The Spread of COVID-19 in the 100 Largest US Cities

Introduction

It is not news to anyone that the spread of COVID-19 has been devastating to much of the world. Globally, more than 1.5 million people have died of COVID-19; almost 280,000 of those deaths have happened in the United States.[1] But how did we get here? What conditions allowed a pandemic that broke out in central China to cause 3,800 deaths in Wisconsin within a year?[2] The answer is complex; there are myriad factors that have contributed to the spread of COVID-19. In order to better understand a piece of the puzzle, this paper focuses on one such broad question: How does climate affect the spread of COVID-19?

It isn't hard to imagine that climate could be related to the spread of COVID-19. Warmer weather leads to people spending more time outside and less time indoors, which could lead to fewer new COVID-19 cases. Colder weather might allow the virus to survive for longer periods on surfaces. Small particulate matter in the air could be an avenue for the virus to travel from person to person. More wind throughout the day could lead to less localized air stagnancy and therefore less likelihood of an individual coming into contact the virus. These are only a few of the many ways weather could impact the spread of COVID-19. Indeed, the relationship between climate and the spread of COVID-19 has been demonstrated in multiple papers published since the beginning of the pandemic.

Zhang et al (2020) find an inverse relationship between average temperature and coronavirus transmission and that air pollution indicators are positively correlated with new confirmed cases. Muhammad et al (2020) conclude that average temperature as well as maximum and minimum temperature are associated with the transmission of COVID-19. Many other researchers have approached this problem and come to similar conclusions.[3]

The relationship between climate and the spread of COVID-19 across the entirety of the United States has not been fully explored in the extant research. While there have been some country-wide studies on the United States,[4] most have been focused either internationally or on subregions within the United States. Further, those studies that did consider the United States as a whole were written early in the pandemic, limiting the amount and quality of data to which they had access. We expand on the extant literature by considering cities across the United States and by using more extensive data than previously available. Hence, our refined statistical question is: do changes in ambient temperature (maximum, minimum, average, and average squared), air-quality, wind speed, and relative humidity have a significant relationship on the spread of COVID-19 in the 100 largest cities in the US when controlling for state fixed effects, population density, and time?

Our findings largely support those that have been already been published. We find that average, minimum, and maximum temperature, average relative humidity, and average wind speed are inversely related to the spread of COVID-19 while average temperature squared is directly related to the spread. Additionally, we find evidence that the density of small particulate matter is positively related to the spread of COVID-19, although our models yield inconsistent results.

Data

---

[1] As of 12/5/20, see "The Coronavirus Outbreak," *New York Times*, https://1drv.ms/w/s!AkFlokYzR3chgncORgD1O32m2Drs?e=QdKQAH

[2] As of 12/5/20, see "The Coronavirus Outbreak," *New York Times*, https://www.nytimes.com/interactive/2020/us/wisconsin-coronavirus-cases.html

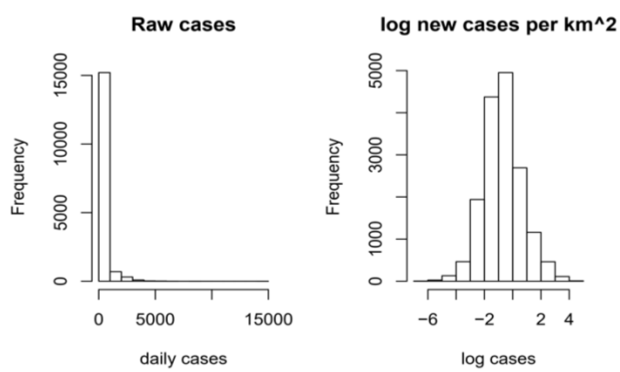
[3] For a comprehensive list of available literature, see "What environmental data are relevant to the study of infectious diseases like COVID-19?," *Climate.gov*, https://www.climate.gov/covid19

[4] See, e.g., Behnood (2020), Chen (2020), Manzak (2020)

Our base unit of measurement for this paper is the 100 largest cities in the United States. These cities span 31 states and represent a comprehensive geographic cross-section of the United States, allowing for significant climatic variation. The list of the 100 largest cities in the United States pulls from the United States Census Bureau. The cities are ordered based on their estimated population as of July 1, 2019. We include cities in the model starting on the day in which they reach 100 daily new cases. Our regression data extends to 10/31/2020.

Our data on the spread of COVID-19 comes from the *New York Times*. The newspaper has engaged in a comprehensive effort to compile data on COVID-19 cases, deaths, and other variables from state and local government agencies into easy-to-use datasets. The paper began this effort in January and has become one of the definitive sources of data on COVID-19. It is relied on by many researchers and companies, including Google. One major drawback of this data is it is reported at the county level in most cases, not the city level. This means that by using this data we are doing a bit of an apples to oranges comparison, as our other analysis is calculated based on the city. We expect this effect to be minimal as these cities represent the bulk of their counties and there is no indication that any bias would be systematic. In our models, we use the use the log of daily new cases per square kilometer as the dependent variable. As **Figure A** shows, the log transformation of the dependent variable is used because the raw numbers of new cases exhibit a non-normal distribution. Some researchers have used the number of new deaths as the dependent variable, we decline to use deaths as the dependent variable because the impact mechanism is less well established. A death from COVID-19 could come days, weeks, or months after the individual initially contacted the virus, so it isn't clear what time period to use to tie climate to deaths. Further, deaths are a function of more than just the rate of COVID-19 spread; we expect that the underlying health of the population of a city and how many treatments had been developed to have large effects on the number of deaths.

Figure A: Raw and Log Values of New COVID-19 Cases per Day by City



We gather data on the air quality index ("AQI") from Airnow.gov. The site is run by the federal government as a central clearinghouse for air quality data. It combines data from federal agencies like the EPA and NOAA as well as tribal, state, and local agencies. We use a standard metric called PPM 2.5, which is the number of particulates smaller than 2.5 microns per cubic meter of air. A higher PPM 2.5 score means there is more particulate matter in the air, meaning there is a lower air quality. While the standard has evolved over time, the EPA currently considers a PPM 2.5 higher than 12 to be unhealthy.[5] To assign AQI data to each city, we match each city with the closest station (based on the latitude and longitude) that has available data.

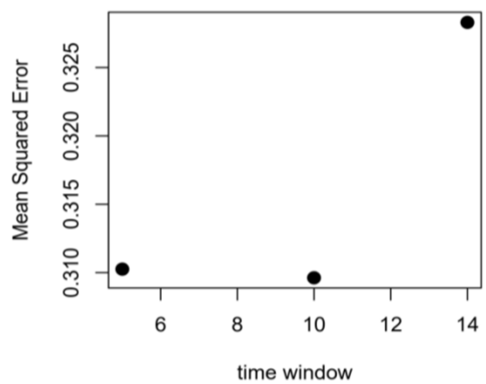
The rest of the weather data comes from the Iowa Environment Mesonet ("Mesonet"), which contains data from Automated Surface Observing Stations ("ASOS") which are located at airports in the United States. These observations are reported at various intervals, but some as frequently as every five minutes. We access this

[5] "What are the Air Quality Standards for PM?," *EPA*, https://www3.epa.gov/region1/airquality/pm-aq-standards.html

data through the “riem” package in R. We match each city with the closest ASOS, which is a different station than the one used for AQI data.

Our data processing consisted of downloading the weather data for the appropriate station for each city, taking the log of the number of new COVID-19 cases, and calculating the maximum/minimum/average of the weather variables. All weather variables (maximum, minimum, and average) are measured over 10-day rolling windows. Pedrosa (2020) uses the 10-day window but also tests the 12-day and 15-day windows, finding that there was no relevant impact on the results of the models. Zhang et al (2020) don’t specify the window over which they average temperature but use 1-, 3-, and 5-day windows for their AQI averages. Gupta et al (2020) consider both 10-day and monthly averages. We tested our model on 5- and 14- day windows by cross validation and found that that 10-day window yielded the lowest mean squared error.

Figure B: Average MSE in cross validation for models on 5-, 10-, 14- day windows.



Results and Discussion

We analyze the impact of climate factors on the spread of COVID-19 using two linear regression models. Both models express the log of daily new cases per square kilometer as a function of a basket of climate variables and three control variables: state fixed effects, population density, and log of lagged cases. **Figure C** below shows the results of estimating the two regression models.

Figure C: Effects of Climate Variables on Log of Daily New Cases per Square Kilometer.

Variables	(1)	(2)
Average temperature		-0.02179***
Maximum temperature	-0.00446 ***	
Minimum temperature	-0.00268**	
Average temperature-Squared		0.00013***
Average AQI	-0.00035	0.00501*
Average AQI*Average temperature		-0.00008*
Average wind speed	-0.00795**	-0.01432 ***
Average relative humidity	-0.00220***	-0.00152*
State	0.08745-1.14249***	0.18878- 1.32620***
Population density	0.00001*	0.0002***
Log of Lagged Cases	0.95676***	0.94195***

Adjusted R^2	0.8393	0.834

*** Significant at the 1% level ** Significant at the 5% level * Significant at the 10% level

For both models, we include the same control variables. The first control variable is a vector of fixed effects for each state. The lack of a coordinated response to the pandemic by the federal government meant that the decision of which, if any, mitigation practices to enact was left largely to the state governments. Some states rapidly adopted stringent mitigation efforts like mask mandates, active messaging around how to maintain social distance, and the closing of businesses that were linked to the spread of COVID-19. Other states were late to, or never, enacted mitigation policies. An analysis by the *New York Times* demonstrated a strong link between the stringency of containment measures and the rate of COVID-19 infection.[6] We attempt to control for this heterogeneity by including a dummy variable for each state. An interpretation of the various coefficients estimated for the states is outside of the scope of this paper and is left to future researchers.

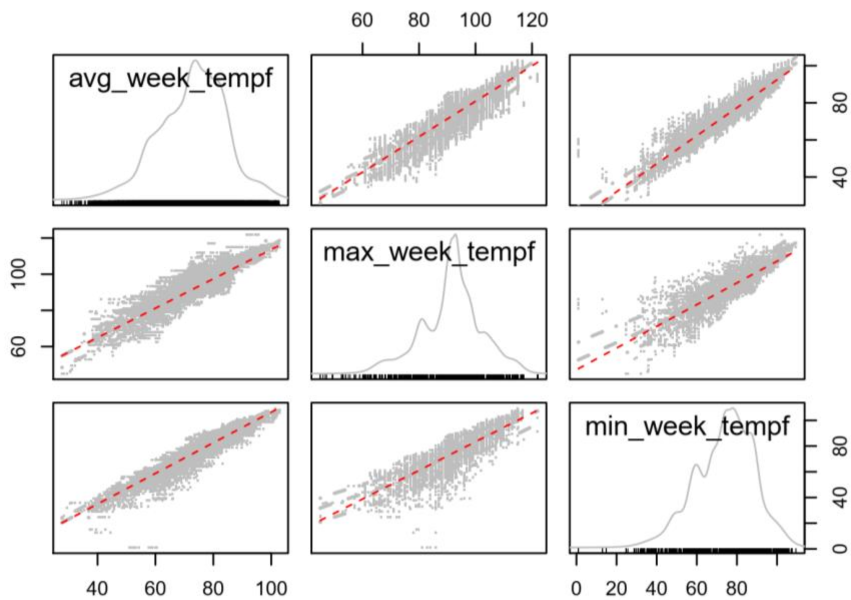
We also include the population density per square kilometer as an explanatory variable to control for the expected positive relationship between population density and the spread of COVID-19. As expected, we estimate a highly significant, positive coefficient for population density. As the population density increases, so too does the spread of COVID-19.

Finally, we include the log of average number of new cases per day within the previous ten days. As the rate of infection in an area increases, we expect that, absent intervention, the rate of spread in the future will likely be higher. Similarly, if there has been very little spread of COVID-19 in a city over the past few days, it is unlikely that there will be a large spike on any given day. The log of cases over the previous ten days is effective at controlling for this momentum. We estimate a statistically significant positive coefficient for this variable that is slightly less than one in both models.

The difference between the two models is which temperature variables are considered. Model (1) includes the maximum and minimum temperature. On the other hand, Model (2) includes the average temperature and the square of the average temperature. We have estimated these models separately because the temperature values are highly correlated as shown in **Figure D.** While the models are estimated separately, the estimated coefficients can be interpreted together. For both average temperature and maximum temperature, we estimate a negative relationship with the spread of COVID-19; for average temperature, the estimated coefficient is highly statistically significant. The estimated coefficient of -.02179 for average temperature means that for one degree increase in average temperature over the prior 10 days, we estimate a 2.179% decrease in the number of new COVID-19 cases per square kilometer. Taken together, these estimated coefficients support our prior belief, and the findings of existing literature, that new COVID-19 cases decrease as the temperature increases but the effect is lessened at extreme temperatures. For minimum temperature, we estimate a highly significant negative coefficient. This matches our expectations as, intuitively, a lower minimum temperature should lead to people spending less time outdoors and therefore having a higher risk of infection. Finally, we estimate a significant, positive coefficient on the square of average temperature. Taken in conjunction with the negative coefficient on average temperature we conclude that, as the average temperature increases, the dampening effect on the spread of COVID-19 is lessened. Intuitively this result makes sense. A change in temperature from 45 degrees to 55 degrees is likely to have a different effect on your willingness to go outside than a change from 85 degrees to 95 degrees.

[6] "States That Imposed Few Restrictions Now Have the Worst Outbreaks," *New York Times*, November 18, 2020.

Figure D: Correlations in Minimum, Maximum, and Average Temperature.



Both models include average AQI; in Model (1) we estimate an insignificant negative coefficient and in Model (2) we estimate a weakly significant positive coefficient. The inconsistency of the estimated coefficients means that estimated effect of AQI on the spread of COVID-19 is dependent on the functional form of your model. Given that previous researchers have found a significant, direct relationship between AQI and the spread of COVID-19, we believe that this is an important topic for future research. Further, in Model (2), we include the interaction between average AQI and the average temperature. The weakly statistically significant negative coefficient estimated for this variable indicates that changes in AQI have less of an impact on new COVID-19 cases at higher temperatures.

We include two more weather variables for which we estimate statistically significant negative coefficients: average wind speed and average relative humidity. We expect that average wind speed should be inversely related to the spread of COVID-19 as an increase in wind movement should keep droplets containing COVID-19 from lingering. For relative humidity, the mechanism is less clear. Prior researchers have also found an inverse relationship, but there is not a consensus on why this should be the case.

We validate our choice of variables using a forward-selection method. Overall, both models do well in accounting for the variation in the spread of COVID-19. Each of the models has an adjusted-R^2 value of about .83, meaning that the models each explain about 83% of the variation in changes in COVID-19 cases. While in this paper we are more interested in mechanisms of impact rather than the predictive ability of the model, the high adjusted- R^2 gives us comfort that we have not missed a major control component when building the model.

Limitations and Future Work

A limitation of this analysis is the use of state fixed effects to control for differences in mitigation efforts. While it is true that many policies are set at the state level, some policies are decided on by cities. So, to the extent that there are multiple cities from the same state, we might not be including enough specificity. Further, by using state fixed effects instead of explicit data on mitigation practices, we may be confounding the effects of the other explanatory variables. Additionally, the inconsistent coefficients that we estimate for average AQI cast some doubt on the conclusions drawn from this model. Given that other researchers have found a stronger impact of AQI, we will need to investigate this result more closely before fully relying on the results of the model.

A clear path for future research is to extend these analyses with additional data. As noted earlier, the rate of spread of COVID-19 has been worse over the past two months than at any time earlier in the pandemic. How much of this worsening situation can be explained by the decreasing temperatures? Future researchers could also use the conclusions drawn from this study to implement real world policy decisions and measure their effectiveness. Recommendations for future policy decisions are covered in the next section.

Conclusions and Implications

The results of the analyses described in this paper represent an important contribution to the literature on climate and the spread of COVID-19. We find that average, minimum, and maximum temperature, average wind speed, and average relative humidity are inversely related to the spread of COVID-19 while average temperature squared, and average AQI all exhibit a significant direct relationship with the spread of COVID-19. The magnitude of these results is also of merit. By far, the largest coefficient we estimate is for average temperature. Our model estimates that one degree increase in average temperature leads to a more than two percent decrease in new COVID-19 cases. This means that a ten degree increase in average temperature, not an unusual thing to happen, would lead to a more than 20% decrease in new COVID-19 cases. That is a large impact that should not be overlooked. While not novel, these findings are important because they confirm the existing understanding on a larger and more comprehensive dataset than has previously been used.

Our findings, and the wealth of research on which they draw, have clear policy implications for both the current pandemic and future viral spreads of disease. Policy makers should take into account the weather when creating mitigation efforts. By relaxing policies during lower risk (especially higher temperature) periods, policy makers may be able to alleviate some of the quarantine fatigue felt by many people over the past nine months. This could increase compliance with mitigation measures during high-risk periods. Additionally, these findings give one more reason to address the worsening air quality problem in many of the world's major cities. Even if long term solutions to air quality problems aren't politically feasible, law makers could enact temporary measures (like localized factory shut-downs or traffic bans) in times of viral disease spread.

Rubric Matching List

Item 1: Page 1 to 6

Item 2: Page 6 Lines 26-Page 7, Lines 10

Item 3: Page 9

Item 4: Page 1 to 6

Item 5: Page 1 to 6

Item 6: Page 1, Lines 3-8

Item 7: Page 1, Lines 7-8

Item 8: Page 1, Lines 9-30

Item 9: Page 1, Lines 9-30

Item 10: Page 1, Lines 27-30

Item 11: Page 1, Lines 36-Page 3, Lines 12

Item 12: Page 1, Lines 36-Page 2, Lines 12

Item 13: Page 3, Lines 3-10

Item 14: Page 3, Lines 14-Page 5, Lines 20

Item 15: Page 4, Lines 2-Page 5, Lines 20

Item 16: Page 4, Lines 2-9; Page 5, Lines 16-20

Item 17: Page 5, Lines 16-20

Item 18: Page 3, Lines 14-Page 5, Lines 20

Item 19: Page 5, Lines 3-5, 26-28; Page 6, Lines 7-16

Item 20: Page 6, Lines 7-24

Item 21: Page 6, Lines 13-16

Item 22: Page 6, Lines 7-16

Item 23: Page 4, Lines 27-38

Item 24: Page 3, Lines 16-Page 4, Lines 1; Page 4, Lines 24-38

Item 25: Page 5, Lines 21-28; Page 6, Lines 8-14

References

Abouk, Rahi and Babak Heydari, “The Immediate Effect of COVID-19 Policies on Social Distancing Behavior in the United States,” April 8, 2020, available at SSRN: https://ssrn.com/abstract=3571421

Behnood et al. “Determinants of the infection rate of the COVID-19 in the U.S. using ANFIS and virus optimization algorithm (VOA),” *Chaos, Solitons & Fractals*, vol. 139, October 2020, 110051.

Chen, Lung-Wen and Lung-Chang Chien, “Meteorological impacts on the incidence of COVID-19 in the U.S.,” *Stochastic Environmental Research and Risk Assessment*, vol.34, October 2020.

Gupta et al., “Effect of weather on COVID-19 spread in the US: A Prediction model for India in 2020,” *Science of the Total Environment*, vol.728, 1 August 2020, 138860.

Iowa Mesonet Data, *Iowa State University*, accessed through “riem” package in R.

“List of United States cities by population,” *Wikipedia,* accessed November 1, 2020.

Manzak, Ali and Dilek Manzak, “Analysis of Environmental, Economic, and Demographic Factors Affecting COVID-19 Transmission and Associated Deaths in the U.S.A,” July 6, 2020, available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3644677

Mehmet Sahin, “Impact of weather on COVID-19 pandemic in Turkey,” *Science of Total Environment*, vol.728, August 1, 2020, 138810.

Muhammad Farhan Bashir, “Correlation between climate indicators and COVID-19 pandemic in New York, USA,” *Science of Total Environment*, vol.728, August 1, 2020, 138835.

“NYT COVID-19 Data,” *New York Times*, accessed from https://github.com/nytimes/covid-19-data.

Pedrosa, Renato H. L., “The dynamics of Covid-19: weather demographics and infection timeline,” *State University of Campinas*, May 10, 2020, (Not Peer Reviewed).

Spencer et al., “Environmental Weather Conditions and Influence on Transmission of SARS-CoV-2," *The Centre for Evidence-Based Medicine,* July 3, 2020.

"States That Imposed Few Restrictions Now Have the Worst Outbreaks," *New York Times*, November 18, 2020.

"What are the Air Quality Standards for PM?," *EPA*, https://www3.epa.gov/region1/airquality/pm-aq-standards.html

"Which states had the best pandemic response," *Politico*, October 14, 2020.

Zhenbo Zhang et al., "Effects of meteorological conditions and air pollution on COVID-19 transmission: Evidence from 219 Chinese cities," *Science of Total Environment*, vol.741, November 1, 2020, 140244.