Understanding COVID-19: National Trends and Local Impact

MSL

2024-12-05

Before you kint

Required Packages

This analysis requires several R packages to be installed and loaded for proper execution. Before running the code, please ensure the following packages are installed on your system: `tidyverse`, `ggplot2`, `knitr`, `scales`, `tigris`, `sf`, `broom`.

To check whether a package is installed, you can use the `require() function`. If the package is not installed, `require()` will return `FALSE`. You can then install the missing package using the `install.packages()` function.

Here is a code snippet to automate this process. Copy and paste the code in the console to execute it

```r
## this chunk will not be executed when ran as part of this notebook
## please run the code in your console to install these packages

# required packages
required_packages <- c("tidyverse", "ggplot2", "knitr", "scales", "tigris", "sf", "broom")

# Install missing packages
for (pkg in required_packages) {
  if (!require(pkg, character.only = TRUE)) {
    install.packages(pkg)
  }
}

# Load all packages
lapply(required_packages, library, character.only = TRUE)
```

Note on Code Chunks

For enhanced readability, the code chunks in this report have been hidden in the final output when the document is knit. While the underlying code is essential for data processing, analysis, and visualization, hiding the code allows the focus to remain on the findings and interpretations. Readers interested in the full code can refer to the source R Markdown file, which includes all the code necessary to reproduce this analysis. This approach ensures a clean and professional presentation of the results while maintaining transparency and reproducibility.

Introduction

The primary goal of this analysis is to explore the temporal and spatial patterns of COVID-19's impact in the United States, with a specific focus on the state of Oregon. At the national level, the analysis seeks to understand the progression of cases and deaths over time, examining trends in cumulative confirmed cases, cumulative deaths, and the death-to-case ratio. At the state level, the focus shifts to exploring disparities in lethality rates across Oregon's counties, with a particular interest in how these rates have evolved over time. The study also investigates whether the lethality rate in Multnomah County, Oregon's most populous county, has significantly declined since the launch of a large-scale immunization campaign in December 2020. This question of interest aims to evaluate the effectiveness of public health interventions and vaccination efforts in reducing the severity of outcomes during the pandemic.

About the Data

This analysis utilizes two daily time series datasets tracking the progression of COVID-19 in the United States, reported at the county level. The datasets include `time_series_covid19_confirmed_US.csv`, which records confirmed cases, and `time_series_covid19_deaths_US.csv`, which captures reported deaths. Both datasets are structured to provide cumulative counts over time and are updated regularly to address any identified inaccuracies. These datasets offer a granular view of COVID-19 trends, enabling detailed analysis of regional impacts and temporal patterns.

Reshape and clean the data

The datasets were transformed from a wide to a long format to facilitate analysis, with each row representing a unique county-date combination. Columns such as latitude and longitude were converted to numeric for geographic computations, while categorical variables, including country, state, and county identifiers, were recast as factors to streamline grouping and summarization. The date column was converted to a standard Date format, and rows with missing dates were removed to ensure temporal consistency. The FIPS code was ensured to be numeric, accounting for potential missing values, and negative values in the cases column were replaced with zeros to correct any data inconsistencies. These cleaning steps ensure the datasets are structured, accurate, and ready for analysis.

Following these steps, both us_cases_clean and us_deaths_clean contain 3,819,906 rows and 14 columns. Geographic identifiers include UID, FIPS, and Combined_Key for unique record identification, while iso2, iso3, and code3 provide standardized country codes. Spatial attributes, such as Admin2 (county name), Province_State (state name), and Country_Region (set to "US"), are paired with latitude (Lat) and longitude (Long) for precise mapping. The date column records the daily timestamp, formatted as a standard Date, while the cases and deaths columns capture cumulative counts of confirmed cases or deaths, depending on the dataset.

```
## |                                                                          |
```

Visualize the Data

To understand the dynamics of COVID-19's progression and impact in the United States, this analysis presents two key levels of visualizations: national trends and state-specific insights for Oregon. At the national level, visualizations include trend analyses of cumulative confirmed cases, cumulative deaths, and death-to-case ratios over time, offering a long-term perspective on the pandemic's growth, its lethality, and its timeline. At the state level, a heatmap of Oregon highlights county-specific variations in lethality rates, providing insights into spatial disparities and the local impact of the pandemic. Oregon was chosen for this state-specific focus because it is where I currently reside, making it a region of personal and professional interest.

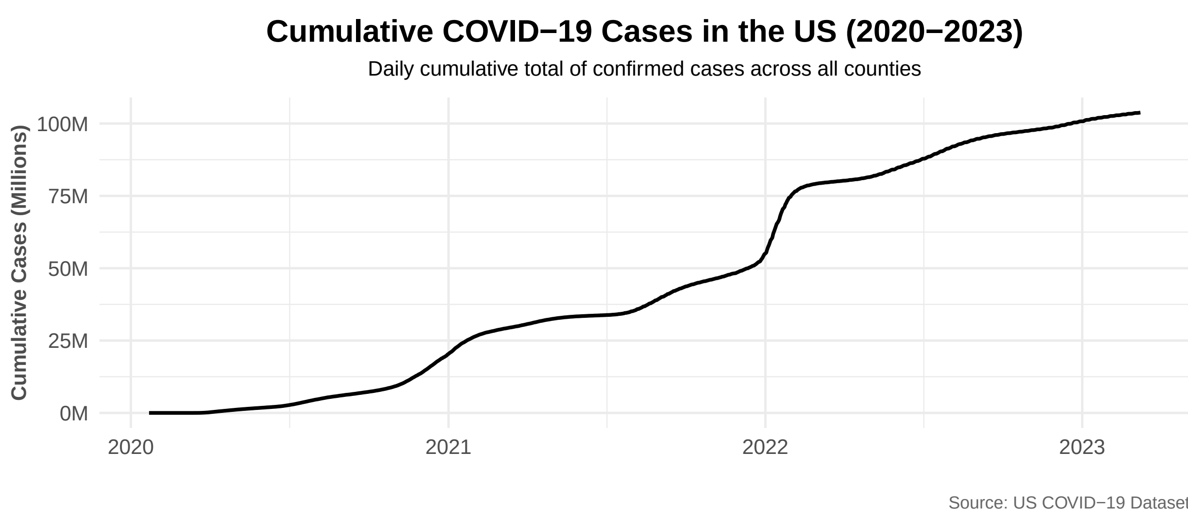
Together, these visualizations reveal critical temporal and geographic patterns, shedding light on the scale and severity of the pandemic across the country and within Oregon. These findings provide valuable context for resource allocation, public health interventions, and policymaking, particularly in addressing healthcare disparities and improving outcomes during future public health crises.

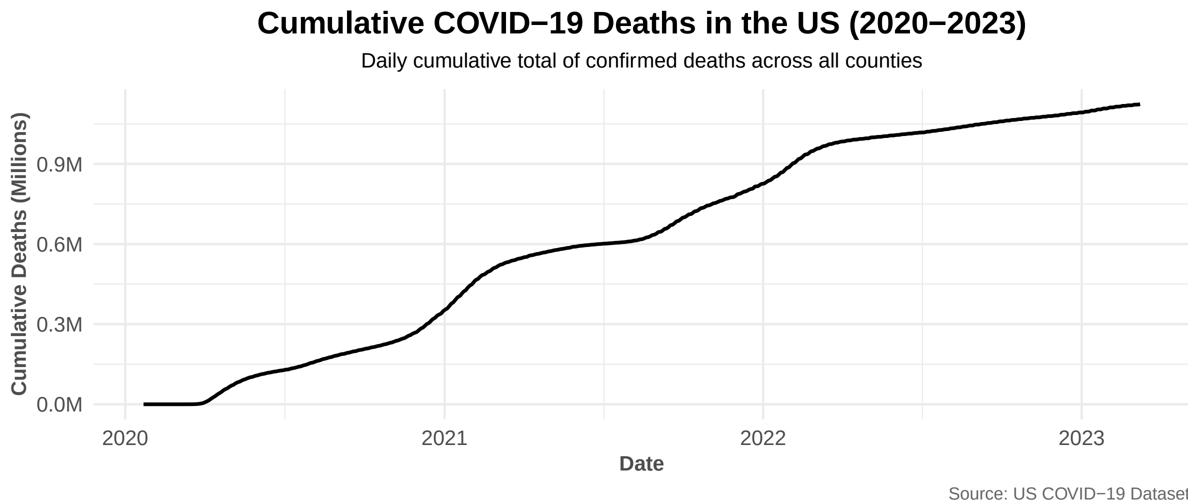
National Level

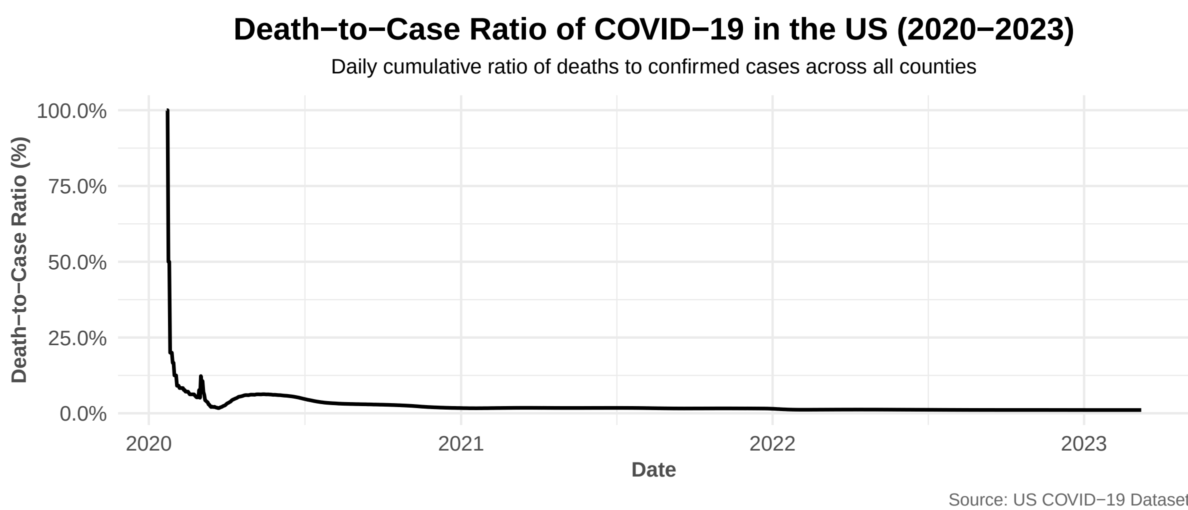
At the national level, the visualizations highlight the rapid and widespread progression of COVID-19 cases and deaths in the United States from 2020 to 2023. Cumulative confirmed cases surpassed 103.8 million by March 2023, with cumulative deaths exceeding 1.1 million. The death-to-case ratio, a measure of the virus's lethality, peaked at 100% during the pandemic's earliest days due to single case-death events, then stabilized to a median of 1.67% over time.

Key findings include sharp increases in both cases and deaths during late 2020 and mid-2021, corresponding to major waves of the pandemic that strained healthcare systems nationwide. Over time, declines in the death-to-case ratio reflect advancements in medical treatments, vaccination efforts, and public health measures, which collectively reduced the lethality of the virus.

These visualizations provide a comprehensive overview of the pandemic's scale and evolution, underscoring the importance of timely interventions, robust public health infrastructure, and continued monitoring to mitigate future health crises effectively.




Cumulative COVID−19 Deaths in the US (2020−2023)
Daily cumulative total of confirmed deaths across all counties
Cumulative Deaths (Millions)
0.0M
0.3M
0.6M
0.9M
2020
2021
2022
2023
Date
Source: US COVID−19 Dataset

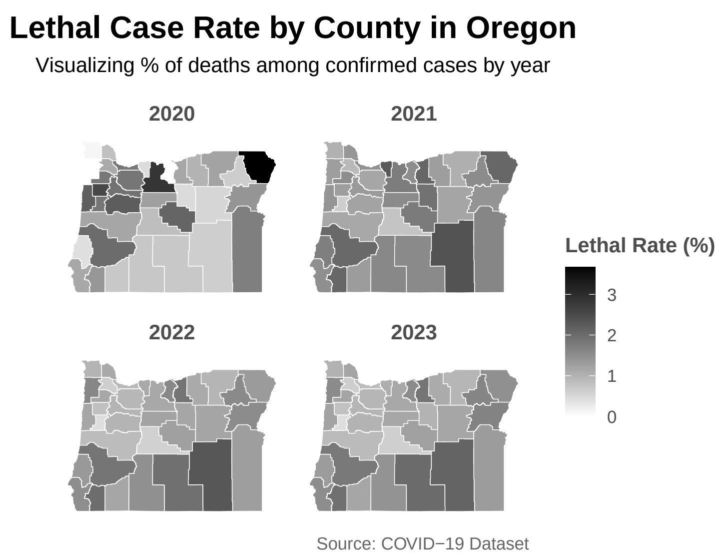

Death−to−Case Ratio of COVID−19 in the US (2020−2023)
Daily cumulative ratio of deaths to confirmed cases across all counties
Death−to−Case Ratio (%)
0.0%
25.0%
50.0%
75.0%
100.0%
2020
2021
2022
2023
Date
Source: US COVID−19 Dataset

State of Oregon

The heatmap visualization of lethality rates (% deaths among confirmed cases) across Oregon counties from 2020 to 2023 reveals distinct spatial and temporal patterns. In 2020, counties such as Wallowa (3.67%), Wasco (2.86%), and Polk (2.55%) exhibited the highest lethality rates, likely reflecting the initial impact of the pandemic in rural areas with limited healthcare access. Over subsequent years, many counties experienced a general decline in lethality rates, with urban counties like Multnomah and Washington reporting significantly lower rates by 2023 (0.86% and 0.63%, respectively).

Certain rural counties, including Harney (2.37%) and Josephine (2.08%), maintained elevated rates in 2021, suggesting ongoing challenges in these regions. By 2023, lethality rates had stabilized across much of the state, with counties like Lane (0.89%) and Deschutes (0.63%) achieving some of the lowest rates, indicative of effective public health measures and vaccine rollout efforts. However, counties such as Lake (2.01%) and Douglas (1.80%) continued to report higher lethality rates, underscoring the need for further investigation into underlying causes, including disparities in healthcare access and population vulnerabilities.

The detailed summary by county can be found in the Appendix.



Statistical Modeling: Lethality Rate in Multnomah County

To assess whether there has been a statistically significant decline in the lethality rate (% of deaths among confirmed cases) in Multnomah County (Oregon's most populated county), we performed a linear regression analysis. The model examines the relationship between the lethality rate and the date, focusing on the period starting December 14, 2020. This date marks the launch of Oregon's large-scale immunization campaign, as healthcare facilities began receiving the Pfizer-BioNTech vaccine. By analyzing daily data points, we aim to determine if the lethality rate has significantly decreased over time, potentially reflecting the impact of vaccination efforts and other public health measures.

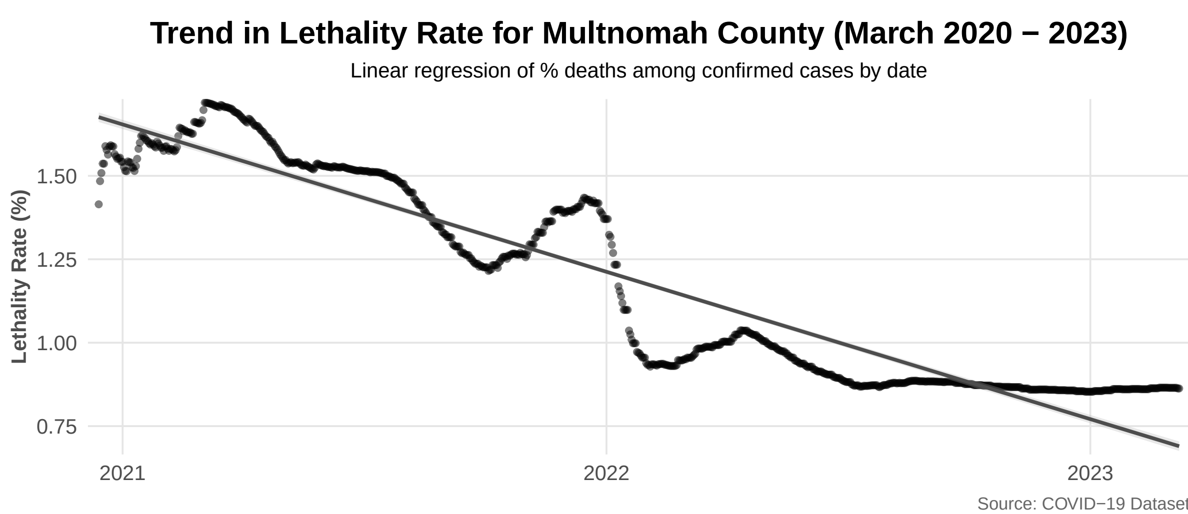
The regression model reveals a clear and statistically significant decline in the lethality rate over time. The intercept of the model, estimated at 24.19, represents the lethality rate at the beginning of the analyzed period. This high initial value reflects the severity of outcomes during a time when vaccine coverage was minimal, and public health measures were still ramping up.

The slope of the regression line, estimated at -0.0012, indicates that the lethality rate decreased by approximately 0.12 percentage points per day over the analyzed period. This decline is statistically significant, with a p-value of virtually zero, confirming that the downward trend is not due to random variation. The strong negative relationship between time (date) and lethality rate, evidenced by the high t-statistic (-78.56), highlights the robustness of this trend.

These findings align with the timeline of increased vaccination coverage and improvements in COVID-19 treatment protocols. The results suggest that the state's immunization campaign and other public health measures played a critical role in reducing the lethality rate over time. The consistent and substantial decline underscores the importance of vaccination efforts in mitigating the severity of COVID-19 and improving health outcomes in Multnomah County.

Table 1: Regression Results for Lethality Rate in Multnomah County

Term	Estimate	Std. Error	t-Statistic	p-Value
(Intercept)	24.1890	0.2929	82.5904	0
date	-0.0012	0.0000	-78.5579	0



Conclusion and Discussion

This analysis provides an exploratory examination of COVID-19's progression and impact at both the national and state levels, with a specific focus on Oregon. At the national level, the trends reveal the scale and severity of the pandemic, with over 103.8 million cumulative cases and 1.1 million deaths by March 2023. The decline in the death-to-case ratio over time underscores the critical role of advancements in treatments, public health interventions, and vaccination campaigns in mitigating the lethality of the virus. In Oregon, spatial analysis highlighted significant disparities in lethality rates between urban and rural counties, with rural areas facing persistent challenges due to limited healthcare access and other factors. The regression analysis for Multnomah County demonstrated a statistically significant decline in lethality rates following the initiation of mass immunization efforts, reflecting the potential positive impact of public health measures.

These findings emphasize the importance of timely and sustained public health efforts in controlling pandemics and reducing mortality. While vaccination campaigns and healthcare advancements significantly lowered lethality rates, the disparities observed across counties highlight the need for targeted interventions, particularly in underserved rural areas. The analysis also underscores the importance of robust and accurate data collection to inform decision-making during public health crises. However, limitations such as data and model biases must be acknowledged. These biases remind us to interpret the results cautiously and to complement statistical findings with qualitative insights. This study provides valuable lessons for future public health responses, emphasizing the need for equitable healthcare access and ongoing monitoring to ensure effective management of health crises.

Potential Sources of Bias

1. **Data Bias**:

- The datasets rely on reported cases and deaths, which may underrepresent the true impact of COVID-19 due to underreporting or discrepancies in data collection methods. Factors such as limited testing capacity during early pandemic stages and variations in state and county reporting practices could influence the accuracy of the data.
- Missing data, particularly in rural areas, may disproportionately affect insights into counties with limited healthcare infrastructure, skewing lethality rate estimates.
- Cumulative data may fail to capture nuanced shifts, such as sudden surges or drops in cases and deaths, potentially smoothing over important short-term trends.

2. **Model Bias**:

- The linear regression model assumes a consistent, linear relationship between time and the lethality rate, which may oversimplify non-linear dynamics, such as abrupt changes due to policy interventions or new COVID-19 variants.
- The focus on a single explanatory variable (date) omits other potential contributors to lethality trends, such as vaccination rates, healthcare capacity, or demographic shifts.
- Excluding data prior to December 14, 2020, provides a cleaner analysis focused on the vaccination era but may introduce bias by omitting early pandemic dynamics.

3. **Analyst Bias**:

- The choice to focus on Multnomah County may reflect personal interest but excludes broader state-level variations that could offer more comprehensive insights.
- Emphasis on visualizations and temporal patterns may have overlooked the role of socioeconomic or policy-driven factors influencing outcomes.
- Interpretation of statistical significance, while data-driven, might downplay potential confounding factors not accounted for in the model.

Mitigation of Bias

1. **Data Validation**

Efforts were made to clean and preprocess the data, including handling missing values, correcting anomalies, and ensuring consistency in variable formats. These steps aim to reduce inaccuracies that could distort findings.

2. **Robust Analytical Approach**

Multiple visualization techniques were employed, such as line charts, regression analysis, and heatmaps, to provide diverse perspectives on the data. Using different methods ensures a more comprehensive understanding and minimizes reliance on a single analytical framework.

3. **Transparency of Assumptions**

The limitations of the linear regression model and the cumulative dataset are explicitly acknowledged to avoid overstating causal inferences. Additionally, the analysis emphasizes exploratory findings and does not assert definitive conclusions about the impact of specific interventions.

Appendix

```r
sessionInfo()
```

```
## R version 4.4.2 (2024-10-31)
## Platform: aarch64-apple-darwin20
## Running under: macOS Sonoma 14.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib;  LAPACK ve
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/Los_Angeles
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] broom_1.0.7     sf_1.0-19       tigris_2.1      scales_1.3.0
##  [5] knitr_1.48      lubridate_1.9.3 forcats_1.0.0   stringr_1.5.1
##  [9] dplyr_1.1.4     purrr_1.0.2     readr_2.1.5     tidyr_1.3.1
## [13] tibble_3.2.1    ggplot2_3.5.1   tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] gtable_0.3.6       xfun_0.49          lattice_0.22-6     tzdb_0.4.0
##  [5] vctrs_0.6.5        tools_4.4.2        generics_0.1.3     curl_5.2.3
##  [9] parallel_4.4.2     proxy_0.4-27       fansi_1.0.6        pkgconfig_2.0.3
## [13] Matrix_1.7-1       KernSmooth_2.23-24 uuid_1.2-1         lifecycle_1.0.4
## [17] compiler_4.4.2     farver_2.1.2       munsell_0.5.1      htmltools_0.5.8.1
## [21] class_7.3-22       yaml_2.3.10        pillar_1.9.0       crayon_1.5.3
## [25] classInt_0.4-10    nlme_3.1-166       tidyselect_1.2.1   digest_0.6.37
## [29] stringi_1.8.4      labeling_0.4.3     splines_4.4.2      fastmap_1.2.0
## [33] grid_4.4.2         colorspace_2.1-1   cli_3.6.3          magrittr_2.0.3
## [37] utf8_1.2.4         e1071_1.7-16       withr_3.0.2        backports_1.5.0
## [41] rappdirs_0.3.3     bit64_4.5.2        timechange_0.3.0   rmarkdown_2.28
## [45] httr_1.4.7         bit_4.5.0          hms_1.1.3          evaluate_1.0.1
## [49] mgcv_1.9-1         rlang_1.1.4        Rcpp_1.0.13-1      glue_1.8.0
## [53] DBI_1.2.3          rstudioapi_0.17.1  vroom_1.6.5        R6_2.5.1
## [57] units_0.8-5
```

% of Deaths Among Confirmed Cases by County and Year in Oregon

Table 2: % of Deaths Among Confirmed Cases by County and Year in Oregon

County	2020	2021	2022	2023
Baker	1.41	1.43	1.54	1.66
Benton	1.85	0.63	0.42	0.42
Clackamas	1.86	1.17	0.94	0.95
Clatsop	0.11	1.03	0.98	1.00
Columbia	0.82	1.38	1.13	1.19
Coos	0.41	1.72	1.35	1.30
Crook	2.10	1.77	1.25	1.24
Curry	1.14	1.49	1.51	1.52
Deschutes	0.84	0.77	0.59	0.63
Douglas	1.99	2.04	1.86	1.80
Gilliam	1.16	2.09	1.88	1.87
Grant	0.53	1.18	1.19	1.17
Harney	0.63	2.37	2.31	2.11
Hood River	0.45	2.27	1.14	1.06
Jackson	0.71	1.31	1.18	1.16
Jefferson	1.25	1.54	1.21	1.15
Josephine	1.38	2.08	1.98	1.94
Klamath	0.73	1.59	1.48	1.44
Lake	0.71	1.56	1.94	2.01
Lane	1.17	1.15	0.87	0.89
Lincoln	2.19	1.40	1.17	1.21
Linn	2.22	1.21	0.98	0.99
Malheur	1.72	1.61	1.28	1.31
Marion	1.90	1.34	0.98	0.96
Morrow	0.99	1.29	1.11	1.10
Multnomah	1.84	1.45	0.92	0.86
Polk	2.55	1.28	0.82	0.81
Sherman	0.00	1.52	1.59	1.55
Tillamook	0.00	1.28	1.61	1.55
Umatilla	1.21	1.05	0.97	0.96
Union	0.64	1.53	1.57	1.63
Wallowa	3.67	2.07	1.32	1.48
Wasco	2.86	1.75	1.14	1.15
Washington	1.12	0.91	0.63	0.63
Wheeler	0.47	1.91	1.17	0.99
Yamhill	1.81	1.51	1.16	1.18

References

1. Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). COVID-19 Data Repository. Available at: https://github.com/CSSEGISandData/COVID-19.

2. Wickham, H. et al. (2019). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. R package version 3.3.3. Available at: https://ggplot2.tidyverse.org.

3. Pebesma, E. (2018). *Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal*, 10(1), 439-446. Available at: https://r-spatial.github.io/sf.

4. Centers for Disease Control and Prevention (CDC). COVID-19 Data Tracker. Available at: https://covid.cdc.gov/covid-data-tracker.

5. Oregon Health Authority. COVID-19 in Oregon: Data and Trends. Available at: https://www.oregon.gov/oha.