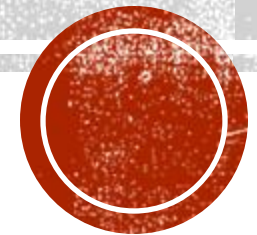


GRAND CHALLENGES IN COMPUTER SCIENCE — PROJECT 1 FINAL PRESENTATION



Understanding COVID-19 in the US, under the context of community health, socio-economic factors and mobility trends

Presented by – Pragya Kaushik



Calls for health funding to be prioritised as poor bear brunt of Covid-19

Most deprived areas of England and Wales have 55.1 deaths per 100,000 people, compared with 25.3 in affluent areas

- [Coronavirus - latest updates](#)
- [See all our coronavirus coverage](#)

Helen Pidd, Caelainn Barr and Aamna Mohdin

Sat 2 May 2020 04:12 AEST



widening inequality

CAROLINA SÁNCHEZ-PÁRAMO, RUTH HILL, DANIEL GERSZON MAHLER, AMBAR NARAYAN & NISHANT YONZAN | OCTOBER 07, 2021

This page in: [English](#) | [Français](#) | [Español](#) | [中文](#) | [العربية](#)



World ▾ Business ▾ Legal ▾ Markets ▾ Breakingviews Technology ▾ Investigations More ▾

April 4, 2022
9:46 PM GMT+9:30
Last Updated 5 months ago

United States

U.S. poor died at much higher rate from COVID than rich, report says

Reuters

Understanding COVID-19 in the US, under the context of community health, socio-economic factors and mobility trends

PROBLEM RELEVANCE

- COVID has impacted the society unequally; some were more vulnerable than others due to health and socio-economic factors
- As we are paving our way out of the COVID-19 pandemic, governments all around the world, including the US, need to construct policies that target and help the ones in need.
- Through my project, I have investigated the relative associations between COVID-19 cases, mortality, mobility trends, as well as sociodemographic, health and economic factors
- Such analysis allows policy makers to get an insight into how the COVID pandemic has impacted lives from different socio-economic backgrounds, and thus guide them to make better targeted policies.



METHODS

Combined 8 different datasets from 5 major sources:

- Centres for Disease Control and Prevention (US Government agency)
- Google COVID-19 Community Mobility Report
- U.S. Department of Agriculture
- County Health Rankings (by the University of Wisconsin-Madison)
- USAFacts

Keeping population density in mind, I chose to analyse the 30 most populated counties in the US.

Wrote code in Python; imported libraries like Seaborn, Matplotlib, NumPy and Pandas

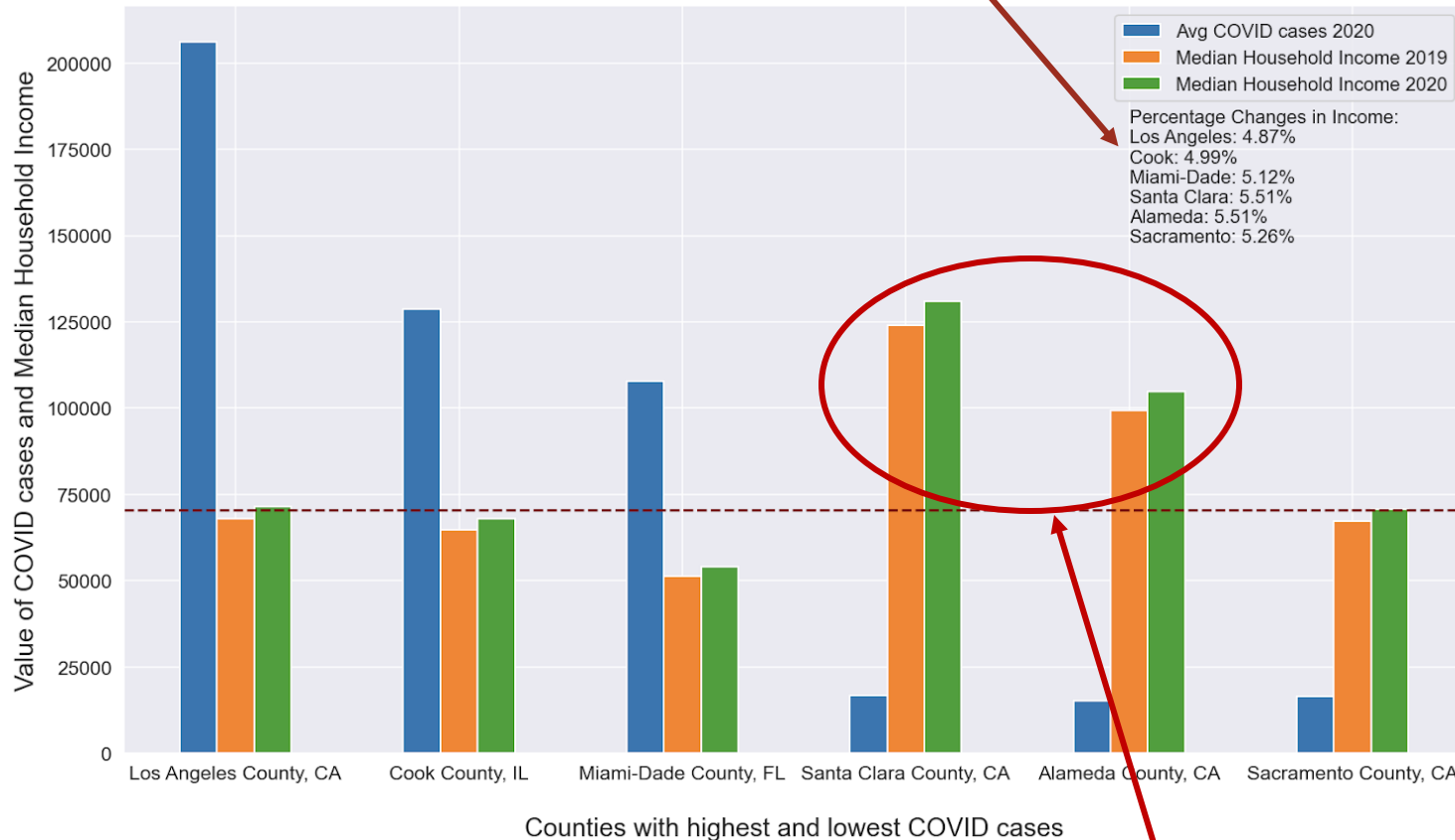
Since I had to filter out data of the same 30 counties from different datasets, I used one common dictionary to make things easier.

```
dictionary = {'Los Angeles County, California': 6037,  
             'Cook County, Illinois': 17031,  
             'Harris County, Texas': 48201,  
             'Maricopa County, Arizona': 4013,  
             'San Diego County, California': 6073,  
             'Orange County, California': 6059,  
             'Miami-Dade County, Florida': 12086,  
             'Kings County, New York': 36047,  
             'Dallas County, Texas': 48113,  
             'Riverside County, California': 6065,  
             'Queens County, New York': 36081,  
             'Clark County, Nevada': 32003,  
             'King County, Washington': 53033,  
             'San Bernardino County, California': 6071,  
             'Tarrant County, Texas': 48439,  
             'Bexar County, Texas': 48029,  
             'Broward County, Florida': 12011,  
             'Santa Clara County, California': 6085,  
             'Wayne County, Michigan': 26163,  
             'Alameda County, California': 6001,  
             'Middlesex County, Massachusetts': 25017,  
             'Sacramento County, California': 6067,  
             'New York County, New York': 36061,  
             'Philadelphia County, Pennsylvania': 42101,  
             'Suffolk County, New York': 36103,  
             'Palm Beach County, Florida': 12099,  
             'Hillsborough County, Florida': 12057,  
             'Bronx County, New York': 36005,  
             'Orange County, Florida': 12095,  
             'Nassau County, New York': 36059,  
             }
```



Percentage Changes in Income (from 2019 to 2020)

Bar chart: Counties with highest and lowest COVID cases and impact on Median Household Income



Higher income counties had lower number of COVID-19 cases

GRAPH 1

- In general, counties with lower income have higher number of cases and counties with higher income have lower number of cases
- Los Angeles, which is the county with the most cases, had the lowest increase in income from 2019 to 2020.
- Whereas, in Santa Clara and Alameda counties, which have the lowest number of cases, had the most increase in income.



CALCULATING AVERAGES

1

date

2020-02-15

2020-02-16

2020-02-17

2020-02-18

2020-02-19

...

2020-12-27

2020-12-28

2020-12-29

2020-12-30

2020-12-31

sub_region_1	sub_region_2	census_fips_code	date	retail_and_recreation_percent_change_from_baseline	year	month	day
Arizona	Maricopa County	4013.0	2020-02-15	2.0	2020	2	15
Arizona	Maricopa County	4013.0	2020-02-16	5.0	2020	2	16
Arizona	Maricopa County	4013.0	2020-02-17	9.0	2020	2	17
Arizona	Maricopa County	4013.0	2020-02-18	1.0	2020	2	18
Arizona	Maricopa County	4013.0	2020-02-19	4.0	2020	2	19
...
Washington	King County	53033.0	2020-12-27	-41.0	2020	12	27
Washington	King County	53033.0	2020-12-28	-30.0	2020	12	28
Washington	King County	53033.0	2020-12-29	-30.0	2020	12	29
Washington	King County	53033.0	2020-12-30	-31.0	2020	12	30
Washington	King County	53033.0	2020-12-31	-28.0	2020	12	31

3

	month	Los Angeles County, California	Cook County, Illinois	Harris County, Texas	Maricopa County, Arizona	San Diego County, California	Orange County, California	Miami-Dade County, Florida	Kings County, New York	Dallas County, Texas	...	Middlesex County, Massachusetts	Sacramento County, California	New York County, New York
0	Mar	-21.064516	-20.516129	-14.548387	-12.870968	-21.193548	-21.935484	-21.129032	-24.193548	-15.322581	...	-23.935484	-18.290323	-44.19354
1	Apr	-51.766667	-48.100000	-37.333333	-37.933333	-53.866667	-54.000000	-51.400000	-63.100000	-39.100000	...	-55.866667	-44.833333	-85.70000
2	May	-45.741935	-38.322581	-21.548387	-26.193548	-46.903226	-47.000000	-40.387097	-49.387097	-24.612903	...	-45.322581	-38.967742	-80.16129
3	Jun	-34.033333	-23.733333	-16.633333	-21.533333	-34.200000	-34.400000	-28.533333	-36.766667	-18.066667	...	-32.166667	-28.033333	-73.13333
4	Jul	-32.096774	-16.322581	-20.903226	-26.935484	-31.419355	-33.258065	-30.967742	-27.193548	-20.516129	...	-25.129032	-28.354839	-65.19354
5	Aug	-31.290323	-17.032258	-18.806452	-25.548387	-29.354839	-31.193548	-31.580645	-25.000000	-19.064516	...	-23.387097	-29.225806	-61.74193
6	Sep	-30.500000	-18.033333	-17.700000	-22.366667	-27.066667	-28.300000	-28.433333	-22.500000	-18.000000	...	-22.300000	-27.433333	-58.26666
7	Oct	-29.677419	-21.129032	-16.354839	-19.677419	-25.709677	-27.161290	-25.322581	-28.290323	-18.096774	...	-25.225806	-24.774194	-57.80645
8	Nov	-31.300000	-29.466667	-17.766667	-19.766667	-29.133333	-29.200000	-25.666667	-30.266667	-20.633333	...	-28.900000	-27.700000	-60.36666
9	Dec	-34.354839	-28.645161	-16.032258	-18.387097	-32.290323	-32.903226	-21.193548	-32.838710	-19.967742	...	-30.709677	-29.548387	-62.96774

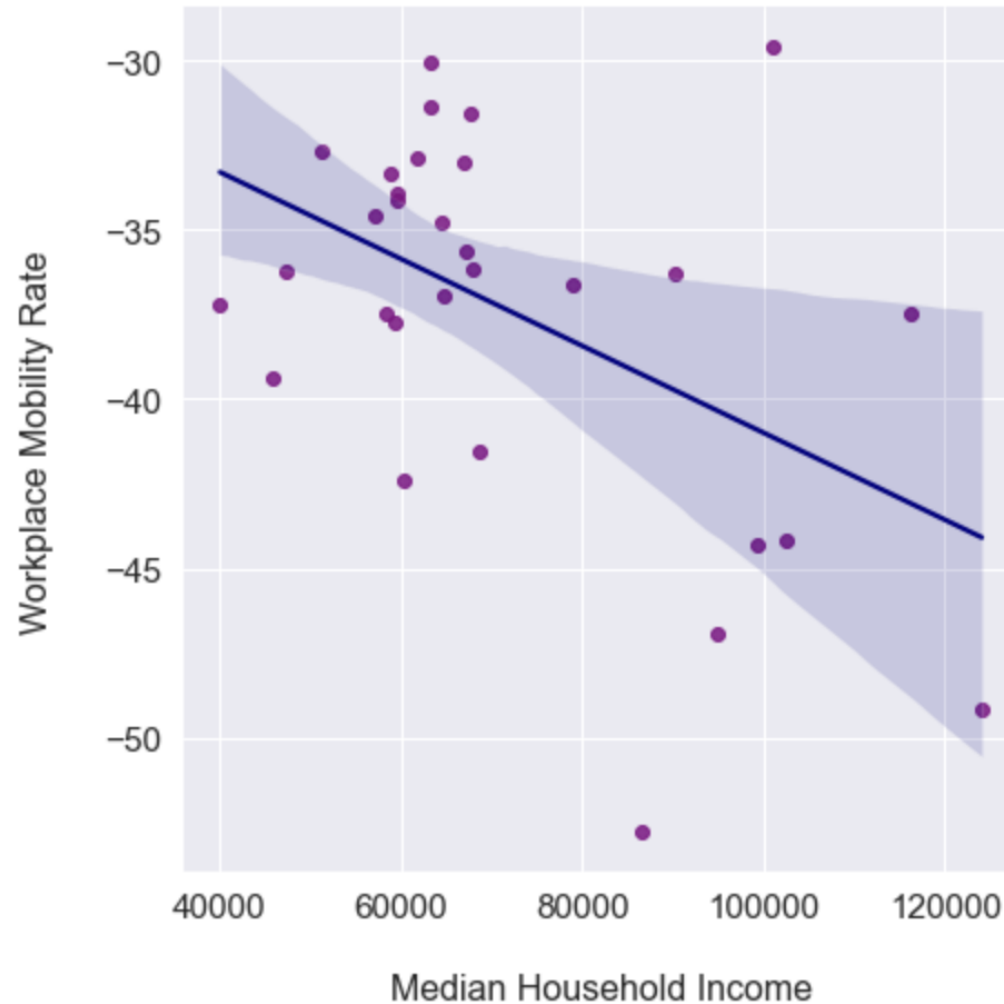
Horizontal mean
would give
monthly average
of all 30 counties

Vertical
mean
would give
yearly
average

2



Scatterplot: Workplace mobility rate Vs Median Household Income,
for all 30 counties

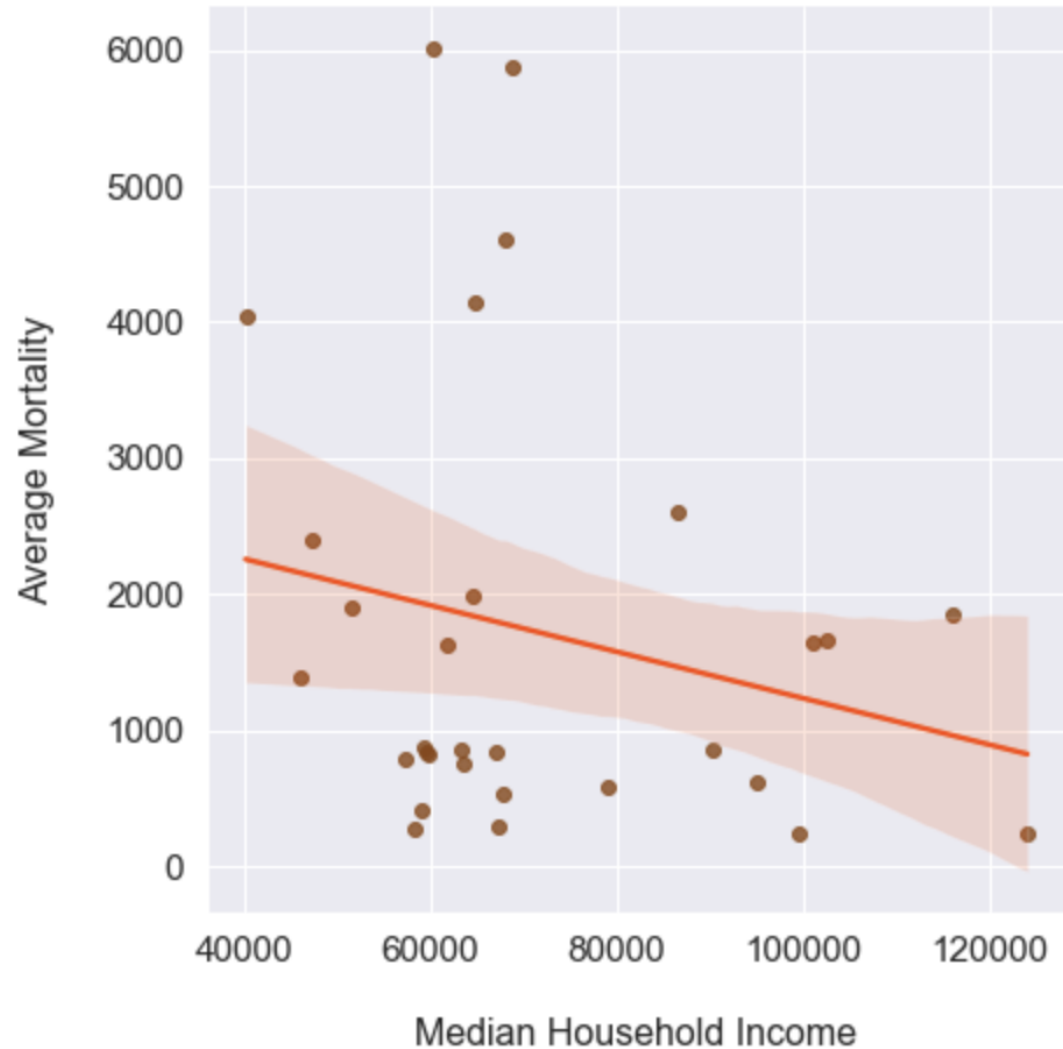


GRAPH 2

- Observe a negative correlation
- Suggests that the higher the income of a county was, the more people avoided coming to the office.
- While in counties with lesser income, the mobility at the workplace didn't decrease as much. Hence, people continued to go to their workplace and risked their lives.



Scatterplot: Median Household Income Vs Mortality, for all 30 counties

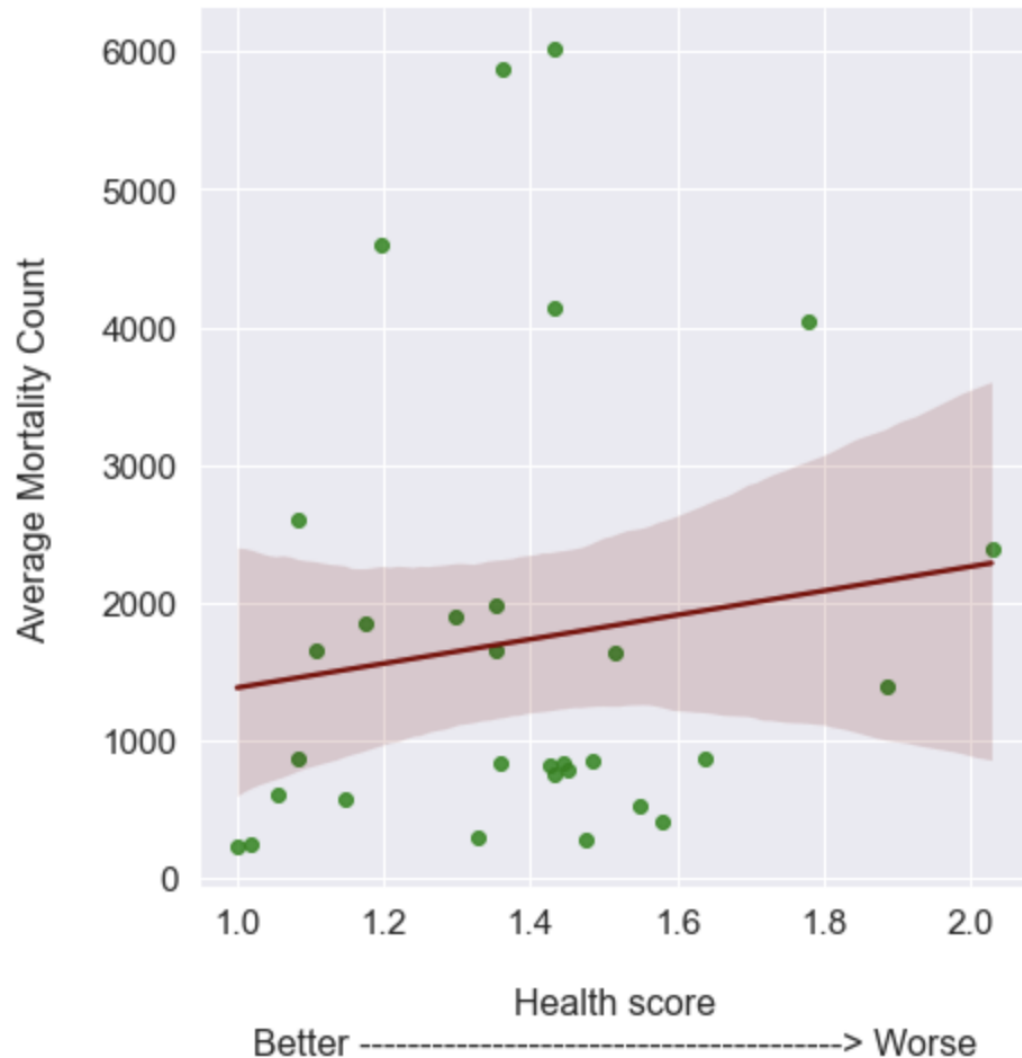


GRAPH 3

- Observe a weak negative correlation
- Suggests that the counties with higher income witnessed lesser deaths.
- While counties with lesser income witnessed more deaths.



Scatterplot: Health Score Vs Mortality Count, for all 30 counties



GRAPH 4

- Depicts a weak positive correlation
- Suggests that counties with worse health conditions experienced more COVID-19 deaths.
- While counties with better health conditions witnessed lesser deaths caused by COVID-19.



WHAT'S HEALTH SCORE?



	county	fips_code	percentage_of_adult_smokers	percentage_of_copd	percentage_of_diab	sum	index
0	Los Angeles County, California	6037	11	4.3	9.1	24.4	1.196078
1	Cook County, Illinois	17031	15	5.2	9.0	29.2	1.431373
2	Harris County, Texas	48201	15	5.5	10.4	30.9	1.514706
3	Maricopa County, Arizona	4013	14	5.6	8.0	27.6	1.352941
4	San Diego County, California	6073	11	4.6	7.8	23.4	1.147059
5	Orange County, California	6059	10	4.1	8.0	22.1	1.083333
6	Miami-Dade County, Florida	12086	14	5.4	7.1	26.5	1.299020
7	Kings County, New York	36047	14	5.8	9.4	29.2	1.431373

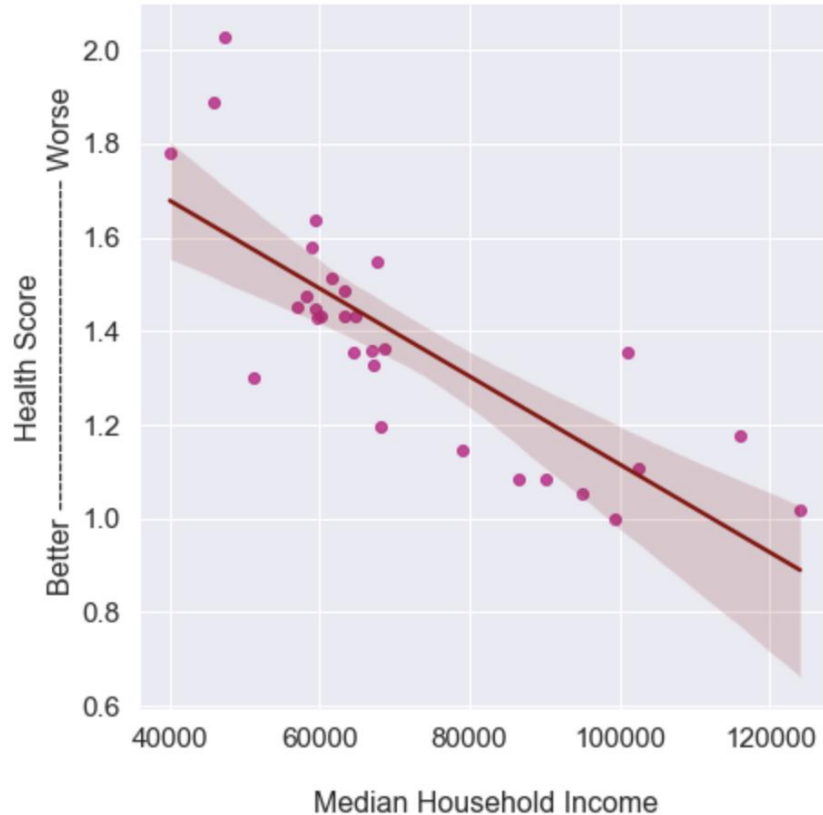
(This is a snippet of one of the dataframes)

Dividing
all sums
by the
smallest
sum value



GRAPHS 5 & 6 (Have same axes)

Scatterplot: Median Household Income Vs Health score, for all 30 counties



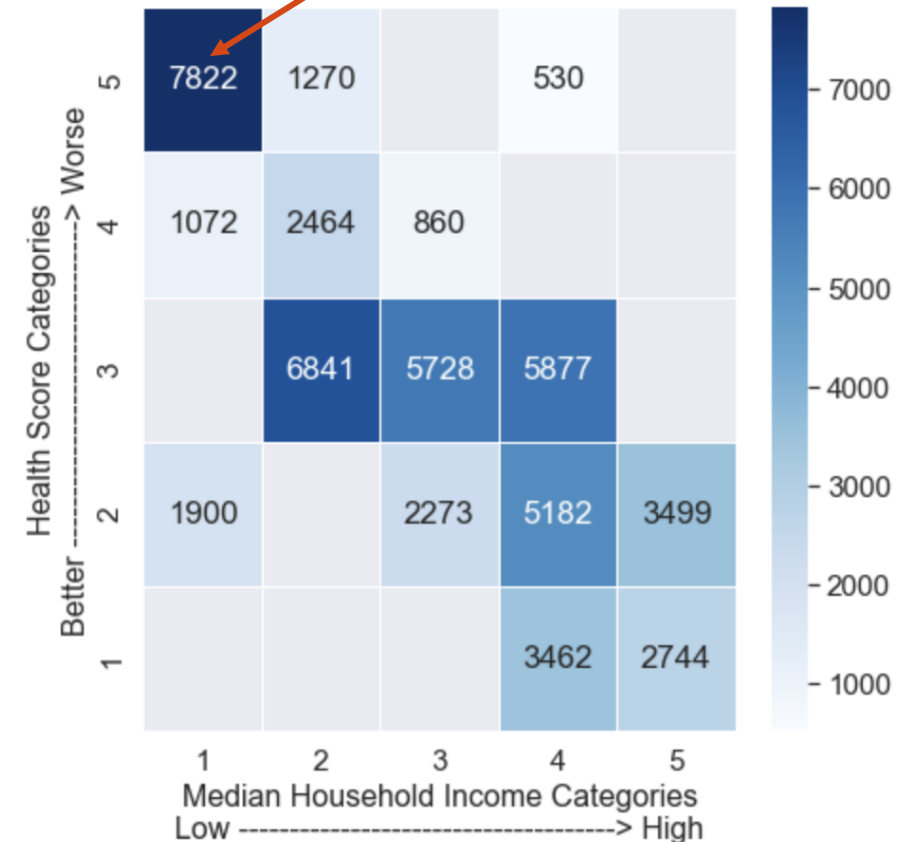
5

Has strong negative correlation; suggests that counties with higher income have better health, whereas counties with lower income have worse health conditions

6

Highest death count; low income and bad health conditions

Heatmap: Median Household Income Vs Health Score, and their impact on mortality count



Highest death count can be observed where there's low income and bad health conditions



MAKING THE HEATMAP

	fips_code	index	mortality	median_hh_income	health_score_cat	income_cat
0	6059	1.083333	863.457957	90234.0	1	4
1	36061	1.083333	2598.315806	86553.0	1	4
2	53033	1.053922	611.165054	94974.0	1	5
3	6085	1.019608	241.244731	124055.0	1	5
4	6001	1.000000	233.551398	99406.0	1	5
5	25017	1.107843	1658.239785	102603.0	1	5
6	12086	1.299020	1900.388280	51347.0	2	1
7	4013	1.352941	1985.038710	64468.0	2	3
8	6067	1.328431	288.284194	67151.0	2	3
9	6037	1.196078	4602.587419	68044.0	2	4
10	6073	1.147059	579.854516	78980.0	2	4
11	36103	1.352941	1646.290430	101031.0	2	5

```
table = pd.pivot_table(heatmap_df, index=['health_score_cat'], columns=['income_cat'], values='mortality', aggfunc=np.sum)
table
```

		income_cat				
		1	2	3	4	5
health_score_cat	1	NaN	NaN	NaN	3461.773763	2744.200968
	2	1900.388280	NaN	2273.322903	5182.441935	3499.019892
	3	NaN	6840.978280	5728.277634	5876.514409	NaN
	4	1072.142366	2464.287312	859.739355	NaN	NaN
	5	7821.686667	1269.842151	NaN	530.420430	NaN

Data in matrix format!

Columns created
based on percentile



CONCLUSION

After considering various factors, my visualisations suggest the poor were at a disadvantage and were more vulnerable to COVID-19.

People from worse socio-economic backgrounds; lesser income and worse health conditions were affected more.

The US government should construct policies that targeted towards them and help them recover from the pandemic.



**THANK YOU FOR
LISTENING!**

