COVID-19 Lockdowns and Waste Water Analysis Review

Kari Bellew

2025-07-25

Contents

Introduction

This is a project for "DTSA-5301: Data Science as a Field" on Coursera offered by the University of Colorado at Boulder. This project looks at COVID-19 data.

QUESTION 1: Were COVID Lockdowns Effective at Saving Lives?

This question was investigated using:

- Two different data sets from Data.gov,
- Pandemic response information by state on Wikipedia (https://en.wikipedia.org/wiki/U.S._state_and_local_government_responses_to_the_COVID-19_pandemic)
- Population data from Data.gov (https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html).

I arbitrarily picked states that had strict policies and states that had more relaxed policies. This method ignores factors such as population density, weather, and mobility of population (how often people went from one place to another).

The data was reported at different intervals and states had differing reporting timelines. To help quash differences this caused in numbers, I summarized counts to quarterly. Even with that, the data should not be compared on the same timelines state-to-state because the disease moved through our nation rather than hitting every place at the same time with the same intensity.

State Lockdown information

- California issued a stay-at-home order in mid-March of 2020 and did not lift it.
- Oregon issued a stay-at-home order in mid-March of 2020 and did not lift it.
- Texas issued a stay-at-home order in mid-March of 2020 and lifted it April 30.
- Wisconsin declared a stay-at-home order to be unconstitutional and, therefore, did not issue such an order.

The states have widely varying populations. Death rates were divided by population in an effort to modulate the numbers.

This approach ignores things like "Most employers switched to work-from-home regardless of the employee state of residence."

Population Information

State	Population
California	39,555,674
Oregon	4,237,224
Texas	29,149,458
Wisconsin	5,894,170

Dataset: Provisional Death Counts for Influenza, Pneumonia, and COVID-19

Data description

Provided by the U.S. Department of Health & Human Services.

Description from the Data.gov web site:

"Deaths counts for influenza, pneumonia, and COVID-19 reported to NCHS by week ending date, by state and HHS region, and age group."

Analysis

Load the data.

```
## Rows: 50400 Columns: 17
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (9): Data As Of, Start Week, End Week, Week Ending Date, Group, Indicato...
## dbl (8): MMWRyear, MMWRweek, COVID-19 Deaths, Total Deaths, Pneumonia Deaths...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Wrangle the Data

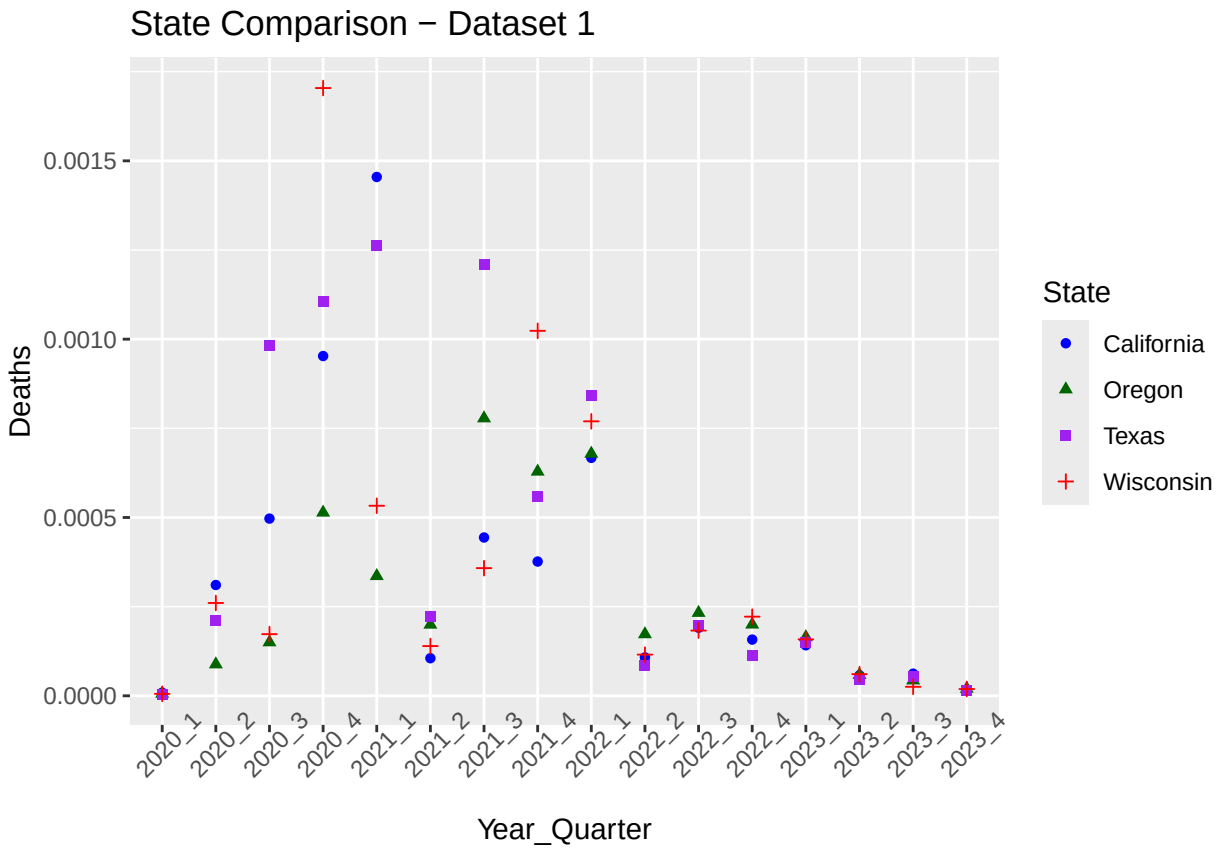
- Cleanse dates.
- Extract Year_Quarter from dates.
- Replace NA values with 0s.

```
##  Jurisdiction            Year      Year_Quarter           Deaths
##  Length:50400       Min.   :2020   Length:50400       Min.   :    0.0
##  Class :character   1st Qu.:2020   Class :character   1st Qu.:    0.0
##  Mode  :character   Median :2021   Mode  :character   Median :   10.0
##                     Mean   :2021                      Mean   :  136.3
##                     3rd Qu.:2022                      3rd Qu.:   63.0
##                     Max.   :2023                      Max.   :25974.0
```

Extract Data for States

- Extract data for states.
- Sum data per Year_Quarter.

```
## `summarise()` has grouped output by 'Year_Quarter'. You can override using the
## `.groups` argument.
```



Visualization

Conclusion

This data set shows some evidence that the tight policies, such as lock downs, saved large numbers of lives.

The data visualization shows spikes in deaths in March and April of 20202 for all states. Lock downs were issued in March of 2020. After Texas reopened it had a spike in deaths that was not seen in states with lock downs. Wisconsin numbers were similar to Texas for quarters 2-4 in 2021. The first vaccine was available at the end of 20202 and numbers leveled out and rates were similar across all states regardless of lockdown policy.

Dataset: Provisional COVID-19 death counts, rates, and percent of total deaths, by jurisdiction of residence

Data description

Provided by the U.S. Department of Health & Human Services.

Description from the Data.gov web site:

"This file contains COVID-19 death counts, death rates, and percent of total deaths by jurisdiction of residence. The data is grouped by different time periods including 3-month period, weekly, and total (cumulative since January 1, 2020). United States death counts and rates include the 50 states, plus the District of Columbia and New York City. New York state estimates exclude New York City. Puerto Rico is included in HHS Region 2 estimates.

"Deaths with confirmed or presumed COVID-19, coded to ICD–10 code U07.1. Number of deaths reported in this file are the total number of COVID-19 deaths received and coded as of the date of analysis and

may not represent all deaths that occurred in that period. Counts of deaths occurring before or after the reporting period are not included in the file.

"Data during recent periods are incomplete because of the lag in time between when the death occurred and when the death certificate is completed, submitted to NCHS and processed for reporting purposes. This delay can range from 1 week to 8 weeks or more, depending on the jurisdiction and cause of death.

"Death counts should not be compared across states. Data timeliness varies by state. Some states report deaths on a daily basis, while other states report deaths weekly or monthly.

"The ten (10) United States Department of Health and Human Services (HHS) regions include the following jurisdictions. Region 1: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont; Region 2: New Jersey, New York, New York City, Puerto Rico; Region 3: Delaware, District of Columbia, Maryland, Pennsylvania, Virginia, West Virginia; Region 4: Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee; Region 5: Illinois, Indiana, Michigan, Minnesota, Ohio, Wisconsin; Region 6: Arkansas, Louisiana, New Mexico, Oklahoma, Texas; Region 7: Iowa, Kansas, Missouri, Nebraska; Region 8: Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming; Region 9: Arizona, California, Hawaii, Nevada; Region 10: Alaska, Idaho, Oregon, Washington.

"Rates were calculated using the population estimates for 2021, which are estimated as of July 1, 2021 based on the Blended Base produced by the US Census Bureau in lieu of the April 1, 2020 decennial population count. The Blended Base consists of the blend of Vintage 2020 postcensal population estimates, 2020 Demographic Analysis Estimates, and 2020 Census PL 94-171 Redistricting File (see https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020-2021/methods-statement-v2021.pdf).

"Rates are based on deaths occurring in the specified week/month and are age-adjusted to the 2000 standard population using the direct method (see https://www.cdc.gov/nchs/data/nvsr/nvsr70/nvsr70-08-508.pdf). These rates differ from annual age-adjusted rates, typically presented in NCHS publications based on a full year of data and annualized weekly/monthly age-adjusted rates which have been adjusted to allow comparison with annual rates. Annualization rates presents deaths per year per 100,000 population that would be expected in a year if the observed period specific (weekly/monthly) rate prevailed for a full year.

"Sub-national death counts between 1-9 are suppressed in accordance with NCHS data confidentiality standards. Rates based on death counts less than 20 are suppressed in accordance with NCHS standards of reliability as specified in NCHS Data Presentation Standards for Proportions (available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf.)."

Analysis

Load the data

```
## Rows: 37765 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (6): data_as_of, Jurisdiction_Residence, Group, data_period_start, data_...
## dbl (8): COVID_deaths, COVID_pct_of_total, pct_change_wk, pct_diff_wk, crude...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
##   data_as_of        Jurisdiction_Residence    Group
##  Length:37765       Length:37765           Length:37765
##  Class :character   Class :character       Class :character
##  Mode  :character   Mode  :character       Mode  :character
##
```

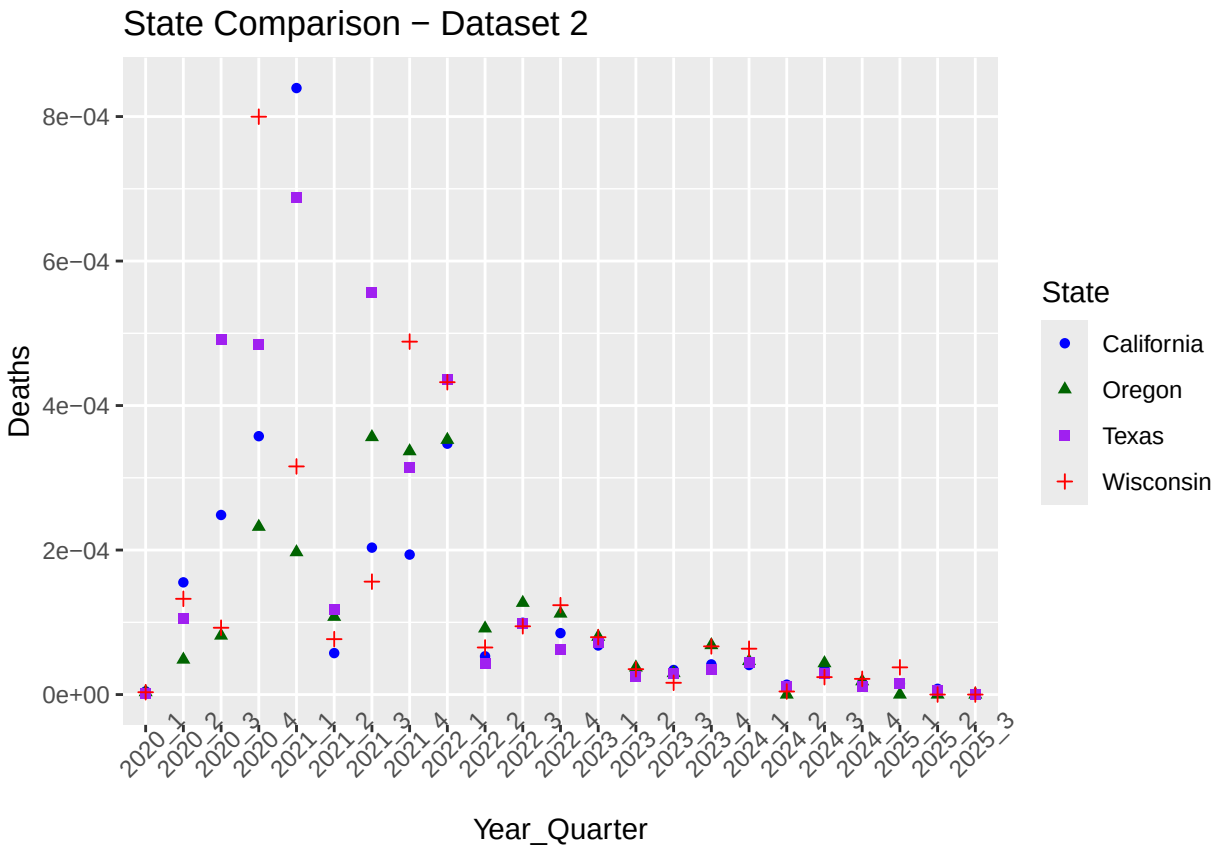
```
##
##
##
##  data_period_start  data_period_end     COVID_deaths     COVID_pct_of_total
##  Length:37765       Length:37765       Min.   :      0   Min.   : 0.000
##  Class :character   Class :character   1st Qu.:     54   1st Qu.: 3.400
##  Mode  :character   Mode  :character   Median :   1481   Median : 7.400
##                                        Mean   :  23270   Mean   : 8.003
##                                        3rd Qu.:  16416   3rd Qu.:10.900
##                                        Max.   :1230294   Max.   :67.000
##                                        NA's   :4393      NA's   :4393
##  pct_change_wk       pct_diff_wk      crude_COVID_rate aa_COVID_rate
##  Min.   :-100.000   Min.   :-13.100   Min.   :  0.00   Min.   :  0.00
##  1st Qu.: -14.400   1st Qu.: -0.600   1st Qu.:  1.00   1st Qu.:  0.80
##  Median :   0.000   Median :  0.000   Median : 97.75   Median : 78.85
##  Mean   :   3.449   Mean   : -0.007   Mean   :155.40   Mean   :127.79
##  3rd Qu.:  15.100   3rd Qu.:  0.500   3rd Qu.:312.20   3rd Qu.:258.27
##  Max.   : 900.000   Max.   : 26.500   Max.   :547.00   Max.   :466.20
##  NA's   :24792      NA's   :24066     NA's   :6911     NA's   :6911
##  crude_COVID_rate_ann aa_COVID_rate_ann   footnote
##  Min.   :   0.00      Min.   :   0.00   Length:37765
##  1st Qu.:  24.20      1st Qu.:  19.80   Class :character
##  Median :  48.10      Median :  40.30   Mode  :character
##  Mean   :  63.85      Mean   :  52.51
##  3rd Qu.:  68.80      3rd Qu.:  55.90
##  Max.   :3099.00      Max.   :2518.40
##  NA's   :6911         NA's   :6911
```

Wrangle the Data

- Cleanse dates.
- Extract Year_Quarter from dates.
- Replace NA values with 0s.

Extract data for states

```
## `summarise()` has grouped output by 'Year_Quarter'. You can override using the
## `.groups` argument.
```



Visualization

Conclusion

This data set shows convincing evidence that the tight policies, such as lock downs, saved large numbers of lives. The data visualization shows in the comparison of these four statest that Texas had the highest death rates throughout the pandemic and Wisconsin had the second highest after 2021.

QUESTION 2: Is Testing Waste Water for COVID an Effective Method to Predict Active Cases?

Dataset: COVID-19 National Wastewater Data

Data description:

Provided by the CDC. (https://www.cdc.gov/nwss/rv/COVID19-statetrend.html)

Excerpt of description from the CDC web site

"Wastewater monitoring can detect viruses spreading from one person to another within a community earlier than clinical testing and before people who are sick go to their doctor or hospital. It can also detect infections without symptoms. If you see increased wastewater viral activity levels, it might indicate that there is a higher risk of infection. See how to protect yourself from respiratory viruses.

"About the Wastewater Viral Activity Level: The wastewater viral activity level indicates whether the amount of virus in the wastewater is very low, low, moderate, high, or very high. The wastewater viral activity levels may indicate the risk of infection in an area. The wastewater viral activity levels are categorized as follows:

- "Up to 1.5 – Very Low
- "Greater than 1.5 and up to 3 – Low
- "Greater than 3 and up to 4.5 – Moderate
- "Greater than 4.5 and up to 8 – High
- "Greater than 8 – Very High"

Analysis

Load the data

```
## Rows: 13559 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (4): State/Territory, Data_Collection_Period, WVAL_Category, Coverage
## dbl  (3): State/Territory_WVAL, National_WVAL, Regional_WVAL
## dttm (1): date_updated
## date (1): Week_Ending_Date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Wrangle the Data

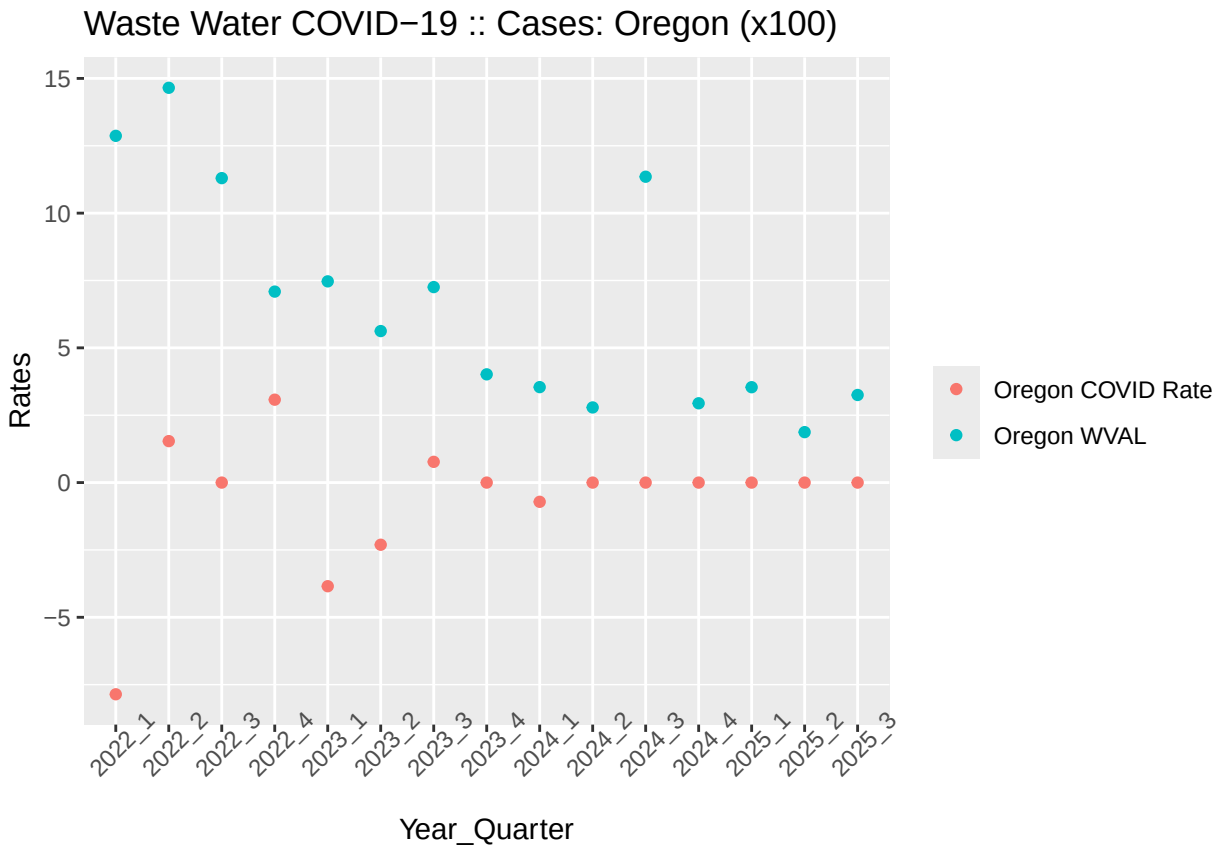
- Cleanse dates.
- Extract Year_Quarter from dates.
- Average rate data by Year_Quarter.

```
##  Year_Quarter         State_WVAL     National_WVAL   Regional_WVAL
##  Length:15          Min.   : 1.873   Min.   :1.902   Min.   : 2.239
##  Class :character   1st Qu.: 3.396   1st Qu.:2.794   1st Qu.: 3.347
##  Mode  :character   Median : 5.624   Median :5.996   Median : 5.847
##                     Mean   : 6.638   Mean   :5.243   Mean   : 5.681
##                     3rd Qu.: 9.384   3rd Qu.:7.353   3rd Qu.: 7.004
##                     Max.   :14.652   Max.   :9.231   Max.   :11.494
```

Extract and wrangle Oregon case data from the "Provisional COVID-19 death counts, rates, and percent of total deaths, by jurisdiction of residence" data already loaded

- Cleanse dates.
- Extract Year_Quarter from dates.
- Average COVID case rate.

```
##  Year_Quarter       crude_COVID_rate
##  Length:23          Min.   :-19.16667
##  Class :character   1st Qu.: -0.74176
##  Mode  :character   Median :  0.00000
##                     Mean   :  0.01433
##                     3rd Qu.:  0.76923
##                     Max.   : 24.61538
```



Visualization

Conclusion

The waste water contamination rate does not match the number of reported cases. It would seem that waste water is not a good indicator of the number of cases. We are seeing symptoms decrease in severity so that infected people have little or no symptoms and, therefore, cases are not reported. This data on its own seems to indicate that waste water is not a good measure. The waste water data without disease symptom or severity information definitely has less value that it would have if it included these things.