# Viruses and the Weather: a Correlation Study Between Different Weather Variables and COVID-19 Cases in Sweden

## First Year Project

FRIDAY, 4TH MARCH, 2022

**Authors:**

Altea Fogh - alfo@itu.dk

Justė Dilytė - judi@itu.dk

Krzysztof Parocki - krpa@itu.dk

Liam O'Flaherty - liof@itu.dk

Raquel Marques - raqm@itu.dk

IT UNIVERSITY OF COPENHAGEN

# Introduction

At the end of 2019, the first cases of the novel COVID-19 virus (SARS-CoV-2) were detected in Wuhan, China. Not long after, the first cases started appearing in other countries around the world, including Sweden. In March, The World Health Organization (WHO) declared it to be a worldwide pandemic. COVID-19, being a novel virus, is especially dangerous due to the lack of information about it. It is hard to predict how the virus will spread and how to respond to it, without knowing more about the virus itself. Thus, ever since the first detected case of the virus, researchers have been actively trying to find out what factors play a role in the spread of COVID-19.

Some of the previous studies have touched upon the environmental factors and their impact on the spread of COVID-19. In fact, the regression models obtained in a study by Kadi and Khelfaoui (2020) found a strong correlation between the population density and the number of COVID-19 cases. Different weather factors, especially the temperature and the humidity, were also found to be influencing the spread of the virus (McClymont, 2021). However, it has also been suggested that the correlation might not be a direct one, but influenced by a third variable. For example, even though a negative correlation has been found between the precipitation level and the COVID-19 cases, it might be the social behavior influenced by the precipitation that affects the amount of cases, rather than the precipitation itself (Menebo, 2020).

It is clear that knowing the contributing factors in the spread of COVID-19 is vital for the proper countermeasures to be put in place. Since weather is continuously confirmed to be one of those factors, it is important to analyze the situation on a geographically smaller scale. For example, in regards to a single country. Thus, to analyze the subject further, we asked the following question: *"Which weather variable has the most impact on the* COVID-*19 cases in Sweden? And how does the impact vary depending on the stringency index?".*

# Data

To explore our research question, we used the latest data set of all infections in the year 2020 provided by the Swedish government, and weather data for the country. The full dataset included different weather variables and COVID-19 cases by date and by region. *Relative surface humidity*, *solar radiation*, *surface pressure*, *temperature above ground*, *total precipitation*, *UV index* and *wind speed* were used as the weather variables.

In addition, we collected the stringency index through the official Oxford website. The data is based on nine response indicators, including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 to record the strictness of the government's policies.
All datasets were checked for *Not Available* (NA) values, and these were filtered out before using the datasets further. We then merged our individual datasets into one master data frame to allow for easier access and manipulation.

# Results

Firstly we plotted the cases aggregated by month, to get an overview of the behavior of COVID-19 cases in each of the Swedish regions. This plot shows clearly a correlation between the cases and the regions, with a peak in December 2020.

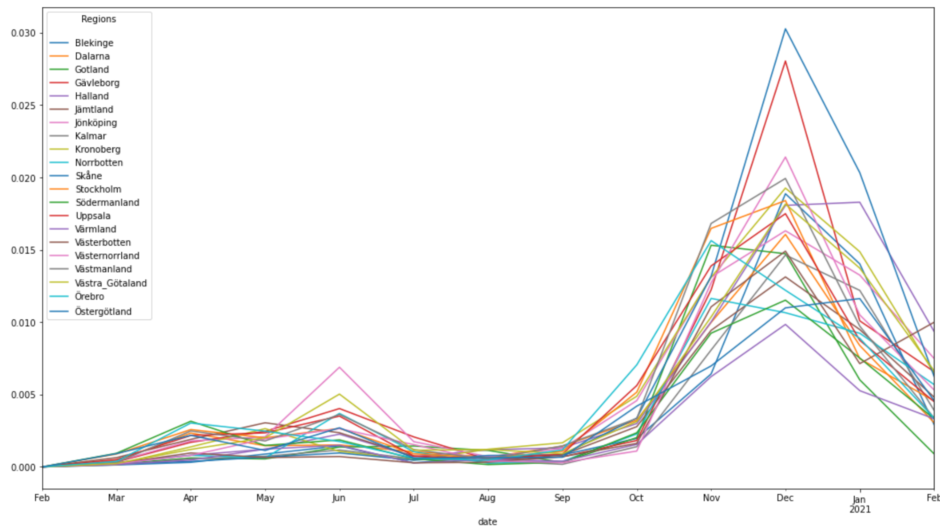


Figure 1 - Corona cases in different Swedish regions over time

After comparing each weather variable to cases per capita, we found what seemed to be a positive correlation with surface humidity, as illustrated in Figure 2.

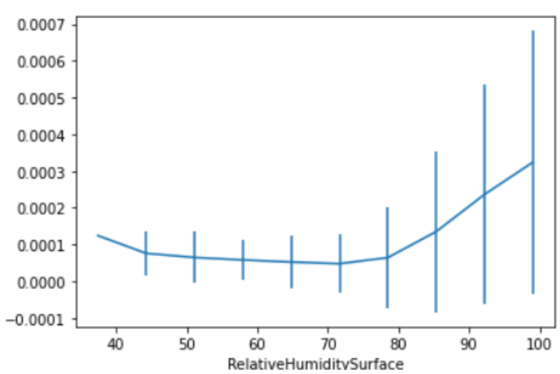


Figure 2 - Surface humidity in relation to the COVID-19 cases per capita

We then ran a multivariate regression analysis using both Pearson linear and log transformations, in order to try to authenticate this correlation.

The regression analysis returned an R-squared value of 0.199 and 0.2 for the linear and log respectively, and humidity returned a p-value of 0.00 on both. This lent further weight to our hypothesis that humidity could have a correlationary effect on COVID-19 cases per capita. Both of these analyses can be seen in Table 1.

| | Regression analysis between COVID-19 cases per capita and the weather variables (linear) | | Regression analysis between COVID-19 cases per capita and the weather variables (logarithmic) | | Regression analysis between COVID-19 cases per capita and the weather variables, together with the stringency index (linear) | | Regression analysis between COVID-19 cases per capita and the weather variables, together with the stringency index (logarithmic) | |
|---|---|---|---|---|---|---|---|---|
| R-squared | 0.199 | | 0.2 | | 0.317 | | 0.317 | |
| Adj. R-squared | 0.199 | | 0.199 | | 0.316 | | 0.316 | |
| | coef | p | coef | p | coef | p | coef | p |
| Relative Humidity Surface | 4.761e-06 | 0.000 | 4.761e-06 | 0.000 | 4.377e-06 | 0.000 | 4.375e-06 | 0.000 |
| Solar Radiation | 3.945e-13 | 0.722 | 3.945e-13 | 0.722 | 2.315e-12 | 0.024 | 2.313e-12 | 0.024 |
| Surface Pressure | 6.353e-10 | 0.000 | 6.353e-10 | 0.000 | 2.951e-10 | 0.000 | 2.95e-10 | 0.000 |
| Temperature Above Ground | -5.415e-06 | 0.000 | -5.415e-06 | 0.000 | -2.351e-06 | 0.000 | -2.35e-06 | 0.000 |
| Total Precipitation | -0.0031 | 0.001 | -0.0031 | 0.001 | -0.0038 | 0.000 | -0.0038 | 0.000 |
| UV Index | -2.537e-06 | 0.000 | -2.537e-06 | 0.000 | -5.745e-06 | 0.000 | -5.743e-06 | 0.000 |
| Wind Speed | -2.733e-06 | 0.154 | -2.733e-06 | 0.154 | 6.038e-06 | 0.001 | 6.035e-06 | 0.001 |
| Stringency Index | - | - | - | - | 5.993e-06 | 0.000 | 5.991e-06 | 0.000 |

Table 1 - The results from the regression analysis

In order to ensure that cases per capita was a better variable to use in our research than cases, we decided to plot four sets of data on a map of Sweden - population, population density, cases, and cases per capita, all in linear scale. Figure 3 illustrates the impact population and population density might have on cases - the maps look very similar. However, they are not exactly alike, which could mean that something more complex occurs here. Cases per capita map, in turn, looks completely different. The cases are more evenly distributed, as the variable inherently allows for population - however, it doesn't allow for population density. Nevertheless, we still managed to denoise our data considerably by choosing cases per capita.

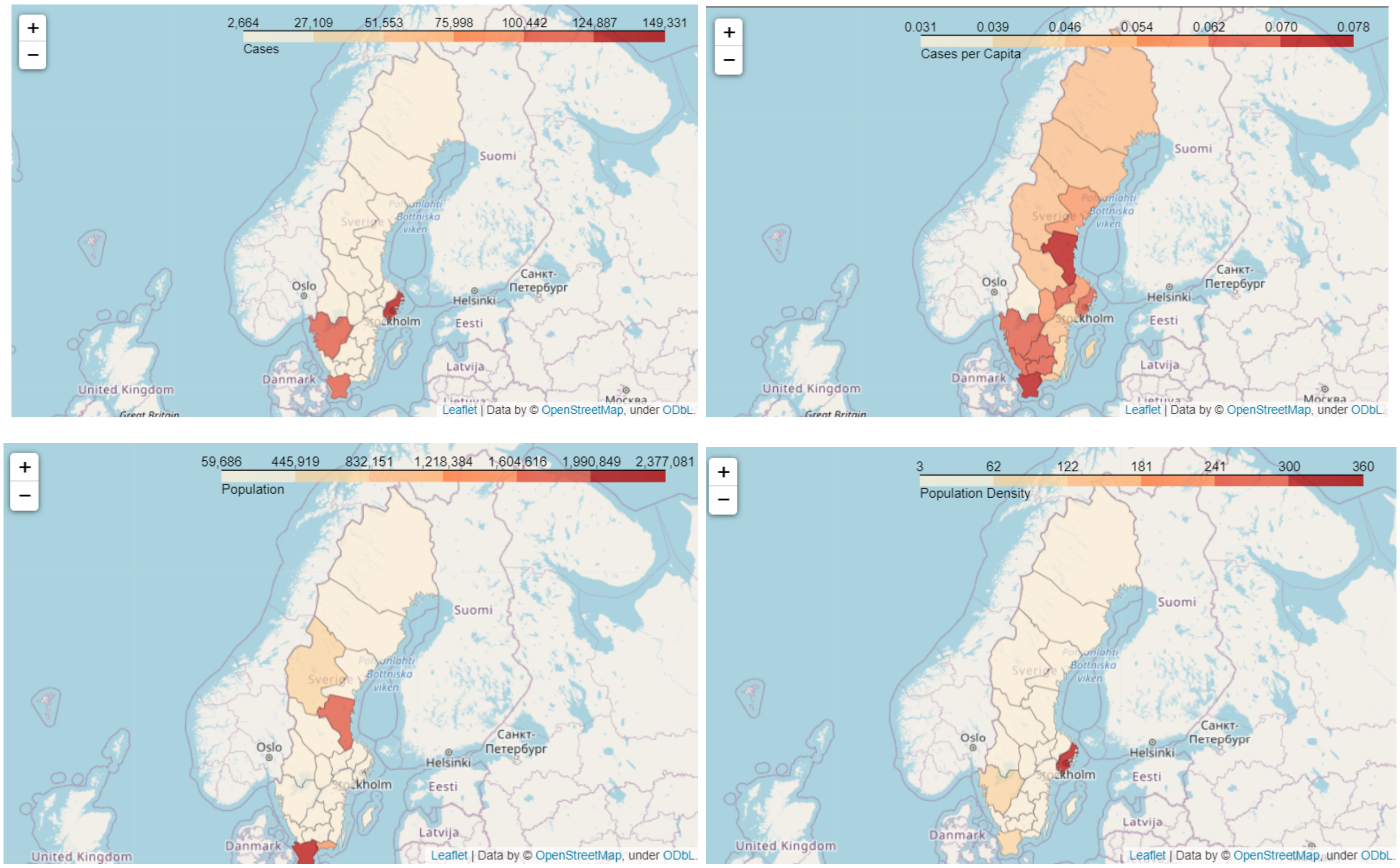


Figure 3: Merged map with the cases of COVID-19, cases per capita, population and population density (linear scale)

To answer the second part of our research question, we added the stringency index to our regression analysis to measure whether or not this index had any impact on both our R-squared value as a whole, and on humidity in general. The results are shown in Table 1. The low p-value (0.00) of the stringency index indicates a significant relationship with the COVID-19 cases per capita, although the relationship, indicated by the coefficient, is not strong. However, the increased R-squared and adjusted R-squared respectively show that the weather, together with the stringency index, represent the COVID-19 effects more accurately than the weather variables on their own.

# Discussion

Through the regression analysis, we found a small, but statistically significant positive correlation between the *Relative Humidity Surface* index and COVID-19 cases per capita. This correlation was, as previously stated, very small in number (4.76*e-06). Therefore, even though the correlation was found, it cannot be considered highly impactful. It's also worth noting that this might be just a correlation - our results don't point unambiguously to a relationship of causation.

Moreover, although the relative humidity surface turned out to be the weather variable with the highest coefficient, there are also other variables showing a very similar correlation. Temperature above ground and total precipitation provided very similar behaviors, but with even smaller potential impact, and were thus not chosen to be the most relevant ones.

Another factor supporting our variable choice is the stringency index, which increased R-squared value while maintaining the coefficient of humidity. This hints that the humidity impact was not influenced by a third variable (in this case, stringency index), but it was indeed having an effect.

# Limitations

It is relevant to underline a few key factors that are limiting the accuracy of this research.

First and foremost, it is necessary to remind the reader that any of the weather variables and the COVID-19 cases do not necessarily have a direct relation of causation. I.e., even if the correlation is found, it does not necessarily imply the causation.

In addition, a correlation might also be influenced by third variables, and we couldn't have accounted for all of them in our research. For example, social factors influenced by the weather or different COVID-19 variants could potentially be indirectly influencing the findings (Katella, 2022).

Notably, we did not account for the delay in efficacy of the government measures, nor the possible delay between the weather's impact and the COVID-19 cases. This is a factor that would be relevant in a future iteration of this research. Moreover, the data collection that was used for this research provided a collection of COVID-19 cases per region per date, but did not account for false positives. This implies that some of the cases could actually be not real and thus not impacted by the weather, but by the reliability of the testing institution.

# Conclusion

This report is relevant for the government, as it provides a proof of correlation with regard to the impact of restrictions on COVID-19 cases per capita. However, we haven't formulated any guidelines for the government related to the weather, as the impact was either too vague or too insignificant to be taken into account as an important factor when applying new policies. The research might have produced different results if we took another approach to measuring COVID-19 cases, for example a percentage change of cases compared to previous days. In the future, it could also be useful to account for more third variables apart from stringency index, such as age distribution. The analysis might be therefore extended in various directions.

## Disclosure statement

This report and research was done as a collaboration of five students. Coding was done mainly collectively or in pairs, for which each person contributed in the most relevant way in relation to their strengths. From the Git-log provided, it will be possible to see how only some people pushed and pulled, due to issues with system compatibility and the machine that was used in the pair coding.

## References

Harvard Medical School (2022, January 26). If you've been exposed to the coronavirus. *Harvard Health Publishing, Harvard Medical School*. https://www.health.harvard.edu/diseases-and-conditions/if-youve-been-exposed-to-the-coronavirus

Kadi, N., Khelfaoui, M. (2020). Population density, a factor in the spread of COVID-19 in Algeria: statistic study. *Bull Natl Res Cent 44, 138*. https://doi.org/10.1186/s42269-020-00393-x

Katella, K. (2022, February 28). Omicron, Delta, Alpha, and More: What To Know About the *Coronavirus Variants. Yale Medicine*. https://www.yalemedicine.org/news/covid-19-variants-of-concern-omicron

McClymont, H. (2021). Weather Variability and COVID-19 Transmission: A Review of Recent Research. *Int. J. Environ. Res. Public Health 18(2), 396*. https://doi.org/10.3390/ijerph18020396

Menebo, M. M. (2020). Temperature and precipitation associated with Covid-19 new daily cases: A correlation study between weather and Covid-19 pandemic in Oslo, Norway. *Science of The Total Environment, 737*. https://doi.org/10.1016/j.scitotenv.2020.139659

University of Oxford (2022, December). *COVID-19 Government Response Tracker.* Blavatnik School of Government, University of Oxford. https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker