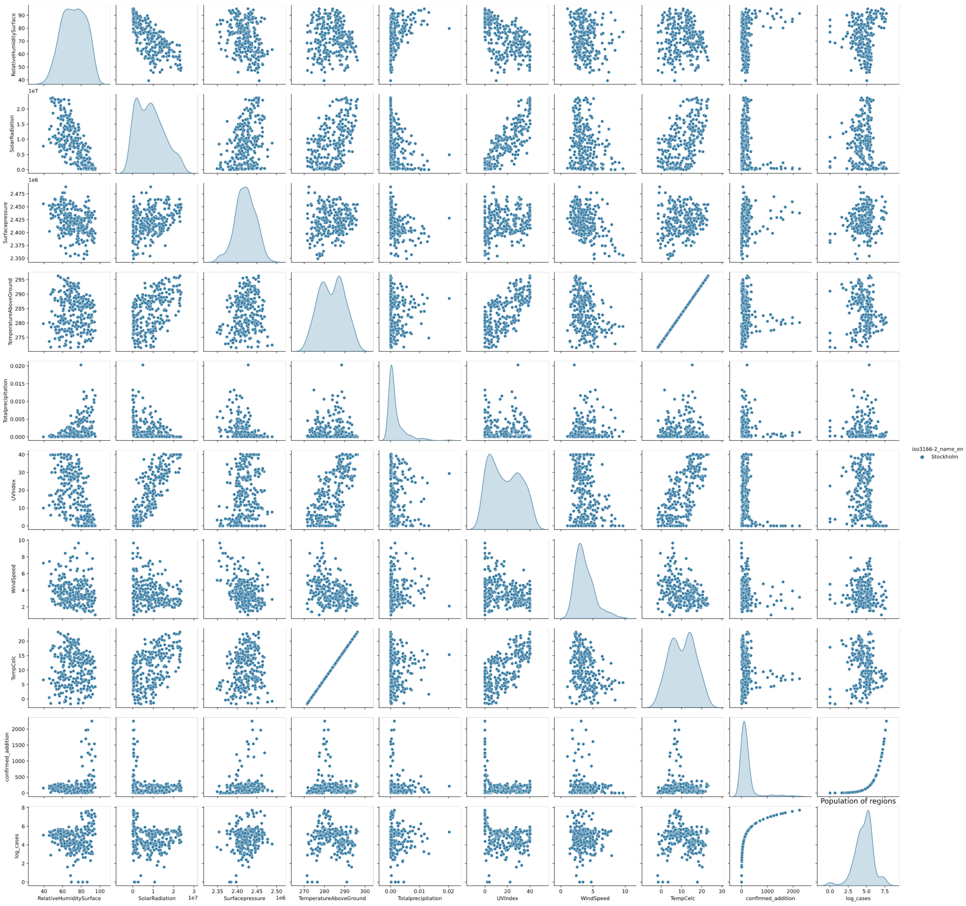
COVID-19 outbreak and Weather in Sweden for 2020

Project 2 Report

Ada Matilde Gige (adgi@itu.dk)
Moneeca Abru Iftikhar Latif (abml@itu.dk)
Sabrina Fonseca Pereira (sabf@itu.dk)
Sofie Bengaard Pedersen (besp@itu.dk)
Yasmin Sarkhosh (yasa@itu.dk)



ORDINARY EXAM
FIRST YEAR PROJECT

Department of Computer Science, IT University
Copenhagen, Denmark

Date: March 19 2021

1 Introduction

Influenza or common cold virus' are widely related to seasonal weather changes. Incidents increase during colder months [1]. Factors such as physical components affecting the durability of the virus and environment of social interaction, relating to people staying inside in less ventilated and more crowded areas. These factors are speculated to apply for COVID-19 too. However the seasonality of the virus has not yet been strongly validated. Therefore, we want to investigate whether there is a statistical association between different weather conditions and confirmed COVID-19 cases in Sweden:

"How is the relationship between the intensity of the COVID-19 outbreak and the weather conditions for Sweden in 2020?"

This study will also look into the age and gender of the COVID-19 confirmed cases, the cases who was admitted to intensive care and the people who died with COVID-19. Exploring external factors, that might affect the rate of spread could help the Swedish government to contain the virus more effectively in the future, that both will be economically and socially beneficial.

2 Data

The data analysed includes 4 supplied datasets; (1) the daily sum of the number of confirmed COVID-19 cases by region, (2) daily weather conditions for several countries which was filtered to Sweden only, (3) Sweden's metadata which includes the population, region codes and the names of each region and (4) geographic location information with coordinates for mapping the regions of Sweden. 1, 2 and 3 were combined into one dataset which was used throughout the analysis. Merging caused an expected loss of data that was necessary for the analysis, as we need both COVID-19 cases and weather data to conduct the analysis. The final dataset included observations from February to November 2020 in Sweden.

External datasets were included for investigating the relationship between COVID-19 cases and demographics. They included; (5) the weekly sum of confirmed cases, admissions to intensive care and deaths, (6) the total count of cases , admissions and deceased for different age groups and (7) the same but divided by gender. [2] All data sets were inspected for missing data, there were no missing values.

3 Results and discussion

3.1 COVID-19 confirmed cases and weather conditions

We explored potential association between weather and confirmed COVID-19 cases by running Person's and Spearman's correlation tests on all weather variables against the total number of cases. We also ran Pearson's correlation tests after log transforming the cases variable to

[1] https://en.wikipedia.org/wiki/Common_cold
[2] https://www.folkhalsomyndigheten.se/folkhalsorapportering-statistik/

account for skewed data. As the relationships between variables in our data were not linear, we focused on the results from the Spearman's correlation tests as it's appropriate for both continuous and discrete ordinal variables independent of their distribution. Testing the data

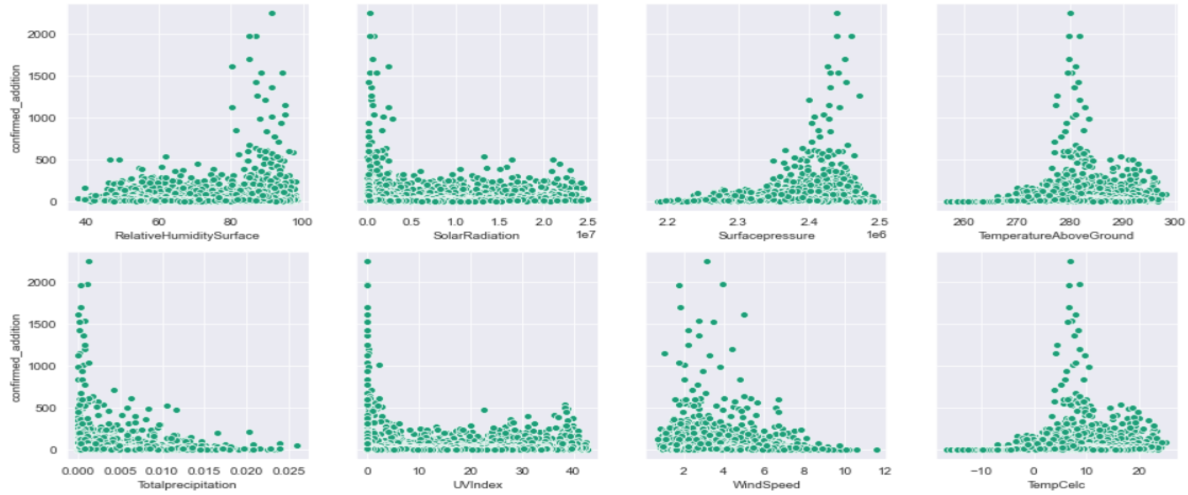
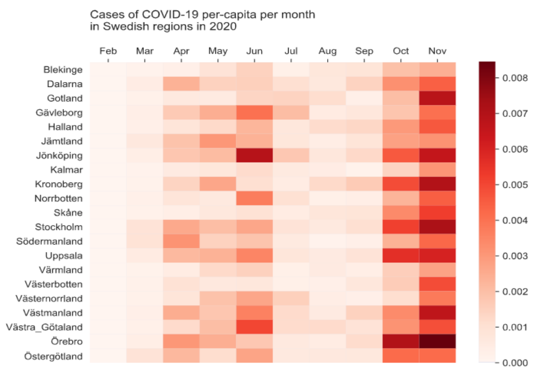
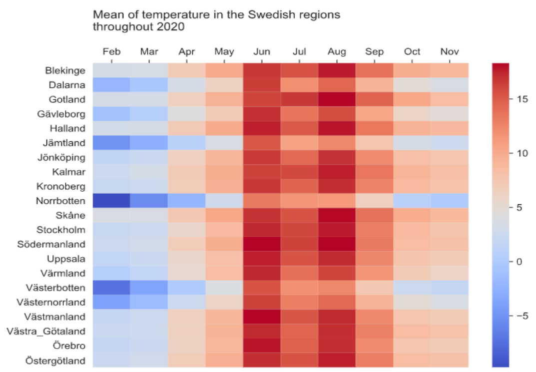


Figure 1: Relationship COVID-19 cases and weather variables for Sweden in 2020

with Pearson's test the correlations varied from 0.16 to -0.09, thus quite close to 0. Spearman's test had correlations varying from 0.22 to -0.9, which is also close to 0. When using Pearson's test of correlation on the logarithmic transformed data the correlations varied from 0.14 to -0.27, meaning correlations are still relatively low. This low correlation is illustrated in figure 1, where there is no clear visual correlation. The variable, which had the highest correlation in both Spearman's and Pearson's test for correlation, was the surface pressure which also had p-values close to zero [3]. There was a correlation between the surface pressure and the cases, but it was relatively low. In relation to the association tests a multivariate regression was made on the weather variables. The R-squared of the regression was 0.054, which means that the data points where spread far away from the regression line. From the low R-squared it was concluded that the multivariate regression analysis was not usable.



(a) Monthly total cases per capita



(b) Monthly mean temperature

Figure 2: Heatmap of COVID-19 cases and temperature in Sweden by region in 2020

Even with relatively weak correlation coefficients, when plotting the evolution of the num-

[3] Pearson's test correlation = 0.16, p-value = 7.617738171132507e-35, Spearman's test correlation = 0.22

ber of cases per-capita by Swedish region throughout 2020 and comparing it to an identical plot with the mean of the temperature variable for the same time period, we can clearly see that there is a relationship as case numbers lowers significantly through the warmer summer months (Figure 2). No new restrictive measures were imposed in the months before this drop. We believe this can can be related to reduced number of tests conducted through the summer months (Fiore, 2020), students being home from school and most activities being moved outdoors (Fiore, 2020).

3.2 COVID-19 cases and Population of Sweden

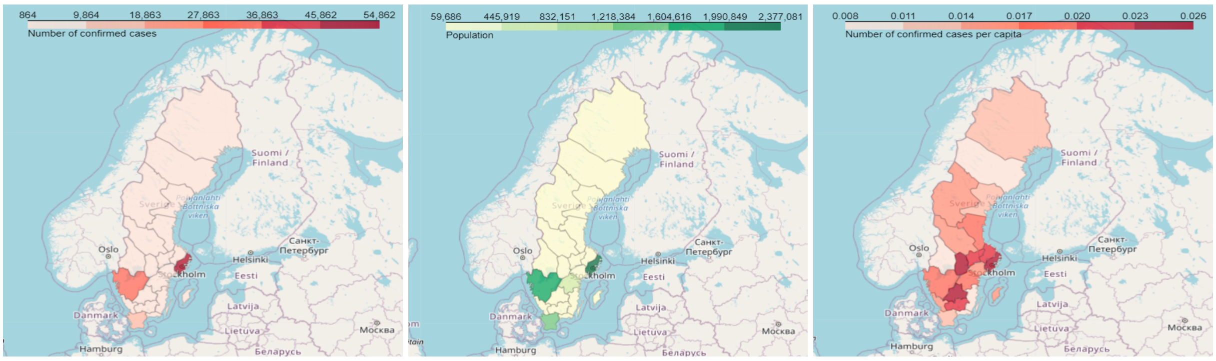


Figure 3: COVID-19 cases in total, population and COVID-19 cases by region in 2020

Although we found a large correlation between the number of people living in a region and the number of confirmed cases, as expected, we also found some unexpected results when looking at cases per-capita (table 1). The regions with biggest variation between population size and cases-per-capita are Skåne and Kronoberg. Skåne ranks 3 in population size and 16 in cases per-capita while Kronoberg ranks 18 in population size and 6 in cases per-capita.

Confirmed cases	Spearman Correlation
Total	0.920
Series Reaction	0.494

Table 1: Correlation of population size, confirmed COVID-19 cases and cases per-capita

3.3 Demographic of COVID-19 cases in Sweden in 2020

As no strong associations were found between weather variables and confirmed COVID-19 cases, we looked into different groups of the population and how they were affected by the virus. We compared age and gender differences with the total number of confirmed cases, intensive care admissions and number of deceased. This was done using an external set containing data of (1) cases by gender and age, (2) intensive care patients, and (3) deceased due to COVID-19.

The bar plots in figure 4 shows differences in both age and gender. In age groups the number of deceased over 80 years and the number from same group, that were admitted to intensive care are surprising: only 364 was admitted to intensive care, while 9.149 deaths was recorded with COVID-19. The chart also shows a majority of men being admitted to intensive care.

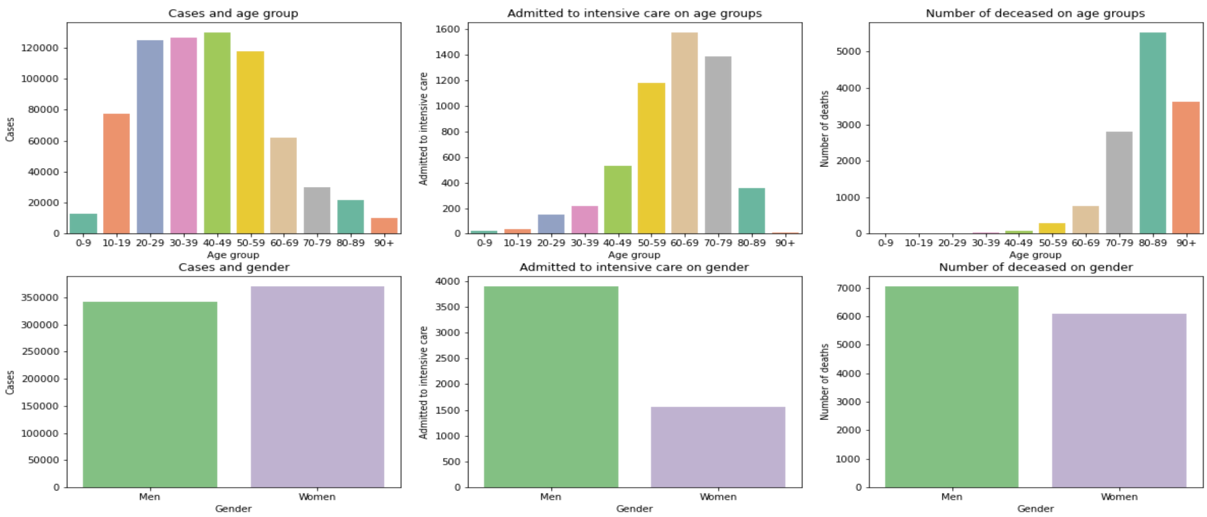


Figure 4: Swedish Demographic of Age and Gender for 2020

This massive difference between admitted cases and deaths could be due to residents in elderly homes were treated at home instead of being transferred to a hospital. This was part of a strategy to protect hospitals from being overwhelmed beyond its capacity (Erica Yarmol-Matusiak et al, 2020) (Peckham et al, 2020).

As we aim to find correlations it is important that our sample size is representative of the whole population. It is difficult to factor these differences into the analysis of the relationship between cases and the weather. Therefore, we decided to implement age and gender demographics of Sweden to get another perspective of the COVID-19 outbreak in 2020. Since the vast majority of cases spans between the age group 20-59, the percentage of severe cases are among men greater than 50 years and cases of deaths are mostly related to elderly at senior care homes. This knowledge makes it even more difficult to conclude a strong correlation between COVID-19 and weather.

Our findings show a low correlation between weather variables and number of cases. This is relevant since the restrictions imposed on the population to reduce COVID-19 cases should not depend on different weather variables. We recommend looking further into age groups, who are most likely to contract COVID-19, whilst looking into groups, who are more likely to be admitted to intensive care, to help facilitate major outbreaks and prevent overburdening hospitals.

4 Limitations

The newly outbreak of COVID-19 contains many unknown variables, which makes it hard to predict and find strong correlations between the confirmed COVID-19 cases and weather conditions. Firstly, we were not given details about how the COVID-19 cases were collected from each region nor do they divide cases into e.g. regular exposure or work-related exposure. Our data set spanned from February 2020 to November 2020 making it challenging to establish a complete analysis of seasonal weather changes and COVID-19 outbreak. Additionally, we do not know if the data set for confirmed COVID-19 cases in Sweden contains false-positives or false-negatives.

The low number of confirmed cases recorded in Sweden in the beginning of the outbreak could be a result of Sweden's test-strategy, lack of enforced government restrictions (Erica Yarmol-Matusiak et al, 2020), and recommendations for people to self-isolate instead of getting tested. Additionally, COVID-19 symptoms can take 4-6 days to manifest, contributing to a delay of confirmed cases, meaning they might not coincide with the day they tested positive. Therefore, the data set appears skewed, and weather data does not necessarily match with the confirmed cases, that has been registered for that particular date.

5 Concluding remarks and future work

The relationship between the COVID-19 outbreak and the weather conditions are weakly related for Sweden in February 2020 to November 2020. Our findings show a small correlation between surface pressure and temperature. However, based upon our limitations for our research we can not confirm a strong relationship between the number of COVID-19 cases and seasonal weather changes.

5.1 Future work

The research conducted raised many other questions about the relationship between COVID-19, the weather and how it affects different demographics. Applying the same methods used for this research when more data is available, would allow us to correlate the same months and seasons in different years and see how the patterns change over time.

Collecting and implementing data such as the number of tests conducted per day overall and per region could explain how the number of cases fluctuate on a national and regional level. Localising infection-occurrence may clarify any misrepresentation in the data, e.g. healthcare workers and nursing homes residents are not representative of general community infections and may misrepresent the spread.

Looking further into the consequences of COVID-19 in different sections of the population could help understand the much lower rate in intensive care given to women and people over 80 years. Especially to understand why 28.59% of people admitted into intensive care were women when they account for 46.25% of COVID-19 casualties in Sweden.

6 Disclosure statement

All the code in the final notebook was organised by Sabrina. However all group members contributed to it through writing code in older versions of the main notebook and their own personal notebooks.

7 Bibliography

Erica Yarmol-Matusiak, A., Cipriano E. L., Stranges S. (2020, November). *A comparison of COVID-19 epidemiological indicators in Sweden, Norway, Denmark, and Finland.* Scandinavian Journal of Public Health, 2021; 49: 69–78. [online]. Available: https://journals.sagepub.com/doi/pdf/10.1177/1403494820980264

Peckham H., Gruijter de M. N., Raine C., Radziszewska A., Ciurtin C., Wedderburn R. L., Rosser C. E., Webb K., Deakin T. C. (2020, December 09). *Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission.* Nature Commonications 11, Article no. 6317 (2020). [online]. Available: https://doi.org/10.1038/s41467-020-19741-6

World Health Organization. (2020, April).*Coronavirus disease (COVID-19): Climate change.* [online]. Available: https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-climate-change

Menebo, M. (2020, October 1). *Temperature and precipitation associate with Covid-19 new daily cases: A correlation study between weather and Covid-19 pandemic in Oslo, Norway*. Science of The Total Environment, vol. 737. [online]. Available: https://doi.org/10.1016/j.scitotenv.2020.139659

Rohde, R. (2020, April 24). *The relationship between coronavirus (COVID-19) spread and the weather*. Berkeley Earth. [online]. Available: http://berkeleyearth.org/coronavirus-and-the-weather/

Pearson, H. (2020, April 10). *Coronavirus: Can cold weather impact the spread of COVID-19?*. Health, Global News. [online]. Available: Coronavirus: Can cold weather impact the spread of COVID-19? — Globalnews.ca

Folmer, K., Katersky, A., Margolin, J. (2020, May 11). *Does warmer weather slow coronavirus?* abc NEWS, Coronavirus Health and Science. [online]. Available: https://abcnews.go.com/US/warmer-weather-slow-coronavirus/story?id=70612437

Fiore, K. (2020, July 29). *How Did Sweden Flatten Its Curve Without a Lockdown?* MedPage, Infectious Disease, COVID-19. [online]. Available: https://www.medpagetoday.com/infectiousdisease/covid19/87812

Independent. (2020, May 03). *Does weather affect the spread of the coronavirus outside?* [online]. Available: https://www.independent.co.uk/news/does-weather-affect-the-spread-of-the-coronavirus-outside-weather-coronavirus-weather-virus-people-b1553150.html