Spatial Data Modelling

Population Science

Francisco Rowe

2025-10-19

Table of contents

Welcome

This is the website for the course: "Spatial Data Modelling for Population Science using R". The course was designed and is delivered by Professor Francisco Rowe from the Geographic Data Science Lab at the Department of Geography and Planning from the University of Liverpool, United Kingdom.

> **Note**
>
> A PDF version of this course is available for download here.

Contact

Francisco Rowe - fcorowe [at] liverpool.ac.uk - Professor in Population Data Science - Roxby Building, University of Liverpool - 74 Bedford St S, Liverpool, L69 7ZT, UK.

Part I

Introduction

Overview

This course provides an introduction to Spatial Data Modelling for Population Science using R. It has been designed to be delivered in 15 hours of teaching, split into five sessions of 3 hours each, including breaks, practical examples and questions. This course provides an intuitive and practical explanation of methods, modelling frameworks and theories for analysing spatial data, as well as demographic and social processes represented in space. Spatial data have become ubiquitous. Data containing location have become increasingly common in both volume and frequency with the development and use of new sensor, satellite technilogy and GPS-based technologies, such as smartphones, social media platforms and contactless payment systems. At the same time, social and demographic processes occur within a geographic context that conditions or triggers their evolution. Therefore, these processes must be studied explicitly, considering the spatial context in which they take place. For these reasons, techniques that enable the analysis and modelling of spatial data are essential to understand demographic and social processes to inform policy and business decision-making.

Aim

The objective of the course is to expose students to a variety of techniques for the analysis and modelling of spatial data. The course discusses and illustrates the types of spatial data and the challenges they present, introducing a range of modelling and analytical approaches that explicitly account for space. It also aims to develop programming skills in the R language. These skills are important for several reasons: programming is in high demand, it teaches logical and methodical thinking, and it allows ideas to be materialised into tangible products such as maps. The course also aims to introduce generative

artificial intelligence through the use of ChatGPT for code writing, debugging and reasoning.

Learning Outcomes

By the end of the module, students should be able to:

- identify some key sources of spatial data and resources of spatial analysis and modelling tools;
- explain the advantages of taking spatial structure into account when analysing spatial data;
- apply a range of computer-based techniques for the analysis of spatial data, including mapping, correlation, kernel density estimation, spatial interaction modelling and spatial econometrics;
- apply appropriate analytical strategies to tackle the key methodological challenges facing spatial analysis – spatial autocorrelation, heterogeneity, and ecological fallacy; and,
- select appropriate analytical tools for analysing specific spatial data sets to address emerging social issues facing the society.

Teaching Approach

The course combines lectures with a strong practical component and independent work. Each session will cover one of the following components:

- Types of spatial data.
- Manipulation and visualisation of spatial data.
- Analysis of location data.

- Analysis of spatial relationship data.
- Models that capture spatial correlation.

The course: - Introduces these techniques in an intuitive and applied manner rather than through theoretical mathematical or statistical derivations.

- Focuses on explaining when and why a technique is appropriate, followed by its application.
- Uses several practical examples based on real data and research questions.
- Is built around the R programming language and uses structured computational notebooks compiled in a book, including code with explanations that can be reused with other datasets.

Environment

To reproduce the code in the book, you need the following software packages:

- R version 4.5.1 (2025-06-13)

- RStudio Version 2025.09.1+401 (2025.09.1+401)
- Quarto 1.8.25
- the list of libraries in the next section

To check your version of:

- R and libraries run `sessionInfo()`
- RStudio click `help` on the menu bar and then `About`
- Quarto check the `version` file in the quarto folder on your computer.

To install and update:

- R, download the appropriate version from The Comprehensive R Archive Network (CRAN)
- RStudio, download the appropriate version from Posit
- Quarto, download the appropriate version from the Quarto website

Dependency list

The list of libraries used in this book is provided below:

- `arm`
- `car`
- `corrplot`
- `devtools`
- `FRK`
- `gghighlight`

- ggplot2
- ggmap
- GISTools
- gridExtra
- gstat
- hexbin
- jtools
- kableExtra
- knitr
- lme4
- lmtest
- lubridate
- MASS
- merTools
- plyr
- RColorBrewer
- rgdal
- sf
- sjPlot
- sp
- spgwr
- spatialreg
- spacetime
- stargazer
- tidyverse
- tmap
- tufte
- viridis
- basemapR

Copy, paste and run the code below in your console. Ensure all packages are installed on your computer.

```
# package names
packages <- c(
    "arm",
    "car",
    "corrplot",
    "devtools",
    "FRK",
    "gghighlight",
```

```
    "ggplot2",
    "ggmap",
    "gridExtra",
    "gstat",
    "hexbin",
    "jtools",
    "kableExtra",
    "knitr",
    "lme4",
    "lmtest",
    "lubridate",
    "MASS",
    "merTools",
    "plyr",
    "RColorBrewer",
    "sf",
    "sjPlot",
    "sp",
    "spgwr",
    "spatialreg",
    "spacetime",
    "stargazer",
    "tidyverse",
    "tmap",
    "tufte",
    "viridis"
)

# install packages not yet installed
installed_packages <- packages %in% rownames(installed.packages())
if (any(installed_packages == FALSE)) {
  install.packages(packages[!installed_packages])
}

# packages loading
invisible(lapply(packages, library, character.only = TRUE))
```


> **Note**
>
> To install the library `basemapR`,
>
> you need to install from source by running:
> `library(devtools)`
>
> `install_github('Chrisjb/basemapR')`

Assessment

The course assessment will consist of two components.

Assignment 1: Spatial Data Infographic (40%)

Description
The first assignment involves the creation of an *infographic* that visually communicates a spatial data story. The goal is to synthesise spatial data, analytical results and visual design into a clear, evidence-based narrative that can fit within a single PowerPoint slide. The infographic must demonstrate understanding of spatial concepts, analytical methods and effective data communication principles.

Instructions

1. **Topic** – Select a spatially explicit social, demographic or environmental issue (e.g., population change, internal migration, population decline, population ageing).
2. **Data** – Use the geospatial datasets on internal migration or population described in the Chapter dataset, or a comparable open dataset. Clearly cite all sources.
3. **Content** – Present a concise spatial narrative integrating:
    - A key research question or problem.
    - Relevant spatial analysis or modelling results (maps, graphs and / or summary statistics).
    - Interpretation of spatial patterns and implications.
    - References or data sources.

4. **Format** –
    - One PowerPoint slide (16:9 ratio).
    - All text must be legible when projected on screen.
    - Include a clear title and author name.
    - Visuals may include maps, charts, icons or minimal text boxes.
    - Use consistent colour-blind friendly palette and typographic hierarchy.
5. **Submission** – Upload the PowerPoint (.pptx) file and a PDF version.

Evaluation Criteria (100 points total)

Criterion	Description	Weight
Relevance and clarity of question	Defines a meaningful spatial issue and communicates the core message clearly.	20 %
Analytical rigor	Demonstrates appropriate use of spatial data and analysis; accuracy of methods and results.	25 %
Interpretation	Interprets spatial patterns logically and links findings to broader demographic or social processes.	20 %

Criterion	Description	Weight
Design and visual communication	Effective visual hierarchy, readability, layout, and use of colour; slide visually balanced and professional.	25 %
Data and source transparency	Correct data citation and ethical use of sources.	10 %

Purpose
This assignment develops skills in communicating spatial evidence succinctly and visually. This is an essential competence for academic, policy and professional audiences.

Assignment 2: Computational Essay (60%)

Description
The second assignment consists of a *computational essay* that integrates narrative, code and visualisation to explore a spatially explicit research question. A computational essay is a document in which the *story, analysis and computation are woven together*: The text explains the logic of the analysis, the code performs it, and the figures or tables present the results. This format reflects the way spatial data scientists communicate reproducible, evidence-based insights.

Objectives
The computational essay should:

1. Demonstrate the ability to apply spatial analysis or modelling techniques learned in the course.
2. Integrate theory, data and computation into a coherent narrative.
3. Present clear, reproducible and interpretable results using R.

4. Communicate findings effectively through maps, graphics and concise text.

Instructions

1. **Topic Selection**
   Choose a spatial problem relevant to the social, demographic, or environmental domains (e.g., population change, internal migration, population decline, population ageing). You may use the geospatial datasets on internal migration or population described in the Chapter dataset, or a comparable open dataset.
2. **Structure**
   - **Introduction (300–400 words):** State the research question and explain its spatial relevance.
   - **Data and Methods (400–600 words):** Describe the data sources, spatial structure and analytical or modelling techniques used. Include annotated R code chunks performing the analysis.
   - **Results and Discussion (600–800 words):** Present and interpret results using visualisations and tables. Discuss implications, limitations and potential extensions.
   - **Conclusion (200–300 words):** Summarise key findings and their significance.
   - **References (not included in the word count):** Include all data and bibliographic sources using a consistent citation style.
3. **Format and Submission**
   - Maximum **2,000 words** (excluding code, references and captions).
   - Include **R code** directly within the essay in Quarto format.

- Include up to **five figures** (a figure may contain multiple maps or plots if they form a single integrated result).
- Include up to **one table**.
- Submit both the rendered **PDF or HTML output** and the corresponding **.qmd source file**. Ensure the active the option `self-contained` in the YALM of the .qmd file. This ensures all figures are embedded within the final HTML file.
- All analyses must be fully reproducible using the submitted code and data. Share data via a link to the platform where they are stored such as Google Drive, Dropbox or GitHub.

Evaluation Criteria (100 points total)

Criterion	Description	Weight
Relevance and research question	Clearly defined, spatially meaningful question grounded in theory or application.	15%
Analytical rigor and reproducibility	Correct and appropriate use of spatial data and methods; code runs and reproduces results.	25%
Integration of code and narrative	Narrative, computation, and visuals are well-linked; code is readable and annotated.	20%

Criterion	Description	Weight
Interpretation and critical insight	Results are interpreted thoughtfully, demonstrating understanding of spatial processes.	20%
Presentation and communication	Clarity of writing, structure, figure quality, and adherence to limits.	15%
Transparency and referencing	Complete and accurate citation of data and literature sources.	5%

Purpose

This assignment develops core competencies in *computational thinking*, *reproducible spatial analysis*, and *scientific communication*. It trains students to present spatial data analyses as coherent narratives supported by transparent, executable code. This is a skill essential for modern population and spatial data science.

Part II

Content

1 Spatial Data

2 Spatial data manipulation and visualisation

3 Point data analysis

4 Spatial interaction modelling

5 Spatial econometrics

Part III

Supplement

Datasets

References