Spatial patterns of reported crime in Italy before, during and after the COVID-19 pandemic

Michele Brunelli

February 9, 2026

Abstract

Some text here later

1 Introduction

The analysis of the spatial distribution of crime represents an important topic in applied geography and spatial data science, as criminal phenomena are intrinsically linked to the territorial, social, and institutional contexts in which they occur. Crime, in fact, is not randomly distributed in space, but tends to exhibit patterns of concentration, persistence, and clustering, shaped by socio-economic factors, urban structure, and the opportunities offered by different territorial contexts.

The COVID-19 pandemic constitutes a major shock that has profoundly altered everyday life, social interactions, and economic activities in recent years. Lockdown measures, mobility restrictions, and changes in people's routines have likely affected both the overall volume of reported crimes and their spatial distribution. From a strictly geographical perspective, the pandemic can be interpreted as a natural experiment that allows observation of how crime patterns respond to sudden changes in human behaviour across space.

This project analyses the spatial patterns of reported crimes in Italy over the period 2014–2023, comparing three distinct phases: the pre-pandemic period (2014–2019), used as a baseline; the pandemic phase (2020–2021); and the recovery phase (2022–2023). The aim is to assess whether the observed changes in crime were homogeneous across the national territory or whether they assumed a spatially differentiated configuration, characterised by areas with specific local dynamics.

The analysis is conducted across three different geographical levels. A macro level, defined by five macro-areas dividing Italy into North-East, North-West, Centre, South, and Islands; a meso level, corresponding to the 20 Italian regions; and a micro level, at which the 107 provinces into which the country is divided are analysed. Given that Italy is characterised by territories with markedly different demographic sizes crime rates are considered to achieve an appropriate result.

From a methodological perspective, the project adopts a reproducible analysis pipeline implemented in Python, integrating official statistical data with geographic boundaries and interactive visualisation tools. The analysis relies on exploratory spatial analysis techniques, including thematic cartography (choropleth maps) and local indicators of spatial association (LISA), aimed at identifying spatial patterns and clusters. To allow users to explore the distribution of crime across space and time, a web-based application has also been developed.

The objective of this work is not to establish causal relationships between the COVID-19 pandemic and crime trends, but rather to provide an exploratory and spatially explicit analysis of the transformations observed during a period of profound social disruption. The project therefore seeks to highlight both persistent—and pre-existing—territorial inequalities and local variations associated with the pandemic period.

2 Data

This section describes the data used for the analysis presented above. Official statistics on reported crimes were obtained by querying the SDMX APIs exposed via web services by Istat[1]. SDMX (Statistical Data and Metadata eXchange) is a standard widely adopted by statistical institutions to provide access to time series in a standardized and reproducible manner.

Data on crime rates are normilized per 100k inhabitants, and made available by Istat in several formats. For convenience, the data were downloaded in CSV format and subsequently aggregated—one dataset per year—into a single Parquet file. Parquet is a binary, columnar, and typed data format. This structure allows access to only the relevant columns without scanning the entire file each time, thereby minimizing input/output operations. Parquet is widely used in data analysis because it maximizes performance while minimizing both memory usage and storage requirements. Although the CSV files are described as containing annual data, they include observations for both the reference year and the subsequent year. For this reason, a consistency check was performed on the aggregated Parquet file in order to avoid duplicated values.

For map creation, Istat dataset were merged with geographic data provided by Eurostat GISCO[2]. Eurostat supplies geometries in GEOJSON format, organized by territorial levels. The levels used in this analysis are NUTS1 (macro-areas), NUTS2 (regions), and NUTS3 (provinces). Istat data rely on the 2006 version of the NUTS classification. As a result, it was necessary to manually map provinces that underwent administrative changes after 2006, most notably in Sardinia.

The data were aggregated into three time periods. The first corresponds to the pre-pandemic period (2014–2019) and reports the six-year average in order to obtain a robust baseline. The second period covers the years of acute pandemic impact (2020–2021) and reports their average. The third corresponds to the post-pandemic recovery phase and reports the average for the subsequent two years (2022–2023).

As mentioned above, the spatial analysis was conducted at three nested geographic levels. The macro level considers five areas of the Italian peninsula: North-East, North-West, Centre, South, and Islands. The meso level corresponds to regions, while the micro level corresponds to provinces.

The dataset containing crime rates classify offenses into 55 categories. Although the analytical pipeline and the web-based platform process all crime categories provided by Istat, the results discussed in this report focus on a subset of categories that are most relevant to the research question, as shown in Table 1. Crime categories were grouped based on the predominant modality of interaction involved in the offense, distinguishing between crimes requiring direct physical contact (*contact crimes*) and crimes primarily mediated by digital infrastructures (*digital crimes*). Based on changes in daily routines induced by the pandemic, a decrease in contact crimes is expected during the second analyzed period, while an increase in digital crimes is anticipated.

The dataset provided by Istat, however, presents several limitations that reduce the effectiveness of the analysis. Some of these issues could be addressed by improving data consistency, while others reflect structural biases[3] typical of large-scale administrative data. Annual data do not allow the identification of short-term or event-specific dynamics, but only the observation of broad trends. For instance, they do not make it possible to analyse in detail what occurred

[1]Istat (Italian National Institute of Statistics) is the official statistical office of Italy. Official data are available at `https://www.istat.it/banche-dati/`.

[2]Eurostat GISCO (Geographic Information System of the Commission) provides harmonized geospatial data for the European Union. Administrative boundaries and related spatial datasets are available at `https://ec.europa.eu/eurostat/web/gisco`.

[3]In this context, bias refers to a systemic distortion in the observed data that arises from the data genereation process itself, rather than from measurement odr processing errors.

Table 1: Crime categories discussed in the report (ISTAT codes)

ISTAT code	**Category (description)**	**Group**
PICKTHEF	Pickpocketing	Contact
BAGTHEF	Snatch theft	Contact
BURGTHEF	Residential burglary	Contact
STREETROB	Street robbery	Contact
RAPE	Sexual assault	Contact
SWINCYB	Online fraud and cyber scams	Digital
CYBERCRIM	Cybercrime	Digital
PORNO	Child sexual abuse material (CSAM) offences	Digital

between March and May 2020, the period of full lockdown in Italy. The most significant source of bias concerns crime reporting. By construction, the dataset includes only crimes reported to the authorities. Sociological theory, however, shows that not all individuals are equally able to report victimisation (Biderman and Reiss, 1967). Baumer (2002) demonstrates that the propensity to report crimes varies according to several factors, including social class, socio-economic status, and the urban context in which individuals live. As a result, crimes experienced by socially marginalised groups are systematically underreported. This issue is particularly pronounced for crimes related to sexual violence (Koss, 1992) and child abuse. Victims of these offences often live in the same physical environment as their abusers, which substantially increases the difficulty of reporting violence. It is therefore plausible that the incidence of such crimes increased significantly during the pandemic. However, the data may depict a different reality, driven by the practical impossibility of reporting abuse while sharing everyday spaces with the perpetrator.

3 Methodology

The analysis was conducted on administrative territorial units, using the NUTS classification across three nested spatial levels. The adoption of this spatial framework makes it possible to identify spatial patterns that may emerge, persist, or disappear across different geographical scales. Administrative units were selected because Istat provides crime statistics consistently at these levels, allowing for meaningful territorial comparisons. To account for the structural discontinuity introduced by the COVID-19 pandemic, crime data were aggregated into three distinct temporal periods. The first period represents the pre-pandemic phase (2014–2019) and is computed as a six-year average in order to reduce annual variability and obtain a robust baseline. The second period corresponds to the acute impact phase of the COVID-19 pandemic (2020–2021) and is calculated as a biennial average. The third and final period captures the post-pandemic recovery phase (2022–2023), also computed as a two-year average. This aggregation strategy enables a clear comparison between structurally different phases while limiting the influence of short-term fluctuations. The analysis relies on crime rates rather than absolute counts of reported offences. As shown in formula 1, crime rates are calculated as the number of reported crimes per 100,000 inhabitants, allowing comparability across territorial units and years.

$$\text{Rate}_{i,t} = \frac{\text{Number of reported crimes}_{i,t}}{\text{Residente population}_{i,t}} \times 100{,}000 \tag{1}$$

The analysis is structured into three components, corresponding to the three analytical pages of the web application. The first page presents an interactive map showing the average variation in reported crimes relative to the baseline period. The map allows users to explore differences between the pandemic and post-pandemic phases, as well as variations across the three spatial levels (macro, meso, and micro). This descriptive and exploratory step makes it possible to observe how the intensity of criminal phenomena changes over time and across space. The map helps identify general patterns, geographical gradients, and areas characterized by marked increases or decreases. At this stage, the spatial structure of the phenomenon is not formally tested; rather, the analysis provides an interpretative foundation for subsequent spatial analyses.

The second page introduces an analysis of global spatial autocorrelation using Moran's I. This index assesses whether observed crime rates follow a random spatial distribution or instead display systematic spatial clustering (Moran, 1950). In other words, Moran's I measures the degree of similarity between neighboring territorial units, allowing the identification of spatial concentration of high or low values. Applying this index to the data is particularly relevant because crime is an intrinsically spatial phenomenon and strongly dependent on territorial context. Examining Moran's I across different temporal periods also allows for an assessment of whether the pandemic altered the overall level of spatial structuring of crime. This may occur, for example, through the attenuation of pre-existing patterns or the emergence of new ones as a consequence of mobility restrictions and changes in daily routines.

The final page deepens the analysis through local indicators of spatial autocorrelation (LISA). Unlike Moran's I, LISA statistics (Anselin, 1995) enable the identification of local clusters and spatial outliers. This is achieved by distinguishing between areas characterized by high values surrounded by high values (High-High), low values surrounded by low values (Low-Low), and spatially dissimilar configurations. LISA is particularly useful because it allows for the precise localization of areas where crime tends to concentrate or disperse, highlighting intra-regional differences and dynamics that are not captured by global statistics. The use of LISA also makes it possible to assess whether the pandemic affected not only the intensity of criminal activity, but also the spatial location of crime clusters. This provides a more fine-grained interpretation of processes of territorial concentration and fragmentation. Local indicators of spatial autocorrelation (LISA) are therefore used to identify local clusters and spatial outliers. The analysis specifically aims to assess whether the pandemic modified the spatial distribution of crime clusters, with particular attention to contact-related crimes and digital offences. In doing so, it highlights potential processes of territorial concentration or spatial dispersion associated with the pandemic period.

4 Results

5 Conclusions

References

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

Baumer, E. P. (2002). Neighborhood disadvantage and police notification by victims of violence. *Criminology*, 40(3):579–616.

Biderman, A. D. and Reiss, A. J. (1967). On exploring the "dark figure" of crime. *The Annals of the American Academy of Political and Social Science*, 374:1–15.

Koss, M. P. (1992). The under detection of rape: Methodological choices influence incidence estimates. *Journal of Social Issues*, 48(1):61–75.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1–2):17–23.