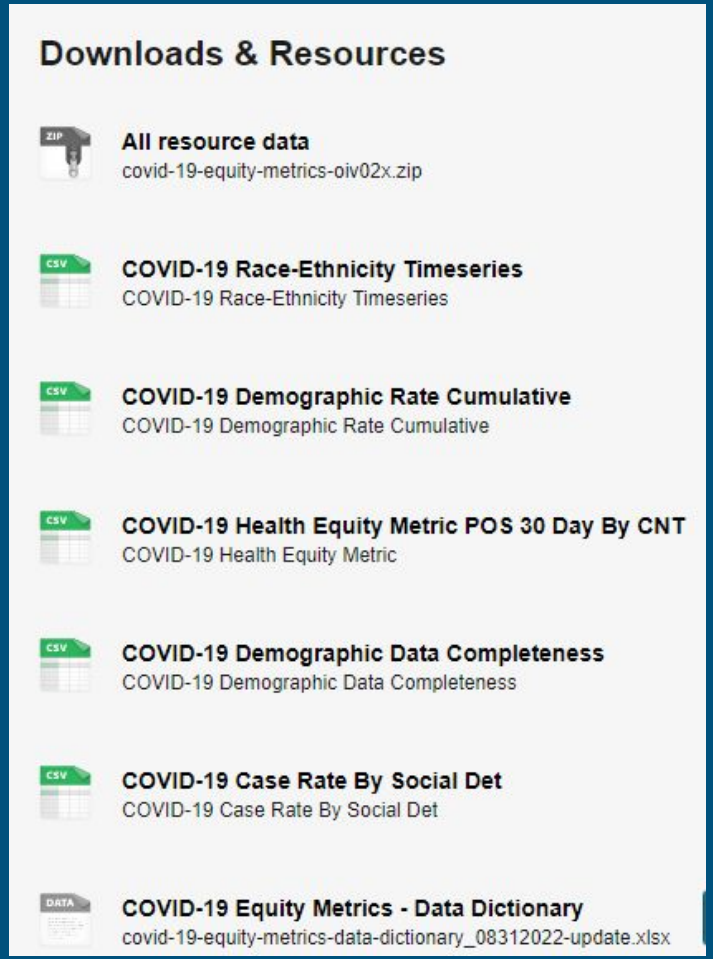
An analysis of the impact of social determinants of health on Covid-19 case rates in California

Kelly Le
Justine Pile
Christina Dang

Dataset: COVID-19 Equity Metrics

Data was obtained from:
https://catalog.data.gov/dataset/covid-19-equity-metrics-b21df

These datasets seemed interesting as Covid-19 is something that impacts so many people and so many aspects of life and the question of what might drive case rates higher is an important one



The Covid-19 Equity Metrics includes different datasets for this state of California program with data typically going from 2020 to 2021

https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/COVID-19/CaliforniaHealthEquityMetric.aspx

Do social determinants affect Covid-19 case rates?

Is there a difference between various groups of people with differing levels of health insurance, housing crowding, and income?

If there are differences, are they consistent or is there any variance?

What are social determinants of health?

HHS (US Health and Human Services) defines social determinants as “the conditions in the environments where people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks.”

HHS goes on to say that social determinants “have a major impact on people’s health, well-being, and quality of life.”

Data exploration and engineering

Setup dependencies and create a dataframe from CSV

```
# Dependencies and Setup
import matplotlib.pyplot as plt
import pandas as pd
import scipy.stats as st

# Study data files
case_rate_path = "Resources/covid19case_rate_by_social_det.csv"

# Read the mouse data and the study results
case_rate = pd.read_csv(case_rate_path)
```

```
# Display a preview from the social determinants of health
case_rate.head()
```

	date	social_det	social_tier	sort	cases_7dayavg_7daysago	population	case_rate_per_100k	state_case_rate_per_100k	case_rate_per_100k_30_days_ago	rate_diff_30_days
0	2023-01-01	income_cumulative	$60k$–80k	NaN	2.247823e+06	8630488	26045.143681	26997.831502	25521.326256	523.817425
1	2023-01-01	income_cumulative	$80k$–100k	NaN	1.278122e+06	5567034	22958.760446	26997.831502	22458.260539	500.499907
2	2023-01-01	income_cumulative	above $120K	NaN	4.710180e+05	2744855	17160.032133	26997.831502	16764.164227	395.867906
3	2023-01-01	income_cumulative	below $40K	NaN	2.252980e+06	7290434	30903.235665	26997.831502	30442.714384	460.521280
4	2023-01-01	income	above $120K	0.0	3.414286e+02	2744855	12.438856	15.048195	11.418766	1.020090

Drop unnecessary columns from dataframe and identify what social determinants are in the dataset

```
# Drop unncessary columns from dataset
case_rate = case_rate.drop(columns=["sort", "cases_7dayavg_7daysago",
                                    "population",
                                    "state_case_rate_per_100k",
                                    "case_rate_per_100k_30_days_ago",
                                    "rate_diff_30_days"])
case_rate.head()
```

	date	social_det	social_tier	case_rate_per_100k
0	2023-01-01	income_cumulative	$60k—80k	26045.143681
1	2023-01-01	income_cumulative	$80k—100k	22958.760446
2	2023-01-01	income_cumulative	above $120K	17160.032133
3	2023-01-01	income_cumulative	below $40K	30903.235665
4	2023-01-01	income	above $120K	12.438856

```
# Determine all the social determinants in the dataset
case_rate.social_det.unique()
```

```
array(['income_cumulative', 'income', 'insurance', 'crowding'],
      dtype=object)
```

```
# Group case_rate by social determinants and view the data
grouped = [group for _, group in case_rate.groupby("social_det")]
grouped
```

```
[           date social_det   social_tier  case_rate_per_100k
 12    2022-11-27   crowding     10% - 15%           13.108560
 13    2022-11-27   crowding     15% - 20%           13.068887
 14    2022-11-27   crowding       2% - 5%           11.868938
 15    2022-11-27   crowding      5% - 10%           12.712180
 16    2022-11-27   crowding  less than 2%           10.786352
 ...          ...        ...           ...                 ...
 8017  2023-01-01   crowding     15% - 20%           15.857626
 8018  2023-01-01   crowding       2% - 5%           15.158523
 8019  2023-01-01   crowding      5% - 10%           16.057027
 8020  2023-01-01   crowding  less than 2%           14.068706
 8021  2023-01-01   crowding      over 20%           14.691394

 [2082 rows x 4 columns],
           date social_det    social_tier  case_rate_per_100k
 4     2023-01-01     income    above $120K           12.438856
 5     2023-01-01     income     below $40K           13.516734
 18    2022-11-27     income  $100k - $120k           11.712117
 19    2022-11-27     income    $40k - $60k           12.377717
 20    2022-11-27     income    $60k - $80k           12.649508
 ...          ...        ...            ...                 ...
 8009  2022-11-27     income     below $40K           11.529786
 8022  2023-01-01     income  $100k - $120k           15.721080
 8023  2023-01-01     income    $40k - $60k           15.534227
 8024  2023-01-01     income    $60k - $80k           16.613859
 8025  2023-01-01     income   $80k - $100k           16.238449

 [2082 rows x 4 columns],
           date          social_det    social_tier  case_rate_per_100k
 0     2023-01-01  income_cumulative    $60k - $80k        26045.143681
 1     2023-01-01  income_cumulative   $80k - $100k        22958.760446
 2     2023-01-01  income_cumulative    above $120K        17160.032133
 3     2023-01-01  income_cumulative     below $40K        30903.235665
 22    2022-11-27  income_cumulative  $100k - $120k        20059.993671
 ...          ...                ...            ...                 ...
 8005  2022-11-27  income_cumulative   $80k - $100k        22252.765117
 8006  2022-11-27  income_cumulative    above $120K        16606.851728
 8007  2022-11-27  income_cumulative     below $40K        30236.773284
 8026  2023-01-01  income_cumulative  $100k - $120k        20723.675956
 8027  2023-01-01  income_cumulative    $40k - $60k        28336.804918

 [1782 rows x 4 columns],
           date social_det   social_tier  case_rate_per_100k
 6     2023-01-01  insurance     10% - 15%           15.966100
 7     2023-01-01  insurance     15% - 25%           15.937163
```

Group data by social determinants

```
# Put data into four datasets for each social determinant
crowding = grouped[0]
income = grouped[1]
income_cumulative_df = grouped[2]
insurance = grouped[3]
```

```
# View a sample from the crowding dataset
crowding.head(6)
```

	date	social_det	social_tier	case_rate_per_100k
12	2022-11-27	crowding	10% - 15%	13.108560
13	2022-11-27	crowding	15% - 20%	13.068887
14	2022-11-27	crowding	2% - 5%	11.868938
15	2022-11-27	crowding	5% - 10%	12.712180
16	2022-11-27	crowding	less than 2%	10.786352
17	2022-11-27	crowding	over 20%	11.231895

```
# View a sample from the income dataset
income.head(6)
```

	date	social_det	social_tier	case_rate_per_100k
4	2023-01-01	income	above $120K	12.438856
5	2023-01-01	income	below $40K	13.516734
18	2022-11-27	income	$100k$–120k	11.712117
19	2022-11-27	income	$40k$–60k	12.377717
20	2022-11-27	income	$60k$–80k	12.649508
21	2022-11-27	income	$80k$–100k	11.988944

```
# View a sample from the insurance dataset
insurance.head(6)
```

	date	social_det	social_tier	case_rate_per_100k
6	2023-01-01	insurance	10% - 15%	15.966100
7	2023-01-01	insurance	15% - 25%	15.037163

Create four new dataframes for each group of social determinants

View tiers in the insurance dataset and verify each tier has the same amount of data

In [10]:
```
# Find all insured tiers in insurance data_set
insurance["social_tier"].unique()
```

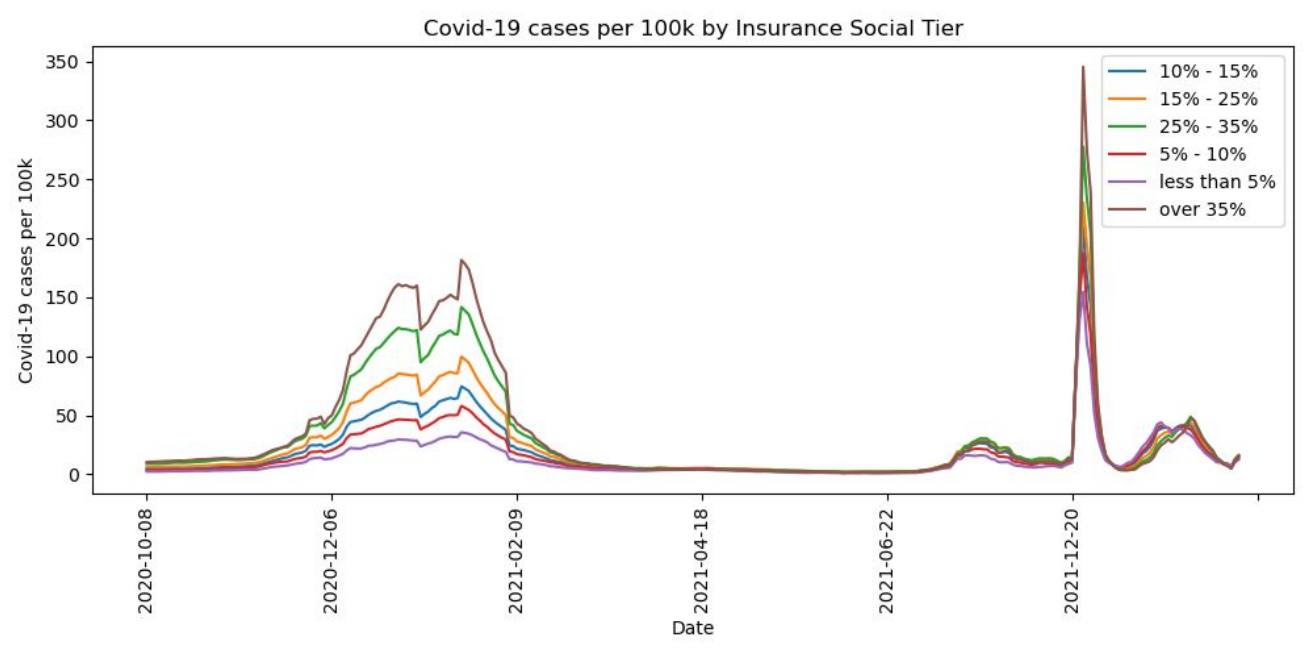
Out[10]:
```
array(['10% - 15%', '15% - 25%', '25% - 35%', '5% - 10%', 'less than 5%',
       'over 35%'], dtype=object)
```

In [11]:
```
# Verify that each insured tier has the same amount of data
counts = insurance["social_tier"].value_counts()
print(counts)
```

```
10% - 15%       347
15% - 25%       347
25% - 35%       347
5% - 10%        347
less than 5%    347
over 35%        347
Name: social_tier, dtype: int64
```

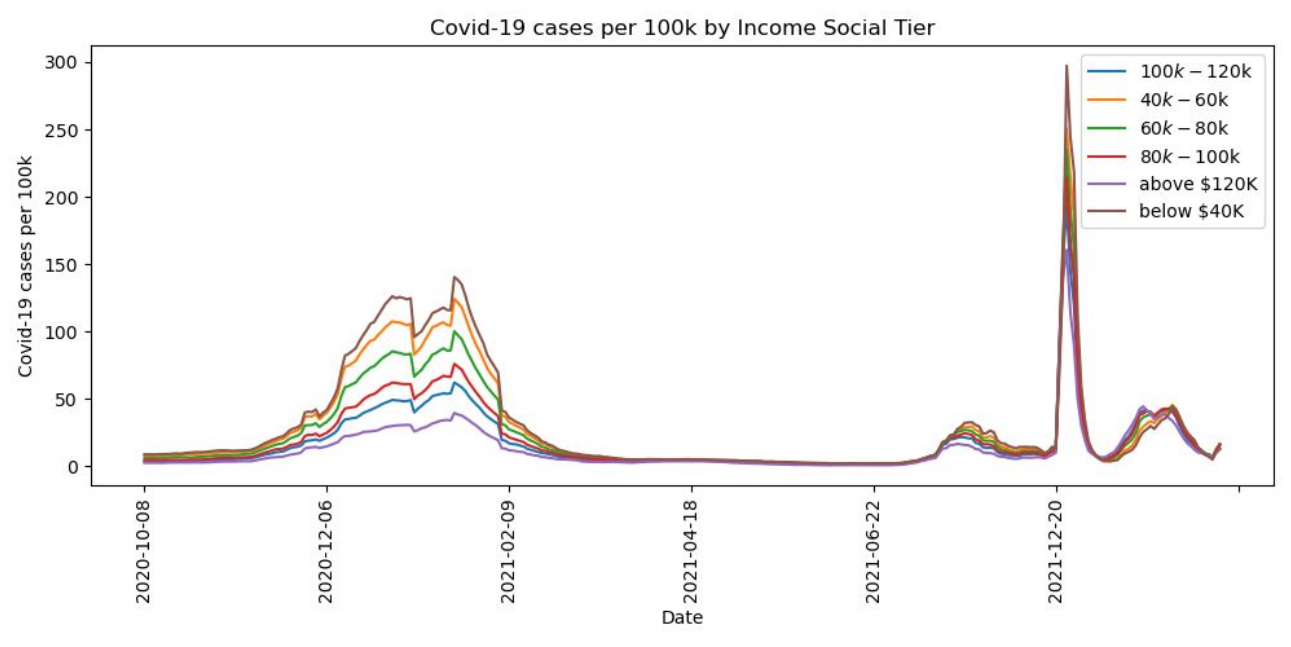
Visualizations

Case rate by insurance social tier



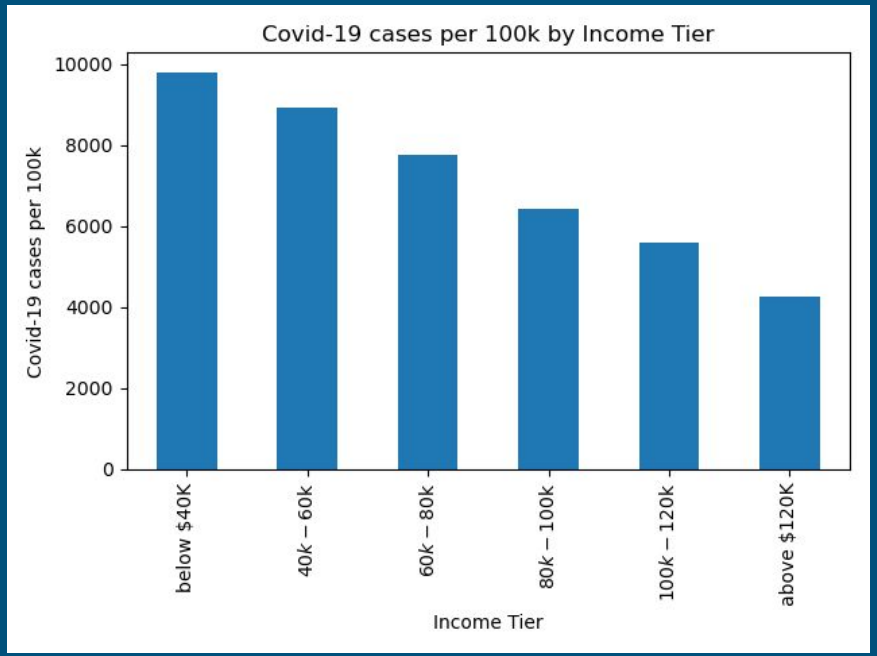
Starting at the over 35% tier and working towards the less than 5% tier, cases decrease consistently from one group to the next

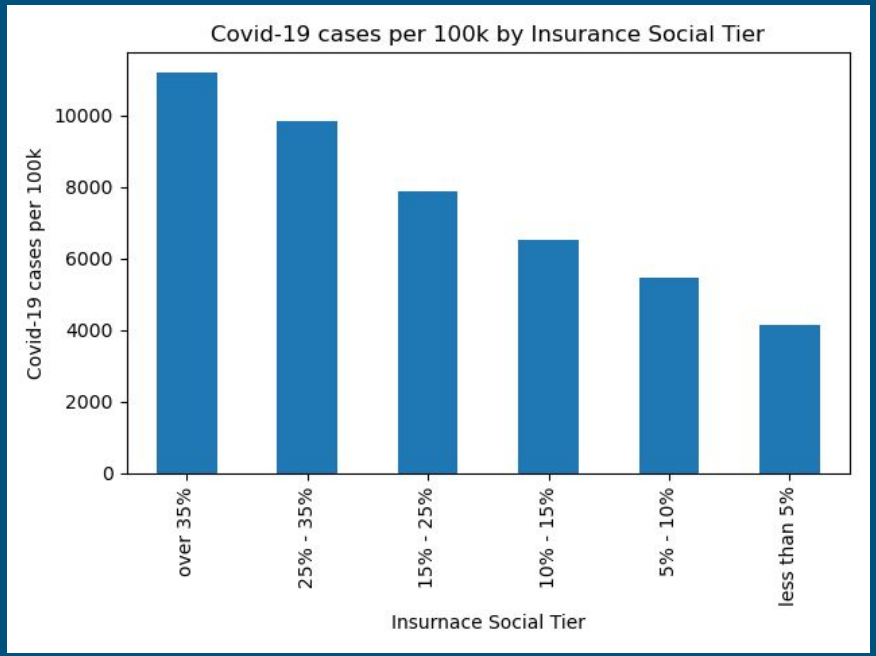
Case rate by income social tier



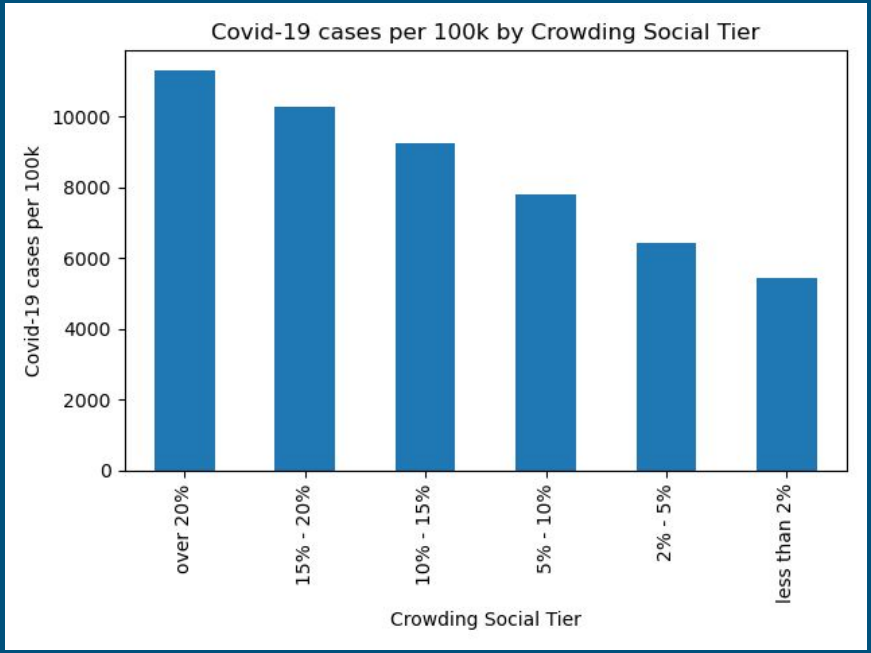
Starting at the below $40K tier and working towards the above tier $120K tier, cases decrease consistently from one group to the next

Case rates by income / insurance tiers



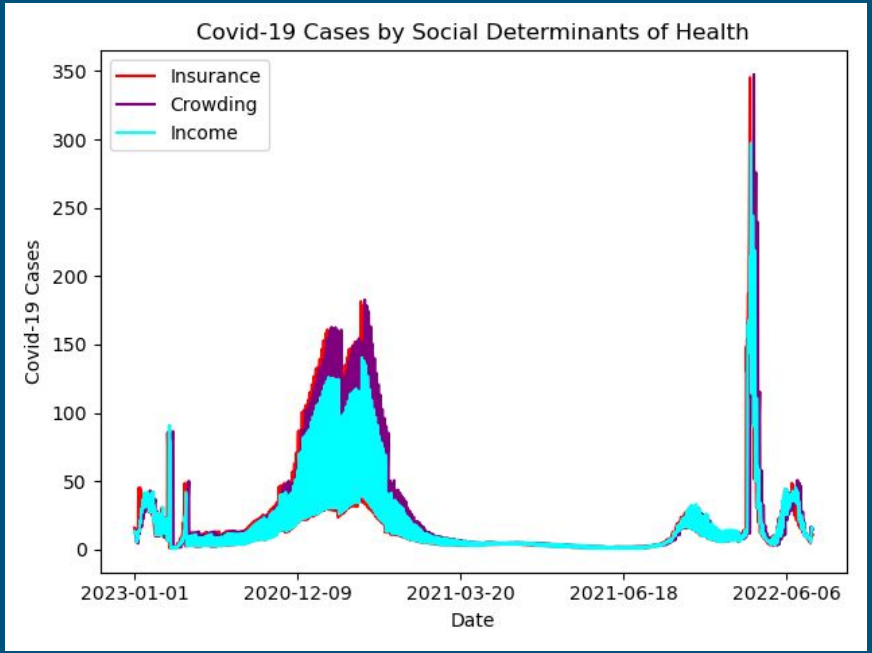


Case rate by housing crowding tiers



For the income, insurance, and crowding social determinants, cases decrease as the tiers increase from least desirable to most desirable

Case rate by social determinants of health



Analysis

Final Data Analysis

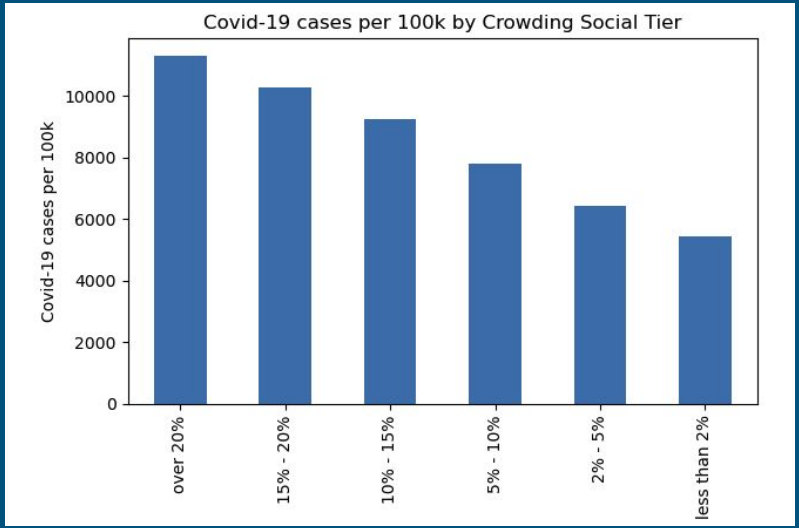
	date	social_det	social_tier	case_rate_per_100k
0	2023-01-01	income_cumulative	$60k$—80k	26045.143681
1	2023-01-01	income_cumulative	$80k$—100k	22958.760446
2	2023-01-01	income_cumulative	above $120K	17160.032133
3	2023-01-01	income_cumulative	below $40K	30903.235665
4	2023-01-01	income	above $120K	12.438856

First looking at a preview from the social determinates of health between all populations, we are analyzing all the social determinants, tiers, population, and case rates between the state, population, and between 7 days ago to 30 days to see how much of an impact it is on Covid 19 infections. Measuring the case rate by calculating the average daily number of COVID-19+ cases (excluding certain cases) over 7 days, divided by the number of people living in the county/region/state.The adjusted case rate is calculated as the case rate multiplied by a case rate adjustment factor that is based on the difference between the county testing volume (testing volume, tests per 100,000 per day) and the median testing volume calculated across all counties.

The first analysis that was done was by grouping the dataset by social determinates, social tier, and what the case rate was. In this dataset, it was shown that those with an income of less than 40k had the highest case rate per 100k at 30,903 per 100k people. As the social tier increases, we see that it drops from 26,045 with an income at 60k-80k to a case rate of 17,160 with those making above 120K. This allows us to conclude that those with lower income are more exposed to Covid-19 infection which may be due to an increase in careers that force them to come in contact with multiple people in their days which could include jobs that require customer service or alot of crowding in condensed areas.

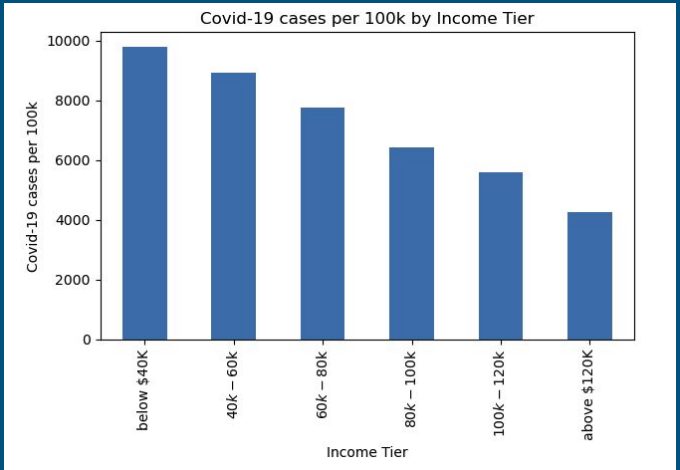
When split up into multiple tiers of social determinants, including, crowding, income, cumulative income, and insurance, we are able to more thoroughly review the difference in case rates between the social tiers.

CROWDING



When the social determinant is due to crowding, which essentially is due to a dense number of occupants in a space, it seems like the case rate per 100k does not necessarily make a difference when it comes to social tier percentages from as little as 0% up to over 20% with the case rate still ranging from 10.7 to 16.6 regardless of the tier of income However in the bar graph depicting the Covid 19 cases per 100k by Crowding Social Tier, when crowding social tier is above 20% there is a high number of cases per 100k at about 15,000 cases when there is a high percentage of crowding. As the social tier percentages decrease, we also see the bar graph slowly decrease linearly as well at about 3,000 less Covid Cases per group of social tier percentages.

INCOME



When the social determinant is due to income, we can definitely see a definitive difference in case rates per 100k as mentioned above. In November of 2022, numbers were not as high ranging from the upper class at about a 11.7 case rate while the lower class making below 40k about 13.5 case rate per 100k people. However, in January 2023 of this year, case rates increased by about 3-4% for about all groups and increased more in the higher social tier. This may be due to wealthier individuals being able to spend more time with families or going to social gatherings which may lead to an increase of Covid 19 infections. The upper tier making more than 60k had a case rate of atleast 16.2 to 16.6 case rate per 100k. In the line graph depicting Covid 19 cases per 100k by Income Social Tier, there was a peak at both times during December of 2020 to December of 2021 within incomes less than 40k. This could be due to this population once again having to work more in public spaces with a lot of foot traffic and chances of infection.

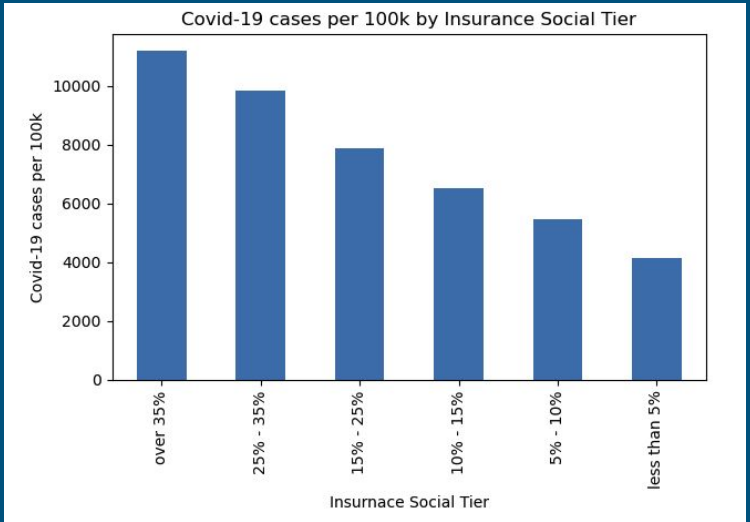
CUMULATIVE INCOME

[2082 rows x 4 columns],

	date	social_det	social_tier	case_rate_per_100k
0	2023-01-01	income_cumulative	$60k - $80k	26045.143681
1	2023-01-01	income_cumulative	$80k - $100k	22958.760446
2	2023-01-01	income_cumulative	above $120K	17160.032133
3	2023-01-01	income_cumulative	below $40K	30903.235665
22	2022-11-27	income_cumulative	$100k - $120k	20059.993671
...	...	...	...	...
8005	2022-11-27	income_cumulative	$80k - $100k	22252.765117
8006	2022-11-27	income_cumulative	above $120K	16606.851728
8007	2022-11-27	income_cumulative	below $40K	30236.773284
8026	2023-01-01	income_cumulative	$100k - $120k	20723.675956
8027	2023-01-01	income_cumulative	$40k - $60k	28336.804918

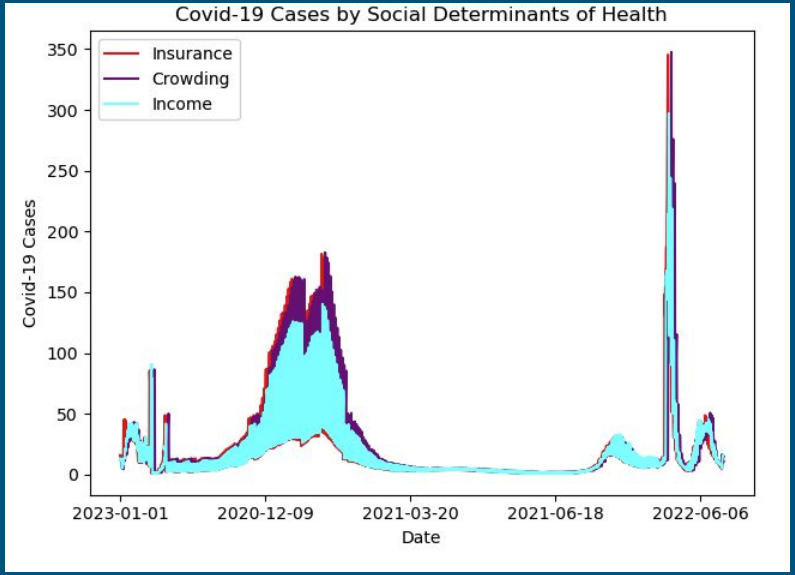
When the social determiniant is between a cumulative income, the case rates are basically doubled, mirroring single income populations.These rates starts at as little as 17,160 for dual income households making more than 120k a year to as high as 30,903 rate for households making less than 40k a year. This could be due to once again, working more in public spaces than wealthier indivoduals and having a higher chance of exposure to Covid. Once they come in contact with the virus, lower income households could have multiple family members and not have the luxury to self quarantine like others.

INSURANCE



When the social determinant is due to insurance, it seems like the higher the social tier, the higher the case rate per 100k is. With the peaks being in December-January 2020 at about 160k to almost 200k for those with a social tier over 35%. The social tier can be due to those with less access to insurance or unable to utilize their insurance benefits.

ANALYSIS CONCLUSION



In conclusion, with the three social determinants of health being insurance, crowding, and income we see the line graph follow the same trends between all three determinants. For example, insurance and crowding being very similar through December of 2020 to March of 2021 exponentially increasing from 50 cases to a little under 200 cases to back down to around 50 cases around February 2021. Income did not play as big of a factor as insurance and crowding did with the numbers being about 50-60 cases less than the two while following the same trends.

IMPLICATIONS

Major Findings Summary

The major findings discovered through out this data set was a repeating pattern. Through the creation of dataframes and graphs made, we were able to analyze what the major social determinants of health were on Covid 19 Case Rates. The top one was primarily due to crowding in households which we saw repeating trends on higher case rates when there is higher percentages of crowding involved at over 20% of a crowding social tier with higher than 15,000 case rates per 100k. We also see that the peak of Covid 19 Cases fluctuate between low income populations and high income populations and this could be due to multiple reasons. For the higher income population, cases could have peaked within December to January for all three continuous years due to large social gatherings and events for the holidays. For lower income populations, cases could have peaked due to household crowdings during the holidays and to be able to afford cheaper rent. Anotjer factor could include working jpbs in high density public places where Covid infection chances are much higher. If one person in a large household gets infected, it is much more difficult to self quarantine as well due to limited personal space in lower income social tiers.

In conclusion, we can analyze and see that these three social determinants play a huge factor within lower social tiers and its relation with chances or likelihood of Covid 19 infection rates.

THANK YOU FOR LISTENING

QUESTIONS?