SAHLGRENSKA AKADEMIN
INSTITUTIONEN FÖR MEDICIN

THE INFLUENCE OF OVERCROWDING AND SOCIOECONOMY ON THE SPATIO-TEMPORAL SPREAD OF COVID-19 – A SWEDISH REGISTER STUDY

Mia Söderberg[1], Ottmar Cronie[2], Martin Adiels[3], Annika Rosengren[4]
[1]Occupational and Environmental Medicine, School of Public Health and Community Medicine, Sahlgrenska academy, University of Gothenburg
[2]Applied Mathematics and Statistics, Department of Mathematical Sciences, Chalmers University of Technology & University of Gothenburg
[3]School of Public Health and Community Medicine, Sahlgrenska academy, University of Gothenburg
[4]Department of Molecular and Clinical Medicine, Sahlgrenska academy, University of Gothenburg

REPORT NO. 2022:1

FROM SCHOOL OF PUBLIC HEALTH AND COMMUNITY MEDICINE

Utgiven av Avdelningen samhällsmedicin och folkhälsa, Göteborgs universitet
2022-10-14
ISBN 978-91-86863-28-9

amm@amm.gu.se
031-342 30 40
GU rapporter, Box 414, 405 30 Göteborg

Hemsidor: www.amm.se och gupea.ub.gu.se/handle/2077/34412

Table of contents

ABSTRACT

Household overcrowding, which primarily occur in groups characterised by low socioeconomy, is a known risk factor for the spread of infectious diseases. Still, little is established whether overcrowding, in itself, or in combination with other disadvantageous sociodemographic factors has affected the incidence of COVID-19 infection over time and geographical areas. This register study investigated the effect of overcrowded housing, and its interaction with various markers of low socioeconomy, as predictors for the spread of COVID-19 infection, by using regressive spatial and spatio-temporal statistical methods.

Through the Swedish Tax Agency database, we could identify all legal residents in Sweden, alive at 1st of January 2020 (baseline) and by using Sweden's personal identification number system we could link this cohort to data from other national registers. Through Statistics Sweden, we gained access to information on several sociodemographic variables relevant for this study, including variables to calculate overcrowding, as well as individual information on income, immigration background, education, occupation, and car ownership. To this, we added data from the Public Health Agency of Sweden's register SmiNet, a register for communicable diseases, through which we had access to all positive PCR-test results for COVID-19 in Sweden until the 30th of June 2021.

Based on a definition of overcrowding proposed by Eurostat, and the data at hand, we defined overcrowding as more than one person per number of rooms in a household, with the exceptions of adult couples in a relationship, children under a certain age or anyone living in a villa, detached or semi-detached house. The spatial (geographical) aggregation level was determined by Sweden's so called DeSO zones ("DEmografiska Statistik Områden", translated: Demographical statistical areas), which divide Sweden into 5984 subregions of varying geographical size, with each zone being inhabited by 700–2700 people. The temporal unit of the data was calendar month. As each person appearing in the statistics Sweden's register has a DeSO identifier, we could generate monthly counts of infection for each DeSO zone.

We started the statistical analyses by evaluating correlations among the spatial covariates and their interaction terms, as well as correlations between the spatial covariates and the log-counts per DeSO, both monthly and accumulated over the whole study period (18 months). This highlighted that the dynamics of infection incidence over time and that certain groups with several highly dependent covariates had a pronounced impact on infection. Our spatial analyses were conducted using an elastic-net regularised Poisson regression approach where the DeSOs' population sizes were used as off-sets and the model selection was carried out by means of cross-validation. In the spatio-temporal analyses, a dummy variable was added for each month, while keeping the rest of the covariates, including the interaction terms, as in the spatial analyses. This approach allowed us to interpret the fitted models as models for the risk that a generic individual from any kind of DeSO zone tested positive (at a given time point), while also adjusting for collinearity and carrying out variable selection in our models to achieve parsimony.

The descriptive results, which we visualized by graphical illustration of the spatially aggregated data, showed clear co-existence of overcrowded housing, low education, low income and having an foreign background in several geographical zones, especially in some of the boroughs of Sweden's largest cities. The analyses focusing on geographical areas' vulnerability (spatial risks) revealed higher risks in areas with a high occurrence of overcrowding, especially in interaction with a high proportion of inhabitants with a foreign

background, an income below the national median or persons in health and social care work. When incorporating time in the models (spatio-temporal risk), overcrowding appeared as a predictor for COVID-19 infection, however, only during the time periods of April, May, August, and November 2020. Overcrowding otherwise seemed to foremost constitute a risk factor when interplaying with other disadvantageous socioeconomic variables, thus indicating that general socioeconomic vulnerability constituted a risk enhancer. Else, being of foreign background or being employed in a low-income job during the second wave of the pandemic, were notable predictors for the risk of testing positive for the disease.

By studying overcrowding and socioeconomic factors, we identified vulnerable groups per geographical area and over the duration of the COVID-19 pandemic. Identified risk factors were clearly more prevalent in groups whose structural living conditions meant less possibilities to protect themselves, and which also already displayed markedly worse health. Targeted interventions towards ill-disposed group and geographical areas are therefore of importance in the still on-going pandemic or in the event of future widespread diseases.

INTRODUCTION

Across history and societies, household overcrowding has been associated with the risks of contracting infectious diseases, such as tuberculosis and measles (1, 2). This is presumably caused by increased disease transmission through close physical proximity to others, sharing of objects and surfaces which are not regularly sanitized, or difficulties to isolate sick household members.

While some limited occurrence of overcrowded living relates to personal preferences or traditions, it is predominantly an indicator of poverty. Social inequalities have had a striking impact on COVID-19-related health outcomes, as markers of low socioeconomic status or an immigrant background have been strongly associated with augmented risks for infection, severe disease and mortality (3, 4). Increased prevalence of severe outcomes is partly explained by a higher prevalence of pre-existing diseases. However, the risk of becoming infected is tied to aspects such as less possibilities to work from home and having to travel to work by public transport due to the lack of means to own a car (5, 6). Housing conditions, being related to becoming infected within one’s own home, is another potential contributing component in a multifactorial cluster of exposure variables from which the possibility to protects oneself varies greatly between societal groups.

There are various definitions of overcrowding and the most intuitive parameter for it is perhaps the living area per number of household members. However, with regard to disease transmission, it is more likely that the number of rooms has a stronger impact on the risk of becoming infected. In this context, more than one member per room, i.e., needing to share a room with the exceptions of rooms shared between adults in a partnership or children under a certain age, is commonly used as the definition of overcrowding. Overcrowded housing is prevalent in all European countries, but with notable differences between and within countries. Sweden is the Nordic country with the highest proportion of overcrowded households, amounting to roughly 16% of all residents, compared to Norway (6%) and Finland (8%) (7). Sweden also displays among the largest within-country gaps in housing depravation in Europe, as 30% of all foreign-born persons reside in overcrowded households, compared to 9% of Swedish-born persons (7).

Over the COVID-19 pandemic, overcrowding has been associated with an increased risk of infection and mortality. In the British Virus Watch study, utilizing monthly SARS-CoV-2 antibody tests in roughly 10 000 adults, those most overcrowded in their home displayed the highest proportion of COVID-19 (6.6%) compared to the least crowded (2.9%) (8). Mixed effects logistic regression models, adjusted for age, sex, ethnicity and household income showed a more than three-fold odds ratios (3.70; 95% confidence interval (CI) 1.92-7.13) for a positive antibody test in those overcrowded, compared to those in under-occupied homes. In another study, with aggregated US data per county, the counties with the highest household crowding (16.8 households per 100 000) contained the highest death rates of COVID-19 (4.9 cases per 100 000) (9). A multilevel negative binomial regression model confirmed an increased mortality rate of about 18% (95% CI 13-27%) in overcrowded counties during June-July 2020, although there were no differences between counties in August-October the same year. Yet, the highest mortality rates were, notably, found in counties with a larger proportion of Afro-Americans (36%; 95% CI 30-43%) or Hispanic Americans (28%; 95% CI 22-35%), adjusting for overcrowding, age, level of education and poverty.

A common feature of a pandemic driven by an infectious disease is an infection pattern where a small amount of cases quickly generate a large number of secondary infections (10), often exuberated by groups with certain sets of characteristics. Living in an overcrowded home is likely such a characteristic. In addition to exposure risks within the home, these households are commonly also concentrated to the same geographical areas, creating a circular infection route between households and the local community.

Due to the nature of disease transmission, disease exposure is a spatio-temporal process which tends to be "non-separable" in the sense that high/low disease rate regions are displaced over time, with the number of cases occurring within a given region fluctuating over time. How a particular region is affected over time depends on two essential components: 1) the between-region transmission, i.e., to what degree there is disease transmission from neighbouring regions by road-travel proximity or other means of transportation, such as trains and airplanes, 2) the region's general vulnerability, depending on its inhabitants' overall health and living conditions, which determines the inhabitants' ability to shield themselves from disease exposure or to decrease the general transmission through isolation. One typically refers to the former as spatial(-temporal) interaction and the latter as spatial risk. Through spatio-temporal statistical analyses it is possible to analyse which variables (predictors or covariates) that drive the spatial risk. Based on previous studies of the COVID-19 pandemic, in addition to overcrowding, spatial risks are likely driven by a combination of factors, such as work in high-risk occupations, low income, short formal education or being of an immigrant background (11-13). Adding to this, different infection risk variables are not only likely to interplay, but also to enhance the effect of each other.

The aim of this study was to examine overcrowding as a predictor for testing positive for COVID-19 over time and geographical subregions of Sweden. To this end, we chose a spatio-temporal modelling approach which, in addition to overcrowding and multiple markers of low socioeconomy, incorporated the interactive effects between these potential individual drivers. This setup is derived from our hypothesis that, in terms of spatially delineated covariates, it is not simply one single covariate, e.g., overcrowding, which primarily drives the risk of being affected by COVID-19, but rather the combined effects from a cluster of covariates.

METHODS

Population

Through Statistics Sweden's registry of the total population, kept by the Swedish Tax Agency, we could identify all legal residents in Sweden alive on January 1st, 2020, who were followed until the 31st of June 2021.

Data

By using Sweden's personal identification number system, it was possible to link data from several national registers. Through Statistics Sweden's register, the Longitudinal integrated database for health insurance and labour market studies (LISA) register, we obtained sociodemographic data, e.g., income, occupation, and housing, last updated in November 2018. Using this information, we excluded individuals in elderly homes or housing for persons with lowered physical abilities (n=307 249), since their housing situation and, thus, infection risks differ from the population-wide pattern. Lastly, we linked data from the Public Health Agency of Sweden's register SmiNet, which contains data over communicable diseases in Sweden. This register study has been approved by the Swedish Ethics Review Authority (2020–02019).

Variables

We estimated overcrowding, our main explanatory variable of interest, based on information on type of dwelling, number of household members, number of rooms (only available for apartments) and relationship between the household members. Our definition starting point was provided by Eurostat (https://ec.europa.eu/eurostat/statistics-explained/index.php). It defines an overcrowded housing in terms of the number of rooms per person in the household while taking into account the age and relationship of the cohabitants, stating that a household is overcrowded if it has fewer rooms than the total of:

- one room per couple in the household
- one room for each single person aged 18 or more
- one room per pair of single people of the same gender aged 12-17 years
- one room for each single person aged 12-17 years, and not included in the previous category
- one room per pair of children under 12 years of age

However, the available data from the LISA register did not form a sufficient basis to recreate the definition. For example, we could only define members as being in a partnership if being married or in a registered civic partnership or being the joint parents/legal guardians of a child. Other partnerships could not be identified. The Eurostat definition's age categories could neither be recreated exactly. Hence, we used a slightly altered definition of overcrowding. Our definition states that a person does **not** live in an overcrowded household if:

- (s)he lives in a villa, detached house or semi-detached house
- (s)he lives in an apartment and constitutes a single household without any children living in the household
- (s)he lives in an apartment of at least two rooms and belongs to a household of two without any children living in the household
- (s)he lives in an apartment where the household has at least three adults, no children are living in the apartment (adult or non-adult) and the apartment has at least as many rooms as there are people + 1
- (s)he does not belong to any of the groups above, lives in an apartment where there are children in the household and the number of rooms in the apartment amount to at least:
    - one common (living) room for the household
    - one bedroom for the parents (if being registered partners or a single parent)
    - one bedroom per extra adult, i.e., adults who are not children of the parent(s) of the household
    - one room per pair of children under 11 years of age
    - one room per every child of age 12 and older

Concerning the spatial (geographical) aggregation level available to us, all data considered in this study follow the DeSO delineation ("DEmografiska Statistik Områden", translated: Demographical statistical areas), which subdivides Sweden into 5984 zones. The DeSO zones, which have been created with consideration to municipality borders, as well as natural geographical dividers such as rivers and forests, are of varying sizes and have population sizes, ranging from 700 to 2700 inhabitants.

Through the SmiNet register we obtained information on all confirmed COVID-19 cases, here defined as positive polymerase chain reaction (PCR) test for COVID-19 infection, including the date of testing positive. Using the DeSO codes of all these individuals, we extracted the counts of cases (per month) for each DeSO zone.

Overcrowding most likely interplays with other sociodemographic variables, which we wanted to account for in our analyses. Among the considered covariates, some were obtained as DeSO-aggregated binary (yes/no) variables and reported as either proportion of the inhabitants of a DeSO or absolute counts. Some variables were prepacked variables from Statistics Sweden, while others were generated from LISA data. These variables were all based on data from 2018, for which demographic changes from 2018 and onwards can be assumed to be negligible. The complete list of individual covariates we considered for each DeSO were:

- proportion of overcrowded people (Overcrowding),
- average m^2 per person (m^2/person),
- proportion of people with an economic standard below the national median (Low economic standard),
- proportion of people without secondary education or higher (Lack (post-) secondary education),
- number of person cars (Person cars),
- proportion of gainfully employed people (Gainfully employed),
- proportion of people with a foreign background, defined as being foreign-born or born in Sweden with both parents being foreign-born (Foreign background),
- proportion of health and social care workers (Health care workers), based on occupational classification according to the "Standard för svensk yrkesklassificering", the Swedish adaption of the International Classification of Occupations 2008 (14).

Note that the variable m^2/person was included to elucidate whether the lack of living space had an enhancing effect on overcrowding. We also chose to let a high value for employment represent a high employment rate, since carrying out a job was viewed as an exposure variable. Car ownership represent the total number of cars in a DeSO rather than the number of cars per person, since we view each individual car as an opportunity to avoid public transportation. The chosen temporal resolution for our spatio-temporal analyses was calendar months, which provides a fine enough aggregation level to detect relevant trends and changes, while providing sufficiently large counts to yield stable statistical modelling.

Statistical analysis

The statistical methods used are presented here is in brief, while a more detailed description and reasoning behind the chosen methods is found in Appendix I. Given that the outcome variable, i.e., the number of confirmed cases, represents counts per DeSO zone (per month), we chose a Poisson regression approach for its modelling, which in addition to being natural for count data, also results in more easily interpretable models. In order to obtain an understanding of the underlying dependencies, as well as to guide the subsequent modelling approach, we conducted several exploratory analyses. We started by evaluating correlations between the spatial covariates, as well as correlations between the spatial covariates and the log-counts per DeSO (accumulated over the whole study period). We further estimated correlations between the spatial covariates, as well as correlations between the spatial covariates and the log-counts per DeSO, per month, excluding observations with count 0. These analyses concluded that there are clear collinearities present and that the influence of a given (interaction term) covariate on the log-counts tends to change with time.

To fit our models to the data, we chose a prediction-based approach, as opposed to a classical statistical inferential approach. More specifically, since there were clear collinearities present and we wanted to find out which single, or combinations of covariates, that had an actual predictive influence on the counts, we used an elastic net regularised regression approach

(15). This enabled fitting the chosen Poisson regression structure while simultaneously i) adjusting for collinearity and ii) carrying out variable selection, i.e., covariates with little predictive effect are excluded from the models. Hence, simplified, we assumed that dependencies between DeSO zone counts (over time) with e.g., result from disease transmission, are solely driven by the underlying covariate structure (i.e., a kind of conditional independence).

In order to enter all variables on a comparable scale into the models, they were standardised by subtracting the individual means and then dividing by the individual standard deviations. Considering the observed collinearity between this list of covariates, we additionally included all (second order) interaction terms. In order to obtain a model for a person's individual risk, given the DeSO-based data at hand, we modelled the DeSO counts (y_i) using Poisson regression models with the population size (n_i) as offset: y_i/n_i = *exp(linear combination of the individual DeSO-based covariates and all of their interaction terms)*. As a result, the exponential term essentially describes the *risk* that an individual, who lives in a DeSO zone with a certain set of characteristics, becomes infected. Since our observation time was 18 months, the individuals in a DeSO zone may become re-infected, and this risk will not necessarily fall between 0 and 1. Our spatio-temporal extension was achieved by adding a dummy variable for each calendar month, while retaining the rest of the covariates and interaction terms in the model. This approach allows us to study both individual covariate's potential influence on the counts and how interplays between the covariates affect the predicted risk of having high counts, relative to the population size.

We used K-fold cross-validation for the hyperparameter selection, which here included both the penalty scaling parameter and the internal elastic net parameter. The latter determines how much weight to prescribe to lasso penalisation, on the one hand, and ridge penalisation, on the other hand, the former determines how parameter-rich a proposed model will be. We here followed standard practice and set the number of folds to K=10. In the spatio-temporal setting, we used penalty factors to force the individual month dummy variables to be included in the final model. The modelling based on standardised covariates allowed us to study variable importance in the fitted models, i.e., the larger a fitted parameter turned out, the more the covariate in question contributes to the fitted risk.

RESULTS

Descriptive

In total n= 10 205 571 persons were included in the cohort at baseline. During our observation period 1st of January 2020 to 30th of June 2021, a total count of 1 006 696 confirmed cases occurred, which were used to obtain the DeSO counts. A graphical illustration of the spatially aggregated data of cases in Sweden in Figure 1.

The structure of the DeSO zones, with consideration to the number of inhabitants, means that larger cities tend to contain several, often smaller, DeSO zones. We will therefore present enlargements of Sweden's three largest cities, Stockholm (population n ≈ 976,000), Gothenburg (n≈579,000) and Malmö (n≈344,000), in figure 2a-c. The Stockholm region is located on the eastern central peninsula while Malmö is located in the very south-west of the country and Gothenburg is located just south of the western-most part of the country. Illustrations of the monthly count data of cases for Sweden, Stockholm, Gothenburg and Malmö can be found in Appendix II.

Figure 1. Total counts of cases per DeSO zones in Sweden

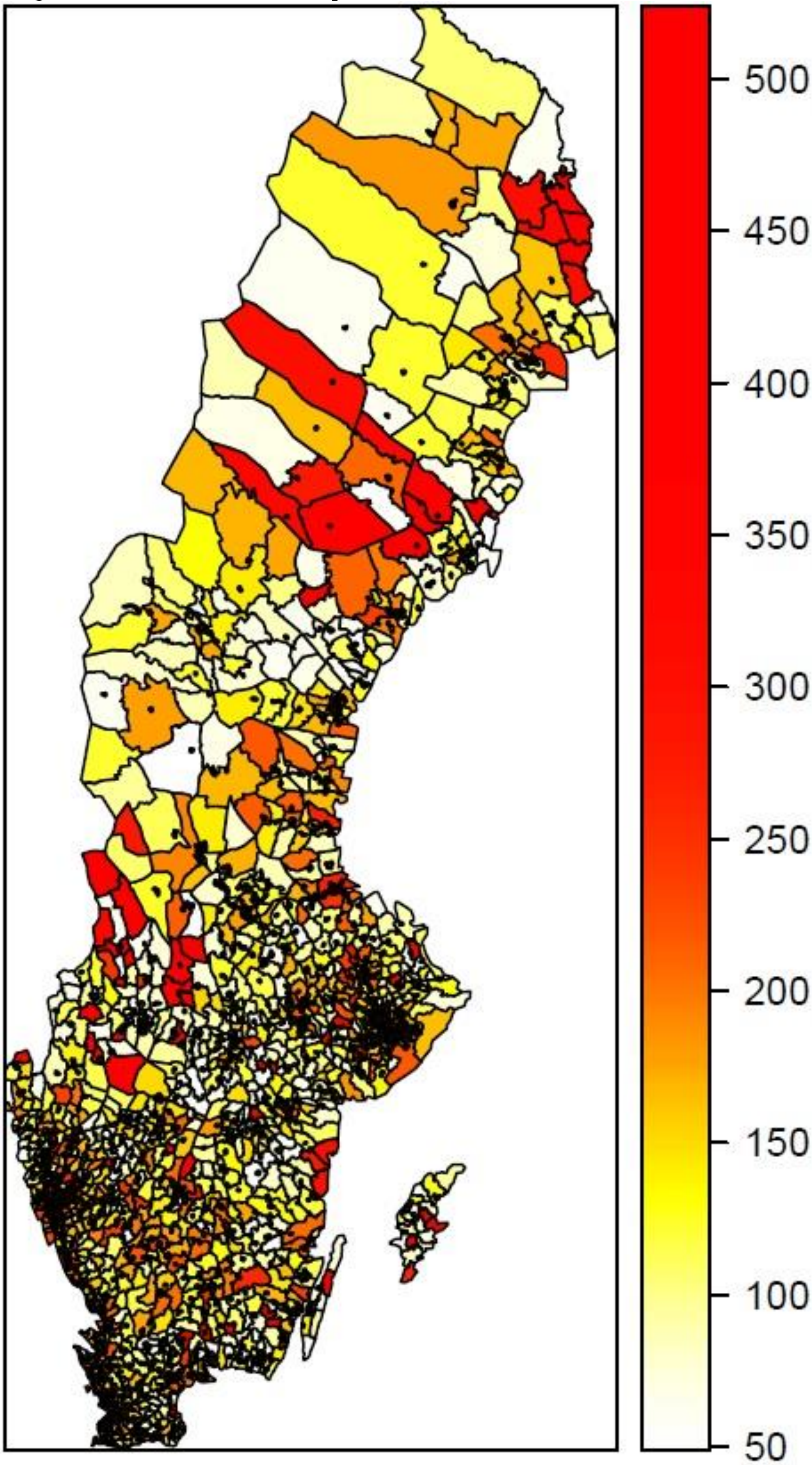


Figure 2a. Total counts of cases per DeSO zones in Stockholm

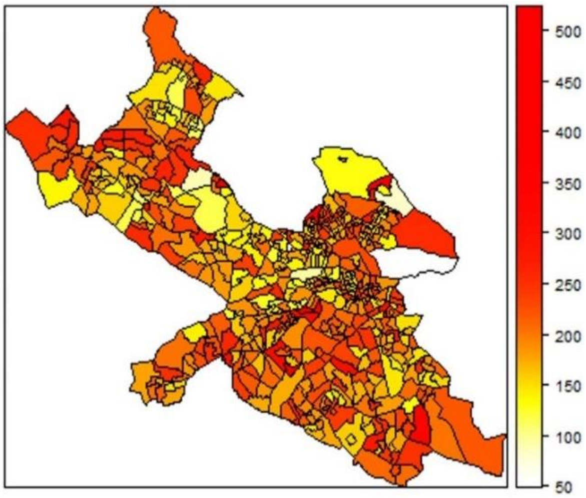


Figure 2b. Total counts of cases per DeSO zones in Gothenburg

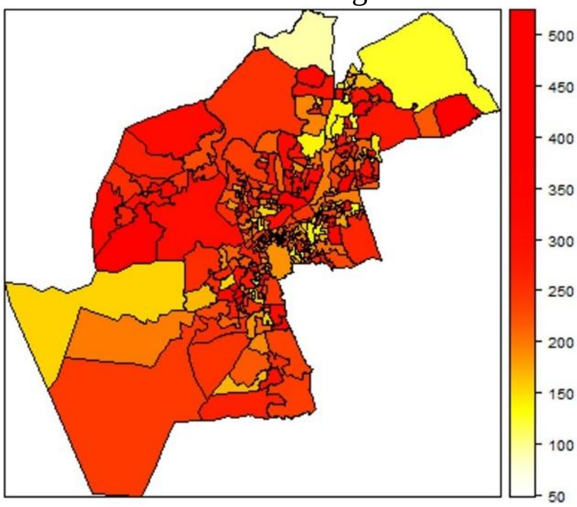


Figure 2c. Total counts of cases per DeSO zones in Malmö

The proportion of overcrowded inhabitants per DeSO zones varies greatly between different regions of Sweden (Figure 3) and between different boroughs in the larger cities (figure 4a-c). It may be of note that since the DeSO delineation has a population size restriction, several DeSOs will be geographically large, while still having few inhabitants, and are greatly affected by having one larger habitual area, in which overcrowding occurs.

Figure 3. Proportion of overcrowding in Sweden per DeSO zones

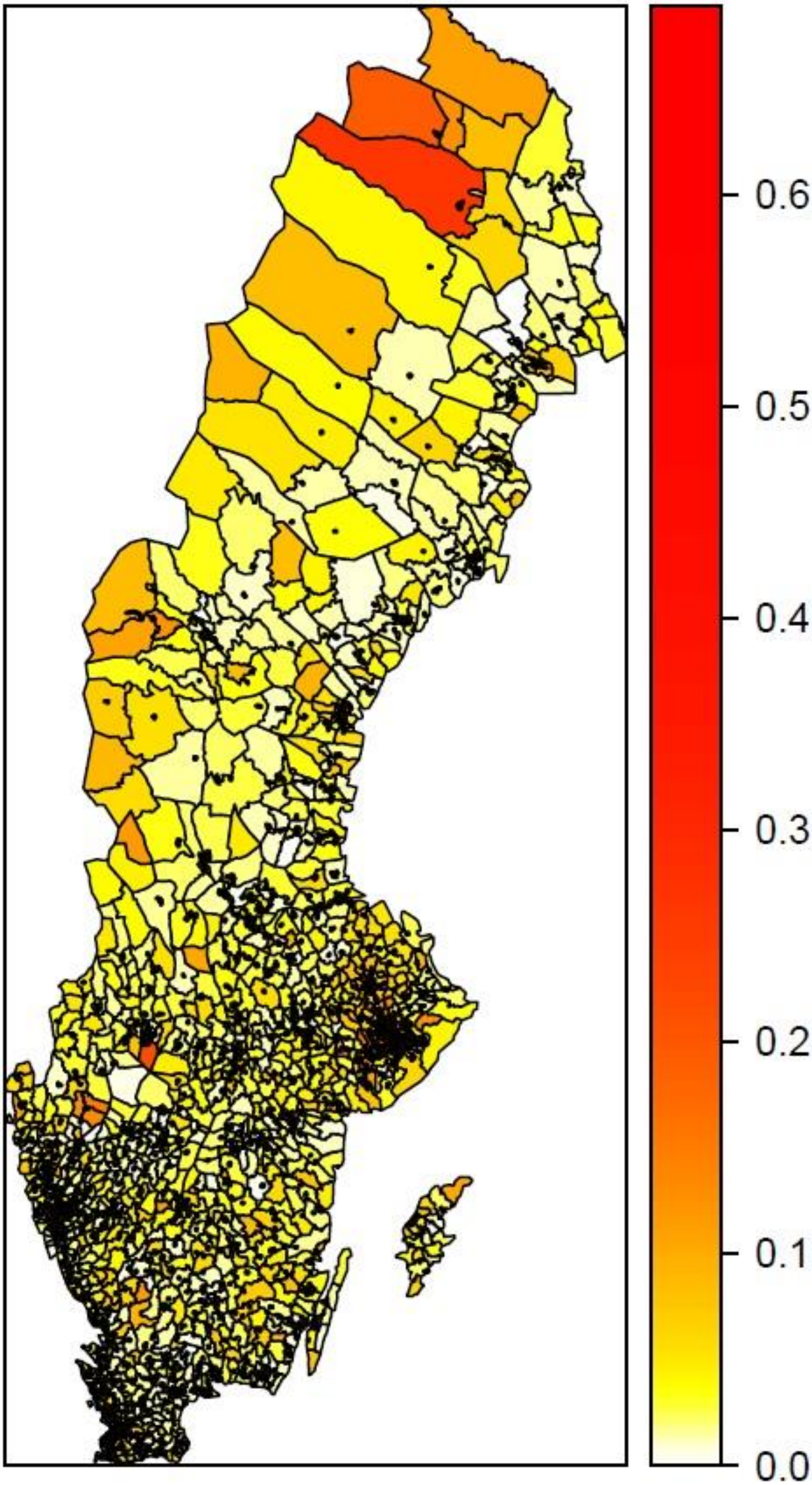


Figure 4a. Proportion of overcrowding in Stockholm per DeSO zones

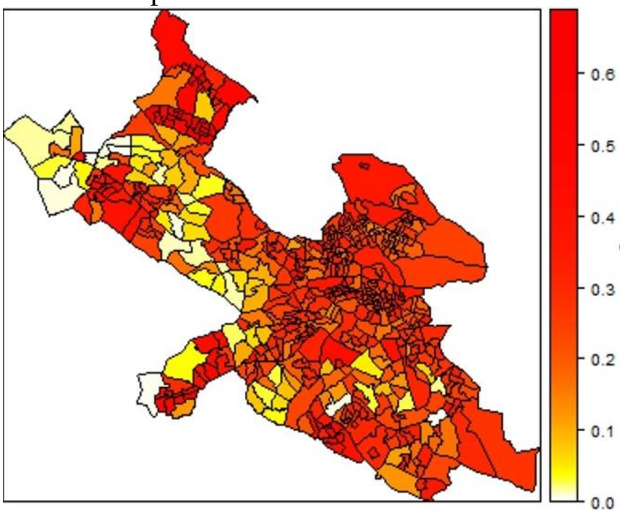


Figure 4b. Proportion of overcrowding in Gothenburg per DeSO zones

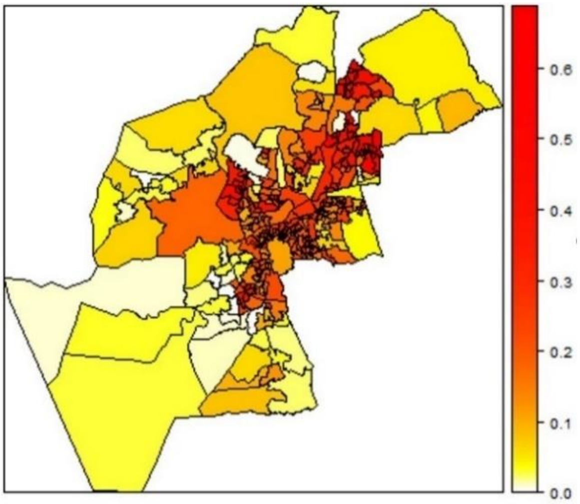
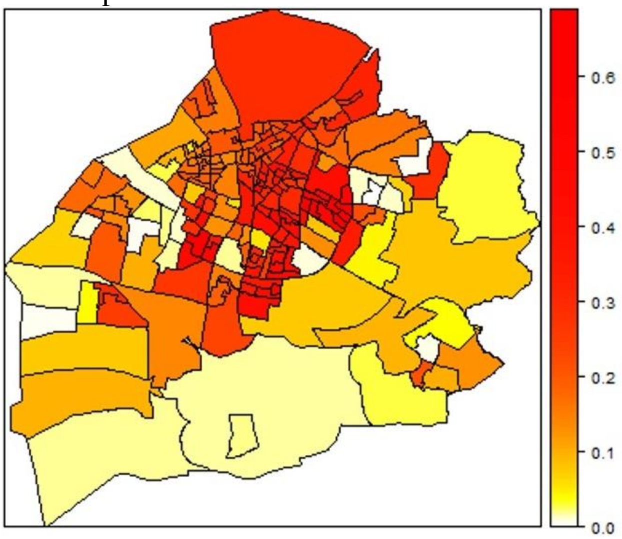
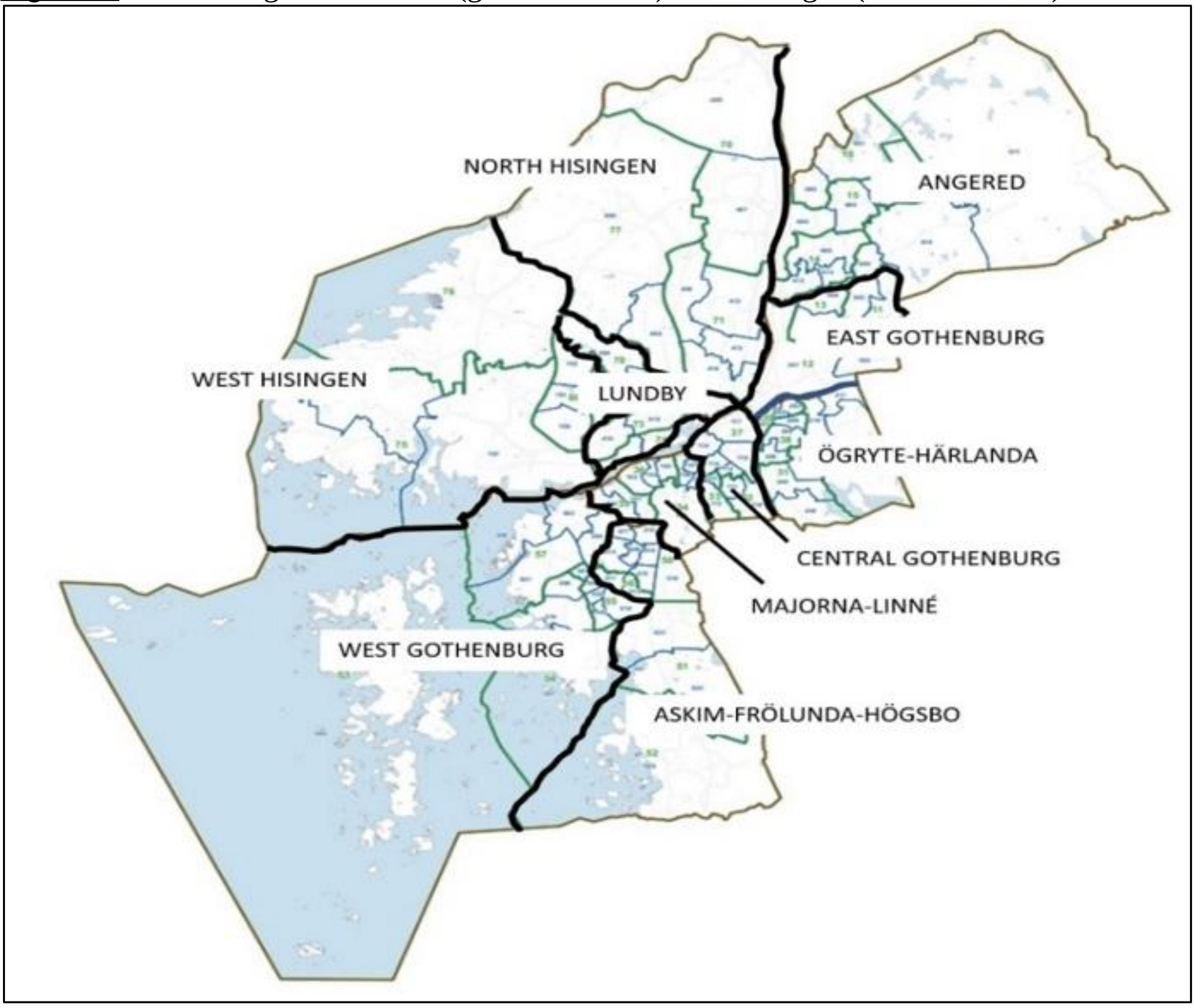


Figure 4c. Proportion of overcrowding in Malmö per DeSO zones



Graphical illustration of the spatially aggregated data for our chosen covariates indicated that overcrowding and most of the socioeconomy markers of seem to correspond. To better illustrate this, we focused on one city Gothenburg (figure 5), for which we show the geographical presentation of economical standard below the country median, lacking secondary or higher education, being gainfully employed, having a foreign background, work in health or social care and passenger cars per DeSO zones, in figure 6a-f. Results for these variables for Sweden, Stockholm and Malmö can be found in Appendix III.

Figure 5. Gothenburg DeSO-zones (green boarders) and boroughs (black boarders)



Angered, North Hisingen and Lundby, commonly known to be disadvantageous boroughs, were characterized by a higher proportion of persons with overcrowded living, lower income, short formal education, unemployment and having a foreign background. Meanwhile, South-West Gothenburg (archipelago), Askim-Frölunda-Högsbo and West-Hisingen are more well-to-do areas, with almost mirrored results compared to Angered, Hisingen and Lundby, in most of the examined sociodemographic variables. DeSO zones within the same borough foremost displayed similar results, with Angered deviating slightly. This may be explained by that Angered consists of areas with both high-rise apartment buildings and areas which are dominated by forest, fields and villas/detached houses.

Figure 6a. Proportion of people with an economic standard below the national median per DeSO zones in Gothenburg

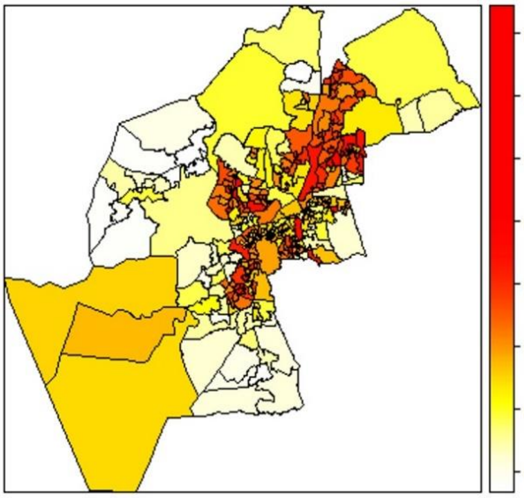


Figure 6b. Proportion of people without (post-) secondary education per DeSO zones in Gothenburg

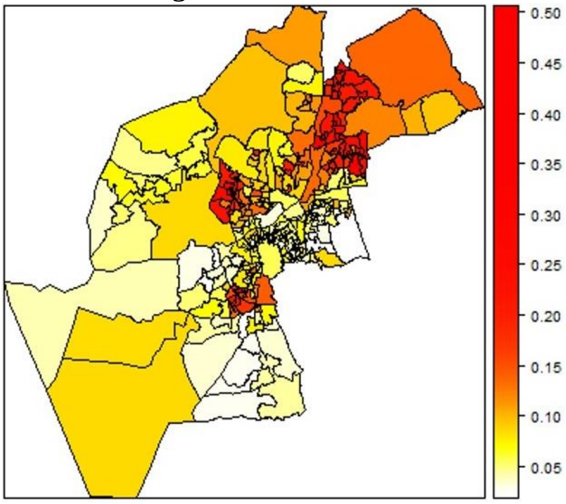


Figure 6c. Proportion of gainfully employed people per DeSO zones in Gothenburg

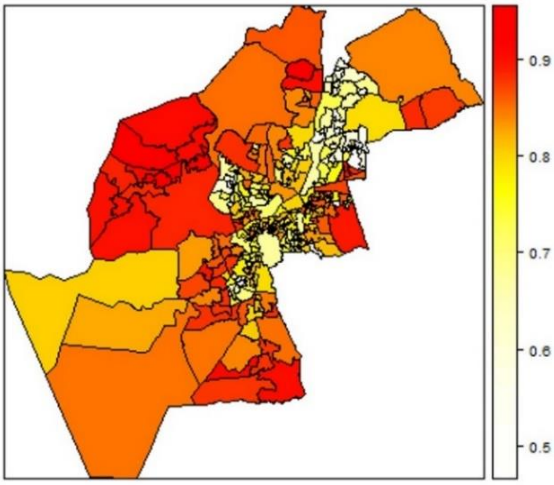


Figure 6d. Proportion of people with a foreign background per DeSO zones in Gothenburg

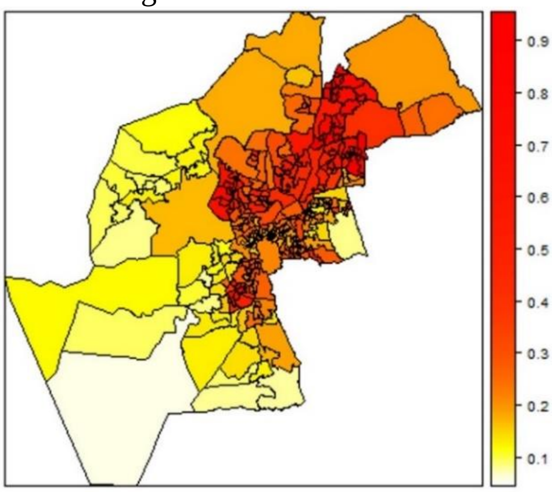


Figure 6e. Proportion of people in health care work per DeSO zones in Gothenburg

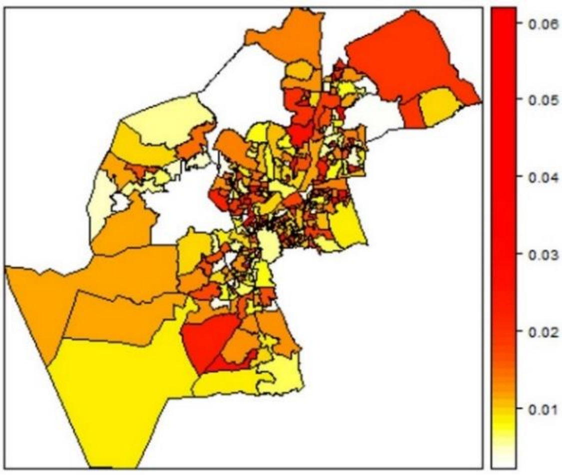
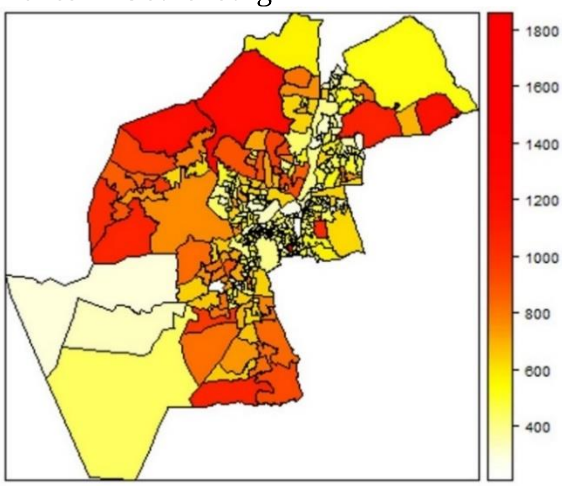


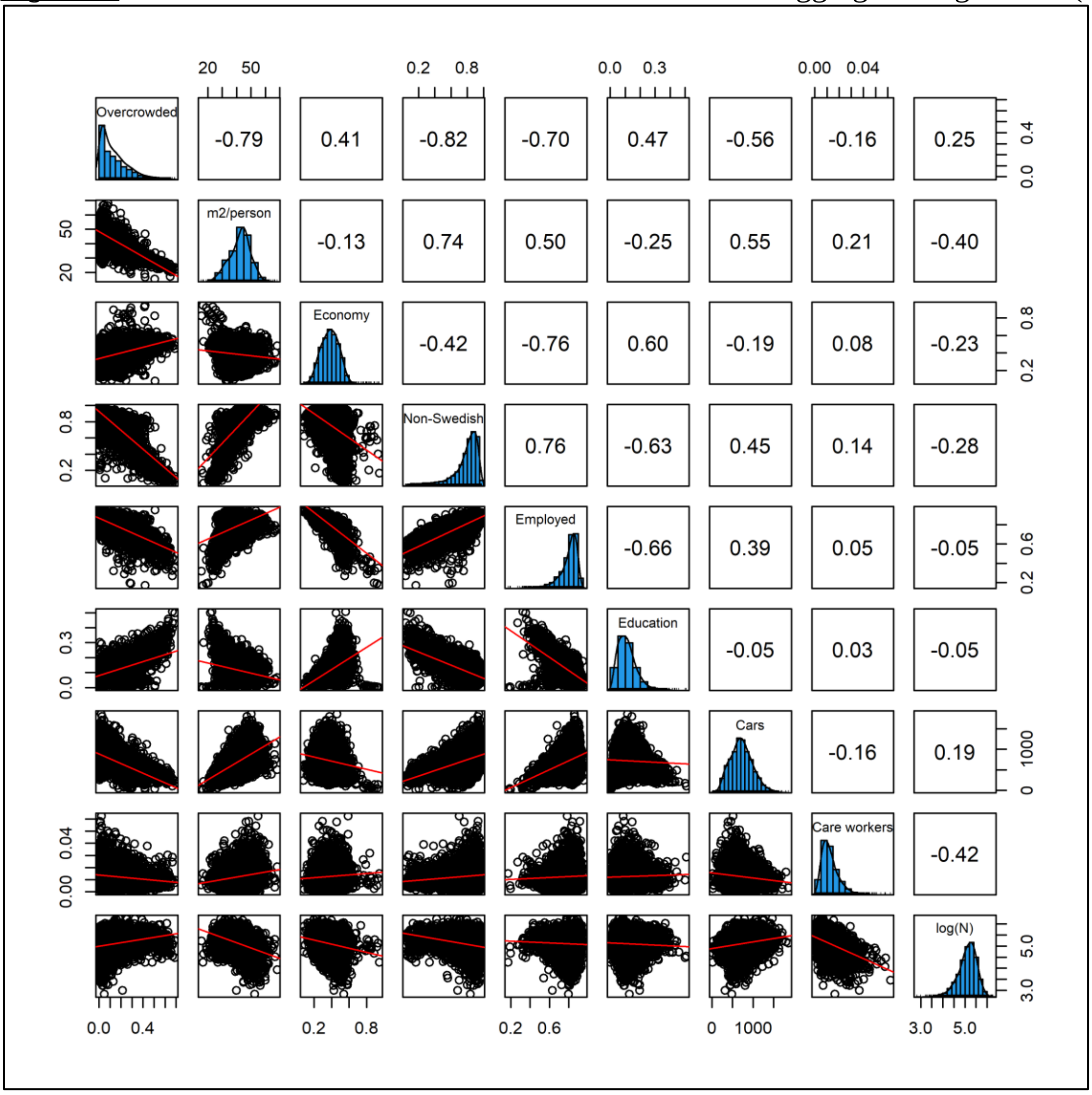
Figure 6f. Counts of person cars per DeSO zones in Gothenburg



Spatial analysis

To decipher how the individual predictors depend on each other and to study to what degree they correlate with the log-counts, we analysed all related correlations (Figure 7). Aside from finding several correlations between the covariates (collinearity) we found some dependencies between the aggregated log-counts and individual covariates, most notably a low proportion of health care workers and a low average m^2/person. Note, however, that none of the correlations are very strong.

Figure 7. Correlations between individual covariates and the aggregated log-counts (log(N))



We next turn to the purely spatial regression modelling where the response variable is given by the total count of all cases occurring in a DeSO zone, throughout the study period. We obtained an internal elastic net parameter of 0.5, which means that we put equal weight on ridge regression and lasso regression, and a penalisation parameter value of 0.4039689, using the 1 standard error rule. Table 1 reports the obtained model, which due to the elastic net induced variable/model selection included only 21 covariates (out of a total of 8+28+8=44 covariates) plus an intercept term.

Table 1. Spatial model

Variable name	Estimate	Exponentiated estimate
Overcrowding*Foreign background	0.043	1.044
Overcrowding*Low economic standard	0.019	1.019
m2/person	0.017	1.017
Overcrowding*m2/person	0.015	1.015
Low economic standard*Gainfully employed	0.008	1.008
Overcrowding*Health care workers	0.007	1.007
Lack (post-) secondary education*Person cars	0.005	1.005
Gainfully employed*Person cars	0.004	1.004
m2/person*Gainfully employed	0.003	1.003
Person cars*Health care workers	0.003	1.003
Overcrowding*Person cars	0.003	1.003
Gainfully employed*Lack (post-) secondary education	0.000	1.000
Lack (post-) secondary education*Health care worker	0.000	1.000
Low economic standard*Health care worker	-0.003	0.997
Foreign background*Lack (post-) secondary education	-0.007	0.992
m2/person*Foreign background	-0.010	0.989
Low economic standard*Foreign background	-0.011	0.989
Foreign background *Person cars	-0.012	0.987
Overcrowding*Gainfully employed	-0.014	0.985
m2/person*Low economic standard	-0.016	0.983
Overcrowding	-0.149	0.860
(Intercept)	-2.120	0.120

Overcrowding seemed to have the most prolific predictive effect on the spatial risk, provided that a high degree of overcrowding within a DeSO occurred in conjunction with 1) a high rate of persons with a foreign background and 2) a high rate of people with an economic standard below the national median. Interaction with overcrowding and whether there was a high rate of health care workers also entered high on the list. DeSO zones with a higher number of people with a low economic standard, but gainfully employed also experienced increased risks. Moreover, complex results for car ownership were found, as an increased degree of persons cars, in combination with either decreased education level, or an increased employment level, or overcrowding in a DeSO, added to the risk. We also see that DeSO zones with a high rate of people living in households with few m^2/persons, i.e., a decrease in the less refined measure of overcrowding, increased the risk to a large degree, both individually and in interaction with other covariates, this was particularly the case when combining the two measures, i.e., overcrowding in combination with a small living space. Note the relatively small intercept, which we interpret as the obtained model capturing a relatively large degree of the features of the data.

Spatio-temporal analysis

We next turn to the spatio-temporal analysis. Correlation plots for the monthly log-counts in Appendix IV. The internal elastic net parameter was 0.6 while the penalisation parameter became 0.009060382, using the 1 standard error rule. The 40 highest ranked estimates obtained estimates are presented here in Table 2. The full table, containing all included predictors and their corresponding coefficient estimates, can be found in Appendix V.

Table 2. Highest ranked obtained estimates of the parameters in the spatio-temporal model

Variable Name	**Estimate**	**Exponentiated estimate**
Month13	2.172	8.781
Month12	1.546	4.695
Month14	1.092	2.982
Month11	0.921	2.513
Month15	0.739	2.095
Month17	0.708	2.030
Month18	0.669	1.952
Month16	0.477	1.612
Foreign background*Month17	0.194	1.214
Foreign background*Month18	0.185	1.203
Person cars*Month19	0.171	1.187
Gainfully employed*Month16	0.156	1.168
Foreign background*Month16	0.148	1.160
Foreign background*Month19	0.142	1.152
Lack (post-) secondary education*Month17	0.141	1.151
Low economic standard*Month14	0.139	1.149
Lack (post-) secondary education*Month16	0.126	1.135
Lack (post-) secondary education*Month18	0.115	1.122
Overcrowding*Month5	0.115	1.121
Foreign background *Month15	0.111	1.117
m2/person*Month14	0.110	1.117
Person cars*Month14	0.093	1.097
Low economic standard*Month19	0.089	1.094
m2/person	0.089	1.093
Overcrowding*Month4	0.085	1.089
Gainfully employed*Month17	0.083	1.087
Lack (post-) secondary education*Month5	0.072	1.074
Gainfully employed*Month7	0.063	1.065
Low economic standard*Month13	0.060	1.062
Gainfully employed*Month12	0.060	1.062
Gainfully employed*Month15	0.059	1.061
Person cars*Month18	0.058	1.060
Health care workers*Month7	0.056	1.058
Lack (post-) secondary education*Month4	0.054	1.056
Low economic standard*Month17	0.053	1.055
Lack (post-) secondary education*Month15	0.053	1.055
Person cars*Month13	0.053	1.055
Lack (post-) secondary education*Month7	0.051	1.052
Health care workers*Month19	0.046	1.047
Overcrowding* Foreign background	0.044	1.045

The months 11 –18 (November 2020 – June 2021), which constitute the period of the second COVID-19 wave, had the biggest impact on a DeSO zone being at an increased risk. During this time period there were also notable effects for the interaction between time and being of foreign background, lacking (post-) secondary education and low income. The first ten months and the 18th month seem to contribute less to the risk, although the first five months most likely reflect the low PCR-testing outside health care workers or hospital admitted persons. Overcrowding seemingly interplayed foremost with time, as different months had larger fitted coefficients for the corresponding interaction terms. In particular, the effect of overcrowding seemed to have been less pronounced during 2021, than during 2020, where the months April, May, August and November of 2020 stood out. As mentioned above, testing was mostly performed for health care workers and in severe cases and could possibly represent special vulnerability in certain groups outside health care workers. High overcrowding seems to additionally increase the risk in combination with a high rate of low economic standard, health care workers, people gainfully employed, or people with less than secondary education. Lastly, car ownership seems to have an interplay with time, in particular during 2021 and a higher proportion of people of foreign background also seems to increase the risk when occurring jointly with overcrowding, lacking secondary education and low income, during 2021.

DISCUSSION

The conducted analyses of risk factors within a geographical area (spatial risks), indicate that areas with a higher proportion of overcrowding, especially when occurring jointly with a higher proportion of inhabitants with a foreign background, an economic standard beneath the national median or health care workers, ran a higher risk of having inhabitants infected with COVID-19. When adding the dimension of time, i.e., conducting spatio-temporal analyses, the largest effect on a DeSO area's risk was time itself, most conspicuously during the second wave of the COVID pandemic, where interactive effects with having a foreign background or low rates of higher education, produced the largest risks. Overcrowding foremost interacted with time, and mainly during the first wave. Otherwise, overcrowding contributed the most to risks in DeSO zones marked by low economic standard, low education levels, or a high proportion of health care workers or gainfully employed persons. Through the descriptive results it was clear that in the three larger cities, spatial risk factors co-exist in certain boroughs and DeSO zones. Our results then suggest the following chain of events: healthcare workers and low salaried workers who can't work from home, tend to get exposed to a higher degree. Out of this group, those who are overcrowded and/or have a foreign background, tend to a higher degree live in poorer neighbourhoods, contributing to a larger degree of secondary infections (in their DeSO zones) due to more exposure sources and less ability to protect themselves.

Several joint factors may increase disease prevalence and mortality in the outbreak of a pandemic. Overcrowded households have been identified as important contributors to the presence of COVID-19, but the determining of factors for transition into, and within, households is still largely unknown, as are interactive effects from aspects of low socioeconomy. Within the home, physical closeness, sharing several surfaces, or lack of possibility to isolate persons who display symptoms, are the most probable causes. Since overcrowded households are related to poverty and poorer health (16), both baseline health and possibly housing conditions (poor ventilation or exposure to mould or rot) could further contribute to increased risks of severe illness (7). Determining factors outside the household are a high frequency of social contacts, especially if involving potentially infected persons, e.g., working health care jobs, commuting to work with public transport, or social activities, whereas only the latter can be considered as truly voluntary. An additional example could be other risk occupations which may be exuberated in vulnerable zones, e.g., taxi drivers who, during the beginning of the pandemic picked up infected passengers from at the airports and then returned back to their neighbourhoods and homes, increasing infection risks by a route of introducing infection from outside the DeSO zone or even from outside Sweden.

Our results emphasize that the predictive value of overcrowding clearly will have the largest effect in combination with other risk-variables. Two Swedish registers studies have for example noted that occupations with many social contacts, such as taxi driving, were not necessarily associated with increased mortality in older cohabitants (>67 years) (17, 18), not even in adjusted models. This indicates that risks mainly occur in the presence of other risk factors, presumably overcrowding, low education or if having a foreign background, which may relate to differences in living conditions and access to adequate information on disease protection. One may also think of university students or persons living in high rent cities, who often share accommodation to be able to live in central locations but will lack relation to other hazardous dimensions such as working in a risk occupation. This could explain why Stockholm, where the average rent per m^2 is the highest throughout the whole of Sweden, displays overcrowding in almost all DeSO-zones.

Aside from overcrowding and higher participation in jobs that cannot be performed from home, there may be particular aspects that increase exposure risks in persons with a foreign background. When COVID-19 was declared a full-scale pandemic in the spring of 2020, the awareness of the disease was high throughout society, but several international studies showed that crucial information about protection and risks of infection, did not reach everyone equally well. Through an interview-study with foreign-born workers in high risk occupations conducted by our project group (19), the lack of health literality, meaning access to and understanding of health information, seemingly played a part in having less ability to protect oneself and others. This ties to several aspects. An initial root cause was lack of information in other languages than Swedish, but also less knowledge of reliable sources to obtain such information or trust in official outlets. Studies on media consumption conducted in different boroughs in Gothenburg, showed that persons in geographical areas with low educational level and a high proportion of foreign-born persons, reported a higher usage of social media, foreign news media or social networks, as the main information source, especially in people who lacked proficiency in the Swedish language (20). This difficulty to access important societal information is sometimes referred to as "the knowledge gap", based on limited language skills or trust in authorities in the new country (21).

Due to a higher prevalence of existing poor health in areas with low socioeconomy, there are not only disparities in infection, but also larger risks for severe COVID-19 and post-COVID. According to the stages of disease theory, as proposed by Clouston (22), when new diseases arise, they transition through phases marking distinct patterns in mortality inequality, that emerge following the development of new information and mitigation strategies. More advantaged communities, such as those with less household overcrowding, can better implement resources that curb the spread and lethality of COVID-19. A widespread disease will therefore hit more disadvantaged groups the hardest at both the entrance of a pandemic, but also in the long-term aftermath, increasing already prevalent societal disparities in health.

Methodological considerations

Regarding modelling considerations, a question which emerged was whether it is the general overcrowding or socioeconomic situation in a person's DeSO which affects its risk of testing positive for COVID-19, or rather the person's household structure, or the combination of the two. To reveals such relationships, one could step away from interpreting the exponential term in the proposed Poisson regression setup as an individual's risk and instead consider e.g., a logistic regression model where the response variable is the indicator whether a given individual has tested positive for COVID-19. As covariates one would then include the spatial covariates which correspond to each individual, as well as different person-level covariates, most notably the dichotomised variable indicating whether a given individual lives in an overcrowded household or not. Here too, one could use an elastic-net regularised regression approach. Finally, the proposed approach was chosen to obtain an easily interpretable model in combination with variable/model selection and collinearity adjustment. This implies that we assume that all dependencies presented can be prescribed to the underlying covariate structure. However, since we are dealing with disease transmission for an infectious disease, the underlying DeSO covariates likely cannot explain the complete dynamics of the spread of the disease. Consequently, a classical spatio-temporal modelling approach which is based on models incorporating dependencies in the temporally evolving (spatial) multivariate response variable (a discrete random field) would be one way forward. Still, successfully incorporating variable selection techniques into such models remains a challenge.

Strengths and limitations

One of the larger limitations of our study is the misrepresentation of testing in the early stages of the pandemic (spring-summer 2020), which was almost exclusively conducted in health care workers or sever cases of COVID-19 that needed hospitalization. Those included in the early phase of the analyses might consequently constitute a sub-cohort of particularly vulnerable persons, due to an underreporting in the first wave of the pandemic. Another methodological limitation is that data on socioeconomic variables are based on information in 2018, and we cannot know to what extent the living conditions for individual persons and within DeSO zones have changed over the years 2020-2021. Furthermore, according to organizations for undocumented immigrants in Sweden, it has been approximated that about 10 000-35 000 illegal immigrants are residing in Sweden, most likely concentrated to the larger cities, Stockholm, Gothenburg and Malmö, where there are more possibilities to work and to live anonymously. While some are known to live in dwellings provided by illegal contractors e.g., in the construction industry, many rent rooms or live with relatives, which can be highly assumed to be more common in immigrant dense areas. It is, therefore, a limitation that our analyses are based on official statistics, likely underestimating overcrowding in high risk DeSO areas.

While most studies on overcrowding and infection risks are based on cohort data of a selection of a country’s inhabitants, our study include all registered residents in Sweden with associated spatial data on decisive socioeconomic factors such as housing, income, education, occupation, origin and car ownership. We also have access to all positive PCR-tests in Sweden and date for test results. Additionally, unlike many similar studies, we also have information on marital status/civic partnership and on age of all housing members and make calculations using a definition of overcrowding similar to the Eurostat-standard. The deviance in age boundaries between Eurostat's definition and our current definition is related to the format of Statistics Sweden's data, and we consider this difference to have rather small implications for the results.

In sum

Sweden has generally issued interventions or protective advice aimed at the Swedish population at large. However, already since the beginning of the pandemic there have been obvious differences in the prevalence and possible transmission mechanisms between societal groups. Many risk factors are therefore resulted by structural inequalities and involuntary exposure. Improved knowledge of sub-population characteristics may provide tailored strategies, targeting a multiple source of risk factors in high-risk geographical areas, hopefully contributing to protecting the inhabitants and slowing the spread of an emerging disease.

ACKNOWLEDGEMENT

This study was funded by The Swedish Research Council special grant for register-based research on the consequences of the COVID-19 pandemic (grant number: 2020-05792), and by the Swedish research council for heath, working life and welfare (grant number 2021-00326).

REFERENCES

1. Shannon H, Allen C, Clark M, Dávila D, Fletcher-Wood L, Gupta S. Web Annex A: Report of the systematic review on the effect of household crowding on health 2018 (WHO/CED/PHE/18.02). Geneva, Switzerland: World Health Organization.
2. World Health Organization. WHO housing and health guidelines. Geneva: World Health Organization. 2018. Report No.: 9241550376.
3. Magesh S, John D, Li WT, Li Y, Mattingly-App A, Jain S, et al. Disparities in COVID-19 outcomes by race, ethnicity, and socioeconomic status: a systematic-review and meta-analysis. AMA network open. 2021;4(11).
4. Hawkins RB, Charles EJ, Mehaffey JH. Socio-economic status and COVID-19–related cases and fatalities. Journal of Public health. 2020;189:129-34.
5. Baker MG. Who cannot work from home? Characterizing occupations facing increased risk during the COVID-19 pandemic using 2018 BLS data. Journal of Medrxiv. 2020.
6. Marmot M, Allen J. COVID-19: exposing and amplifying inequalities. Journal of Epidemiology and Community Health. 2020;74(9):681-2.
7. Statistiska Centralbyrån (Statistics Sweden). Sverige har flest trångbodda i Norden https://www.scb.se/hitta-statistik/artiklar/2021/sverige-har-flest-trangbodda-i-norden/ 2021
8. Aldridge RW, Pineo H, Fragaszy E, Eyre M, Kovar J, Nguyen V, et al. Household overcrowding and risk of SARS-CoV-2: analysis of the Virus Watch prospective community cohort study in England and Wales. Journal of medRxiv. 2021.
9. Kamis C, Stolte A, West JS, Fishman SH, Brown T, Brown T, et al. Overcrowding and COVID-19 mortality across US counties: Are disparities growing over time? Population Health. 2021:100845.
10. Galvani AP, May RM. Dimensions of superspreading. Nature. 2005;438(7066):293-5.
11. Alderling M, Albin M, Ahlbom A, Alfredsson L, Lyström J, Selander J. Risken att sjukhusvårdas för covid-19 i olika yrken. Rapport 2021:02. Stockholm: Centrum för arbets- och miljömedicin; 2021.
12. Molenaar J. The COVINFORM project - Migrants as a ‘vulnerable group’ in the COVID-19 pandemic: A European perspective. Migration and Health seminars 2021.
13. Chen JT, Krieger N. Revealing the unequal burden of COVID-19 by income, race/ethnicity, and household crowding: US county versus zip code analyses. Journal of Public Health Management Practice. 2021;27(1):S43-S56.
14. International Standard Classification of Occupations 2008 (ISCO-08): Structure, group definitions and correspondence tables. Geneva: International Labor Organization; 2012.
15. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: Springer; 2013.
16. Krieger N, Waterman PD, Chen JT. COVID-19 and overall mortality inequities in the surge in death rates by zip code characteristics: Massachusetts, January 1 to May 19, 2020. American journal of public health. 2020;110(12):1850-2.
17. Billingsley S, Brandén M, Aradhya S, Drefahl S, Mussino E, Andersson G. Deaths in the frontline: Occupation-specific COVID-19 mortality risks in Sweden. University of Stockholm; 2020:36.
18. Billingsley S, Brandén M, Aradhya S, Drefahl S, Andersson G, Mussino EH. COVID-19 mortality across occupations and secondary risks for elderly individuals in the household: A population register-based study. Scandinavian Journal of Work, Environment and Health. 2021;48(1):52-60.
19. Söderberg M, Swaid J, Jakobsson K, Magnusson M, Rosengren A. En intervjustudie bland utlandsfödda arbetare i högriskyrken för covid-19. Göteborgs universitet; 2021.

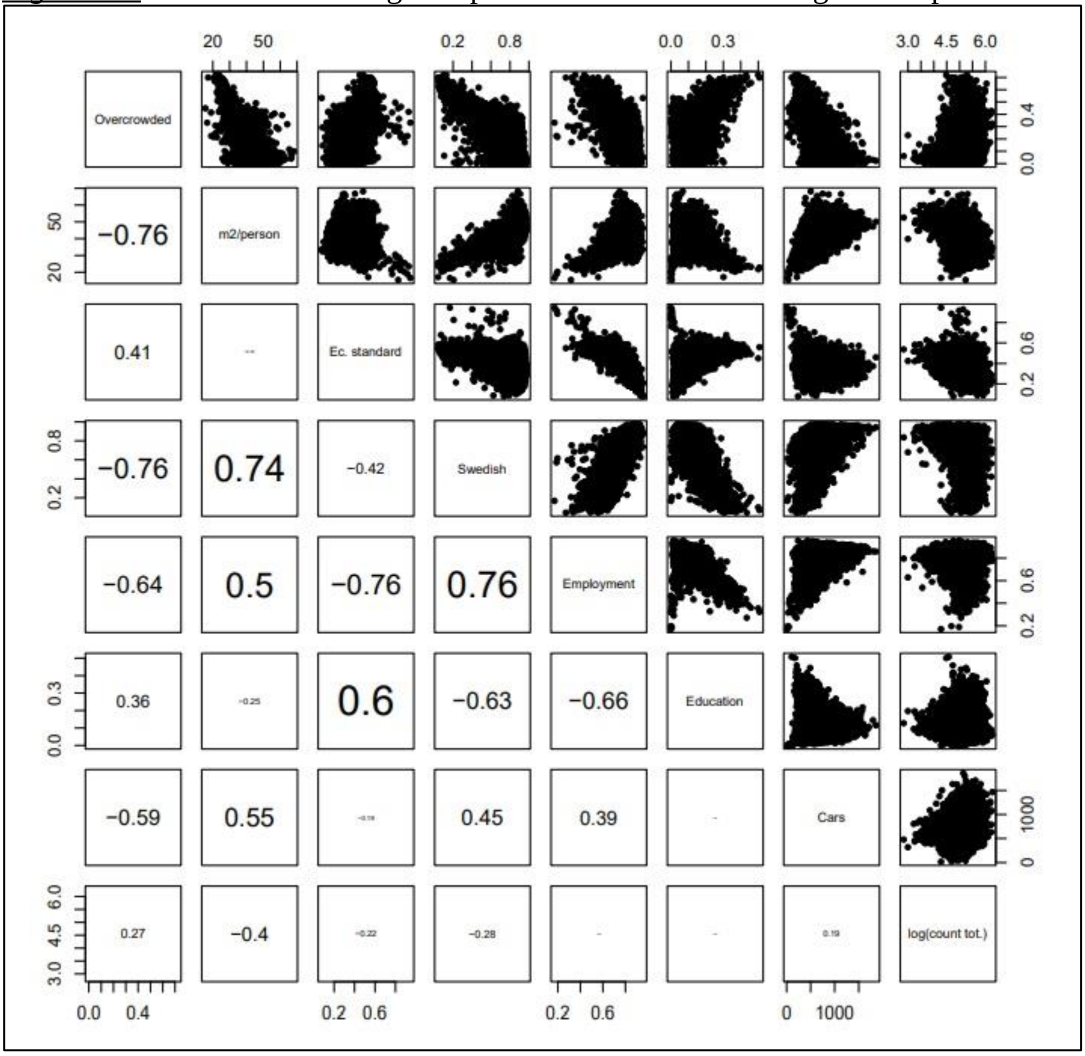
20. Esaiasson P, Johansson B, Ghersetti M, Sohlberg J. Kriskommunikation och segregation i en pandemi. Hur boende i utsatta områden informerade sig om coronaviruset våren 2020. Institutionen för journalistik, medier och kommunikation, Göteborgs universitet; 2020.
21. Spence PR, Lachlan KA, Burke JA. Differences in crisis knowledge across age, race, and socioeconomic status during Hurricane Ike: A field test and extension of the knowledge gap hypothesis. Journal of Communication Theory. 2011;21(3):261-78.
22. Clouston SA, Rubin MS, Phelan JC, Link BG. A social history of disease: Contextualizing the rise and fall of social inequalities in cause-specific mortality. Demography. 2016;53(5):1631-56.

APPENDIX I

Expanded statistical methodology with reasoning for chosen methods

As an exploratory analysis, we started by evaluating dependencies among the spatial covariates, as well as between the log-counts per DeSO zone and these covariates, both accumulated over the whole study period and broken down monthly, excluding DeSO zones with zero counts. The idea of analysing the log-counts stems from the idea of evaluating how well the covariates would fit the raw counts using a Poisson regression model. The results can be found in Figure A1, where we see that there is a high degree of collinearity among the covariates, which is to be expected, and that the correlations between the log-counts and the individual covariates are quite small, implying that it is unlikely that only the spatial covariates would explain the log-counts over time. This is in line with our hypothesis that, in terms of spatial covariates, it is the combination of several different factors rather than specific individual factors that drive the risk of becoming affected. Concerning the collinearity, we see, for instance, that overcrowding is more prevalent in DeSO zones where there is a higher degree of socioeconomic vulnerability, reflected by e.g., a high proportion of persons that are unemployed, with low economic standard and with an education level of ≤9 years.

Figure A1. Correlations among the spatial covariates and the log-counts per DeSO zone.



Proceeding to the correlations between the log-counts and the interaction terms (not reported), we found that some interaction terms correlated more strongly with the log-counts than the individual covariates, thus motivating their inclusion in the regression modelling. Note that the inclusion of interaction terms tends to introduce additional collinearity.

In Appendix IV, we further illustrate the same kind of correlation analysis but with the monthly counts. Here the message is essentially the same, but we see that the dependencies between the non-zero log-counts and the different covariates change with time.

Turning to the spatio-temporal setting, we further estimated correlations between the spatial covariates, as well as correlations between the spatial covariates and the log-counts per DeSO, per month, excluding observation with count 0, see Appendix IV. Also, here we saw that the influence of a given covariate on the log-counts tends to change with time.

Hence, the overall conclusion in our correlation analyses was that there are clear collinearities present and that the influence of a given (interaction term) covariate on the log-counts tends to change with time.

Regression

To have all covariate variables on a comparable scale, they were standardised by first subtracting the individual means and then diving by the individual standard deviations. Moreover, considering the observed collinearity between this list of covariates, we additionally included all (second order) interaction terms. In order to obtain a model for a person's individual risk, given the DeSO-based data at hand, we modelled the DeSO counts (y_i) using Poisson regression models with the population size (n_i) as offset: $y_i = n_i$ *exp(linear combination of the individual DeSO-based covariates and all of their interaction terms)*. The underlying idea here was that, within a given DeSO zone, there is a given number of people (n_i), who can become infected, and the exponential term then essentially describes the "risk" that an individual gets infected. However, as we studied the first 18 months of the pandemic, during which one or more of the n_i individuals in the i^{th} DeSO may become re-infected, this risk representation should be interpreted somewhat loosely and the risk will not necessarily fall between 0 and 1. We included all possible interaction terms since we want to study how, in addition to an individual covariate's potential influence on the counts, interplays between the covariates affect the predicted risk of having high counts, relative to the population size.

Since there were clear collinearities present, which we could deduce from the correlation plots, we needed to adjust for collinearity. Moreover, one of our main objectives was to find out which single, or combinations of covariates, that had an actual predictive influence on the counts. To fit our models to the data, we chose a prediction-based approach, as opposed to a classical statistical inferential approach, since, arguably, this most naturally corresponds to the notion of a person's individual risk. In order to handle these two challenges jointly, we carried out elastic net regularised regression. This enabled fitting a Poisson regression model while simultaneously i) adjusting for collinearity and ii) carrying out variable selection, i.e., covariates with little/negligible predictive power are excluded from the model. Somewhat simplified, this implies that we assumed that dependencies between DeSO zone counts, which e.g., result from disease transmission, were solely driven by to the underlying covariate structure (one may think of this as a kind of conditional independence).

A benefit with this approach is that we obtained a familiar and easily interpretable Poisson regression model, in contrast to e.g., different machine learning approaches which exhibit even more flexibility, although at the cost of interpretability, or a classical spatial (-temporal) areal data/discrete random field model, which does not allow for the same kind of joint collinearity adjustment and variable selection. A Poisson regression model, as such, which is motivated by the study of independently occurring events, is arguably a bit simplistic for fully describing our data. However, by including an elastic net penalty and framing the resulting regularised regression problem within the framework of cross-validation-based statistical learning, we ensured that we fitted the Poisson regression model in such a way that it predicts hold-out data as well as possible, within the limits of the Poisson regression model framework's limitations.

Further, since all spatial covariates and their interactions were standardized, we were able to study variable importance in the fitted models. Hence, a larger estimated coefficient would be interpreted as a higher contribution to the fitted Poisson regression model for the expected counts, since all standardised covariates were entered with the same scale. Note further that the K-fold cross-validation, which is used for the hyperparameter selection, was used both for the penalty scaling parameter and the internal elastic net parameter, alpha, which determines how much weight is prescribed to lasso penalisation, on the one hand, and ridge penalisation, on the other hand. Here, we followed standard practice and set K=10; we have also evaluated K=5, which yielded worse out of sample prediction performance.

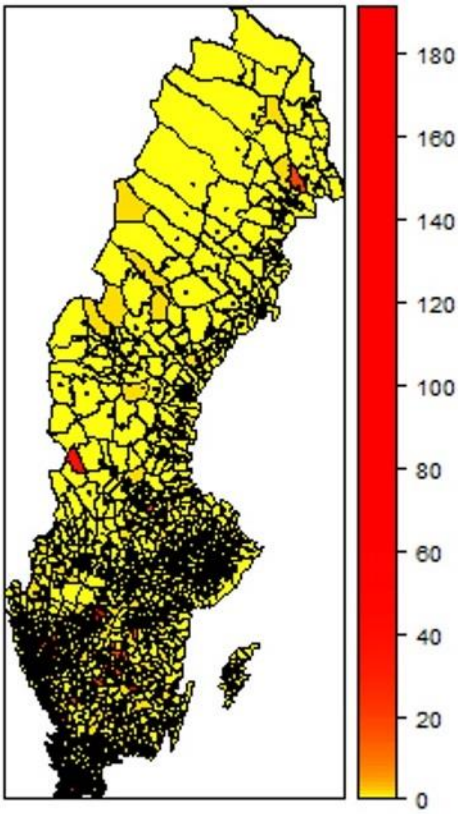
In the purely spatial analysis, where a count y_i corresponded to the aggregation of cases occurring in the underlying DeSO zone throughout the study period, we obtained an internal elastic net parameter of 0.5, which meant that we put equal weight on ridge regression and lasso regression, and a penalisation parameter value of 0.4039689, using the 1 standard error rule. Our spatio-temporal extension was achieved by adding a dummy variable for each calendar month, while retaining the rest of the covariates and interaction terms in the model. We emphasise that the chosen approach allows us to study both individual covariate's potential influence on the counts, as well as how interplays between the covariates affect the predicted risk of having high counts, relative to the population size. Here the internal elastic net parameter became 0.6 while the penalisation parameter became 0.009060382, using the 1 standard error rule.

One of the main messages here was that the dependence between the monthly log-counts and the individual covariates tend to change over time, thus indicating that a spatio-temporal analysis may reveal intricacies and that the statement that a single factor has a strong constant impact on the risk for an individual may be too simplistic.

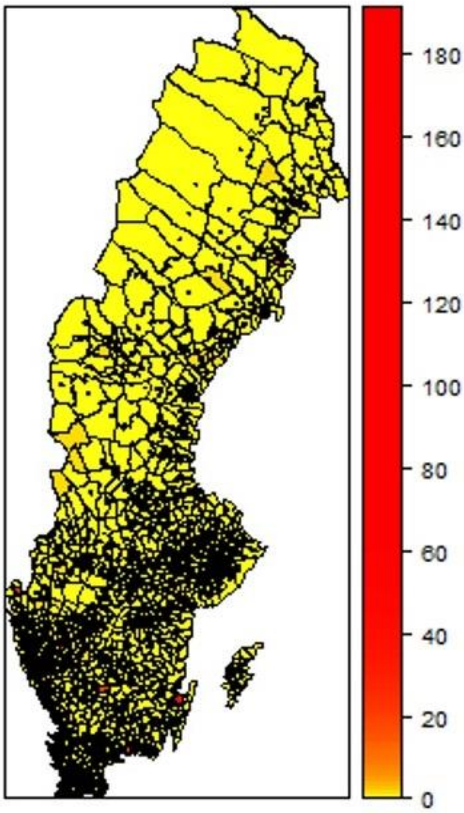
APPENDIX II

Monthly counts of COVID-19 cases in Sweden (January 2020 - June 2021)

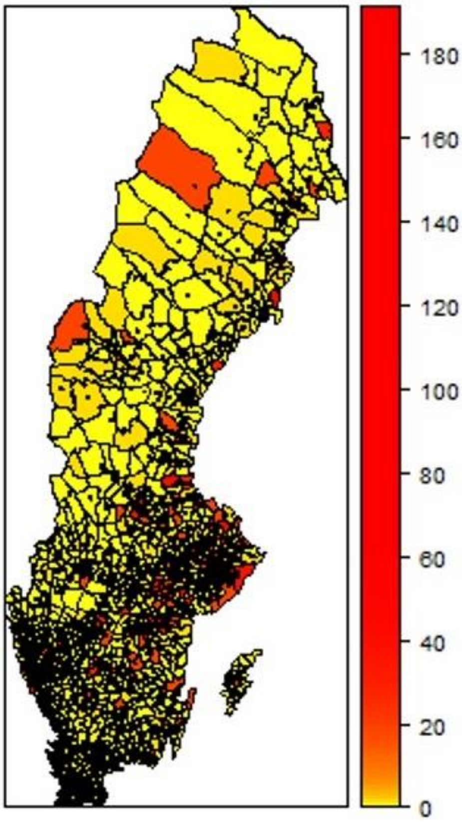
Month 1 (January 2020)



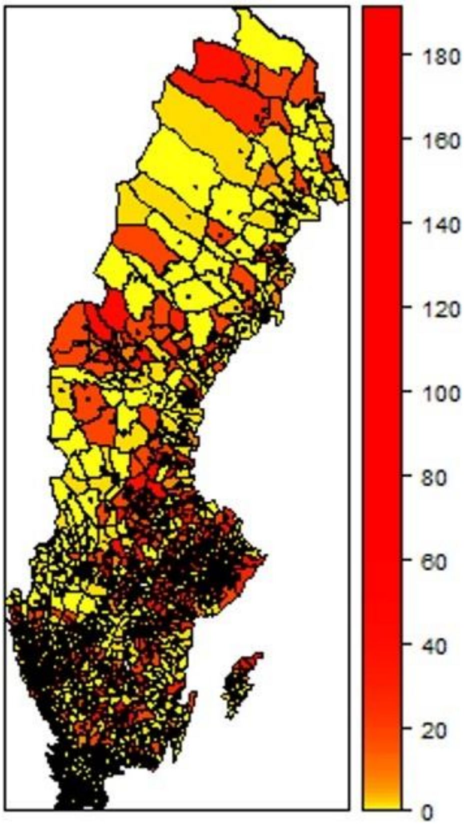
Month 2 (February 2020)



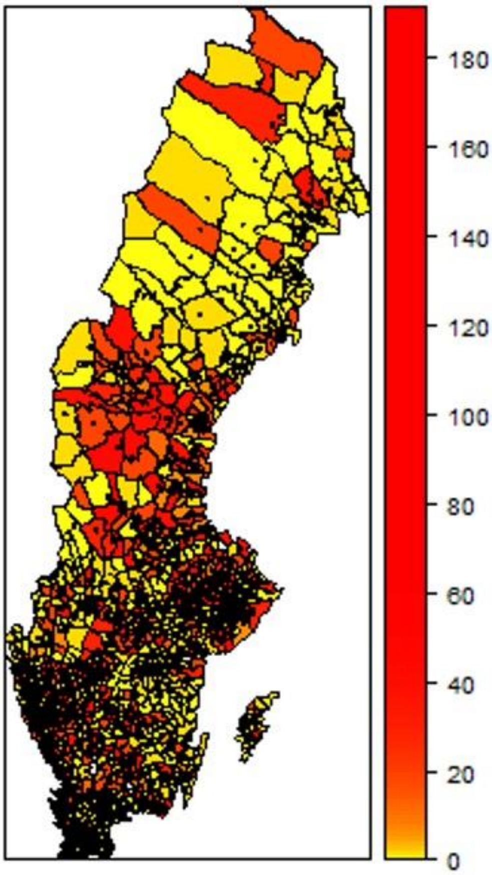
Month 3 (March 2020)



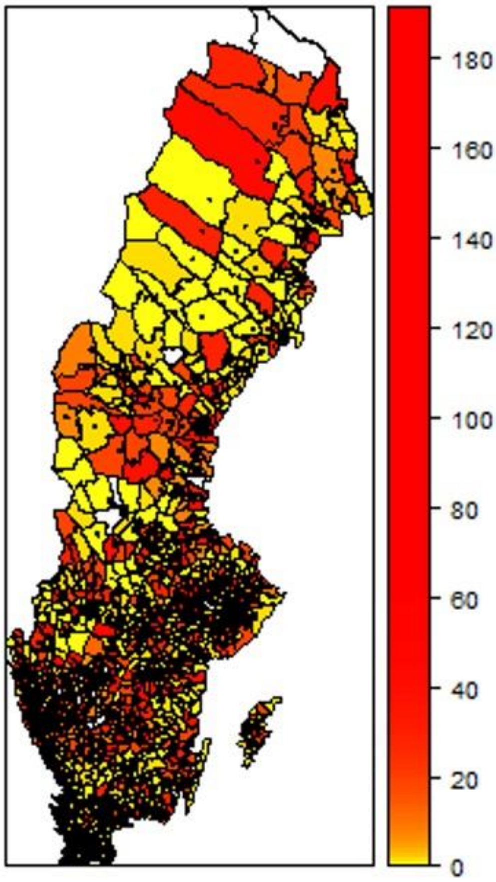
Month 4 (April 2020)



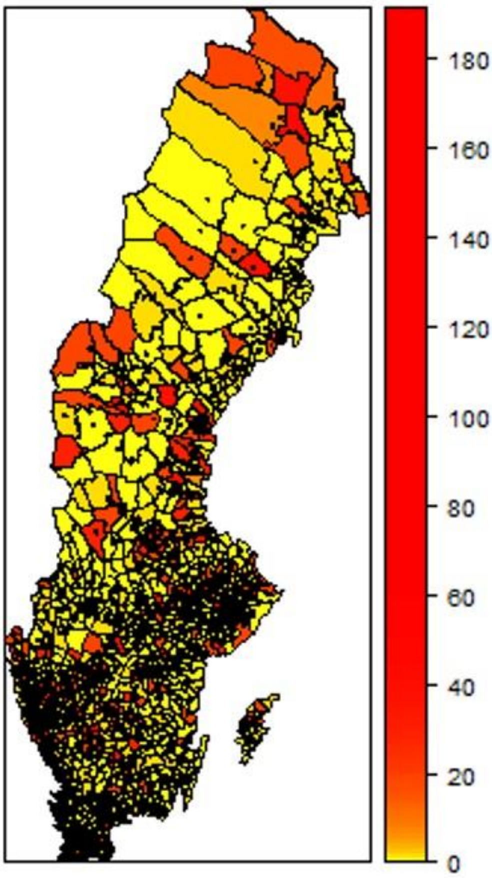
Month 5 (May 2020)



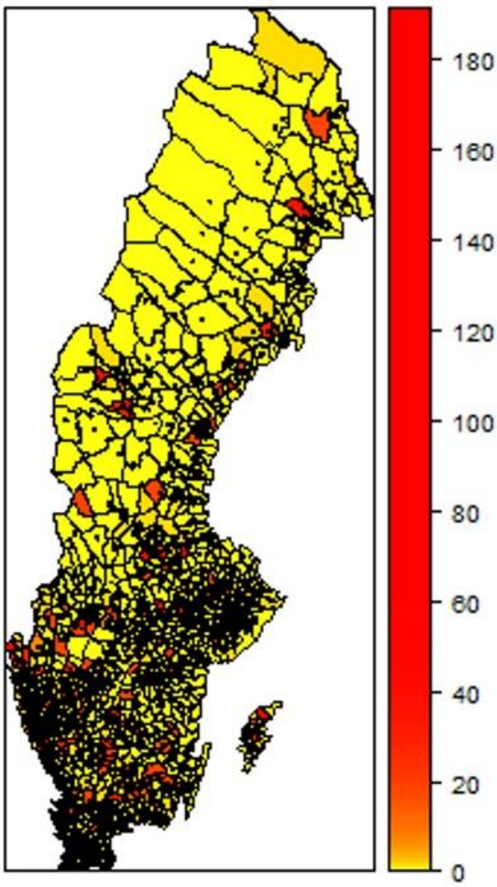
Month 6 (June 2020)



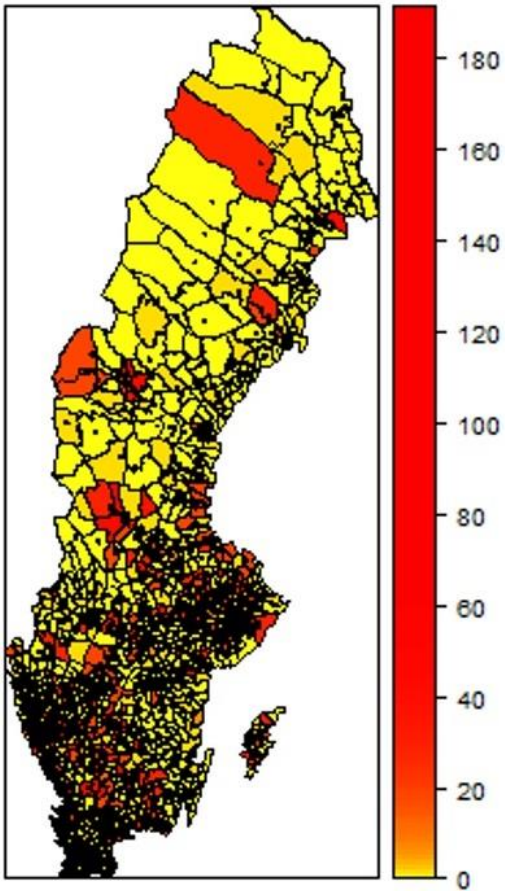
Month 7 (July 2020)



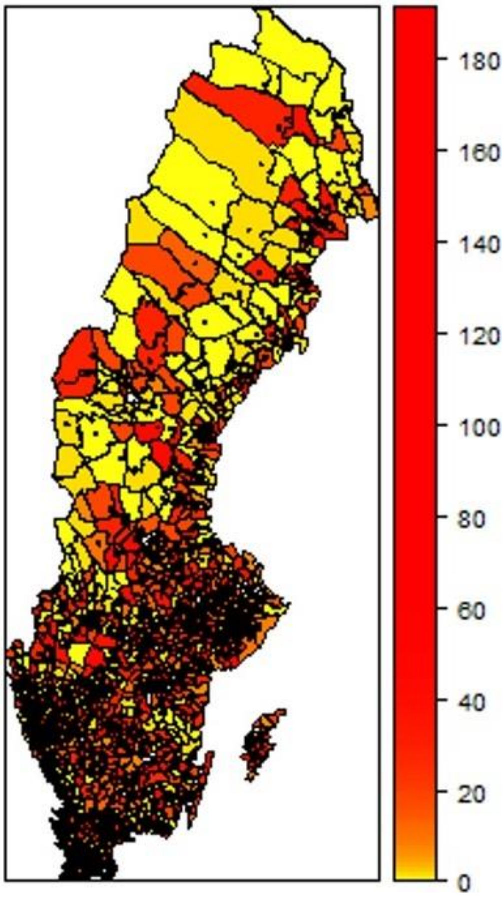
Month 8 (August 2020)



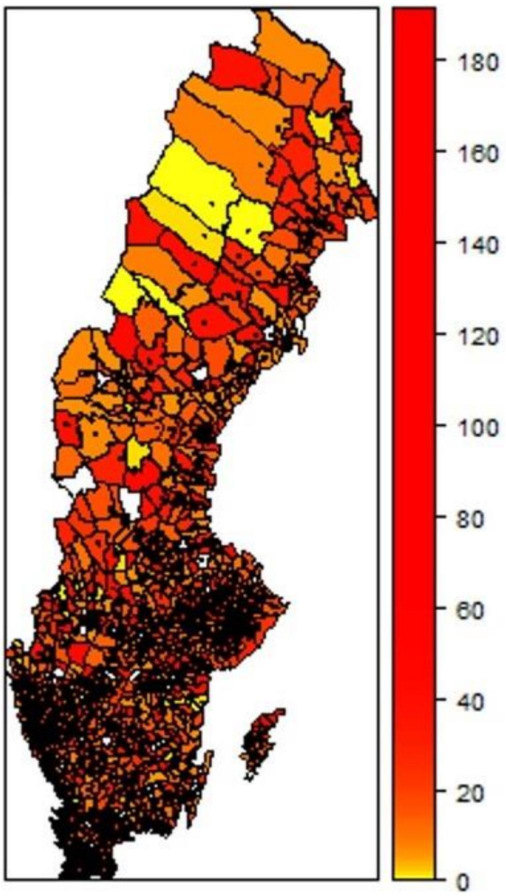
Month 9 (September 2020)



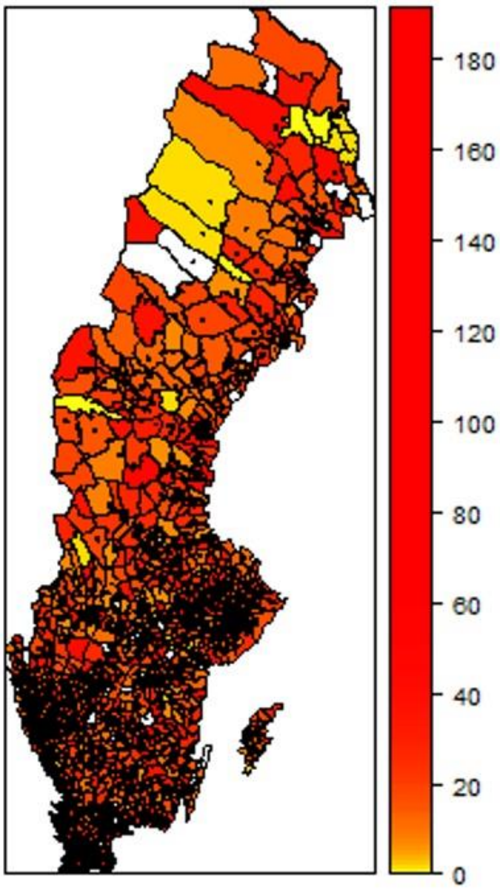
Month 10 (October 2020)



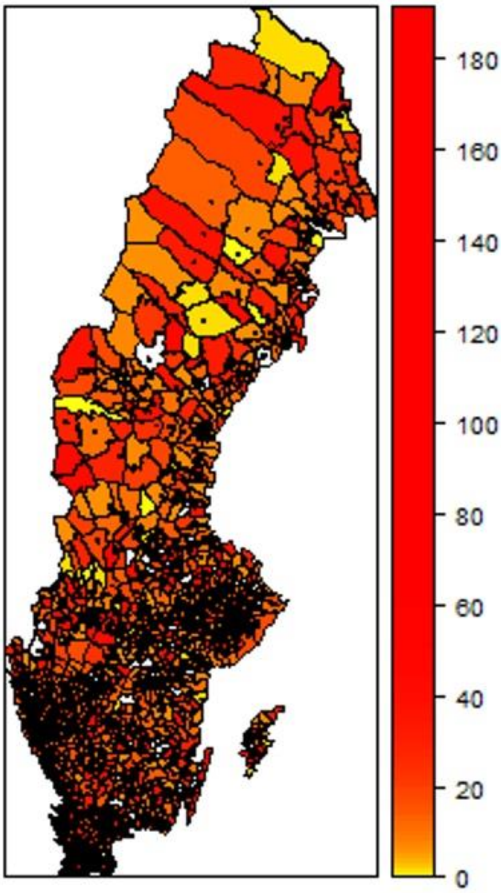
Month 11 (November 2020)



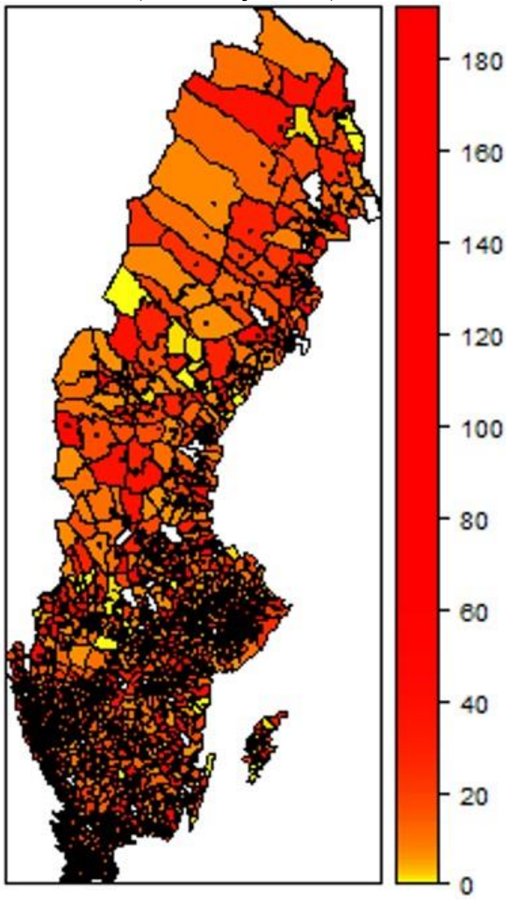
Month 12 (December 2020)



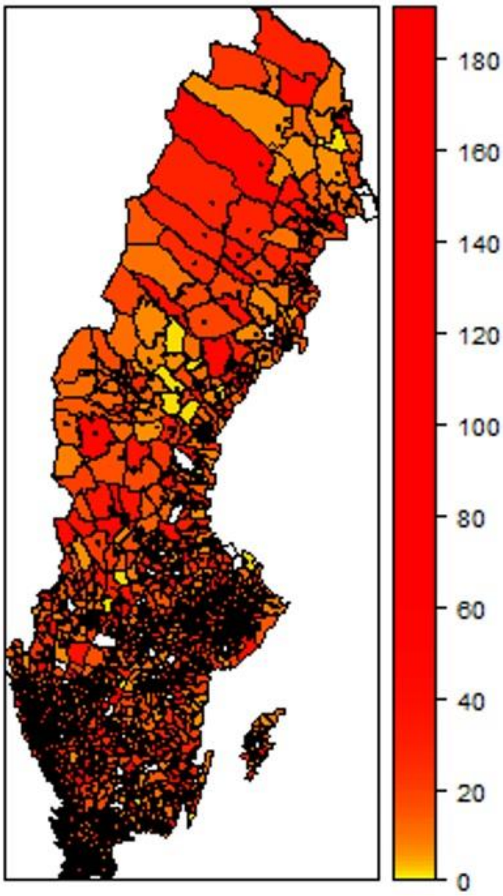
Month 13 (January 2021)



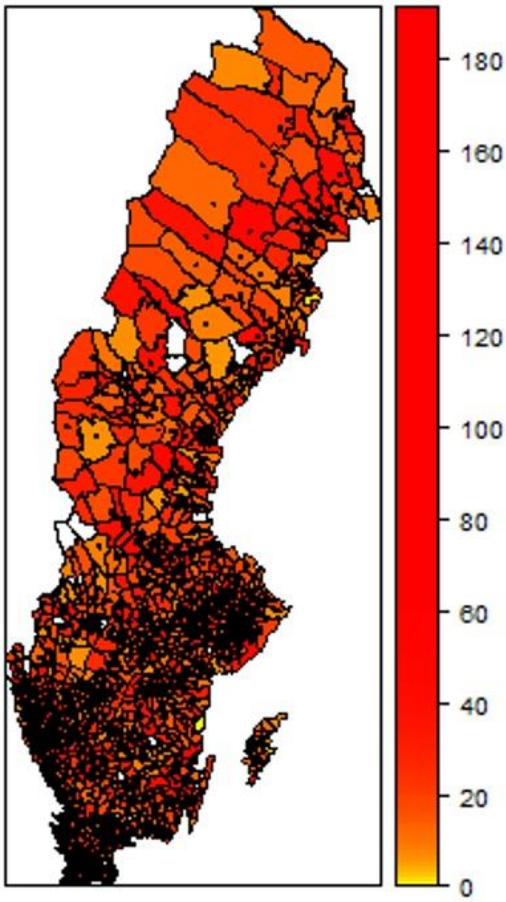
Month 14 (February 2021)



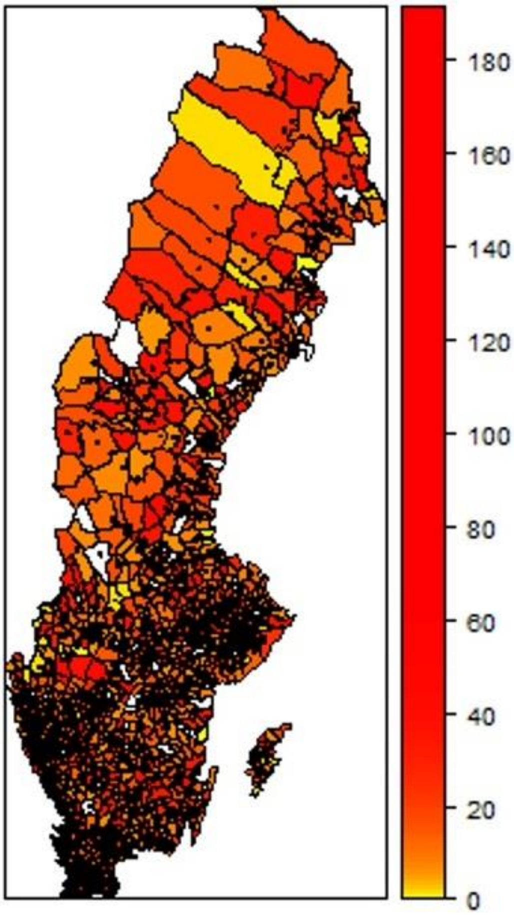
Month 15 (March 2021)



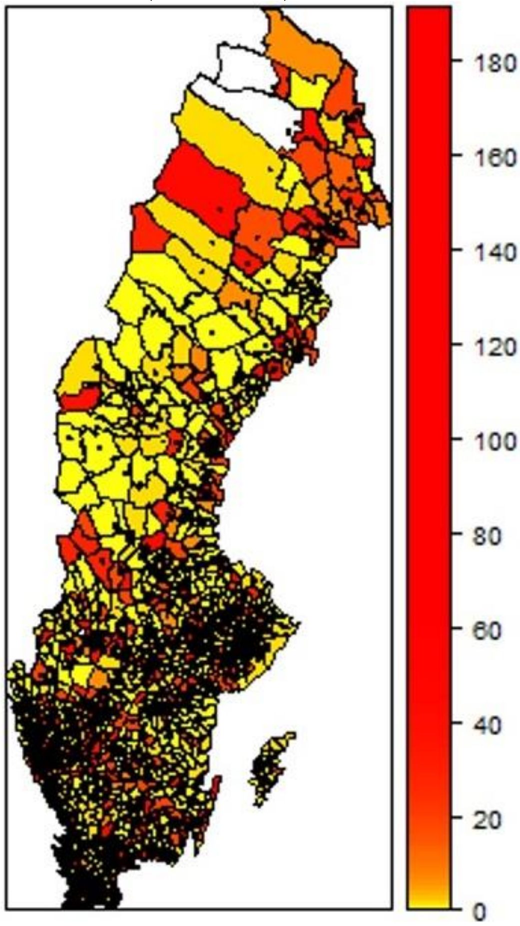
Month 16 (April 2021)



Month 17 (May 2021)

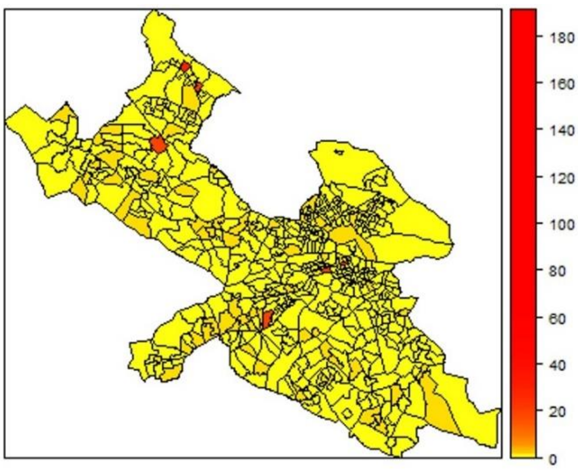


Month 18 (June 2021)

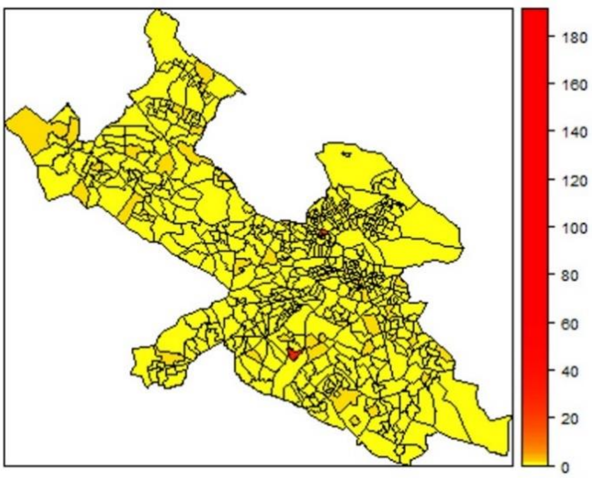


Monthly counts of COVID-19 cases in Stockholm (January 2020 - June 2021)

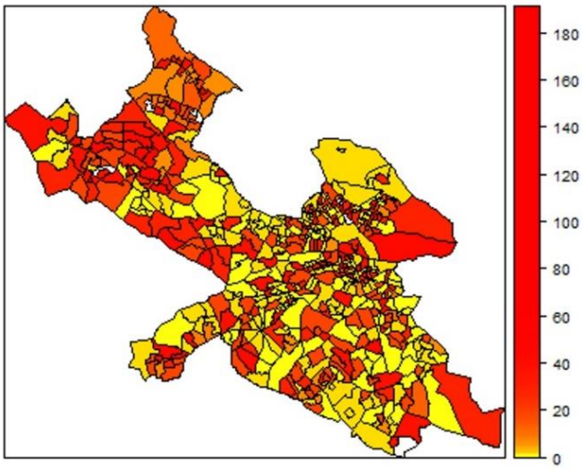
Month 1 (January 2020)



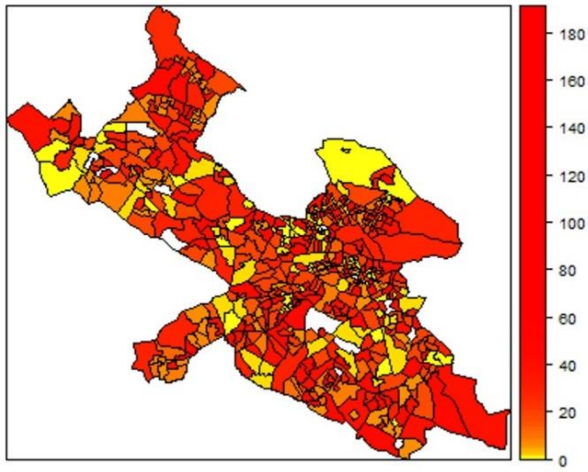
Month 2 (February 2020)



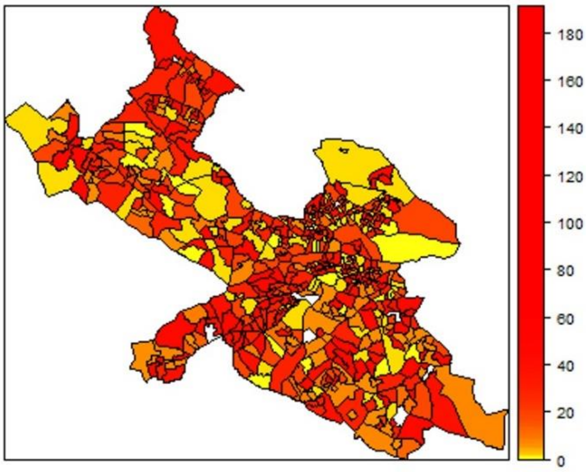
Month 3 (March 2020)



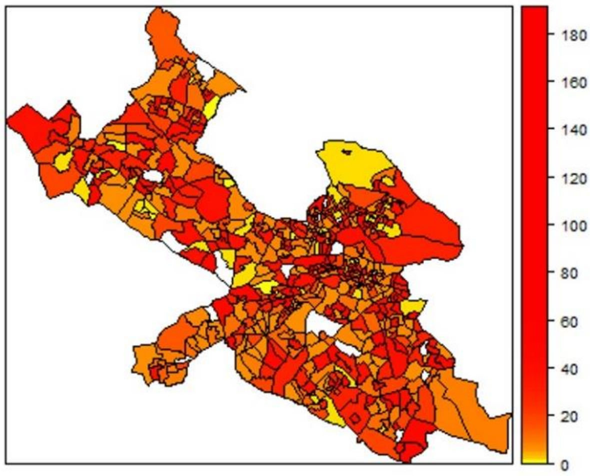
Month 4 (April 2020)



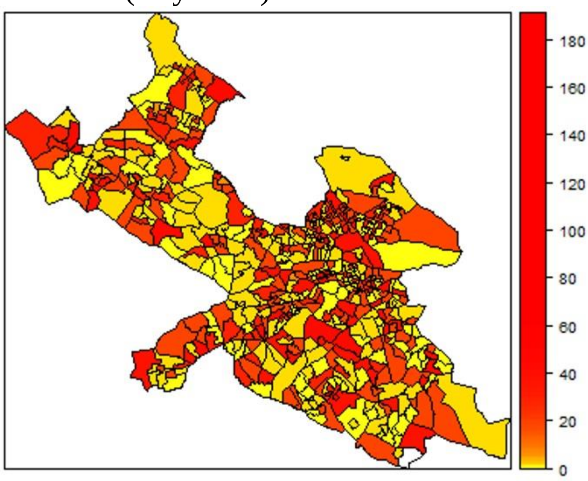
Month 5 (May 2020)



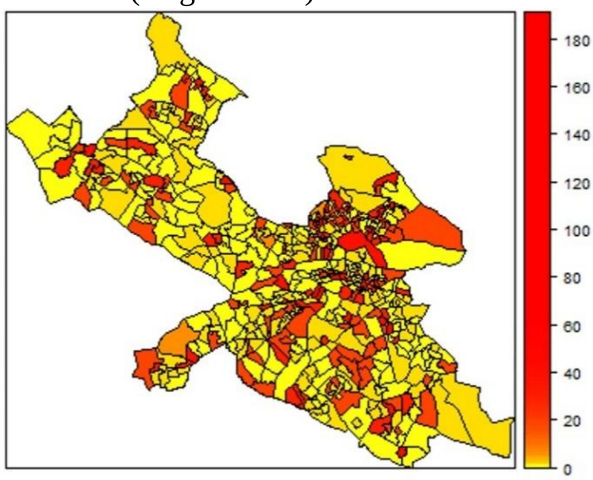
Month 6 (June 2020)



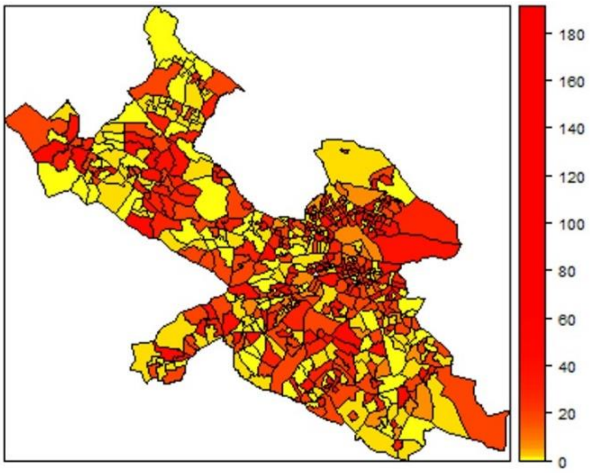
Month 7 (July 2020)



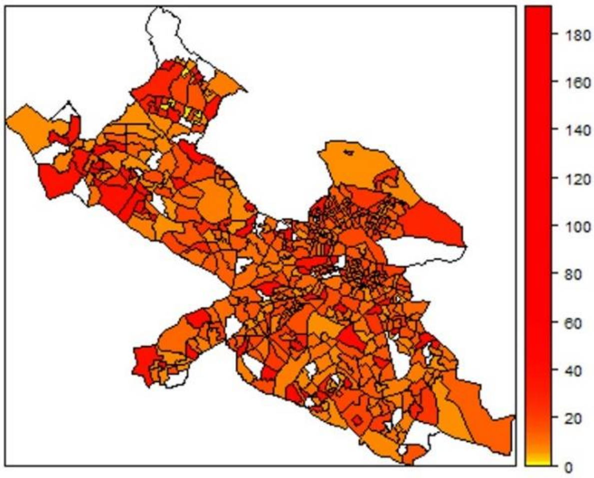
Month 8 (August 2020)



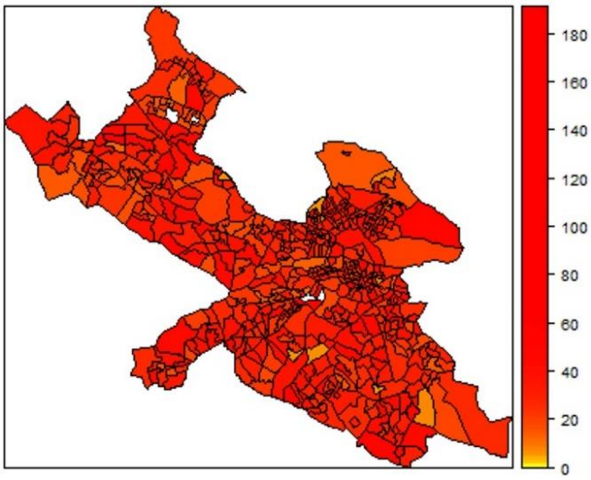
Month 9 (September 2020)



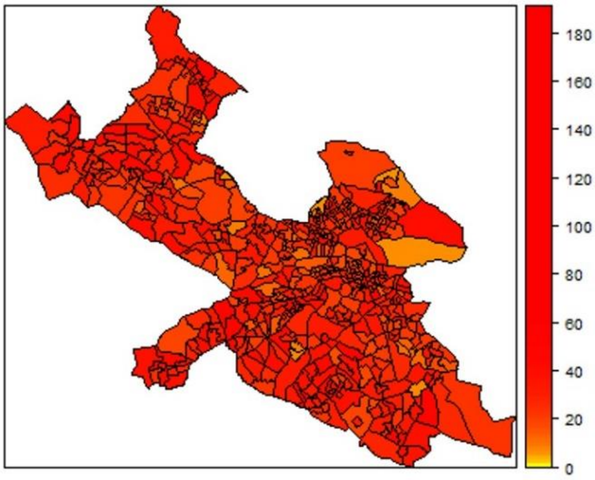
Month 10 (October 2020)



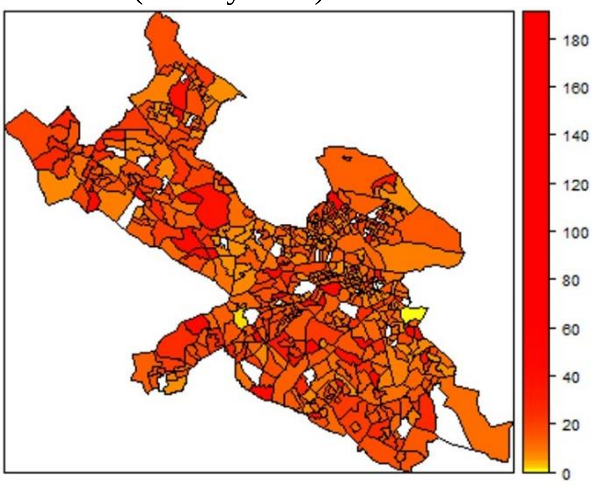
Month 11 (November 2020)



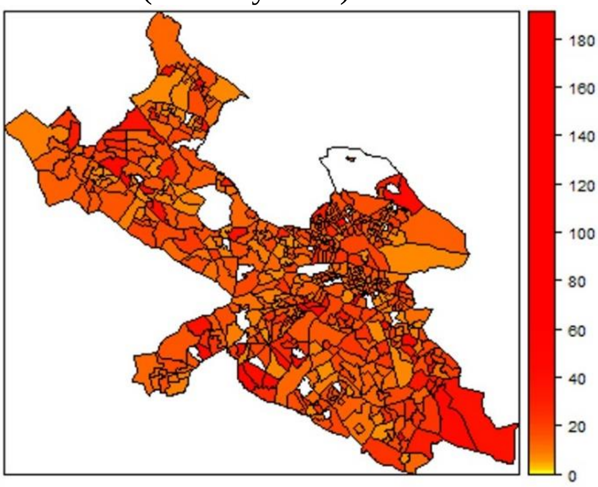
Month 12 (December 2020)



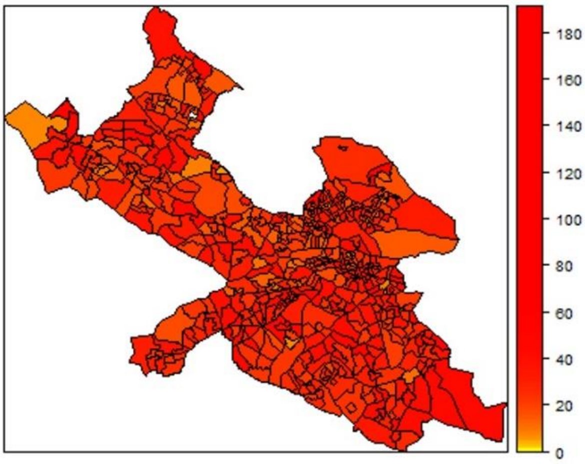
Month 13 (January 2021)



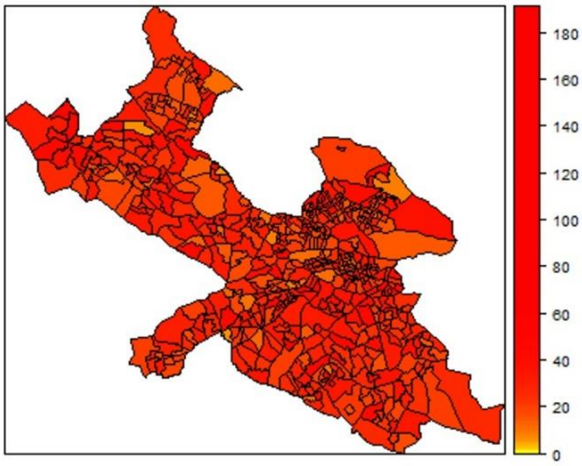
Month 14 (February 2021)



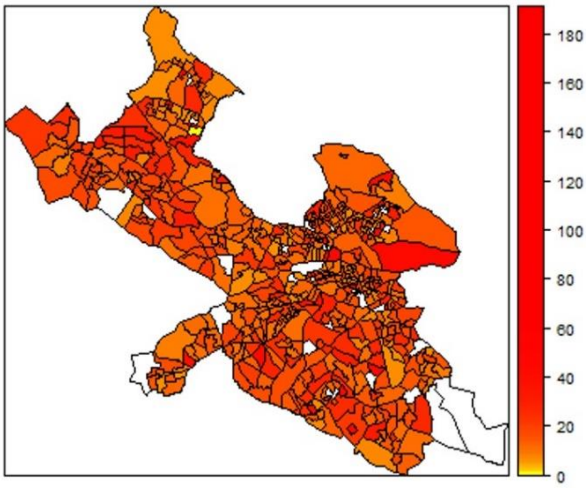
Month 15 (March 2021)



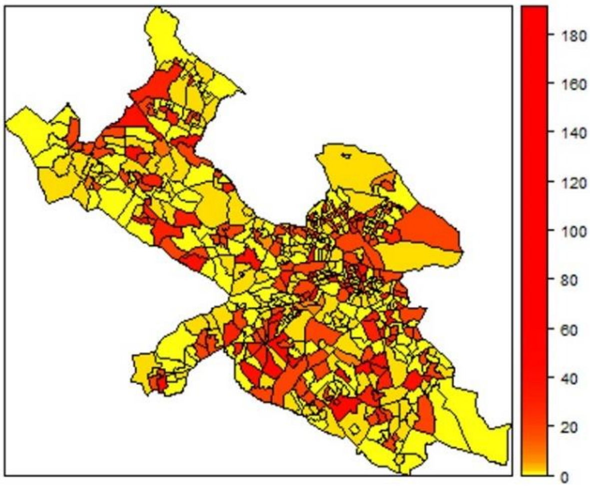
Month 16 (April 2021)



Month 17 (May 2021)

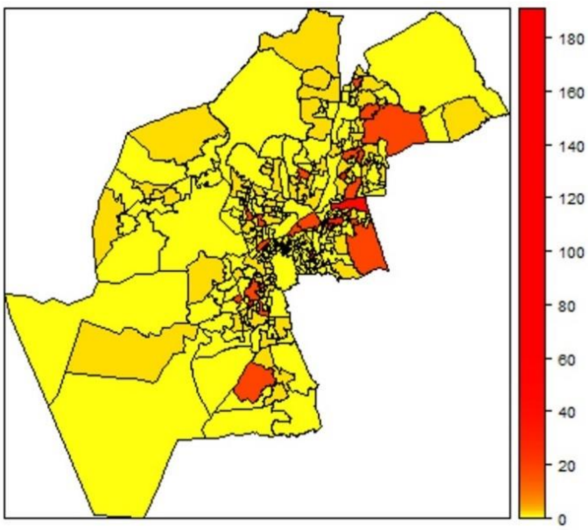


Month 18 (June 2021)

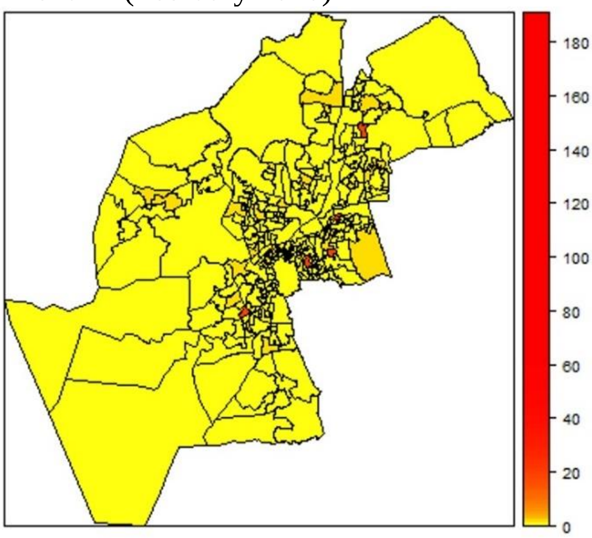


Monthly counts of COVID-19 cases in Gothenburg (January 2020 - June 2021)

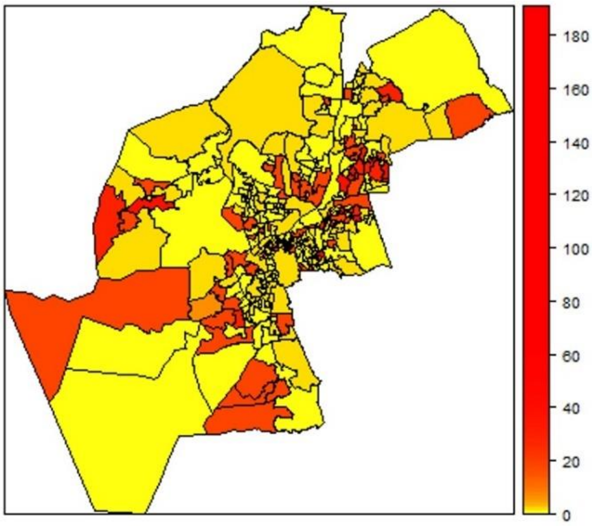
Month 1 (January 2020)



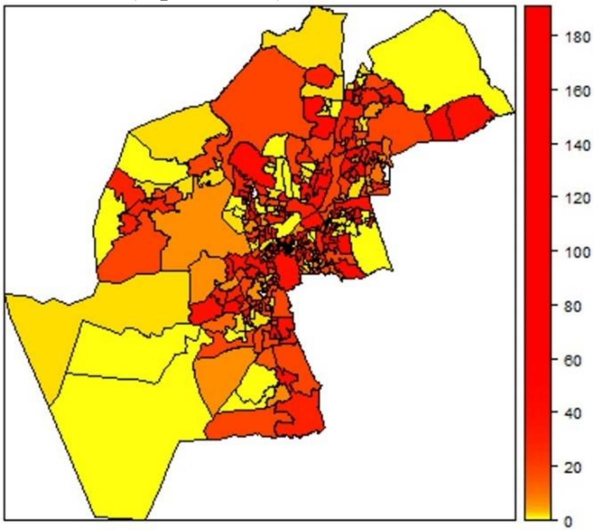
Month 2 (February 2020)



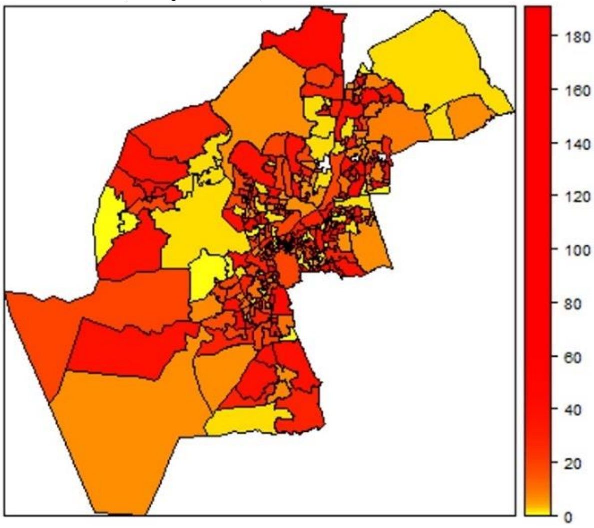
Month 3 (March 2020)



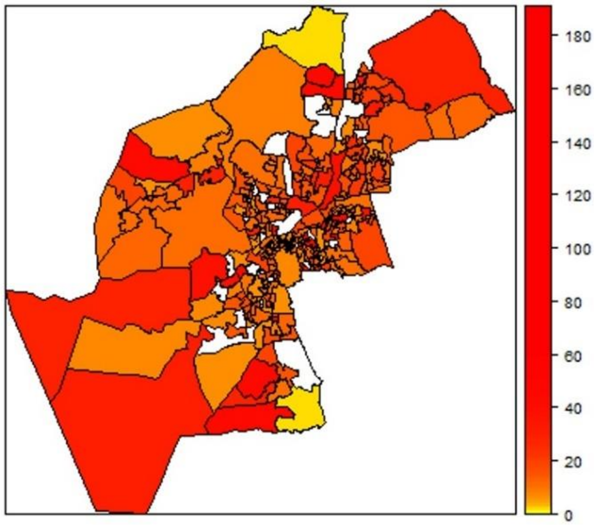
Month 4 (April 2020)



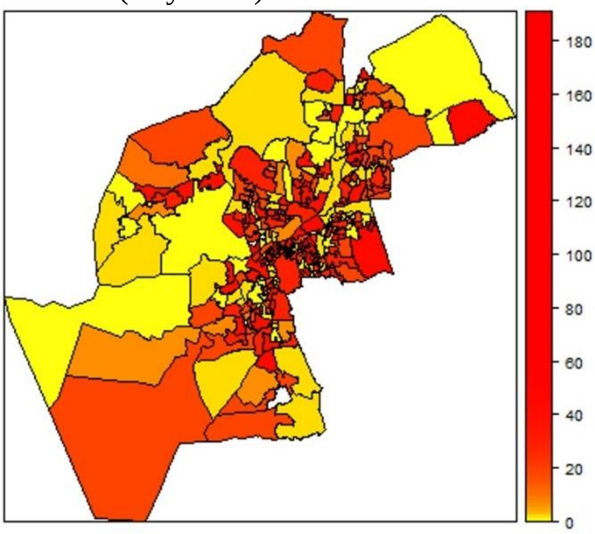
Month 5 (May 2020)



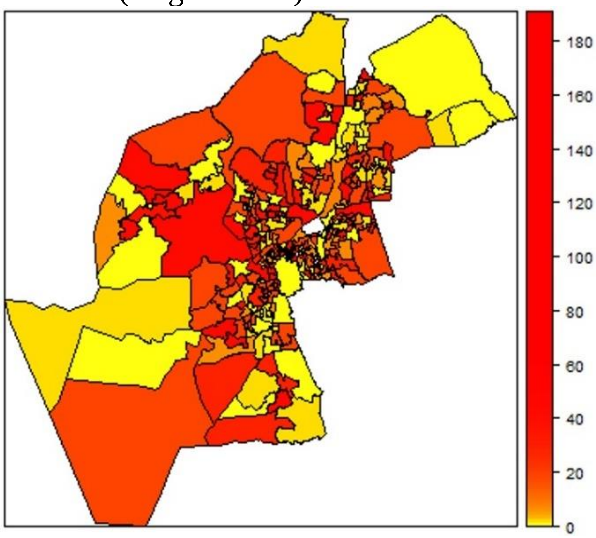
Month 6 (June 2020)



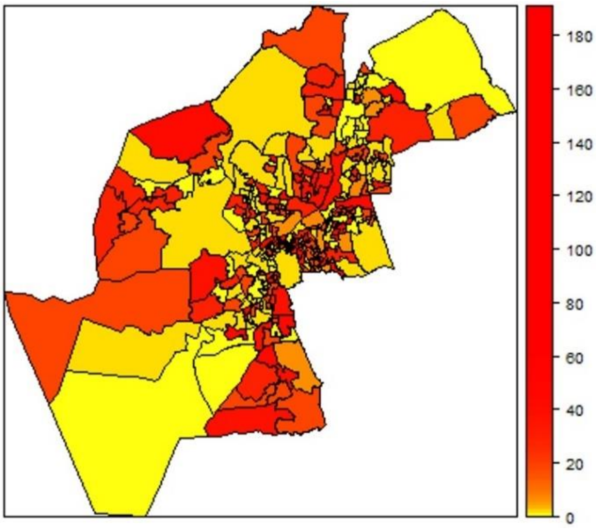
Month 7 (July 2020)



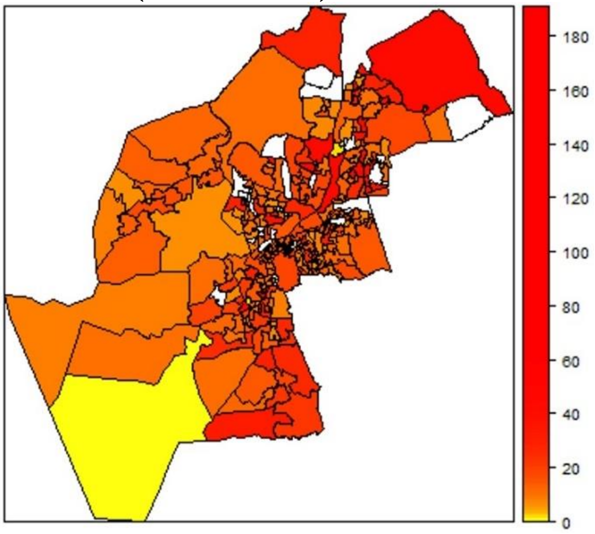
Month 8 (August 2020)



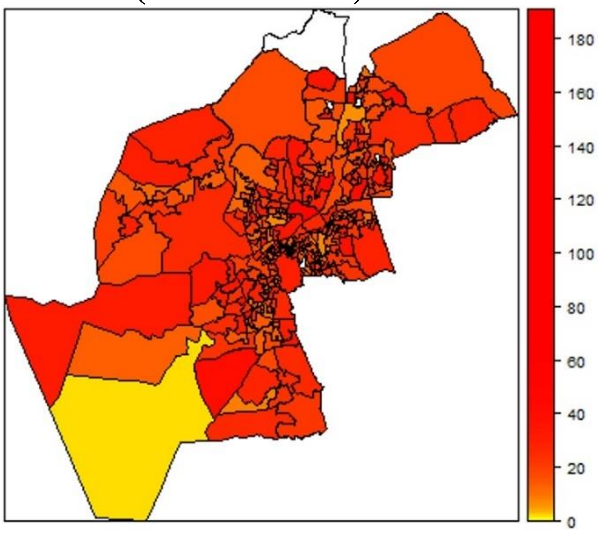
Month 9 (September 2020)



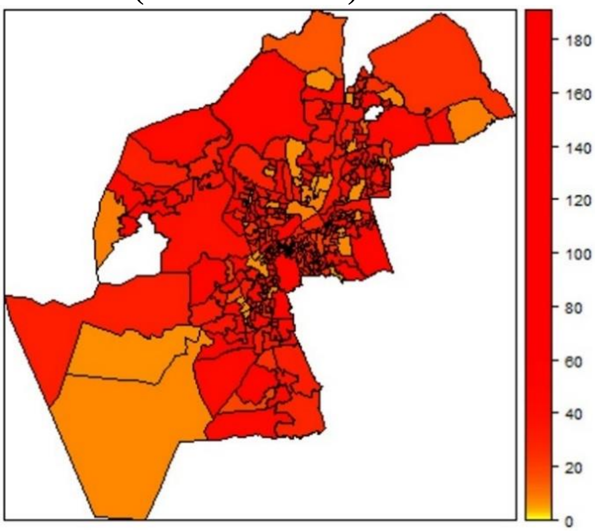
Month 10 (October 2020)



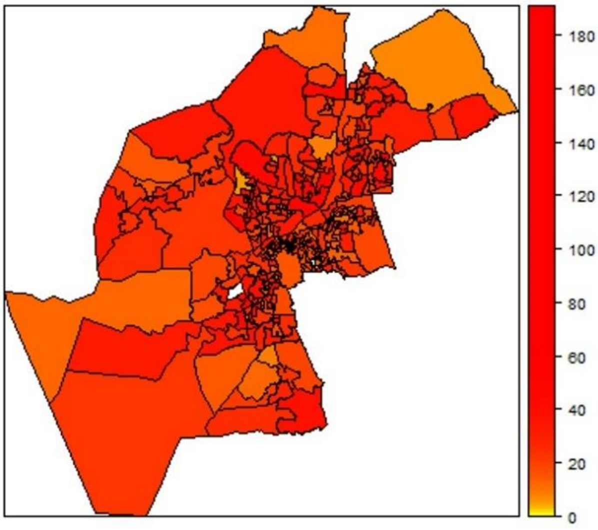
Month 11 (November 2020)



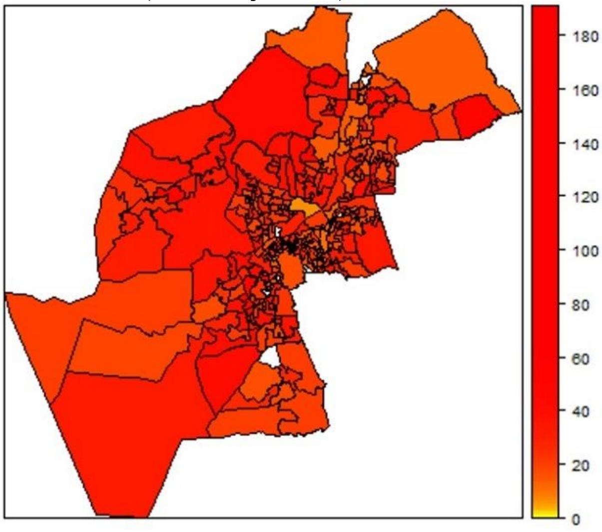
Month 12 (December 2020)



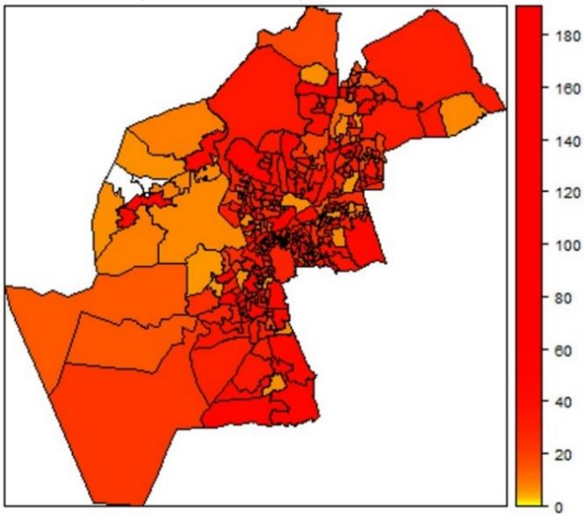
Month 13 (January 2021)



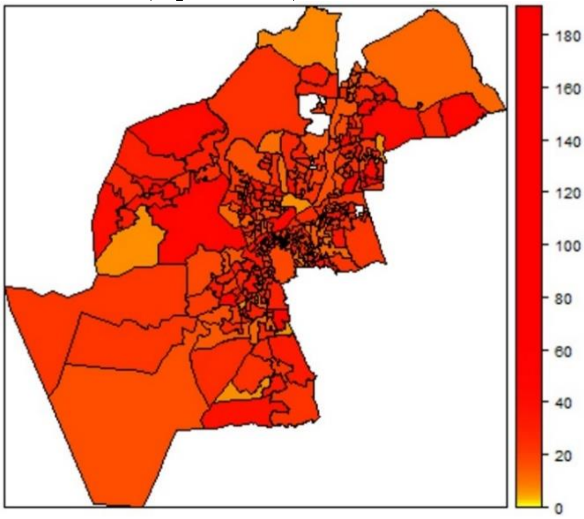
Month 14 (February 2021)



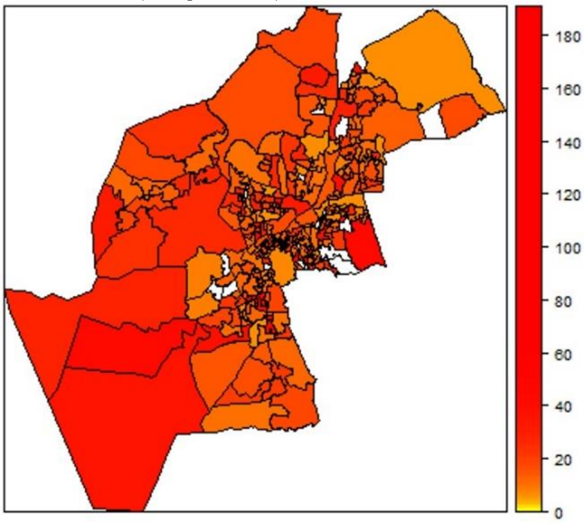
Month 15 (March 2021)



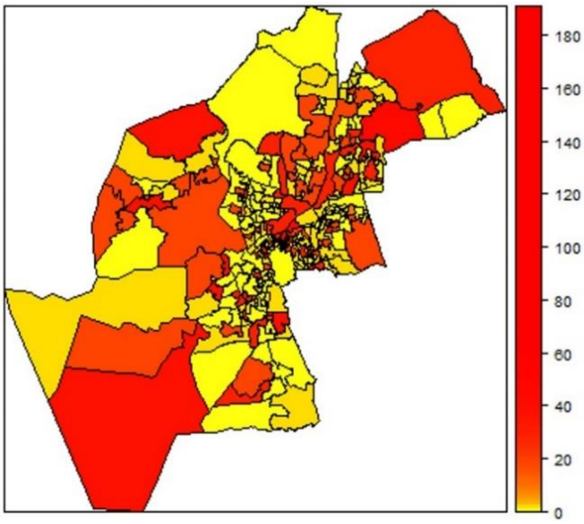
Month 16 (April 2021)



Month 17 (May 2021)

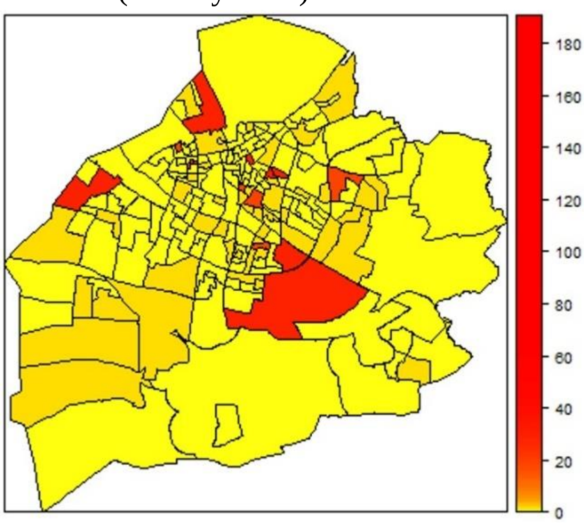


Month 18 (June 2021)

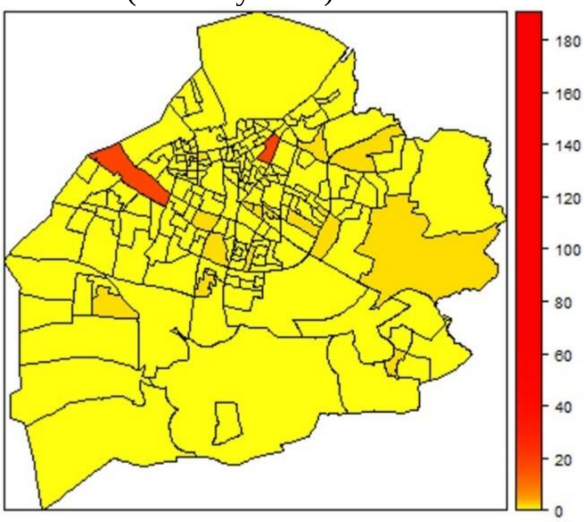


Monthly counts of COVID-19 cases in Malmö (January 2020 - June 2021)

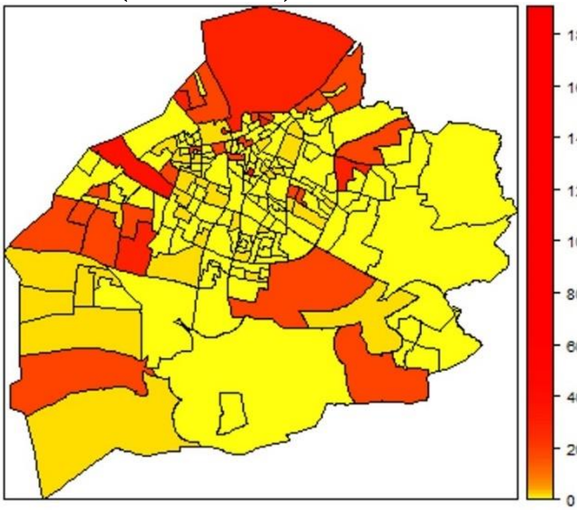
Month 1 (January 2020)



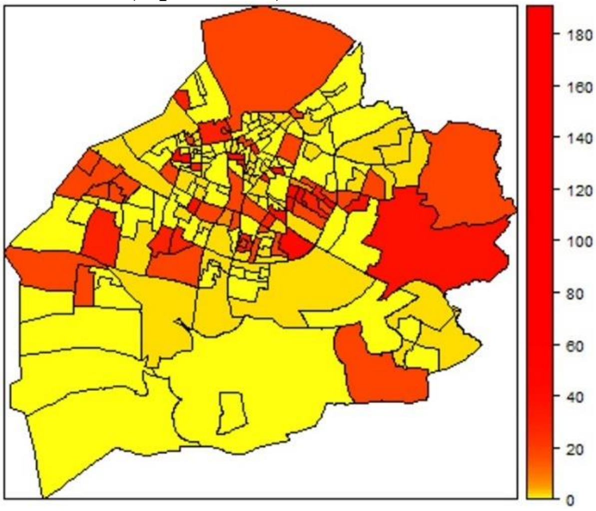
Month 2 (February 2020)



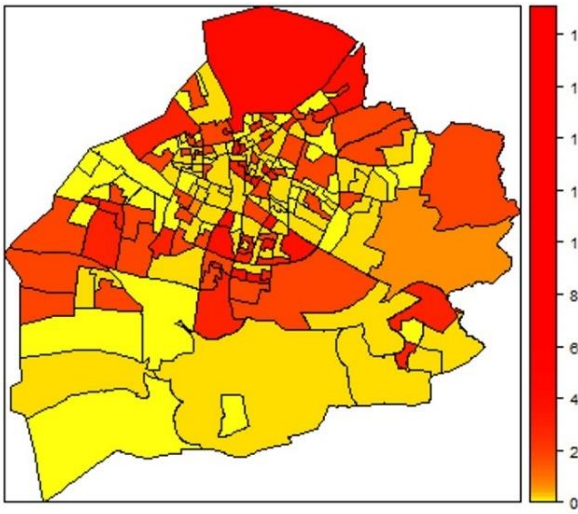
Month 3 (March 2020)



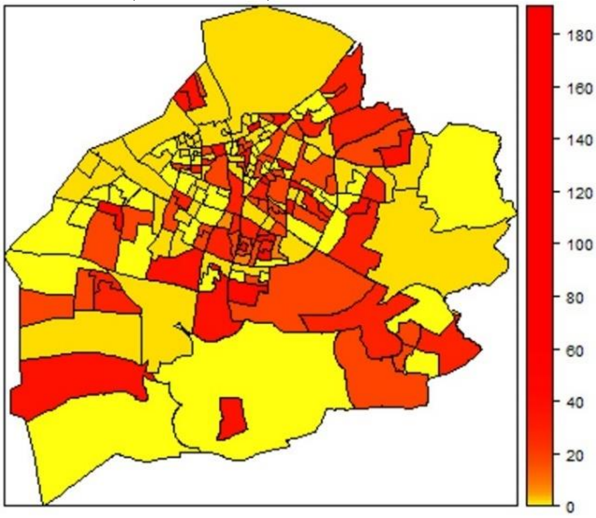
Month 4 (April 2020)



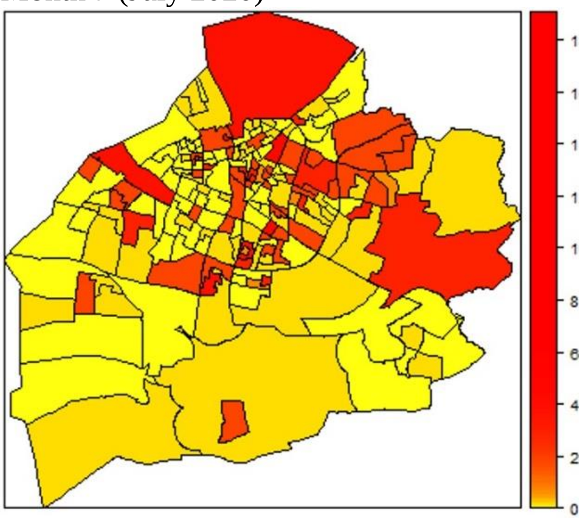
Month 5 (May 2020)



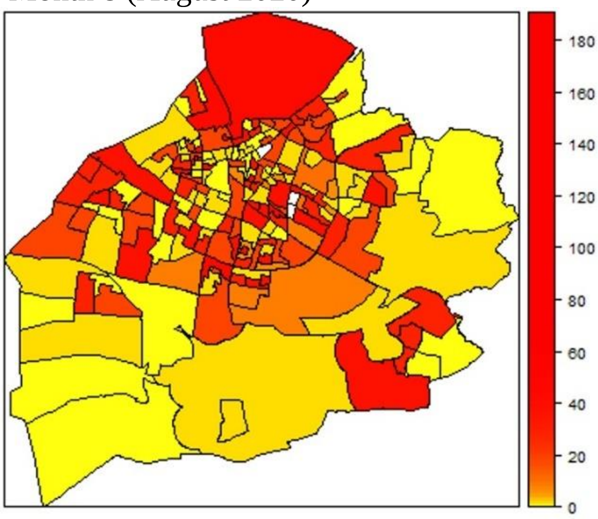
Month 6 (June 2020)



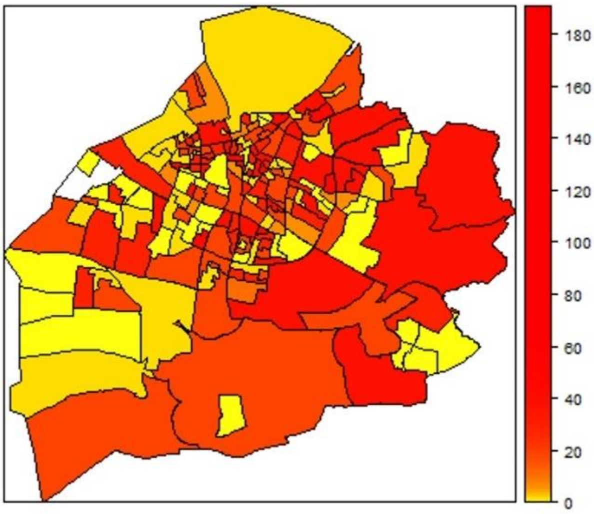
Month 7 (July 2020)



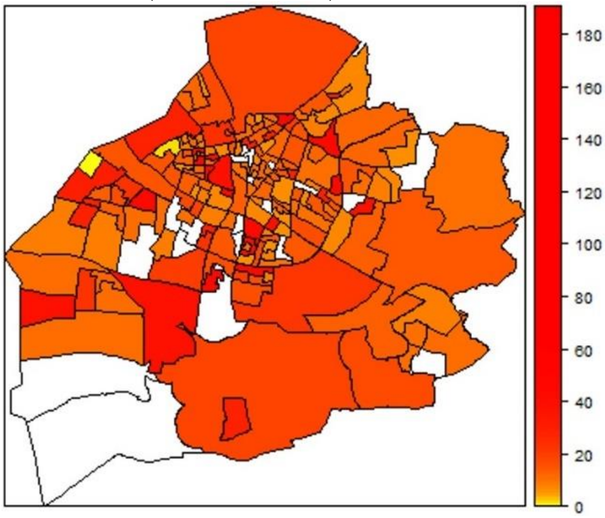
Month 8 (August 2020)



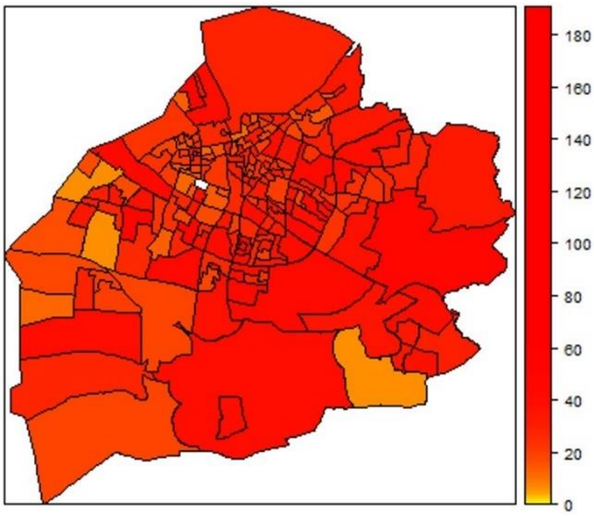
Month 9 (September 2020)



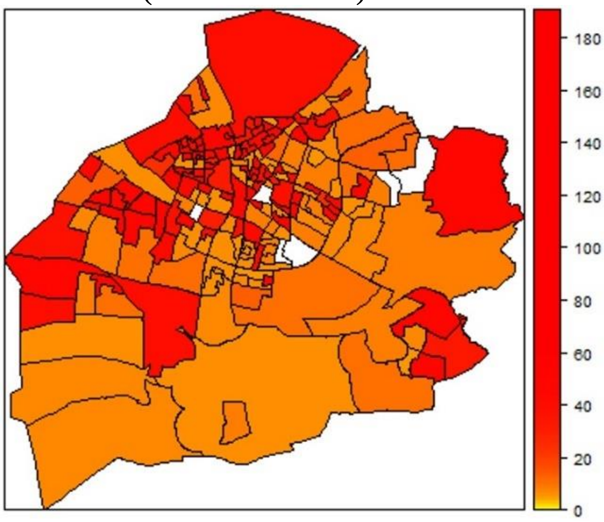
Month 10 (October 2020)



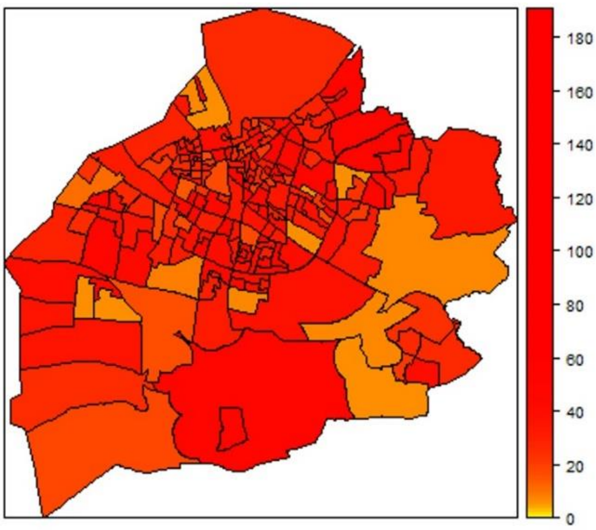
Month 11 (November 2020)



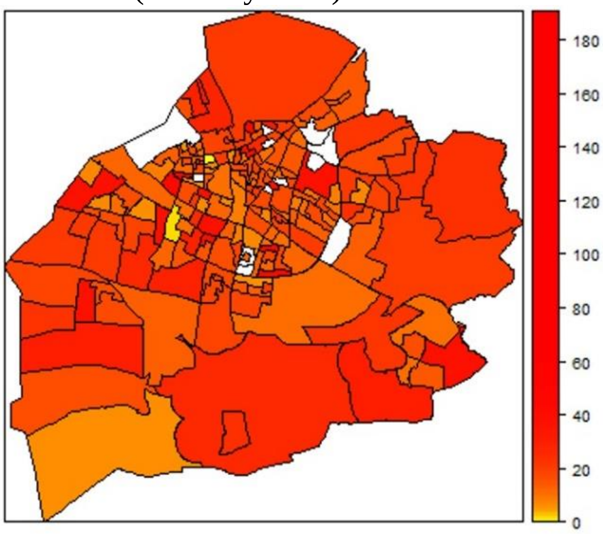
Month 12 (December 2020)



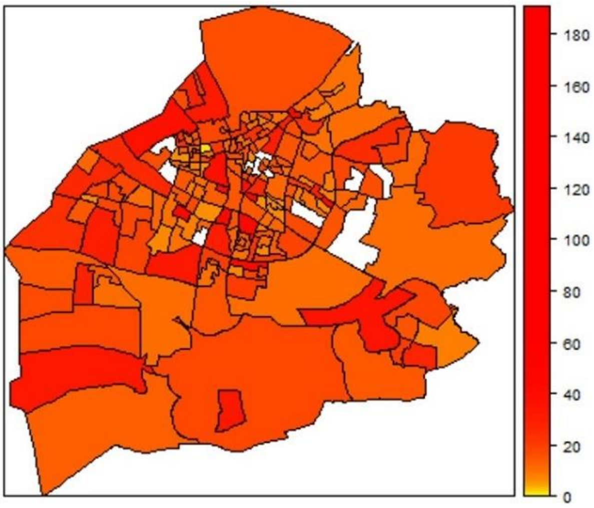
Month 13 (January 2021)



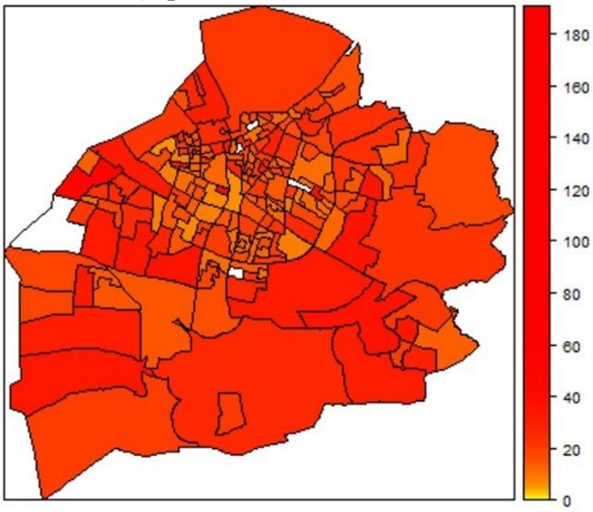
Month 14 (February 2021)



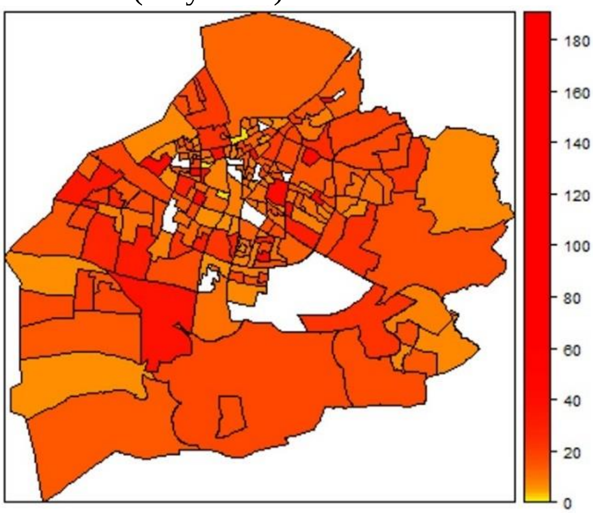
Month 15 (March 2021)



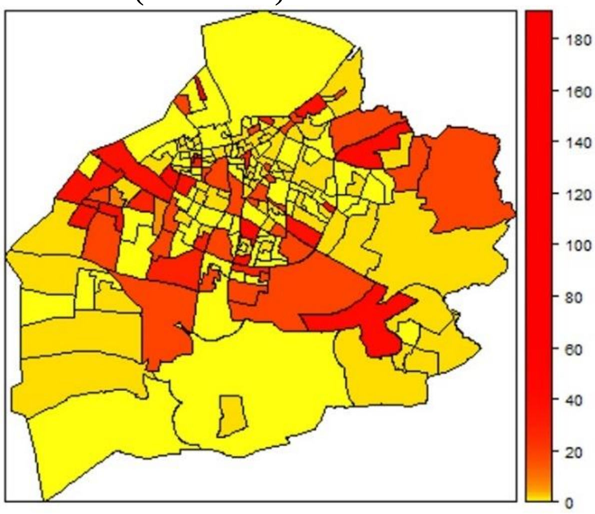
Month 16 (April 2021)



Month 17 (May 2021)



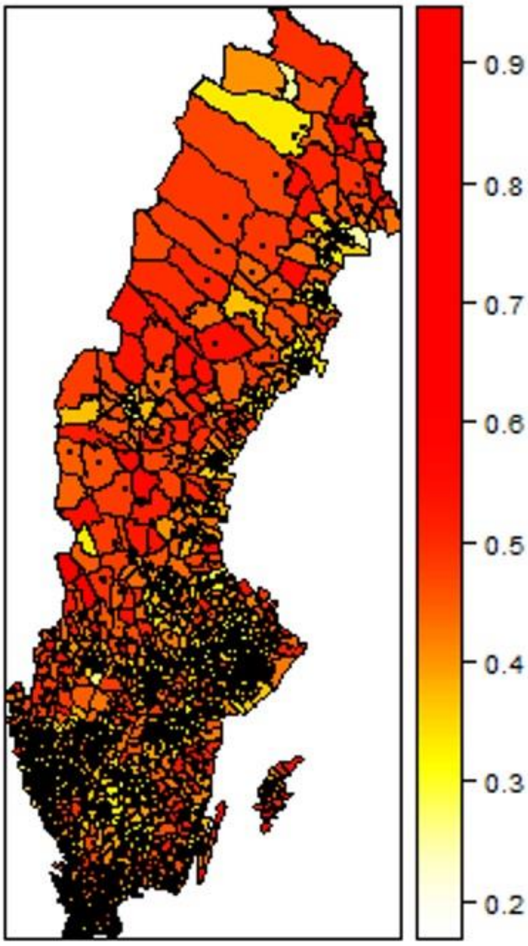
Month 18 (June 2021)



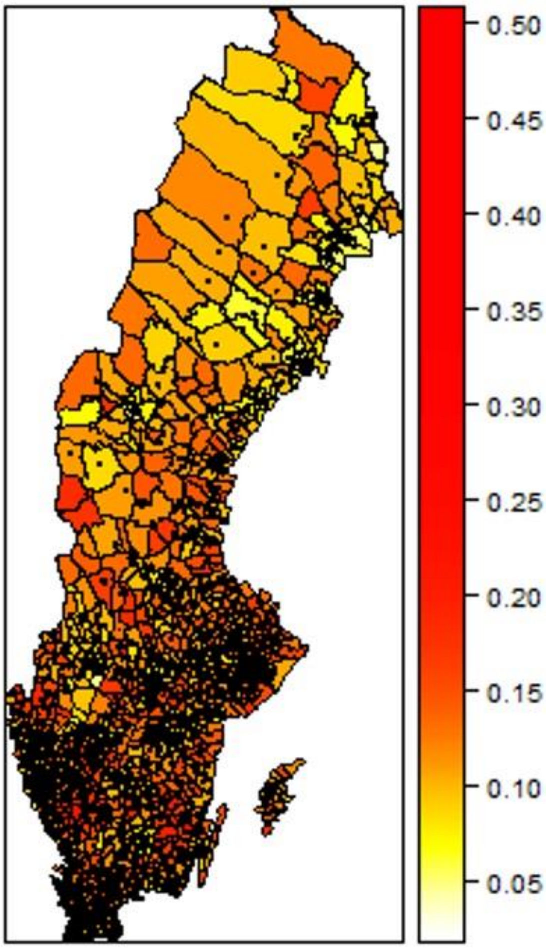
APPENDIX III

Geographical presentation of chosen covariates for Sweden per DeSO-zones

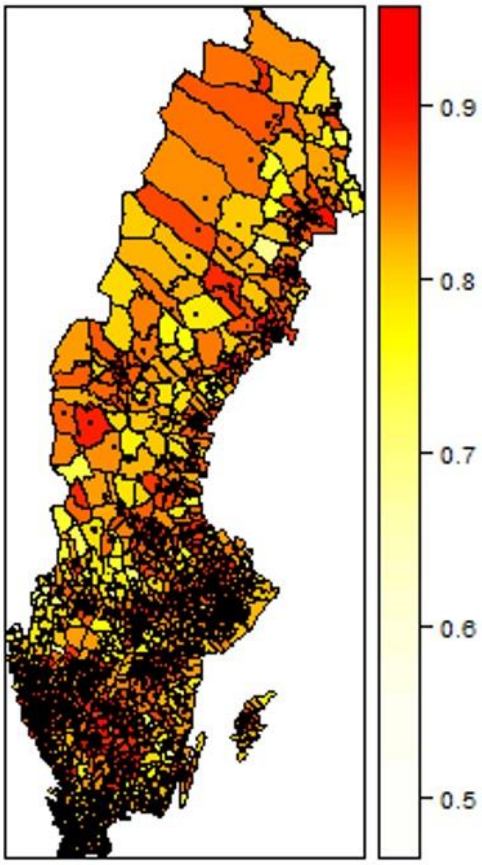
Proportion of people with an economic standard below the national median



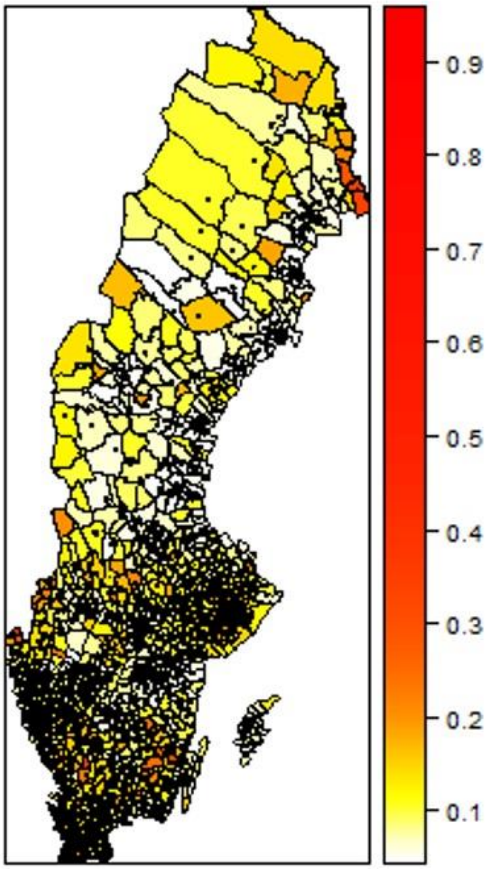
Proportion of people without (post-) secondary education



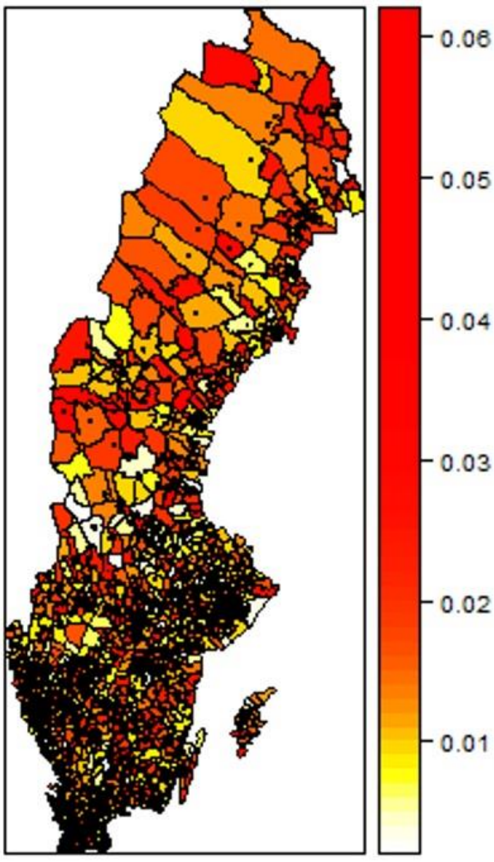
Proportion of gainfully employed people



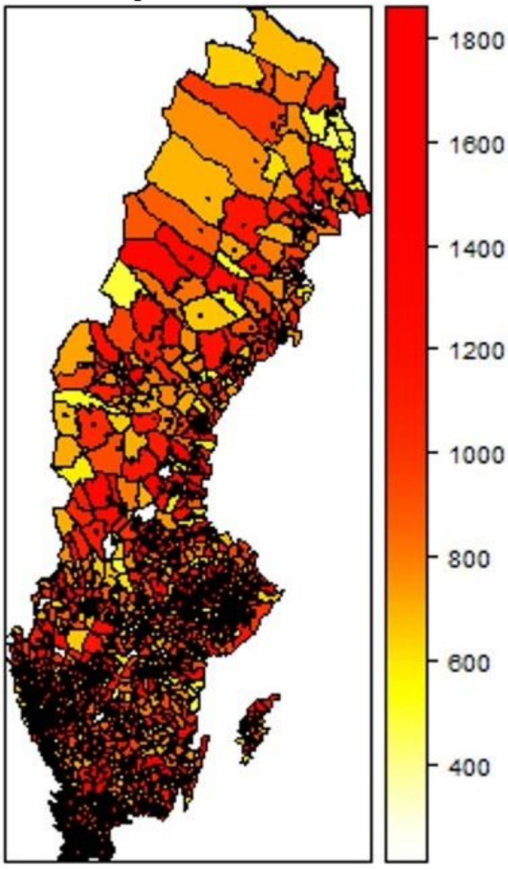
Proportion of people with a foreign background



Proportion of people in health care work

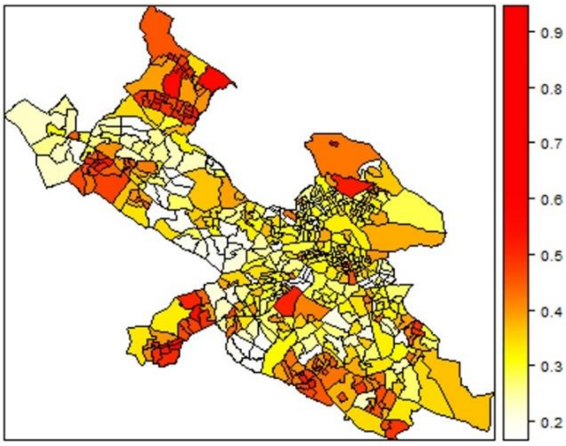


Counts of person cars

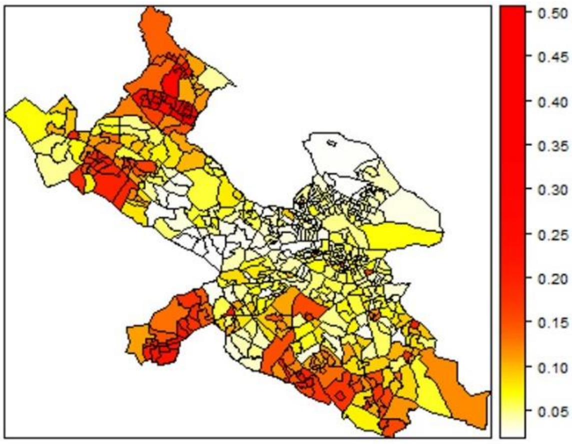


Geographical presentation of chosen covariates for Stockholm per DeSO-zones

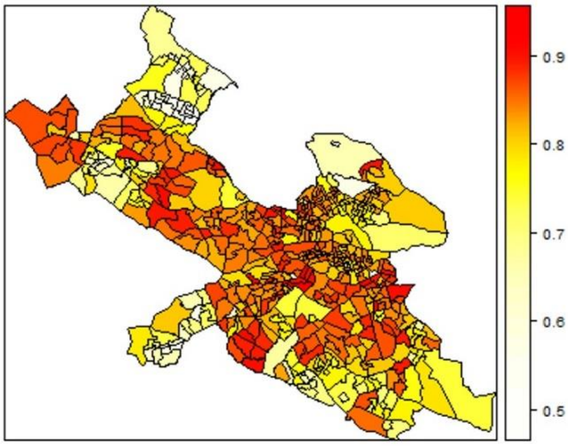
Proportion of people an economic standard below the national median



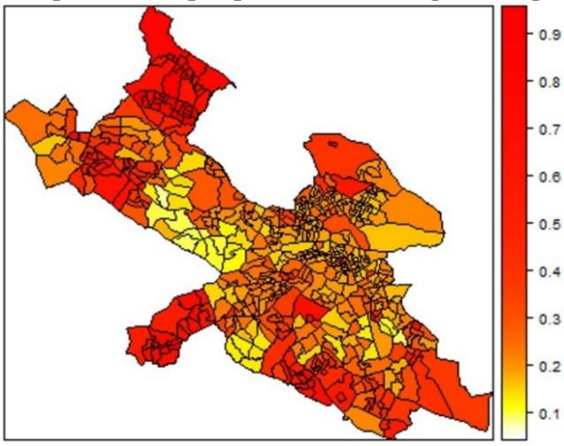
Proportion of people without (post-) secondary education



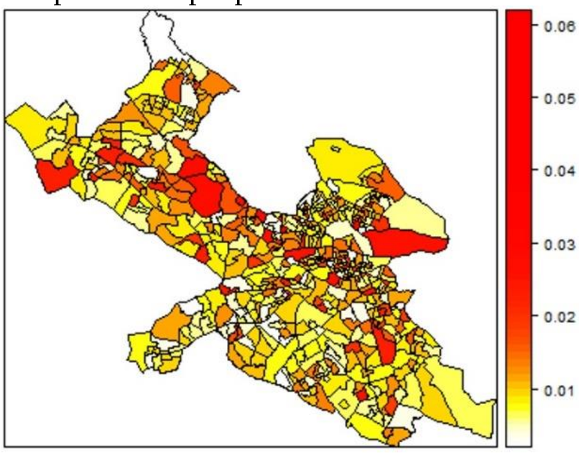
Proportion of gainfully employed people



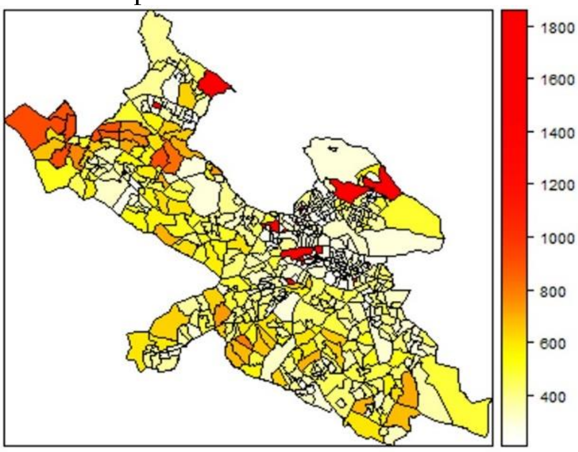
Proportion of people with a foreign background



Proportion of people in health care work

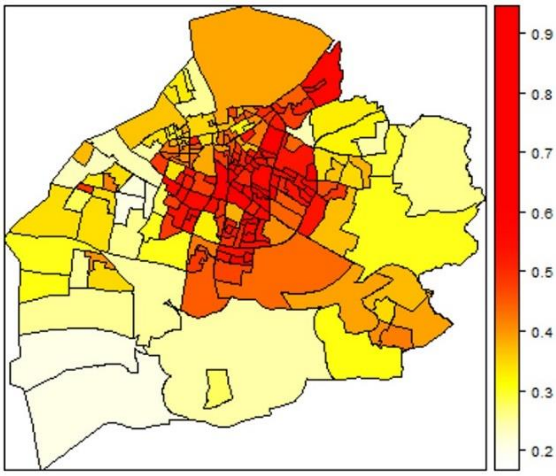


Counts of person cars

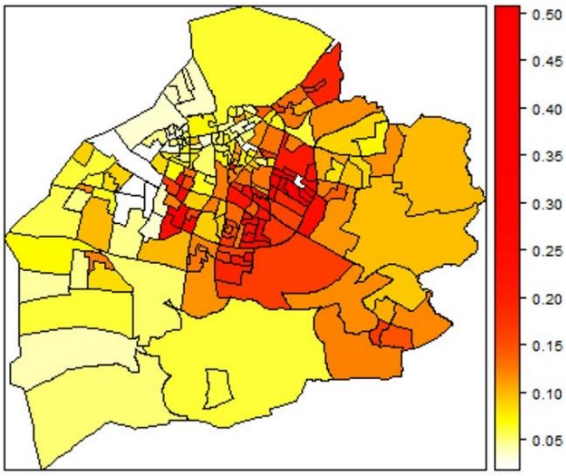


Geographical presentation of chosen covariates for Malmö per DeSO-zones

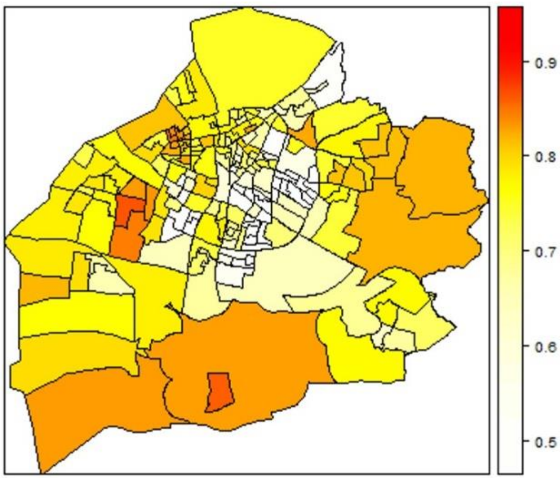
Proportion of people with an economic standard below the national median



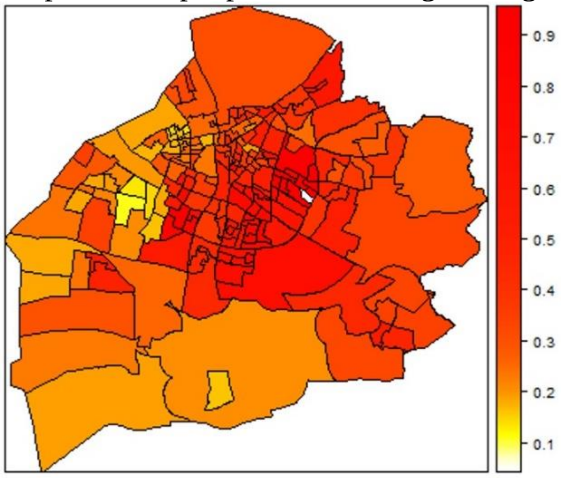
Proportion of people without (post-) secondary education



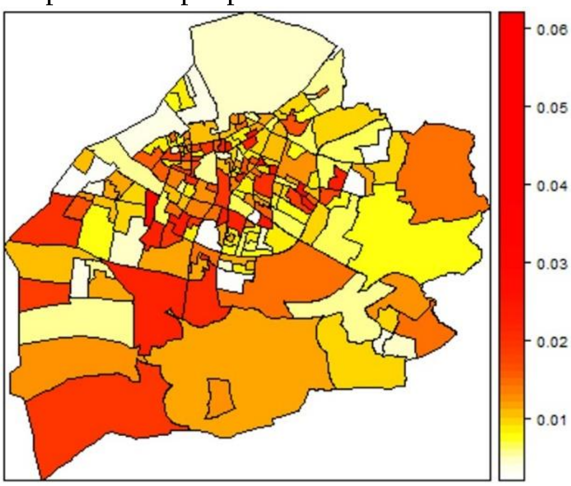
Proportion of gainfully employed people



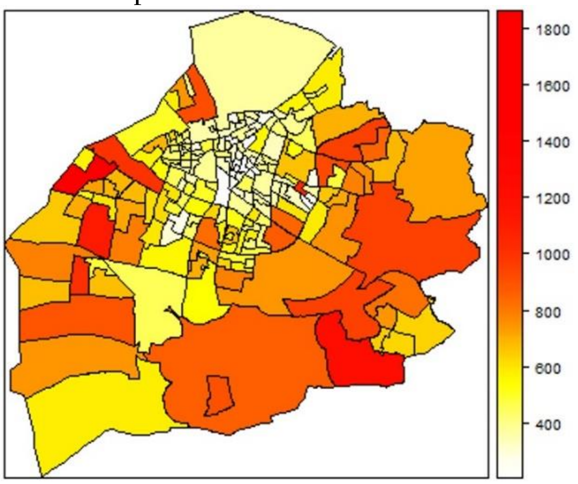
Proportion of people with a foreign background



Proportion of people in health care work

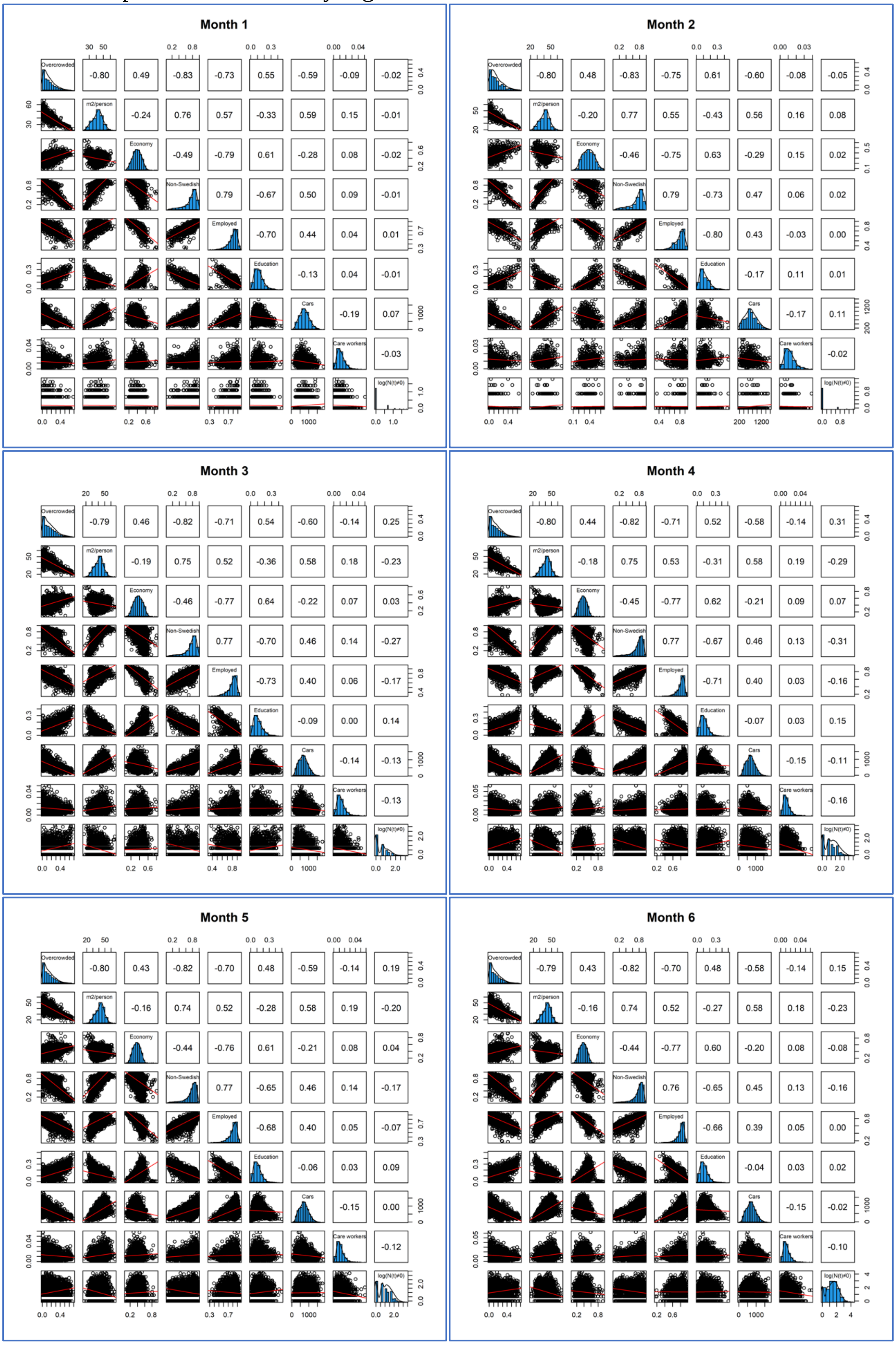


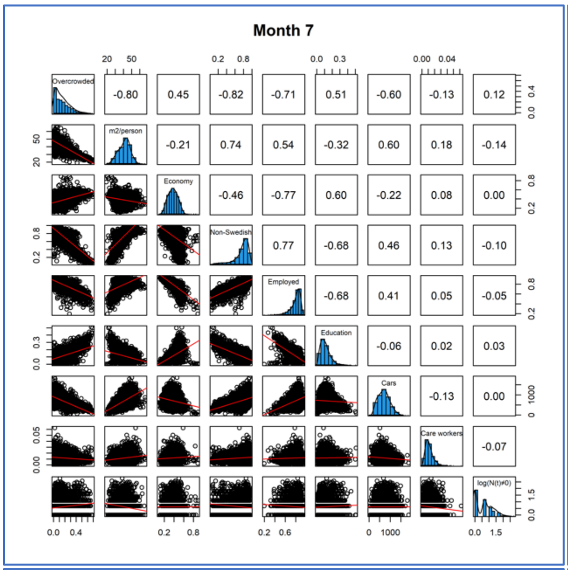
Counts of person cars

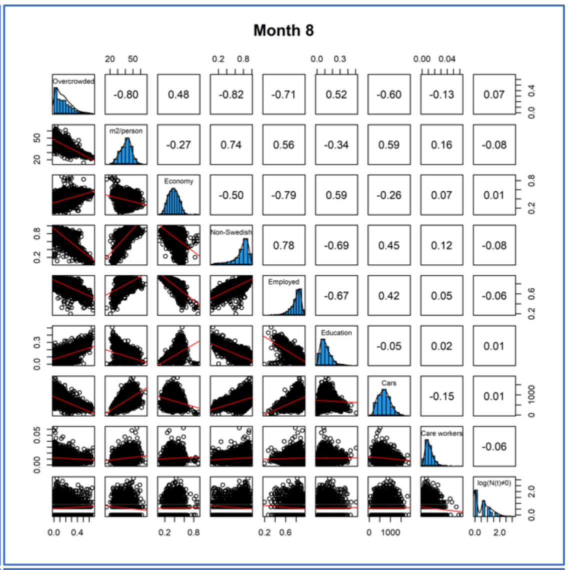


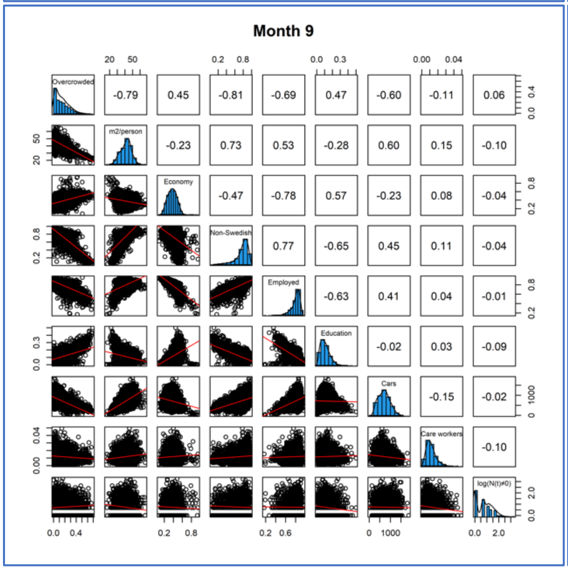
APPENDIX IV

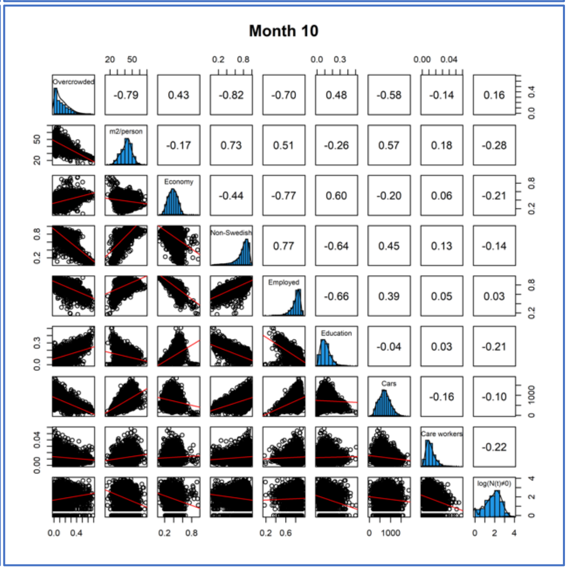
Correlation plots for the monthly log-counts

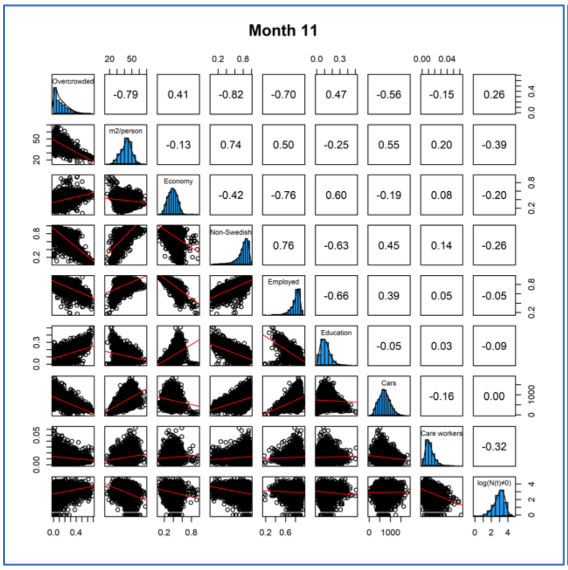


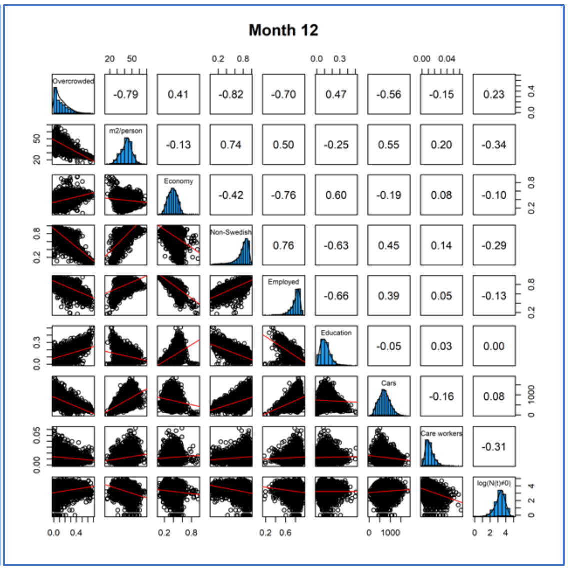

Month 7
Overcrowded
m2/person
Economy
Non-Swedish
Employed
Education
Cars
Care workers
log(N(t)≠0)


Month 8
Overcrowded
m2/person
Economy
Non-Swedish
Employed
Education
Cars
Care workers
log(N(t)≠0)

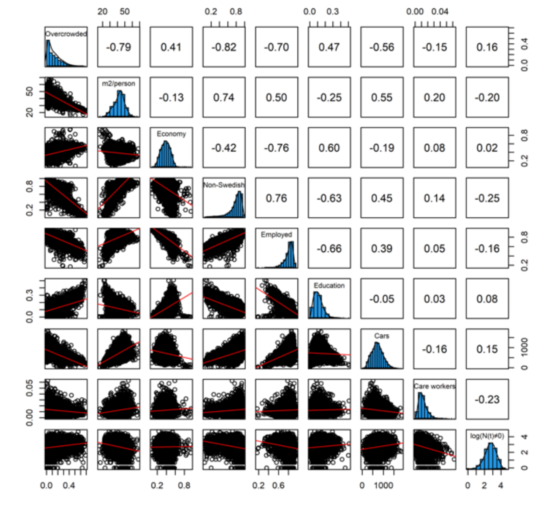

Month 9
Overcrowded
m2/person
Economy
Non-Swedish
Employed
Education
Cars
Care workers
log(N(t)≠0)


Month 10
Overcrowded
m2/person
Economy
Non-Swedish
Employed
Education
Cars
Care workers
log(N(t)≠0)

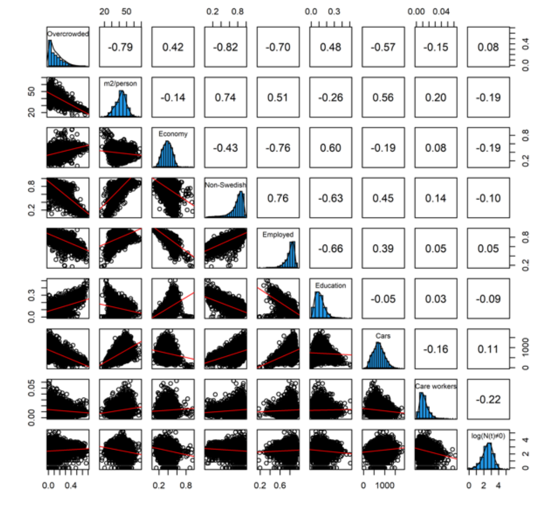

Month 11
Overcrowded
m2/person
Economy
Non-Swedish
Employed
Education
Cars
Care workers
log(N(t)≠0)


Month 12
Overcrowded
m2/person
Economy
Non-Swedish
Employed
Education
Cars
Care workers
log(N(t)≠0)

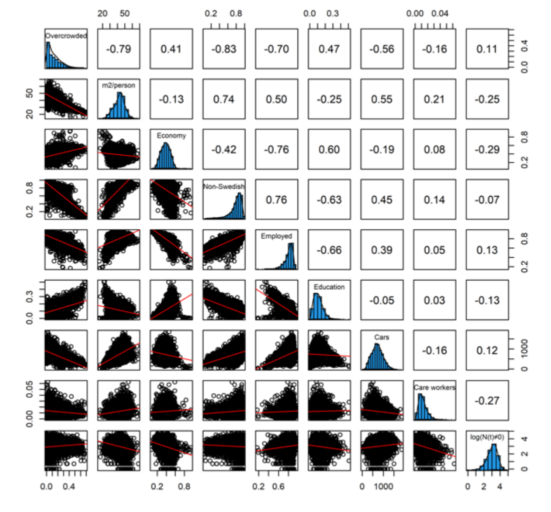
Month 13



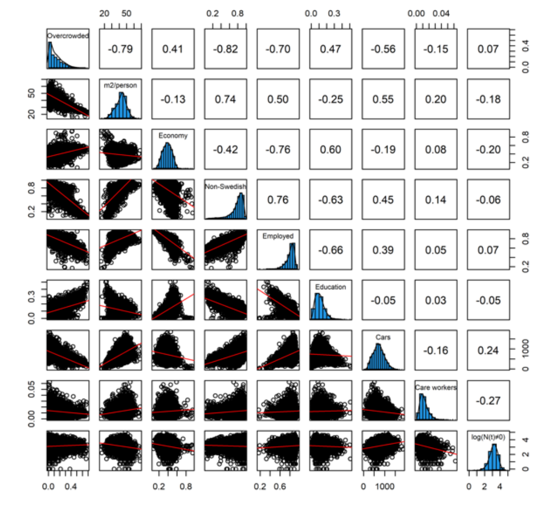
Month 14



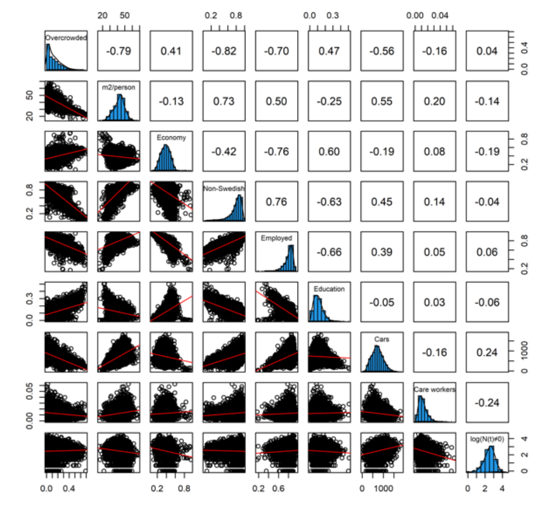
Month 15



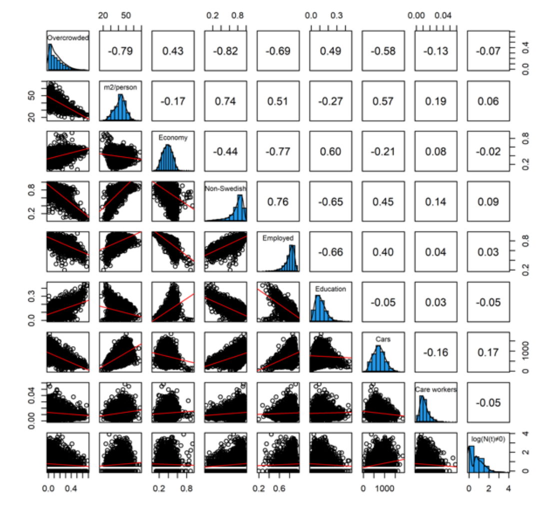
Month 16



Month 17



Month 18



APPENDIX V

Results for the full spatio-temporal models

Variable name	Estimate	Exponentiated estimate
Month13	2.172	8.781
Month12	1.546	4.695
Month14	1.092	2.982
Month11	0.921	2.513
Month15	0.739	2.095
Month17	0.708	2.030
Month18	0.669	1.952
Month16	0.477	1.612
Foreign background*Month17	0.194	1.214
Foreign background*Month18	0.185	1.203
Person cars*Month19	0.171	1.187
Gainfully employed*Month16	0.156	1.168
Foreign background*Month16	0.148	1.160
Foreign background*Month19	0.142	1.152
Lack (post-) secondary education*Month17	0.141	1.151
Low economic standard*Month14	0.139	1.149
Lack (post-) secondary education *Month16	0.126	1.135
Lack (post-) secondary education *Month18	0.115	1.122
Overcrowding*Month5	0.115	1.121
Foreign background*Month15	0.111	1.117
m2/person*Month14	0.110	1.117
Person cars*Month14	0.093	1.097
Low economic standard*Month19	0.089	1.094
m2/person	0.089	1.093
Overcrowding*Month4	0.085	1.089
Gainfully employed*Month17	0.083	1.087
Lack (post-) secondary education*Month5	0.072	1.074
Gainfully employed*Month7	0.063	1.065
Low economic standard*Month13	0.060	1.062
Gainfully employed*Month12	0.060	1.062
Gainfully employed*Month15	0.059	1.061
Person cars*Month18	0.058	1.060
Health care workers*Month7	0.056	1.058
Lack (post-) secondary education*Month4	0.054	1.056
Low economic standard*Month17	0.053	1.055
Lack (post-) secondary education*Month15	0.053	1.055
Person cars*Month13	0.053	1.055
Lack (post-) secondary education*Month7	0.051	1.052
Health care workers*Month19	0.046	1.047
Overcrowding*Foreign background	0.044	1.045
m2/person*Month13	0.043	1.043
Lack (post-) secondary education *Month6	0.039	1.040
Person cars*Month17	0.034	1.035
Low economic standard*Month15	0.032	1.032
Low economic standard*Month6	0.029	1.030
Gainfully employed*Month11	0.027	1.028

Person cars*Month15	0.027	1.028
Overcrowding*m2/person	0.024	1.025
Health care workers*Month14	0.024	1.024
Overcrowding*Low economic standard	0.023	1.024
Overcrowding*Month8	0.022	1.023
m2/person*Month12	0.021	1.021
Foreign background*Month6	0.017	1.017
Overcrowding*Month11	0.017	1.017
Low economic standard*Month18	0.016	1.016
Health care workers*Month18	0.015	1.015
Gainfully employed*Month18	0.014	1.014
Health care worker*Month17	0.013	1.013
Health care worker*Month6	0.013	1.013
Low economic standard*Gainfully employed	0.013	1.013
m2/person*Month18	0.011	1.011
Health care worker*Month15	0.009	1.009
Health care worker	0.008	1.009
Health care worker*Month8	0.008	1.008
m2/person*Month11	0.008	1.008
Car ownership*Month16	0.007	1.007
Overcrowding *Health care workers	0.006	1.007
Person cars*Month6	0.006	1.007
Gainfully employed*Person cars	0.005	1.005
Car ownership*Health care workers	0.004	1.005
Lack (post-) secondary education*Health care workers	0.003	1.003
Health care workers*Month16	0.003	1.003
Lack (post-) secondary education*Person cars	0.002	1.002
m2/person*Month17	0.002	1.002
m2/person*Month15	0.001	1.001
Gainfully employed*Health care workers	0.001	1.001
Lack (post-) secondary education *Month12	0.0005	1.000
Gainfully employed*Month5	0.0002	1.000
Overcrowding *Person cars	0.00002	1.000
Car ownership	0.00001	1.000
Gainfully employed*Month14	-0.001	0.999
Foreign background*Gainfully employed	-0.001	0.999
Person cars*Month12	-0.001	0.999
Low economic standard*Month5	-0.002	0.998
Low economic standard*Car ownership	-0.002	0.998
Car ownership*Month8	-0.002	0.998
Gainfully employed*Month13	-0.002	0.998
Month7	-0.003	0.997
Foreign background*Health care workers	-0.004	0.997
Health care workers*Month5	-0.004	0.996
Overcrowded* Lack (post-) secondary education	-0.004	0.996
Low economic standard*Foreign background	-0.004	0.996
Overcrowded*Month16	-0.005	0.995
m2/person*Month7	-0.005	0.995
Low economic standard*Month8	-0.006	0.994
m2/person*Month16	-0.007	0.993
Low economic standard*Health care workers	-0.008	0.992
Health care workers*Month11	-0.012	0.988

Foreign background	-0.0151	0.984
Health care workers*Month12	-0.0170	0.983
Foreign background*Lack (post-) secondary education	-0.0176	0.982
Overcrowding*Gainfully employed	-0.0179	0.982
m2/person*Foreign background	-0.0181	0.981
Lack (post-) secondary education*Month13	-0.0199	0.980
Health care workers*Month9	-0.0203	0.979
Gainfully employed	-0.0211	0.979
Foreign background*Car ownership	-0.0216	0.978
Person cars*Month5	-0.0218	0.978
Overcrowding*Month3	-0.0236	0.976
m2/person*Low economic standard	-0.0335	0.967
Person cars*Month11	-0.0362	0.964
Lack (post-) secondary education*Month9	-0.0413	0.959
Gainfully employed*Month10	-0.0425	0.958
Overcrowding*Month17	-0.0461	0.954
Gainfully employed*Month9	-0.0529	0.948
Foreign background*Month13	-0.0593	0.942
Foreign background*Month4	-0.0623	0.939
Health care workers*Month4	-0.0660	0.936
Overcrowding*Month15	-0.0727	0.929
Overcrowding*Month13	-0.0787	0.924
Overcrowding*Month18	-0.0803	0.922
Month10	-0.0818	0.921
Foreign background*Month14	-0.0853	0.918
Lack (post-) secondary education*Month19	-0.0876	0.916
Overcrowding*Month14	-0.0933	0.910
Person cars*Month4	-0.1145	0.891
Overcrowding*Month19	-0.1165	0.890
Month5	-0.1167	0.889
Low economic standard*Month4	-0.1182	0.888
Gainfully employed*Month2	-0.1249	0.882
Gainfully employed*Month19	-0.1311	0.877
Low economic standard*Month3	-0.1311	0.877
Overcrowding	-0.1328	0.875
Lack (post-) secondary education *Month10	-0.1359	0.872
Lack (post-) secondary education *Month11	-0.1464	0.863
Gainfully employed*Month3	-0.1525	0.858
Overcrowding*Month2	-0.1792	0.835
Month6	-0.1919	0.825
Month4	-0.2332	0.791
Month9	-0.3959	0.673
Month19	-0.4356	0.646
Month8	-0.7162	0.488
Month2	-1.2805	0.277
Month3	-1.7242	0.178
(Intercept)	-6.0024	0.002

Utgiven av Avdelningen samhällsmedicin och folkhälsa, Göteborgs universitet
2022-10-14
ISBN 978-91-86863-28-9

amm@amm.gu.se
031-342 30 40
GU rapporter, Box 414, 405 30 Göteborg

Hemsidor: www.amm.se och gupea.ub.gu.se/handle/2077/34412