US COVID-19 Data Analysis

Part – 1:

By my analysis of US COVID-19 Datasets, I have answered following questions.

1. Plot the weekly number of new cases of covid in the following states: California, Oregon, Washington, and Nevada. A week starts on Monday and ends on Sunday.
2. Do those states follow the same trends?
3. Plot the weekly number of covid deaths in the following states: California, Oregon, Washington, and Nevada.
4. How do the deaths and the new cases compare?
5. Compute the number of new covid cases per 100,000 population in each county per week. What are the 20 highest rates achieved? Show the date, county, and the rate.
6. Compute the number of covid deaths per 100,000 population in each county per week. What are the 20 highest rates achieved? Show the date, county, and the rate.
7. What is the relation between the results in #5 and #6?
8. Find the week that each county achieved their highest rate of new covid cases per 100,000 population. If a county reaches the peak multiple ties pick the earliest one. Plot via a bar chart the number of countries that reached their peak each week. What does the plot indicate about the pandemic?
9. Compute the total number of covid cases in each county and the number of covid deaths. Compute the percent of the population in each county that is still alive and has had covid. Show the top 20 rates. Is any county getting close to herd immunity, that is 70% or more people immune?
10. Produce a Violin and box plots of all the values computed in #8. What do the plots show about the pandemic in the USA?

Part – 2:

One question about the pandemic is when we will reach herd immunity. Many places are tracking the vaccination rate. But those that have recovered from covid-19 are also immune. The goal here is to combine vaccine data with the covid data. The data sets involved are described below. The vaccine data, us_state_vaccinations.csv, is reported differently than the covid cases data.

1. In us_state_vaccinations.csv there are some gaps in the data. Will those gaps affect the results of the following questions? If no, why not. If yes explain the affect.

US COVID-19 Data Analysis

2. The vaccination data is reported by the entire county, state, US territory, and various groups like the department of defense, long term care and others. Presumably the US data includes the state data and the special group data but not the US territory data. Is the US territory data included in the US totals?
3. We would like to compute the percent of the population in each state that has been fully vaccinated. Presumably the data reported got the various groups is not included in the individual state data. If we distributed the number of people fully vaccinated in those groups to the states proportionally by population how much would that change the percent of people vaccinated in each state?
4. Produce a chart or table showing the percent of people in each state that have been fully vaccinated. Sort the data by the percent of people vaccinated. Your notebook should down-load the dataset so when it is run, we get the most recent data.
5. We want to see the progression from the start of the pandemic to now the percent of people that have some immunity in the US. Produce a plot of weekly data that contains three lines. One showing the percent of the living people have or had covid. One showing the percent of people that have had at least one covid vaccine shot. The third line showing the sum of the two to estimate the number of people that have some immunity. Over time sum will be an overestimate as people who had covid get a vaccine shot. As in #4 the notebook should download the data to up-to-date results.
6. Produce the same plot as in #5 for the states. You should have a function that takes the two-letter abbreviation for the state and produces the plot for the state.

Data

The data used is from USAFacts (https://usafacts.org). The data sets can be downloaded on the assignment page of the course website.

covid_confirmed_usafacts

Column Labels

State

2 letter abbreviate for the State

County Name

US COVID-19 Data Analysis

Name of the County

stateFIPS

Federal ID number for the state

countyFIPS

Federal ID number for the county

2020-01-22

Column contains the number of new cases.

covid_deaths_usafacts

This data set from USA Facts contains the covid deaths in each county in the country each day from the beginning of the pandemic.

Column Labels

State

2 letter abbreviate for the State

County Name

Name of the County

stateFIPS

Federal ID number for the state

countyFIPS

Federal ID number for the county

2020-01-22

Column contains the number of deaths.

US COVID-19 Data Analysis

covid_county_population_usafacts

Column Labels

countyFIPS

Federal ID number for the county

County Name

State

population

Population of the indicated county.

us_state_vaccinations.csv

Information about the data set:

https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations

URL for dataset:

https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/vaccinations/us_s-tate_vaccinations.csv

Column Labels

location:

Name of the state or federal entity.

date:

Date of the observation.

total_vaccinations:

Total number of doses administered. This is counted as a single dose, and may not equal the total number of people vaccinated, depending on the specific dose regime (e.g. people receive multiple doses). If a person receives one dose of the vaccine, this metric goes up by 1. If they receive a second dose, it goes up by 1 again.

total_vaccinations_per_hundred:

US COVID-19 Data Analysis

Total_vaccinations per 100 people in the total population of the state.

daily_vaccinations_raw:

Daily change in the total number of doses administered. It is only calculated for consecutive days. This is a raw measure provided for data checks and transparency, but we strongly recommend that any analysis on daily vaccination rates be conducted using daily_vaccinations instead.

daily_vaccinations:

New doses administered per day (7-day smoothed). For countries that don't report data on a daily basis, we assume that doses changed equally on a daily basis over any periods in which no data was reported. This produces a complete series of daily figures, which is then averaged over a rolling 7-day window. An example of how we perform this calculation can be found here.

daily_vaccinations_per_million:

Daily_vaccinations per 1,000,000 people in the total population of the state.

people_vaccinated:

Total number of people who received at least one vaccine dose. If a person receives the first dose of a 2-dose vaccine, this metric goes up by 1. If they receive the second dose, the metric stays the same.

people_vaccinated_per_hundred:

People_vaccinated per 100 people in the total population of the state.

people_fully_vaccinated:

Total number of people who received all doses prescribed by the vaccination protocol. If a person receives the first dose of a 2-dose vaccine, this metric stays the same. If they receive the second dose, the metric goes up by 1.

people_fully_vaccinated_per_hundred:

People_fully_vaccinated per 100 people in the total population of the state.

total_distributed:

Cumulative counts of COVID-19 vaccine doses recorded as shipped in CDC's Vaccine Tracking System.

US COVID-19 Data Analysis

total_distributed_per_hundred:

Cumulative counts of COVID-19 vaccine doses recorded as shipped in CDC's Vaccine Tracking System per 100 people in the total population of the state.

share_doses_used:

Share of vaccination doses administered among those recorded as shipped in CDC's Vaccine Tracking System.