

# Geographic Data Science – Lecture III

## Spatial Data

Dani Arribas-Bel

# “Day 1”

- Introduced the (geo-)data revolution
  - What is it?
  - Why now?
- The *need* of (geo-)data science to make sense of it all

# Today

- Types of (geo-)data: refresher
- Traditional and new sources of spatial data
- New ways for traditional approaches

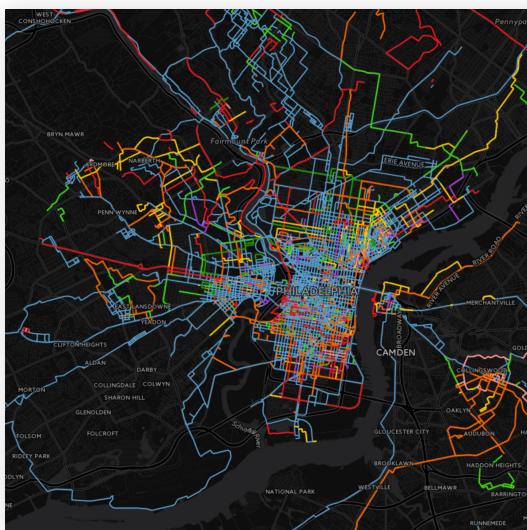
# Representing the World Digitally

# GIS Data Models

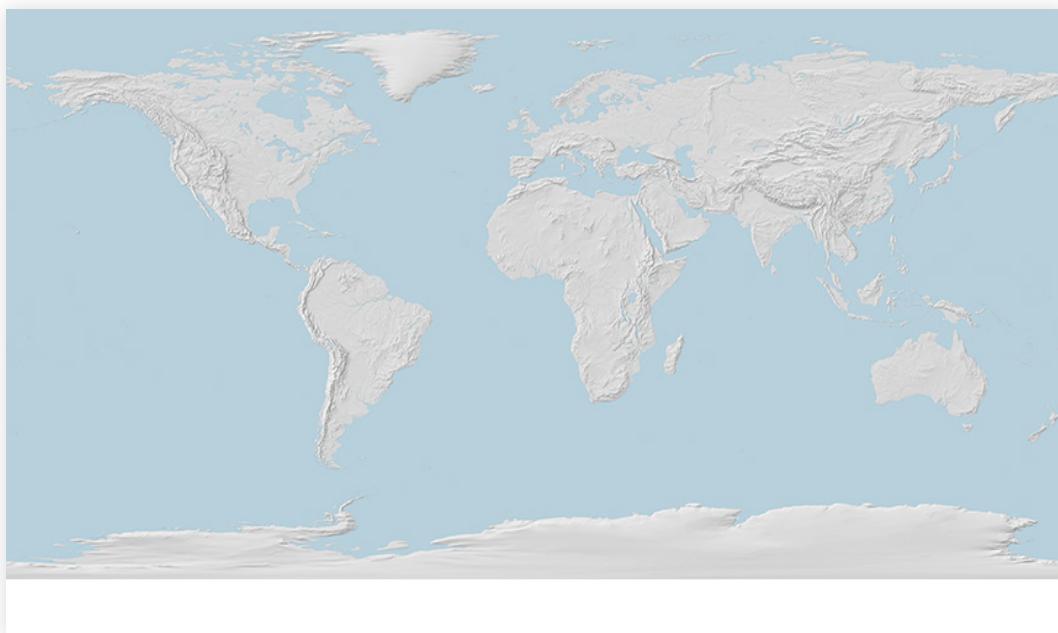
Traditionally, geographic information is represented as:

- Vector finite set of entities (shapes/geometries)
- Raster images encoding surfaces (values, colours, etc.)

# Vector



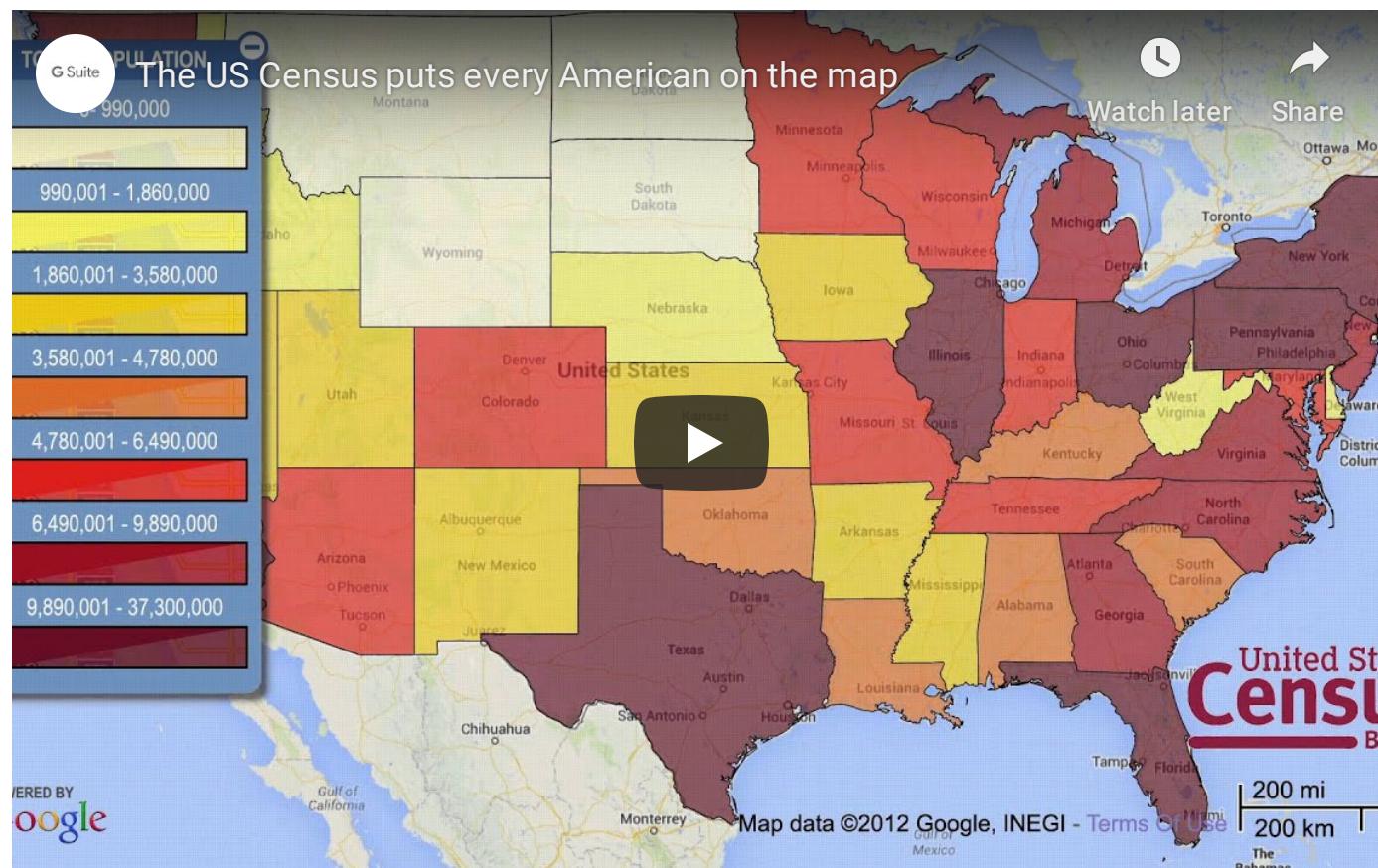
# Raster



*Good old spatial data*

# *Good old spatial data*

# [source]



# *Good old spatial data (+)*

Traditionally, datasets used in the (social) sciences are:

- Collected for the purpose -> carefully designed
- Detailed in information (“*...rich profiles and portraits of the country...*”)
- High quality

# *Good old spatial data (-)*

But also:

- Massive enterprises ("*...every single person...*) -> costly
- But coarse in resolution (to preserve privacy they need to be aggregated)
- Slow: the more detailed, the less frequent they are available

# Examples

- Decennial census (and census geographies)
- Longitudinal surveys
- Customly collected surveys, interviews, etc.
- Economic indicators
- ...

# New sources of (spatial) data

# New sources of (spatial) data

Tied into the (geo-)data revolution, new sources are appearing that are:

- **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

Different ways to categorise them...

# 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”

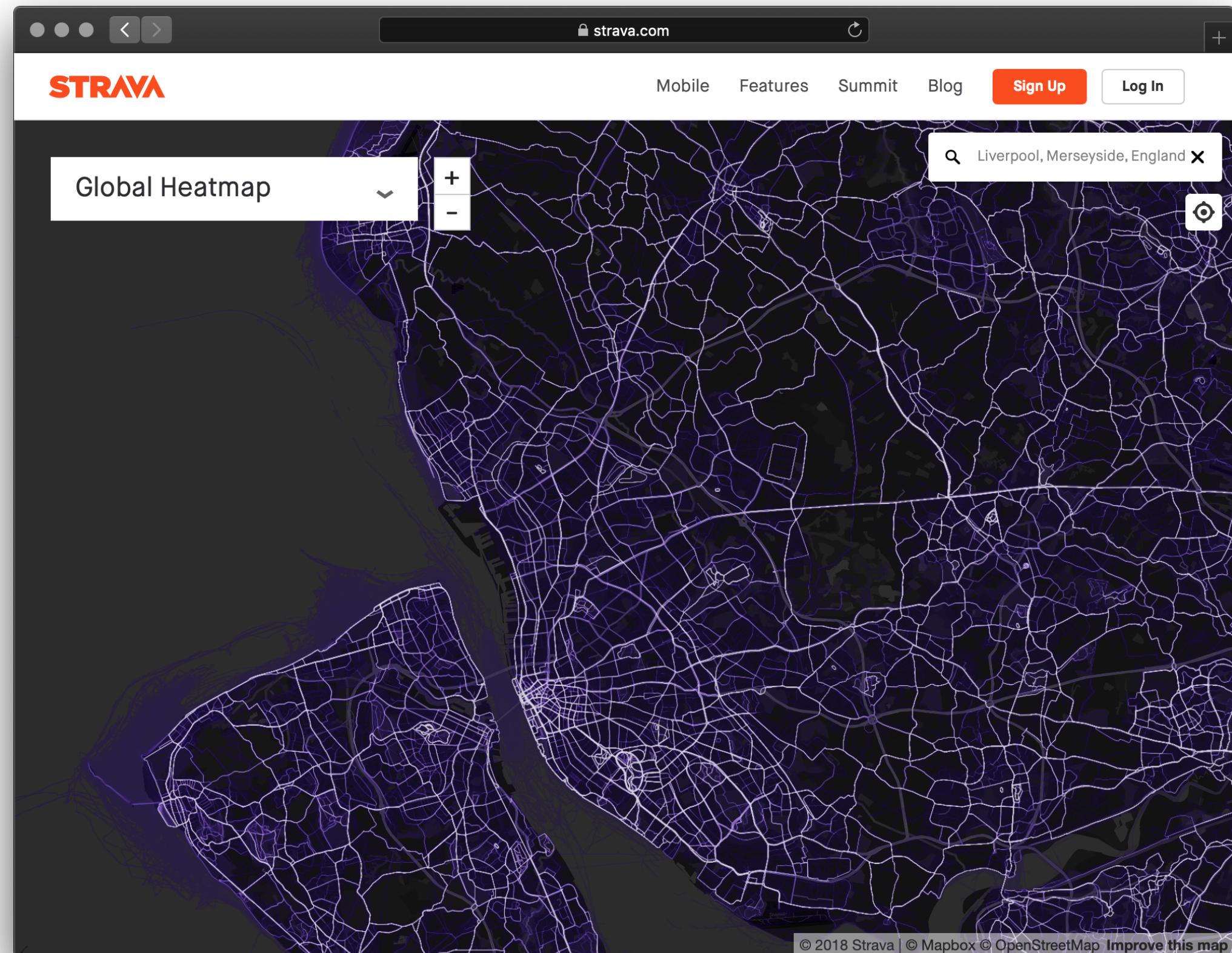
Old/New, raster/vector...

# Old/New, raster/vector...

Traditional **approaches** to represent the world in a computer are **blending** thanks to new forms of data

Keep an **open mind** to tools, approaches, and methods

Buildings from satellite (NYTimes)



GHSL



Geographic Data Science'19 by Dani Arribas-Bel is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.