Investigating the efficiency of measures taken by the Cypriot government for the prevention of Covid-19 pandemic using spatio-temporal analysis

Maria Loizou

Master of Science in Data Science and Statistics
The University of Bath
September 2021

Investigating the efficiency of measures taken by the Cypriot government for the prevention of Covid-19 pandemic using spatio-temporal analysis

Submitted by: Maria Loizou

Declaration

This dissertation is submitted to the University of Bath in accordance with the requirements of the degree of Bachelor of Science in the Department of Computer Science. No portion of the work in this dissertation has been submitted in support of an application for any other degree or qualification of this or any other university or institution of learning. Except where specifically acknowledged, it is the work of the author.

Abstract

Investigating the efficiency of measures taken by the Cypriot government for the prevention of Covid-19 pandemic using spatio-temporal analysis

Author: Maria Loizou

The aim of the project is to investigate the performance of Covid-19 response measures imposed by the Cypriot government during one year period since the start of the pandemic in the island on 9th of March 2020. Data are retrieved by the Ministry of Health after approval. The investigation is based on spatial analysis methods and techniques from which results inferred explaining the propagation of the pandemic in accordance with response measures and their duration in force. Exploratory and modelling spatial methods are applied to the main dataset containing information about Covid-19 in Cyprus. At last, project findings will be presented and any topics of future work based on spatial analysis of Covid-19 in Cyprus will be recommended.

Supervisor: Dr. Evangelos Evangelou

Acknowledgements

”The data used in this study was collected by the Epidemiological Surveillance Unit of the Medical and Public Health Services of the Cyprus Ministry of Health. The ideas and opinions expressed herein are those of the author. Endorsement of these ideas and opinions by the Ministry of Health of Cyprus is not intended nor should it be inferred.”

I would like to extend my gratitude to my supervisor Dr. Evangelos Evangelou for his in-valuable guidance and help throughout every step of the project. I would also like to thank my family for their support and help throughout this tough journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

On this chapter the background and motivation of investigating the Covid-19 pandemic using spatial analysis are discussed. The theory behind spatial statistics is introduced briefly and more precisely areal/lattice component which is the one to be used. At last, the aims and objectives of the project are presented.

1.1 Introduction

On 31 December 2019, the coronavirus disease (SARS-CoV-2) first makes its appearance in the Chinese city, Wuhan. The Chinese government first reports the outbreak of the disease, while the coronavirus spreads rapidly in Chinese cities. After the excessive worldwide spread of the disease, the World Health Organisation (WHO) declared it as a pandemic on 11 March of 2020. From this day until now, several sectors show their interest on investigating this novelous disease and thus numerous research papers have been addressed so far. Therefore, after one year and a half, the rise of awareness for Covid-19 leads different nations and sectors from all over the world to corporate such that the end of this pandemic will start to emerge.

Undoubtedly, Cyprus country could not be an exception. The disease is still within the borders of the country and many people get infected every day. As surveillance and early notice are important for the prevention of infectious disease outbreaks, the need for developing epidemiological models and making forecasts is a vital for the prevention and management of Covid-19 worldwide and particularly in the country. Since Covid-19 first appearance the government has introduced various measures in order to prevent the excessive propagation of the pandemic. However, to what extend were these measures efficient? Are there any possible limitations in them? Or the lack of controlling the pandemic in the island depends on other concerns in the health sector? The current project will give answers to these questions.

Governments worldwide, including Cyprus, are still using the common measures to prevent the propagation of the pandemic such as facial masks, social distancing, massive testing and tracing, or more severe restrictions like quarantine, curfew, or lockdowns. Therefore, a deeper understanding of the Covid-19 dynamics would be a great step in the improvement of the measures taken, giving some reliable conclusions to help governments handle the pandemic. This process will also lead us to some relevant factors that affect the spread of the pandemic like the efficiency of the health system, individual responsibility, or pressure from the economic sector on the incidence distribution and spatiotemporal pattern.

1.2 Motivation

Data for infectious diseases, such as Covid-19, are crucial as their investigation can respond appropriately when and why cases occur. Such data can be analysed using a variety of methods. However, nowadays the huge variety of georeferenced data about public health, demographic variables or either satellite capabilities that affect disease activity have facilitated a new generation of the investigations based on spatial and spatio-temporal variations of a disease. There is a wide range of spatial and spatio-temporal methods for disease surveillance including methods for disease mapping, clustering, and geographic correlation studies. Spatial mapping of disease risk is also beneficial as it provides an immediate visual summary of the data and identifies any patterns within data. Such maps are crucial for describing the spatial and temporal variation of the disease, identifying areas of unusually high risk, formulating etiological hypotheses, measuring inequalities, and allowing better resource allocation.

Spatial data is extensively important when dealing with diseases and spatial epidemiology. Firstly, they contribute to present the theory and methods concerned with the study of spatial patterns of disease incidence and mortality. Maps are a very useful and interactive tool in spatial epidemiology to gain insights and illuminate potential causes of disease. A coordinated approach for developing an early detection and response system for disease prevention are vital. Mapping the infections offers an immediate visualization of the extent and magnitude of the public health problem. Secondly, sophisticated algorithms used, and modelling are powerful tools for the prediction of disease patterns. Spatial statistics when investigating infectious diseases overlay the rest of the statistical methods regarding their tools defined above.

The distribution of the diseases is discussed with respect to demographic risk factors such as population, population per age group, population per gender, employment status, educational attainment etc. Spatial autocorrelation, disease mapping, clustering is some of the concepts to be investigated. An infectious disease such as Covid-19 is propagating rapidly in the community and its investigation can reduce the number of cases by imposing the right measures. Therefore, spatial analysis, maps and spatial models are valuable in strengthening the whole process of incidence surveillance information management and analyses.

For the current dissertation project I have chosen to investigate Covid-19 pandemic in Cyprus which is my country in order to contribute with my own way to fight against the virus. The whole project is based on spatial data analysis and more specifically in spatiotemporal analysis. For the project R software is used as it provides excellent tools which can accompany the whole process of the study. Aims and objectives about investigating this topic are followed.

1.3 Aims and Objectives

Several approaches have been used so far for analysing and modelling Covid-19 cases. From classic statistics, epidemiological models to spatial analysis. The main aim of this project is to examine and explore the pandemic in the country of Cyprus using spatial or either spatio-temporal analysis and supplementary to evaluate in depth to what extend measures taken by the government were efficient enough. Spatio-temporal dissemination of the pandemic in Cyprus will be investigated, according to some variables like time/date, districts, municipalities, age groups, gender etc. Models and methods introduced in spatial statistics will be applied in a dataset referring to Covid-19 in Cyprus in order to allow inferences to be drawn. Prior to the development of the modelling process, a review of relevant literature surrounding these methods in some other real-world datasets, and through simulation studies is essential.

The project will examine the strengths and more importantly, limitations associated with the different measures imposed, as improved understanding into the current measures, for facing the pandemic has vast benefits on knowledge of the causes and progression of the disease, and therefore society as a whole.

1.4 Background

Spatial statistics is an area of statistics which analyses data that has a spatial (location) characteristic to it. This type of data analysis tries to identify any patterns or correlation among observations of some process that occurs across a space. Various sectors like epidemiology, biology, economics, geography, and many more have shown a great interest on tools and methods introduced in spatial analysis and statistics. This type of data can be divided into three main components, the so-called geostatistical data, areal/lattice, and point-pattern data. Hence, when considering a study an essential initial step is to recognize the type of data under study. It is important to distinguish between the discreteness or continuity of the spatial indices on which variables are measured and, on the discreteness, and continuity on variable values themselves. The focus of this project will be on areal/lattice data analysis of Covid-19 cases at municipality/community level in Cyprus accompanied with different risk factors such as age, gender etc.

The fundamental rule used in spatial analysis is the spatial dependence among observations. As expressed in Tobler's First Law of Geography, ”everything is related to everything else, but near things are more related than distant things.” Nearby observations tend to be more similar than those far apart. This patterning is most known as spatial autocorrelation, and it can be found useful and challenging simultaneously as some data assume mutual independence of observations. ”In order to interpret what ”near” and ”distant” mean in a particular context, observations on the phenomenon of interest need to be referenced in space, e.g., in terms of points, lines or areal units” (Anselin, 1989). Therefore, a brief definition of spatial analysis is that techniques and methods are used to draw inferences about the data using the spatial referencing associated with each data value. Conclusions reached describe any spatial relationships within the data, why they exist and their interpretation.

Observations in spatial problems come from a spatial process defined as:

$$X = \{X_s, s \in S\}$$

where S is called the spatial set and provides a set of indices which represents the locations/sites where values are measured. The random variable X_s displays the value of X at the s location. The variables X_s are taking values in a state space E. Spatial set, S, can include fixed or random observation sites depending on the type of spatial data examined as it is discussed next. In practice, S is often a two-dimensional subset, $S \subseteq R^2$. However, it can be one-dimensional for instance measuring variables with time indices.

1.4.1 Areal/lattice data

During the current study the second element of spatial data is considered, which is the areal/lattice data. In areal data the observations are associated with a fixed number of areal units that represent a regular lattice or it can be a set of irregular areas like counties, districts or even countries. Therefore, S is a fixed discrete non-random set, usually $S \subseteq R^d$ and X is observed at points in S. Random variables in each s, can be either discrete or continuous and the areal objects to which the variable is attached are fixed.

Within a lattice it is assumed that there are n lattice measurements each one denoted by $X_i, X_j, X_k, \ldots$ for the n lattice points $i, j, k, \ldots$. Except from the certain measurement at a specific point, a lattice point i is associated with its neighbouring points which are denoted by N_i. Figure 1.1 is an example of a lattice area.

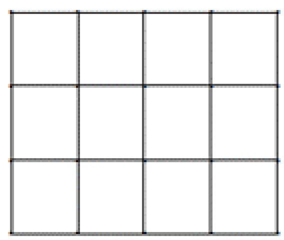


Figure 1.1: Lattice area (Ott, 1994).

Symmetric property for neighbourhood relationship:

$$j \in N_i \text{ if and only if } i \in N_j$$

Within this spatial data type a model used considers all possible joint probabilities $p(x_1, x_2 \ldots, x_n)$ which define a probability mass or density function depending on the context of data. Alternatively, using just one conditional probability the same result is being derived, $p(x_i | x_j, j \neq i)$.

The most important class of models are the Markov random fields models, and the Hammersley-Clifford theorem can be used to provide consistent systems of probability distributions. The main property of a Markov random field is that all conditional probabilities take the form,

$$p(x_i | x_j, j \neq i) = p(x_i | x_j, j \in N_i),$$

where N_i is the set of neighbours for the i lattice point in the lattice structure considered. Some reasonable models used to fit lattice data are the so-called by Besag auto-models, are a particularly simple subclass of Markov random fields that are useful in spatial statistics.and includes a function q of the form,

$$q(x) = \sum_{i=1} x_i g_i(x) + \sum_{i<j} \beta_{ij} x_i x_j,$$

where $\beta_{ij} = 0$ unless i and j are neighbours. These models are suitable for situations where measurements are clustered over some spatial areas for example in epidemiology the number of cases of a disease per district. A real example of lattice data is shown in Figure 1.2. In this study the fixed region of interest, is the country of Cyprus which has been partitioned into a finite fixed number of sub-regions (municipalities/communities) at which outcomes have

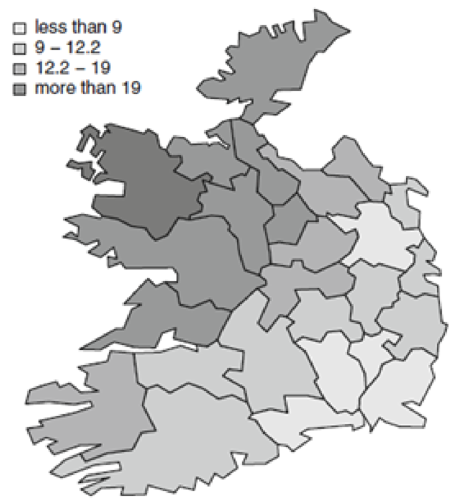


Figure 1.2: Real example Areal data.Percentage of people with blood group A in the 26 counties of Ireland (eire dataset from the spdep package) (Gaetan and Guyon, 2010a).

been aggregated. At each region we observe the number of Covid-19 cases in a daily basis accompanied with the age and gender of the cases. Cyprus and its sub-regions are displayed in the below Figure 1.3.

Figure 1.3: Polygons structure map of Cyprus, divided in municipalities/communities (link here).

Chapter 2

Literature Search

2.1 Literature Search Process

Prior to conducting my literature search for investigating covid-19 using mainly spatiotemporal analysis it was vital to gain appropriate knowledge of the whole context, statistics and methods used in spatial data analysis in general. Thus, by reading some recommended books on spatial data I initiated myself and became familiar with spatial data analysis and statistics and its three main components. More specifically, 'Spatial Statistics and Modelling' (Gaetan and Guyon, 2010b) book first introduced me to the concept of spatial data with some good preface describing briefly and comprehensively the three main elements of spatial statistics. In the following chapters of the book the definitions of random field, variogram, stationarity, intrinsic and anisotropy are discussed in depth. However, as the context in this book was excessively mathematically a look at the 'Environmental Statistics' (Ott, 1994) book made these definitions clear and explicit. A reading combination using the certain books offer me the basic insights, methods and techniques used in each of the three components (geostatistical, areal/lattice, point-pattern) of spatial data. Consequently, I was able to understand the theory and methods of spatiotemporal analysis. At the same time, a brief reading of the book 'Applied Spatial Data Analysis with R' (Bivand et al., 2013) gave me insights on the practical part of spatial analysis using R programming environment.

After gaining the appropriate knowledge on spatial analysis, I started searching and reading articles and papers from Google Scholar. I first searched for papers investigating the spatiotemporal propagation of covid-19 in other countries, in relation to actions taken from governments to prevent it, in order to familiarize myself with the concept of my chosen dissertation topic and gain some insights on models and methods used in this situation to investigate the evolution of covid-19.

In particular, (Gross et al., 2020) refers to the first wave of the pandemic in China and how the country attained to stabilize in just 20 days the infections using strict measures, such as quarantines. Data compare results between cities and spatial scaling is used. Paper (Kergassner et al., 2020) deals with covid-19 in Germany modelling the spatiotemporal outbreak dynamics of it with an SIQRD model, coupled to a network model that allows for spatially distributed cross-county infections. Another model used in simulating diseases is the SAIR compartmental model which presented in the (Aràndiga et al., 2020) paper for investigating the spatiotemporal evolution of covid-19 in various areas of Spain in order to analyse the influence of mobility on the evolution of a disease after a lockdown period.

Another approach presented on (Spassiani, Sebastiani and Palù, 2021) uses a multivariate time-series mixed-effects generalized linear model in sub-groups of the general area of interest, for describing the spatiotemporal evolution the disease in Veneto, Italy. The paper (Kim and Castro, 2020) is the one of great interest as it presents how South Korea government respond to the pandemic. South Korea instead of imposing strict measures, like quarantines, introduced the '3T-Test, Trace, and Treat' method for aggressive and accurate tracking of infected individuals and their potential contacts. Methods used are the global Moran's I statistic, a retrospective space-time scan statistic method (Kulldorff, 1998), a discrete Poisson model, to identify whether the number of infections in each district was spatially correlated, to include time dimension into the analysis and to analyse spatiotemporal clusters of covid-19 cases, respectively. Global Moran's I statistic is also used in (Arauzo-Carod, 2021) (Xiong et al., 2020) papers along with local spatial autocorrelation (LISA) measures to explore the disease in Catalonia (Spain) and Hubei Province, China. For spatial analysis of covid-19 in Iran (Ramírez-Aldana, Gomez-Verjan and Bello-Chavolla, 2020) again the Moran's I statistic and LISA measures are being using again with some spatial multivariate linear models. Two kinds of spatial models were compared the spatial-lag model considers the spatially lagged response as an additional explanatory variable, whereas the spatial-error model considers that the error is a linear function of a spatially lagged error plus another error term. Another paper (Cordes and Castro, 2020) that I have read about spatial analysis of covid-19 in New York City is also using the Moran's I test for spatial autocorrelation and to assess the occurrence of local clusters in each outcome is used a discrete Poisson spatial scan statistic. The certain study aimed to the spatial distribution of covid-19 testing rates, positivity rates, and proportion of positive tests.

As a second step of the literature search process, I decided to find and read some papers referring to spatial modelling and techniques that have been applied in other diseases to gain some insights in the general methods used in spatial analysis for diseases. Firstly, I read about Ebola, an epidemic outbreak in West Africa in the paper (Santermans et al., 2016). Two models, Integrated Nested Laplace Approximation and Markov chain Monte Carlo (MCMC) have been used to compare growth patterns over time among districts. In addition to these models SEIR compartmental models used to fit daily cases and deaths in order to investigate the expansion of the disease over time in a district level. Another paper for the cholera epidemic in Haiti (Piarroux et al., 2011) provides some idea on how to model spatiotemporal data. In this paper a software for temporal and space-time statistics, Sat Scan™, appeared useful tin spatiotemporal modelling process to identify clusters of cholera cases. As data on cholera cases were non-normally distributed and thus violated basic assumptions for linear regression, the model used to predict the number of cases in each commune is the generalized additive model (GAM). A spatiotemporal approach has been applied in malaria epidemic as well as shown in the paper (Taddese, Baraki and Gelaye, 2019). The purpose of this study was to examine the spatial distribution of malaria in northwest Ethiopia. Again, the software Sat Scan™ along with ArcGIS one used to detect clusters and spatial pattern within the data using the Kulldorf method. For clustering and spatial pattern was detected by using ArcGIS and Sat Scan™ software, using the Kulldorf method. Spatial autocorrelation is tested using again global and local Moran's I test statistic.

Last step of the literature search process was to gain some deeper knowledge on the methods used in the papers above, for spatiotemporal analysis. A first look at the paper (Auchincloss et al., 2012) familiarized me with the most used spatial methods in Epidemiology from 2000-2010. These spatial tools from the papers reviewed in the decade were calculations of distances (proximity calculation), estimation of summary measures across prespecified geographic areas (aggregation methods), assessment of various forms of clustering, spatial smoothing and interpolation methods, and spatial regression. From the previous literature search it is obvious that some methods are used very often in such disease data. Specifically, global, and local Moran's I statistic in combination with some epidemiological models are used.

Consequently, a further investigation in the mathematical background of these techniques is essential. I found book 'Applied Spatial Statistics for Public Health' (Waller and Gotway, 2004b) very useful, as it describes all the mathematical and statistical background of spatial statistics, for example Moran's I statistic. It provides a deeper understanding on how spatial statistics can be used in public health. Also, the book 'Compartmental Models in Epidemiology' (Brauer, 2008) initiated me with some basic models used in epidemiology to model disease data. It presents a clear explanation of the simplest epidemiological compartmental model SIR, in order to understand more complex compartmental models referring in some papers. Paper (Cheng et al., 2016) refer to a time series modelling using a negative binomial distribution and some logarithmic scale linear predictors. On Table 2.1 and Table 2.2 the conclusions and finding results of Literature Search process are shown respectively.

2.2 Conclusions from the papers

Conclusions
China managed to handle the pandemic in 20 days. Strict quarantine was efficient and led to an exponential decay of the infection rate over a time period. Temporal infection, recovery and death rate are associated to the health system efficiency as well. There is a relation between population migration and the disease spreading. (Gross et al., 2020)
The probability of future covid-19 case and high propagation of it is larger in populated urban environments. Traveling and large events are two main reasons of the widespread of covid-19 in Germany. (Kergassner et al., 2020)
It presents the SAIR compartmental model is fitted for a region, which is divided into small areas called mobility areas as their mobility variability is known. Numerical results show that the model can qualitatively represent the tendencies of small outbreaks.(Aràndiga et al., 2020)
Infection number in each municipality is associated with the corresponding population size, thus higher intensity for higher incidence. The mathematical analysis in combination to some mobility, hospital admissions, availability in hospital data is useful to address the spread. It suggests that such analysis will influence measures taken suited to local conditions. (Spassiani, Sebastiani and Palù, 2021)
Spatial autocorrelation analysis identified some spatial clusters of covid-19 across the country of South Korea. Spreading events before measures had imposed contributed to this spreading. Social distancing and 3T measures prevent covid-19 dissemination without the need of a lockdown. Healthcare of the country was the backbone to handle covid-19. However, 3T method needs strong trust in the government as access on several personal information is needed and thus cannot be adopted from most of the countries. (Kim and Castro, 2020)

At early staged of covid-19 it was difficult to predict its evolution and the efficient of measures taken as it was something new and governments did not know what the most efficient way was to deal with it. Maps are useful to display cases in order to understand relevant instruments to understand the spatial diffusion process of the pandemic. Public data resources are the most appropriate. (Arauzo-Carod, 2021)
There is evidence that covid-19 pandemic in Hubei province, China had a significant global spatial autocorrelation and had an extremely significant association with social and economic factors. (Xiong et al., 2020)
In Iran provinces covid-19 cases are spatially correlated. Urbanization, population age, age group, education, average temperatures, number of physicians, and mobility in each province are factors strongly related with the case numbers amongst Iranian provinces. (Ramírez-Aldana, Gomez-Verjan and Bello-Chavolla, 2020)
In New York City spatial clusters of high testing rates, high positivity rates, and high proportion of positive tests, as well as low test rate values have been identified. Social and health inequalities are significant factors. Poor social and health status is associated with an increased number in covid-19 cases. Public transportation also increases the number of infections. (Cordes and Castro, 2020)
There is a strong temporal and spatial variability in a subnational level of the ebola disease in West-Africa. Transmission within cities plays an important role in measures taken to prevent the disease. Map visualizations combined with compartmental models to predict the reproduction number of ebola. Asymptomatic cases within the population partially explain why the epidemic did not reach the expected incidence as predicted by models ignoring them. (Santermans et al., 2016)
The spread of cholera in Haiti was associated to the Artibonite River. This result, as well as the simultaneity of the outbreak onset in 7 communes of Lower Artibonite on October 19, is in accordance with contamination of the Artibonite River in a way that could infect thousands, and kill hundreds, of persons within a few days. (Piarroux et al., 2011)
Malaria incidence proportion in each region in Ethiopia was higher than in the country. Many spatial clusters were identified. Most of the northern part after modelling is predicted to have lower incidence. (Taddese, Baraki and Gelaye, 2019)
The spatial tools used in epidemiology over a decade are presented. Paper concludes that space and location are the main dimensions of epidemiology. Nowadays, there is a wide range of spatial tools to model public health data. (Auchincloss et al., 2012)
This paper presents the mathematical background of a multivariate time series model using Negative Binomial distribution to predict the parameters in the model. (Cheng et al., 2016)

Table 2.1: Conclusions from papers read as a prior step of literature review

Searching Results					
No.	Database	Search Statement	Number of Results Returned	Number of articles read	Date Searched
1	Google Scholar	covid-19 spatio temporal	30,500	4	20/04/2021

2	Google Scholar	”covid 19” spatio temporal analysis	28,100	4	21/04/2021
3	Google Scholar	spatial evolution of covid-19	205,000	3	21/04/2021
4	Google Scholar	moran’s i statistic covid-19	16,300	1	22/04/2021
5	Google Scholar	moran’s I statistic and diseases	21,500	1	22/04/2021
6	Google Scholar	ebola virus spatial modelling	6,350	1	22/04/2021
7	Google Scholar	cholera epidemic spatial analysis	26,600	1	22/04/2021
8	Google Scholar	malaria spatial modelling	63,500	1	22/04/2021
9	Google Scholar	epidemics spatial analysis	260,000	1	22/04/2021
10	Google Scholar	spatial epidemiology methods	1,080,000	1	22/04/2021
11	Google Scholar	spatio-temporal spread of covid-19	10,400	1	23/04/2021
12	Google Scholar/From 11 referencing list	spatio-temporal spread of covid-19	10,400	2	23/04/2021

Table 2.2: Papers read as a prior step of literature review

2.3 Useful R packages

During the implementation of the project, R statistical programming language will be used. R provides a huge variety of tools for spatial data analysis and advances in data representation and methods. R contains several packages which may be useful for the project.

Initially, some basic packages needed are ‘tidyverse’, ‘dplyr’, ‘ggplot2’. An important part of spatial research is to visualize the results in a way that insights can be drawn easily. A map is a great representation of such data, thus R package ‘tmap’ (Tennekes, 2018) is a powerful and flexible map-making tool. It is somehow like ’ggplot2’, however it has the unique capability to generate static and interactive maps using the same code via ’tmapmode()’ and it accepts a wider range of spatial classes. Another package is the ‘sp’ (Bivand, Pebesma and Gomez-Rubio, 2013a) which includes methods and classes for spatial data types. Using this package, we can create points, lines, polygons and grids to work with them. The package ‘raster’ (Hijmans et al., 2015) has several capabilities. It is used for reading, writing, manipulating, analysing and modeling of spatial data. This package is suitable for raster data especially when raster files are extremely large. As for spatiotemporal modeling and methods for epidemics, such covid-19, there is a very useful package called ‘surveillance’ (Meyer, Held and Höhle, 2017). This package includes statistical methods that allow modelling, as well as techniques for working with time series of counts, proportions, categorical data or continuous over time. All these methods aim to model epidemic phenomena. It is appropriate for implementing future evolution of a disease and detection such as the negative binomial method of Held and Paul. In order to evaluate the spatial autocorrelation between regions or within a region the package ‘spdep’ (Bivand, Pebesma and Gomez-Rubio, 2013b) contains a collection of functions to create spatial weights matrix, a collection of tests for spatial autocorrelation, such as global Moran’s I statistic and many more different functions.

2.3.1 R-INLA package

The integrated nested Laplace approximation is the full name of INLA. R-INLA package (Moraga, 2019) in R contains a lot of useful models fitting spatial data. As in all models fitted in R the linear predictor of the model is written as a formula object in R and then the model is executed using inla() function which takes as main arguments the formula, the family and the data. Output of the model displays the information of the fitted model including summaries, posterior marginals of the parameters, linear predictors, and the fitted values. The posteriors are evaluated through functions provided by the package. The key advantages of R-INLA are the ease with which complex models can be created and modified, without the need to write complex code, and the speed at which inference can be done even for spatial problems with hundreds of thousands of observations.

Also, R-INLA provides some Bayesian modelling performance and selection criteria. Hence, the fit of a model in the data and the selection of the most appropriate model could be achieved easily. These criteria must be applied when calling the formula inla() of calculating the summary information of the model. In this study three criteria are used the Conditional predictive ordinate (CPO), the Deviance information criterion (DIC) and the Widely applicable Bayesian information criterion (WAIC). Generally smaller values of the three criteria indicate better fit of the model.

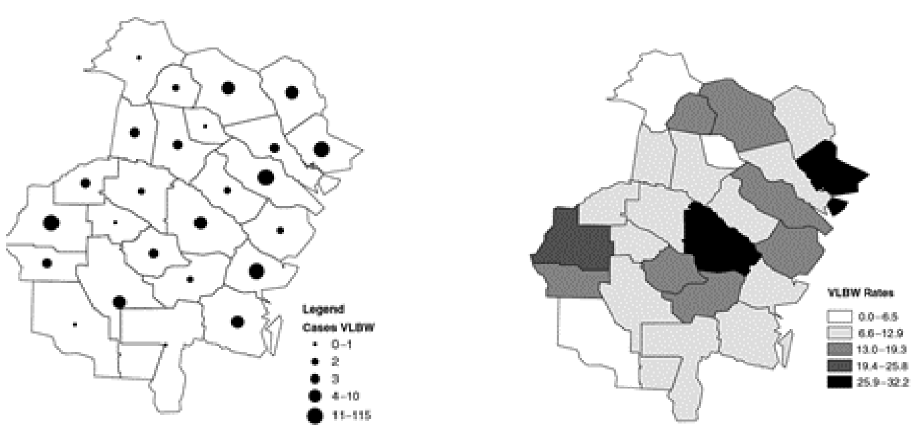
The aforementioned packages are just a small sample of the huge variety of packages that R provides in the study of spatial data analysis.

Chapter 3

Literature Review

3.1 Areal data visualizations

There are various ways to visualize spatial areal/lattice data including classed symbol maps, choropleth maps, three-dimensional display, dot density maps and other maps. Thus, choropleth maps are the most common type of areal data visualization is the one chosen for the purposes of this project. Classed Symbol Maps are suitable for displaying disease rates or exposure values which differ in space, appropriate for Covid-19 data. In such maps each area is coloured according to the category in which the corresponding value falls. Some choropleth visual examples are shown in Figures 3.1.



(a) Fig 1 (b) Fig 2

Figure 3.1: Fig 1: Graduated symbol map of the number of cases of very-low-birth-weight (VLBW) per county in Georgia Health Care District 9 (GHCD9) (Waller and Gotway, 2004a). Fig 2: Choropleth of VLBW rates per 100 people per year in GHCD9.
(Waller and Gotway, 2004c)

3.2 Spatial Autocorrelation Measurements

Correlation reflects the relationship between two variables. For instance, usually variable values retrieved from individual observations are related. The prefix auto- means self. Hence, autocorrelation may be defined loosely as the relationship among observations of a single variable. The adjective spatial describes the source of this selfcorrelation, which is the two-dimensional ordering of data values. Therefore, spatial autocorrelation refers to the dependencies that exist among observations that are attributable to the relative locations, or underlying twodimensional ordering, of variable values in geographic space. In turn, these dependencies produce clustering of similar (positive spatial autocorrelation) or dissimilar (negative spatial autocorrelation) values, and hence induce some map pattern.

Spatial autocorrelation studies the correlation between values which measure the same quantity at different locations within the study area. There are two main types of spatial autocorrelation which are the so-called global and local spatial autocorrelation.

3.2.1 Global Spatial Autocorrelation

Aims to measure to what extend similar observations tend to occur near each other. Very large values of it in one direction indicate positive spatial autocorrelation, while very low ones indicate negative spatial autocorrelation. A global index of spatial autocorrelation provides a summary over the entire study area. The basic form of global index of spatial autocorrelation is given by,

$$\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} sim_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}},$$

where sim_{ij} is the similarity between data values Y_i and Y_j, w_{ij} denotes a weight describing the proximity between locations i and j , for i and $j = 1,...,N$.

3.2.2 Method Moran's I

Moran's I index, denoted by I, is the most common used global index of spatial autocorrelation. It is structured as the basic form defined above. The mathematics are accessed from (Waller and Gotway, 2004d). The formula of testing the similarity between regions i and j is defined as,

$$I = \frac{1}{s^2}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(Y_i-\bar{Y})(Y_j-\bar{Y})}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}},$$

where $s^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i-\bar{Y})^2$ is the sample variance for vector of observations Y for i and $Y = \frac{1}{N}\sum_{i=1}^{N} Y_i$. I is a random variable with a sampling distribution depending on the interaction between the Y_i. Positive values of I indicate that neighbouring regions are having similar values, otherwise are clustered. On the other hand, negative values indicate that values of neighbouring regions are different. If values share similar properties, it means that observations are not independent. This is the key difference from classical statistics in which data assumes independence among observations. No correlation leads to among nearby values leads to the expectation of I,

$$E[I] = \frac{-1}{N-1}$$

which tends to zero as N increases. Values of Moran's I statistic are constrained within the interval $[-1,1]$. Variance of I is given by,

$$Var(I) = E[I^2] - E[I]$$

If $I = -1$ (perfect dispersion) it specifies that there is a perfect clustering of different values. On the contrary, $I = 1$ (perfect clustering) indicates a perfect clustering of alike values, while $I = 0$ (perfect randomness) shows that no autocorrelation exists and there is perfect randomness. Figures 3.2 and 3.3 show these patterns.

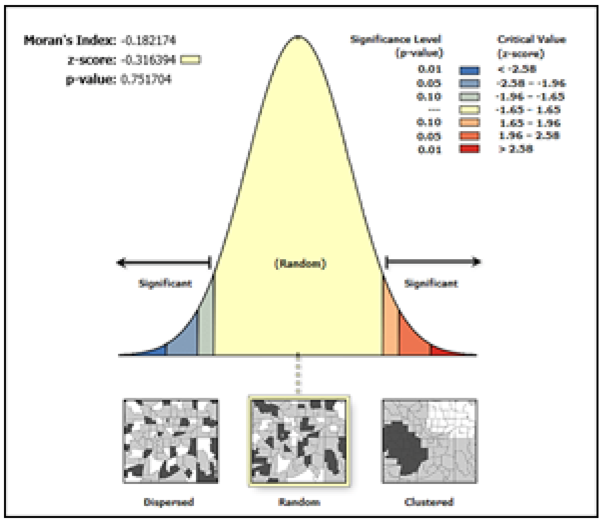


Figure 3.2: Report of the Moran's I (Spatial autocorrelation) statistic showing that the pattern was clustered. p-Value: probability; z-score: standard deviation (Papa, Sidira and Tsatsaris, 2016).

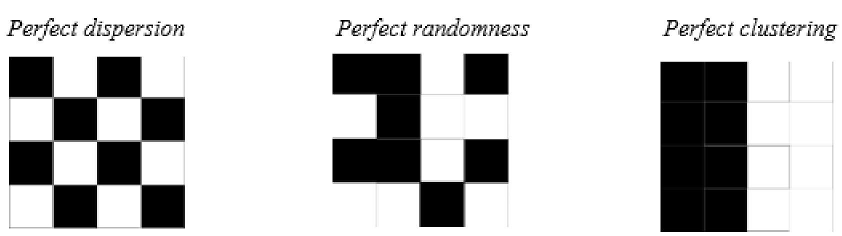


Figure 3.3: The images illustrate the spatial patterns using Moran's I statistic (Glen, n.d.).

A test of significance will derive the result of the Moran's I statistic. The hypothesis test calculates a z-score and its corresponding p-values. The null hypothesis indicates that data is randomly disbursed. The alternative hypothesis implies that the data is spatially clustered and there are two possible outputs. If z-value > 0 then data is spatially clustered in some way or if z-value < 0 then data is clustered in a competitive way.

$$\text{z-score statistic: } z = \frac{I - E[I]}{\sqrt{Var(I)}}$$

3.2.3 Local Spatial Autocorrelation Measurements

The purpose of global indexes is only to detect spatial associations averaged over the entire study area. In addition, a global index can suggestion clustering but cannot identify individual clusters. So local forms are needed, Anselin (Anselin, 1995) termed local indicators of spatial association (LISAs). Such indicators aim to provide a local measure of similarity between each region's associated value (in our case, a count, or an incidence proportion) and those of nearby regions. LISA values for each region are elements of the global indicator as they sum to the global index for spatial autocorrelation. Therefore, LISAs are usually derived from their corresponding global form. More often maps designed to display each region LISA values to allow comparisons between neighbouring regions to be drawn. An example illustration of LISA is shown in Figure 3.4 .

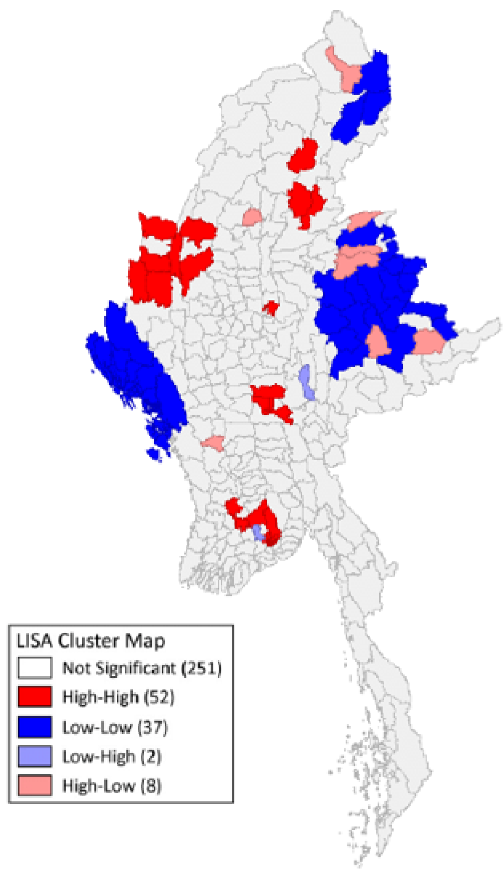


Figure 3.4: Local Indicators of Spatial Association (LISA) Cluster Map of Adult Literacy at the Geographic Level of Myanmar Townships (MacTavish, Rémillard and Davison, 2018).

The basic form of the LISA for a region i is given by,

$$\sum_{i=1}^{N} w_{ij} sim_{ij},$$

where sim_{ij} is the measure of similarity between regional observations.

3.2.4 Method: Local Moran's I

It is the most common used LISA indicator. Local Moran's I for a region i is denoted by I_i and given by,

$$I_i = \sum_{j=1}^{N} w_{ij} sim_{ij} = \sum_{j=1}^{N} w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y}) = (Y_i - \bar{Y}) \sum_{j=1}^{N} w_{ij}(Y_j - \bar{Y}),$$

Most authors (e.g., (Anselin, 1995), (Lee and Wong, 2001)) divide each deviation from the overall mean by the overall variance of the Y_i values, yielding,

$$I_{i,std} = \frac{(Y_i - \bar{Y})}{s} \sum_{j=1}^{N} (w_{ij} \frac{(Y_j - \bar{Y})}{s},$$

where s is the square root of the sample variance of the vector Y of observations in i region, the 'std' subscript implies that standardized scaling has been applied for each region.

3.3 Models

Modelling the data and observing the evolution of a disease (Covid-19 pandemic), evaluating the basic reproduction number we can initially investigate the efficiency of the several measures taken by the government and at the same time to introduce new possible effective measures that can curb Covid-19 spreading.

3.3.1 Poisson Autoregressive Model

The log-linear version of Poisson autoregressive model (Ott, 2018) is considered. The model allows us to predict the number of infections at a time t, given past information up to $t-1$. The model is given by,

$$y_t|F_{(t-1)} \sim Poisson(\lambda_t)$$
$$\log(\lambda_t) = w + \alpha \log(1 + y_{t-1}) + \beta\, log(\lambda_{t-1}),$$

where $\{y_0, ..., y_t\}$, $y_t \in N, w, \alpha, \beta \in R.$, w is the intercept, and α, β demonstrate how the number of new infections λ_t relates to the past ones.

- α: represents the short-term dependence on the previous time point.
- β: represents a trend component, that is, the long-term dependence on all past values of the observed process.

The model can be estimated by a maximum likelihood method. The model can be used for every region for example a city, a country, a municipality etc.

3.3.2 Standardized Incidence Ratio

In such epidemiological diseases' investigation, we often wish to provide a metric of disease risk estimation in each of the areas of the main study region. Therefore, we will introduce a method which will give us some Covid-19 infectious risk estimates in each municipality/-community in Cyprus.

One simple measure of disease risk is the standardized incidence ratio (SIR) (Moraga, 2019). For each area $i, i = 1, \ldots, n$ (municipality/community $i, i = 1, \ldots, 615$) the SIR is defined as the ratio of observed (Covid-19 cases) counts to the expected (Covid-19 cases) counts.

$$SIR_i = \frac{Y_i}{E_i}$$

The expected counts E_i represent the total number of cases that one would expect if the population of the area behaved the way the standard (or regional) population behaves. E_i can be calculated using indirect standardization as,

$$E_i = m \sum_{j=1}^{m} r_j^{(s)} n_j^{(i)}$$

where $r_j^{(s)}$ is the rate (number of cases divided by population) in stratum j in the standard population, and $n_j^{(i)}$ is the population in stratum j of area i. In applications where strata information is not available, we can easily compute the expected counts as,

$$E_i = r^{(s)} n^{(i)}$$

where $r^{(s)}$ is the rate in the standard population (total number of cases divided by total population in all areas), and $n^{(i)}$ is the population of area i. SRI_i indicates whether area i has higher ($SRI_i > 1$), equal ($SRI_i = 1$) or lower ($SRI_i < 1$) risk than expected from the standard population. When applied to mortality data, the ratio is known as the standardized mortality ratio (SMR).

Strengths

The SIR is simple to evaluate and is not only a simple ratio. It is often adjusted for factors/stratum such as gender, age, race e.t.c.

It is considered the best method when using multiple stratum, for time series analyses.

Mapping of trends as time passes is relatively simple to implement.

Limitations

SIR calculation is a great and useful tool in spatial disease risk estimation. However, its reliability is decreasing when the population of an area is small or there is a rare disease. Cyprus's municipalities/communities have small number of residents according to the rest of European countries. Hence, using SIR the expected cases per week on each municipality/community may be low and the result will be misleading and insufficient.

Moreover, the evaluation of just the SIR is not enough. In order to determine the efficiency of SIR and whether the observed number of cases is significantly different or not from the expected number, a 95% confidence interval (CI) is good to be calculated for each SIR. More precisely, a 95% CI is the range of estimated SIR values that have a 95% probability to include the actual values of SIR for the population.

This method requires the stratum to be available for each area in the examined dataset. However, sometimes this is not the case and a pre-processing is needed before applying SIR.

3.4 Bayesian Models

Therefore, the use of Bayesian models is preferable to estimate disease risk. Models are reliable enough as they can hold information of neighbouring regions and incorporate covariates information resulting in the smoothing or shrinking of extreme values based on small sample sizes. Bayesian models can even model the complexity of the spatial data structure and they can face any weaknesses of SIRs.

For each area, $i, i = 1, ..., n$ the observed counts, Y_i, are modelled using a Poisson distribution with mean $E_i\theta_i$, where E_i is the expected counts in the i^{th} area and θ_i is the relative risk (RR) in the i^{th} area. The logarithm of the relative risk θ_i is expressed as the sum of an intercept that models the overall disease risk level, and random effects to account for extra-Poisson variability. The relative risk θ_i quantifies whether area i has higher ($\theta_i > 1$) or lower ($\theta_i < 1$) risk than the average risk in the standard population. For example, if $\theta_i = 2$, this means that the risk of area i is two times the average risk in the standard population.

The model is given below,

$$Y_i \sim Poi(E_i\theta_i), i = 1, ..., n$$

$$\log(\theta_i) = a + u_i + v_i, i = 1, ..., n$$

- Y_i: represents the observed counts in the i^{th} area,
- a: defines the overall risk in the study's region,
- u_i: represents the random effect of the i^{th} area for modelling spatial dependence among relative risks,
- v_i: is the uncorrelated noise of the model, where $v_i \sim N(0, \sigma_v^2)$.

For a more precise model a virtuous approach is the addition of covariates admissible to risk factors and supplementary random effects for better explanation of the variability in the model. For example, $log(\theta_i)$ can be expressed as,

$$\log(\theta_i) = d_i\beta + u_i + v_i$$

where $d_i = (1, d_{i1}, ..., d_{ip})$ is the vector of the intercept and p covariates corresponding to the i^{th} area, $\beta = (\beta_0, \beta_1, ..., \beta_p)'$ is the vector of the coefficients.

Thus, the interpretation of a coefficient β_j is that for one unit increase in covariate $d_j, j = 1, ..., p$ the relative risk increases by a factor of $exp(\beta_j)$ when the other covariates are hold constant.

3.4.1 BYM model

One of the most popular and usable models in spatial statistics investigating disease propagation is the Besag-York-Mollié (BYM) model (Riebler et al., 2016).This model was firstly introduced by Clayton and Kaldor and then extended by Besag, York, and Mollie and its formulation is described as: In this model the spatial random effect follows a Conditional Autoregressive (CAR) (Cramb et al., 2018) distribution which is based on queen neighbours which share same boundaries. The CAR distribution aims to smooth the data according to a queen criteria neighbourhood. The spatial random effects follow a normal distribution given below,

$$u_i|u_{-i} \sim N(\bar{u_{\delta_i}}, \frac{\sigma_u^2}{n_{\delta_i}})$$

$\bar{u_{\delta_i}} = n_{\delta_i}^{-1} \sum_{j \in \delta_i} u_j$ and δ_i represents the set of neighbours of the i^{th} area and, n_{δ_i} represents the number of neighbours of the i^{th} area. The noise follows a normal distribution with zero mean and variance $\sigma_v^2, v_i \sim N(0, \sigma_v^2)$.

The CAR model has been proposed by Besag et al. (Best, Richardson and Thomson, 2005) hence during the implementation in R via R-INLA package an argument defines the distribution has the name 'besag'.

Strengths

It is one of the most popular Bayesian models and has the ability to generate more reliable estimates as it draws information from neighbouring areas.

Also according to a paper by Best et al. (Best, Richardson and Thomson, 2005) the BYM model is the only model which have been used excessively in real word problems and especially disease mapping applications.

Limitations

As this model is often used accompanied with CAR distribution sometimes does not fit well on irregular areal/lattice data applications where entities are of different sizes and shapes.

The BYM characterizes of its strong smoothing effect which sometimes leads to inaccurate relative risks as is unable to detect potential increases in locally risks.

In the BYM model the elements v and u are not interpreted as independent components as the independent random effect can only be shown in the case where in spatial random effect partially when no spatial dependence occurs. To resolve this issue, priors should be heavily dependent.

3.4.2 BYM2 model (different parametrization)

A lot of research has been achieved and various papers have been published regarding spatial methods and models. Hence, in 2016 a new parametrization of the BYM model has been introduced by Simpson et al. (Riebler et al., 2016) named BYM2 which makes parameters of the model interpretable and incorporates meaningful Penalized Complexity (PC) priors. Following this model there is a scaled spatially structured component u_* and an unstructured component v_*.

$$b = \frac{1}{\tau_b}\sqrt{1-\phi}v_* + \sqrt{\phi}u_*$$

where, τ_b is a precision parameter > 0 which controls the marginal variance contribution of the weighted sum of u_* and v_* and ϕ is the mixing parameter, $0 \leq \phi \leq 1$ which measures the proportion of the marginal variance explained by the structured effect u_*.

For $\phi = 1$ the BYM2 model equals to an only spatial model and for $\phi = 0$ the BYM2 model equals to an only unstructured spatial noise (Moraga, 2019).

Strengths

In the BYM2 model the spatial structured effect is scaled. The model makes use of the penalized complexity framework and hence, interpretation and transition of priors is achievable.

The marginal precision of v and u is interpreted independently.

3.5 Use of Space and Time in Covid-19 data

Measured values can either be recorded at some locations (points, areas) or times (moments, intervals). The main interest of a study or a matter of convenience specify the type of indices selected. When dealing with some disease cases, like covid-19, spatial set can be a collection of points and/or times, their registered locations and times indicate where and when they were or occurred. For covid-19 data is collected daily and locally, thus time and location on which values are collected, will be the set of indices. In this case, spatiotemporal analysis methods are used to analyse such data.

On March 9th of 2020 the first positive cases of covid-19 make their appearance on the Cyprus island. Without any vaccination or medicine, the means of controlling it were limited to quarantine and social distancing. Hence, government of Cyprus had to impose some measures to prevent the rapid spread of coronavirus in the island. During the first wave of covid-19 the first measure taken was the strict quarantine for about two to three months. Inferences about the disease arise by investigating the time evolution of the infection, death, and recovery rates which affected by policies. At the beginning of the pandemic the only way of detecting positive cases was with the clinical PCR test which required some number of days from the laboratories to extract the results. Consequently, the infection rate at the first months of the pandemic is close to a constant probably due to the lack of means to detect infected individuals before infection symptoms are observed. Now, rapid tests which are much faster can detect infected people immediately within some minutes.

Chapter 4

Data Analysis

On this chapter a description of the data is presented. An exploratory analysis is conducted and its main conclusions are shown. A main characteristic of spatial data is the presence of dependency among neighbouring observations/locations which is expected to be weaker during lockdown measures. Therefore, the spatial autocorrelation at municipality/community level of the number of confirmed Covid-19 cases on a weekly basis is measured using Moran's Index based on queen neighbours. Moran's I is the most common measure of global spatial autocorrelation which gives the overall distribution of departures from randomness. Both global and local Moran's I and local indicators of spatial association (LISA) are presented at the municipality/community level to provide information on spatial clusters. LISA allows for the decomposition of global Moran's I inferring the scope of clustering in smaller areas. Similarly, LISA calculation was based on Queen's contiguity neighbours and the statistical significance of the pattern of spatial autocorrelation in each county municipality/community was tested at 5% level of significance. The municipalities spatial dependence was plotted on the cluster map which is colour according to the type of association. Only high-high spatial clusters are presented with red colour and indicate a positive local spatial correlation. All statistical analyses were implemented in R statistical software version 4.0.5. The code used to conduct this data analysis can be found at https://github.com/ml5298/project.

4.1 Data Sources

The project consists of three main datasets from where variables and dataframes are derived. For this study, relevant data were collected from different sources. Demographic/Census data of Cyprus were retrieved from the Statistical Service of Cyprus, polygons spatial coordinated data of Cyprus has been accessed via the official website of Cyprus data with the link https://data.gov.cy/ and Covid-19 data cases across the country were retrieved from the Ministry of Health.

4.1.1 Demographic data

Demographic data is a statistical data collection about the characteristics of the population, e.g., age, gender, and income etc. The Census of Population is one of the main sources of statistical data which is used in policy development in fields such as education, health, employment, housing, rural development etc. Since the Turkish invation in 1974, the north part of the island of Cyprus has been occupied by Turkey and the government of Cyprus is not able to exercise any authority in that part, so it is excluded from our study. The Census was conducted according to the relevant EU Regulation (Regulation (EC) No. 763/2008 of the European Parliament and of the Council) and it covered the Government controlled area of Cyprus, excluding the Turkish occupied north part of Cyprus.

The aim of the Census of Population was the enumeration of the entire population and the collection of information on the demographic and social characteristics of the population, on the size and amenities of it as well as the geographical distribution of the population. The dominant source of demographic data of Cyprus is the Statistical Service of Cyprus and in particular the Demography, Social Statistics, Labour and Tourism department. Data were collected through CAPI (Computer Assisted Personal Interviewing) method, using netbook computers on which the Census of Population Questionnaire was installed. Enumerators of the Statistical Service visited every housing unit and with the assistance of the household members, completed the Census Questionnaire. The reference date of the Census was the 1st of October 2011 with the last updates imposed in 2013. Population census in Cyprus is conducted every 10 years with the last one on 2011. The new census will start conducted on Autumn 2021.

The entire Government-controlled area was delimitated in clear geographical boundaries, as follows: each district is subdivided in the greater urban and rural area, according to the definition provided by the Department of Town Planning and Housing. A further delimitation was conducted with regard to municipalities, communities, and quarters (where this delimitation exists) and moreover, a further delimitation in enumeration blocks took place. For the purposes of the present study data is collected with regards to municipalities and communities. Figure 4.1 and 4.2 display a visual mapping representation of the 6 districts of the Cyprus island with a polygon structure reflecting the municipalities/communitites.

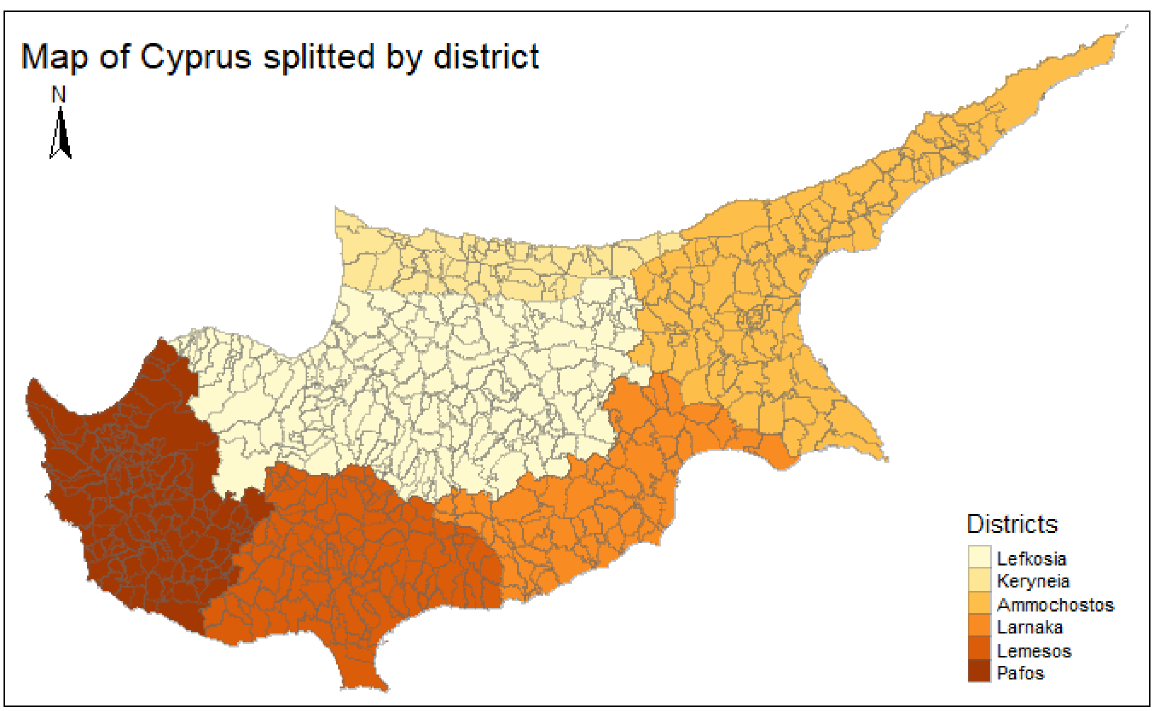


Figure 4.1: Map of Cyprus at municipality/community level split by district.

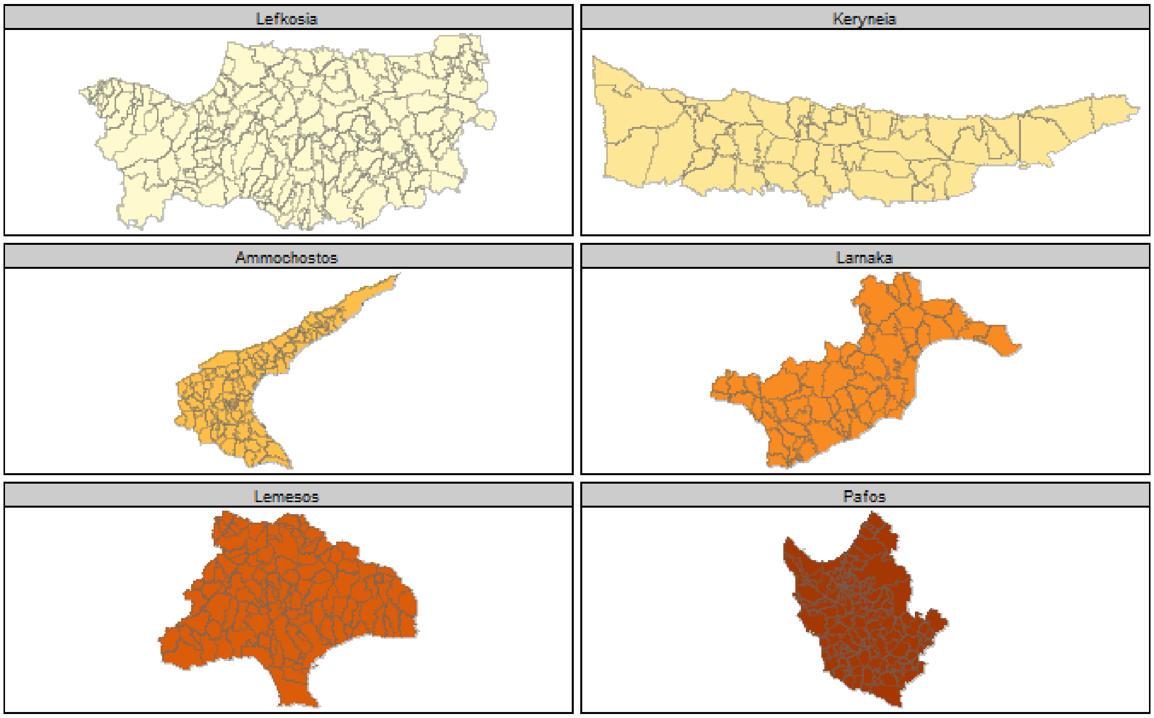


Figure 4.2: Polygons spatial representation of each district of Cyprus at municipality/community level.

The source of assessing the data is the Statistical Service of Cyprus. However data from the official site of CyStat are in the form of tables separately for each category e.g. Education Attainment, Population e.t.c. Therefore, with the use of Microsoft Excel 365 a new data file is created with all the information collected. Census data have the form of a table with rows all the different districts, municipalities, and communities. Each column in the data set represents a different metric of each row. Demographic data consists of 413 records and 82 variables. There are 408 records represent a unique municipality or community in the Republic of Cyprus and 5 records refer to the 5 districts municipalities of the country. It is of great attention that regions belong to the Northern part are not considered in the data. The 82 variables are grouped into seven main categories namely Basic Information, Economic Activity, Nationality, Mean age by gender, Age groups per gender, Educational attainment Male and Educational attainment Female and are presented in the table 4.1 below.

Variable Name	**Description**
Basic Information	
Code	A unique code characterizes the municipality or district.
District Code	A unique code characterizes the district: 1-Lefkosia, 3-Ammochostos, 4-Larnaka, 5-Lemesos, 6-Pafos.
Name	The name in Latin characters of each district/municipality/community.

Area (km²)	The space occupied by each district/municipality/community. Area is measured in square units such as square kilometres.
Economic Activity	
conomically active population	Comprised all persons aged 15 and over who during the week prior to the interview were either employed or unemployed (as defined above).
Total unemployed	Those aged 15 years or older (on the reference date of the Census) who were not employed during the week prior to the interview, but were actively looking for full-time or part-time work, were available for work and were also able to work.
Total employed	Considered those who were 15 years of age or older (on the reference date of the Census), who during the week which preceded the week of the interview, worked for at least one hour for pay, profit, or as unpaid family helper. It also includes persons who had a job but who happened not to work during that particular week because they were on leave, sick leave etc.
In the primary sector	This sector includes Agriculture, Forestry and Fishing, Mining and Quarrying.
In the secondary sector	This sector includes Manufacturing, Electricity, Gas, Steam and Air Conditioning Supply, Water Supply, Sewerage, Waste Management and Remediation Activities, and Construction.
In a tertiary sector	This sector includes Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles; Transportation and Storage; Accommodation and Food Service Activities; Information and Communication; Financial and Insurance Activities; Real Estate Activities; Professional, Scientific and Technical Activities; Administrative and Support Service Activities; Public Administration and Defence; Compulsory Social Security; Education; Human Health and Social Work activities; and other service provision activities.
Not stated	The sector of employment is not stated.
Nationality	

Total Population	Persons who had their usual residence in Cyprus on 1st October 2011, i.e., persons who resided permanently in Cyprus for one year or longer, or persons who had come to Cyprus during the last twelve months (prior to the reference date of the Census) with the intention of residing in Cyprus for a period of at least one year. The total number of populations in the district/municipality/community.
Male Population	The population of males in the district/municipality/community. Persons who had their usual residence in Cyprus on 1st October 2011, i.e. persons who resided permanently in Cyprus for one year or longer, or persons who had come to Cyprus during the last twelve months (prior to the reference date of the Census) with the intention of residing in Cyprus for a period of at least one year.
Female Population	The population of females in the district/municipality/community. Persons who had their usual residence in Cyprus on 1st October 2011, i.e. persons who resided permanently in Cyprus for one year or longer, or persons who had come to Cyprus during the last twelve months (prior to the reference date of the Census) with the intention of residing in Cyprus for a period of at least one year.
Total Cypriots	Residents with Cypriot nationality.
Male Cypriots	Male residents with Cypriot nationality.
Female Cypriots	Female residents with Cypriot nationality.
Total EU citizens	Residents who are from other European countries.
Male EU citizens	Male residents who are from other European countries.
Female EU citizens	Female residents who are from other European countries.
Total Non-EU citizens	Residents who are from other non-European countries.
Male Non-EU citizens	Male residents who are from other non-European countries.
Female Non-EU citizens	Female residents who are from other non-European countries.
Total nationality Not stated	Residents with not available nationality.

Male nationality Not stated	Male residents with not available nationality.
Female nationality Not stated	Female residents with not available nationality.
Mean age by gender	
Total Mean Age	The average age in each district/municipality/community.
Male Mean Age	The male average age in each district/municipality/community.
Female Mean Age	The female average age in each district/municipality/community.
Age groups per gender	
Male 0-4	The population per age group and gender in each district/municipality/community. Same for the below groups.
Female 0-4	
Male 5-9	
Female 5-9	
Male 10-14	
Female 10-14	
Male 15-19	
Female 15-19	
Male 20-24	
Female 20-24	
Male 25-29	
Female 25-29	
Male 30-34	
Female 30-34	
Male 35-39	
Female 35-39	
Male 40-44	
Female 40-44	
Male 45-49	
Female 45-49	
Male 50-54	
Female 50-54	
Male 55-59	
Female 55-59	
Male 60-64	
Female 60-64	
Male 65-69	
Female 65-69	
Male 70-74	
Female 70-74	
Male 75-79	

Female 75-79	
Male 80+	
Female 80+	
Total Age Not stated	
Total Age Male Not stated	
Total Age Female Not stated	
Educational attainment Male	
Up to lower secondary (gymnasium) Male	For each respondent who was 15 years of age or older, the highest level of education which s/he had completed on the reference date of the Census, was recorded. Same for the below educational groups.
Upper Secondary (Lyceum/ Technical/ Vocational) Male	
Post-secondary non-tertiary Male	
Tertiary level (non university) Male	
Tertiary level (University - First degree) Male	
Tertiary level - Post-graduate degree Male	
Tertiary level - Doctorate Male	
Not stated education Male	
Educational attainment Female	
Up to lower secondary (gymnasium) Female	
Upper Secondary (Lyceum/ Technical/ Vocational) Female	
Post-secondary non-tertiary Female	
Tertiary level (non university) Female	
Tertiary level (University - First degree) Female	
Tertiary level - Post-graduate degree Female	
Tertiary level - Doctorate Female	
Not stated education Female	

Table 4.1: Description of the Demographic dataset created, with its information obtained from the Statistical Service of Cyprus.

Modifications/Additions

Except from the above variables, 8 more derived variables are created. More precisely, age grouping is split in just 4 age groups per gender (male and female). For both gender we have children (0-19 age), youth (20-39 age), middle ages (40-59) and seniors (60-80+).
The new variable names are ”children_male”, ”children_female”, ”youth_male”, ”youth_female”,

"middle_ages_male", "middle_ages_female", "seniors_male", "seniors_female", "childre_total", "youth_total", "middle_ages_total", "seniors_total".

The distribution of the population by district is given in Table 4.2 below,

POPULATION BY DISTRICT, AS AT 2011 CENSUS

District	Population (2011 Census)	Males	Females
Lefkosia	326,980	158,262	168,718
Ammochostos	46,629	23,188	23,441
Larnaka	143,192	70,116	73,076
Lemesos	235,330	113,636	121,694
Pafos	88,276	43,578	44,698
Total	**840,407**	**408,780**	**431,627**

(Last Updated 23/05/2012)

Table 4.2: Distribution of the Population per districts derived from the Statistical Service of Cyprus.

The population is distributed in the 5 different districts. The total enumerated population was $840,407$, of which $408,780$ are men and $431,627$ are women. Of the enumerated population the 38.9% resided in Lefkosia district, 28% in Lemesos, 17% in Larnaka, 10.5% in Pafos and 5.55% in Ammochostos. Moreover, there are more women in all districts. From the mapping of Cyprus according to gender population we can definitely observe that there is a spatial dependence among near municipalities/communities. The majority of Cypriot population is distributed in the cities, while mountain villages have a small number of residents. Moreover, maps show that municipalities near the 5 big city municipalities of Cyprus have also a large number of residents. Figures 4.3, 4.4, 4.5, 4.6 visualize a mapping representation of the main characteristics of the population which are the total number of residents and the divided by gender, employed or unemployed and age group at municipality/community level.

4.1.2 Cyprus Shapefile

For a spatial investigation of Covid-19 dissemination in Cyprus a shapefile with municipality/community boundaries is needed. The shapefile is published online, and it can be accessed from the link here. Shapefiles have been introduced and developed from Esri and they have become widely popular due to their convenient utility properties. Shapefile has become the most common file format in GIS. Despite its name indicating a singular file, a shapefile is actually a collection of at least three basic files: .shp, .shx and .dbf, .shp contains the geometry data, .shx: is a positional index of the geometry data that allows to seek forwards and backwards the .shp file, .dbf: stores the attributes for each shape. All three files must be present in the same directory for them to be viewable. More precisely, a shapefile is a geospatial vector data in which geographic features can be represented by points, lines, or polygons (areas).

The shapefile for Cyprus boundaries by municipalities/communities consists of polygon data structure. More precisely, each municipality/community is represented from a polygon. A polygon are 3 or more vertices that are connected and closed. The SpatialPolygonsDataFrame shapefile is imported using the 'rgdal' package and its function readOGR(). The shapefile was published on 18/12/2017 with some updates have been done on 26/05/2021 form its publisher the Department of Land Registry and Surveying. The data of the shapefile includes 615 features and 5 fields. There are 615 different villages/municipalities/communities, one per row and 5 informative variables/columns about them which are given in the below table 4.3.

Cyprus Shapefile	
Variable Name	**Description**
VIL_CCD	A unique code of each municipality/community.
VIL_NM_E	The Latin name of each municipality/-community.
VIL_NM_G	The Greek name of each municipality/-community.
DIST_CODE	The district depending on the district each municipality/community is included.
VIL_CODE	The village codes.

Table 4.3: Description of the data included in the Cyprus shapefile.

Modifications/Additions

1. Initially, the name of the variable 'VILL_NM_E' is replaced with the 'municipality_town_lst' in order to allow merging of dataframes based on identical municipality/community names.

2. The shapefile contains some identical names for two villages and each one belongs to different district. However, the names of these villages are being replaced with their district as an extension. For instance, AGIA VARVARA is renamed as AGIA VARVARA LEFKOSIA.

3. Data in the shapefile contain municipalities/communities from the occupied part of Cyprus. These regions are not considered in the study and any data about them is not available.

4.1.3 Covid data

The Ministry of Health gave me approval on accessing specific variables of the raw data of Covid-19 in Cyprus. Epidemiological and clinical data have been drawn from individuals with confirmed Covid-19 in Cyprus. Variables recorded for each confirmed case are shown in the Table 4.4.

Data cleaning on excel

Data needed some cleaning and wrangling before any use. As the data have been retrieved from the Ministry of Health as an excel file some initial data cleaning and wrangling has been done with Excel 365 before cleaning the data further in R. Initially, within the data there are some municipalities with a name of 0. Thus, these records are dropped from the data file. There are in total 991 such records. Also, there are 175 cases from abroad. As our study focuses only on Cyprus municipalities these records are dropped. British bases which have in total 62 cases are also dropped. Also, the names of municipalities/communities in the data file are rename to those in the shapefile accordingly. This step is extremely important as we are going to merge the two data files by region name. All records with blank district name are filled with the district in which the municipality belongs. There are some municipalities/communities with wrong district name and thus were corrected. Variables are described on Table 4.4

Covid data	
Variable Name	**Description**
Date_of_first_sampling	A unique code of each municipality/community.
Lab_Result_Date	The Latin name of each municipality/-community.
Municipality_Town_lst	The Greek name of each municipality/-community.
District_lst	The district depending on the district each municipality/community is included.
District_code	The village codes.
Age Group	The district depending on the district each municipality/community is included.
Sex_lst	The village codes.

Table 4.4: Description of the covid dataset obtained from the Ministry of Health.

Modifications/Additions

1. A loop has been constructed to identify if any municipalities/communities do not belong in the demographic data. As a result, there was one case assigned to municipality of ”TROODOS” which does not exist. ”TROODOS” is the general name of the mountainous municipalities/communities of Cyprus and includes several communities which are available separately in the Census data. Hence, as there is only one record like this and more possibly has been recorded accidentally is dropped from the dataset.

2. Covid-19 data have some dates which are not included in the study period, thus their corresponding record cases are dropped. There are in total 12 such cases.

3. Age groups of 0-4 and 5-9 are retrieved from the raw data as ’00-04’ and ’05-09’. Hence, in order to be consistent with the demographic data zeros in the first place are dropped and we get ’0-4’ and ’5-9’.

4. Moreover, cases which belong to ’80-84’, ’85-89’, ’90+’ will be included in the ’80+’ age group.

5. Values in some age groups records are of wrong format and thus they extract wrong content. More precisely, there are records with age group as ‘05-Sep’ and ‘Oct-14’. These values correspond to ‘05-09’ and ‘10-14 years’ respectively and are being replaced with the correct ones.

6. There are some Unknown gender values in the dataset that must be replaced with either a ’Female’ or ’Male’ values. Looping through the rows (Covid cases) in the dataset the Unknown values are substituted with either ’Female’ or ’Male’ based on some probabilities. These probabilities are derived as follows: using the function ’sample()’ in R a sample is extracted from the vector c(’Female’,’Male’) with probabilities the number of females and males devided by the total population in the current municipality/community respectively.

7. There are some Unknown age group values in the dataset that must be replaced with an age group depending on the gender of the current case. Looping through the rows (Covid cases) in the dataset the Unknown values are substituted with either ”0-4”, ”5-9”, ”10-14”, ”15-19”, ”20-24”, ”25-29”, ”30-34”, ”35-39”, ”40-44”, ”45-49”, ”50-54”, ”55-59”, ”60-64”, ”65-69”, ”70-74”, ”75-79”, ”80” based on some probabilities. These probabilities are derived as follows: using the function ’sample()’ in R a sample is extracted from the vector contains the different age groups with probabilities the number of residents per age agroup and gender(’Female’ or ’Male’) divided by the total population of ’Female’ and ’Male’ in the current municipality/community respectively.

8. Raw data contain information until the first days of June 2021. However, as information is missing for a lot of dates between April and June 2021 it is chosen to consider a time period until 28/03/2021.

The modified data constists of 42535 cases. The number of people in each category as defined above is extracted from the Demographic/Census dataset of Cyprus.

4.2 Potential Risk factors to be investigated in accordance with the measures imposed

An initial investigation from the scientific community has shown that Covid-19 virus spreads rapidly via human-to-human transmission. The dissemination of Covid-19 starts primarily via respiratory droplets and has an incubation period of up to 14 days in which mild, severe or no symptoms may appear. At the start of the pandemic where no mutations existed many infected people were asymptomatic, and this made it difficult to be identified. However, studies have shown that even asymptomatic people can transfer the virus to others. Treatment thus far has mainly focused on supportive measures.

Governments had imposed several measures such as social distancing, lockdown or even distance learning to reduce the probability of an infected individual to transfer the virus to a healthy one. Covid-19 had many unknowns, and the only immediate measures that governments could take were dominantly based on socializing reduction. During the first wave the elderly individuals, non-white, people with low socio-economic status, people who were living in a very populated area were more prone to be infected. Therefore, Covid-19 was associated with age, population density, health risks and conditions or even ethnicity. Hence some census maps are created in order to identify potential municipalities/communities with high risk of infection.

The majority of the population is gathered at the 5 cities and their closest municipalities/-communities. We can observe that in mountainous villages the population density is small. The distribution of the population per gender in each municipality/community is almost identical and villages seem to contain more females than males. Therefore, we expect more Covid-19 cases in the 5 main districts and their nearest municipalities/communities.

As for demographic data per gender Lemesos has the most children. Youth and middle ages are also gathered most in Lemesos, followed by the capital city Lefkosia. We can definitely detect that there is a minority of young people in mountainous areas while seniors are dominant in such villages. As a general conclusion is that nearby municipalities/districts tend to have similar population density per age group. The 5 main districts primarily have the largest population in each age group, followed from their neighbouring regions. Thus, investigating the impact of age group in the relative risk of infection may be found useful.

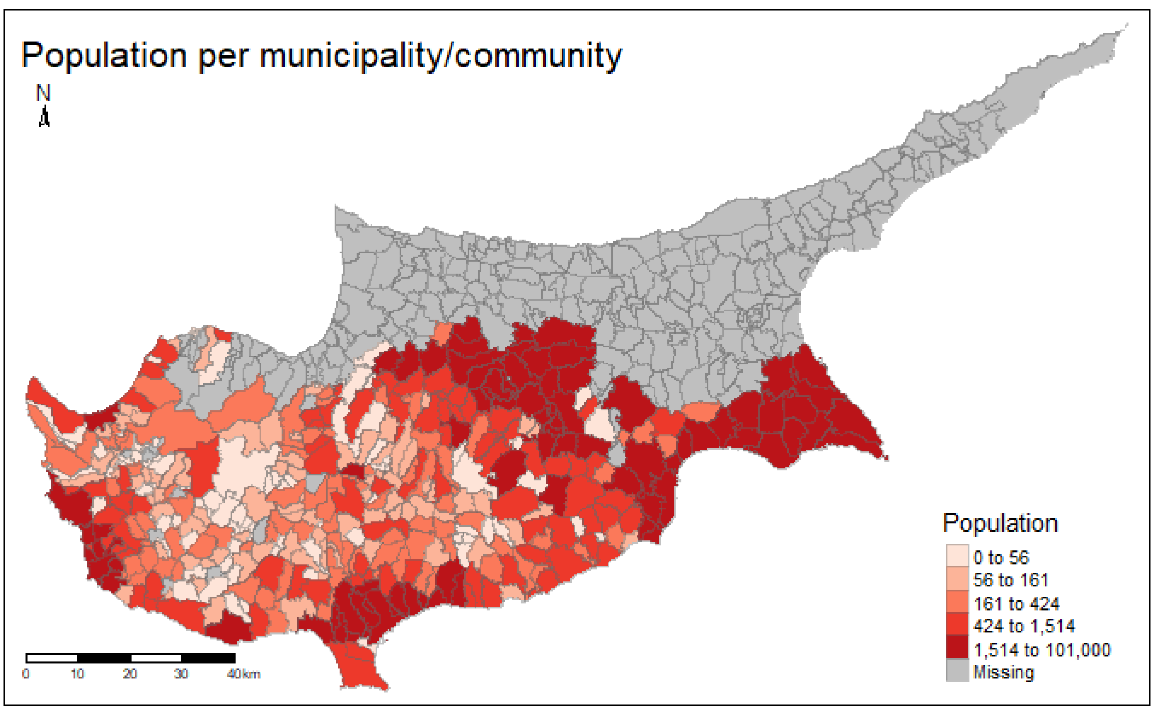


Figure 4.3: Population per municipality/community.

Figure 4.4: Population per gender at a municipality/community level.

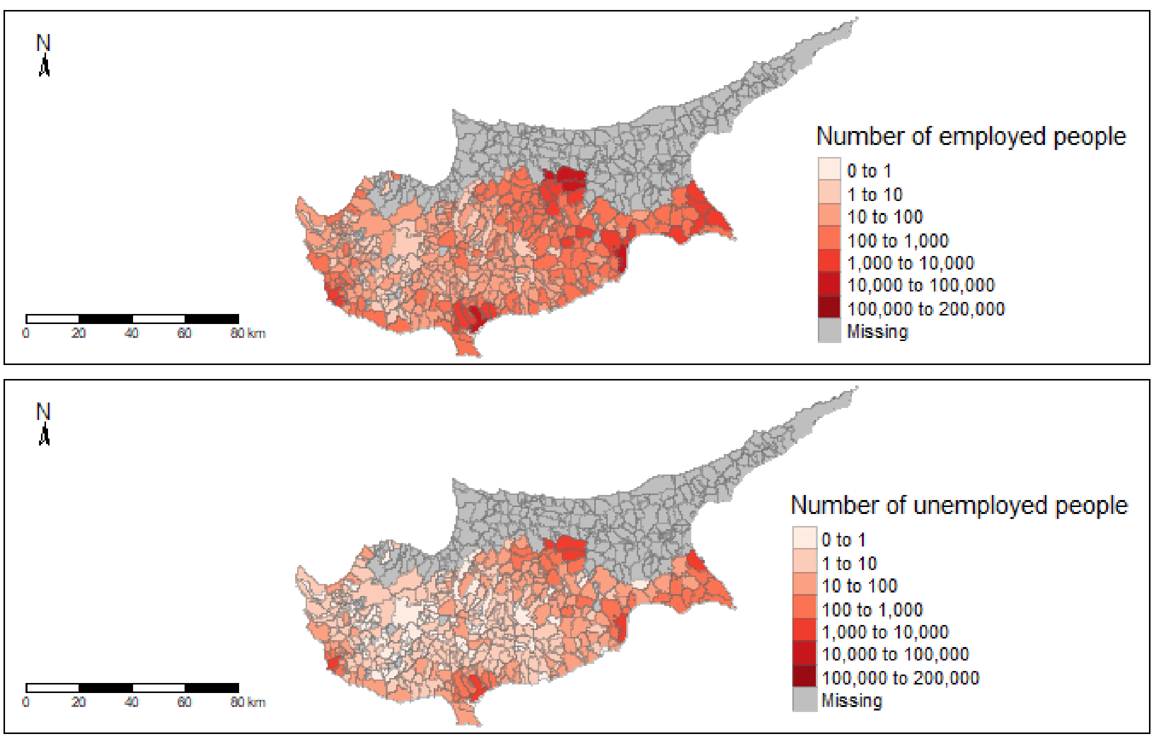


Figure 4.5: Number of employed and unemployed per municipality/community level.

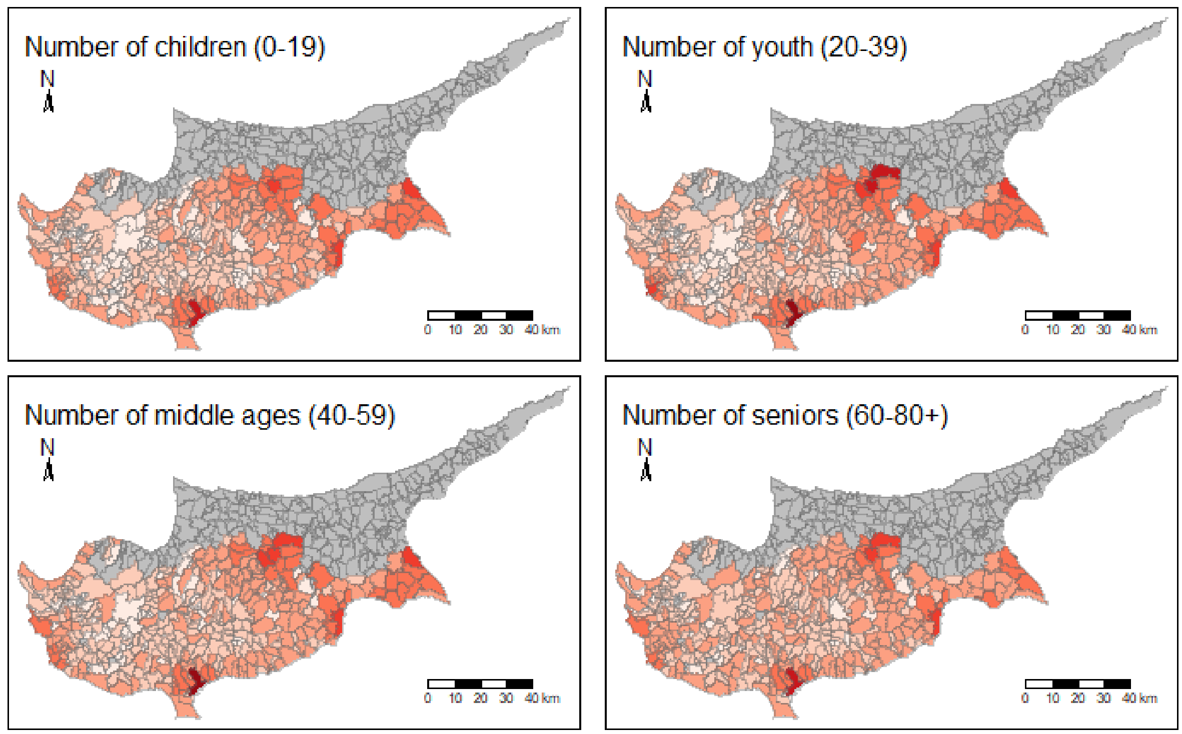


Figure 4.6: Population per age group defined by the four age group categories per municipality/community level.

4.3 Main dataset

The main data of this study is the number of newly daily Covid-19 cases in each municipality/community of Cyprus. After data cleaning and wrangling Covid-19 data in Cyprus has 42535 confirmed cases which were collected from 325 municipalities/communities in Cyprus from a total of 615 municipalities/communities in the country including those in the north occupied part of Cyprus. The spatiotemporal characteristics of Covid-19 were explored using spatial autocorrelation, hot spot, and spatiotemporal scanning statistics.

However, for this study's purposes the variable of interest would be the weekly number of confirmed Covid-19 cases in each municipality/community. The weekly study period is defined from 09/03/2020-28/03/2021. Therefore, there are in total 56 different dates and 55 weeks. As a starting date we have chosen the 9th of March 2020 as on this date the first Covid-19 case in Cyprus was confirmed. In the raw data retrieved from the Ministry of Health there are two date columns named 'date_of_first_sampling' and 'lab_result_date'. For this study the date used is the 'lab_result_date' as on that date possible cases were confirmed. Daily cases define exactly the number of confirmed cases, thus to this extent aggregating cases per 'lab_result_date' would agree with the number of daily cases. At the very beginning of the pandemic in the island laboratory results needed maximum 3 days for extracting a result. Afterwards through Rapid Antigen Tests sampling date was the same as result date.

For the analysis process a new Covid-19 dataset was created based on the initial (data from MoH Cy). In the initial dataset rows represent unique cases in the country. However, for the purposes of the current study it is essential to construct a dataset which will contain Covid-19 information for every municipality/community on a daily basis. Following this idea, a new dataframe named 'ts_covid_data' was created ('ts' stands for time series). The dataframe consists of 236775 rows and 45 columns. Each row displays Covid-19 information for a specific municipality/community on a specific date. We have in total 615 municipalities/communities (including those in North Cyrpus) and 385 unique dates within the study period 09/03/2020-28/03/2021. Although the dataframe is still not capable for analysis. We can notice that a lot of municipalities/communities which are included in the shapefile are missing from the Covid data. This happens because some regions belong to the Northern part of Cyprus, or some regions might not have any Covid-19 cases in this study period.

The dataframe was constructed as follows:

1. Primarily, a dataframe was produced which includes two columns, the name of each municipality/community and the date of the result with $615 \times 385 = 236775$ rows. The name of each municipality/community is repeated 385 times, for the 385 different dates.

2. All variable columns were derived from the initial Covid data after filtering according to some conditions, grouping data by municipality/community and test result date, and aggregating the number of rows. Thereafter, we joined data with the above dataframe to allow any dates on which no cases occur to exist in the dataset as missing values.

3. In R missing values are displayed with NA. However, a discrimination must be done between NA values which concern municipalities/communities from North Cyprus and those included in the Republic of Cyprus. Missing values referring to the number of cases in municipalities/communities in the Republic of Cyprus indicate zero number of cases, while those belong to municipalities/communities in North Cyprus specify missing data. Hence, cases with NA values from North Cyprus are replaced with 99999. Afterwards, the rest of NAs (belong to RoC) are replaced with zeros and 99999 are replaced back with NAs.

4. Cumulative covid-19 cases per municipality/community and date of the result are displayed in a new column.

5. Four new columns are created. The first two are holding the month and year of the result date, while the rest hold the population and cases rate by 1000 people.

6. The main investigation is based on weekly or 14-days data. Therefore, from 'ts_covid_data' two other dataframes were derived by selecting records with regards to a date filter applied. Starting date is defined as the 9th of March 2020. From this day on data date period is splitting per one and two weeks.

4.4 Exploratory spatial data analysis

The dataset includes 40 numeric variables for which a summary statistics, descriptive statistics were presented as standard deviation, mean, maximum and minimum, is shown in Table 4.7 below, Covid data summary statistics

	sd	min	max	mean
cases	2.519	0	127.000	0.340
female	1.350	0	68.000	0.176
male	1.239	0	60.000	0.164
female_0_4	0.071	0	5.000	0.004
male_0_4	0.077	0	5.000	0.005
female_5_9	0.097	0	7.000	0.006
male_5_9	0.095	0	6.000	0.006
female_10_14	0.114	0	6.000	0.008
male_10_14	0.116	0	5.000	0.009
female_15_19	0.134	0	8.000	0.011
male_15_19	0.145	0	8.000	0.012
female_20_24	0.157	0	10.000	0.014
male_20_24	0.152	0	7.000	0.014
female_25_29	0.183	0	10.000	0.017
male_25_29	0.181	0	9.000	0.017
female_30_34	0.175	0	8.000	0.016
male_30_34	0.179	0	10.000	0.016
female_35_39	0.171	0	10.000	0.016
male_35_39	0.154	0	8.000	0.014
female_40_44	0.156	0	9.000	0.014
male_40_44	0.138	0	7.000	0.012
female_45_49	0.152	0	7.000	0.014
male_45_49	0.131	0	7.000	0.011
female_50_54	0.152	0	7.000	0.013
male_50_54	0.130	0	6.000	0.011
female_55_59	0.145	0	8.000	0.012
male_55_59	0.124	0	6.000	0.010
female_60_64	0.123	0	6.000	0.009
male_60_64	0.123	0	6.000	0.010
female_65_69	0.095	0	4.000	0.006
male_65_69	0.091	0	4.000	0.006
female_70_74	0.082	0	5.000	0.005
male_70_74	0.079	0	4.000	0.005
female_75_79	0.064	0	5.000	0.003
male_75_79	0.060	0	3.000	0.003
female_80	0.127	0	19.000	0.005
male_80	0.073	0	6.000	0.004
month	3.447	1	12.000	6.343
year	0.418	2020	2021.000	2020.226
week	15.076	1	53.000	25.901
csum_cases	182.760	0	6903.000	30.197
total_population	8257.222	0	101000.000	2203.628
cases_rate	1.471	0	166.667	0.138

Figure 4.7: Summary statistics of numeric variables in 'ts_covid_data'.

The statistical summaries show that the maximum number of cases on a daily basis in a municipality/community is 127 and the minimum 0. The maximum number of cases in a municipality/community per age and gender occurs in female 80+ and is 19 cases. This may be a propagation of the disease in an elderly care home.

In Figure 4.8 the total cumulative number of Covid-19 cases per month shows the spatio-temporal propagation of the disease in the Republic of Cyprus. The monthly map visulization of the cumulative number of Covid-19 cases in Cyprus at municipality/community level from March 2020 until March 2021 categorizes the cumulative number of confirmed Covid-19 cases as follows: 0-20, 20-40, 40-60, 60-80, 80-100, 100-120, 120-140, 140-160, 160-180, 180-200, 200-500, 500-1000, 1000-1500, 1500-2000, 2000-2500, 2500-4000, 4000-5500, 5500-7000, 7000-8500. Throughout this period 5 main districts of Cyprus and their surrounding municipalities/communities are the most infected. Since the outbreak of the pandemic most cases occurred in the 5 city municipalities of Cyprus and their nearest municipalities/communities. Dimos Lemesou (6903 cases) accounted for 16.2% of the total cases, and Lemesos district (district_code = 5) had the most confirmed cases, accounting for 37.6% of the total cases in the country. Lemesos district is followed by Lefkosia district. However, regarding the population of each district in accordance with the number of confirmed cases Larnaca follows Lemesos with 2.50%. All the above results are shown in Tables 4.5 and 4.6. In terms of spatial distribution, it is observable that municipalities/communities with the higher number of Covid-19 cases are included in the surrounding areas of the 5 main city municipalities indicating that the disease was spreading spatially in the community. As time passes the cumulative number of cases had increased sharply reflecting the fact that the epidemic's geographical scope has expanded significantly. In terms of spatial distribution, it has the highest number of confirmed cases accumulated. Due to the closure of cities in March, the epidemic has been effectively brought under control.

Spatial choropleth plot
Initially, data is visualized using choropleth spatial plot. Below we visualise the monthly number of the cumulative confirmed Covid-19 cases at municipality/community level.

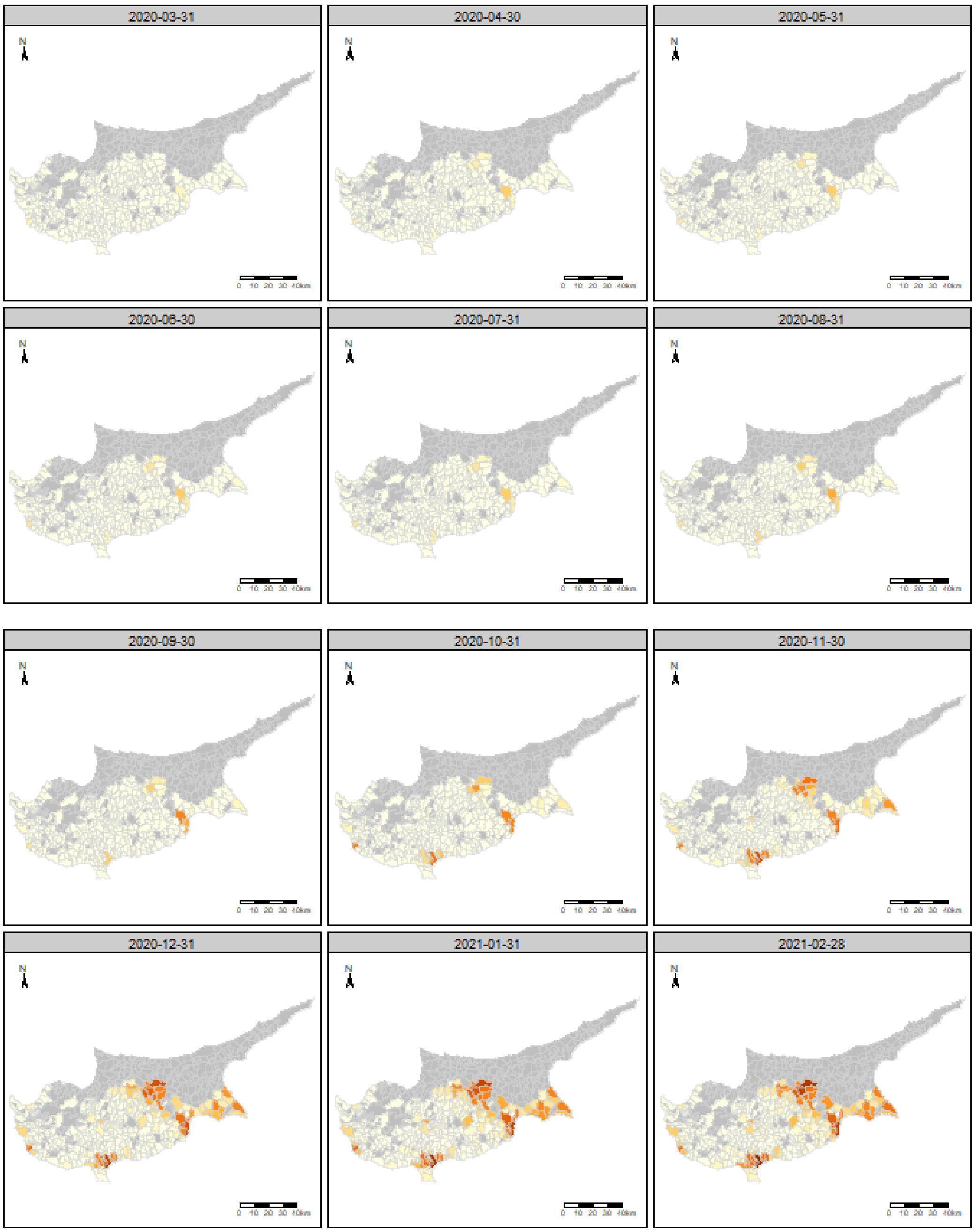


Figure 4.8: Cumulative cases per month

Hovmöller Plots

Another visualization used in exploratory analysis is a Hovmöller plot otherwise heatmap. Y-axis defines space while x-axis defines time. Now the variable of interest is the rate of weekly cases per 1,000 people for better visualizations. Figure 4.9 below displays a Hovmöller plot for municipalities/communities with populations over 3,000 residents. A conclusion drawn form this plot is that the critical period of Covid-19 spread has been during November 2020-April 2021 despite the implementation of a series of social distancing measures by the government.

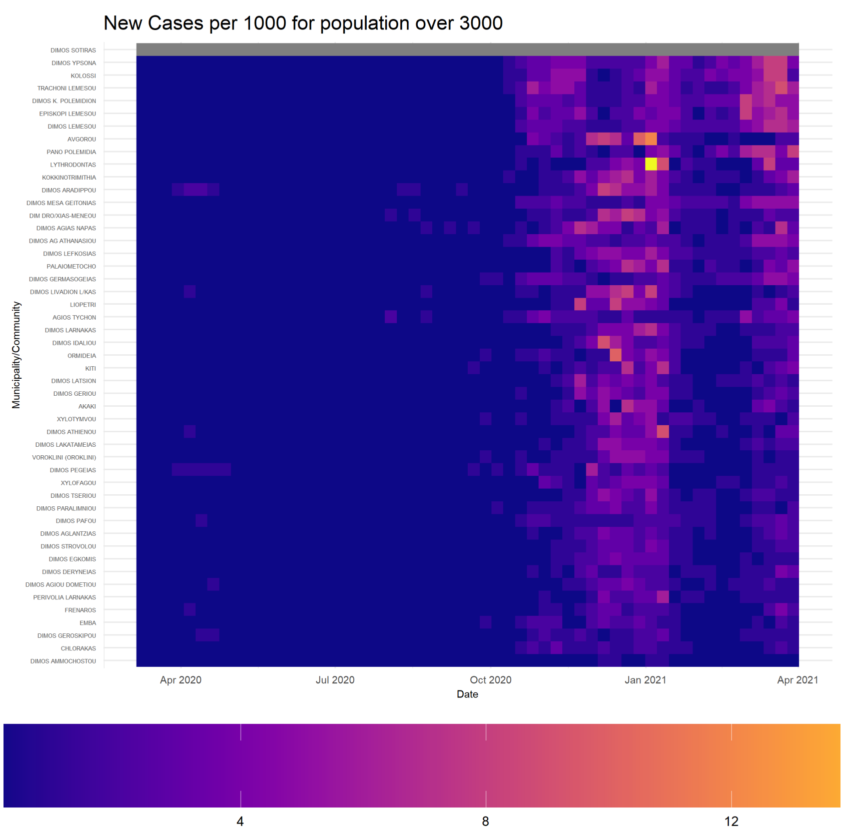


Figure 4.9: Hovmöller Plot for weekly cases in municipalities/communities with total population over 3,000 people

4.4.1 Means

Empirical Spatial Mean

The average number of cases has been taken by averaging over laboratory result date for each municipality/community.

In our case, we can compute the empirical spatial mean by averaging the daily rate of weekly cases for municipalities/communities with population over 3,000. It reveals that Dimos Ypsona, Kolossi and Trachoni Lemesou report an average daily infection rate of over 1.5 new cases per 1,000 people, whereas Dimos Geroskipou, Chlorakas and Dimos Ammochostou display an average of less than 0.5.

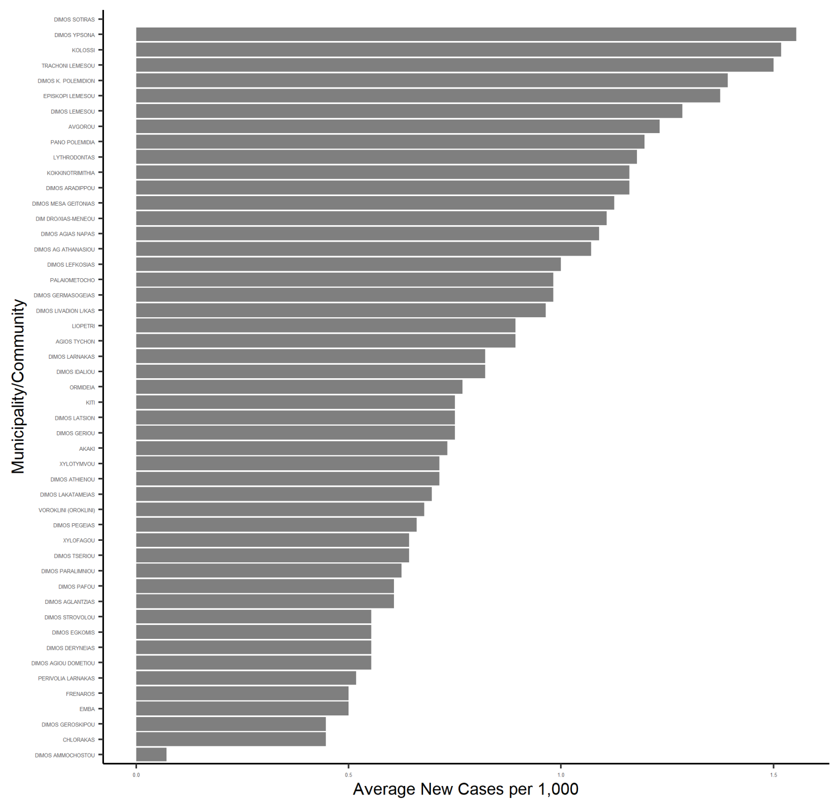
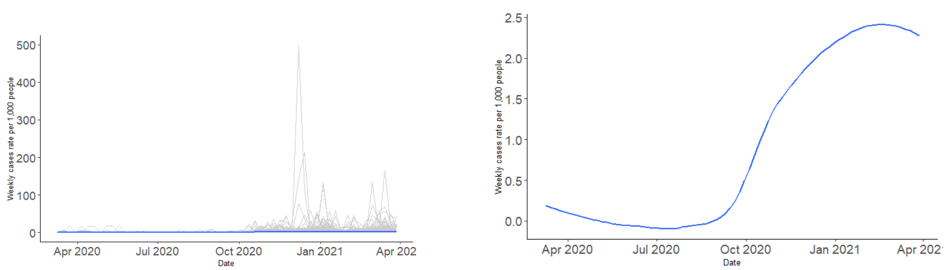


Figure 4.10: Average weekly cases in municipalities/communities with total population over 3,000 people.

Empirical Temporal Mean

The empirical temporal mean for a data set can be obtained by averaging across spatial locations for a time point. In our case, we can compute the empirical temporal mean by averaging the rate of new Covid-19 cases over municipalities/communities by week. The empirical temporal mean is plotted below revealing a peak of 500 number of new cases per 1,000 people by the end of December, and then decreasing. Blue line indicates the general trend of number of people infected.



(a) Fig 1 (b) Fig 2

Figure 4.11: Temporal plot for weekly cases in municipalities/communities with total population over 3,000 people.

Descriptive Tables

Number of confirmed cases per municipality/community

municipality_town_lst	total_cases	percentage_all_cases	population	percentage_population
DIMOS LEMESOU	6903	16.200%	101000	6.800%
DIMOS LEFKOSIAS	3096	7.300%	55014	5.600%
DIMOS LARNAKAS	2600	6.100%	51468	5.100%
DIMOS STROVOLOU	2394	5.600%	67904	3.500%
DIMOS K. POLEMIDION	1697	4.000%	22369	7.600%
DIMOS LAKATAMEIAS	1667	3.900%	38345	4.300%
DIMOS ARADIPPOU	1325	3.100%	19228	6.900%
DIMOS PAFOU	1177	2.800%	32892	3.600%
DIMOS YPSONA	1003	2.400%	11117	9.000%
DIMOS MESA GEITONIAS	951	2.200%	14477	6.600%

Table 4.5: The 10 municipalities/communities with the highest number of Covid-19 cases and their relative percentages regarding the total number of cases in the study period and their population.

Number of confirmed cases per district

DIST_CODE	total_cases_dist	percentage_all_cases	population	percentage_population
5	16505	37.60%	470660	3.50%
1	14177	32.30%	653960	2.20%
4	7237	16.50%	286384	2.50%
6	2508	5.70%	178570	1.40%
3	2108	4.80%	93258	2.30%
2	0	0.00%	NA	NA

Table 4.6: The number of Covid-19 cases per district and their relative percentages regarding the total number of cases in the study period and their population.

From time series plots in Figure 4.9 it is observable that the largest number of cases occurs in December of 2020 followed by March 2021. Also, the above plots show that from October 2020 there is an increase in the number of cases until December of 2020. By then a reduction of cases is observed until February 2021. A sudden change in the number of cases occurs within February and March of 2021, where the number of cases in the country increases for about 6500 cases. These ups and downs in the number of Covid-19 cases may be related with the measures imposed each period and how rigorous they were.

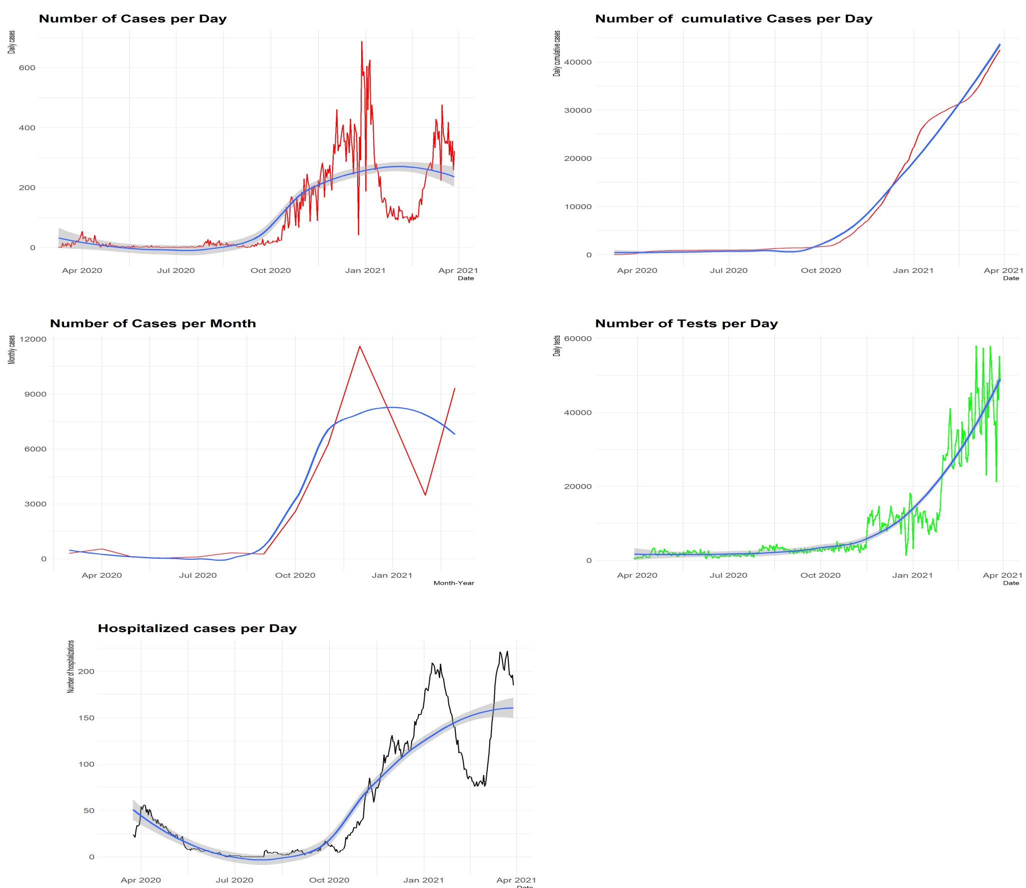


Figure 4.12: Time series plots regarding the number of cases, tests and hospitalizations during the study period

Until mid of November laboratory tests were done by PCR method. At the start of the outbreak in Cyprus on 9th of March 2020 laboratory tests were conducted in the Cyprus Institute of Neurology and Genetics. Therefore, there was a limitation in the number of tests per day as only one institution was legal to perform the coronavirus test. Thereafter, all Covid-19 incidents were confirmed directly through private regional reference laboratories. From mid November 2020 until today in each district there are several sampling points which offer free Rapid Covid-19 test for all residents.

The healthcare system of Cyprus was facing a difficult challenge due to its serious limitations. Its capacities were not able to meet the needs of infected Covid-19 patients. Until 1st of April 2020 the total number of Intensive Care Units in public Cypriot hospitals was in total 90-100 beds with the 30 of them used for Covid-19 cases. As a general trend the number of hospitalizations increase as time passes.

In time series plots blue line indicate the empirical smoothing fit to the data. Data for time series plots has retrieved from https://www.data.gov.cy.

On November 18th a local lockdown was imposed in Lemesos and Pafos districts as cases were high. Figure 4.13 and Figure 4.14 show the weekly cases and weekly case rate per population respectively. It seems that local lockdown was not necessary in Pafos.

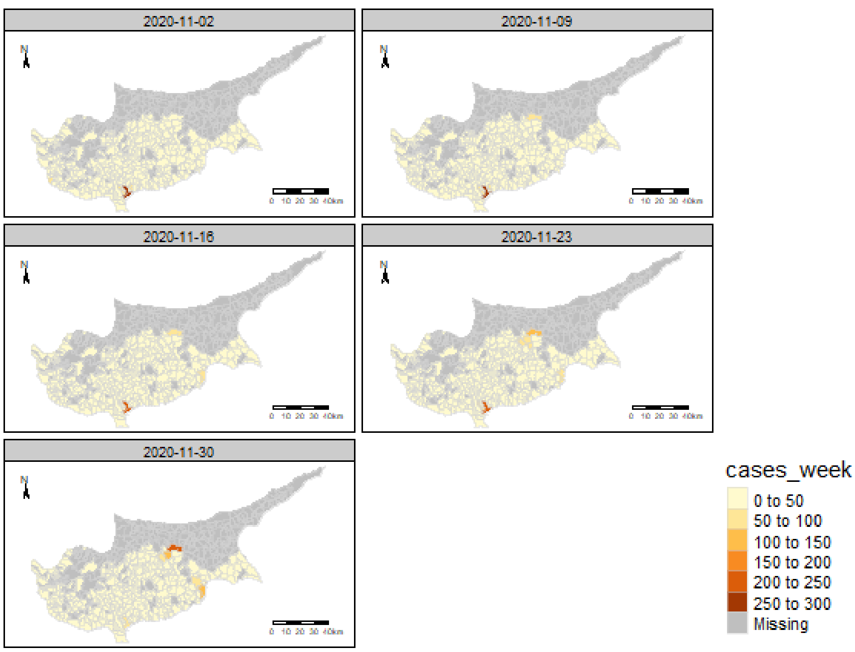


Figure 4.13: The number of weekly Covid-19 cases during November.

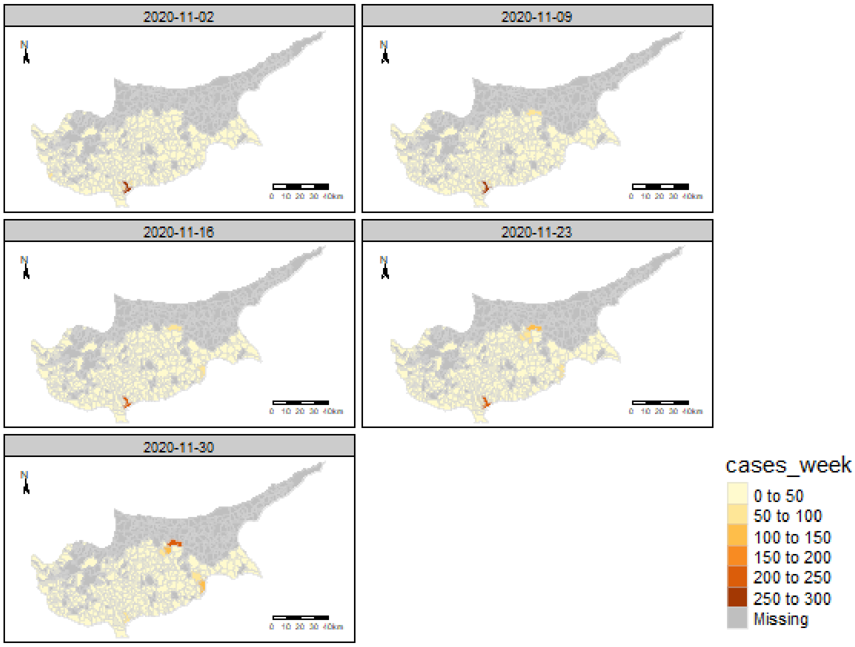


Figure 4.14: The number of weekly Covid-19 cases during November.

In accordance with the numeric data, we retrieved data on response measures applied in Cyprus. Data has been accessed from ECDC's official website via the link **here**, however the main source of it is the official covid-19 website of Cyprus which shares the daily measures and announcements of the pandemic in the country. Measures were imposing rapidly with multiple daily modifications. ECDC holds the main measures without any details and their corresponding time periods. Details about any measures imposed can be read from the official public website of Covid-19 in Cyprus. For the current study a combination of the two sources will be considered and measures discussed during the analysis are presented in detail on A.

4.5 Spatial correlation characteristics

4.5.1 Spatial Autocorrelation

To check for any spatial autocorrelation between the weekly number of covid cases within municipalities/communities the global and local Moran's I spatial autocorrelations were calculated. In order to evaluate the Moran's I we need information about the neighbours of each municipality/community. A neighbour can be defined in several ways. For instance, two regions are neighbouring if they share boundaries, and we refer to them as neighbours by contiguity. There are two criteria of contiguity, the rook criteria which refers to regions which are sharing common boundaries and the queen criteria for regions with common boundaries and corners. Referring to contiguity neighbours another aspect to be established is the order of contiguity. First order contiguity refers to regions immediately contiguous. Second order means that we consider neighbours only those areas that are immediately contiguous to our first order neighbours and you could go on and on. However, we can talk about neighbours by distance, which do not have any common boundaries but are closed to each other. The two kind of neighbours as well as their combination can be shown in Figure 4.15.

The definition of adjacency is necessary in the evaluation of the spatial autocorrelation and the most commonly known adjacent regions are the ones who have common boundaries. Thus, we tend to use contiguity neighbours and queen concept is used in the current study. Therefore, using the poly2nb() from the spdep package we can build a neighbours list based on municipalities/communities with contiguous boundaries. In this function there is an argument called queen which defines the contiguity neighbouring criteria. The default is queen = TRUE and will return a list of first order neighbours using the queen criteria.

A summary of queen and rook criteria neighbours is given using the summary function.

Queen neighbours:

```
Neighbour list object:
Number of regions: 615
Number of nonzero links: 3434
Percentage nonzero weights: 0.9079252
Average number of links: 5.58374
```

Rook neighbours:

```
Neighbour list object:
Number of regions: 615
Number of nonzero links: 3380
Percentage nonzero weights: 0.893648
Average number of links: 5.495935
```

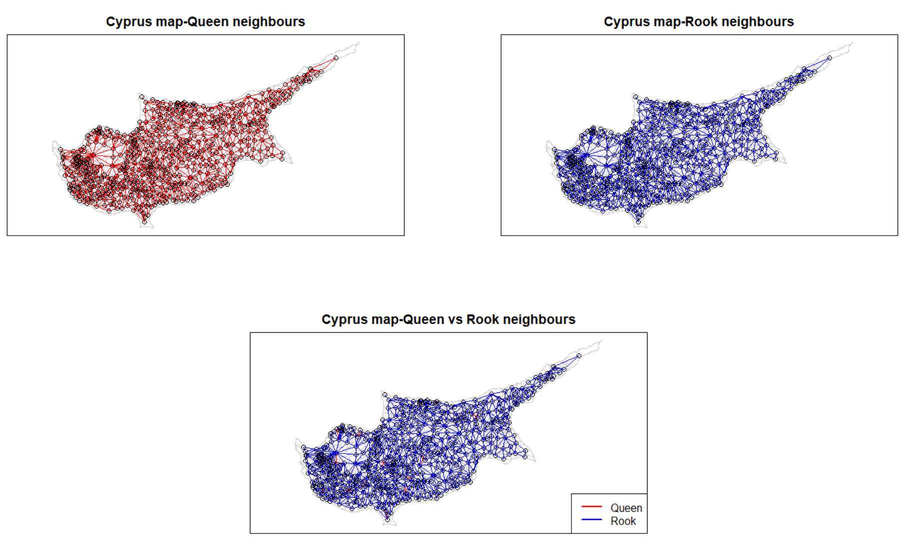


Figure 4.15: Queen and Rook neighbours mapping visualization.

Hence, we can say that each polygon (municipality/community) has on average of 5.6 queen neighbours and 5.5 rook neighbours and that all areas have at least one neighbour. Therefore, few diagonal neighbours exist. The link number extracts the number of links, neighbours, per region. For this study queen neighbours are going to be used. From this step and onwards queen neighbours would be referred for simplicity as neighbours. Then using the function nb2listw() we supplement a neighbours list with spatial weights for the chosen coding scheme. For a total number of 615 regions there are 3434 links.

4.5.2 Global spatial correlation characteristics

The global spatial correlation score can take values between -1 and 1 where 1 determines perfect positive spatial autocorrelation (so our data is clustered), 0 identifies the data is randomly distributed and -1 represents negative spatial autocorrelation (so dissimilar values are next to each other). Moran's I statistics obey normal distribution and test significance based on z-score is conducted. A hypothesis test is examined to choose between the null hypothesis where our data is not cluster and the alternative in which data is clustered, We can consider the p-value as a measure of the statistical significance of the model and since p-value < 0.05 we do reject the null hypothesis in some 5% significance level; therefore, our data is spatially clustered or autocorrelated.

As foretold the total number of weekly confirmed Covid-19 cases in each municipality/-community was taken as a variable. For a spatial autocorrelation investigation between the weekly Covid-19 cases in a municipality/community level in Cyprus, Moran's I statistic and Z-score on each date within the weekly study period were evaluated.

The evaluation was achieved using a function which has been developed in R. Firstly, a neighbours list was created using poly2nb from "spdep" package in R. Hence, we obtained the neighbours of each municipality/community as a list which contains each municipality/community and the indices of their neighbours. Then using function nb2mat() a weights matrix for the neighbours list with spatial weights with "B" , for binary weights taking values zero or one (only one is recorded), was generated.

Spatial weights matrix is often called as spatial neighbourhood matrix and is denoted by W. Its elements are given by w_{ij} (weights) where i and j specify areas within the examined region that are connected. The larger the value of w_{ij} the closer the two areas are, otherwise, areas are farther away. The simplest neighbourhood definition matrix is provided in a binary structure as follows, for areas with shared boundaries $w_{ij} = 1$, otherwise $w_{ij} = 0$. Binary structure is used in the current study. Therefore, the weight matrix is symmetric with diagonal elements defined by default with zeros ($w_{ii} = 0$ for weights of an area).

Then we convert the square spatial weights matrix to a weights list object with "W" for row-standardised weights. The list is converted back to a matrix using listw2mat() function. This process leads to the calculation of Moran's I test statistic with moran.test() function from 'spdep' package in R using the spatial weights matrix in weights list form and the variable of interest as input values. Using this function global Moran's I test will be calculated on a weekly basis for the study's period.

In the current study, we take a confidence level corresponding to $|z| > 1.96$ or equivalently to $p < 0.05$ which allows us to state whether clustering or dispersion of our spatial data exists indicated by I. In such case the p-value is statistically significant and the distribution of the variable of interest normal. Therefore, the value of Moran's I statistic determine the pattern of the determine the pattern of the data.

Figure 4.16 shows the Moran's I index and Z statistical scores of the weekly confirmed Covid-19 cases in the Republic of Cyprus at municipality/community level from 9th of March

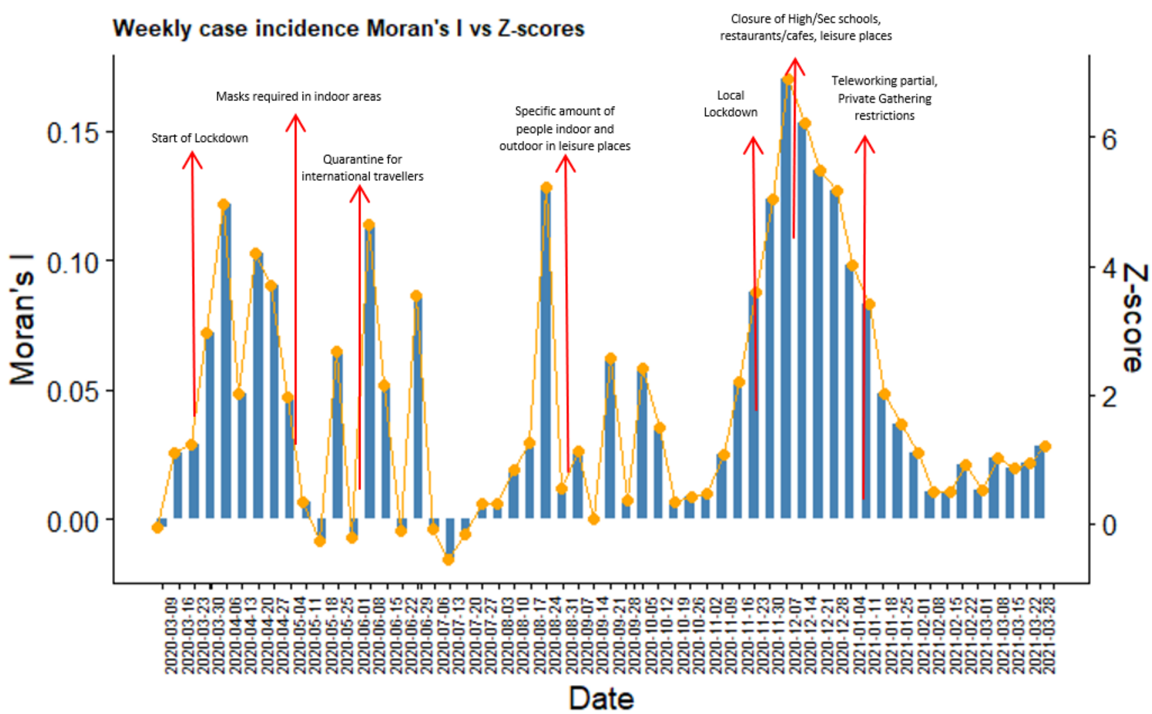


Figure 4.16: Global Moran's I index of the weekly number of cases with Z-scores add to caption policy decisions.

2020 to 29th of March 2021. The global spatial autocorrelation seems to have some ups and downs which can be related with the social distancing measures which were in force at that time. Until the 23/03/2020, the global spatial autocorrelation ($p > 0.05$, $Z < 1.96$) was not statistically significant, while on 30/03/2020 there was a significant global spatial autocorrelation ($p < 0.05$, $Z > 1.96$). From this day on there is some reduction in the value of global spatial correlation until the 25/05/2020. During June and July 2020 some ups and downs occur and by 03/08/2020 the weekly number of cases shows a sequential increase regarding the significance of global spatial autocorrelation until the end of August 2020. In September 2020 there was no important spatial dependency in the number of confirmed cases, however from October 2020 until mid-December number of Covid-19 cases reached the peak value of global Moran's I statistic, showing a significant global spatial autocorrelation and indicating that the weekly confirmed cases at municipality/community level showed a very significant spatial dependence and that there is strong clustering of infections in municipalities/communities sharing common borders and among nearer municipalities/communities.. Those changes in trend characteristics in figure of Moran's I spatial autocorrelation reveal that many turning points occur regarding the dissemination of the pandemic in the country under a spatial propagation. High values of spatial autocorrelation, specifically observed on October-December 2020 indicate the occurrence of hot-spots/clusters where there is a large number of infections on a weekly basis. Ups and downs however, mean that, although the degree of clustering is lower, the global spatial correlation is still dominated by clustering character- istics and tends to develop in a decentralized manner. Only few weeks present no clusters of infections in nearby municipalities/communities. The maximum value of Moran's I reached 0.17 on week 30/11/2020-06/12/2020 indicating a high level of geographic clustering of the disease spread since November 2020. The minimum value of Moran's I reached 0 and occurred on week

20/04/2020-26/04/20 where the first rigorous lockdown imposed. Although as most consecutive low global Moran's I statistics occurred by the of June 2020 until July 2020 we expect few or no clusters within June and July 2020.

4.5.3 LISA

Lisa statistics aim to identify potential clusters of Covid-19 cases within Cyprus. Also, they aim to assess the influence of individual municipalities/communities on the magnitude of the global statistic and to identify "outliers". Local variation of each municipality/community is investigated. Moreover, the Moran's I statistic is not robust to outlier or both skewed data, thus a scaling is applied to weekly Covid-19 cases which is the variable of interest.

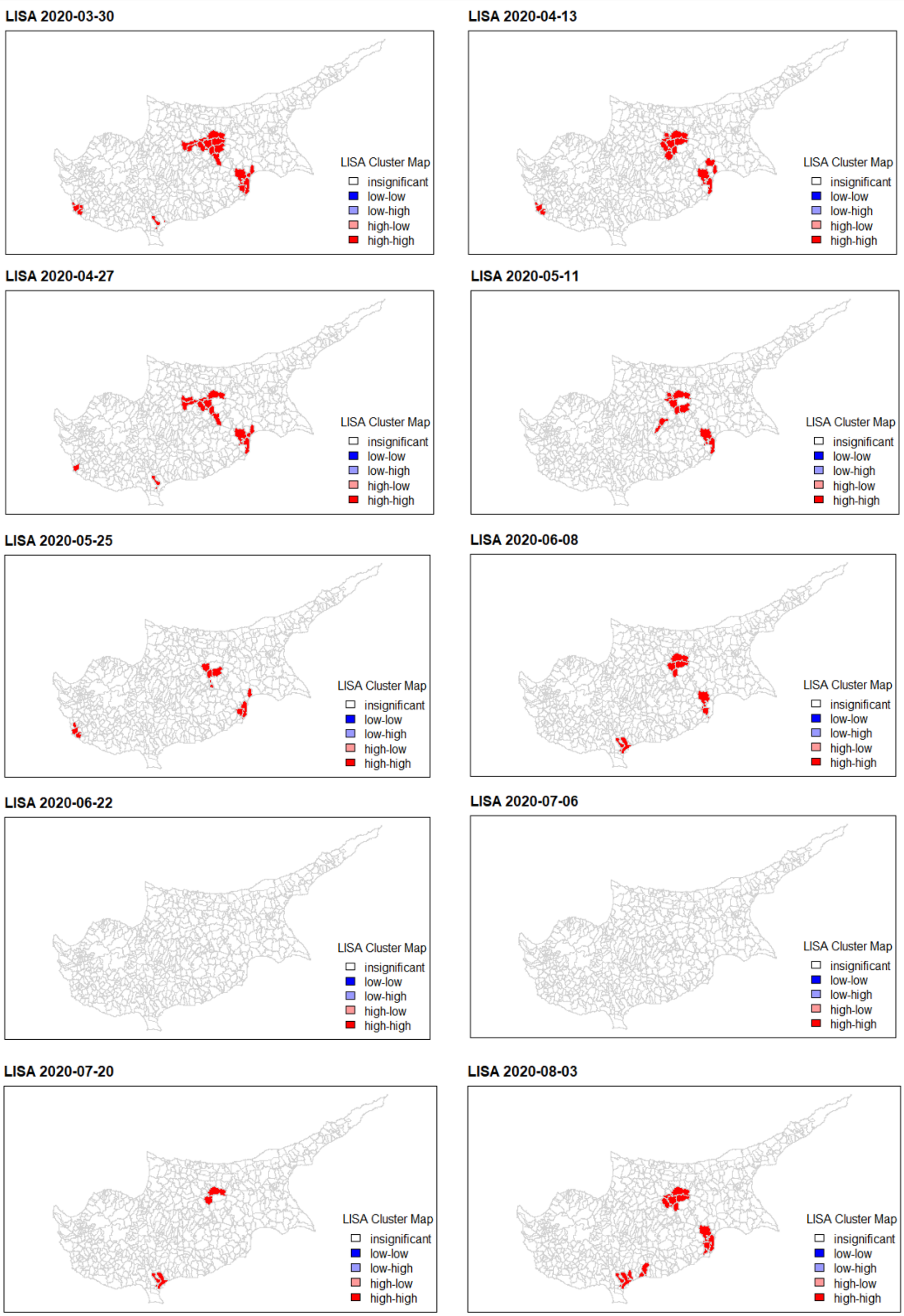
The interpretation of the local Moran's I indicator is essentially the same as that of the global indicator. Thus, positive values of local I indicate a spatial concentration of similar values (low or high) while negative values of local I indicate a spatial concentration of dissimilar values (Anselin, 1995). Unlike global Moran's I, local Moran's can get values greater than +1 and less than -1. A LISA map is constructed in order to get a visual representation of potential Covid-19 spots per week. Subtracting the mean and dividing by the standard deviation weekly cases were scaled. Local Moran's I tool returns five values: the Moran's I Index, Expected Index, Variance, z-score, and p-value. The purpose of this evaluation is to assess whether the pattern of the spatial data is clustered, dispersed, or random. We can observe whether the z-score and p-value of each municipality/community is statistically significant. In such scenario a positive Moran's I index value indicates tendency toward clustering, while a negative Moran's I index value indicates tendency toward dispersion. An interpretation of LISA results is as follows. The LISA value for each municipality/community is determined from its individual contribution to the global Moran's I evaluation. As for any statistical test of significance there is a statistical process that leading to the result. Therefore, a municipality/community is considered statistically significant if its actual value is close enough to the corresponding new value which is recalculating after data is randomly reassigning in all regions in the study.

Local Indicator of spatial autocorrelation are conducted on the data of a weekly confirmed Covid-19 cases in Cyprus at municipality/community level from 9th of March 2020 to 28th of March 2021. LISA Cluster Map is shown in Figure 4.17. The map can display five spatial autocorrelation categories. High-high is represented in red and represents the high-value aggregation class. High-Low is shown in pink, representing high-value elements, and is surrounded by Low-value elements. Low- high is shown in light blue, representing low-value elements, and surrounded by high-value elements. Low-low, shown in dark blue, represents the low-value cluster classes. According to this Figure the 5 main city/municipalities of Cyprus (Dimos Lefkosias, Dimos Lemesou, Dimos Larnakas, Dimos Ammochostou, Dimos Pafou) fall into the "high-high" clustering region as well as their neighbouring municipalities/communities. There are no regions belong to the rest of clustering categories. This suggests that except from the 5 main municipalities and their surrounding municipalities/communities there is weak, otherwise absent clustering trend.

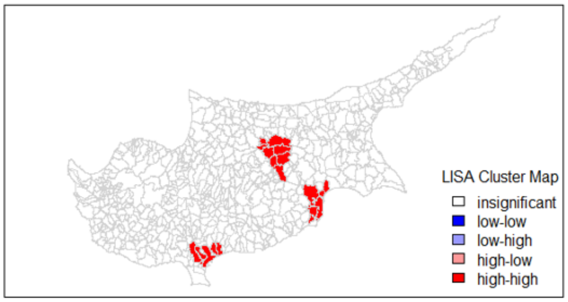
Regarding the interpretation of local Moran results municipalities/communities with a positive value of I means that the existing region has similar number of Covid-19 cases with its neighbours, thus it can be considered as a part of a cluster. A negative value of I indicates that region's neighbours have dissimilar number of Covid-19 cases, thus the region may be a potential outlier. Local Moran statistic is used to discover hot spots of the pandemic in Cyprus according to the variable of interest (weekly cases at municipality/community level).

The identification of clusters will be based mainly in cluster/significance maps. As it has been discussed in Chapter 3 global autocorrelation statistic is used in the hypothesis testing where the null hypothesis indicates spatial randomness and the alternative clustering. However, clustering specifies only if a spatial pattern exists within georeferenced data or not. For a deeper investigation of the location of the clusters local Moran's I is used instead. Local spatial statistic has two main characteristics. First, it provides a statistic for each location with an assessment of significance. Second, it establishes a proportional relationship between the sum of the local statistics and a corresponding global statistic. Using LISA statistics and technique cluster maps are created from 30/03/2020 and for every two weeks. A cluster map is defined as the map that shows neighbouring locations which have similar values (in our case similar number of weekly Covid-19 cases). These maps augment the significant municipalities/communities with the largest number of covid-19 cases and display the type of spatial association.

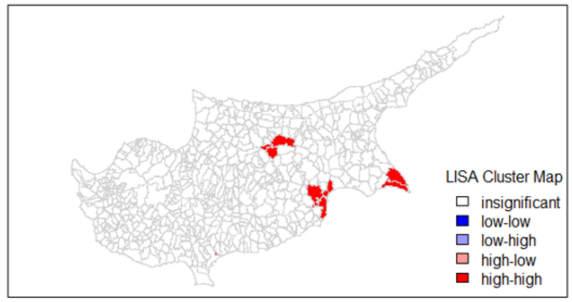
A LISA mapping below Figure 4.17 displays visually the elements described above. The first effects and results of quarantine measures on the Covid-19 diffusion can be shown after at least 14 days, and for thus starting point of LISA maps plotting is the 30th of March 2020.


LISA 2020-03-30
LISA Cluster Map
insignificant
low-low
low-high
high-low
high-high
LISA 2020-04-13
LISA Cluster Map
insignificant
low-low
low-high
high-low
high-high
LISA 2020-04-27
LISA Cluster Map
insignificant
low-low
low-high
high-low
high-high
LISA 2020-05-11
LISA Cluster Map
insignificant
low-low
low-high
high-low
high-high
LISA 2020-05-25
LISA Cluster Map
insignificant
low-low
low-high
high-low
high-high
LISA 2020-06-08
LISA Cluster Map
insignificant
low-low
low-high
high-low
high-high
LISA 2020-06-22
LISA Cluster Map
insignificant
low-low
low-high
high-low
high-high
LISA 2020-07-06
LISA Cluster Map
insignificant
low-low
low-high
high-low
high-high
LISA 2020-07-20
LISA Cluster Map
insignificant
low-low
low-high
high-low
high-high
LISA 2020-08-03
LISA Cluster Map
insignificant
low-low
low-high
high-low
high-high

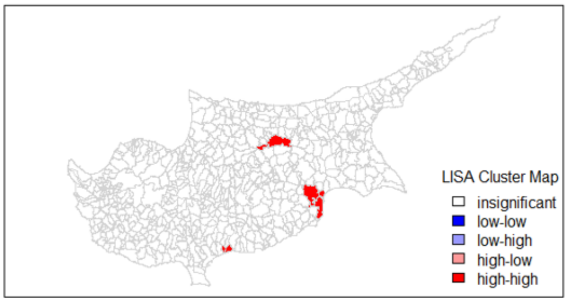
LISA 2020-08-17



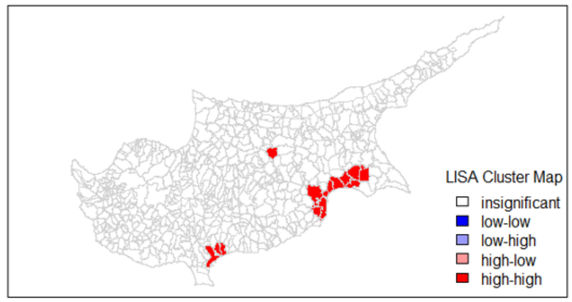
LISA 2020-08-31



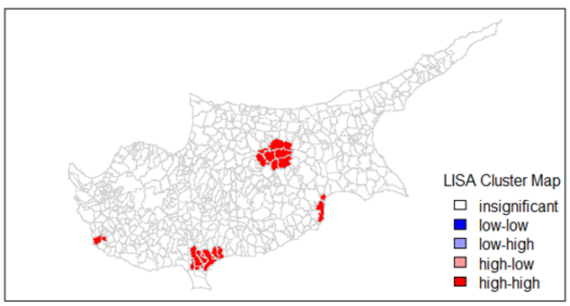
LISA 2020-09-14



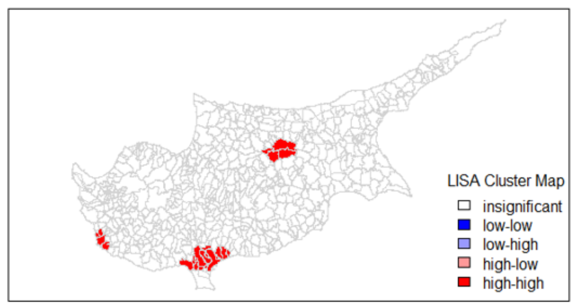
LISA 2020-09-28



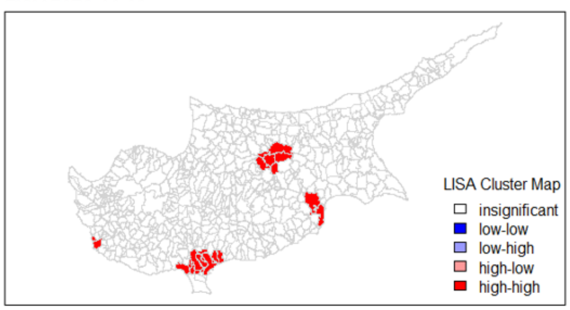
LISA 2020-10-12



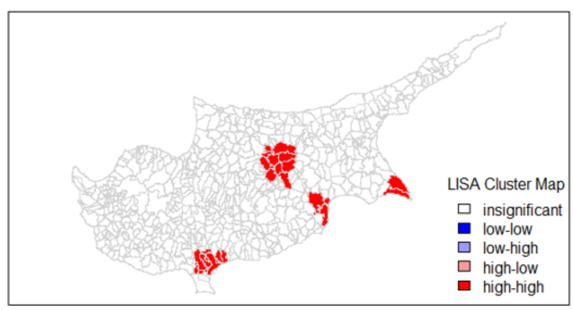
LISA 2020-10-26



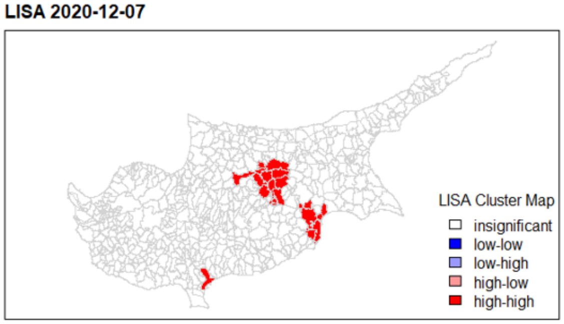
LISA 2020-11-09



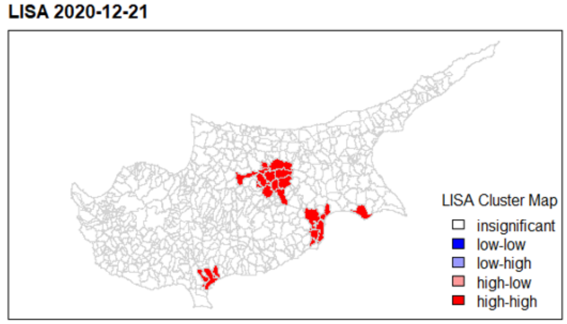
LISA 2020-11-23



LISA 2020-12-07



LISA 2020-12-21



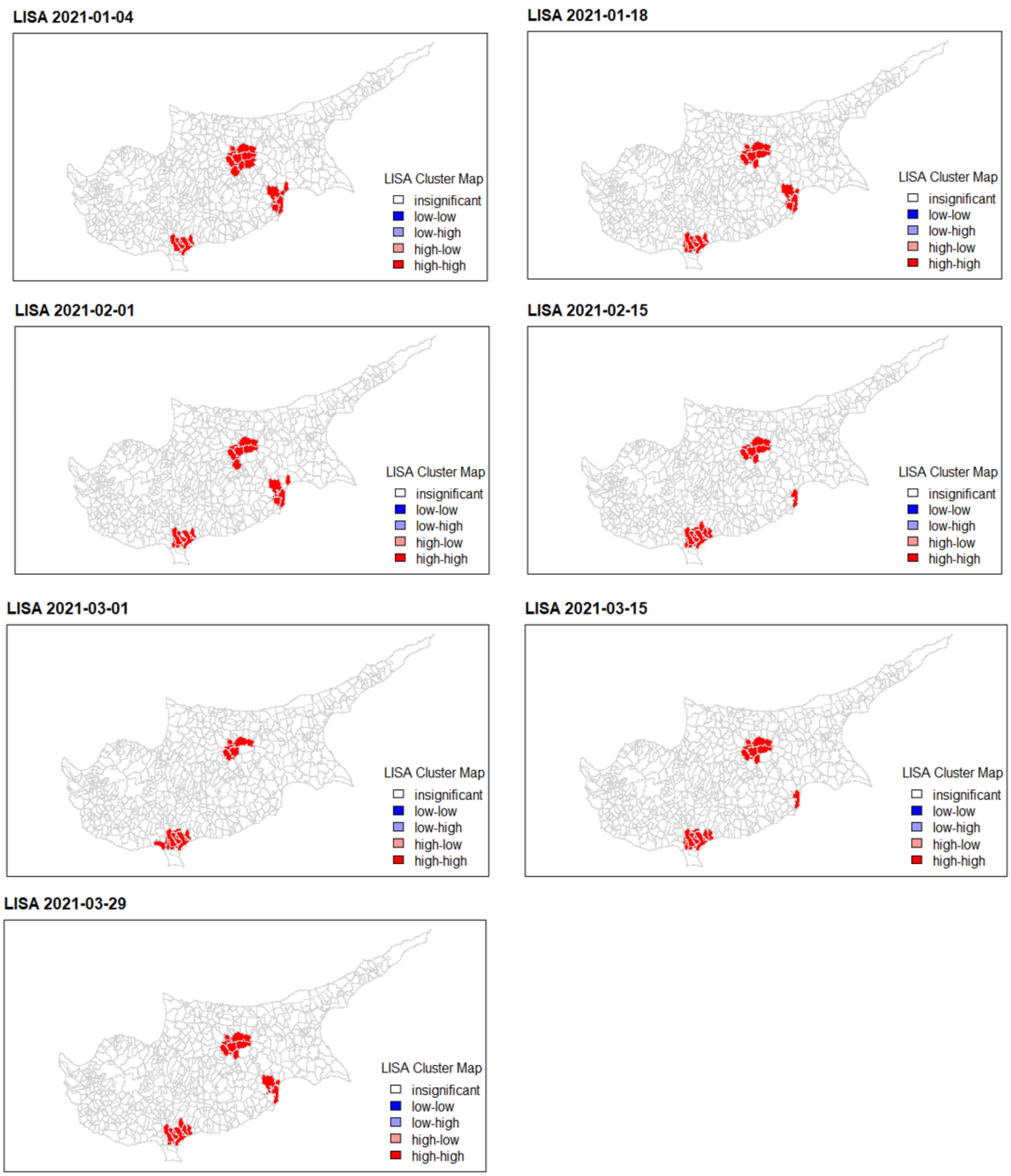


Figure 4.17: Spatial Cluster map associated with COVID-19 cases by municipality/community, Significant Clusters obtained through LISA.

4.6 Standardized incidence ratio Method

SIR Data preparation

For the purposes of this study SIR was evaluated only for the average cases of each municipality/community in the study period. Before analysing the data to evaluate SIRs for each municipality/community a dataframe should be created with columns containing the names of municipalities/communities, the observed and expected number of Covid-19 cases, and the SIRs. For each municipality/community we calculate the SIRs of average cases. Using the demographic data, we create a new column which contains the population in accordance with the age and gender in each municipality/community. Demographic data contains the population of Cyprus at a county level, stratified on gender (female and male) and age (0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79 and 80+). The names of the columns in demographic data which defines the population per municipality/community in each stratum are of the form e.g., female_25_29. In each age group '-' is replaced with '_' and gender choices are become lowercase. Then, a new column is created which is the age group and gender of each row striped. We seek for a data frame which has all the information for each one of the 34 stratum for each municipality/community. The modified Covid data is related with the sir data using left_join.

We obtain the number of cases for all the strata together in each municipality/community, Y, by aggregating the rows of the sir data by municipality/community and adding up the number of cases. We can do this using the functions group_by() and summarize() of the dplyr package. The SIRs are stratified on age group and gender. Therefore, the number of stratum counts to 34 which are,
'female_0_4', 'female_5_9', 'female_10_14', 'female_15_19', 'female_20_24', 'female_25_29', 'female_30_34', 'female_35_39', 'female_40_44', 'female_45_49', 'female_50_54', 'female_55_59', 'female_60_64', 'female_65_69', 'female0_74', 'female_75_79', 'female_80', 'male_0_4', 'male_5_9', 'male_10_14', 'male_15_19', 'male_20_24', 'male_25_29', 'male_30_34', 'male_35_39', 'male_40_44', 'male_45_49', 'male_50_54', 'male_55_59', 'male_60_64', 'male_65_69', 'male0_74', 'male_75_79', 'male_80'.

A new dataframe is derived from the data which includes all the information for all stratum combinations on each date and municipality/community. Hence, the dataframe called 'sir_df_used' consists of 8050350 as there are in total 615 municipalities/communities, 385 dates and 34 stratum, $615 \times 385 \times 34 = 8050350$. The number of Covid-19 cases for all the strata in each municipality/community is obtained by aggregating the rows by municipality/-community and adding up the number of cases per age and gender. The daily number of cases per municipality/community is compared to a predicted number of cases evaluated. Using indirect standardization we calculate the expected number of Covid-19 cases depending on the population per stratum and municipality/community. This calculation can be achieved using the expected() function from the SpatialEpi package in R. This function has three arguments, namely,

- population: vector of population counts for each strata in each area,
- cases: vector with the number of cases for each strata in each area,
- n.strata: number of strata.

Vectors population and cases need to be sorted by area first and then, within each area, the counts for all strata need to be listed in the same order. All strata need to be included in the vectors, including strata with 0 cases. Here, in order to obtain the expected counts, we first sort the data using the order() function where we specify the order as municipality/community, laboratory result date and finally age-gender.

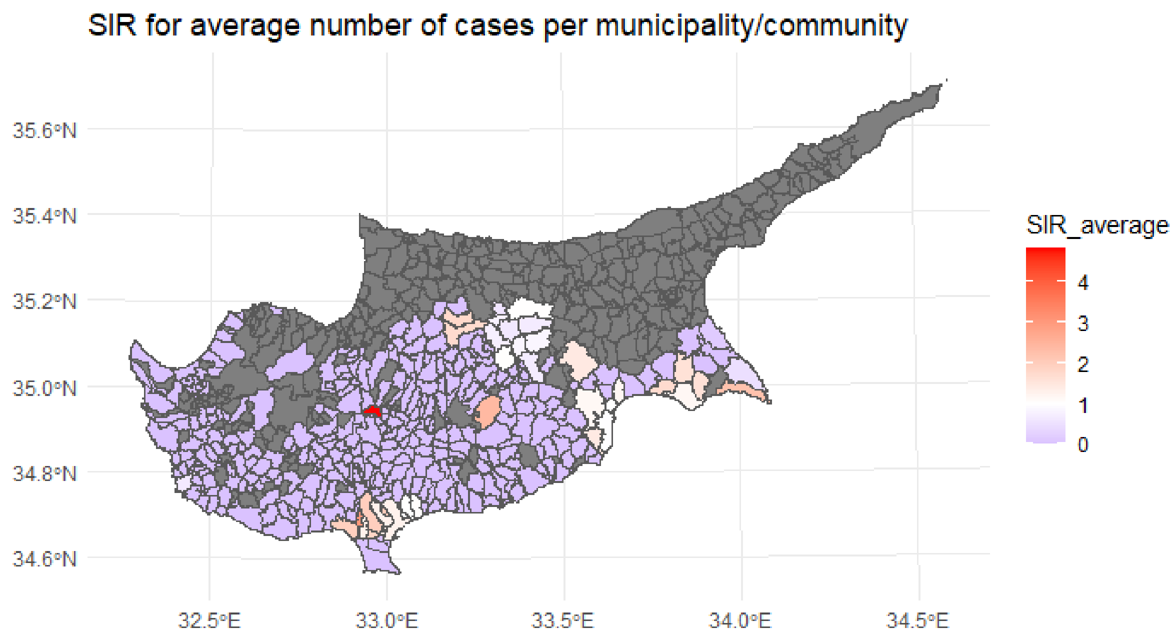


Figure 4.18: SIR for the average number of cases in each municipality/community.

The result of SIR is shown in Figure 4.18 below,

As discussed on Chapter 3, SIR technique involves some limitations on its processing method for investigating SIR in Cyprus as municipalities/communities are of small population. Hence, Bayesian models were implemented and evaluated. During the implementation of spatial modelling the variable of interest is the average number of Covid-19 cases in each municipality/community. There are in total 615 municipalities/communities $i, i = 1, ..., 615$. In the Poisson model as introduced in section (number) Y_i defines the observed average cases in the i^{th} municipality/community, E_i represents the corresponding expected average cases, a defines the overall risk in the study's region which is the country of Cyprus, u_i represents the random effect of the i^{th} municipality/community for modelling spatial dependence among relative risks and v_i is the uncorrelated noise of the model. In all the fitted models external covariates do not exist in order to focus only on the spatial variability of the disease in the Republic of Cyprus. A quick reminder that the linear predictor is specified on the logarithmic scale. The models estimation require the list of neighbours of each municipality/community deriving from the shapefile with the associated information.

4.6.1 Prerequisites-Process

For a spatial modelling of Covid-19 in Cyprus models are implemented in R and the relevant results are extracted. Initially a PC potential prior of the marginal precision τ_b is defined by using $P(\frac{1}{\sqrt{\tau_b}}) > U) = a$ (Moraga, 2019). Also, the neighbourhood structure needs to be defined as it is an input argument in the formula of the model. The neighbourhood object takes the form of a matrix and is calculated through poly2nb() function and then converted into a file with the representation of the matrix as required by R-INLA. The next step is reading the file using the inls.read.graph() function of R-INLA and assign it in the object g. The input data is drawn from the data evaluted for SIR. After the above process the BYM2 model is fitted using the inla() function which has as arguments the formula, the family (in that case 'poisson'), the data. We also set control.predictor equal to list(compute = TRUE) to compute the posteriors of the predictions and control.compute = list(dic = TRUE, waic = TRUE, cpo = TRUE) to evaluate the values of specific selection criteria.

4.7 Implementation of spatial models

4.7.1 BYM model

Two BYM models are considered. A BYM model with and without noise. Before fitting the BYM models using INLA package in R, the formula of the model is essential to be constructed. The formula includes the response variable in the left-hand side which is the number of weekly (or average) Covid-19 cases and the fixed and random effects in the right-hand side. An intercept is included by default in the formula. The random effects of each for each municipality/community are defined through some parameters in the f() function which are the name of the index variable, the model and other options. For the selected BYM model the formula contains a spatially structured component with index variable with name idu which equals to $1,...615$, a model 'besag' with a CAR distribution and a neighbourhood given by the graph g. The option scale.model = TRUE is used to make the precision parameter of models with different CAR priors comparable (Freni-Sterrantino, Ventrucci and Rue, 2018). In order to fit the noise of the model a formula contains the index variable named idv equals to $1,...,615$ and model 'iid'. This model specifies the independent and identically normal distribution with zero mean random effect. The vector variables idu and idv are the same as they define the indices of municipalities/communities. However, they are specified as two different arguments as the R-INLA does not allow the inclusion of the same index variables.

The BYM model is a model for u and is a very famous approach when dealing with spatial mapping disease. However, this model is intrinsic and penalises local deviation from its null space, which is a constant level in the case of one connected component. Thus, a prevention of a possible confounding with the intercept is restricted through the constraint $1^T u = 0$. Hence a more reliable model a prior demanded to combine the intercept, the fixed effects β and the priors for the two precision parameters τ_u and τ_v.

Provided the issues of BYM model discussed on Chapter 3 the BYM2 model accompanied with the Penalised Complexity priors is introduced and is able to solve the aforementioned issues.

4.7.2 Penalized Complexity Priors

Simpson et al. (Riebler et al., 2016) proposed first the creation of appropriate priors which can be used in latent spatial analysis as they are defined for different components with latent effects. Such priors penalize departure from a base model and are defined using probability statements about the parameter.

4.7.3 BYM2 model (different parametrization)

In the case of BYM2 model two models are fitted. The two model differ on their prior distribution. The formula for the BYM2 model in R is defined exactly with the same way as for BYM. In the f() function the first argument given is the vector of the index variable for the random effects which in that cases is the vector of $1,\ldots,615$ the indices of the 615 municipalities/communities. The second argument is the g graph and defines the neighbourhood structure. The intercept is included in the formula by default. The vectors idu, idv and id can be obtained by aggregating the number of rows of the data given from the *sir_ratio* function which is a simple features object with a geometry variable for each municipality/community.

In total 4 different models are fitted in the data and compared within each other to choose the best model. The 4 models considered are the BYM model of the random effects without the noise, the BYM model of the random effects considering the noise as well, the BYM2 model and the BYM2 model with the parametrization of the prior distribution.

4.7.4 Modelling selection criteria

From all models the estimations/fitted values of the number of Covid-19 cases per municipality/community is retrieved and plotted later through an sf data. The estimates are the posterior means. An initial step of the selecting process is to display a plot of the fitted average cases (posterior means) for each model and make a comparison with the observed average cases. Figure 4.19 achieves the aforementioned and is shown below. Observing the plots we can identify that BYM2 model with a defined prior is the one suited best followed by the BYM2 model. It seems like besag model with or without noise fail to extract such good results. Both BYM2 models provide identical posterior estimates for the average Covid-19 cases of each municipality/community. The fitted plots agree very closely with the corresponding observed plots both for the BYM2 and BYM2-prior.

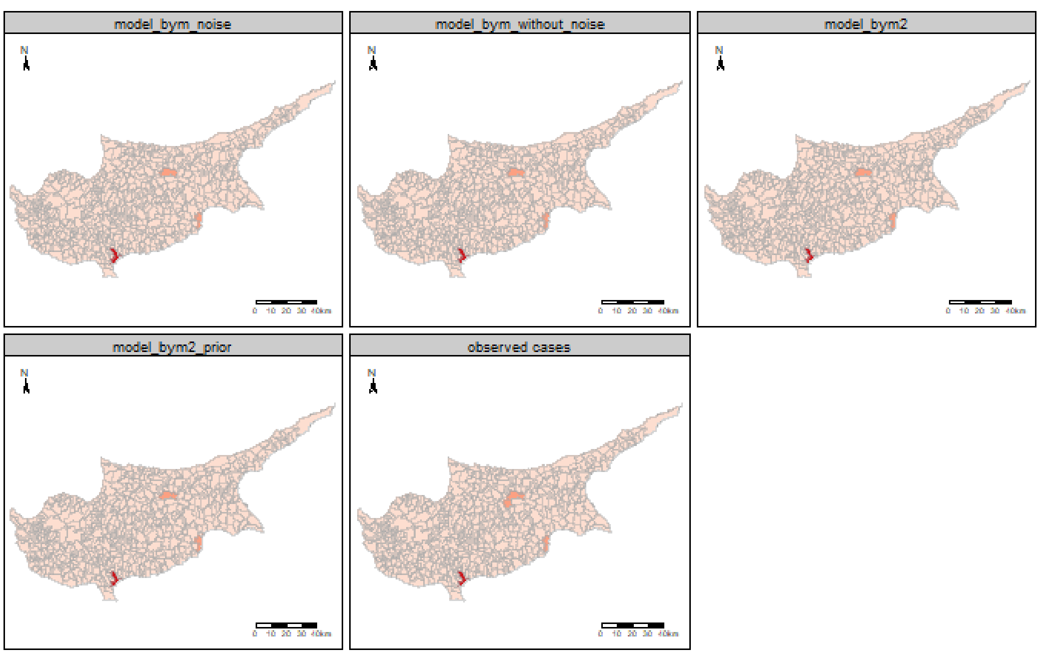


Figure 4.19: Fitted values of models accompanied with observed average cases in each municipality/community.

Modelling selection criteria is applied in the four models and is evaluated using DIC, CPO and WAIC criteria. Table 4.7 shows the results.

Summary of model selection criteria

model	DIC	WAIC	CPO
model_bym_noise	310.6700	63504.3230	25.5496360
model_bym_without_noise	310.8911	181068.6388	30.3301443
model_bym2	298.1136	293.6834	0.2362707
model_bym2_prior	294.7418	298.5989	0.2372407

Table 4.7: Values for each modelling selection criterion of each model.

DIC values are relatively similar. Although, BYM2 prior model has the smallest number of the DIC selection criterion. BYM2 model without the defined prior has larger value of DIC than BYM2-prior but smaller values of WAIC and CPO.

Another plot is shown in Figure 4.20 displaying the relative risks gained based on some color mapping scale to allow for more inferences to be made. Relative risks are the exponential posterior means of the linear predictor.

All models provide relatively similar relative risks. It seems that in the majority of municipalities/communities the expected all period average cases do not differ significantly from the real ones.

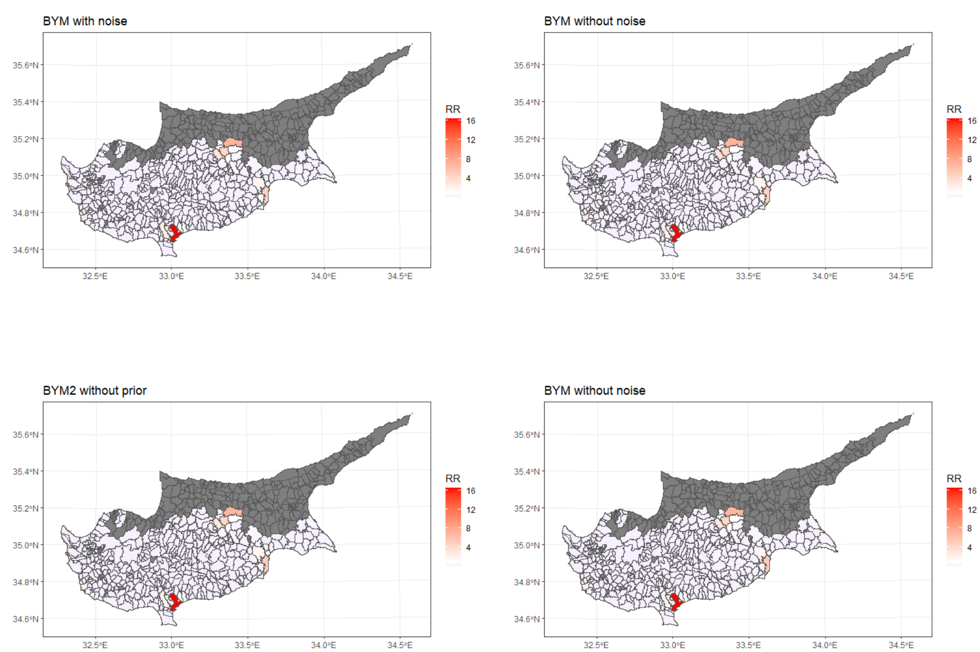


Figure 4.20: Maps of Relative Risks for each model.

Conclusions

According to the population of each municipality/community per gender and age group the overall average number of Covid-19 cases at municipality/community can be compared to the expected. It is observable that there were some municipalities/communities with larger number of confirmed cases than those expected. Considering the BYM2 model with a defined prior such municipalities/communities are shown in the Table 4.8,

Municipalities/communities with RR > 1 -
BYM2 with prior

municipality_town_lst	RR
DIMOS LEMESOU	16.529264
DIMOS LEFKOSIAS	6.897659
DIMOS LARNAKAS	5.373658
DIMOS STROVOLOU	4.440690
DIMOS LAKATAMEIAS	2.886407
DIMOS K. POLEMIDION	2.865418
DIMOS PAFOU	2.358370
DIMOS YPSONA	1.942050
DIMOS ARADIPPOU	1.851722
DIMOS MESA GEITONIAS	1.328182
DIMOS GERMASOGEIAS	1.246002
DIMOS AG ATHANASIOU	1.115443
DIMOS LATSION	1.114595
DIMOS AGLANTZIAS	1.087918
DIMOS EGKOMIS	1.034853

Table 4.8: Municipalities/communities with more confirmed cases than expected.

4.8 Further Work: Spatio-temporal analysis and models

So far our investigation is being focused on the spatial modelling of the propagation of Covid-19 in the community. However, time plays a significant role and it can be taken into account as well. A disease except from its spatial evolution is characterized for its temporal evolution. We have seen cases form all over the world where the curve has being flattened or increases as time passes. Thus, a simultaneous investigation of the change on Covid-19 cases according to space-time will lead us to some deeper conclusions. In Spatio-temporal analysis we deal with two dimension, time and location. In R a package which is reliable for implementing such cases is the so-called 'spacetime' package. Except from spatial dependence among neighbouring areas when referring to diseases outbreak spatial dependence depends also on temporal attributes as new cases may emerge or response measures are imposed. Before fitting temporal models a time-series object should be created with dominant element the laboratory result date with Date format given cycle frequency. Moraga in her book Geospatial Health Data: Modeling and Visualization (Moraga, 2019) introduced a brief theory of spatio-temporal models and presents a spatio-temporal model with parametric time trends that expresses the logarithm of the relative risks as proposed by Bernardinelli et al. (Bernardinelli et al., 1995) and models in which linearity is not required and use non-parametric model for time (Schrödle and Held, 2011). The aforementioned spatio-temporal models can be implemented with R-INLA package as well. Therefore, a step forward to the current study is the addition of a temporal variable which can be the date of laboratory result.

Chapter 5

Discussion and Results

This thesis investigated the spatial propagation of Covid-19 in the Republic of Cyprus in order to draw inferences about the efficiency of response measures taken by the Cypriot government.

The role of spatial analysis information is twofold. Firstly, spatial transmission, confirmed cases hot spots and clustering can be identified. Secondly, spatial relative risk per week at municipality/community level allow the detection of regions with high transmission risk, which is helpful for the scientific prevention and control of the pandemic in the island. The variable of interest was mainly the weekly number of confirmed cases at municipality/community level and it has been shown that municipalities/communities with the highest number of cases were located at the 5 main district municipalities of Cyprus and their surrounding municipalities/communities. Dimos Lemesou and the most populated municipalities/communities of Limassol district were the main centres of diffusion, while municipalities/communities with small number of residents (mainly mountainous villages) had fewer cases and relatively stable coverage. Spatial autocorrelation statistics using Moran's I based on queen neighbour structure were used to measure the spatial autocorrelation among municipalities/communities regarding the number of weekly cases. Both Global and Local Moran's I were evaluated, as Global Moran's I can only indicate whether space is clustered without defining the locations. The local Moran's I is able to display where hot-spots and clusters appear. As shown in subsection 4.5.2, the Global Moran's I index, p-test value, and Z-statistical score were used to determine a significant spatial association among the confirmed Covid-19 cases in the Republic of Cyprus. Applying Local Moran's I during 30/03/2020-28/03/2021 for an interval of 2 weeks only significant clustering municipalities/communities which belong to the high-high clustering category were detected. The aim of LISA was to identify clustered regions of Covid-19 weekly cases at different clustering category and associate them with response measures imposed during the current weeks interval. The pandemic spread not only to the critical municipalities/communities near the 5 main district municipalities, but also within them. The closer to a high-risk area, the greater the risk. This finding is helpful for governments at all levels to take effective classified prevention and control measures. Standardized incidence ratio was used to derive the expected number of cases at municipality/community level depending on a stratum combination of age group and gender. There are in total 34 stratum derived from two gender and 17 different age groups. SIR approach has been applied to the average confirmed cases at municipality/community level during the study period. Due to the limitations arised from SIR, bayesian models were fitted instead. In total four models (BYM, BYM-noise, BYM2, BYM2-prior) were fitted such that the Relative Risk of weekly cases is extracted. BYM2-prior model was the one with the best fit on the data and according to this model Dimos Lemesou, Dimos Lefkosias and Dimos Larnakas were the three municipalities/-communities with the highest Relative Risk values.

It seems that the key factors affecting the dissemination of Covid-19 pandemic in the island were the source of infection, the transmission in neighbouring municipalities/communities, and the population. The propagation of Covid-19 is associated to the prevention measures applied, such as controlling the source of infection and cutting off transmission routes.

In the community the spread of the Covid-19 pandemic is strongly related with human mobility as it can exacerbate the spread of the virus, impose alarm in the public health sector and threaten human lives. First lockdown seemed to has vast benefits as cases were decreased significantly in April and May and no for four weeks no clusters were detected. However, after the withdraw of it the number of cases began to rise gradually. This study elaborates on the investigation of response measures efficiency based on spatial analysis implementation. Therefore, further study is worth it as already implemented measures can be improved.

Main results are going to be presented referring to three sub-periods, the period of the first lockdown, summer months during measures were not as strict and the period of the second wave of the pandemic.

As during the study period no vaccine existed the only immediate solution for flattening the curve was the enforcement of quarantine measures. Initially, following the measures imposed in the rest of Europe a general strict lockdown applied in the Republic of Cyprus despite the social, psychological and economic adverse effects as the main priority was the protection of public health. The first lockdown applied from 16/03/2020 with a step by step withdraw of strict measures starting by the end of May 2020. The first goal was to assess spatial propagation of the virus in Cyprus. The investigation was based on weekly confirmed cases at municipality/community level.

An initial exploratory analysis showed that most cases occurred between November 2020 and January 2021 with the peak in the number of cases arises on December of 2020. Hovmöller plot, Figure 4.9, displays a times series visualization of the weekly cases rate per 1,000 people which concludes that November 2020 until March 2021 were the most highly infected months. The same conclusion is reached by the temporal mean Figure **??**. Also, the evaluation of Global spatial autocorrelation revealed the presence of relatively large number of positive global autocorrelation statistic during the weeks ”21/09/2020-27/09/2020”, ”28/09/2020-05/10/2020”, ”06/10/2020-12/10/2020”, ”13/10/2020-19/10/2020”, ”20/10/2020-26/10/2020”. Values of global autocorrelation were 0.124, 0.170, 0.153, 0.134 and 0.127 respectively which means that during these weeks clusters of confirmed cases are detected. This phenomenon is strongly related with the measures imposed this period. Schools were one of the main sources of contamination and there were remaining opened until December from their starting date on September. The general hygiene rules and Covid-19 protocol measures were applied at schools. However, more precisely at early ages the prevention of human interaction was difficult to handle. Students at schools were sitting with fewer than 2 metres apart. Unfortunately, Cyprus was not prepared enough at schools with the appropriate school furniture or single desks to prevent the propagation during classes. Primary school students did not wear protective masks. Also, from May until the end of November early days of December sport clubs, entertainment venues, cafes and restaurants were opened with the basic restriction and protection rules. Thus, they might become possible sources of the virus. After the withdraw of the first lockdown employees had to return back to their offices. Although, many private companies promote the teleworking of a number of employees in order to minimize the natural presence of many people at the same place. The aforementioned accompanied with the winter period had massive impact on the increase of Covid-19 cases.

During the first lockdown the global spatial autocorrelation was not as high with its peak appeared at first days of April until the first effects of lockdown detected. By mid April there was a reduction in the global spatial autocorrelation. During summer months Cyprus was one of the European countries with the best epidemiological condition and this can be reflected on LISA maps, Figure 4.17, as during summer weeks clusters were significantly minimized. During summer period domestic tourism was showed a great interest, while leisure places were open to public again. Moreover, schools in Cyprus are closed during summer months which might surely contributed to the reduction of cases. The two dominant response measures imposed were the mandatory use of masks at indoor spaces and the prohibition of international travellers.

Local Spatial autocorrelation statistic was used for the detection of clusters. During the study period 5 or less significant clusters are found consisting of the district municipalities and their surrounding municipalities/communities. LISA maps displayed for every two weeks during the period from 30/03/2020-20/03/2021. The choice of the starting date was chosen as response measures need at least two weeks to show beneficial effects as Covid-19 is not only highly contagious but some infected people are asymptomatic and they can spread the virus in the community. Since the general lockdown introduced on 16th of March 2020 the starting date chosen was the 30/04/2020. It is observable that lockdown and absence of human to human communication had vast benefits on the prevention of the pandemic. People were also, afraid during the first wave and thus they were following the rules strictly. Clustering showed a significant reduction as for weeks 22/06/2020-19/07/2020 no clusters detected. It appears that from 10/12/2020 the clustering of incidence remains high across time. During the first wave, 09/03/2020-25/05/2020 the municipalities/communities with a high incidence can be observed through the red colored areas in LISA maps 4.17. It is observable that the size of clusters is reduced as time passes during the above time period which indicates the beneficial effect of the lockdown.

On 18th of November 2020 a local lockdown had been imposed in Lemesos and Pafos districts as were detected as highly infected. Although, from the current study analysis it seems that local lockdown in Pafos was not essential. Figures 4.13 and 4.14 show that most weekly cases and their corresponding rate according to population occurred in Lemesos and Lefkosia. However, a local lockdown had not been applied in Lefkosia. Lefkosia is the capital city of Cyprus and all main public, governmental services and private companies are located there. A possible lockdown in Lefkosia could shrink the economic output. Government has supported economically those affected during first lockdown, although this could not be extended in time. During winter months the Republic of Cyprus was facing the second wave of the pandemic. Weather conditions, use of indoor spaces and flu reasons were undoubtedly some reasons about this situation. Personally, apart from the aforementioned and as results shown in accordance with A.1 the relaxation of measures during summer months was one of the main reasons as citizens thought that life returned back to normal and were not prepared for a second wave.

As discussed previously, December was the pick of confirmed cases indicating a second wave of the pandemic in the Republic of Cyprus. From December 2020 a new decree of measures announced and according to LISA maps, and the reduction of clusters, they had been beneficial. The reduction although was not as significant as during the first lockdown and this might be influenced due to the winter period or reluctance of citizens to cooperate with the measures.

SIR and Bayesian spatial models showed that in the majority municipalities/communities had the same number of actual and expected cases. However, the 4 main district municipalities (DIMOS LEMESOU, DIMOS LEFKOSIAS, DIMOS LARNAKAS, DIMOS PAFOU) and their most populated surrounding municipalities presented more cases than those expected according to their demographic characteristics such as population per age and gender. This observation may be associated with the weakness of controlling the pandemic in high populated areas despite the response measures implemented.

The only way for an early detection of the virus is the vast and immediate number of Covid-19 tests. However, during the first wave in Cyprus only the Institute of Neurology and Genetics was conducting laboratory test and results were available after 2-3 days. Hence, quarantine measures were the only solution in facing the pandemic as neither a vaccine nor a treatment existed. The lockdown implemented did indeed help to slow the spread of Covid-19, but this effect declines as the time passes. Strict lockdowns can be only a temporary solution and implemented extremely rigorous as permanent lockdowns can have adverse effects on the economy of a country. For instance, not all jobs can be implemented through teleworking. This leads to economic losses of individuals, companies and the country itself.

Although the priority was the public health with the main concern of the government associated with the reliability of the healthcare and public health services, causing shortage of hospital beds and delays in the diagnosis of new cases. The first lockdown was imposed mostly to avoid the overload of health system. Government forced several measures, however their effectiveness relies on the ability of individuals and health providers to keep up with the strict quarantine measures.

Chapter 6

Conclusion

Project Strengths and Limitations

Limitations

Covid-19 data retrieved are highly variable and uncertain as at the very beginning of the pandemic in the country laboratory tests were limited and thus the number of cases per day could be underestimated.

Another limitation of the data is that confirmed cases belong to infected people who contacted the Ministry of Health. However, there were cases which were not declared as close contacts of confirmed infections for personal or professional reasons, and thus many cases were undetected.

The study is also limited for future comparisons as now more than the 70% of the population had been vaccinated and Delta variant is propagated within the country and is much more contagious. Thus, the underlying spatial determinants of Covid-19 cases and deaths, relative risks and spatial autocorrelation are more likely to change as the pandemic continues.

Moreover, as the data were gained after an approval by the Ministry of Health a specific number of variables could be obtained. Therefore, a data analysis using additional variables as the number of tests, deaths, hospitalizations, severity of symptoms at municipality/community level might address more accurately the efficacy of measures implemented.

Strengths

The current study can be the framework for future research associated with the spatial analysis of Covid-19 in Cyprus for the detection of possible underlying causes of Covid-19 cases. As the pandemic continues mainly due to the Delta variant, it is highly recommended that a future spatial analysis is conducted to fully understand the spatial distribution of Covid-19 cases and how they vary across geographic regions.

Testing can bring a more accurate epidemiological picture of Covid-19 in Cyprus. Massive tests are conducted daily in Cyprus and the number of confirmed cases have increased. The current study can be the base for a future work which should incorporate testing data into the analysis once these data are made publicly available. Another recommendation is that Covid-19 data can be split at the zip code level providing opportunity for the exploration of the virus in sub-county level.

This study demonstrated the early stages of the Covid-19 pandemic in Cyprus through spatial analysis. Examining the spatial spread in the early stages is very important to prevent further transmission. Covid-19 studies regarding the dissemination of the Republic of Cyprus are limited, especially using spatial analysis and statistics. Therefore, the current project becomes supportive for the country and the scientific community.

Conclusion

Covid-19 pandemic affected several countries in the world and quarantine measures were the only solution before vaccination. The Republic of Cyprus and decisions on response measures implemented by the Cypriot government as discussed on Chapter 5 had both advantages and disadvantages. Pandemic was easier to be handled at the beginning, while after summer months the number of cases increased. Examining the spatial spread of the virus is very important to prevent further transmission, while studies based on spatial analysis of Covid-19 in Cyprus are very limited. Therefore, the current study may become very supportive. Future research, such as a study examining the spatio-temporal spread of Covid-19, will contribute to the control and prevention of this disease.

Ethics approval and consent to participate: Covid-19 data contains anonymous cases in the research presented therefore, ethical approval was not necessary.

Availability of data and materials: The code implemented and the datasets used and analysed during the current study are available through github: https://github.com/ml5298.

Competing interests: The author declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this study.

Bibliography

Anselin, L., 1989. What is special about spatial data? alternative perspectives on spatial data analysis.

Anselin, L., 1995. Local indicators of spatial association—lisa. *Geographical analysis*, 27(2), pp.93–115.

Arauzo-Carod, J., 2021. A first insight about spatial dimension of covid-19: analysis at municipality level. *Journal of public health*, 43(1), p.98–106.

Aràndiga, F., Baeza, A., Cordero-Carrión, I., Donat, R., Martí, M., Mulet, P. and Yáñez, D., 2020. A spatial-temporal model for the evolution of the covid-19 pandemic in spain including mobility. *Mathematics*, 8(10).

Auchincloss, A., Gebreab, S., Mair, C. and Diez Roux, A., 2012. A review of spatial methods in epidemiology, 2000–2010. *Annual review of public health*, 33, p.107–122.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. and Songini, M., 1995. Bayesian analysis of space—time variation in disease risk. *Statistics in medicine*, 14(21-22), pp.2433–2443.

Best, N., Richardson, S. and Thomson, A., 2005. A comparison of bayesian spatial models for disease mapping. *Statistical methods in medical research*, 14(1), pp.35–59.

Bivand, R., Pebesma, E. and Gomez-Rubio, V., 2013a. Applied spatial data analysis with r [Online]. Available from: `https://asdar-book.org/`.

Bivand, R., Pebesma, E. and Gomez-Rubio, V., 2013b. Applied spatial data analysis with r [Online]. Available from: `https://asdar-book.org/`.

Bivand, R., Pebesma, E., Gomez-Rubio, V. and Pebesma, E., 2013. *Applied spatial data analysis with r*, vol. 2. New York: Springer.

Brauer, F., 2008. Compartmental models in epidemiology. *Mathematical epidemiology*. Berlin, Heidelberg: Springer, p.19–79.

Cheng, Q., Lu, X., Wu, J., Liu, Z. and Huang, J., 2016. Analysis of heterogeneous dengue transmission in guangdong in 2014 with multivariate time series model. *Scientific reports*, 6(1), p.1–9.

Cordes, J. and Castro, M., 2020. Spatial analysis of covid-19 clusters and contextual factors in new york city. *Spatial and spatio-temporal epidemiology*, 34, p.100355.

Cramb, S., Duncan, E., Baade, P. and Mengersen, K., 2018. Investigation of bayesian spatial models.

Freni-Sterrantino, A., Ventrucci, M. and Rue, H., 2018. A note on intrinsic conditional autoregressive models for disconnected graphs. *Spatial and spatio-temporal epidemiology*, 26, pp.25–34.

Gaetan, C. and Guyon, X., 2010a. *Spatial statistics and modeling*, vol. 90. New York: Springer.

Gaetan, C. and Guyon, X., 2010b. *Spatial statistics and modeling*, vol. 90. New York: Springer.

Glen, S., n.d. Moran's i: Definition, examples" from statisticshowto.com: Elementary statistics for the rest of us! [Online]. Available from: https://www.statisticshowto.com/morans-i/.

Gross, B., Zheng, Z., Liu, S., Chen, X., Sela, A., Li, J., Li, D. and Havlin, S., 2020. Spatio-temporal propagation of covid-19 pandemics. *Epl (europhysics letters*, 131(5), p.58003.

Hijmans, R., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J., Lamigueiro, O., Bevan, A., Racine, E., Shortridge, A. and Hijmans, M., 2015. Package 'raster'. R package, 734.

Kergassner, A., Burkhardt, C., Lippold, D., Nistler, S., Kergassner, M., Steinmann, P., Budday, D. and Budday, S., 2020. Meso-scale modeling of covid-19 spatio-temporal outbreak dynamics in germany.

Kim, S. and Castro, M., 2020. Spatiotemporal pattern of covid-19 and government response in south korea (as of may 31, 2020. *International journal of infectious diseases*, 98, p.328–333.

Lee, J. and Wong, D.W., 2001. *Statistical analysis with arcview gis*. John Wiley & Sons.

MacTavish, R., Rémillard, L. and Davison, C., 2018. Environmental lead exposure and adult literacy in myanmar: An exploratory study of potential associations at the township level. *International journal of environmental research and public health*, 15(6).

Meyer, S., Held, L. and Höhle, M., 2017. Spatio-temporal analysis of epidemic phenomena using the r package surveillance. *Journal of statistical software* [Online], 77(11), p.1–55. Available from: https://doi.org/10.18637/jss.v077.i11.

Moraga, P., 2019. *Geospatial health data: Modeling and visualization with r-inla and shiny*. Chapman and Hall/CRC.

Ott, W., 1994. *Environmental statistics and data analysis*. CRC Press.

Ott, W.R., 2018. *Environmental statistics and data analysis*. Routledge.

Papa, A., Sidira, P. and Tsatsaris, A., 2016. Spatial cluster analysis of crimean-congo hemorrhagic fever virus seroprevalence in humans, greece. *Parasite epidemiology and control*, 1(3), p.211–218.

Piarroux, R., Barrais, R., Faucher, B., Haus, R., Piarroux, M., Gaudart, J., Magloire, R. and Raoult, D., 2011. Understanding the cholera epidemic, haiti. *Emerging infectious diseases*, 17(7).

Ramírez-Aldana, R., Gomez-Verjan, J. and Bello-Chavolla, O., 2020. Spatial analysis of covid-19 spread in iran: Insights into geographical and structural transmission determinants at a province level. *Plos neglected tropical diseases*, 14(11), p.0008875.

Riebler, A., Sørbye, S.H., Simpson, D. and Rue, H., 2016. An intuitive bayesian spatial model for disease mapping that accounts for scaling. *Statistical methods in medical research*, 25(4), pp.1145–1165.

Santermans, C., E., R., E., G., T., S., B., F., C., Q., C., V.B., W., H., T., K., T., V.R., F. and Testa, M., 2016. Spatiotemporal evolution of ebola virus disease at sub-national level during the 2014 west africa epidemic: model scrutiny and data meagreness. *Plos one*, 11(1), p.0147172.

Schrödle, B. and Held, L., 2011. Spatio-temporal disease mapping using inla. *Environmetrics*, 22(6), pp.725–734.

Spassiani, I., Sebastiani, G. and Palù, G., 2021. Spatiotemporal analysis of covid-19 incidence data. *Viruses*, 13(3), p.463.

Taddese, A., Baraki, A. and Gelaye, K., 2019. Spatial modeling, prediction and seasonal variation of malaria in northwest ethiopia. *Bmc research notes*, 12(1), p.1–6.

Tennekes, M., 2018. tmap: Thematic maps in r. *Journal of statistical software*, 84(6), p.1–39.

Waller, L. and Gotway, C., 2004a. *Applied spatial statistics for public health data*, vol. 368. John Wiley Sons.

Waller, L. and Gotway, C., 2004b. *Applied spatial statistics for public health data*, vol. 368. John Wiley Sons.

Waller, L. and Gotway, C., 2004c. *Applied spatial statistics for public health data*, vol. 368. John Wiley Sons.

Waller, L. and Gotway, C., 2004d. *Applied spatial statistics for public health data*, vol. 368. John Wiley Sons.

Xiong, Y., Wang, Y., Chen, F. and Zhu, M., 2020. Spatial statistics and influencing factors of the covid-19 epidemic at both prefecture and county levels in hubei province, china. *International journal of environmental research and public health*, 17(11), p.3903.

Appendix A

Appendix

A.1 Response Measures

Response measures collected include mass gathering cancellations (for specific events or a ban on gatherings of a particular size); closure of public spaces (including restaurants, entertainment venues, non-essential shops, partial or full closure of public transport etc.); closure of educational institutions (including daycare or nursery, primary schools, and secondary schools and higher education); 'stay-at-home' recommendations for risk groups or vulnerable populations (such as the elderly, people with underlying health conditions, physically disabled people etc.); 'stay-at-home' recommendations for the general population (which are voluntary or not enforced); and 'stay-at-home' orders for the general population (these are enforced and also referred to as 'lockdown'), social circles or bubbles, use of protective masks in all public spaces/on closed public spaces/transport (mutually exclusive voluntary recommendations and mandatory obligations shown separately), and also teleworking recommendations, closure of workplaces and adaptation of workplaces.

First lockdown has been imposed from 16/03/2020 with a step by step withdraw started by the end of May 2020. For about 2 and a half months Cyprus was under strict quarantine measures as the rest of European countries. Alongside with quarantine measures the president of Cyprus following the idea first introduced by the prime-minister of Greece announced on 23rd of May 2020 a new strict measure. This measure aimed to reduce human mobility, thus from the 24th of March until the 13th of April every citizen was restricted to send a message in order to get out of his house. Every citizen had one message per day. A list of possible cases defining human needs was published and every citizen had to select the corresponding number of the list in order to have the licence of getting out of his house. Everyone who had violated the restrictions was imposed a fine of 150 euro which was increased in 300 euro some day after.

During the first lockdown flights were postponed. Although Ministry of Foreign Affairs carried out chartered flights for Cypriots who wanted to repatriate. Every Cypriot citizen arriving in the country was imposed to follow a strict quarantine of 14 days in hotels offered free of charge by the government in order to limit potential cases. This measure was in force until 25th of May 2020.

Cyprus government except from the daily tests of close contacts carried out random testing in the health sector and workplace for an early detection of positive cases as many people appeared asymptomatic. On 16th of November rapid test program started firstly at Lemesos and Pafos districts. The program was addressed to all citizens and especially to those employed in public service areas, delivery services and front-line professionals. The program was conducted at sampling points across the country. Sampling points were announced regularly. Citizens will be informed about their result within 24 hours via mobile messaging and within 15 minutes in hand-written form.

On 18th of November a new local lockdown impose at Lemesos and Pafos districts for 3 weeks.

Social distancing measures' data obtained by the European Centre for Disease Prevention Control

(ECDC) reflect the study period from 09/03/2020-28/03/2021. Staring date has been used to consider the selected measures in that time period.

A timeline has been created to display visually the period and duration of each response measure. Table 2 on A.1 provides the definition of response measures.The timeline is shown in Figure A.1.

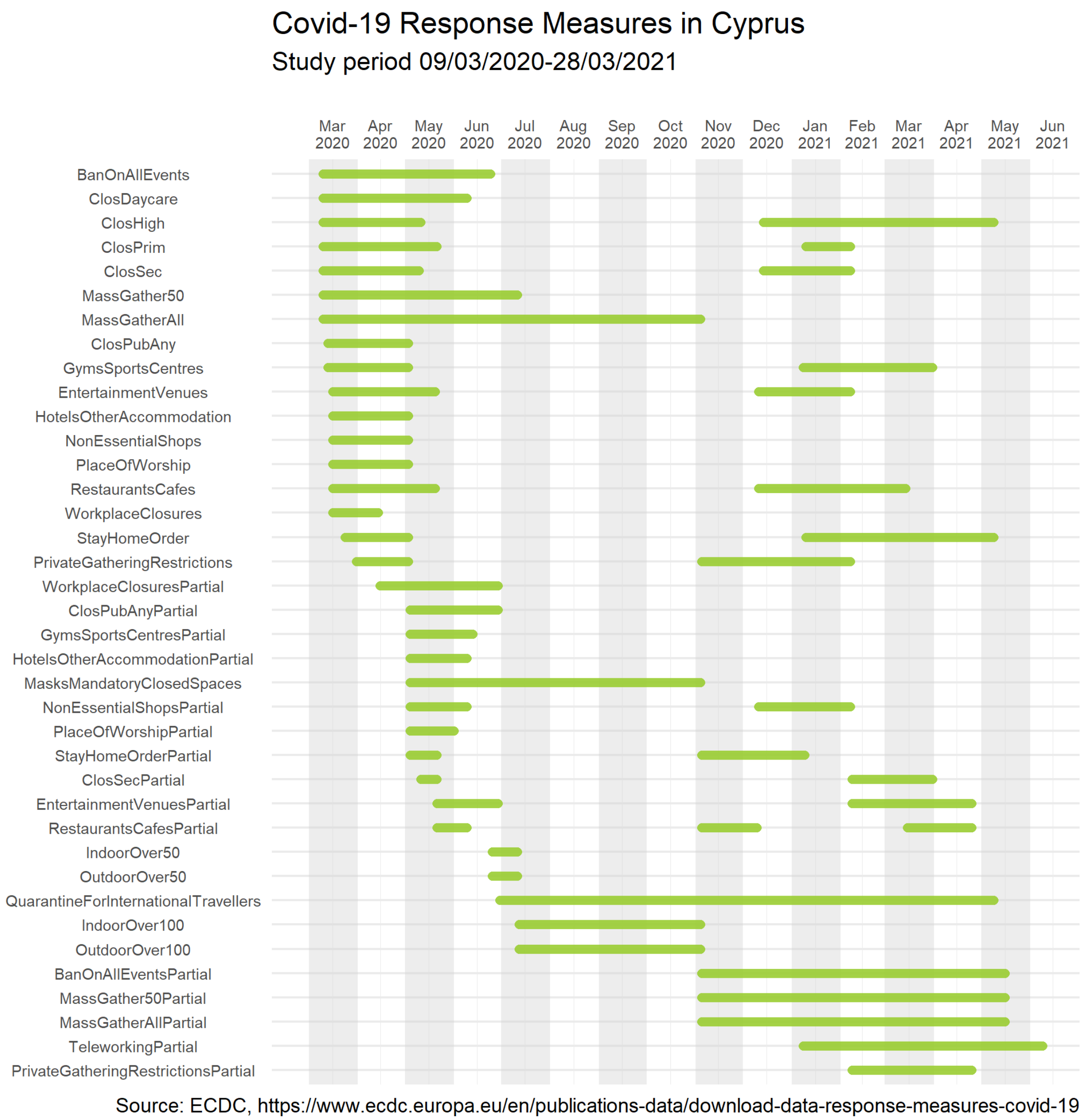


Figure A.1: Timeline of response measures imposed between 09/03/2020-28/03/2021.

Response Measures Definition	
Variable Definition Code	Response_measure
StayHomeOrder	Stay-at-home orders for the general population (these are enforced and also referred to as 'lock-down')
RegionalStayHomeOrder	Regional stay-at-home orders for the general population at least in one region (these are enforced and also referred to as 'lockdown')

StayHomeGen	Stay-at-home recommendations for the general population (which are voluntary or not enforced)
StayHomeRiskG	Stay-at-home recommendations for risk groups or vulnerable populations (such as the elderly, people with underlying health conditions, physically disabled people, etc.)
SocialCircle	Social circle/bubble to limit social contacts e.g. to limited number of households
PrivateGatheringRestrictions	Restrictions on private gatherings
ClosDaycare	Closure of educational institutions: daycare or nursery.
ClosPrim	Closure of educational institutions: primary schools.
ClosSec	Closure of educational institutions: secondary schools.
ClosHigh	Closure of educational institutions: higher education.
MassGatherAll	Interventions are in place to limit mass/public gatherings (any interventions on mass gatherings up to 1000 participants included)
BanOnAllEvents	Interventions are in place to limit all indoor/outdoor mass/public gatherings
IndoorOver50	Interventions are in place to limit indoor mass/public gatherings of over 50 participants
IndoorOver100	Interventions are in place to limit indoor mass/public gatherings of over 100 participants
IndoorOver500	Interventions are in place to limit indoor mass/public gatherings of over 500 participants
IndoorOver1000	Interventions are in place to limit indoor mass/public gatherings of over 1000 participants
OutdoorOver50	Interventions are in place to limit outdoor mass/public gatherings of over 50 participants
OutdoorOver100	Interventions are in place to limit outdoor mass/public gatherings of over 100 participants
OutdoorOver500	Interventions are in place to limit outdoor mass/public gatherings of over 500 participants
OutdoorOver1000	Interventions are in place to limit outdoor mass/public gatherings of over 1000 participants
ClosPubAny	Closure of public spaces of any kind (including restaurants, entertainment venues, non-essential shops, partial or full closure of public transport, gyms and sport centers, etc).
EntertainmentVenues	Closure of entertainment venues
ClosureOfPublicTransport	Closure of public transport
GymsSportsCentres	Closure of gyms/sports centres
HotelsAccommodation	Closure of hotels/accommodation services
NonEssentialShops	Closures of non-essential shops
PlaceOfWorship	Closure of places of worship
RestaurantsCafes	Closure of restaurants and cafes/bars

MasksVoluntaryAllSpaces	Protective mask use in all public spaces on voluntary basis (general recommendation not enforced)
MasksVoluntaryClosedSpaces	Protective mask use in closed public spaces/transport on voluntary basis (general recommendation not enforced)
MasksMandatoryAllSpaces	Protective mask use in all public spaces on mandatory basis (enforced by law)
MasksMandatoryClosedSpaces	Protective mask use in closed public spaces/transport on mandatory basis (enforced by law)
Teleworking	Teleworking recommendation
AdaptationOfWorkplace	Adaptation of workplaces (e.g. to reduce risk of transmission)
WorkplaceClosures	Closures of workplaces
StayHomeOrderPartial	Stay-at-home orders for the general population (these are enforced and also referred to as 'lockdown') – partially relaxed measure
RegionalStayHomeOrderPartial	Regional stay-at-home orders for the general population at least in one region (these are enforced and also referred to as 'lockdown') – partially relaxed measure
StayHomeGenPartial	Stay-at-home recommendations for the general population (which are voluntary or not enforced) – partially relaxed measure
StayHomeRiskGPartial	Stay-at-home recommendations for risk groups or vulnerable populations (such as the elderly, people with underlying health conditions, physically disabled people, etc.) – partially relaxed measure
SocialCirclePartial	Social circle/bubble to limit social contacts e.g. to limited number of households – partially relaxed measure
PrivateGatheringRestrictionsPartial	Restrictions on private gatherings – partially relaxed measure
ClosDaycarePartial	Closure of educational institutions: daycare or nursery – partially relaxed measure
ClosPrimPartial	Closure of educational institutions: primary schools – partially relaxed measure
ClosSecPartial	Closure of educational institutions: secondary schools – partially relaxed measure
ClosHighPartial	Closure of educational institutions: higher education – partially relaxed measure
MassGatherAllPartial	Interventions are in place to limit mass/public gatherings (any interventions on mass gatherings up to 1000 participants included) – partially relaxed measure
BanOnAllEventsPartial	Interventions are in place to limit all indoor/outdoor mass/public gatherings – partially relaxed measure
IndoorOver50Partial	Interventions are in place to limit indoor mass/public gatherings of over 50 participants – partially relaxed measure

IndoorOver100Partial	Interventions are in place to limit indoor mass/public gatherings of over 100 participants – partially relaxed measure
IndoorOver500Partial	Interventions are in place to limit indoor mass/public gatherings of over 500 participants – partially relaxed measure
IndoorOver1000Partial	Interventions are in place to limit indoor mass/public gatherings of over 1000 participants – partially relaxed measure
OutdoorOver50Partial	Interventions are in place to limit outdoor mass/public gatherings of over 50 participants – partially relaxed measure
OutdoorOver100Partial	Interventions are in place to limit outdoor mass/public gatherings of over 100 participants – partially relaxed measure
OutdoorOver500Partial	Interventions are in place to limit outdoor mass/public gatherings of over 500 participants – partially relaxed measure
OutdoorOver1000Partial	Interventions are in place to limit outdoor mass/public gatherings of over 1000 participants – partially relaxed measure
ClosPubAnyPartial	Closure of public spaces of any kind (including restaurants, entertainment venues, non-essential shops, partial or full closure of public transport, gyms and sport centers etc) – partially relaxed measure
EntertainmentVenuesPartial	Closure of entertainment venues – partially relaxed measure
ClosureOfPublicTransportPartial	Closure of public transport – partially relaxed measure
GymsSportsCentresPartial	Closure of gyms/sports centres – partially relaxed measure
HotelsAccommodationPartial	Closure of hotels/accommodation services – partially relaxed measure
NonEssentialShopsPartial	Closures of non-essential shops – partially relaxed measure
PlaceOfWorshipPartial	Closure of places of worship – partially relaxed measure
RestaurantsCafesPartial	Closure of restaurants and cafes/bars – partially relaxed measure
MasksVoluntaryAllSpacesPartial	Protective mask use in all public spaces on voluntary basis (general recommendation not enforced) – partially relaxed measure
MasksVoluntaryClosedSpacesPartial	Protective mask use in closed public spaces/transport on voluntary basis (general recommendation not enforced) – partially relaxed measure
MasksMandatoryAllSpacesPartial	Protective mask use in all public spaces on mandatory basis (enforced by law) – partially relaxed measure

MasksMandatoryClosedSpacesPartial	Protective mask use in closed public spaces/transport on mandatory basis (enforced by law) – partially relaxed measure
TeleworkingPartial	Teleworking recommendation or workplace closures – partially relaxed measure
AdaptationOfWorkplacePartial	Adaptation of workplaces (e.g. to reduce risk of transmission) – partially relaxed measure
WorkplaceClosuresPartial	Closures of workplaces – partially relaxed measure
date_start	Start date of the intervention/response measure DD/MM/YYYY
date_end	End date of the intervention/response measure (NA indicates a measure that is still active on the date of the file/end date in the future)

Table A.1: Descriptions of response measures imposed.