaths Per Thousand	Predicted Deaths
Bronx, New York, US	132.6943	4.655879	2.201670
Queens, New York, US	126.4268	4.448816	2.110398
Kings, New York, US	113.6309	4.113828	1.924054
Wayne, Michigan, US	96.8535	2.981691	1.679729
New York, New York, US	89.0130	2.789331	1.565549



Difference in Timelines This final plot shows drastic differences in the volume of Covid 19 cases for counties who had low deaths per case (blue) and high deaths per case (red).

It appears counties who had high number of deaths per case reached higher volumes of cases earlier in the pandemic. This suggests that treatment was not as effective against positive Covid 19 patients early on in the pandemic versus later on. Therefore, it is likely that having a low rate of deaths per Covid case was not so much a measure of success within the county as much as it was at what stage in the pandemic did cases spike within the county.

Bias Areas

I think the conclusion that I suggested in the above paragraph was a belief that I suspected prior to doing this analysis. I attempted to combat my bias by questioning my belief head on to see if my suspicions were correct. However, there are likely numerous factors in play other than when cases spiked that attribute to high death per case rate. Perhaps vaccination rates by state or county would affect these rates later on in the pandemic. The age, race and underlying population make-up of a location is likely a big factor on the death per case rate as well.

One other belief I held prior to this analysis was that Florida would have higher rates of death per case. This is because of my belief that the age of a typical Florida resident skews older than the age of a typical American. However, when looking at the rates of death per case by state and most populous counties this actually turned out to not be the case. Overall, Florida had one of the better deaths per case ratios amongst heavily populated states. This is most evident in the first visual on page 1.