MANCHESTER
1824
The University of Manchester

Aspect

Methods for Change

Critical Spatial Data Science

Dr Caitlin Robinson
University of Bristol

Corresponding author
Dr Caitlin Robinson
caitlin.robinson@bristol.ac.uk

Critical Spatial Data Science (or Geographic Data Science) analyses quantitative data with some form of spatial identifier - for example, a coordinate, a street name, or a census block – to generate new knowledge.

In carrying out analysis of spatial data, a critical analysis is typically underpinned by theories that help us to understand - and therefore to best represent - complex real-world processes. Critical Spatial Data Science has wide-ranging applications with the potential to provide new insights into the distribution and dynamics of populations and societies across space and time. However, most commonly it seeks to understand and evidence socio-spatial inequalities, for example, inequalities in health, infrastructure, or education. Critical Spatial Data Science is founded on principles of openness, transparency, and reproducibility. At its best, the approach can be used to evidence and challenge injustice and have real-world impacts beyond academia. The field is rapidly expanding, drawing on an increasingly diverse range of spatial methods and data.

Figure 1: Critical Spatial Data Science at the ESRC Festival of Social Science 2022, illustration by Jack Brougham jackbroughamdrawing.com

How does Critical Spatial Data Science create or contribute to change?

Critical Spatial Data Science is designed to be socially-impactful, providing evidence of socio-spatial inequalities that can underpin decision-making to reduce inequalities. Researchers are encouraged to think critically at every stage of the process when analysing spatial data using quantitative methods. Instead of an analysis driven by the data, we start from the research question of interest.

In Critical Spatial Data Science it is important to consider what changes or impacts the research seeks to achieve from the start of the project, engaging relevant stakeholders and communities as part of the process. As a result, the research outputs are likely to better acknowledge gaps and shortcomings in the data and analysis, and to be more representative of the complexity of the real world and stakeholder experiences.

What ideas or concepts influence this approach?

When it comes to Critical Spatial Data Science, it is often necessary to draw on a wide range of theories and concepts to inform the research process. On the one hand, we can think about theoretical underpinnings of the research as the theory of the methods themselves. There is a long-standing history of using quantitative approaches and techniques in human geography. Increasingly, we can also draw on computationally intensive techniques and machine learning methods from wider data science.

Machine learning algorithms can learn and adapt without following specific instructions, allowing them to analyse and draw inferences from patterns in data.

On the other hand – and fundamental to a critical approach to spatial data science - it is also necessary to draw on theory and concepts that help us to understand the social processes that we are trying to represent. These underpinning theories are wide ranging depending on the application area – including theories of social justice or urban change. For example, we might draw on concepts from data justice to help us to understand inequalities in the distribution digital technologies.

Why might I want to use Critical Spatial Data Science?

- To evidence socio-spatial inequalities. For Elizabeth Delmelle, Critical Spatial Data Science has the potential to make a "substantial imprint on some of society's more arduous problems" - whether that be evidencing the uneven distribution of urban heat amongst populations, or place-based health inequalities that accumulate over the course of a lifetime.

- To learn about social processes at scale. Critical Spatial Data Science is a powerful way of extrapolating the detailed insights and knowledge derived from qualitative research, at scales that would otherwise be timely and costly to achieve, for example, across cities globally or analysing change over decades.

- To combine multiple dimensions. Societies are inherently complex and there are often a wide range of dimensions that are important to represent. Spatial data science approaches can combine and analyse a myriad of different pieces of data simultaneously – social, demographic, economic, infrastructural, or political – to provide new insights and information.

- To tell a convincing story with data. Mapping spatial data in new ways is a powerful way of conveying complex information, that many people find intuitive to interpret. New and interactive ways of visualising spatial data can be useful, including interactive dashboards or story-maps.

Step by step guide to using Critical Spatial Data Science:

1. **Deciding on a focus of analysis.** A critical approach to spatial data science starts with that process of interest - whether that be poor air quality or population decline. A strong understanding of the process is integral to the research that follows. A classic example is the construction of a Social Vulnerability Index. Social vulnerability refers to how able populations are to withstand adverse impacts from stressors they are exposed to, including, for example, climate change. There are often multiple factors that shape vulnerability. In assessing how vulnerability varies spatially it is necessary to decide, based on a detailed review of qualitative evidence of the experiences of different populations, which factors are important to represent.

2. **Identifying appropriate spatial data.** Next, it is important to consider how the social phenomena of interest can be best represented using available spatial data. Often, it is not possible to collect new data at scale, which means it is necessary to rely on pre-existing data. Spatial datasets range from traditional Census and survey micro-data, to mobility data, satellite imagery and data from mobile devices. Keep in mind that available data is unlikely to have been collected with the aim of answering our specific research question. When selecting data important considerations include what is likely missing from the dataset (which can be important information in and of itself!), the scale at which it is available, and how closely it matches the process of interest. It is also important to consider whether the research is ethical, especially when using new, less well understood forms of data. Spatial datasets for analysing social and spatial inequality can be accessed from a wide range of sources, but a good place to start in the UK is the Consumer Data Research Centre.

Data feminism

In their book Data Feminism, Catherine D'Ignazio and Lauren Klein argue that when working with data it is important to "consider context". Data is neither neutral or objective and is often the produce of unequal power relations. Understanding of this context is essential for accurate and ethical analysis.

3. **Cleaning the data.** Perhaps the least glamorous, but often most important, stage of the research process is cleaning the data. There are almost always issues with spatial data – whether that be incompleteness, inaccuracy, repetitiveness, inconsistency - especially if you are using multiple data sources with varied characteristics. Data cleaning can therefore take many forms. Spatial data has what is called a projection - the means by which you display the coordinate system and your data on a flat surface. There are many different systems for projection, and when working with multiple datasets it is often necessary to reproject data to ensure datasets align. It is also important to make sure all datasets are of a common scale, checking for geometry issues and removing missing values.

4. **Spatially analysing the data.** Once the most appropriate dataset(s) have been identified, the researcher can decide whether further analysis would generate new insights and understanding. Almost always the answer to this question is yes! Analysis of your spatial data can take many forms – you might be searching for "hot-spots" in a dataset where a variable is especially high, or trying to classify or cluster your data in order to identify areas with similar characteristics. Or, you might be trying to understand the strength of the relationship between different variables across space (e.g. using spatial statistics), or between multiple scales (e.g. using multi-level modelling). To find out about the diversity of analytical approaches in spatial data science see the tutorials in Further Reading.

Dealing with uncertainty

Uncertainty refers to when some aspect of your data or results is imperfect or unknown. Uncertainty can result during a range of stages in the analytical process – whether that be uncertainty in your data (both and manufactured uncertainty) and the methodological approach. Although often framed as an issue, uncertainty can also be framed as "a feature, not a bug". Methods for evaluating uncertainty vary depending on your analytical approach.

5. **Interpreting and visualising the results.** When interpreting the results, the theories that were used to frame the analysis in step one really come to matter again, helping to contextualise the findings and to use them to tell a convincing story to the audience. It is important to avoid "fetishizing [our results] as definitive and final answers" and to be mindful that just because our approach is likely quantitative, the interpretation of results remains highly subjective and subject to methodological limitations. Decisions about how to visualise and present data are hugely important to avoid misleading the audience – as was the case in the presentation of the UK government COVID-19 charts to the public.

6. **Sharing the results.** A key element of the research process in Critical Spatial Data Science is how to share results. As the technical capability of research grows, so too does the sophistication of the outputs. Spatial data science research typically involves powerful visual elements (i.e., maps and charts) and careful consideration needs to be given to making these outputs as accessible and user friendly as possible. The final outputs are, however, just the tip of the iceberg - an iceberg that is underpinned by extensive data, analysis and code that should also be made openly available. Increasingly researchers use software such as GitHub to share these outputs and facilitate interactive software development.

Open and reproducible analysis and outputs

Many researchers in this field are moving away from using proprietary software tools where analysis is carried out in a 'black box' in which the internal workings are not revealed. Instead, there is growing commitment to values of openness, transparency, sharing and reproducibility in which coding is a core research tool. An open framework ensures that research is not conceptually or methodologically flawed, and helps to make visible the work of writing collaborative code. These principles are also increasingly applied to the final outputs. This is exemplified by the move towards open data products, which turn data into accessible information based on open principles.

Examples of Critical Spatial Data Science in social science research

Spatial inequality in the smart city

Researchers: *Prof. Rachel Franklin, Dr Eman Zied, Dr Kate Court (Newcastle University); Dr Caitlin Robinson (University of Bristol); Dr Jack Roberts (The Alan Turing Institute)*

A smart city is a technologically advanced urban area that uses different digital technologies and sensors to collect detailed data, that in turn can be used to make urban processes run more efficiently. Smart cities, and the technologies associated, have a tendency to be unevenly distributed. The aim of the Spatial Inequality in the Smart City project, funded by The Alan Turing Institute, was to investigate inequalities in the distribution of sensors in cities. The project was carried out in collaboration with the Newcastle Urban Observatory, the largest public network of sensors providing real-time information about a city in the UK. In this project we draw on ideas of data justice, that emphasises the importance of power relations as our lives are increasingly datafied.

The researchers integrated a range of spatial data types, including fine-scale demographic data and sensor locations (e.g. air quality, noise and temperature sensors), to identify existing inequalities in sensor coverage for vulnerable populations. We also used spatial optimisation techniques –computational approaches to find the best solutions to geographic decision problems - to prioritise sensor placement based on the location of vulnerable populations. The results are embedded within a decision-support tool that supports stakeholders to evaluate inequality when placing sensors. Importantly, the design of the final decision-support tool was underpinned and shaped by findings from interviews with relevant stakeholders.

Findings from the project have also been shared via traditional academic publications. The techniques and evidence developed are now being used to shape the distribution of sensors in new networks in UK cities.

Ambient vulnerabilities: Mapping air-energy-climate interrelations in cities

Researcher: *Dr Caitlin Robinson (University of Bristol)*

From its temperature and humidity to its toxicity, our immediate ambient environment is essential to health, comfort, and wellbeing. The ambient environment is therefore integral to several important social justice questions facing cities globally including poor air quality, energy poverty, and urban or climate-related heat. The aim of this project is to define and analyse vulnerability to poor quality ambient environments, and the associated negative impacts for health and well-being. The project focuses both on cities across England and a detailed case study of the Liverpool and Bristol areas.

Using Critical Spatial Data Sciences approaches underpinned by stakeholder collaboration, the project analyses how ambient vulnerabilities accumulate in cities, and the uneven impact on different people and places. The project has several collaborating and partner organisations at different scales, including National Energy Action, Met Office, Liverpool City Council, and Merseyside Fire and Rescue Service.

Whilst currently in its infancy, it will produce a range of spatial analyses that show how different forms of ambient vulnerabilities accumulate in cities. This will include building openly available small area indicators of energy poverty, climate-related heat, and indoor air pollution.

Where else could Critical Spatial Data Science be used?

Critical Spatial Data Science has multiple application areas to a wide range of socio-spatial inequalities, beyond the examples presented here. This ranges from evaluation of the uneven distribution of COVID-19 impacts, to the spatial relationships between greenspace and health, and inequalities in access to amenities in cities. Future innovation in critical data science should take multiple forms:

- Innovative application areas. As urban populations grow, the impacts of climate change intensify, and societies become increasingly unequal, a critical approach to spatial data science will become increasingly pertinent. This should also involve reaching beyond data- rich case studies and examples which are often the focus of research (e.g. major urban conurbations in the Global North)

- Innovative methods and data. Machine learning techniques, and new forms of "Big" and non-traditional data have the potential to provide new perspectives on socio-spatial inequalities. However, this needs to be done with care. It is also important to avoid assuming that with "enough volume data [we] can speak for themselves", instead Big Data comes with its own challenges and risks. As a result, it is ever more important to think critically about data and keep in mind wider theoretical understandings of populations and cities.

- Socially impactful research. This is nothing new, however, arguably it is an area in which researchers could be more innovative. Assessments of socio-spatial inequalities should be designed for, and in collaboration with, stakeholders and communities.

Tips

- **Simplicity is key.** This is especially the case when it comes to analysis! Avoid throwing every method in the book at a particular dataset just because you can. Instead, take a step back and think ahead. Ask questions like: what do I want my finished product to achieve? How will my audience interpret the information? Is it going to be easy for them to understand the message I want to convey?

- **Don't be afraid to have a go.** It is often tricky to know where to start with computational and quantitative methods, especially if you have no previous training or experience to draw on. In fact, many open-source materials are highly accessible and assume you have no prior experiencing of coding or the methods they are introducing.

- **Collaborate with other researchers.** There is significant benefit from researchers with different disciplines and backgrounds applying spatial data science methods critically, bringing new perspectives and approaches.

Further reading

There are lots of open-source materials to help you get started analysing, visualising and modelling spatial data, including:

- Lovelace, R., Nowosad, J., and Muenchow, J. (2019) *Geocomputation* with R. Chapman and Hall/CRC. Available at: https://geocompr.robinlovelace.net/
- Rey, S., Arribas-Bel, D., and Wolf, L. (2022) *Geographic Data Science with Python.* Available at: https://geographicdata.science/book/intro.html
- Rowe, F. (2022) *Geographic Data Science for Public Policy.* Available at: https://fcorowe.github.io/udd_gds_course/

To better understand some of the underlying principles of Critical Spatial Data Science see:

- Delmelle, E. C. (2019). Toward a more socially impactful geographical analysis. *Geographical Analysis* 53.1 (2021): 148-156. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/gean.12213 (Reflections on ensuring impactful research that is better able to explain the messiness of social reality.)
- Franklin, R. (2022). Quantitative methods I: Reckoning with uncertainty. *Progress in Human Geography*, 46(2), 689-697. Available at: https://journals.sagepub.com/doi/full/10.1177/03091325211063635 (An overview of the concept of uncertainty in quantitative geography.)
- D'ignazio, C., & Klein, L. F. (2020). *Data feminism.* MIT Press. https://datafeminism.io (An introduction to new ways of thinking about (spatial) data science and ethics informed by ideas of intersectional feminism.)

To reference:

Robinson, C. (2023) 'Critical Spatial Data Science' in Rodekirchen, M., Pottinger, L. Briggs, A., Barron, A., Eseonu, T., Hall, S. and Browne, A.L. (eds.) *Methods for Change Volume 2: Impactful social science methodologies for 21st century problems.* Manchester: Aspect and The University of Manchester.

Funding:

Caitlin Robinson and her research is supported by a UKRI Future Leaders Fellowship MR/V021672/2

Methods for Change

To read about more exciting social science methods, the full range of Methods for Change 'how-to' guides can be found here.