Geographic Data Science

Spatial Data

Elisabetta Pietrostefani & Carmen Cabrera-Arnau

Data, data, data

- Everything produces data
- New dense quantitative representations of the world
- But data is not useful, insights are

Data Science

“gathering data messaging it into a tractable form, making it tell its story and presenting that story to others”

Loukides (2011) What is Data Science?

Spatial Data Science

- A lot of new data is spatial data
- Spatial is special
- We don’t want to reinvent the GIS wheel
- How do we bring both world together?

Good old spatial data (+)

Good old spatial data (+)

Traditionally, datasets used in social sciences are - Collected for the purpose – carefully designed - Detailed and informative ("rich profile and portraits of the country") - High quality

Good old spatial data (-)

But also - Massive enterprises - very costly - Coarse in resolution (to preserve privacy they need to be aggregated) - Slow - the more detailed, the less frequent they are available

Examples

- Decennial census (census geographies)
- Longitudinal surveys
- Custom collected surveys, interviews etc.
- Economic or well-being indicators

New Forms of spatial data

New Forms of spatial data

Tied into the geo-data revolution

- Accidental : created for different purposes but available for analysis as a side effect
- Very diverse in nature: resolution and quality but, potentially much more detailed in both space and time

We will look at this more in a few weeks!

Lazer & Radford (2017)

- Digital life: digital actions (Twitter, Facebook, WikiPedia…)
- Digital traces: record of digital actions (CDRs, metadata…)
- Digitalised life: nonintrinsically digital life in digital form (Government records, web…)

Arribas-Bel (2014)

Three levels, based on how they originate:

- Bottom up: “Citizens as sensors”
- Intermediate: Digital businesses/businesses going digital
- Top down: Open Government Data

Opportunities (Lazer & Radford, 2017)

Massive, passive Nowcasting Data on social systems Natural and field experiments (“always-on” observatory of human behaviour) Making big data small

Challenges (Arribas-Bel, 2014)

Bias Technical barriers Methodological “mismatch”

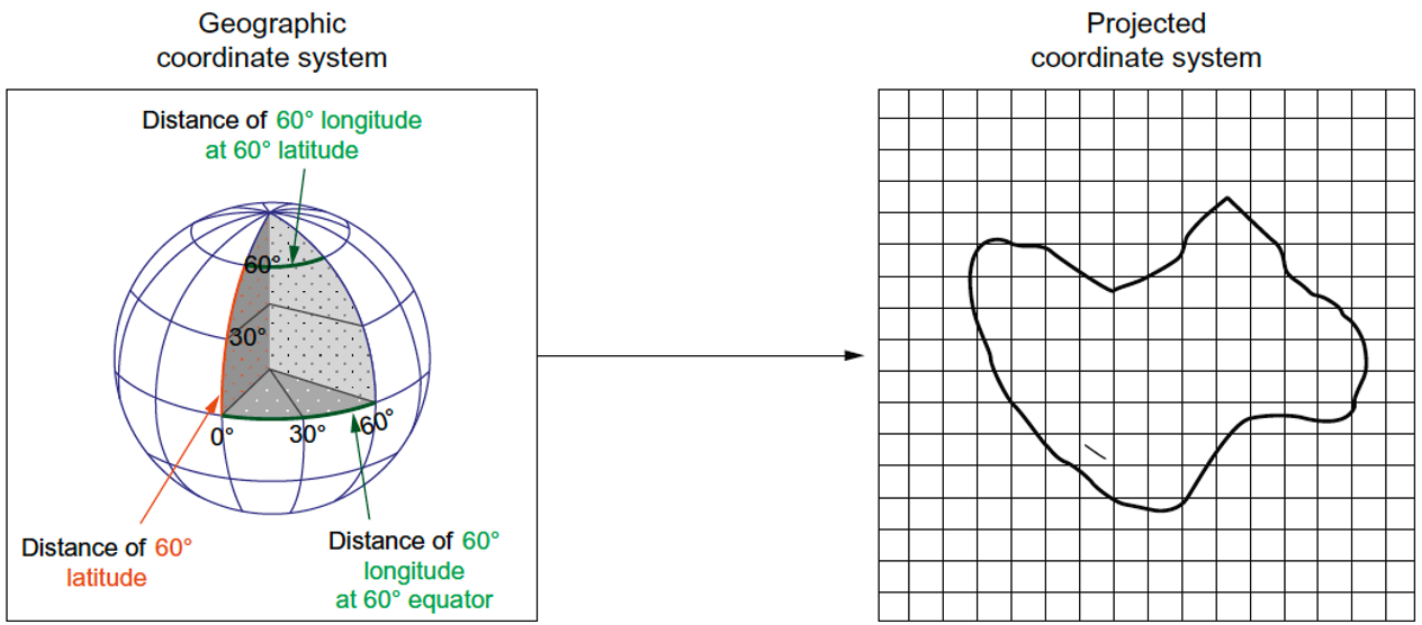
Analysing maps – identifying patterns

Creating a map (or several maps of related phenomena) enables Pattern identification Better understandings of correlations (and possible causations)

Linking Spatial Info

Layers containing different features are linked together by common positional information, which references these features in space across layers

Linking Spatial Info



Questions