JOURNAL OF
EMERGING INVESTIGATORS

Article

Correlation between shutdowns and CO levels across the United States

Arnav Gupta[1] and Daniel Dudek[2]
[1]American High School, Fremont, California, USA
[2]University of Texas Arlington, Arlington, Texas, USA

SUMMARY

We conducted research in Summer 2020 to analyze the effect of the shutdown orders due to the COVID-19 pandemic on the carbon monoxide (CO) levels across the United States. The effects of these lockdowns have not yet been entirely studied, but they have caused wide-spread lifestyle changes. Since the shutdown orders during the pandemic prevented most public places from conducting any kind of business, we hypothesized that CO levels would drop as there was no commute, fewer buildings being used, and fewer people traveling as much as they previously would. We collected publicly available data from the EPA's Daily Air Quality Data and analyzed in R. We selected the 15 states based on the greatest number of coronavirus cases on August 20th, 2020. Each state had a mean value of CO concentration for each date that the data was available in the sites that was an average of multiple recorded values across counties in the state. In almost every state, the CO levels went down starting from February, with the lowest CO levels during the shutdown, indicating that the shutdown likely could have led to a decrease in CO levels. Since some states saw CO levels start to drop before the shutdown orders were enacted, the absence of people on roads and the lack of human activity due to concerns over the spread of the virus in general could have also been a contributing factor. As some states opened and restrictions eased, the CO level started rising then fluctuated again, similar to before the shutdowns.

INTRODUCTION

Humans have contributed significantly to climate change over the past few decades (1). One example is the increase of Carbon Monoxide, a weak greenhouse gas. CO (Carbon Monoxide) reacts with hydroxyl radicals in the atmosphere and thus prevents those radicals from reacting and degrading other more harmful greenhouse gases such as Carbon Dioxide. When wood-based fuels such as coal, oil and wood burn incompletely or inefficiently, CO is produced, which is then spread throughout the lower atmosphere. However, most of the CO produced comes from burning fossil fuels in vehicles, factories and power plants. We researched how the decrease of human activity resulting from the onset of the COVID-19 pandemic affected CO. These lockdowns along with social distancing practices advised by the CDC were to prevent the further spread of the virus in places such as schools and offices. The shutdowns in March and April in 2020 gave us the perfect opportunity to do so.

The research conducted in this study is important because with the effects of climate change starting to appear such as the temperature rising more than 2°F in the last century, the sea level rising, and the ice caps melting, we need to take action. Humans have contributed significantly to the acceleration of climate change through increasing pollution levels, and this presents one of the most substantial issues that face the current generations. Increasing pollution levels is one of the most significant issues that face the current generations. Pollutants such as CO (carbon monoxide), CO2 (carbon dioxide), SO2 (sulfur dioxide), NO2 (nitrogen dioxide), PM 2.5 (fine particulate matter 2.5), PM 10 (fine particulate matter 10), and O3 (ozone) are the most problematic and dangerous to agriculture, human health, and the environment. Air pollution causes many health problems, ranging from minor upper respiratory irritation to chronic respiratory and heart disease, lung cancer, acute respiratory infections in children and chronic bronchitis in adults, aggravating pre-existing heart and lung disease, or asthmatic attacks; also, being in polluted environments is not very safe for people belonging to sensitive groups (2).

In addition to health problems, air pollution also has devastating effects on the environment (3), some of them being acute morbidity in various species of trees and soil and forest contamination (3). Mining, deforestation, factories, power plants, airplanes, cars, and the burning of fossil fuels are the main contributors to the pollution we observe today (4–6). All these activities are carried out every day globally which creates a significant amount of pollution (6). Air pollution is a global problem that is only getting worse (7). With all this increase in pollution, we must also consider the health problems, one example is how CO, one of the pollutants that is very harmful to humans, leads to carbon monoxide poisoning. CO poisoning is taken seriously as when the concentration of CO is higher than the safe limits, out bodies replace the oxygen in the red blood cells with Carbon Monoxide, this leads to serious tissue damage or even death and is also pretty common. With pollution rising and more cars on the road, the CO levels have also gone up in the past decades (8).

Therefore, we hypothesized that the CO levels in the United States would decrease with the decrease in human

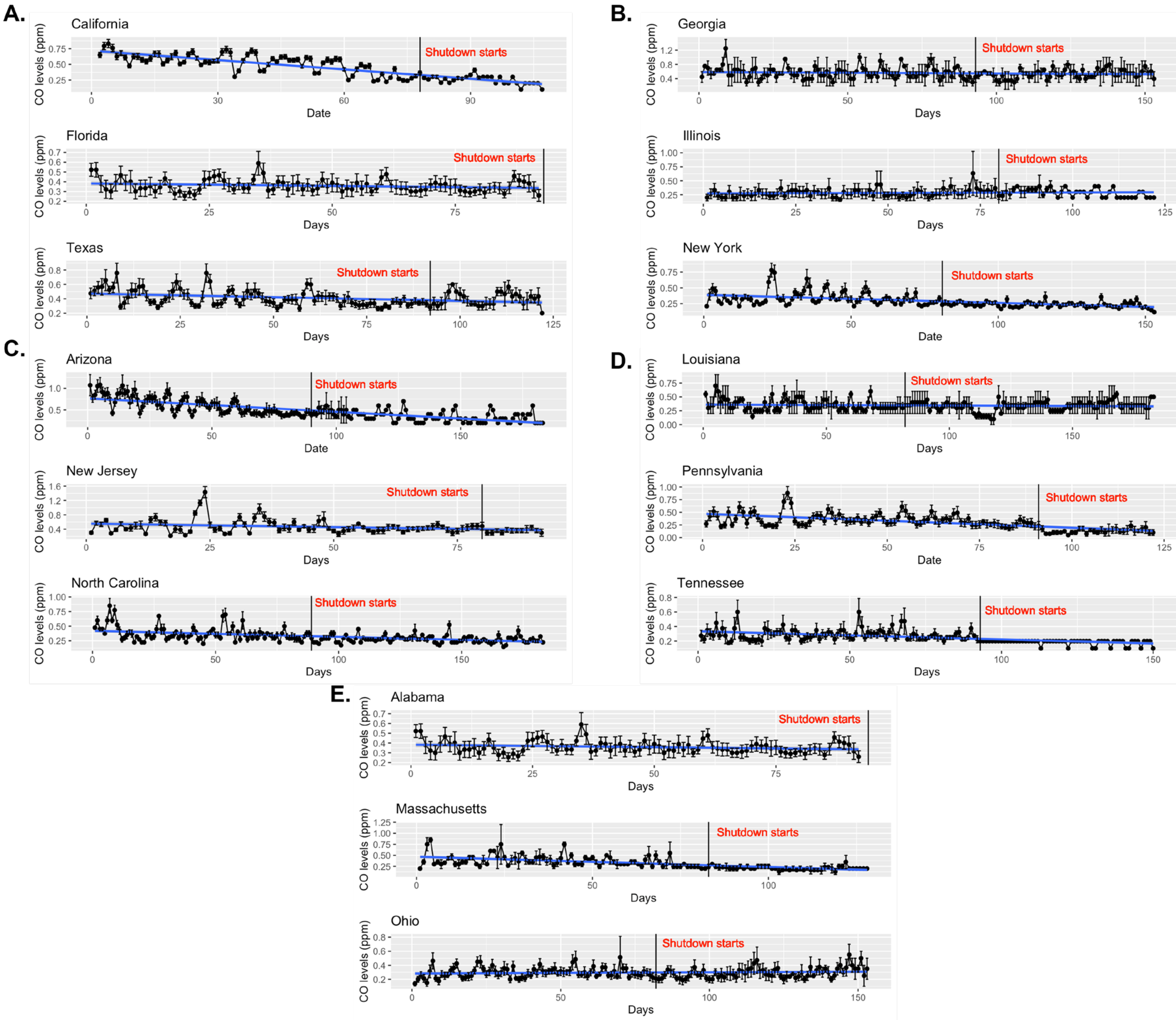

Figure 1. CO levels since January 2020. Y-axes represent CO levels in ppm and X-axes represent Dates into 2020. The line of best fit (blue) is shown along with the date that the shutdowns started in each state. Note: not all states have the same X-axis due to data availability at the time of the study. A) California, Florida, and Texas, B) Georgia, Illinois, and New York, C) Arizona, New Jersey, and North Carolina, D) Louisiana, Pennsylvania, and Tennessee, E) Alabama, Massachusetts, and Ohio.

activity driven by the limitations put in place during the beginning of the COVID-19 pandemic that we are currently enduring. The data analysis we conducted brought us one step closer to inferring how a potential decrease in human activity would affect CO levels. From our analyses, we have found that the lockdowns in early 2020 did have a statistically significant effect on CO levels. We've concluded that human activity as mentioned above does likely contribute to pollution.

RESULTS

The effect of the COVID-19 shutdown on CO levels was examined by analyzing the mean from states that had the most COVID-19 cases. Outdoor sensors in counties of these states measured the daily level of CO in part per million (ppm) and we used these values to calculate the mean of all the sites in each state for each day. This data was collected from the Environment Protection Agency's (EPA) Outdoor Air Quality Data website for 15 states and the mean for each state was calculated for all the data available.

Most of the states showed a decline in CO levels from the beginning of the year (**Figure 1A–E**). We applied a regression model, which showed the overall trend in CO levels with dates, accounting for all the variation in the available data. The data that we analyzed included a period of lockdowns for each state starting in March and April 2020 and all states included data for the dates after the lockdown, but the number

State	Shutdown date	Days since January 2020
California	March 19th	78
Florida	April 3rd	93
Texas	April 2nd	92
Georgia	April 3rd	93
Illinois	March 21st	80
New York	March 22nd	81
Arizona	March 31st	90
New Jersey	March 21st	80
North Carolina	March 30th	89
Louisiana	March 23rd	82
Pennsylvania	April 1st	91
Tennessee	March 31st	90
Alabama	April 4th	94
Massachusetts	March 24th	83
Ohio	March 23rd	82

Table 1. Shutdown dates for each state. Shutdown dates for each state along with the days since January 1st, 2020.

State	Pearson correlation value
California	-0.855
Florida	-0.217
Texas	-0.343
Georgia	-0.112
Illinois	+0.881
New York	-0.542
Arizona	-0.777
New Jersey	-0.283
North Carolina	-0.498
Louisiana	-0.102
Pennsylvania	-0.658
Tennessee	-0.577
Alabama	-0.217
Massachusetts	-0.651
Ohio	+0.103

Table 2. Pearson correlation values of States. Some states have a stronger correlation, such as California while some have weak correlations, such as Illinois.

of dates after the lockdown varied for each state due to the unavailability of data. This data after the lockdown was important as it shows how CO levels rose in certain states. The linear regression line in the data shows the trendline for CO levels in each state; states where this line had the greatest negative slope had the most substantial decrease in CO levels. For example, in California, the shutdown began on March 19th (the 78th day of the year 2020) (**Table 1**). A decline and less fluctuations after the 78th day was seen, indicating that the CO levels dropped and did not fluctuate as they did before the shutdown (**Figure 1A**). The same trend is seen for states such as Pennsylvania, Arizona, New York, Louisiana, and Tennessee (**Figures 1B-D**).

Data from all states was combined and analyzed to determine if a correlation existed between the CO levels and time since the beginning of the year. We then used a regression analysis to determine what proportion of the variation in CO levels was explained by the Dates variable. The Pearson Correlation (Cor function in R) test for all the states analyzed gave a result of -0.3244 which indicates a weak to moderate correlation between the mean of CO values and time within the shutdown, along with that the t-value was -15.461. The Pearson correlation values for each state show how some states had a stronger and more directed correlation between CO levels and days in the year compared to some states that showed an overall increase in CO levels (**Table 2**). The results indicated that the shutdown dates and CO levels had a degree of variation and that the results obtained were statistically significant (Tukey post hoc, adjusted R^2 = 0.105, p value = 2.2e-16). Some states saw a higher Correlation value, indicating that some states had more of a change in CO levels, such as Florida (Table 2, Pearson correlation test = -0.2166). Looking closely at the data, we saw that even though the CO levels did not drop much in some states, the fluctuations in CO values present before the lockdown were greatly reduced if not absent in Florida, New York, Arizona, New Jersey, North Carolina, Tennessee, and Massachusetts (**Figure 1A–E**). While this was the case in some states, other states such as California, Texas, Arizona, New York, and Pennsylvania experienced significant drops in the CO levels. To quantify the differences in data among the states, we performed a one-way ANOVA followed by a Tukey's post-hoc test (**Table 3**).

DISCUSSION

The data analysis of the CO levels in states in early 2020 showed a decrease in CO levels from the start of the year 2020 in January, indicating that the shutdown orders, the earliest one in March, did have a significant effect on the level of CO produced (**Figure 1A-E**). It is likely that the fear of the virus spreading in months before the lockdowns could have also lowered human activity on smaller level, for example, families not going to dinner or not going to parties, etc. The results seen in **Figure 1** suggest that human activity largely contributes to pollution, although there is no concrete evidence of human activity decreasing. The hypothesis that the CO levels will decrease due to human inactivity in context of e.g. traveling and social activities is supported by most of the 15 states (**Figure 1A-E**). This was confirmed using the Pearson correlation, R^2 value, and p value since they all showed a correlation between CO levels and the dates (Pearson = -0.324, R^2 value = 0.105, p value = 2.2e-16). The correlation value was evaluated on a graph with all 15 states, and the result was -0.3244, indicating a weak to moderate correlation. Some states had weaker Pearson correlation values, for example, Florida (-0.217) compared to California (-0.855). States such as Florida likely did not follow the shutdown orders or had shutdowns that were not strongly enforced. However, other states such as California, Texas, Arizona, New York, and Pennsylvania experienced significant drops in the CO levels, potentially because these states were more populated and there was a lot more human activity in

State	States Significantly Different From
Alabama	Arizona, California, Georgia, Illinois, New Jersey, New York, Tennessee
Arizona	Florida, Georgia, Illinois, Louisiana, Massachusetts, New York, North Carolina, Ohio, Pennsylvania, Tennessee, Texas
California	Florida, Georgia, Illinois, Louisiana, Massachusetts, New York, North Carolina, Ohio, Pennsylvania, Tennessee
Florida	Georgia, Illinois, New Jersey, New York, Tennessee
Georgia	Illinois, Louisiana, Massachusetts, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Tennessee Texas
Illinois	Louisiana, New Jersey, Tennessee
Louisiana	New Jersey, Tennessee, Texas
Massachusetts	New Jersey, Tennessee, Texas
New Jersey	New York, North Carolina, Ohio, Pennsylvania, Tennessee
New York	Texas
North Carolina	Tennessee, Texas
Ohio	Texas
Pennsylvania	Texas

Table 3. Significant differences in States analyzed with Tukey post-hoc. The states in the table have statistically different CO levels. For example, Arizona and Alabama have statistically different levels, the combinations of states not included were not significant.

these states which was suspended. Our analysis showed that there was a rise of CO levels in Arizona at around 120 days into 2020; this can be attributed to a reopening in the state and further indicates how human activity may affect pollution levels. Other states Arizona closed on March 31st, around 90 days in 2020. In the graph, a decline in CO levels around 75 days is seen. The CO levels stayed low until day 100 and around day 120, they started fluctuating again like they were until day 75. Surprisingly, Arizona reopened on May 8th, the 129th day of 2020. This included the opening of casinos, food and drink places, recreation, and personal care; this strongly suggests that the CO levels were affected by human activity. The same observation is seen in Ohio, which reopened on May 12, the 132nd day, where the CO levels clearly rose. Other states that saw a rise include Texas, Georgia, and North Carolina. This decrease and increase in CO levels is only seen because there was enough data for Arizona, Ohio, Texas, Georgia, and North Carolina to graph this. This would also have been the case for other states had there been enough data.

There was also a statistically significant difference between the state and CO level as determined by a one-way ANOVA ($F = 68.11$, $p < 2e-16$). We then performed a Tukey post-hoc test following the ANOVA test to look for differences of the mean CO levels in each state and correct for p-value inflation and for multiple comparisons when analyzing differences in significance between individual states. The results showed that several states were statistically different, and these are listed in Table 3 ($p < 0.05$).

Even with all these tests, the data is certainly not perfect, some of the factors that might have affected the data are that people did not follow the shutdown orders, that not enough data was available for each state, and that exceptions in the shutdown orders occurred in individual states. Since the research was done remotely, it would be difficult to measure independent data for each of the 15 states with more than one measurement site in each state, which is why the EPA dataset was chosen. Furthermore, the shutdown orders applied to all public places, offices, schools, and factories. Even with the shutdown orders, there might still have been people who did not follow the orders or continue non-essential tasks. In this case, the shutdown order did not apply to this group of people as their commute/activity still influenced the pollution levels.

The EPA data did not have every date available and the same number of sites for all states, instead, it either had more sites available for some states and fewer sites for others or fewer dates for some states with more dates for other states. For example, Arizona and New York had data for around 150 dates while California had around 90. Even with all uniform data, the factors influencing the data might have affected the research. Despite the shutdown orders, some essential businesses and places needed to stay open, (which meant that their pollution levels likely did not decrease) places like grocery shops, postal services, agriculture, essential travel, hospitals, and certain factories never closed or closed only for a brief period of time. For example, when the US had a shortage of ventilators, General Motors and Ford had their factories redirected to produce ventilators, which means the pandemic did not entirely stop all pollution coming from their factories (9).

However, even with all these factors, the research still shows a significant correlation between human activity and

CO levels. Other experiments and research also show a correlation between pollutants and pollution levels and human activity including a study done in one of the most polluted cities in the world, New Delhi, India (10–14). This study also found reduced levels of another pollutant, PM2.5 and these results indirectly show the adoption of social distancing. Further research on the effect of human activity on CO levels would further contribute and prove this correlation. In the US, more states and more dates are needed to make a more comprehensive and concrete analysis. Furthermore, since this research was done in the first few months of the shutdown, it would need to be done as the states start reopening to ensure the results are the same. Since only 15 states in the United States were analyzed, a future study consisting of more locations throughout the world would enhance the study. All these changes would further determine if human activity plays a role in CO levels. Future applications of research such as this include environmental analysis and climate change studies such as questioning if it is worth shutting down the economy to save the environment or researching how climate change has increased in past years.

MATERIALS AND METHODS

Data Sets

The data that we used in the experiment was obtained from the EPA's Air Quality public data available on its website (15). We then downloaded the CO levels among other pollutants such as SO_2, NO_2, Ozone, Pb, PM 2.5, and PM 10. The region for the data was the state, which included every site available in the state. The number of these sites depended on the state, with more populous states having more sites and less populous states having fewer sites.

Data Analysis

After the data was obtained, we edited the datasets in Microsoft Excel to make the analysis more understandable and cleaner. The original datasets included the specific dates such as 1/1/2020 but these dates were changed to the number of days since 2020, so 1/1/2020 would be Day 1 and 1/31/2020 would be Day 31. We did this for each site to make sure that the data was still accurate since the dates were essential to the hypothesis. After the above steps were done for every state, all the edited datasets were imported to R studio (Version 1.4.1106) to be further processed. The additional packages that we used in R were: ggplot2, dplyr, plotrix, and ggpubr. Then, we filtered the datasets were filtered using R since the original downloads had unnecessary information such as Site ID, County location, County ID. We took the mean of the daily CO values for each state to provide a more accurate representation. After having the data prepped for each state, the datasets were combined into groups of three to make it easier to graph, and then they were plotted as a scatter plot.

Statistical Analysis

The R^2 value was calculated using the lm function in R. The R^2 value quantified the extent of the total variation in the linear graph. The p-value was also calculated using the lm function and it quantified the variation in the linear relationship. To obtain the Pearson Correlation coefficient value for all the states, we used the cor.test function on the dataset that contained all the states. Correlation coefficient values for individual states were also obtained using the cor.test function, but only the specific states' dataset was used. States that have a higher negative value, for example -0.7, saw a greater decrease (they had a stronger and more direct correlation) in CO as the days progressed compared to states with a lower negative or a positive value, such as -0.2 or +0.1. We performed a one-way ANOVA test (aov function) followed by a Tukey's post-hoc test (TukeyHSD function) to look for differences of the mean CO levels in each state and correct for multiple comparisons when analyzing differences in significance between individual states. The code and relevant explanations are provided in the Appendix.

ACKNOWLEDGMENTS

I'd like to thank Spencer Eusden for all his help with the research, review, and publication of the manuscript.

Received: September 17, 2020
Accepted: October 25, 2021
Published: December 05, 2021

REFERENCES

1. "The Causes of Climate Change." NASA, climate.nasa.gov/causes/.
2. Kampa, Marilena, and Eliasz Castanas. "Human Health Effects of Air Pollution". *Proceedings of the 4th International Workshop on Biomonitoring of Atmospheric Pollution (With Emphasis on Trace Elements)*, vol. 151, no. 2, 23 July 2007. <www.sciencedirect.com/science/article/abs/pii/S0269749107002849#!>
3. Smith, William H. "Air Pollution—Effects on the Structure and Function of the Temperate Forest Ecosystem". *Environmental Pollution*, vol. 6, no. 2, Feb. 1974, pp. 111–129. <www.sciencedirect.com/science/article/abs/pii/0013932774900275>
4. Goudie, A. Human impact: man's role in environmental change. United States: N. p., 1982. Web.
5. Lin, C.-Y. Cynthia, *et al.* "The Effects of Driving Restrictions on Air Quality: São Paulo, Bogotá, Beijing, and Tianjin." 2011, ageconsearch.umn.edu/record/103381/.
6. Karl, T. R. "Modern Global Climate Change." *Science*, vol. 302, no. 5651, 2003, pp. 1719–1723., doi:10.1126/science.1090228.
7. Li, Xiangdong, *et al.* "Air Pollution: a Global Problem Needs Local Fixes." *Nature*, vol. 570, no. 7762, 25 June 2019, pp. 437–439., doi:10.1038/d41586-019-01960-7.
8. Ernst, Armin, and Joseph D. Zibrak. "Carbon Monoxide Poisoning." *The New England Journal of*

Medicine, 26 Nov. 1998, www.nejm.org/doi/full/10.1056/NEJM199811263392206
9. Reed Albergotti and Faiz Siddiqui. "Ford and GM Are Undertaking a Warlike Effort to Produce Ventilators. It May Fall Short and Come Too Late." *The Washington Post*, www.washingtonpost.com/business/2020/04/04/ventilators-coronavirus-ford-gm/.
10. Bao, Rui, and Zhang, Acheng . "Does Lockdown Reduce Air Pollution? Evidence from 44 Cities in Northern China." *Science of The Total Environment*, vol. 731, 2020, p. 139052., doi:10.1016/j.scitotenv.2020.139052.
11. Wang, Qiang, and Min Su. "A Preliminary Assessment of the Impact of COVID-19 on Environment – A Case Study of China." *Science of The Total Environment*, vol. 728, 22 Apr. 2020, p. 138915., doi:10.1016/j.scitotenv.2020.138915.
12. Myllyvirta, Lauri. "Analysis: Coronavirus Temporarily Reduced China's CO2 Emissions by a Quarter." *Carbon Brief*, 26 June 2020, www.carbonbrief.org/analysis-coronavirus-has-temporarily-reduced-chinas-co2-emissions-by-a-quarter.
13. Chauhan, Akshansha, and Ramesh P. Singh. "Decline in PM2.5 Concentrations over Major Cities around the World Associated with COVID-19." *Environmental Research*, vol. 187, 5 May 2020, doi:10.1016/j.envres.2020.109634.
14. Kerimray, Aiymgul, et al. "Assessing Air Quality Changes in Large Cities during COVID-19 Lockdowns: The Impacts of Traffic-Free Urban Conditions in Almaty, Kazakhstan." *Science of The Total Environment*, vol. 730, 4 May 2020, doi:10.1016/j.scitotenv.2020.139179.
15. Environmental Protection Agency. (n.d.). EPA. https://www.epa.gov/outdoor-air-quality-data/download-daily-data.