COVID-19 Vaccination Progress

Measured by County in California from 2020 to 2021

Contents

Introduction

Proposal
Hypothesis

Process

High level topic review
Data selection
Narrow and identify topics

Data Review:
CSV Data pull from various locations
Clean data, such as duplicate counties
Identify anomalies in the demographic data
Develop data frames and appropriate graphs to answer questions

What is the correlation between COVID cases and COVID related deaths?

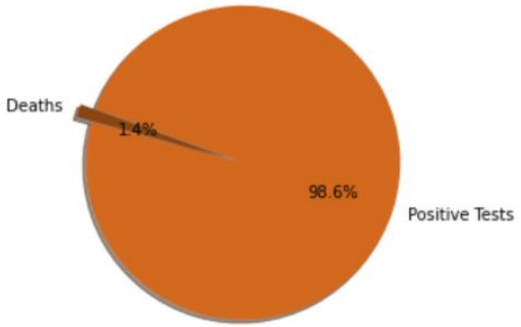
In [267]:
```
total_death = clean_cases['deaths'].sum()
print(total_death)
```

```
120254.0
```

In [282]:
```
pos_tests = clean_cases['positive_tests'].sum()
print(pos_tests)
```

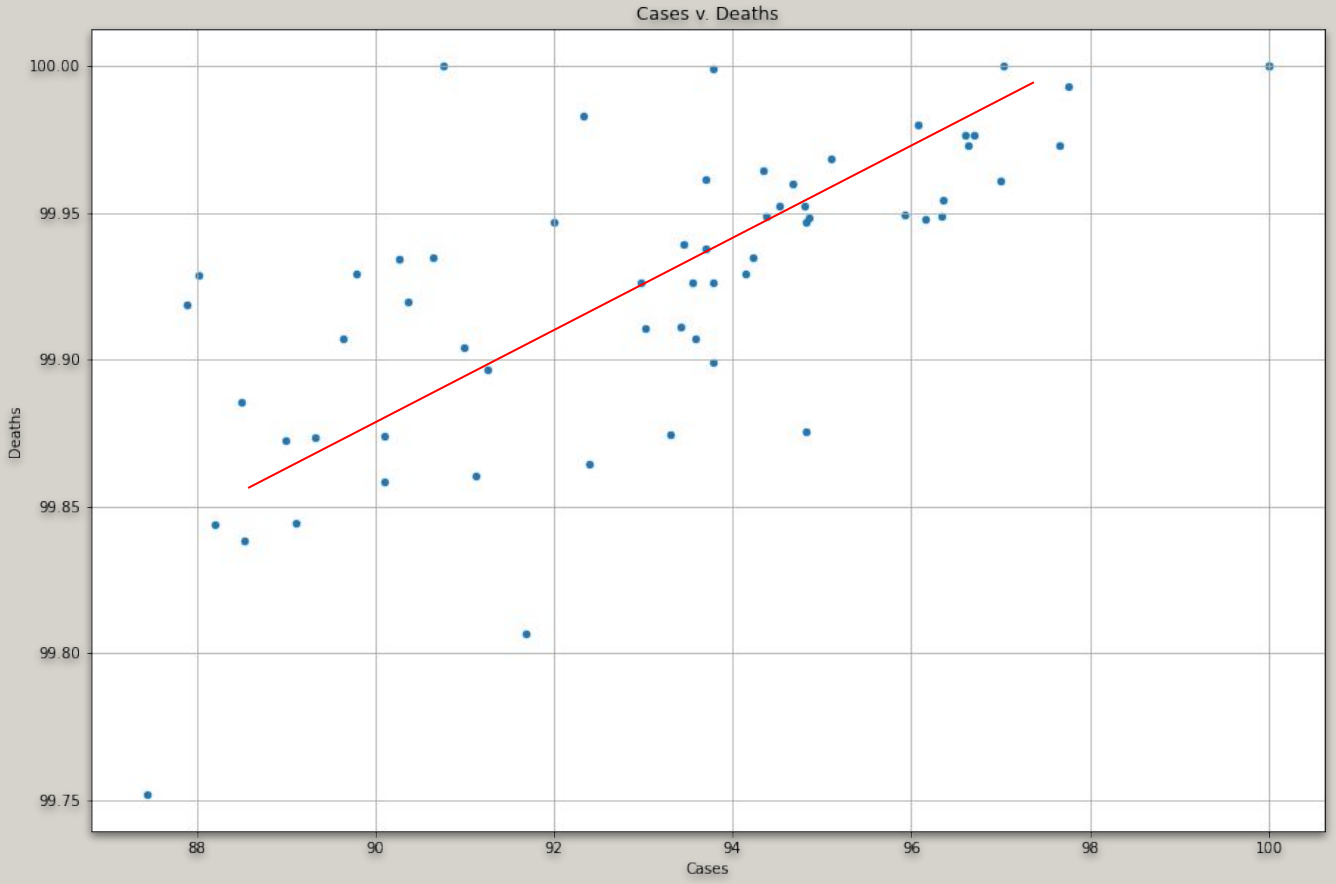
```
8718498.0
```

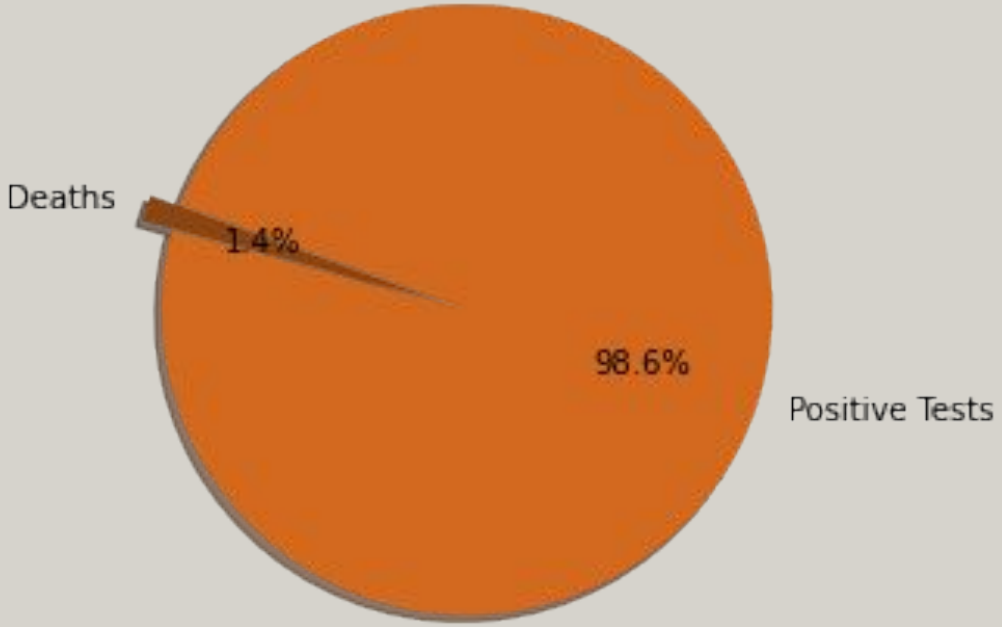
In [283]:
```
labels = ["Deaths", "Positive Tests"]
values = ["120254", "8718498"]
colors = ["saddlebrown", "chocolate"]
explode = (0.1, 0)

plt.pie(values, explode=explode, labels=labels, colors=colors, autopct="%1.1f%%", shadow=True,startangle=160)
plt.axis("equal")

plt.savefig("circle.png")
plt.show()
```



What is the correlation between COVID cases and COVID related deaths?

- Process for creating chart
- Positive correlation between cases and deaths
-





How does vaccination rate affect the case / death rate?

In [15]:
```
vaccine_data = vaccine_data.rename(columns={"administered_date": "date"})

merge_data = pd.merge(cases_data, vaccine_data,
                      on='date',
                      how='left')

# displaying result
merge_data
```

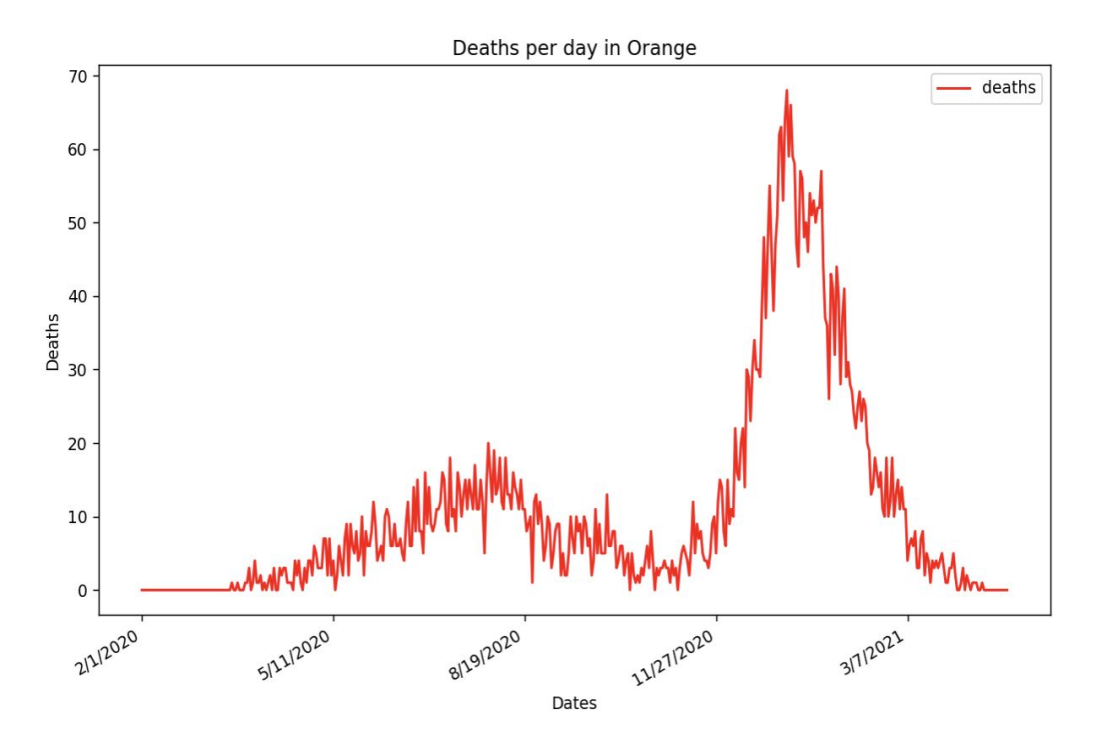
Out[15]:

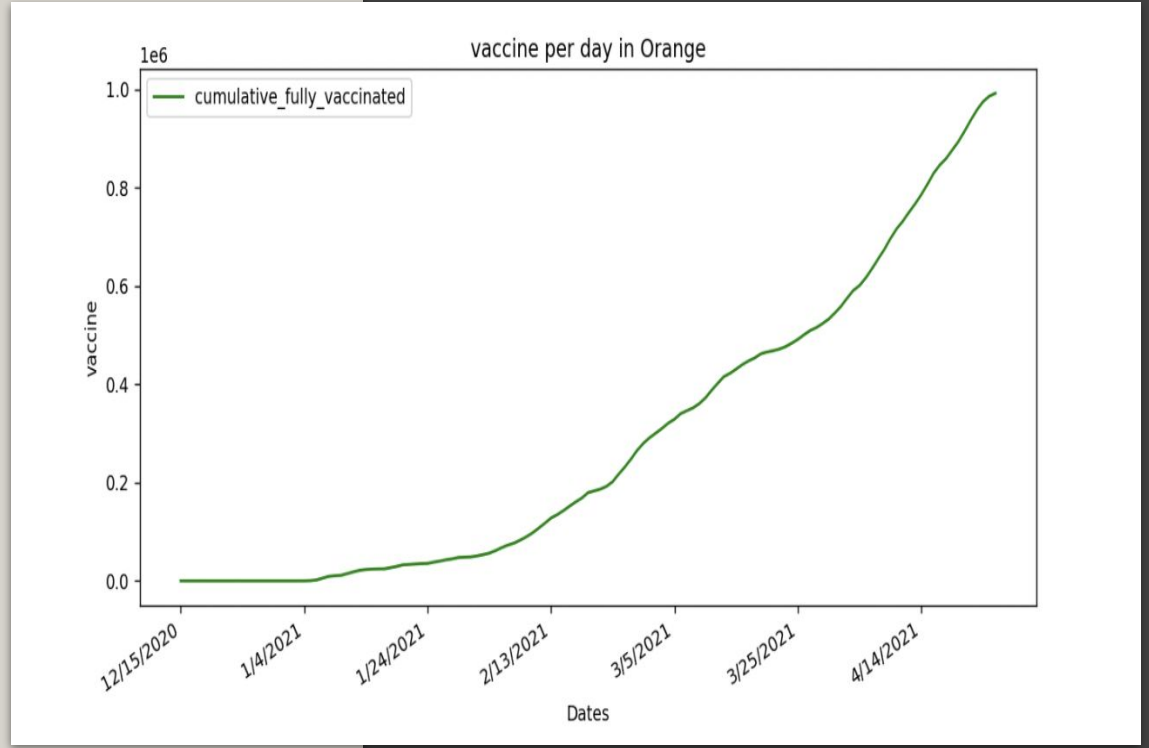
	date	area	area_type	population	cases	deaths	total_tests	positive_tests	reported_cases	reported_deaths	...	cumulative_moderna_dos
0	2/1/2020	Alameda	County	1685886.0	3.0	0.0	4.0	0.0	0.0	0.0	...	N
1	2/1/2020	Alpine	County	1117.0	0.0	0.0	0.0	0.0	0.0	0.0	...	N
2	2/1/2020	Amador	County	38531.0	0.0	0.0	0.0	0.0	0.0	0.0	...	N
3	2/1/2020	Butte	County	217769.0	0.0	0.0	0.0	0.0	0.0	0.0		N

In [43]:
```
deaths_plot = ca_counties_cases_df.plot('date', 'deaths',kind="line", color="red", figsize=(8,5))
plt.title(f"Deaths per day in {county1}")
plt.xlabel("Dates")
plt.ylabel("Deaths")
plt.legend()
plt.gcf().autofmt_xdate()
plt.show()
```

```
<IPython.core.display.Javascript object>
```

How does vaccination rate affect the case / death rate?





Detail of Vaccine vs Death Rate per day in 2021:

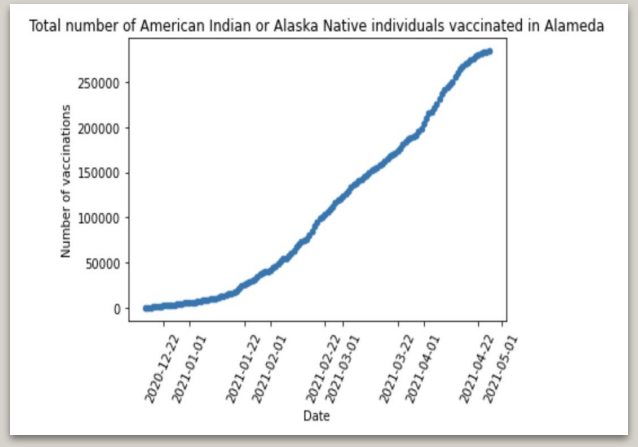
The initial rate of death did not show an initial impact from the vaccinations

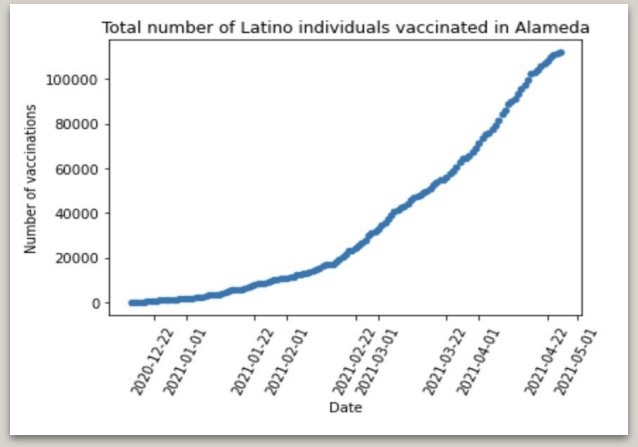
The rate of vaccination increase is in direct correlation to the decline in death

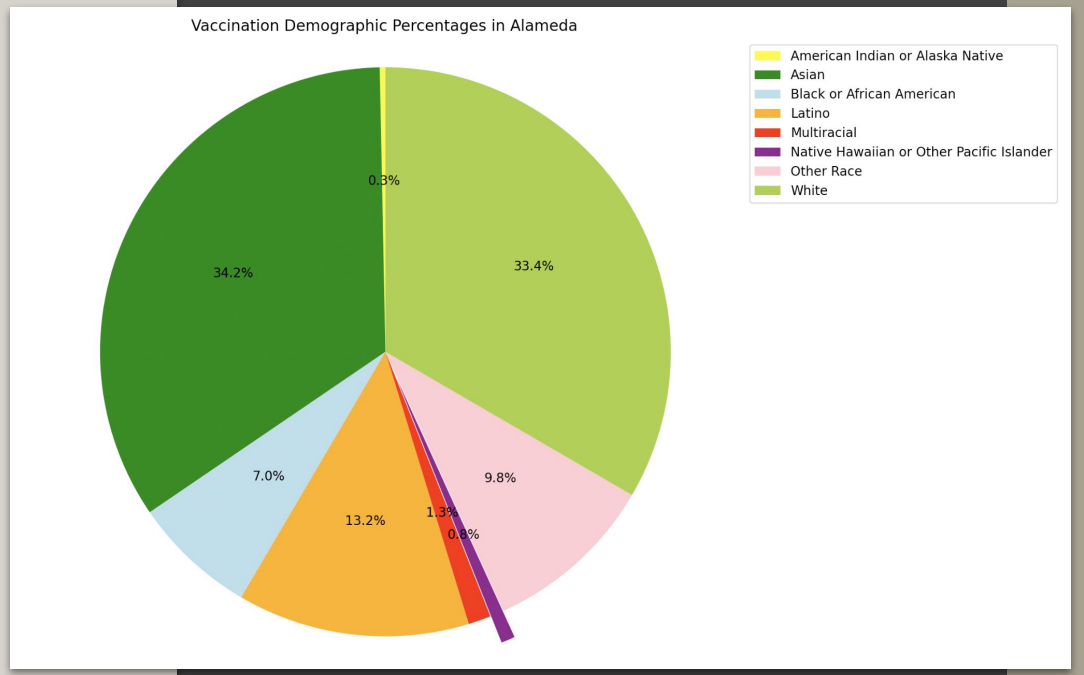
Which demographics have the highest vaccination rate in California?

```
#Use the loc function to index into the specific demographic
demo = county.loc[ca_counties_vaxdemo_df["demographic_value"]== "Asian"]
xaxis = np.arange (1.0, 134.0, 1)

#Generate scatter plot with administered date on the x axis and daily vaccination total on the y ax
demo.plot('administered_date', 'cumulative_at_least_one_dose',kind="scatter")
plt.title(f"Total number of Asian individuals vaccinated in {county2}")
plt.xlabel("Date")
plt.ylabel("Number of vaccinations")

# generate the regression line - the slope is equal to the vaccination rate of the demographic
x_values = xaxis
y_values = demo['cumulative_at_least_one_dose']
(slope, intercept, rvalue, pvalue, stderr) = linregress(x_values, y_values)
regress_values = x_values * slope + intercept
line_eq = "y = " + str(round(slope,2)) + "x + " + str(round(intercept,2))
plt.xticks(rotation = 65)
plt.figure(figsize=(10,6))


print (line_eq)
print (f"The vaccination rate of the Asian demographic in {county2} county is about {slope} people per day")

# to normalize the data we converted overall rate into a percentage of total demographic population in the county
a = demo.iloc[0, 4]
b = ((slope) / (a))*100

print (f"{b} percent of the Asian population is vaccinated per day")

pieb = demo.iloc[-1, 12]

repeat the process for the next seven demographics

# According to this data the White demographic has the highest overall vaccination rate at 2364.76
# individuals vaccinated per day in Alameda.

# After data normalization we found the Asian demographic has the highest percentage of their total population
# vaccinated per day with 0.49891304369930184 percent of their population per day
```

Which demographics have the highest vaccination rate in California?







As of April 2021:

American Indian or Alaskan Native vaccinations have scaled past 250K

Latino vaccinations have scaled past 1million

Conclusion

There is a strong correlation between COVID cases and COVID related deaths

Vaccination rates have a direct impact to the rate of deaths

There is a serious disparity in vaccination rates in some demographics

Recommendation: Improving the rate of vaccinations across all demographics will have a positive impact on the death rate


Closing