Evaluating the relationship between a county's prevalence of COVID-19 to its median income

Kevin Nguyen

Created for USP4 WI23 taught by Soydan Alihan Polat

Is there a relationship between a county's median income and its prevalence of COVID-19 cases? And if so, are there any outstanding trends?

Data Sources

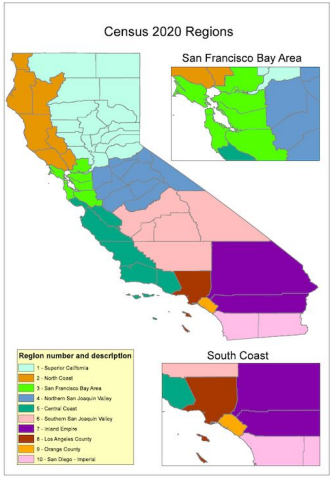
median income (csv) - Census data (ACS 2020 5-Year Estimates)
[S1901 INCOME IN THE PAST 12 MONTHS (IN 2020 INFLATION-ADJUSTED DOLLARS)]

population (csv) - Census data (ACS 2020 5-Year Estimates)
[B01003 TOTAL POPULATION]

COVID-19 cases (csv) - California Health and Human Services Open Data Portal (2020-02-01 to 2023-02-28)
[COVID-19 Time-Series Metrics by County and State]

California counties (shp) - California Open Data Portal (US Census Bureau's 2016 MAF/TIGER database)
[CA County Boundaries]

California regions (list & png) - CA Census 2020
[Regions]



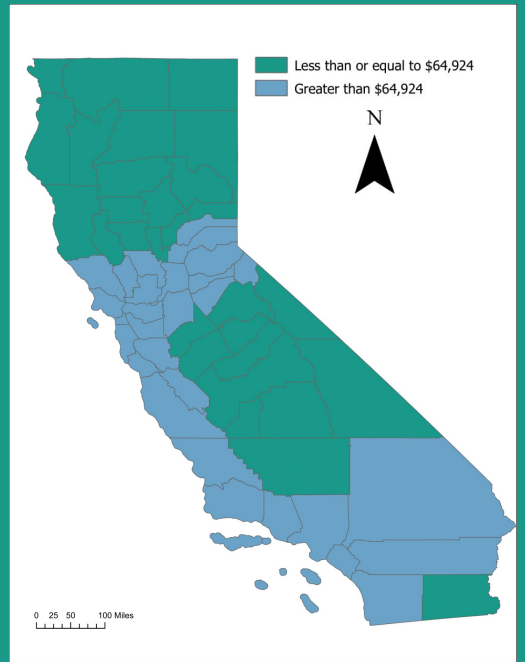
Data Cleaning/Organizing

variables of interest: **county names, total population, median income (dollars), cumulative reported cases, cases per county population per 100 persons, regions, latitude, longitude**

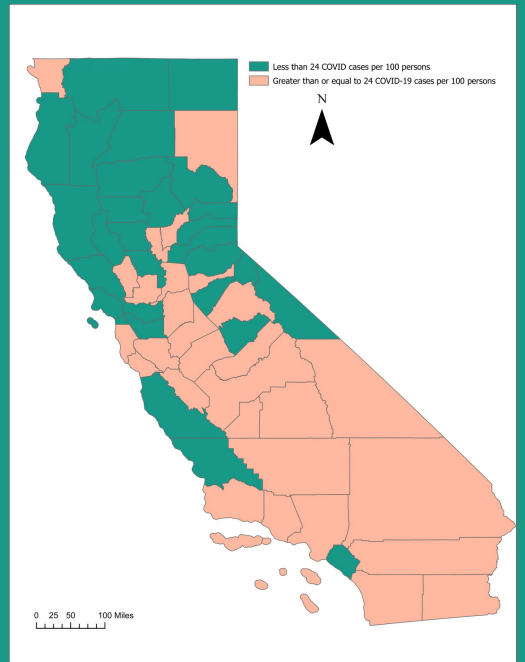
county	total_population	median_income_dollars	cumulative_reported_cases	cases_per_population_per_100_persons	regions	latitude	longitude
Alameda	1,685,886	$109,729.00	378,544	22.45	San Francisco Bay Area	37.6471385	-121.912488
Butte	217,769	$64,738.00	41,655	19.13	Superior California	39.6659588	-121.6019188
Contra Costa	1,160,099	$111,080.00	272,604	23.5	San Francisco Bay Area	37.919479	-121.9515431
El Dorado	193,098	$87,491.00	34,553	17.89	Superior California	38.785532	-120.5343981
Fresno	1,032,227	$63,656.00	294,108	28.49	Southern San Joaquin Valley	36.7610058	-119.6550193
Humboldt	134,098	$54,752.00	23,359	17.42	North Coast	40.7066731	-123.9258181
Imperial	191,649	$51,809.00	69,938	36.49	San Diego - Imperial	33.0408143	-115.3554001
Kern	927,251	$58,217.00	229,772	24.78	Southern San Joaquin Valley	35.3466288	-118.7295064
Kings	156,444	$62,155.00	62,825	40.16	Southern San Joaquin Valley	36.072478	-119.8155301
Lake	64,871	$61,221.00	13,639	21.02	North Coast	39.0948019	-122.7467569

Individual Variables Map

median income (dollars)



COVID-19 cases per county population per 100 persons



Evaluation

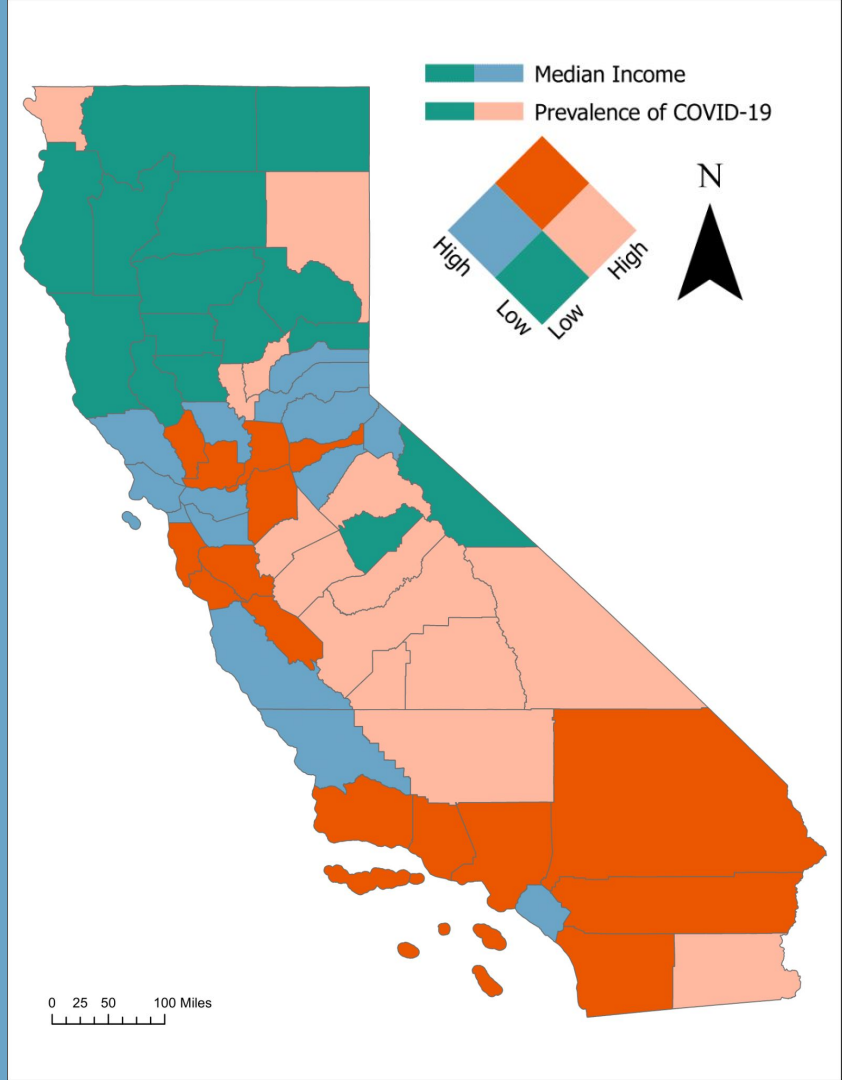
Colors and their meanings:

Light Blue (**high** median income, **low** prevalence of COVID-19 cases)

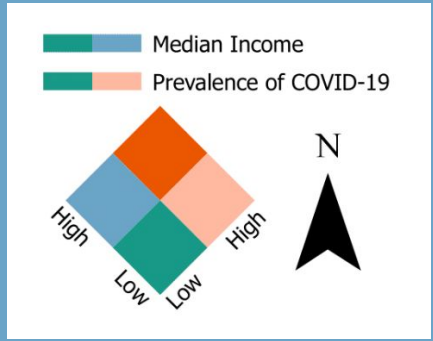
Green (**low** median income, **low** prevalence of COVID-19 cases)

Orange (**high** median income, **high** prevalence of COVID-19 cases)

Light Orange (**low** median income, **high** prevalence of COVID-19 cases)



Analyzing Grouped Regions

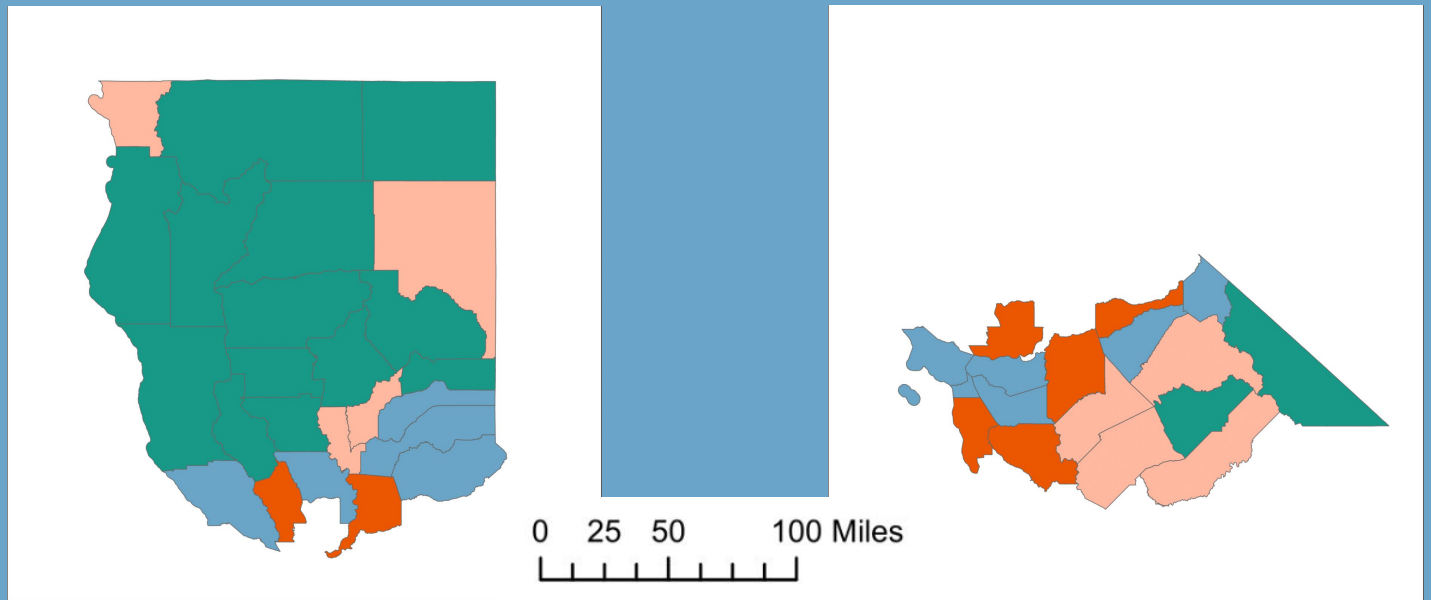


Superior California and North Coast

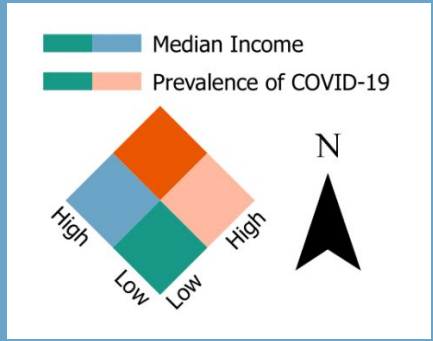
low income | low prevalence

San Francisco Bay Area and Northern San Joaquin Valley

mix of everything



Analyzing Grouped Regions

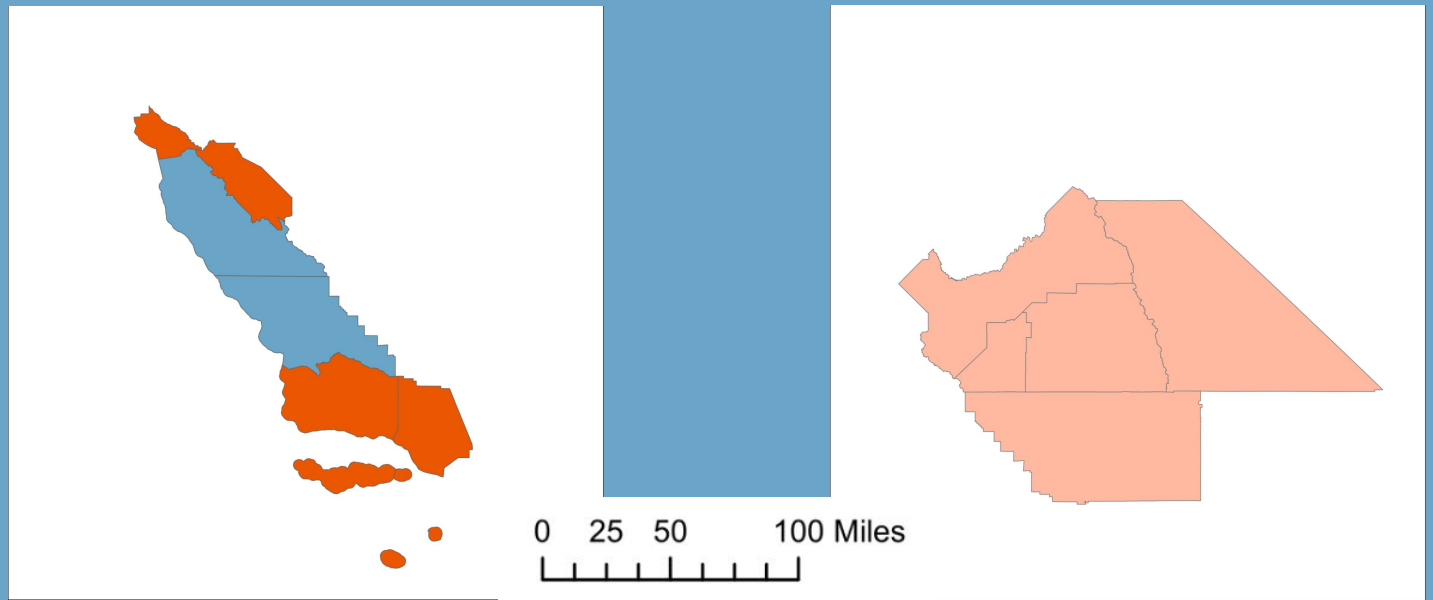


Central Coast

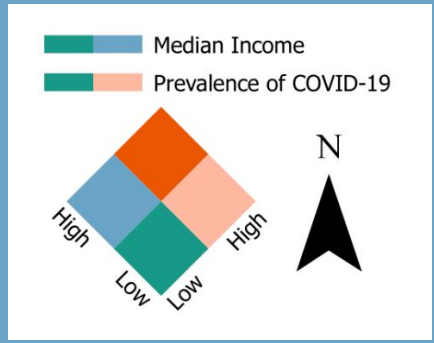
high income | mix high & low prevalence

Southern San Joaquin Valley

low income | high prevalence

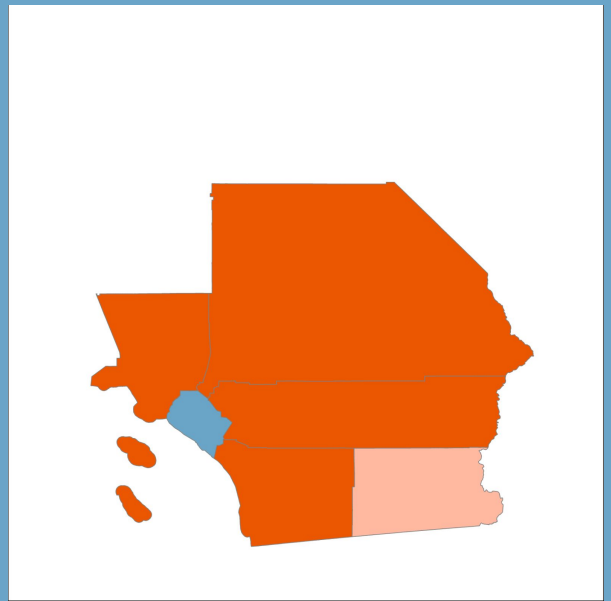


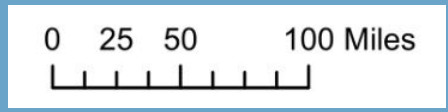
Analyzing Grouped Regions

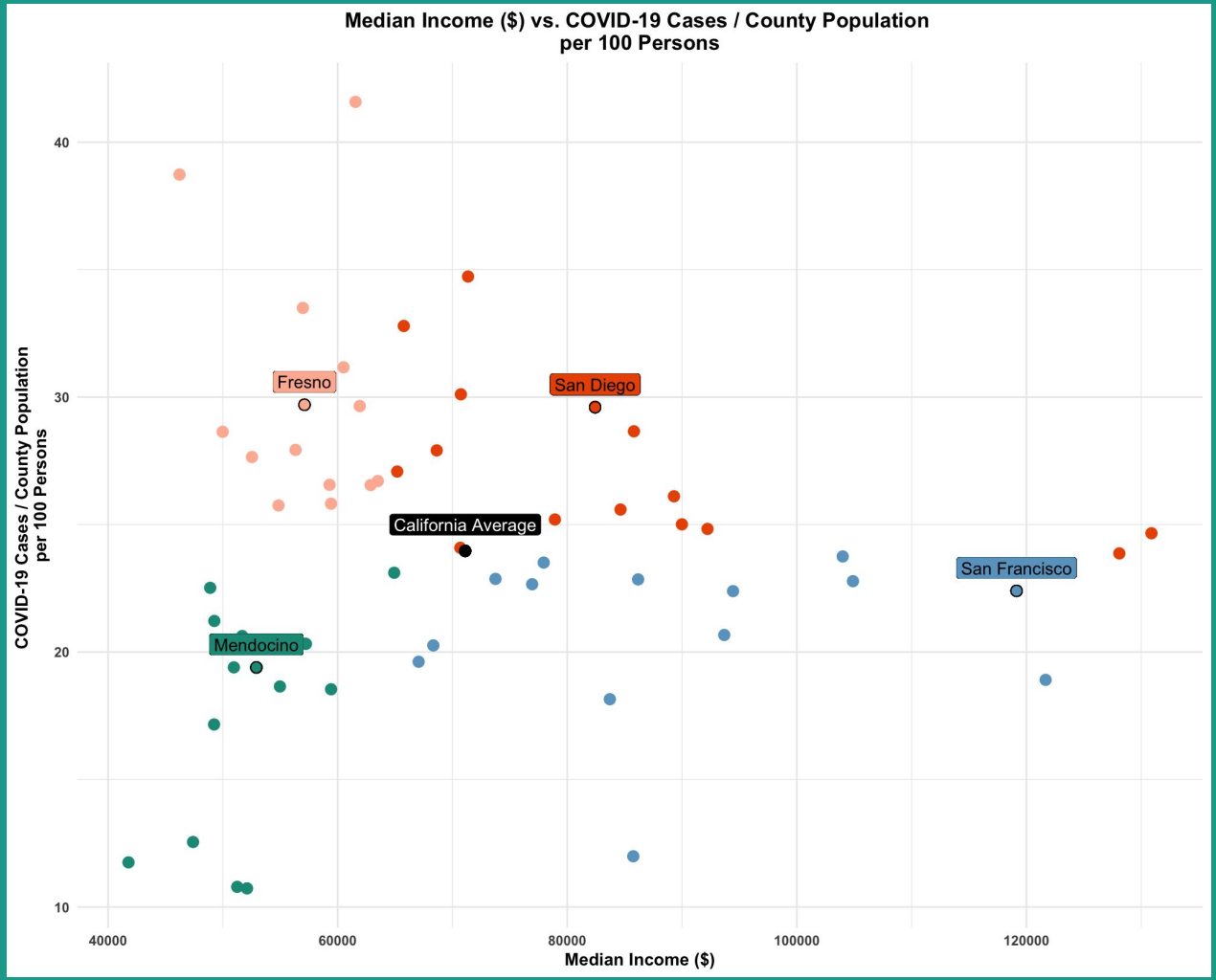


Los Angeles County, Inland Empire, Orange County, San Diego - Imperial

high income | high prevalence






Median Income ($) vs. COVID-19 Cases / County Population per 100 Persons
COVID-19 Cases / County Population per 100 Persons
Median Income ($)
40
30
20
10
40000
60000
80000
100000
120000
Fresno
San Diego
California Average
San Francisco
Mendocino

There is no clear relationship between a county's median income and its prevalence of COVID-19 cases however, they are spatially correlated to the specific region that they are in.

So if a county falls under the Southern San Joaquin Valley region, they are quite likely to have low income and a low prevalence of COVID-19 cases.

If I had more time, I would take a closer look at the counties and dive deeper into the specific ZIP codes per counties

One limitation that I faced when doing the analysis was when I was collecting the data where the US Census data does not report data for counties with populations less than 65,000 people for 1 year aggregates so I had to grab 5 year aggregates.