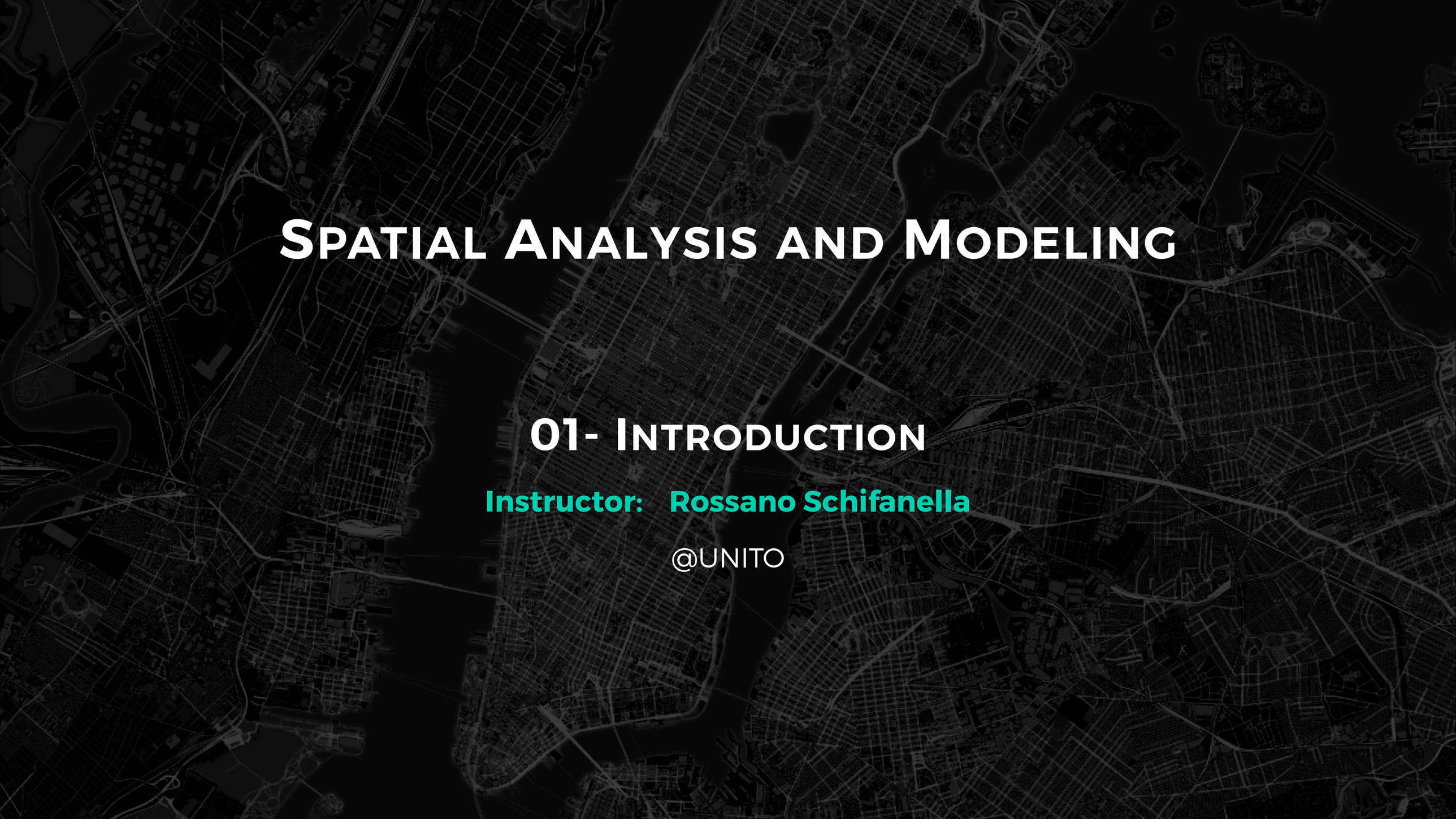


Spatial Analysis and Modeling

01- Introduction

Instructor: Rossano Schifanella

@UNITO


@Nokia Bell Labs
CAMBRIDGE
@IU
BLOOMINGTON
@UNITO + @ISI Foundation
@Yahoo
TURIN
NEW YORK
BARCELONA
@UDD
SANTIAGO

learning objectives

- **Why spatial thinking is important?**
- **Methodologies to**
  - describe and visualize
  - analyze and extract patterns
  - interpret and explain

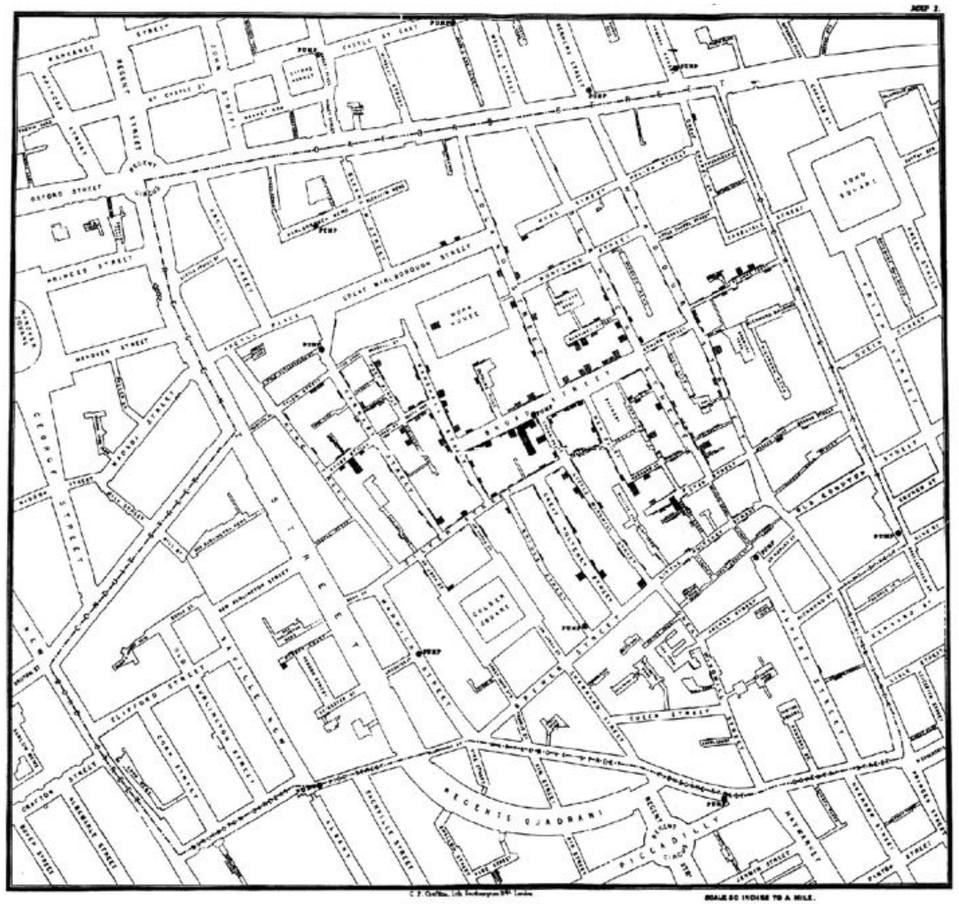
  spatial data in computational disciplines
- **Let you fall in love with spatial analysis and mapping!**

organization

- **Material GitHub repository**
  - **https://github.com/rschifan/SpatialAM-21**
- **Calendar**
  - **see WebEx meetings**
- **Classes**
  - **Theory + Practice** (Python ecosystem + PostGIS)
- **Evaluation**
  - **Data-driven Project + Essay**

Why spatial thinking is important?
a first example

John Snow's cholera map



Original map by John Snow showing the clusters of cholera cases in the London epidemic of 1854.

[see the Ghost Maps TED talk]

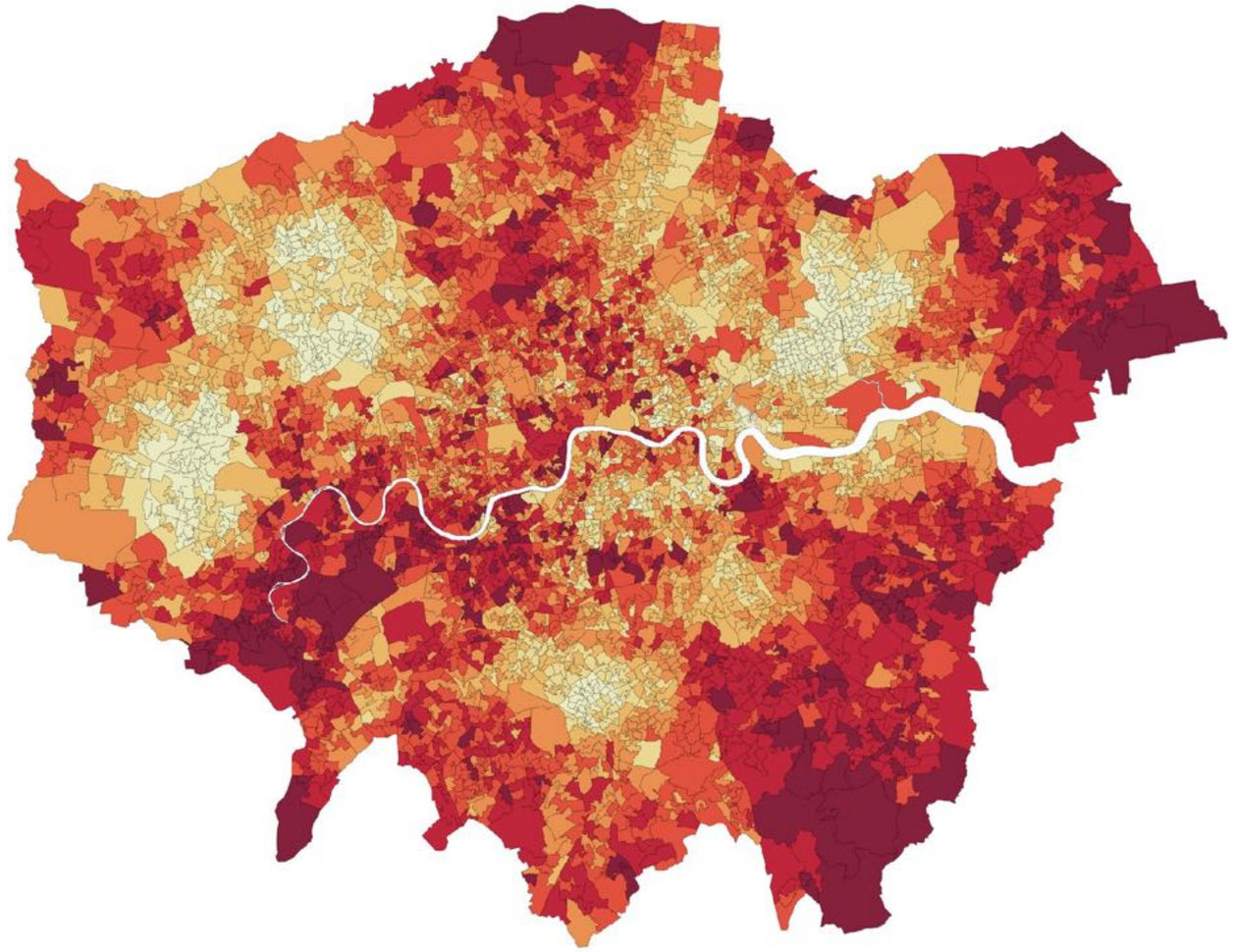
Why spatial thinking is important?

a second example

Spatial patterns matter

Percentage of white people in London (LSOA, census 2011)

darker red means higher concentration



Spatial patterns matter

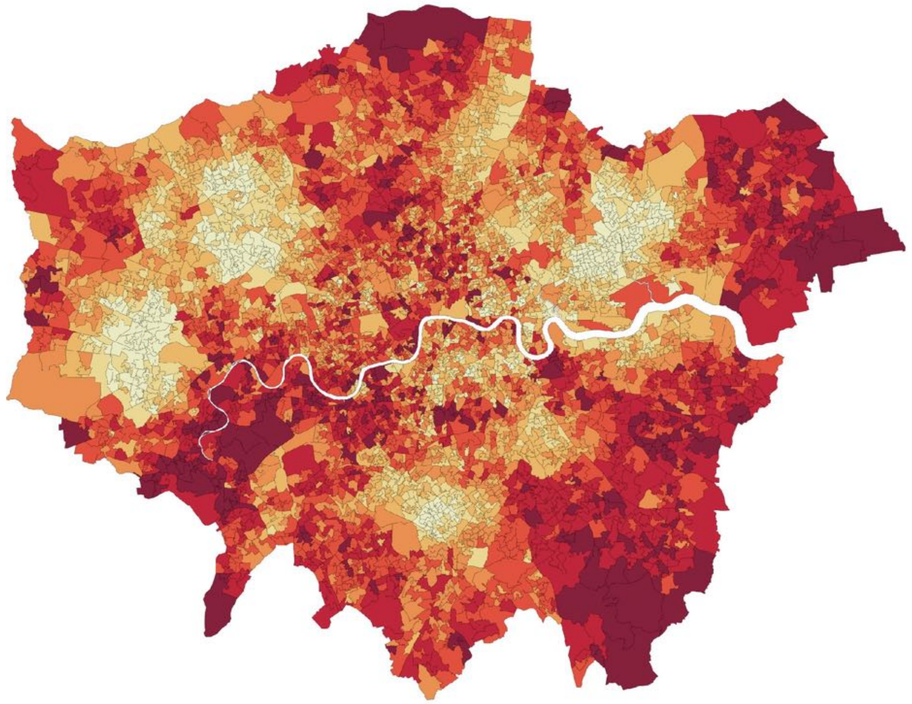
Percentage of white people in London (LSOA, census 2011)

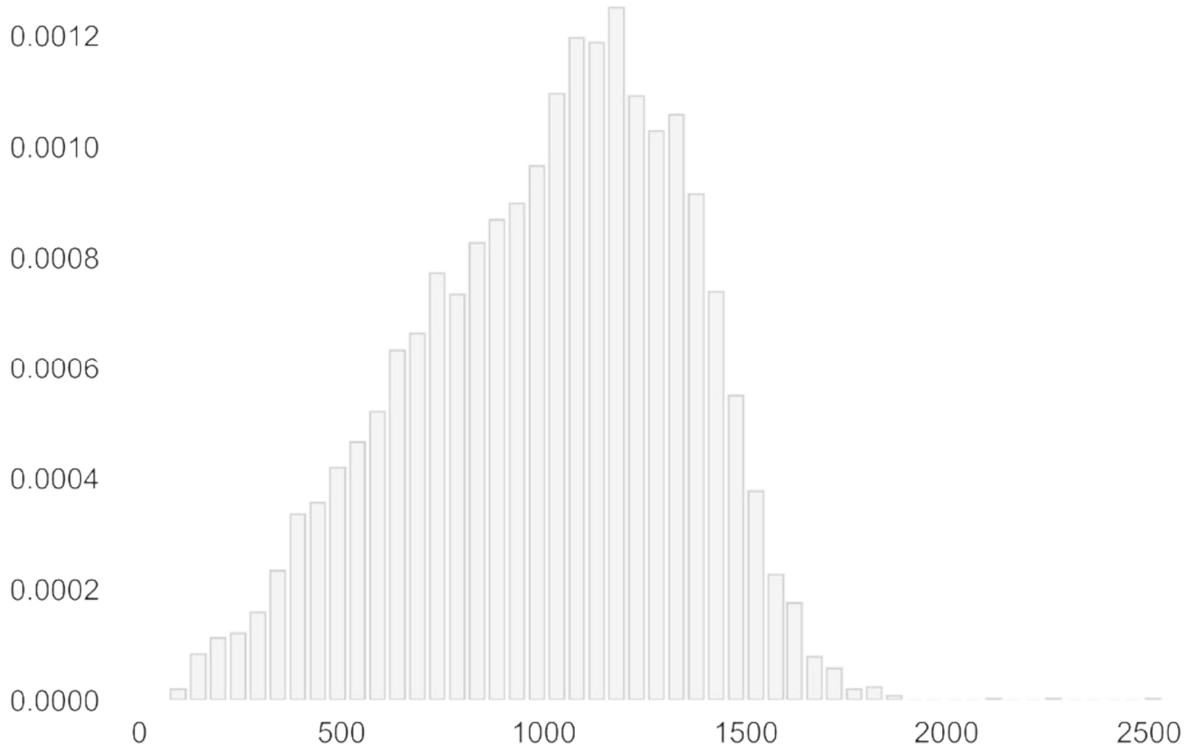
darker red means higher concentration

Spatial patterns matter

Percentage of white people in London (LSOA, census 2011)

darker red means higher concentration

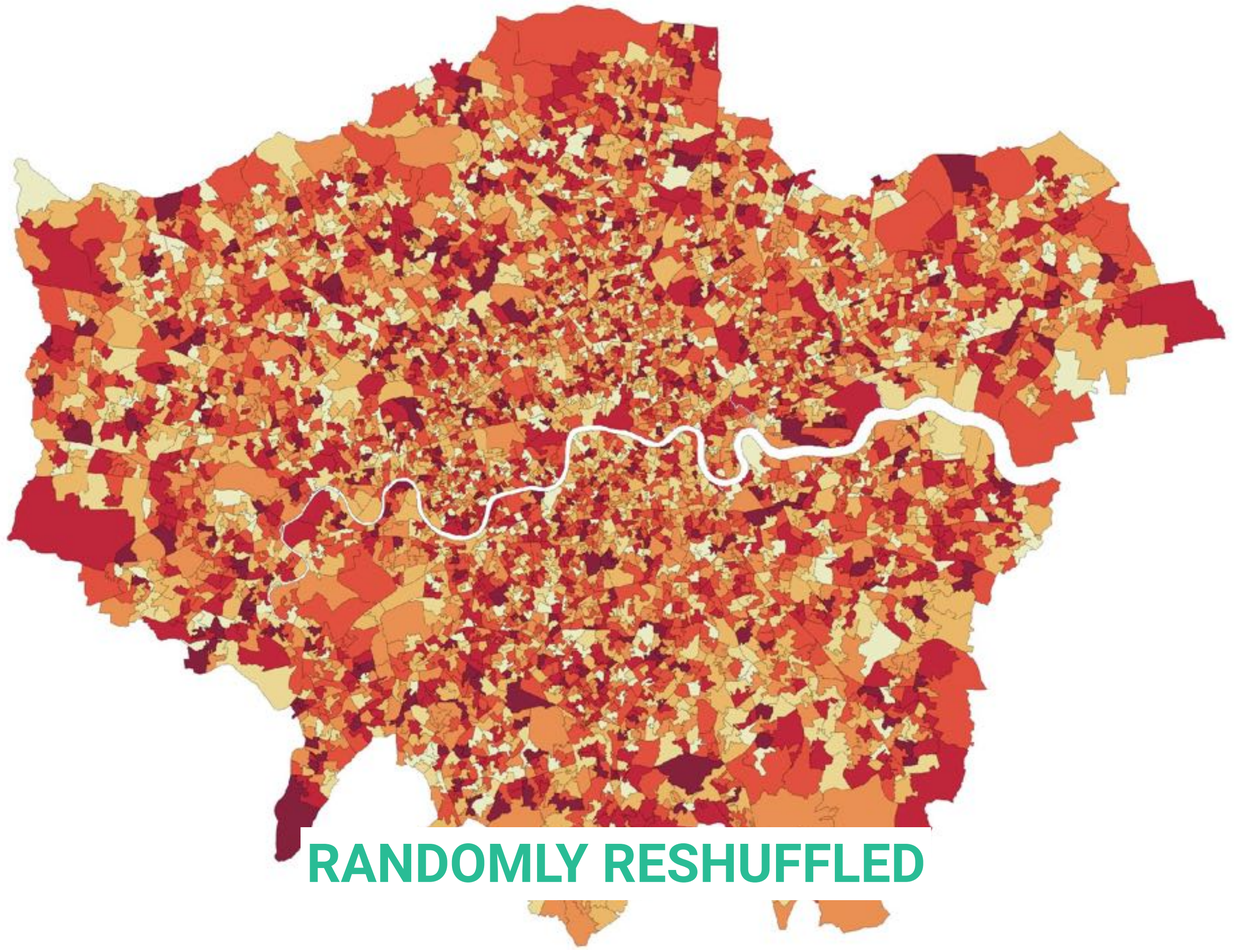




Spatial patterns matter

Percentage of white people in London (LSOA, census 2011)

darker red means higher concentration



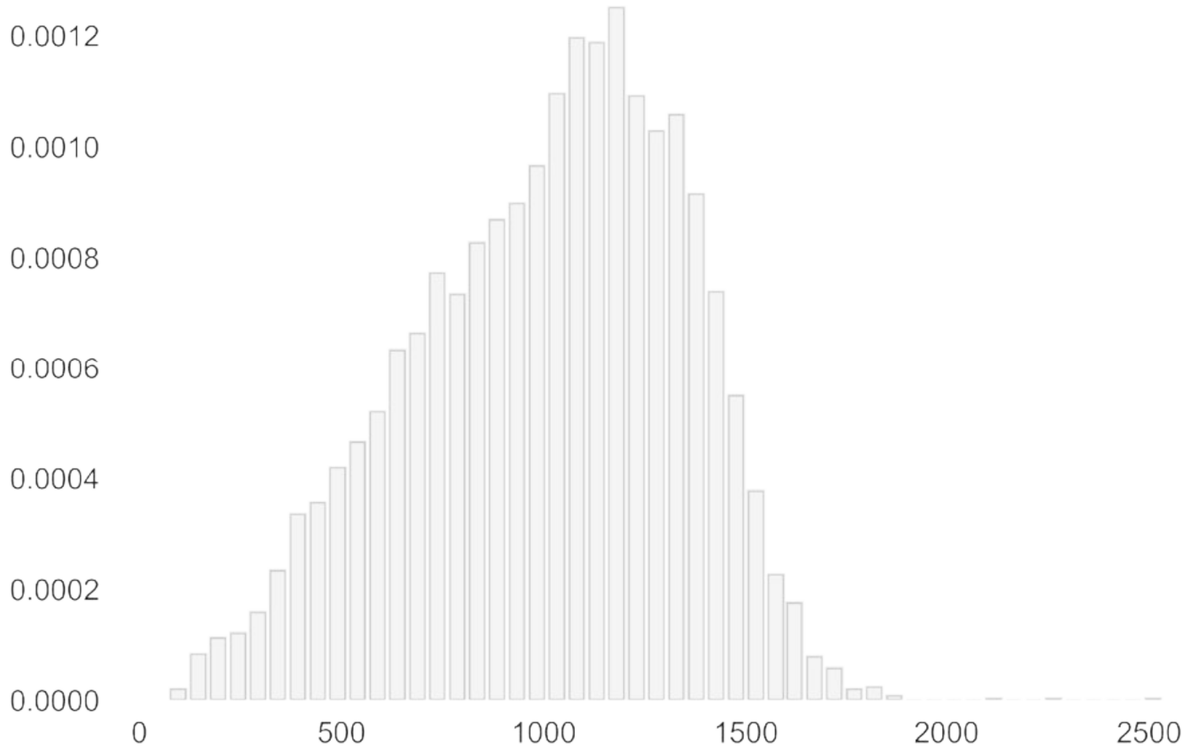
RANDOMLY RESHUFFLED

Spatial patterns matter

Percentage of white people in London (LSOA, census 2011)

darker red means higher concentration

SAME DISTRIBUTION!

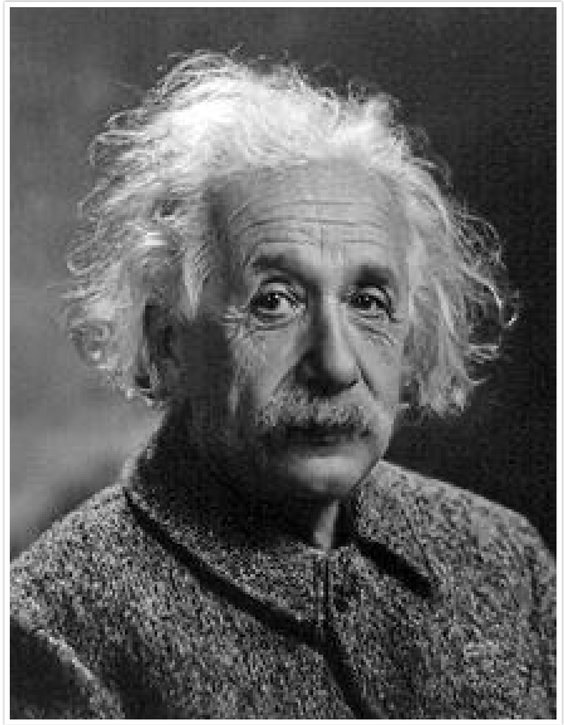


location data is
pervasive

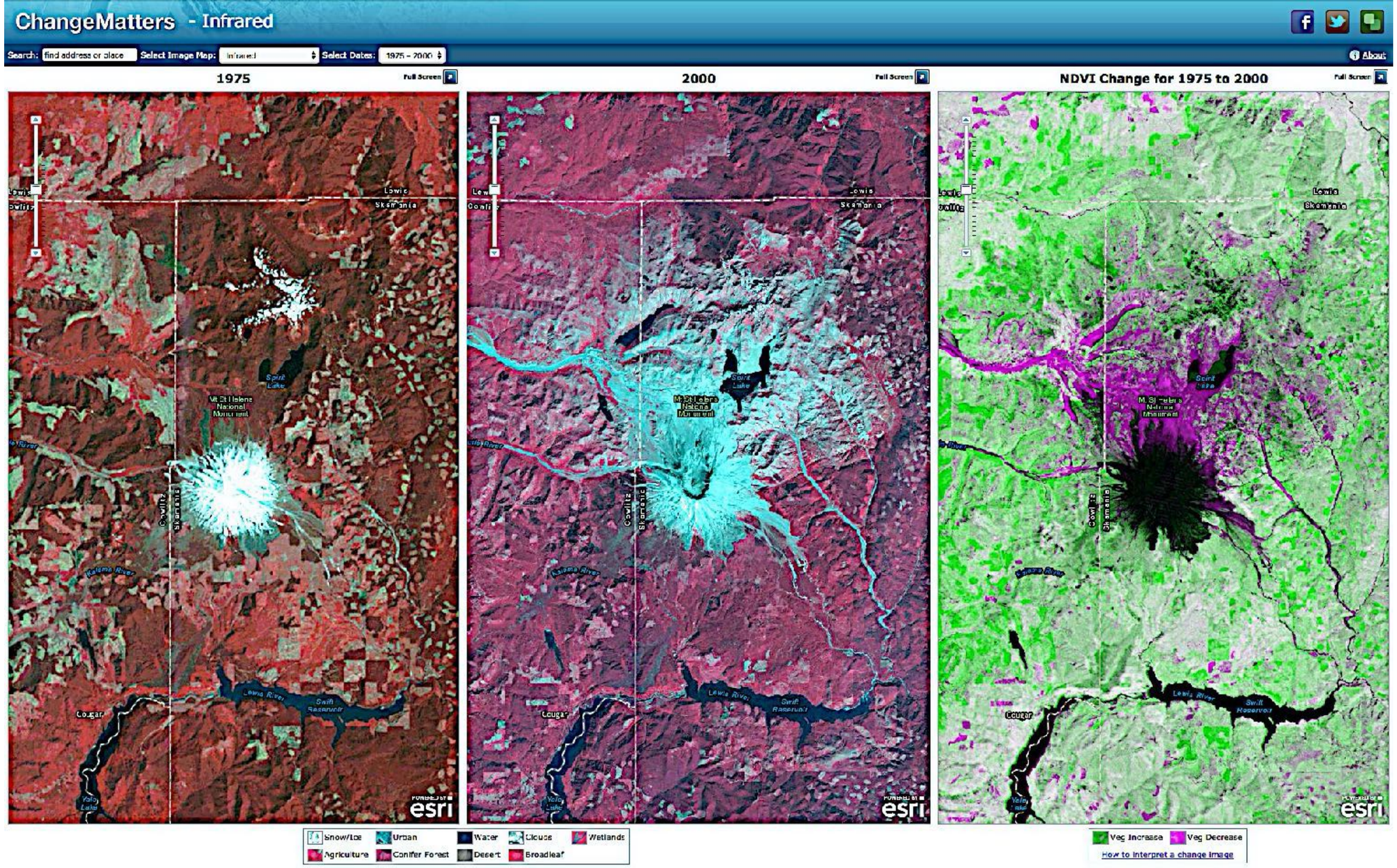
geography

“As a young man my fondest dream was to become a geographer. However, while working in the customs office I thought deeply about the matter and concluded that it was far too difficult a subject. With some reluctance, I then turned to physics as a substitute.”

Albert Einstein (1879-1955) Unpublished Letters



Geographers are masters of processes that unfold across space and time



Almost everything in geography involves a dynamic process of one type or another

What is Geography?

- ***geo*** **meaning “Earth”, *graphy* refers to “writing” and “describing”**

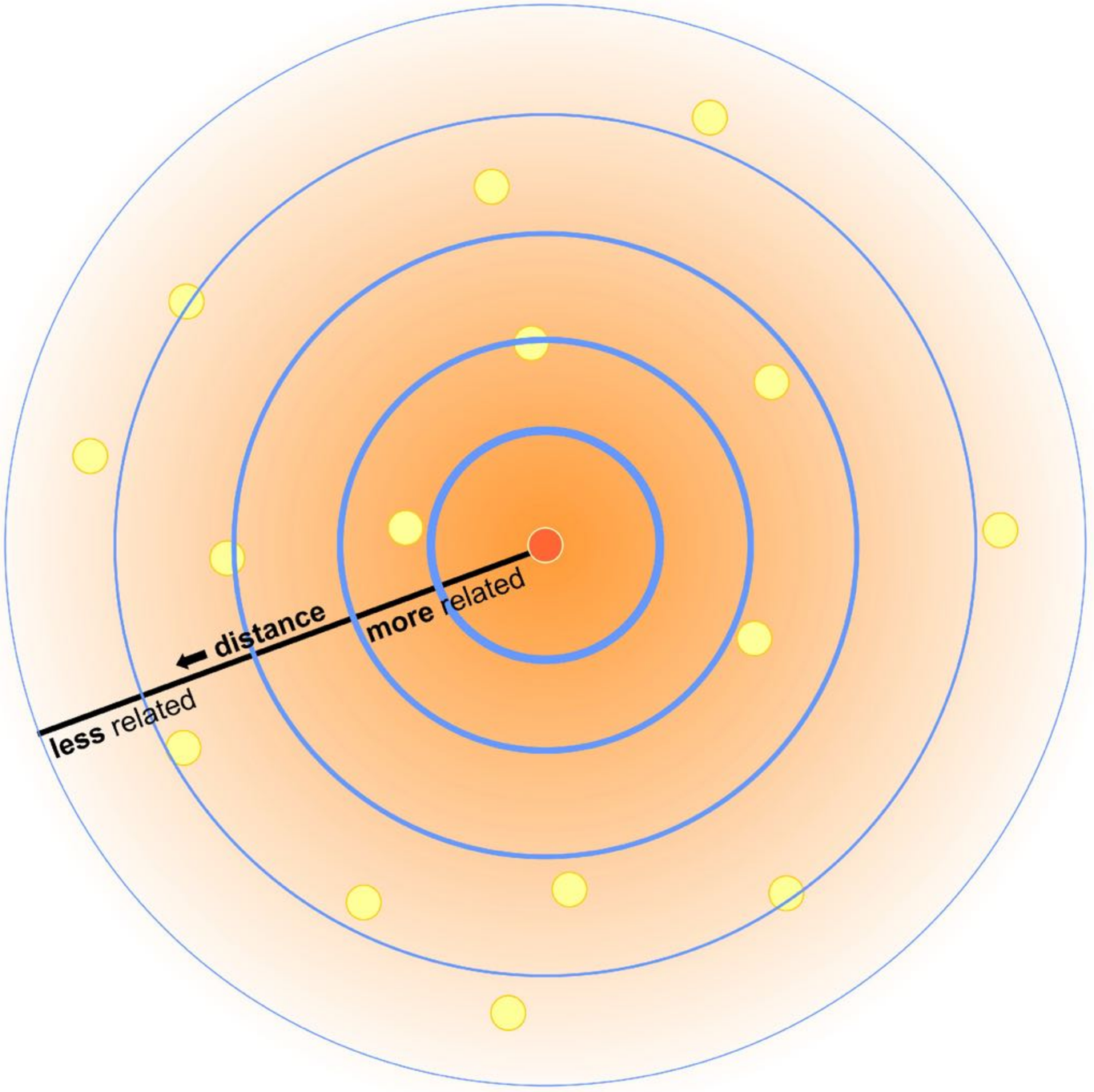
TWO MAIN AREAS

- **Physical (environment)**
  - Branch of **natural science**
    - Environmental geography
    - Hydrology
    - Biogeography
    - Climatology
    - Meteorology
    - Oceanography
    - Environmental geography
    - Landscape ecology
- **Human**
  - Branch of **social science**
    - Culture
    - Development
    - Economic
    - Health
    - Historical
    - Political
    - Population
    - Settlement
    - Urban

key theory #1
Tobler's law

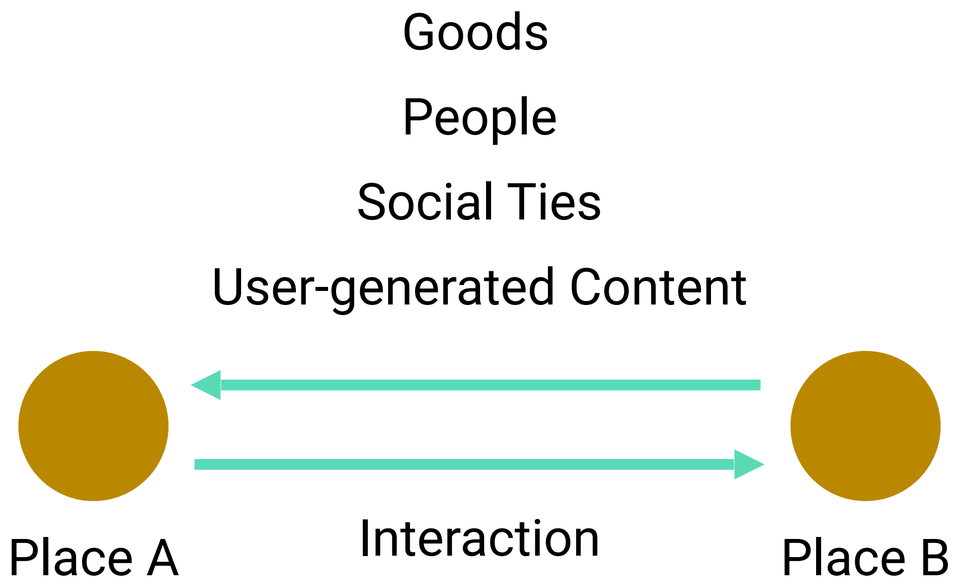
"Everything is related to everything else, but near things are more related than distant things"

Waldo Tobler (1970)


distance
more related
less related

key theory #2
gravity model

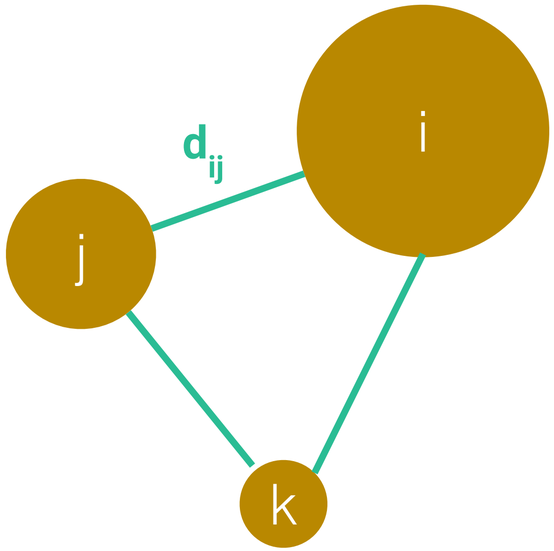
- **Explains interaction between places**
- **Originally proposed to describe population migration between two regions**



key theory #2
gravity model

(Basic form)

$$T_{ij} = k\frac{P_i P_j}{d_{ij}}$$

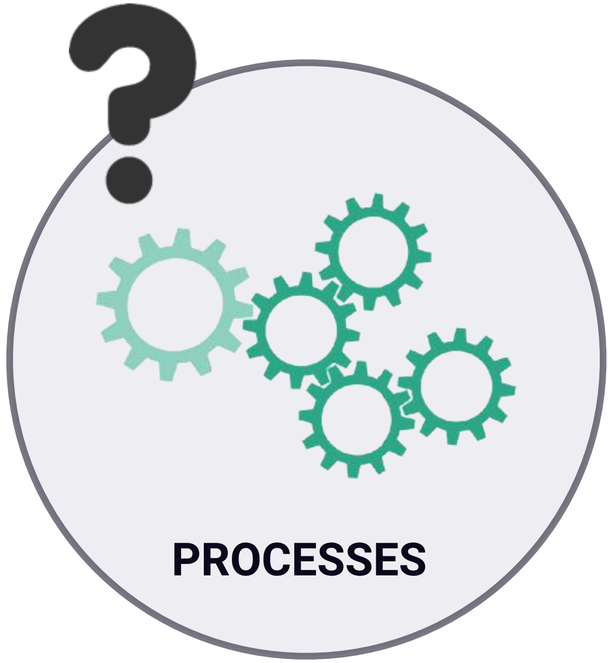


Advanced form

$$T_{ij} = k\frac{P_i^{\lambda} P_j^{\alpha}}{d_{ij}^{\beta}}$$

spatial data analysis

- **Processes operating in space create patterns**
- **Spatial Data Analysis is aimed at:**
  - Identifying and describing the **patterns**
  - Identifying and understanding the **processes**

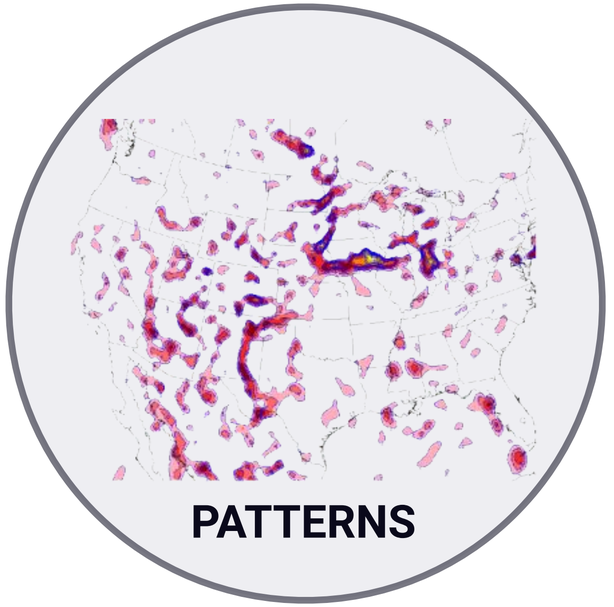


PROCESSES

Often not directly observable

INFER

CREATE



PATTERNS

Often a sample of the data

Spatial Data Analysis:
examples

- **Seismologist** collects data on the regional distribution of earthquakes. Does this distribution show any pattern or predictability over space?
- **Geologist** wish to estimate the extent of a mineral deposit over a particular region, given data on borehole samples taken from locations scattered across the area. How can we make sensible estimates?
- **A groundwater hydrologist** collects data on the concentration of a toxic chemical in samples collected from a series of wells. Can we use these samples to construct a regional map of likely contamination?

Spatial Data Analysis:
examples

- **Public health specialist** collects data on the occurrence of diseases. Does the distribution of cases of a disease form a pattern in space? Is there some association with possible sources of environmental pollution?
- **Police** wish to investigate if there is any spatial pattern to the distribution of certain crime locations. Does the rate of crime in particular areas correlate with socio-economic characteristics of the area?
- **Retailers** wish to use socio-economic data, available for small areas from the population census, to assess the likely demand for their products if they open or expand an outlet. How are we to classify such areas? The same retailers collect information on movements of shoppers from residential zones to stores. Can we build models of such flows? Can we predict changes in such flows if we expand an outlet or open a new one?

spatial data analysis:
successive levels of sophistication

- **Spatial Data Description**
  - Focus is on describing the spatial data and representing it in a digital format (maps, databases)
  - Uses classic GIS capabilities (buffering, map layer overlay, spatial queries, measurement)
- **Exploratory Spatial Data Analysis (ESDA)**
  - Showing and discovering interesting patterns
- **Spatial statistical analysis and hypothesis testing**
  - An extension of traditional statistics into a spatial settings to determine whether or not data are typical or unexpected
- **Spatial modeling**
  - Explaining interesting patterns
  - Optimization, simulation, prediction
  - Involves constructing models to predict spatial outcomes

Spatial data types

Discrete

only found at fixed locations or when the data represent only specific values, e.g., # accidents at crossings

Point Line Polygon

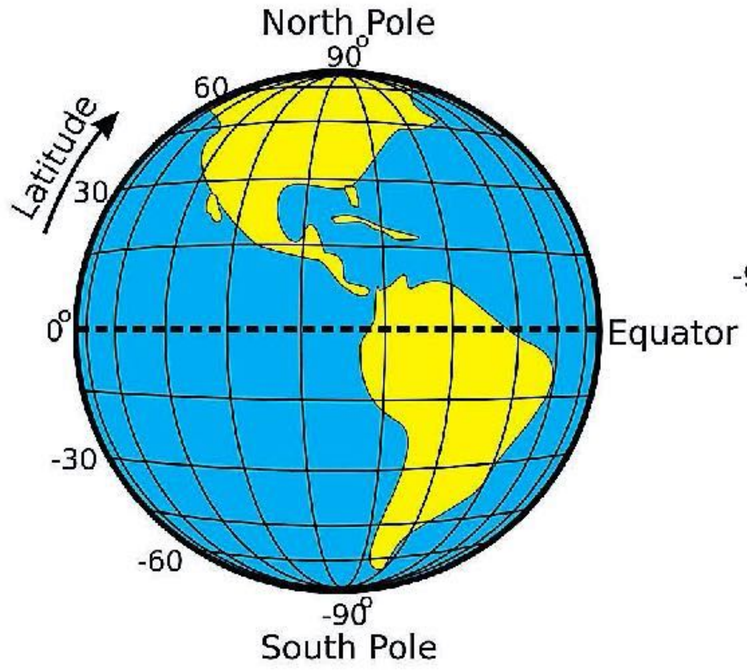
Continuous

seen throughout the mapped area and smoothly transitions from one value to another, e.g., air temperature.

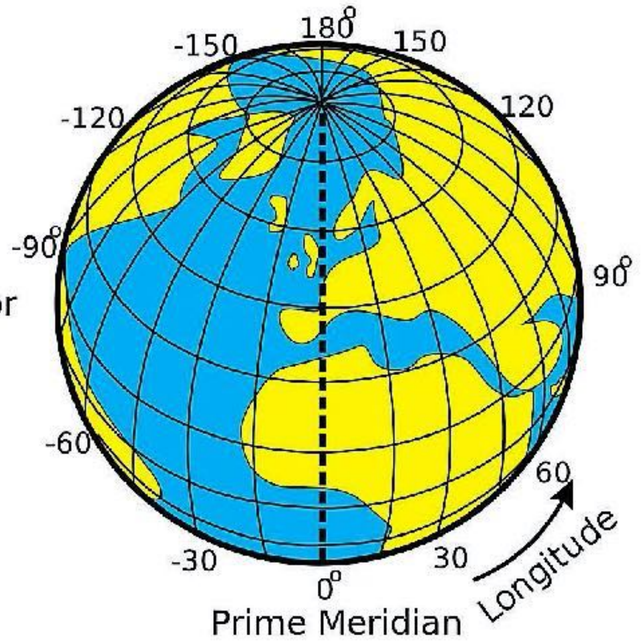
Surface Volume

georeferenced data:
coordinates

- (longitude, latitude) can be associated with
  - altitude, accuracy, timestamp
- can be reversed geocoded to a readable address



VERTICAL

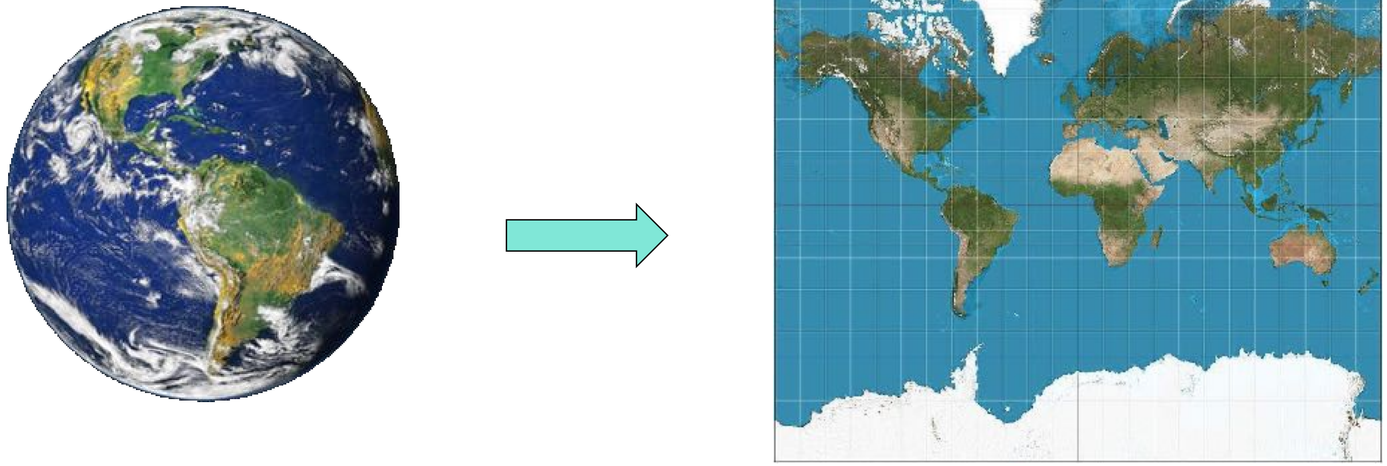


HORINZONTAL

video

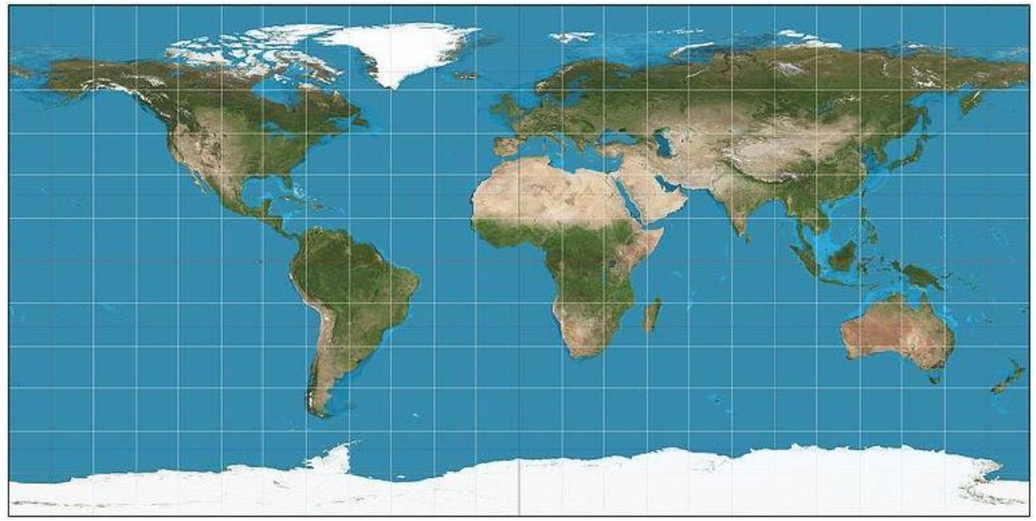
Earth is 3D, maps no!
map projections

- a projection is used to transform the geographic coordinates from the curved surface of our planet to the flat surface of a plane

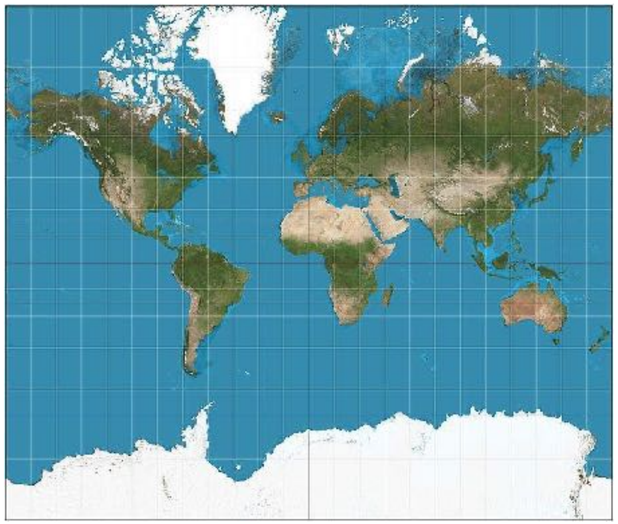


map projections
examples

- **over the years a variety of map projections have been proposed**



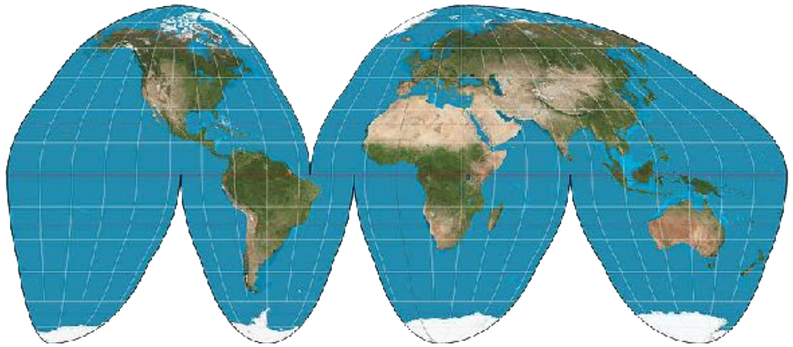
Equirectangular



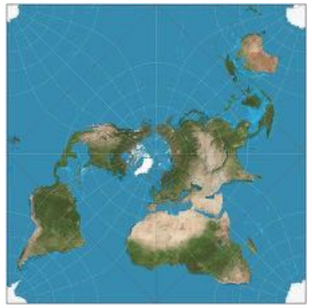
Mercator

map projections
examples

Cassini



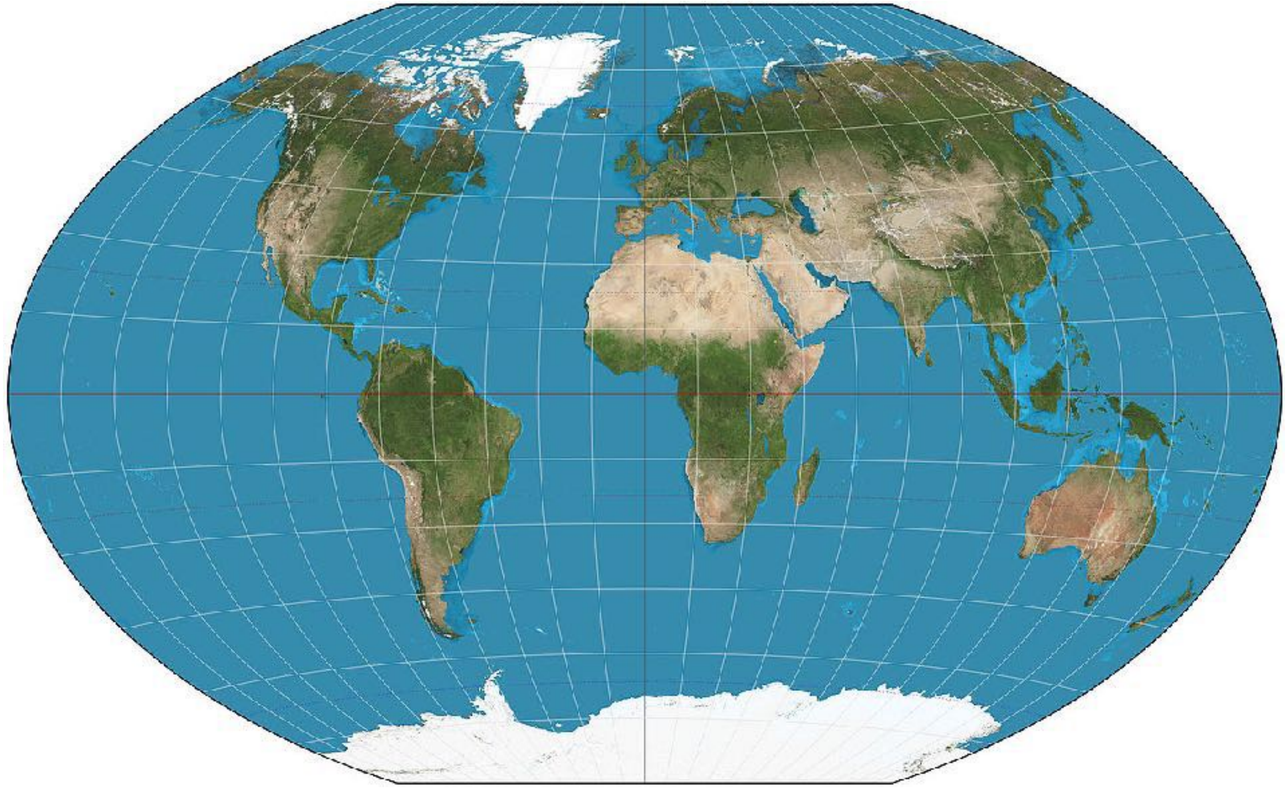
Goode homolosine



Peirce quincuncial

Albers conic

map projections
examples

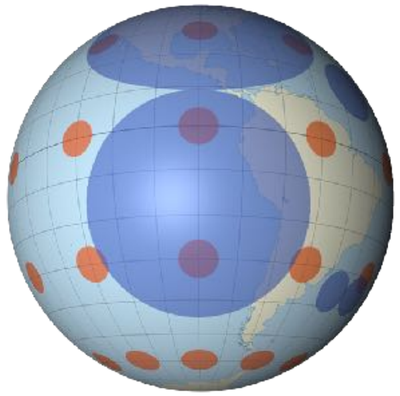


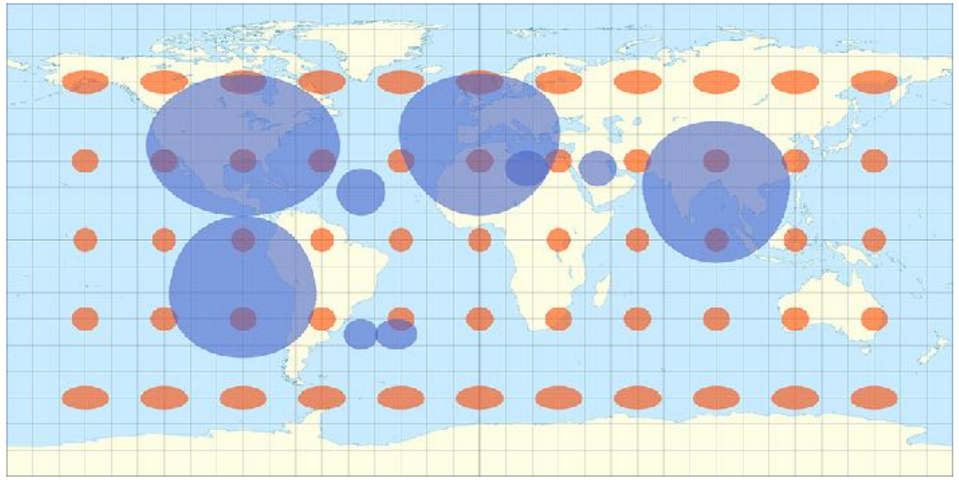
Winkel Triple

adopted by National Geographic

map projections
distortions

- **projections cause distortions**
  - **shape, area, distance, direction**
  - **depending on the application, some projections may be more suitable than others**
  - **Good references:**
    - https://www.youtube.com/watch?v=kIID5FDi2JQ
    - https://www.youtube.com/watch?v=eLqC3FNNOaI

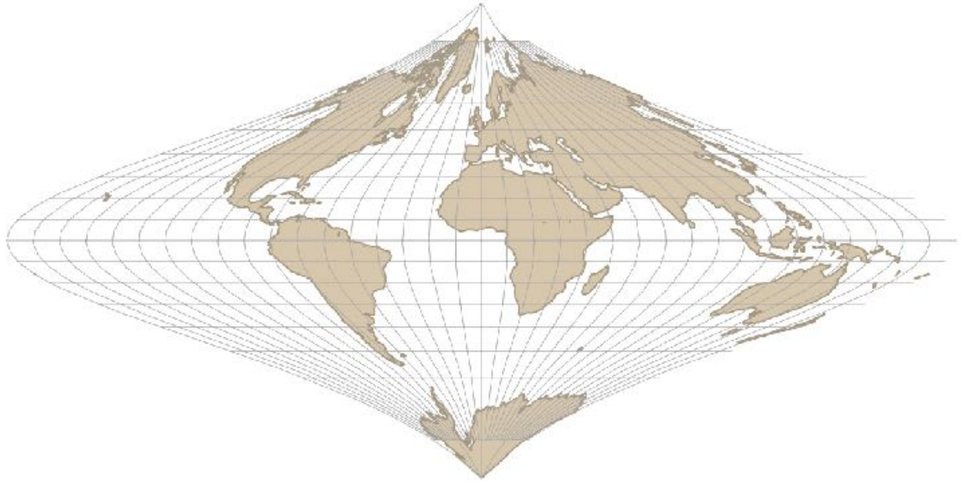


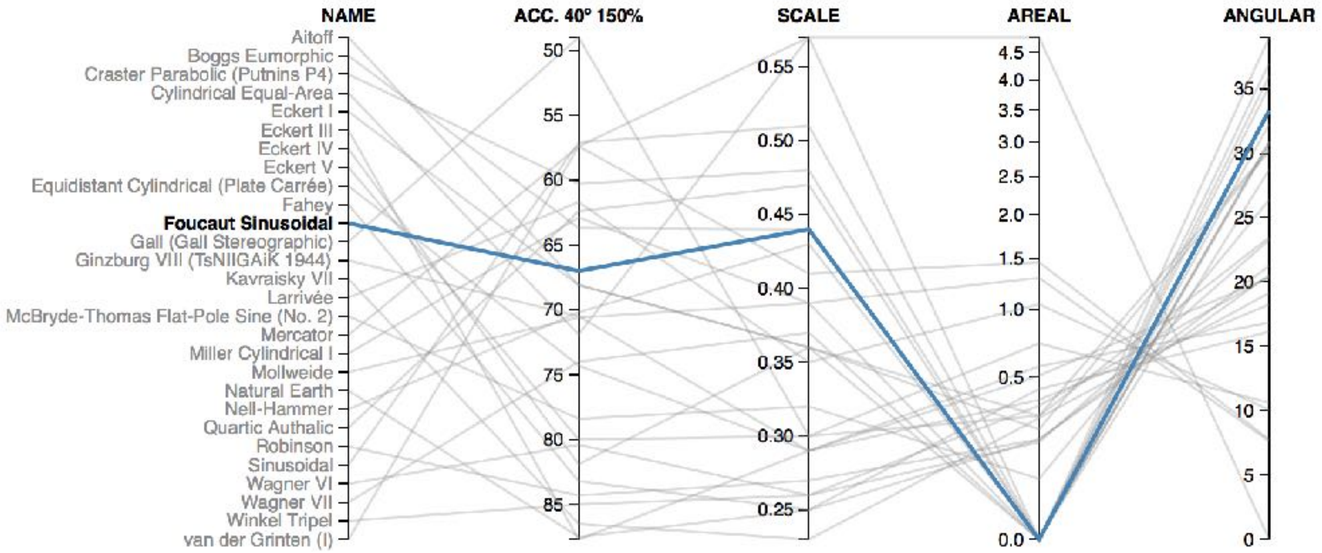


types of projections

- **azimuthal**
    - preserves the azimuth (direction) from center
- **conformal**
    - local angles are correct, preserving small shapes
- **equal-Area**
    - equal-area maps preserve area measure, generally distorting shapes
- **equidistant**
    - distances from center (or along certain lines, like along meridians) are correct

compare projections on d3.js





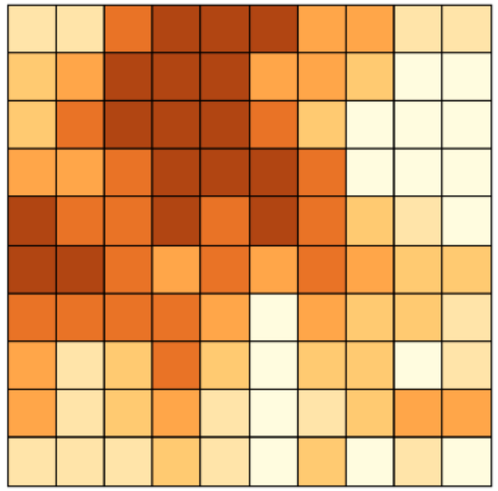
projections can produce societal biases

- See video from "The West Wing" Season 2 Episode 16
  - https://www.youtube.com/watch?v=vVX-PrBRtTY&t
- Other useful references:
  - https://www.youtube.com/watch?v=kIID5FDi2JQ
  - https://www.youtube.com/watch?v=KUF_Ckv8HbE

pitfalls

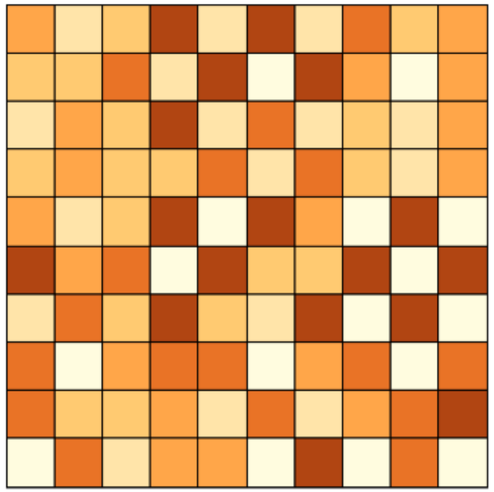
visual experiment
apophenia

A



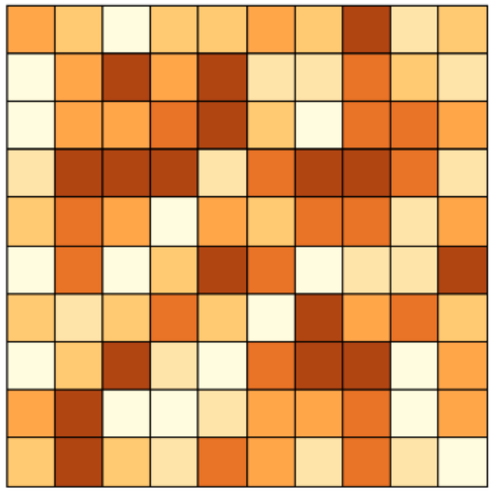
clustered

B



dispersed

C



random

Aphophenia is the tendency to mistakenly perceive connections and meaning between unrelated things

Can be considered a generalization of **pareidolia**

Patternicity
Michael Shermer (2008)

- **The tendency to find meaningful patterns in both meaningful and meaningless noise**
    - Type I error (false positive)
    - Type II error (false negative)
- **Humans are pattern-seeking primates and this behavior is hardly-coded in how our brain works**
- **Related to survival skills**
    - https://www.ted.com/talks/michael_shermer_the_pattern_behind_self_deception

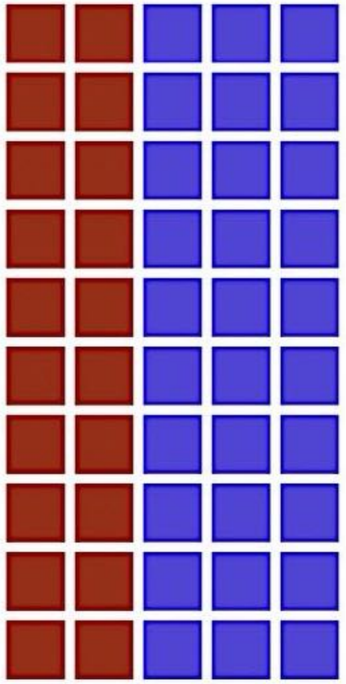
MAUP
Modifiable Area Unit Problem

- **The same basic data yield different results when aggregated in different ways**
    - Nice read: “A million or so correlation coefficient: three experiments on the modifiable area unit problem” (Openshaw and Taylor, 1979)
- **Zonal effect**
    - Similar size and number of units, but different boundaries
        - Zip codes versus census tracts, postal zones versus city neighborhoods
- **Scale effect**
    - Increases size and decreases number of units
        - US counties versus states
    - Global model might be inconsistent with local models
- **The take home message is that how we aggregate the input units will impact the values of the output units**

Gerrymandering, explained

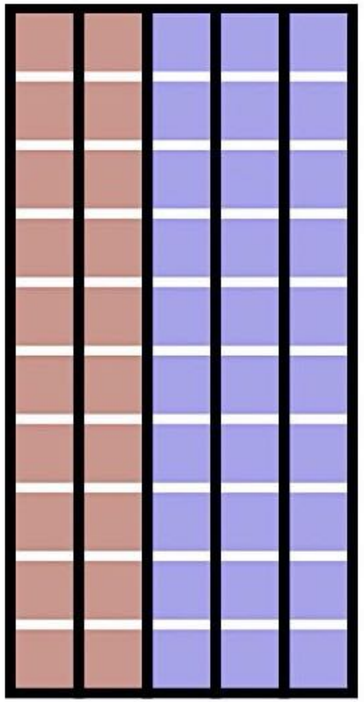
Three different ways to divide 50 people into five districts

50 people



60% blue, 40% red

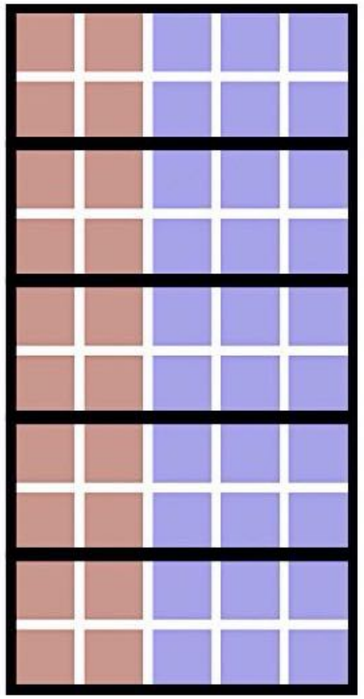
1. Perfect representation



3 blue districts, 2 red districts

BLUE WINS

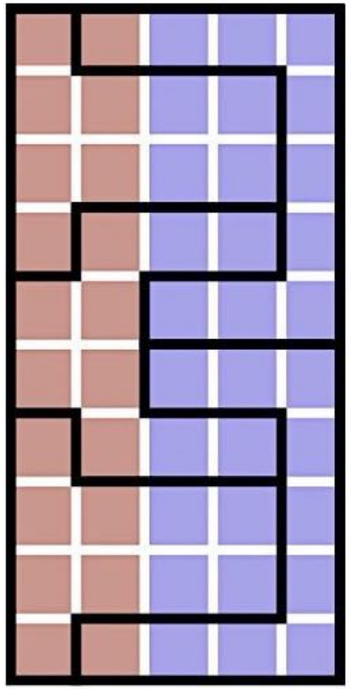
2. Compact, but unfair



5 blue districts, 0 red districts

BLUE WINS

3. Neither compact nor fair



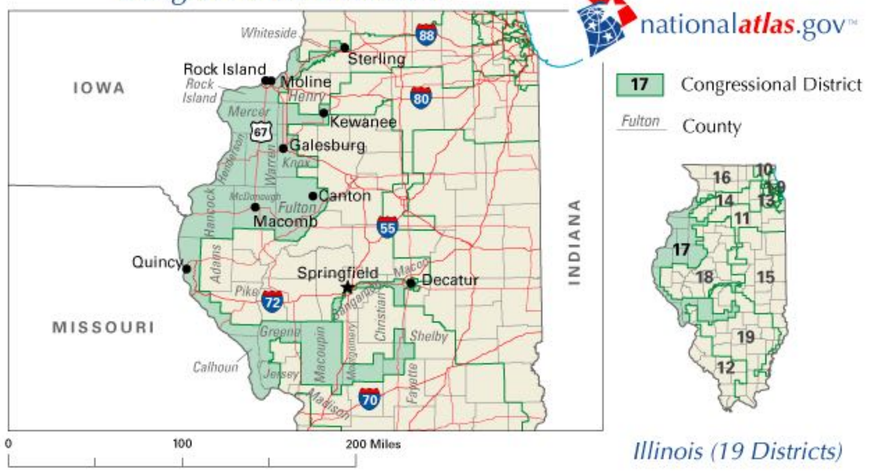
2 blue districts, 3 red districts

RED WINS

a first real example
gerrymandering

- **In the process of setting electoral districts, intended to establish a political advantage for a particular party or group by manipulating district boundaries**
  - **cracking**: diluting the voting power of the opposing party's supporters across many districts
  - **packing**: concentrating the opposing party's voting power in one district to reduce their voting power in other districts

Congressional District 17



nationalatlas.gov™

17 Congressional District

Fulton County

Illinois (19 Districts)

Congressional District 2

nationalatlas.gov™

2 Congressional District

Grand County

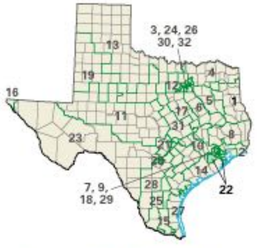
Utah (3 Districts)

Congressional District 22

nationalatlas.gov™

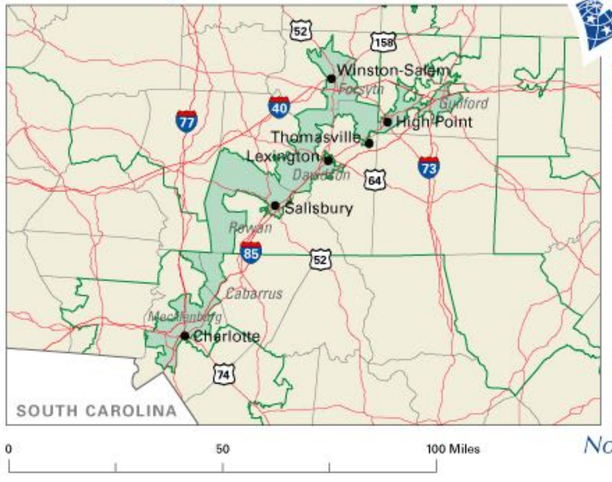
22 Congressional District

Harris County



Texas (32 Districts)

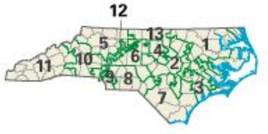
Congressional District 12



nationalatlas.gov™

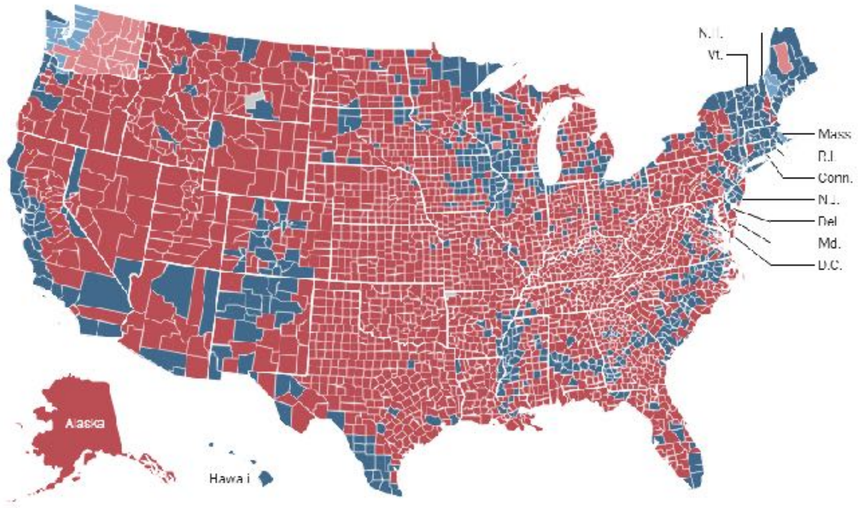
12 Congressional District

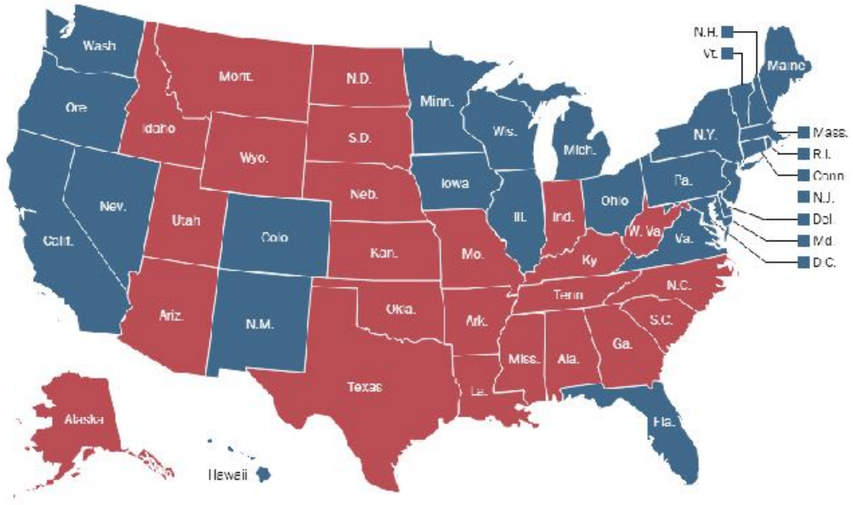
Rowan County



North Carolina (13 Districts)

US election 2012
counties versus states





Ecological Fallacy

- **Individual behavior cannot be explained at the aggregate level**
- **Issue of interpretation**
    - e.g., county homicide rates do not explain individual criminal behavior
    - model aggregate dependent variables with aggregate explanatory variables
    - alternative: multilevel modeling

COSP
Change of Support Problem

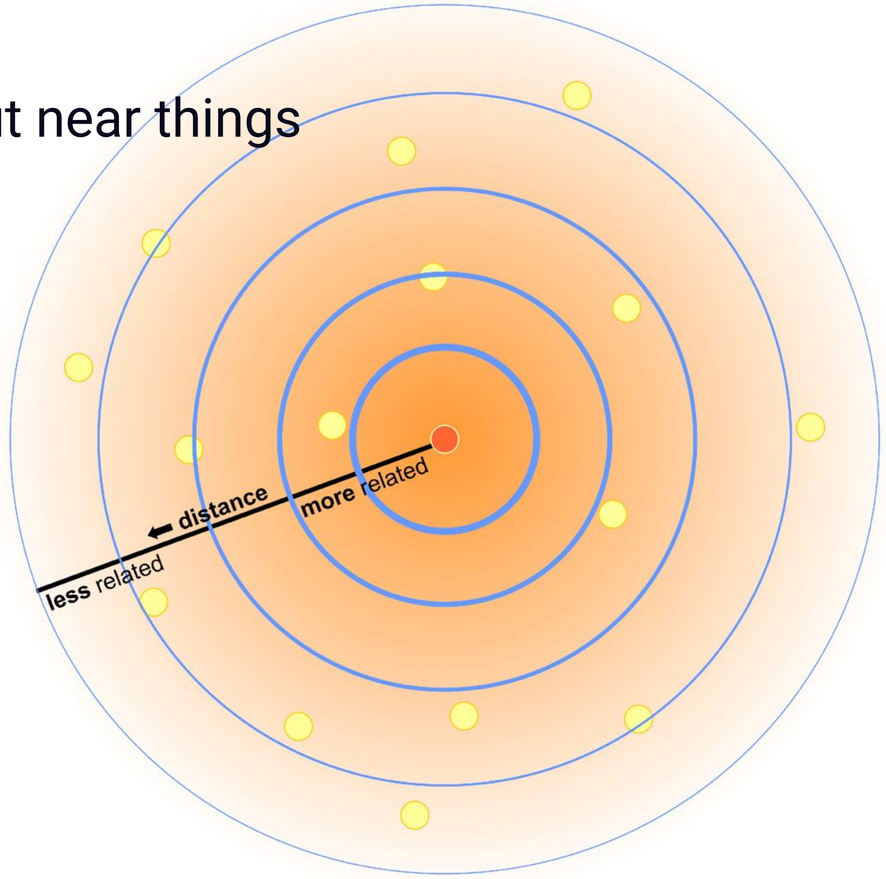
- **Variables measured at different spatial scales**
- **Spatial misalignment**
    - we collected the data on one scale, but need to make inferences on a different scale.
        - How do we change from one spatial scale to another?
    - have different spatial datasets that come to us on different spatial scales.
        - How do we combine data sources?
- **Aggregate up to a common scale (the finest possible)**

other critical issues

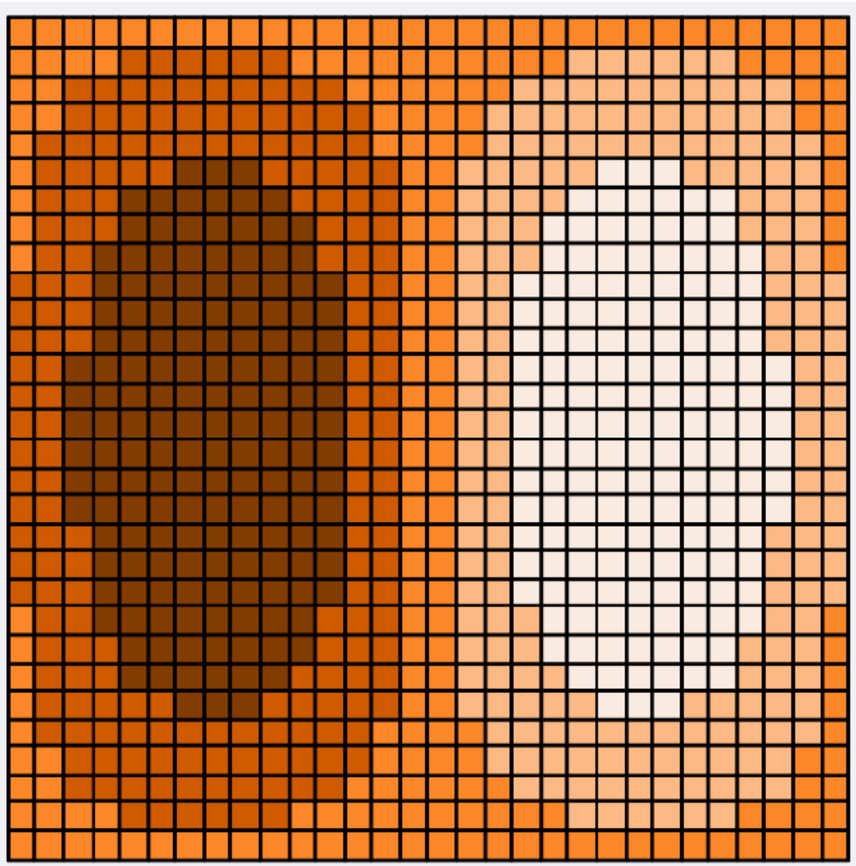
- **Edge issues**
  - Every study region has a boundary (unless you study the entire world!)
  - You do not have data for outside your study region
  - However, the outside data can affect the inside data if there is spatial autocorrelation
  - Consequently, edges of the map, beyond which there is no data, can significantly effect results

other critical issues

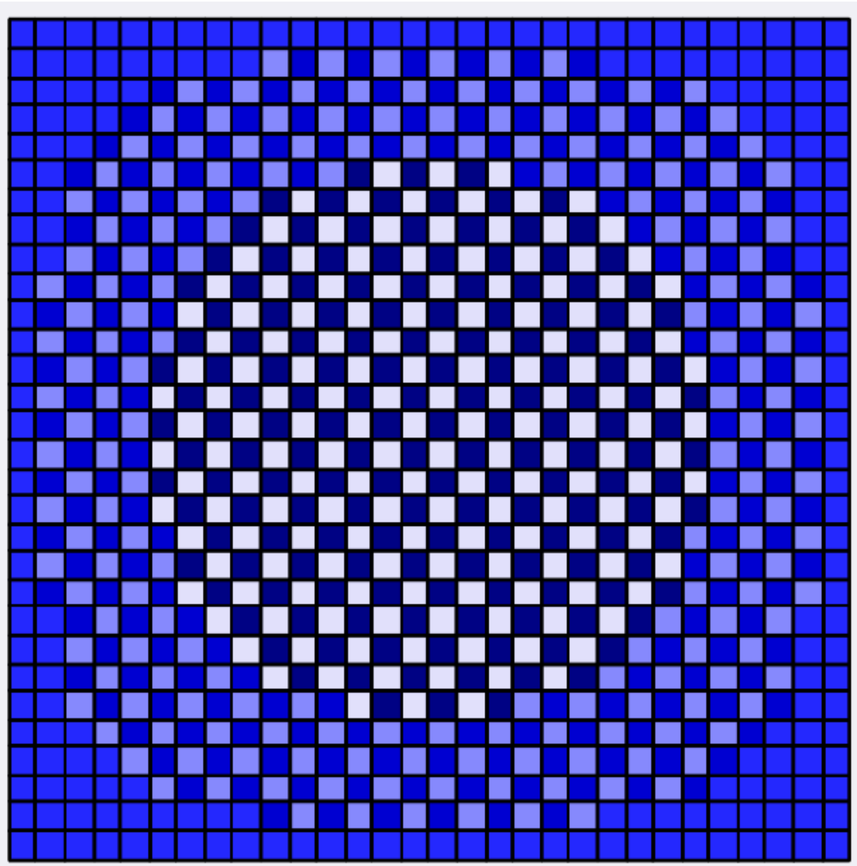
- **Spatial autocorrelation**
  - **Measures the correlation of a variable with itself through space**
  - Related to Tobler’s first law of geography
    - Everything is related to everything else, but near things are more related than distant things.



spatial autocorrelation



positive = clustered



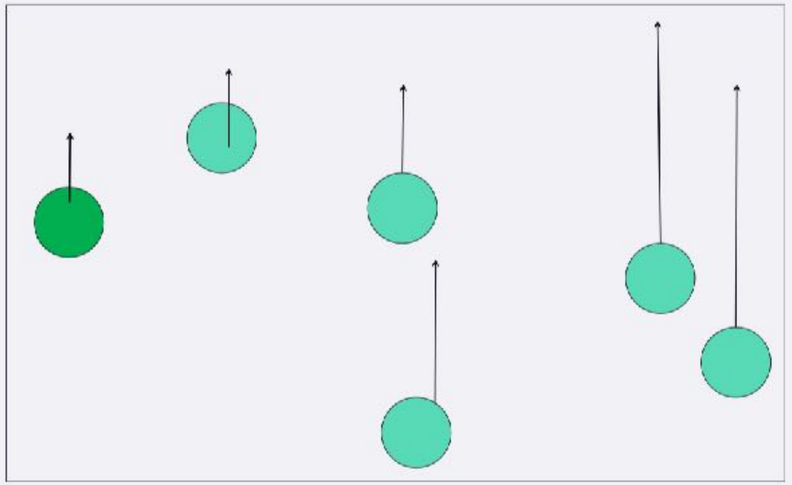
negative = dispersed

why is spatial autocorrelation important?

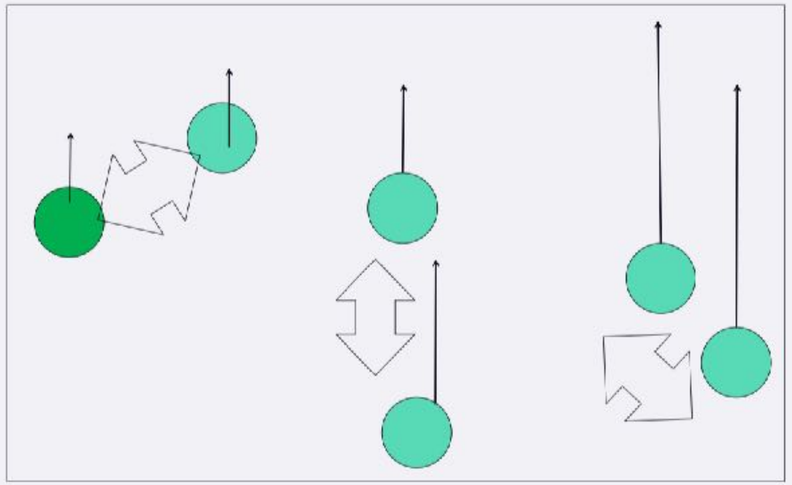
- **It implies the existence of a spatial process**
  - Why are near-by areas similar to each other?
  - Why do high income people live close each other?
  - These are geographical questions.
    - They are about location
- **It invalidates most traditional statistical inference tests**
  - If spatial autocorrelation exists, the results of standard statistical inference tests may be incorrect
  - We need to use spatial statistical inference tests
- **For example**
  - You are more likely to incorrectly conclude a relationship exists when it does not
  - You believe that the relationship is stronger than it really is

why is spatial autocorrelation important?

- **Statistical tests are based on the assumption that the values of observations in each sample are independent of one another**
- **spatial autocorrelation violates this**
  - samples taken from nearby areas are related to each other and are not independent



Values near each other are similar in magnitude.



Implies a relationship between nearby observations

@rschifan

schifane@di.unito.it

http://www.di.unito.it/~schifane