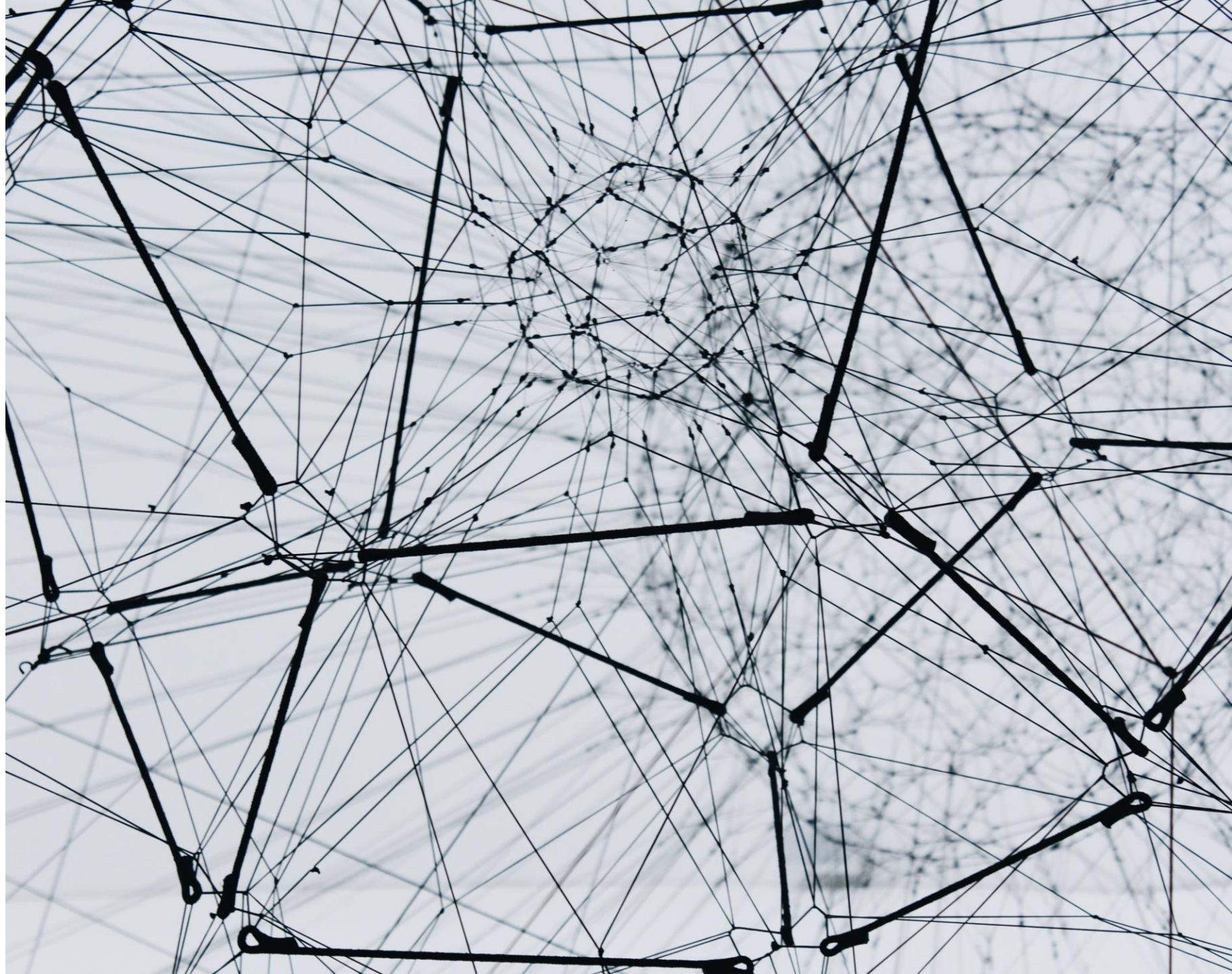


Spatial Analysis

Introductory Session

Orlando Sabogal-Cardona
PhD researcher
University College London UCL

Who
What
Why
How



What is data?

What is statistics?

Variables: characteristics of the individuals

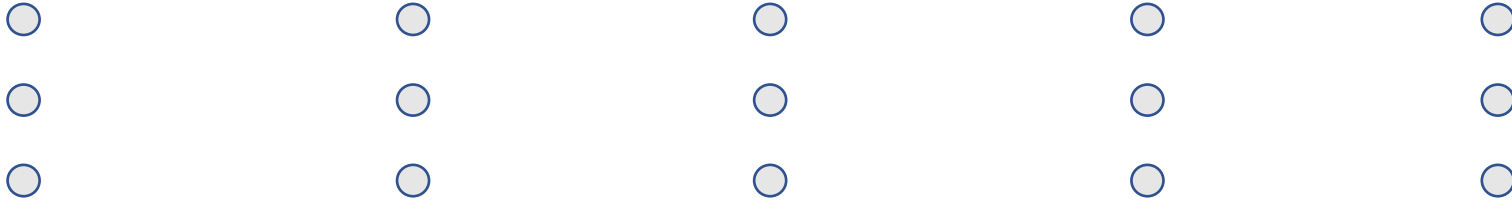
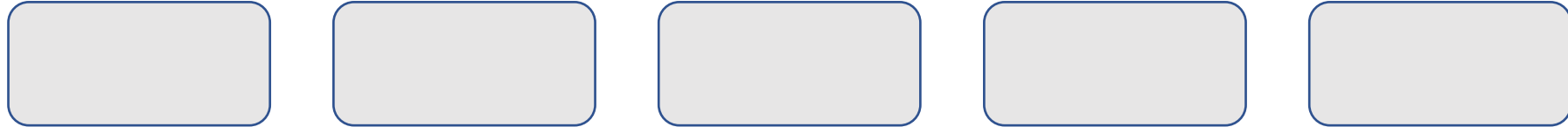
Individual 1



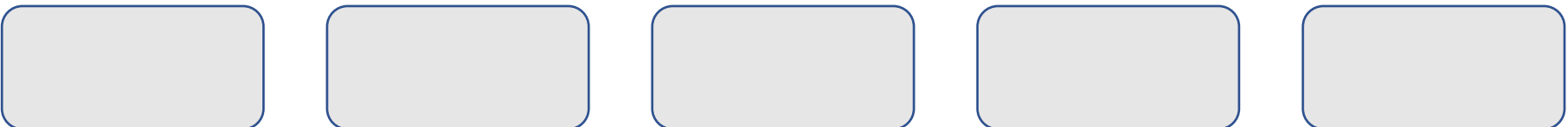
Individual 2



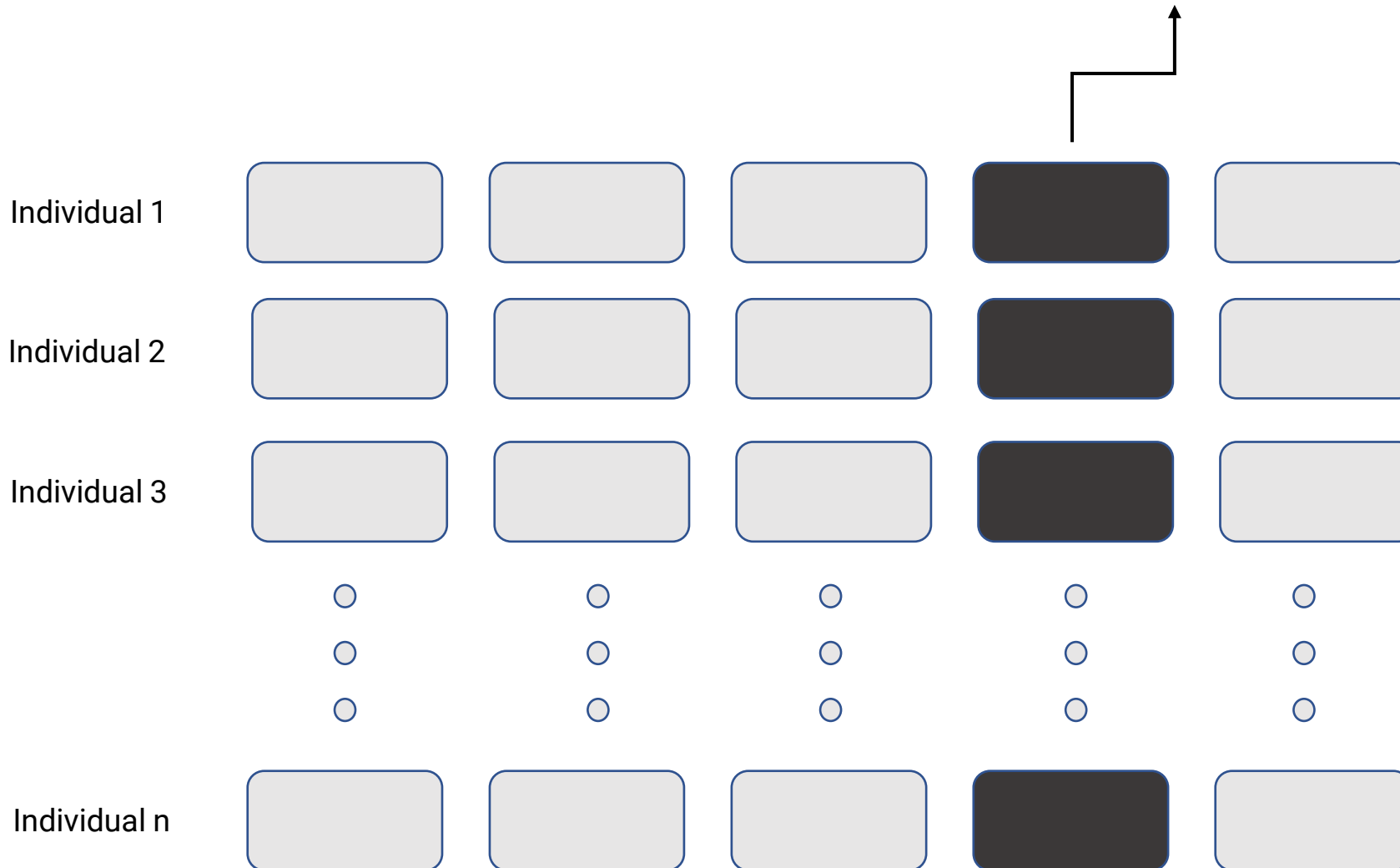
Individual 3



Individual n



One (or more) variables can
refer to a categorical variable



Types:

- Binary
- Nominal
- Ordered

Also:

- Choices
- Rankings
- Hidden (latent) groups
- Treatments

Individuals can have a “spatial” dimension

Individual 1



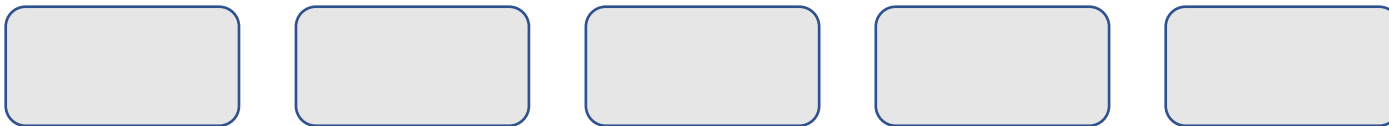
Individual 2



Individual 3



Individual n



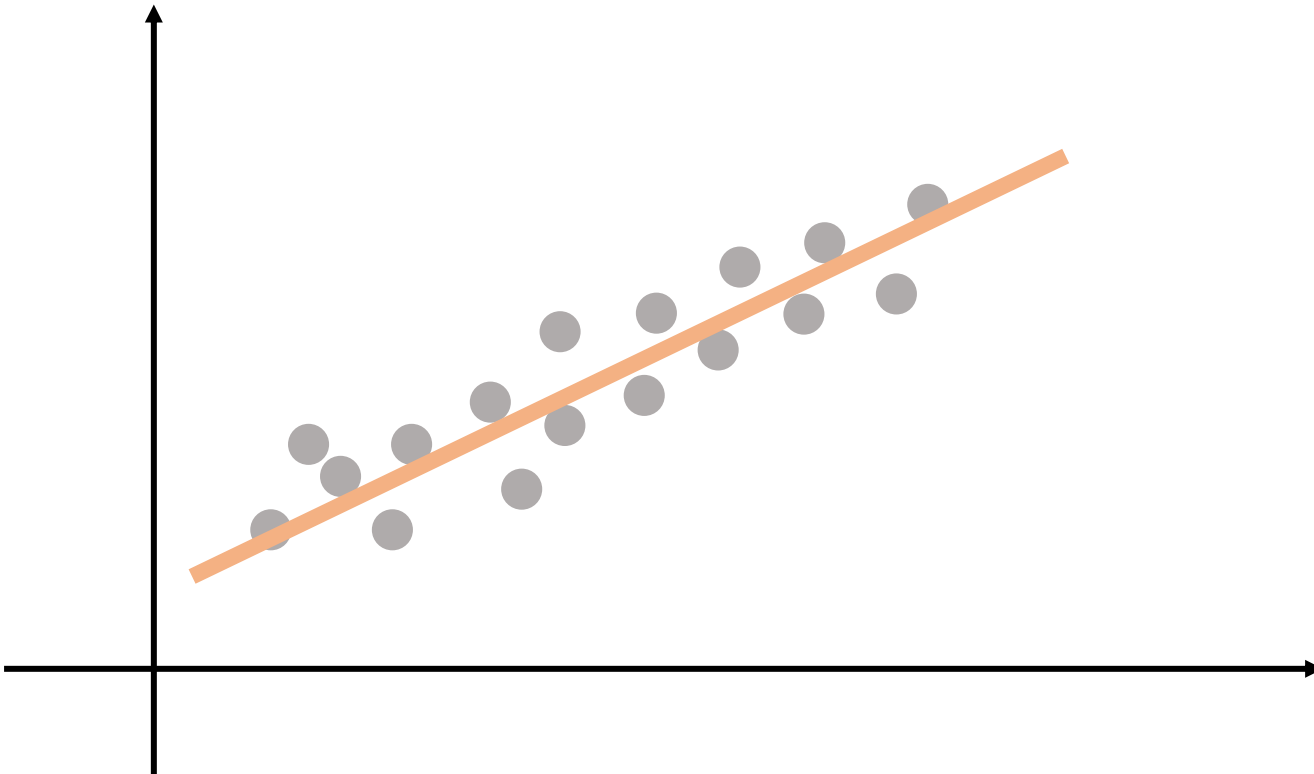
They can be cities,
neighborhoods, TAZ,
counties.
They can also be points
(more on this later).

In short:
geolocation + database

○ You regularly want to:

- Understand a social, physical or natural phenomena
- Describe a process
- Know if two variables are related to each other (association/correlation)
- Test hypothesis
- Assess causality: what is causing what?
- To get insights
- Prediction

A note on linear regression

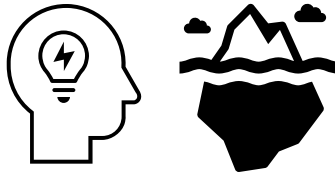


- The line of “best fit”
- “Explain” Y given X
- An abstraction of how the real-world works

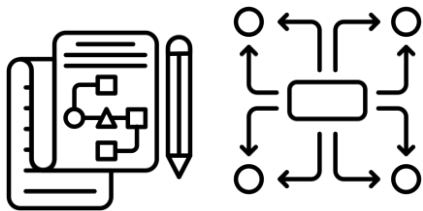
An easy way to think about LR:
You are trying to figure out the ingredients in your food.

Master LR and will conquer the world (of statistics)

Main goals of this course:



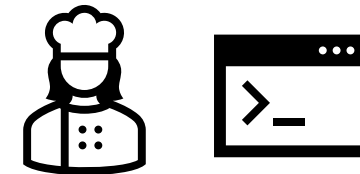
I want you to learn to produce knowledge and insights in the presence of geo-data. To think as a data scientist (?)



Understand the logic behind each tool. Familiarity with key concepts.

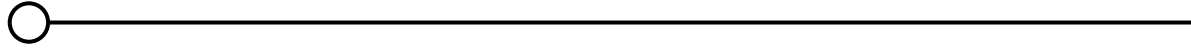


To know what is what. Identify problems (questions) and tools to solve them



I promise to avoid math as much as possible. Instead, we will focus on applications. Hands-on with R

To achieve this you must:



Understand (conceptually) in simple terms
what you are doing

Understand the math behind it

Actually do it

Outline of the course

- A lot of R
- Reproducible research/analysis: Rstudio
- The tidyverse ecosystem
- sf + tmap + leaflet
- Hypothesis testing
- Central limit theorem
- Linear regression
- When and why linear regression “fails”?
- Complete Spatial Randomness, spatial dependence, spatial heterogeneity
- Spatial autocorrelation – Moran's' I
- LISA (spatial clusters)
- Spatial Regimes
- Spatial Lag model (SAR)
- Spatial Lagged X
- Spatial Error Model
- Spatial Durbin Model
- Spatial Durbin Error
- Geographically Weighted Regression GWR
- Spatial Panels
- Point Patter Analysis PPT (brief intro)
- Kriging Models (brief intro)
- Categorical models (brief intro)
- Multilevel models (brief intro)

Geographic data:

Raster

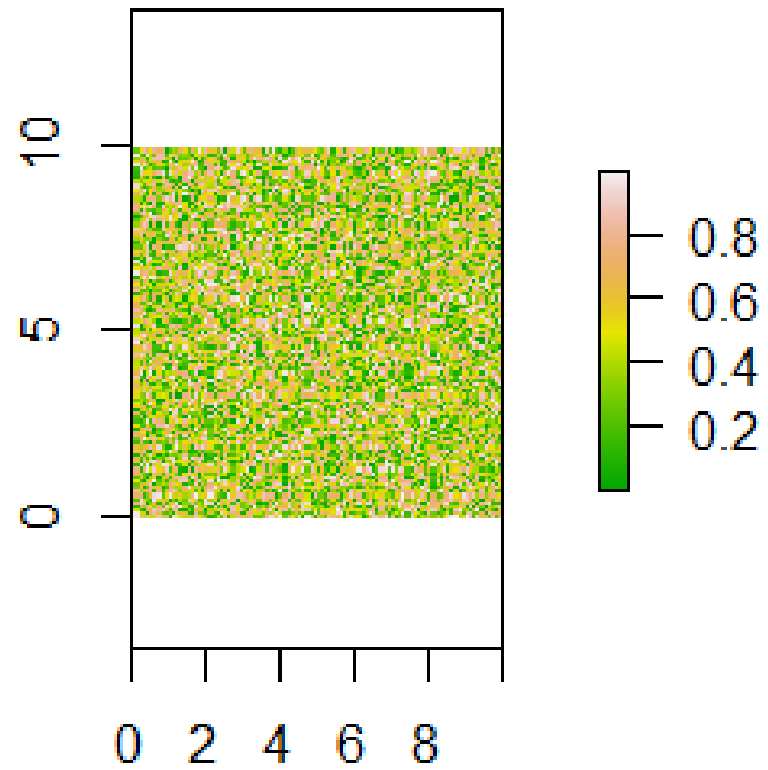
Lines

Points

Areas

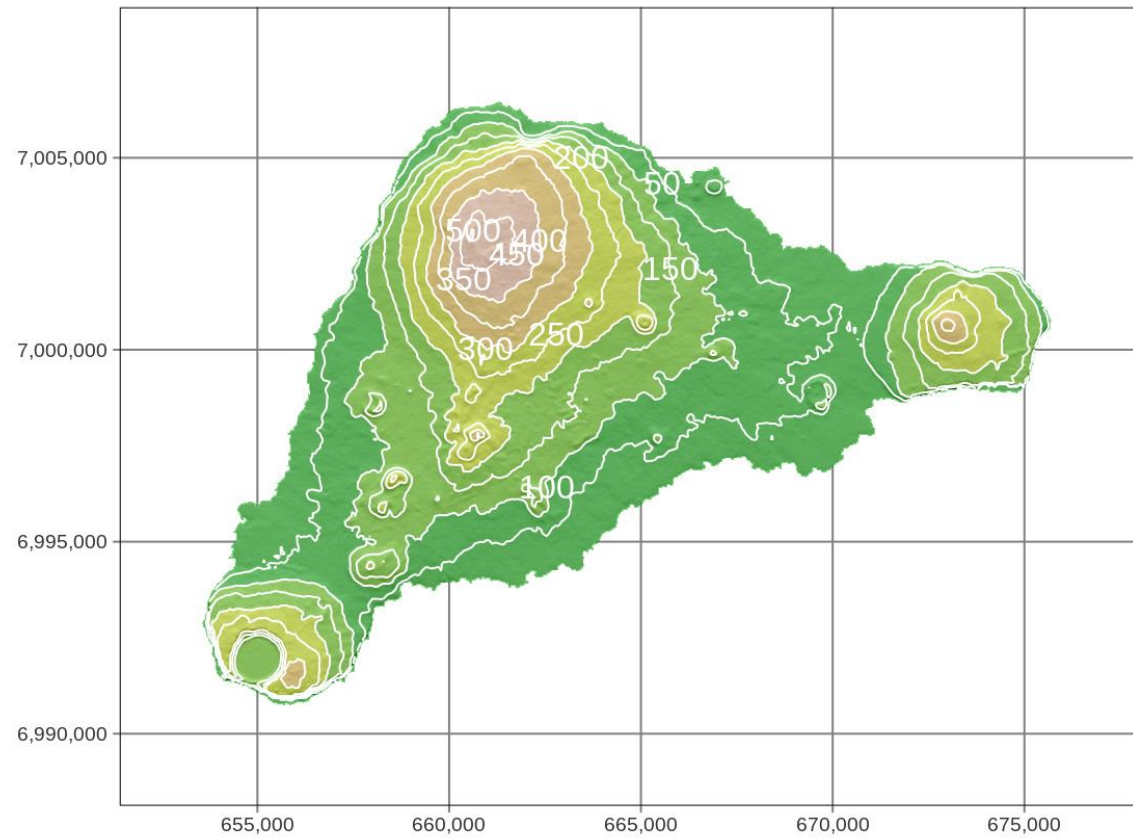
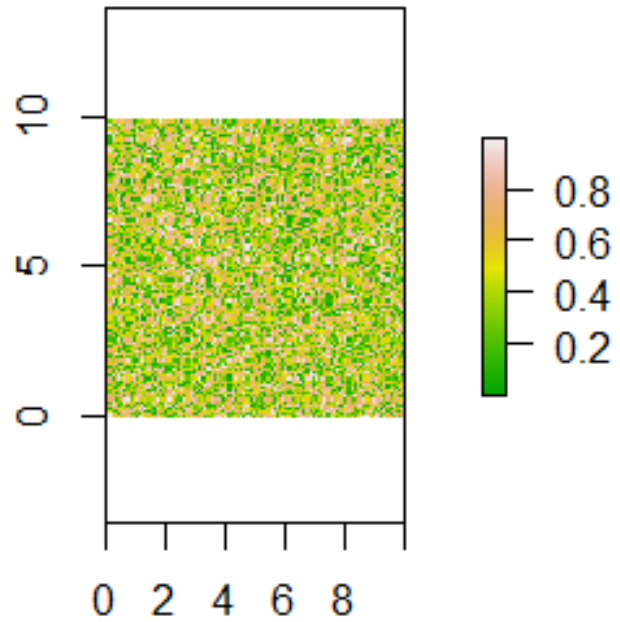
Raster

Example Raster Map



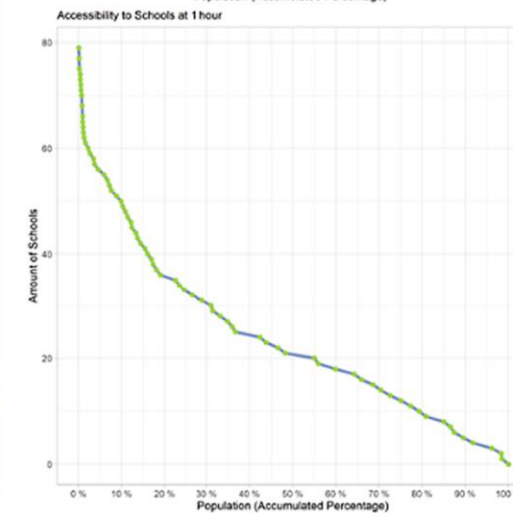
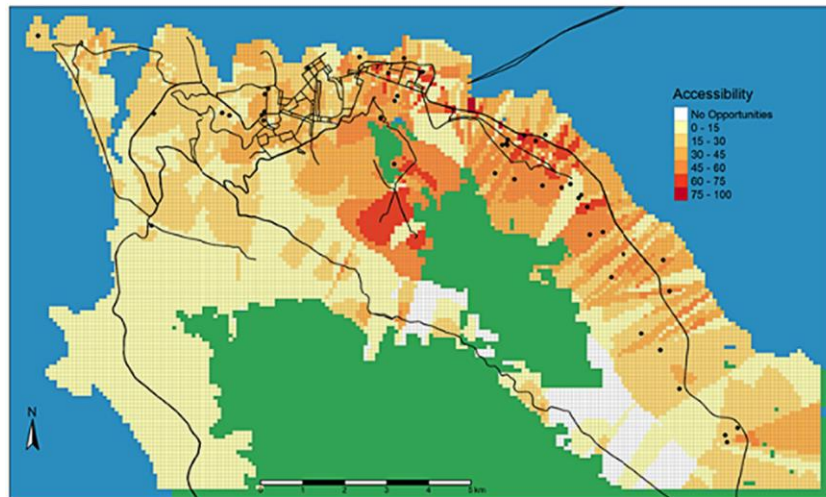
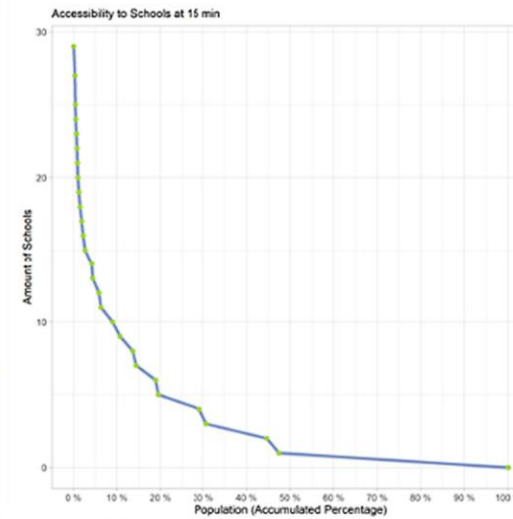
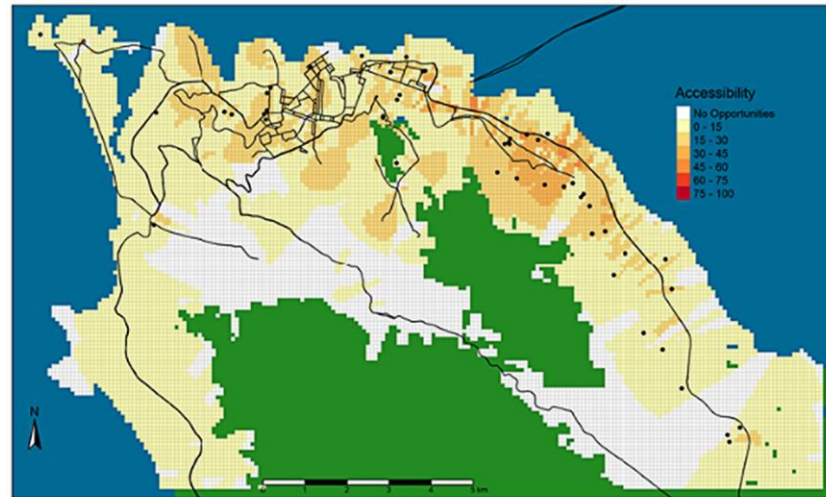
Raster

Example Raster Map



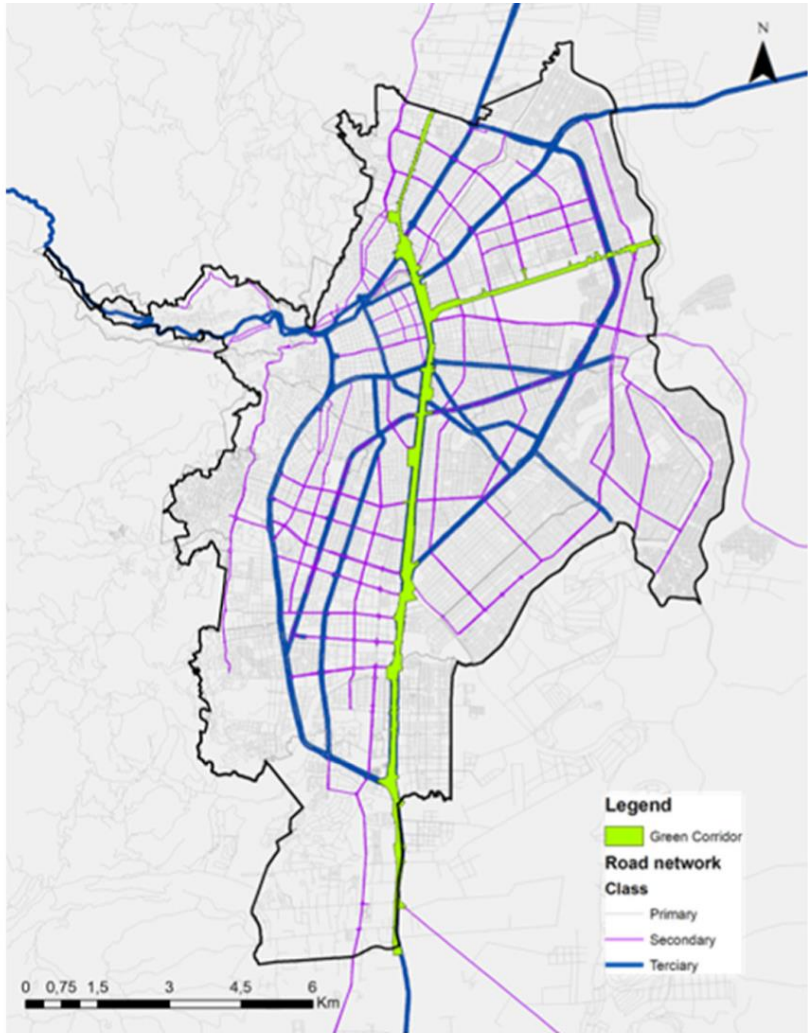
Source: *Martijn Tennekes, and Jakub Nowosad (Elegant and informative maps with tmap, 2021)*
See: <https://r-tmap.github.io/tmap-book/layers.html>

Raster

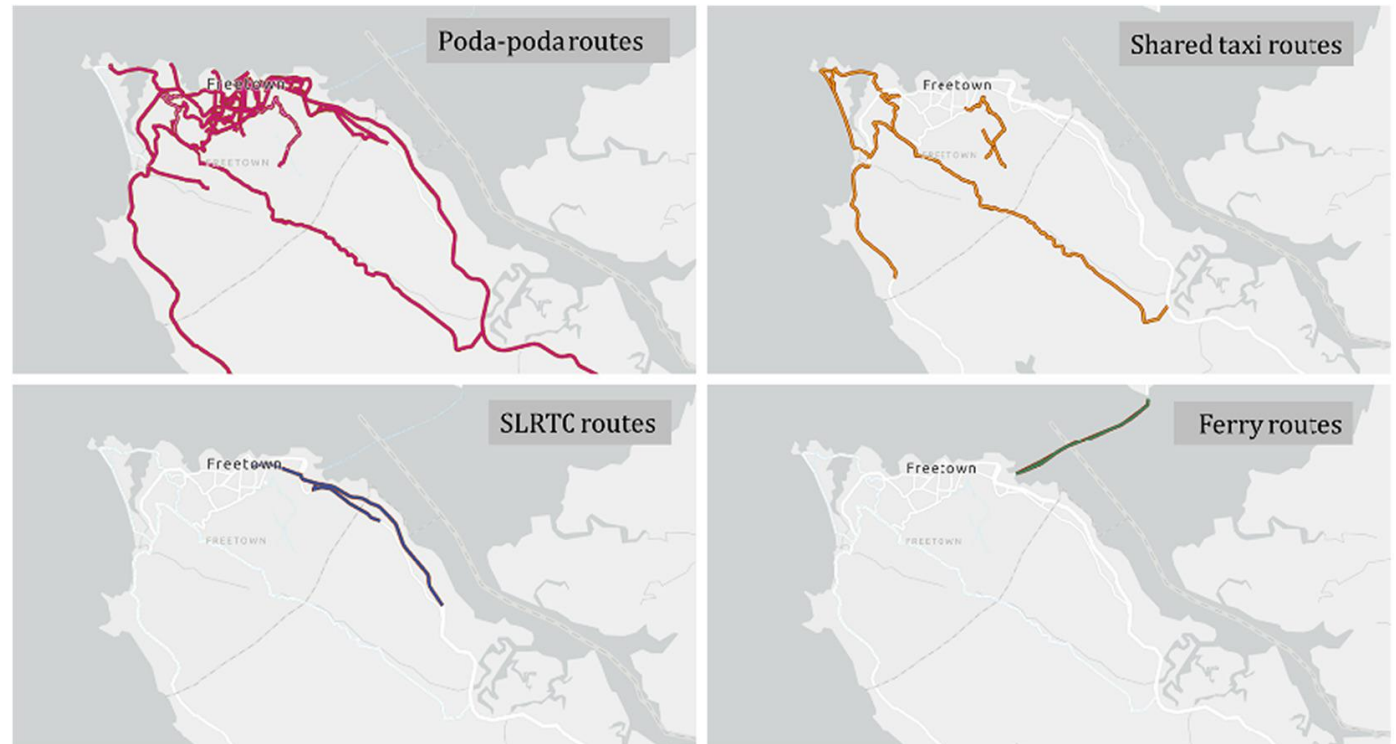


Source: Oviedo et al., (Accessibility and sustainable mobility transitions in Africa: Insights from Freetown 2022)
See: <https://www.sciencedirect.com/science/article/pii/S0966692322001879>

Lines



Source: Oviedo, Sabogal-Cardona, Villamizar-Duarte, Chong (Perceived liveability, transport, and mental health: A story of overlying inequalities, 2022)
See: <https://www.sciencedirect.com/science/article/pii/S2214140522001852>



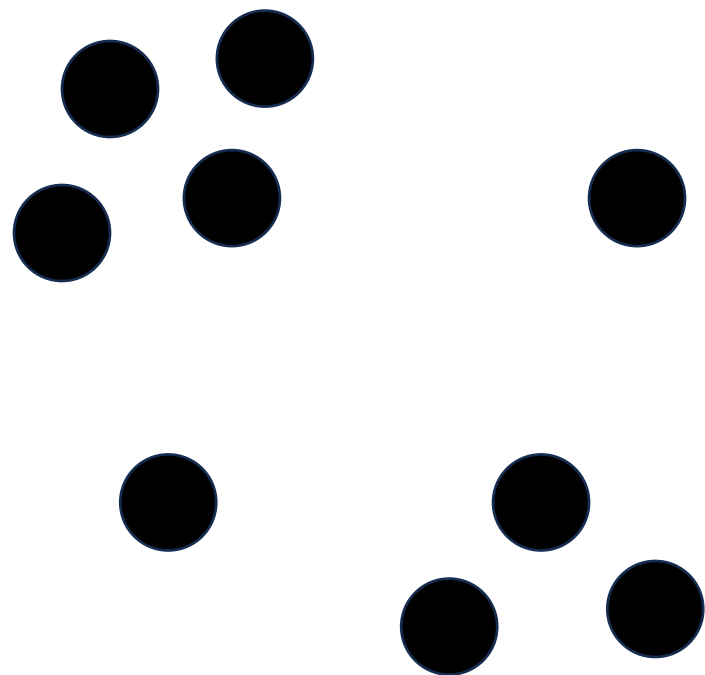
Source: Oviedo et al., (Accessibility and sustainable mobility transitions in Africa: Insights from Freetown 2022)
See: <https://www.sciencedirect.com/science/article/pii/S0966692322001879>

Points

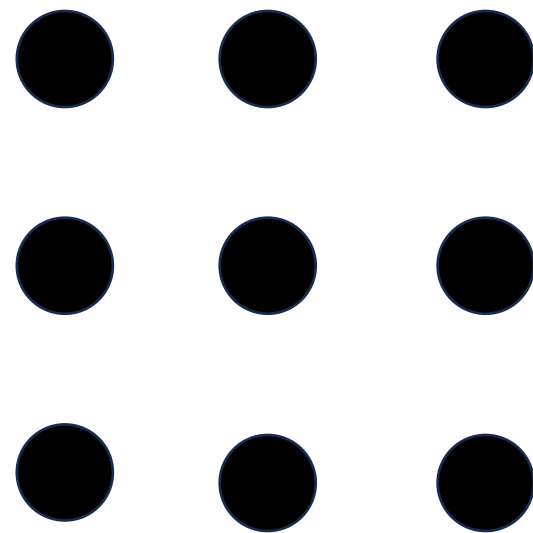


Points

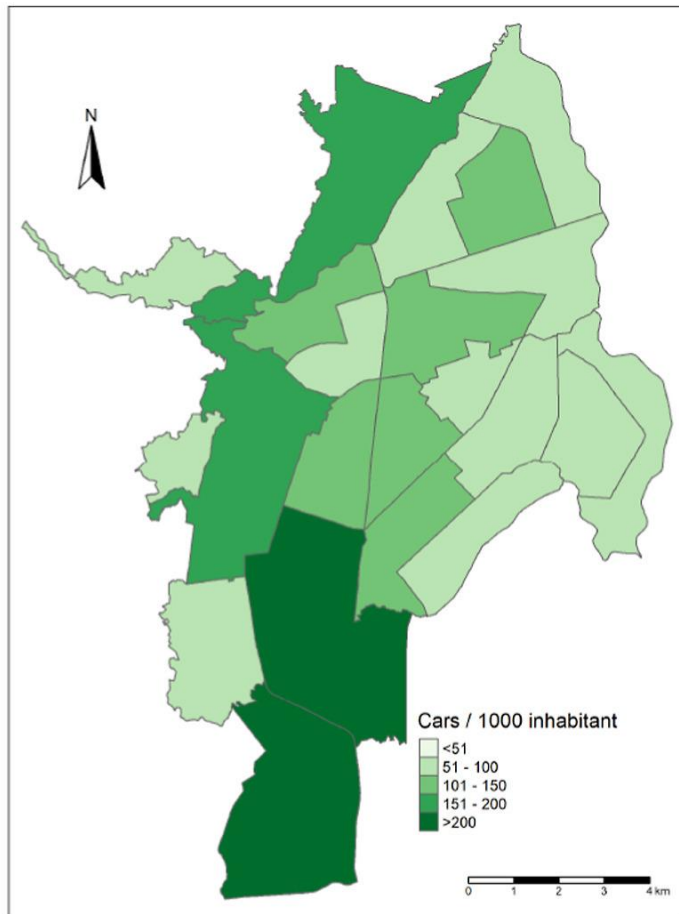
Events (random
process)



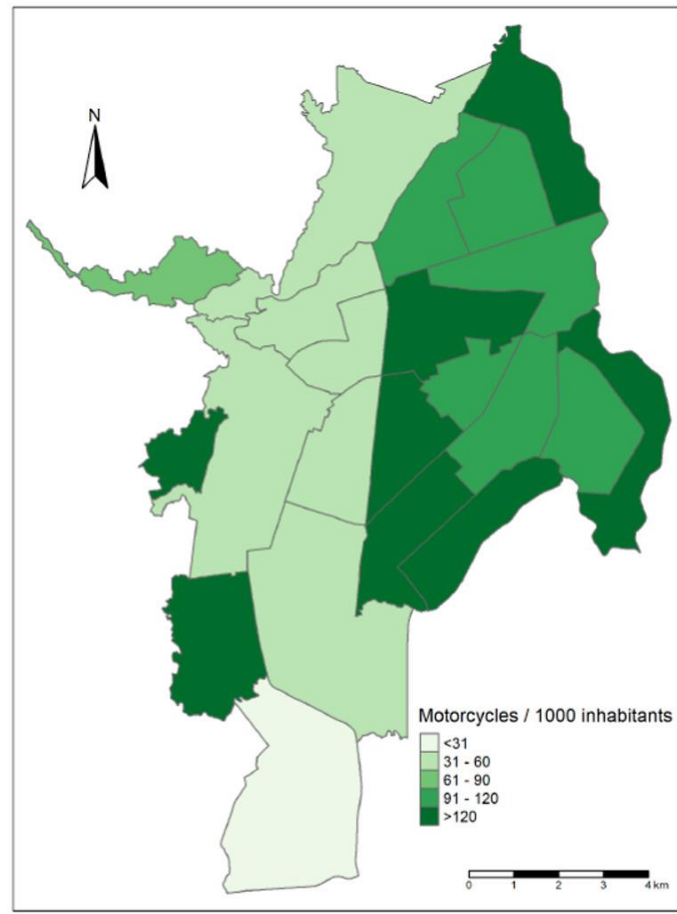
Fixed measurement
locations



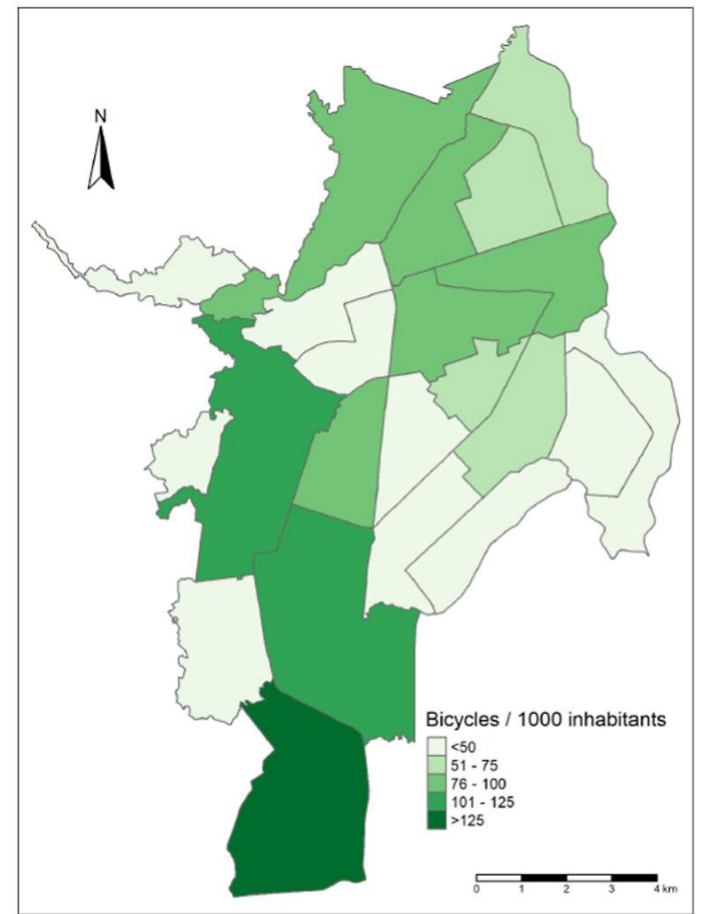
Areas



Panel A: Cali Motorization rates (car)

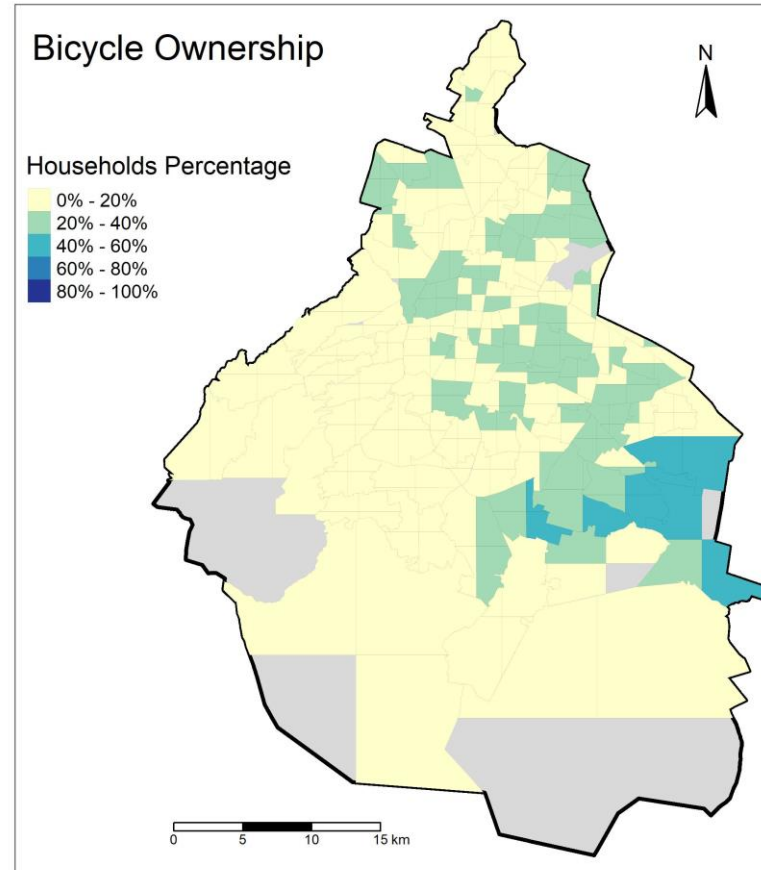
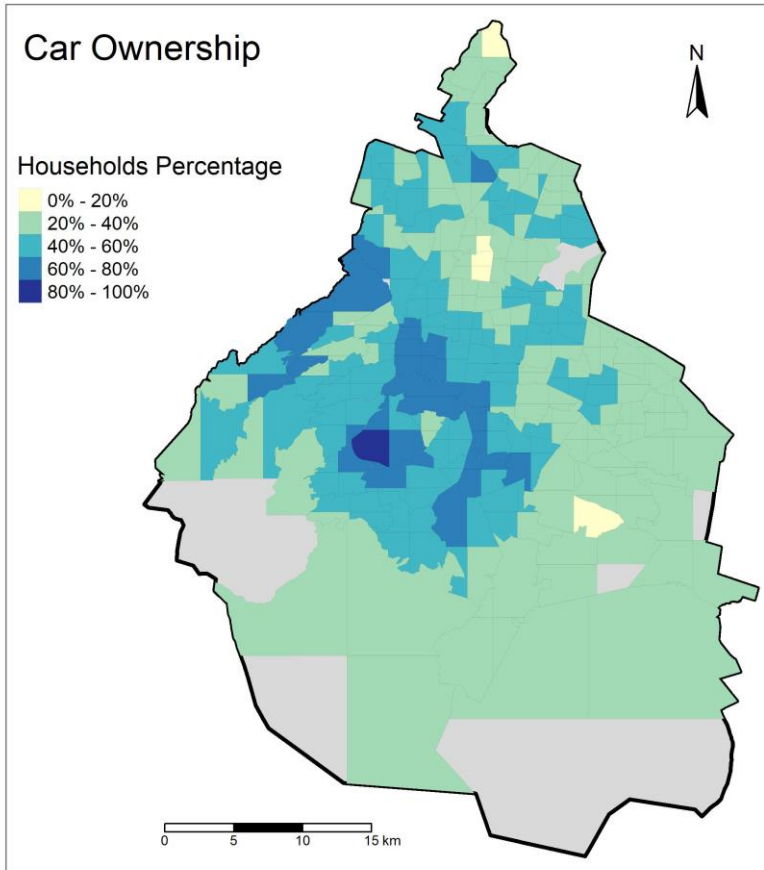


Panel B: Cali Motorization rates (motorcycle)



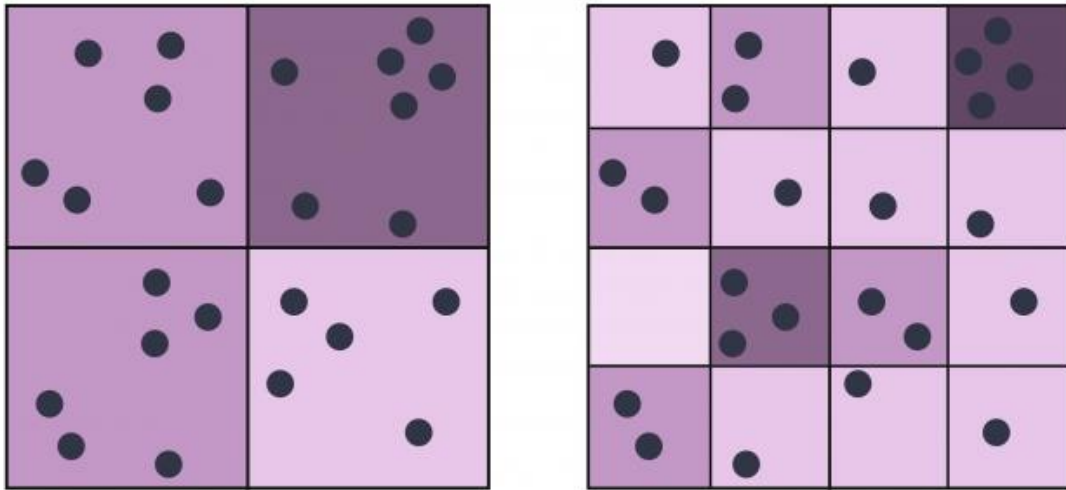
Panel C: Cali bicycle ownership rates

Areas

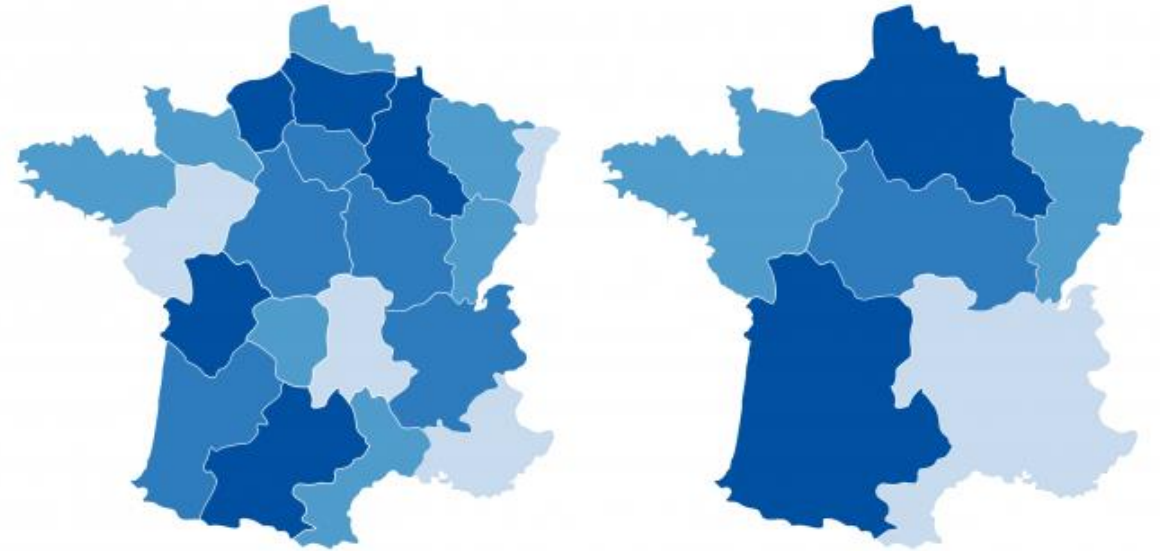


In the context of areas concepts such as Moran's I, spatial regression, and GWR take place

Modifiable Areal Unit Problem



Scale effect



Zonal effect

Modifiable Areal Unit Problem

- Geographic data can be represented at different spatial resolutions or levels of aggregation
 - An analysis can be affected by the arbitrary delineation or aggregation of spatial units or zones

What is data analysis?

- Make sense of data
- Transform, visualize, and model
- **Tell a coherent, well founded, evidence-based story:**
 - The story should have at least one clear take away message.
 - Think about communication!
- Apply statistical methods:
 - What method? Do I have the right data?
 - Are there any assumption in my methods
 - This is a complicated world, yet...



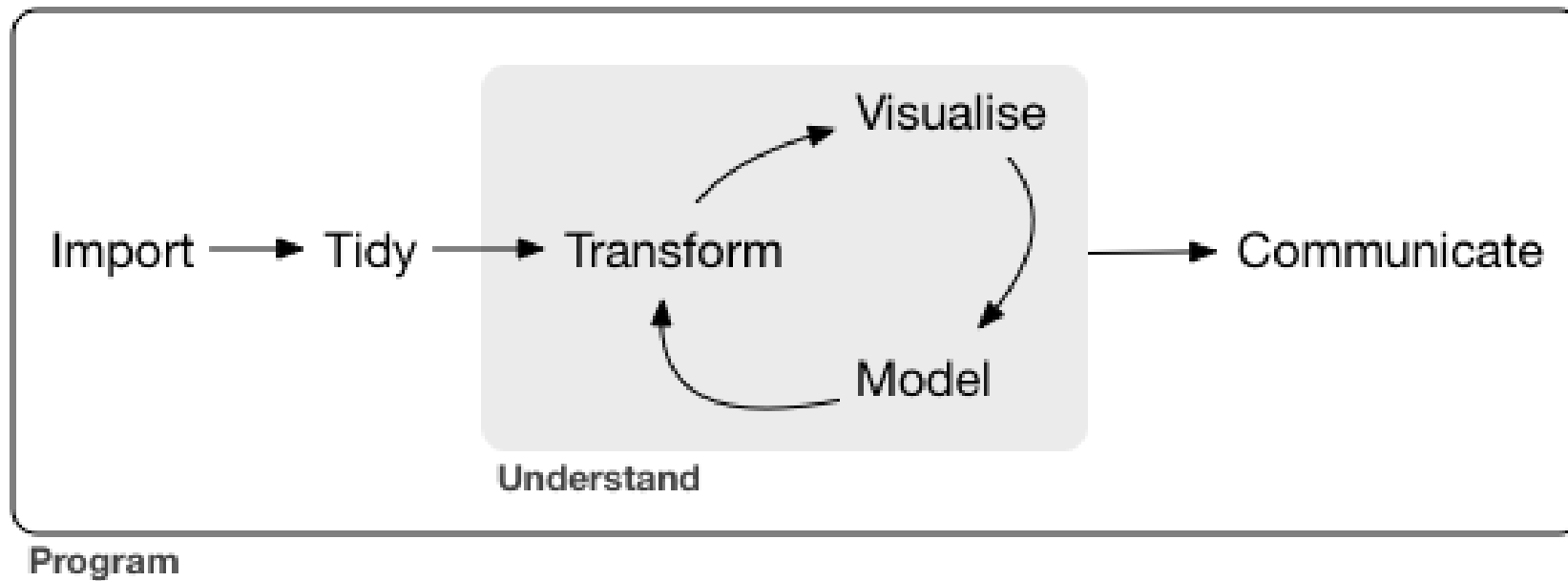
Data science

All the new family: machine learning, data mining, cognitive knowledge (?), business intelligence, analytics, deep learning, Artificial Intelligence.

Data science:

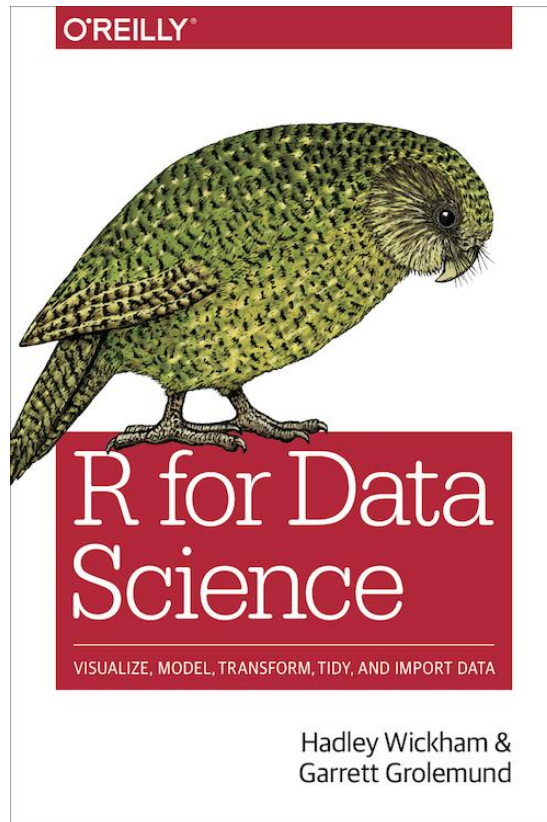
- Computational capabilities +
- Strong statistical background +
- Domain knowledge +
- Communication skills

**It should be about
knowledge production
informed by data**

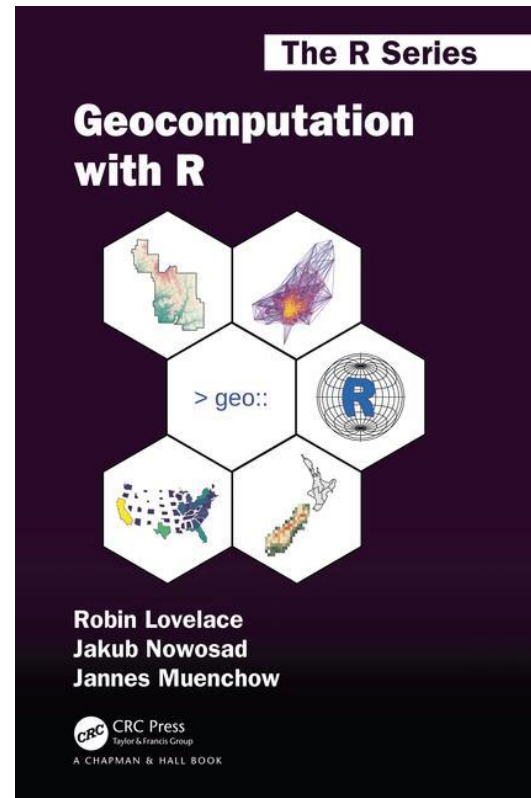


Wickham and Grolemund (2017)

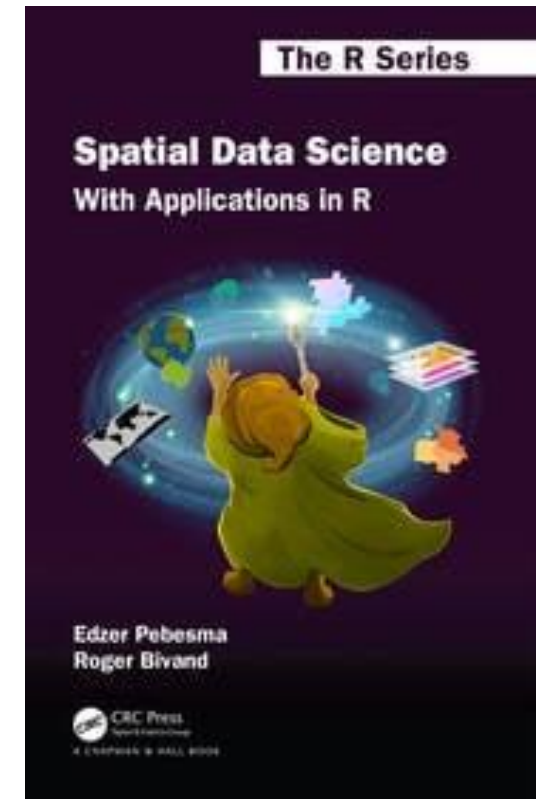
Recommended books (1/3):



Freely available at:
<https://r4ds.had.co.nz/>



Freely available at:
<https://bookdown.org/robinlovelace/geocompr/>



Freely available at:
<https://r-spatial.org/book/>

Recommended books (2/3):

- Handbook of spatial analysis (INSEE, 2018): <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/insee-estat-spatial-ana#:~:text=Handbook%20of%20Spatial%20Analysis%20This%20handbook%20of%20geospatial,been%20compiled%20by%20INSEE%20with%20funding%20from%20EUROSTAT>.
- Elegant and informative maps with tmap (Martijn Tennekes, Jakub Nowosad, 2021): <https://r-tmap.github.io/tmap-book/>
- Spatial Modelling for Data Scientists (Francisco Rowe, Dani Arribas-Bel, 2023): <https://gdsl-ul.github.io/san/>
 - Crime Mapping in R (Reka Solymosi and Juanjo Medina, 2023): https://maczokni.github.io/crime_mapping_textbook/

Recommended books (3/3):

- Handbook of Spatial Analysis in the Social Sciences (edited by Sergio J. Rey and Rachel S. Franklin, 2022)
- Understanding Crime: Analyzing the Geography of Crime (Spencer Chainey, 2020)
- Spatial Analysis Methods and Practice Describe – Explore – Explain through GIS (George Grekousis, 2020):
- Spatial Regression Models for the Social Sciences (Guangqing Chi and Jun Zhu, 2019)
- Spatial Analysis Using Big Data Methods and Urban Applications (Yoshiki Yamagata and Hajime Seya, 2019)
 - Spatial Econometrics (Harry Kelejian and Gianfranco Piras, 2017)
- Spatial Econometrics From Cross-Sectional Data to Spatial Panels (J. Paul Elhorst, 2015)
- Handbook of Quantitative Criminology (Alex R. Piquero and David Weisburd, 2010)

Thank you

Orlando Sabogal-Cardona
PhD researcher
University College London UCL