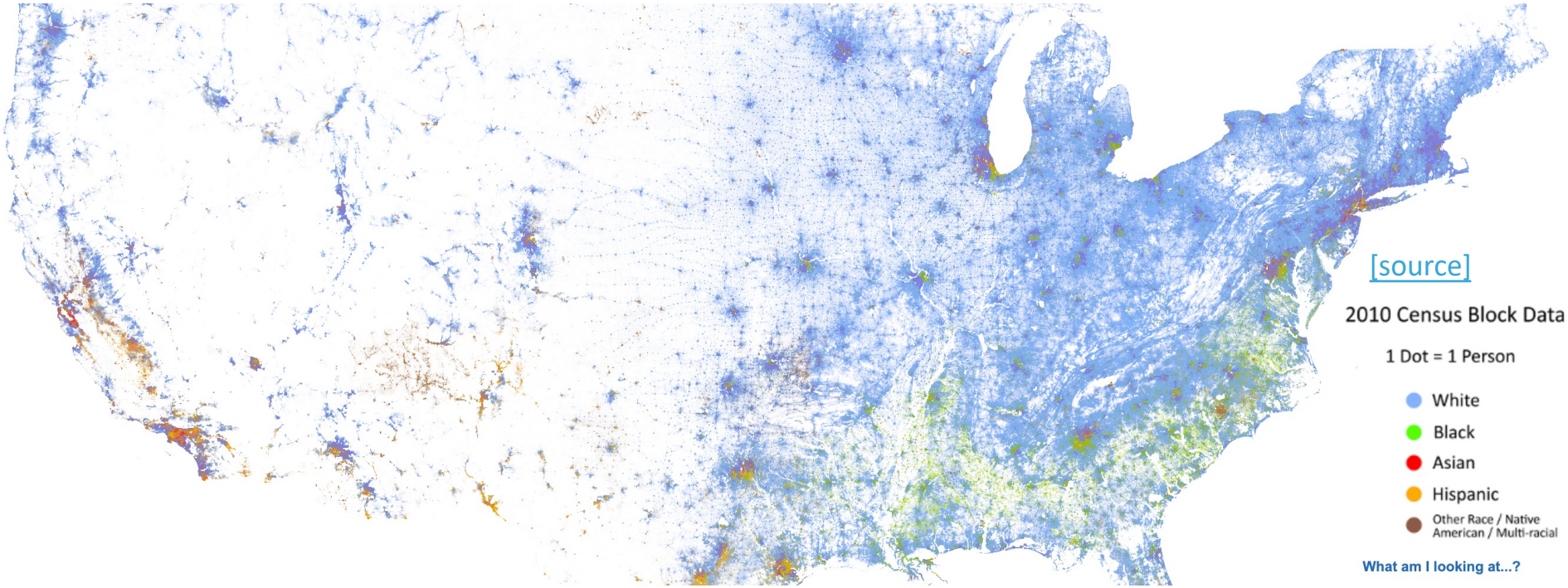
# Introduction to *Urban* Data Science

## Exploring Space in Data

(EPA1316A)

Lecture 8

Trivik Verma




The Racial Dot Map developed by the Demographics Research Group at University of Virginia

# Last Time

- Introduction to Networks
- The need to represent space formally
- Spatial weights matrices
  - What
  - Why
  - Types
- The spatial lag

# Today

- Exploratory Spatial Data Analysis (ESDA)
- Spatial Autocorrelation Measures
  - Global
  - Local

## [Exploratory]

Focus on discovery and assumption-free investigation

## [Spatial]

Patterns and processes that put space and geography at the core

## [Data Analysis]

Statistical techniques

# Questions that ESDA helps with…

Patel, R., Verma, T., Marvuglia, A., Huang, Y., Baustert, P., Shivakumar, A., Nikolic, I. (2021). Quantifying the Consumption-driven Environmental Impact of Households in Cities. In Preparation (In Preparation).

## Answer

- Is the variable I'm looking at concentrated over space?
- Do similar values tend to locate close by?
- Can I identify any particular areas where certain values are clustered?

## Ask

- What is behind this pattern?
- What could be generating the process?
- Why do we observe certain clusters over space?

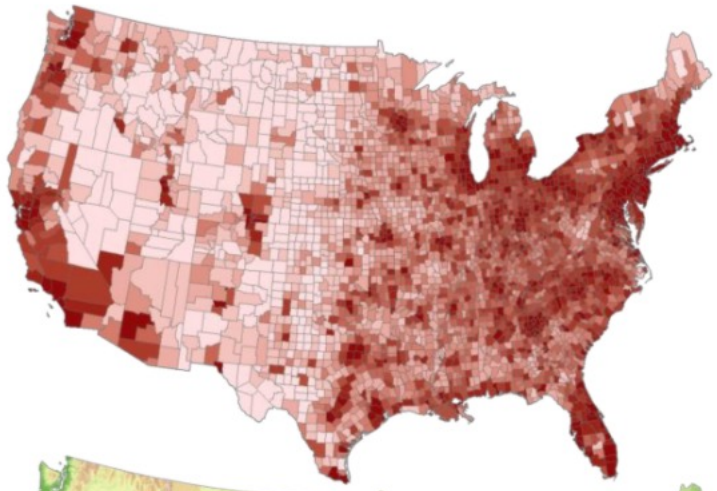

*Net emission reduction in the mobility sector for different neighbourhoods of the Hague under different car parking charging policies ceteris paribus*

The first law of geography:

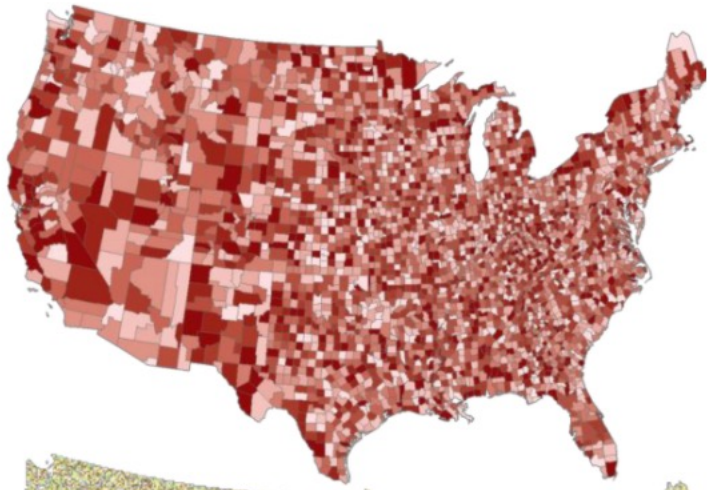
*"Everything is related to everything else, but near things are more related than distant things."*

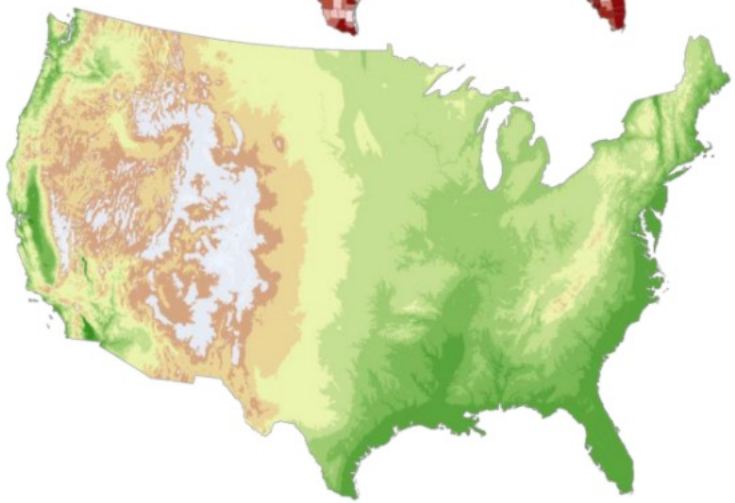
Waldo R. Tobler (Tobler 1970)

If features were randomly distributed



population density map of the US

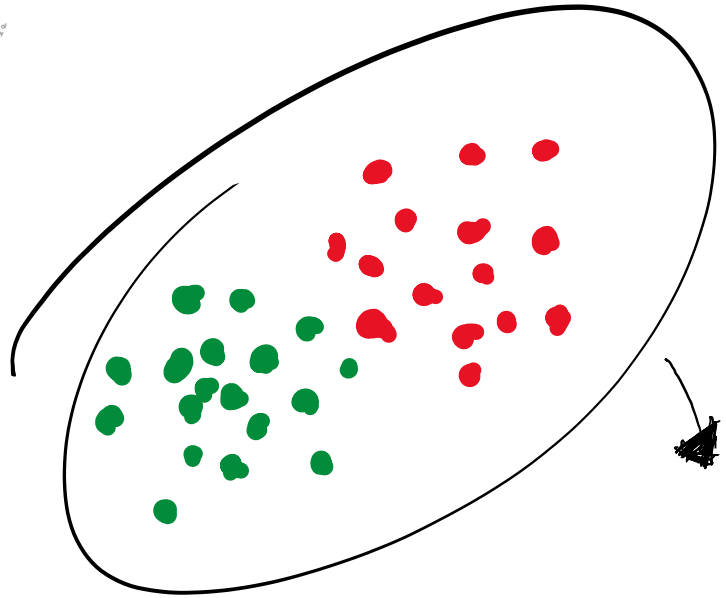




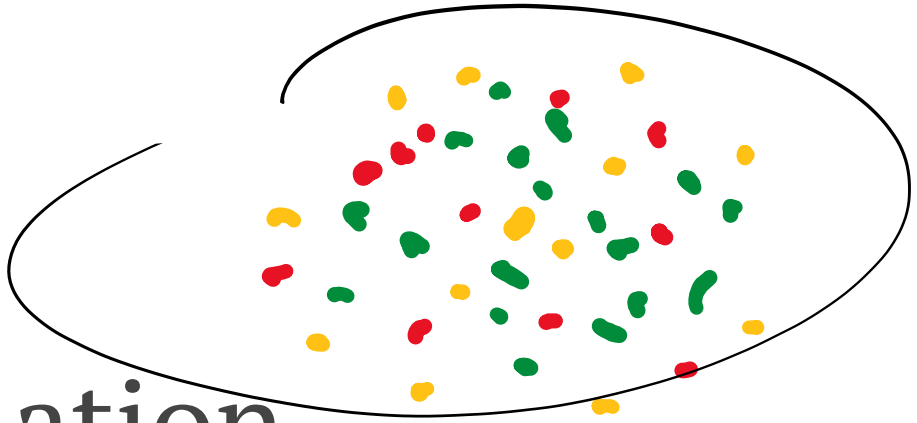
elevation map of the US

Images adapted from: Gimond, Manuel. Intro to GIS and Spatial Analysis.

HOW ARE FEATURES CLUSTERED?



Clustered



non-clustered regions

# Spatial Autocorrelation

1. Quantitative
2. Objective
3. Degree of similarity
4. where does it occur?

# Spatial Autocorrelation

- Statistical representation of Tobler's law
- Spatial counterpart of traditional correlation

***Degree to which similar values are located in similar locations***

# Spatial Autocorrelation

Two flavours:

- Positive: similar values → similar location *(close by)*
- Negative: similar values → dissimilar location *(further apart)*

# Examples

**Positive** SA: income, poverty, vegetation, temperature…

**Negative** SA: supermarkets, police stations, fire stations, hospitals…

# Scales

[Global] Clustering: do values tend to be close to other (dis)similar values?

[Local] Clusters: are there any specific parts of a map with an extraordinary concentration of (dis)similar values?

# Global Spatial Autocorrelation

# Global Spatial Autocorr.

"Clustering"

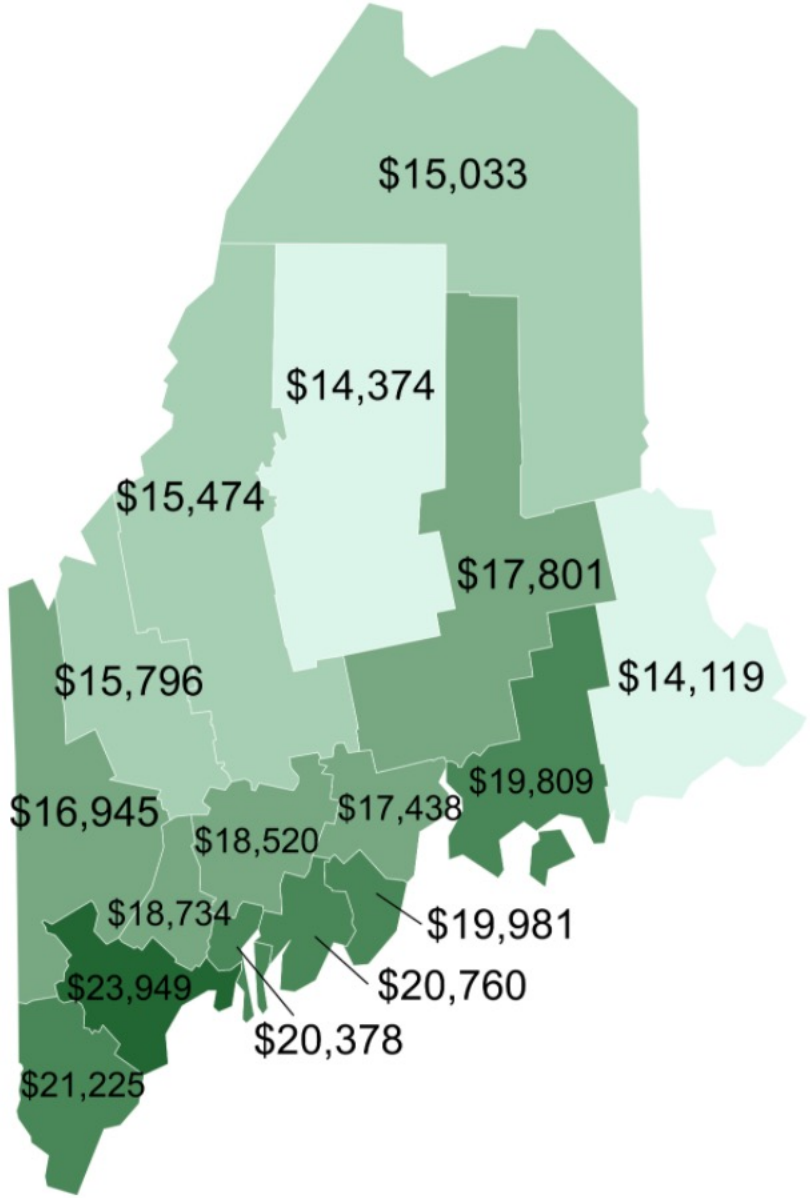
*Overall trend where the distribution of values follows a particular pattern over space*

**[Positive]** Similar values close to each other (high-high, low-low)

**[Negative]** Similar values far from each other (high-low)

How to measure it???

Let's start with a working example: 2010 per capita income for the state of Maine.

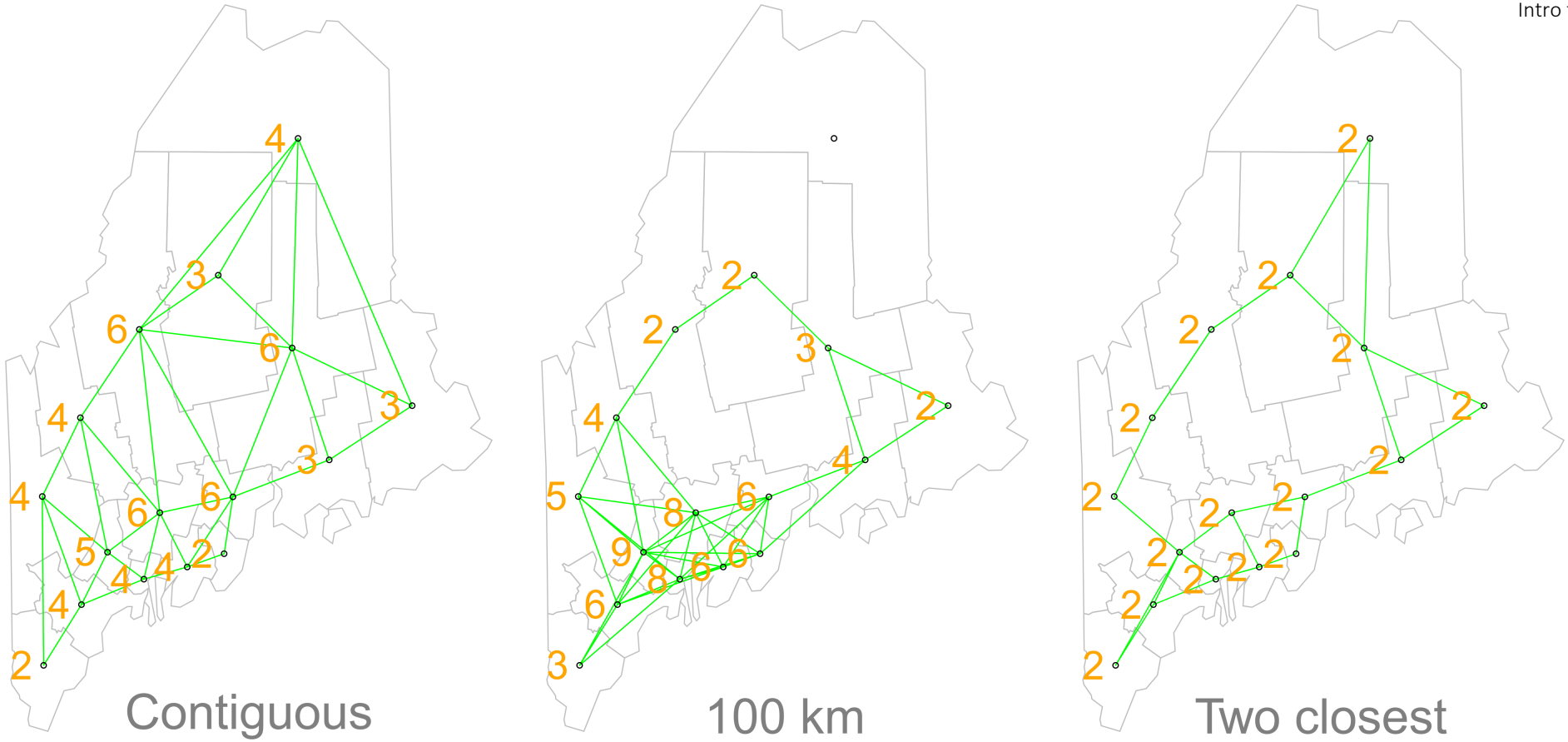



Images adapted from: Gimond, Manuel.
Intro to GIS and Spatial Analysis.

# Moran Plot

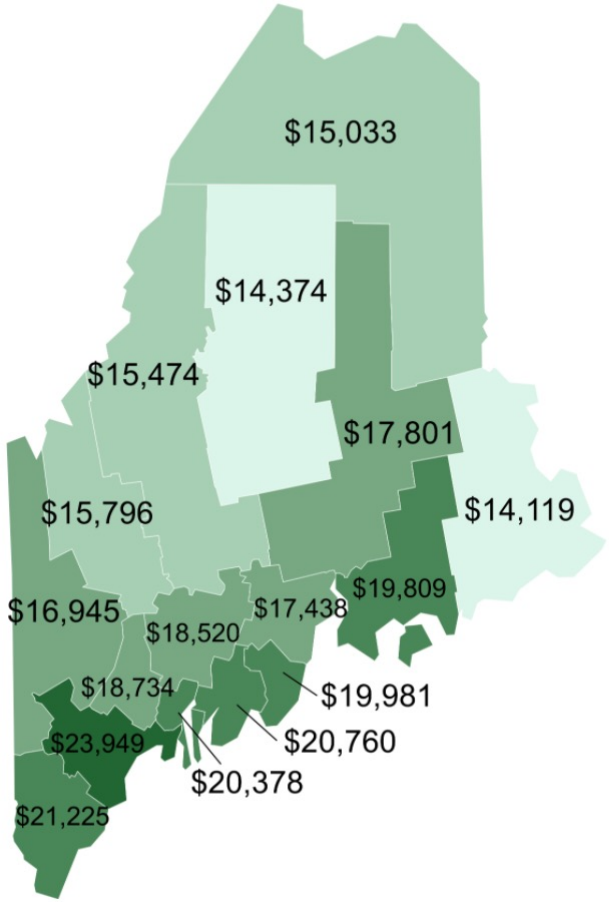
- Graphical device that displays a **variable** on the horizontal axis against **its spatial lag (Y_il – previous lecture)** on the vertical one
- Variable and spatial weights matrix are preferably standardized
- Assessment of the overall association between a variable in each location and, in its *neighbourhood*

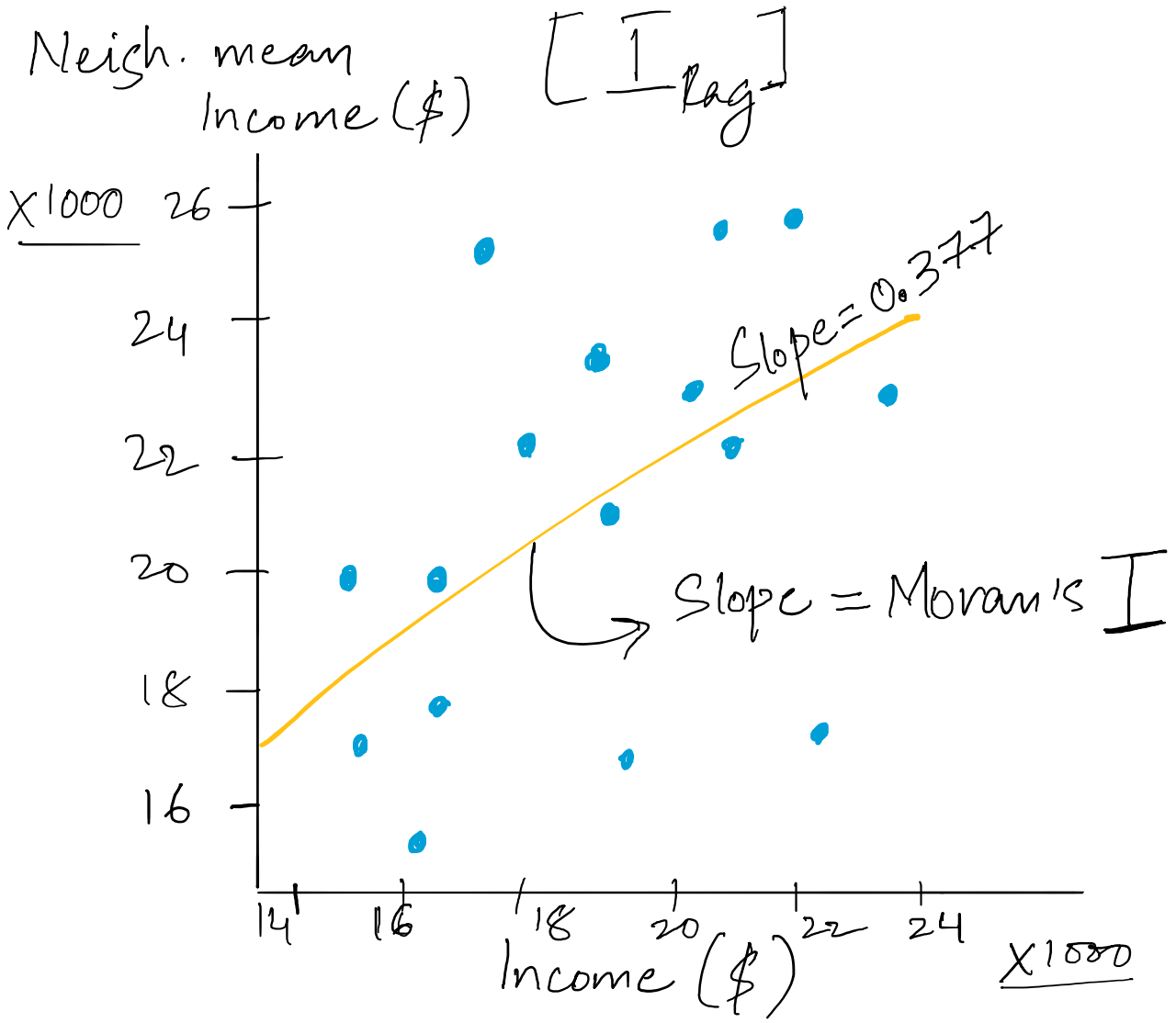
Images adapted from: Gimond, Manuel.
Intro to GIS and Spatial Analysis.




Maps show the links between each polygon and their respective neighbour(s) based on the neighbourhood definition. A contiguous neighbour is defined as one that shares a boundary or a vertex with the polygon of interest. Orange numbers indicate the number of neighbours for each polygon. Note that the top most county has no neighbours when a neighbourhood definition of a 100 km distance band is used (i.e. no centroids are within a 100 km search radius)

Let's start with a working example: 2010 per capita income for the state of Maine.

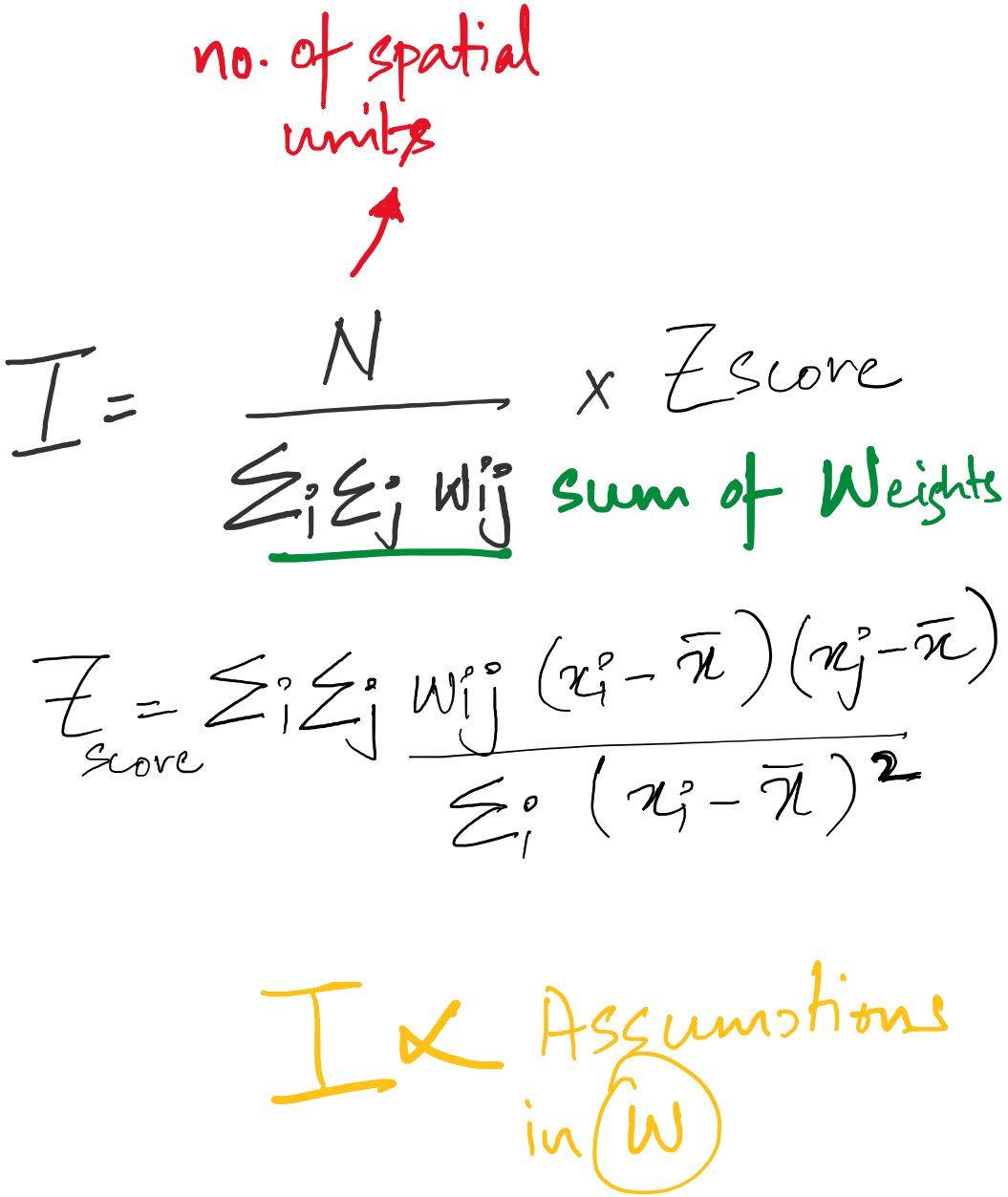


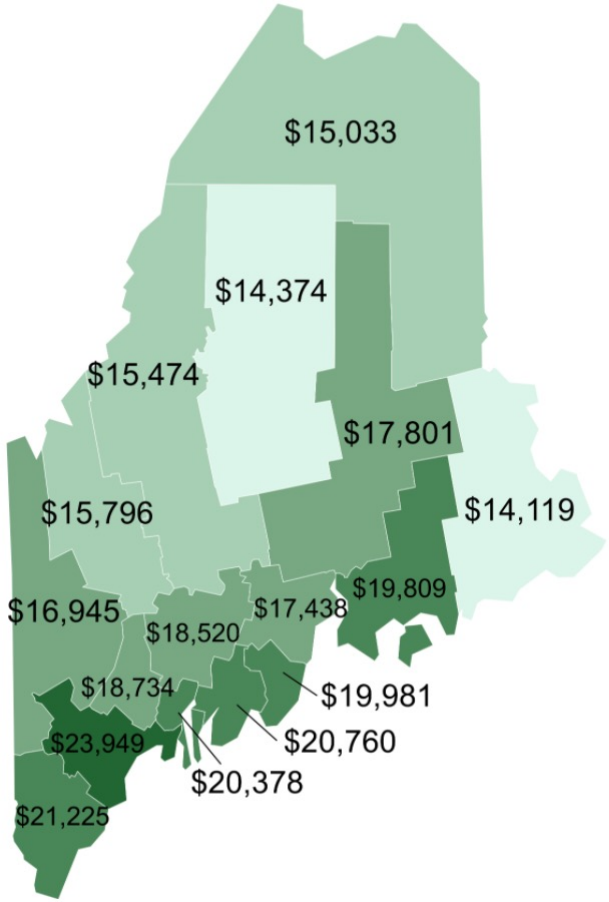

Images adapted from: Gimond, Manuel.
Intro to GIS and Spatial Analysis.

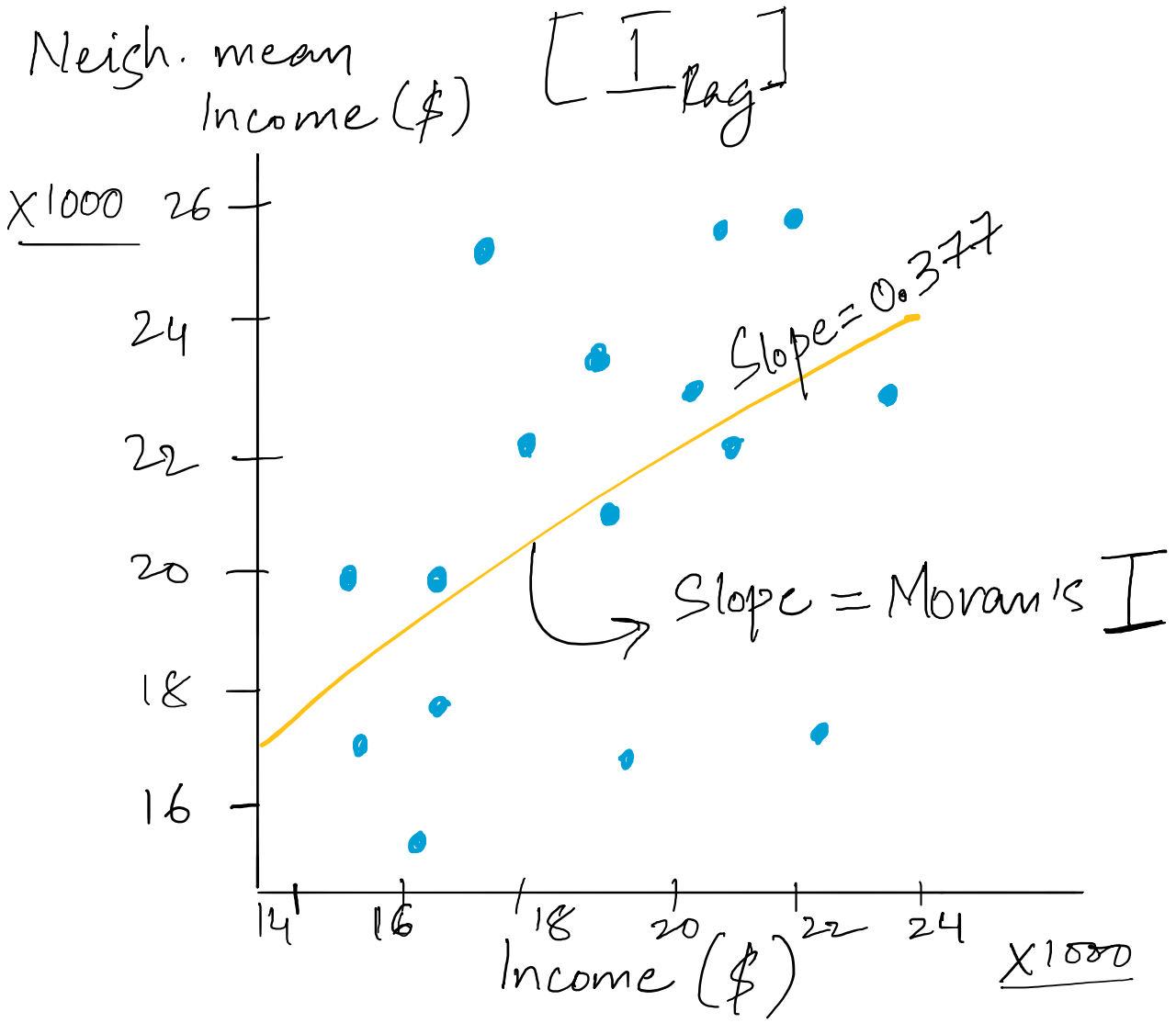
# Moran's I

- Formal test of global spatial autocorrelation
- Statistically identify the presence of clustering in a variable
- Slope of the Moran plot
- Inference based on how likely it is to obtain a map like the observed one from a purely random pattern



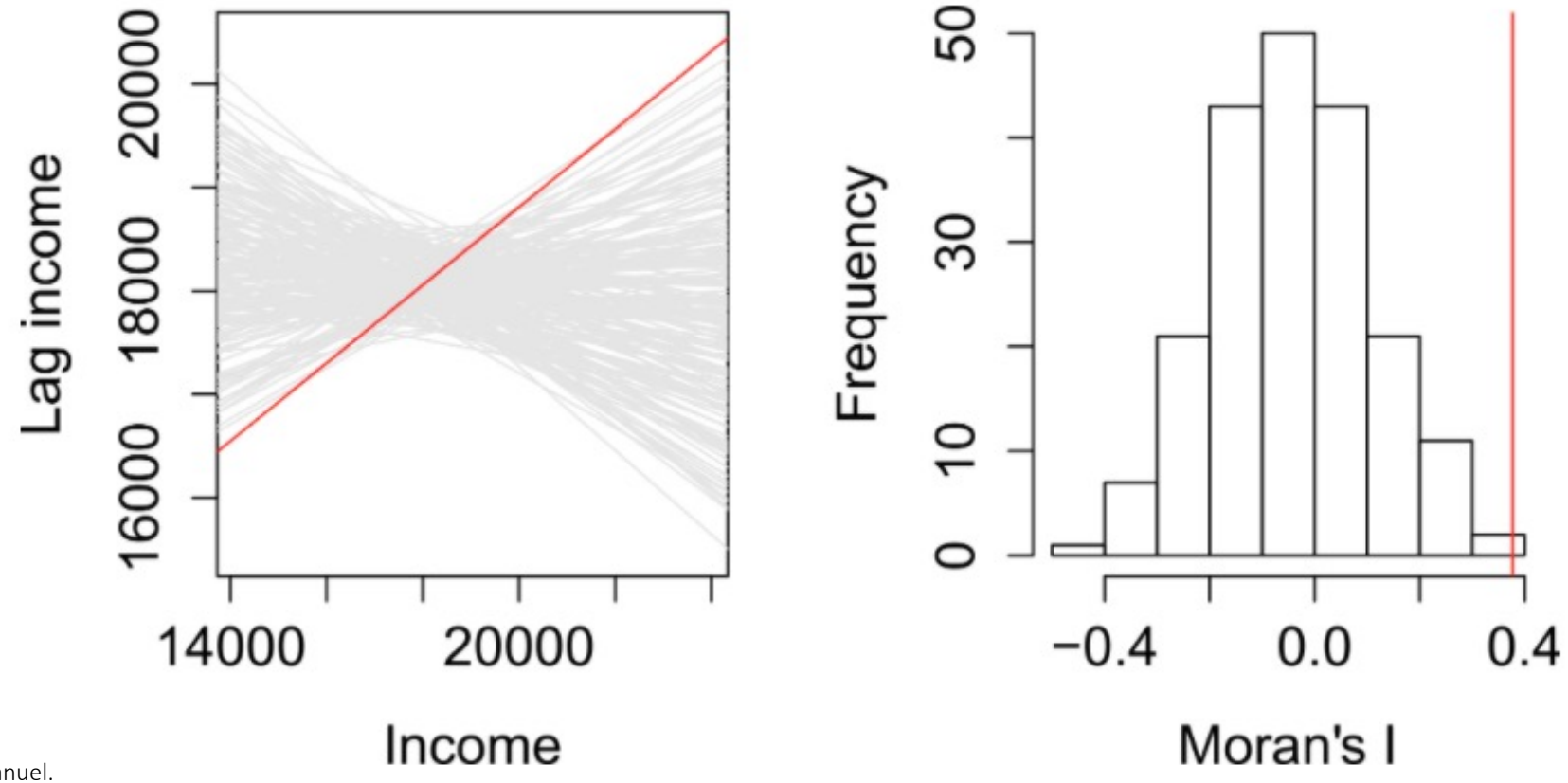
Let's start with a working example: 2010 per capita income for the state of Maine.



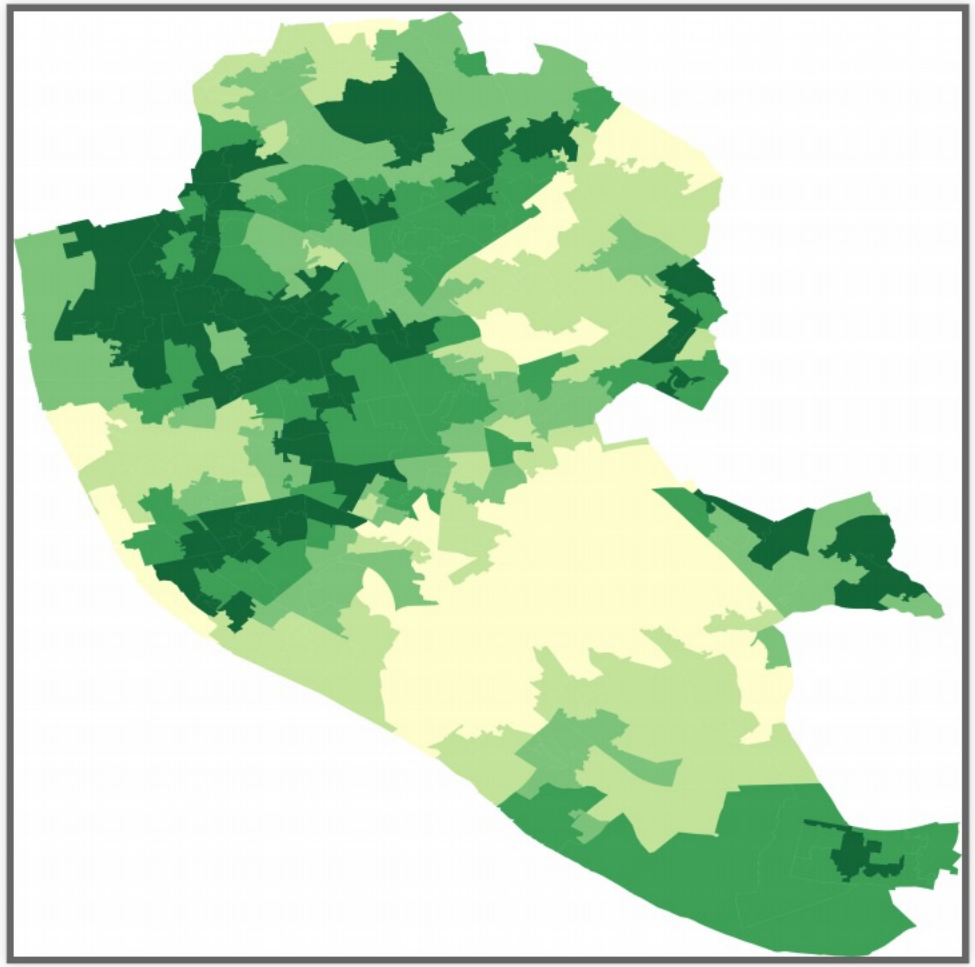

Images adapted from: Gimond, Manuel.
Intro to GIS and Spatial Analysis.

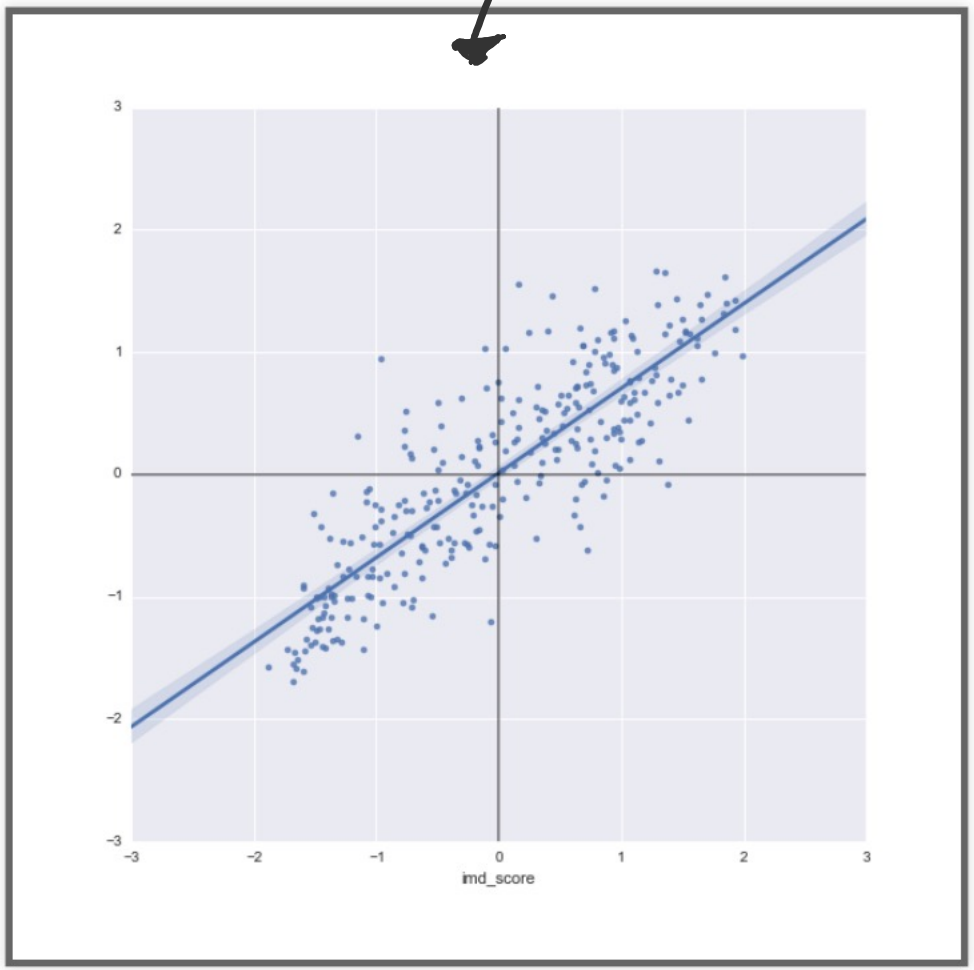
# How significant is this I statistic?




Images adapted from: Gimond, Manuel.
Intro to GIS and Spatial Analysis.

from the lab exercises






Images taken from: Arribas-Bel, D. (2019). A course on geographic data science. *Journal of Open Source Education*, *2*(16), 42.

# Local Spatial Autocorrelation

# Local Spatial Autocorr.

"Clusters"

*Pockets of spatial instability*

Portions of a map where values are correlated in a particularly strong and specific way

**[High-High]** + SA of high values (hotspots)

**[Low-Low]** + SA of low values (coldspots)

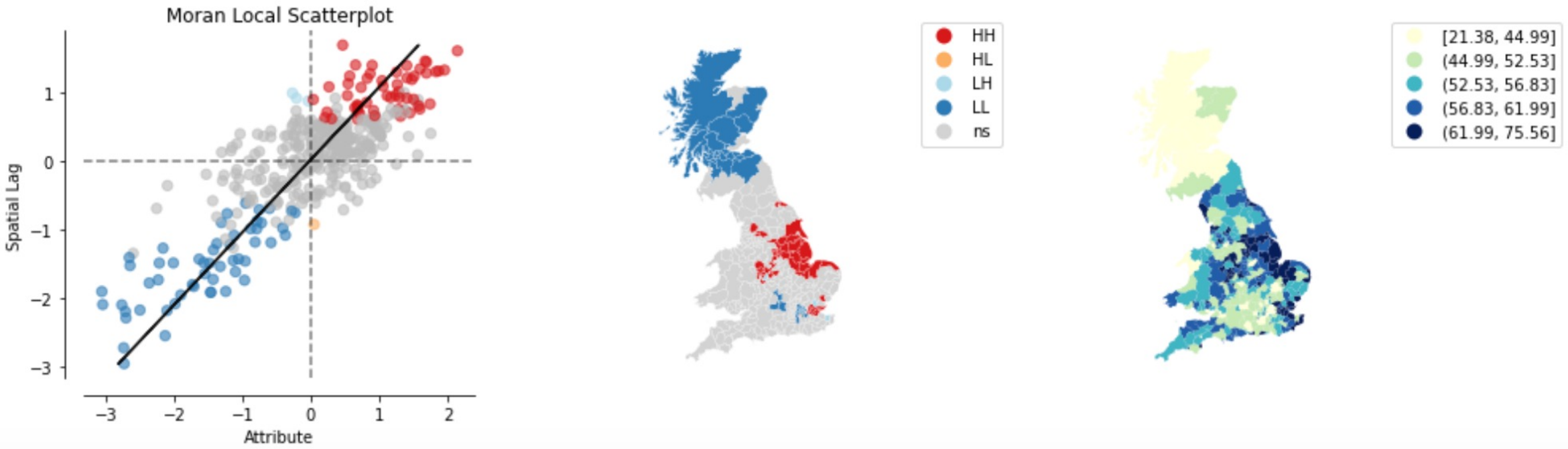
**[High-Low]** - SA (spatial outliers)

**[Low-High]** - SA (spatial outliers)

# What is LISA?

**L**ocal **I**ndicators of **S**patial **A**ssociation

- Statistical tests for ***spatial cluster detection*** → Statistical significance
- **Compares** the **observed** map with many **randomly** generated ones to see how likely it is to obtain the areas of unusually high concentration




Images taken from: Arribas-Bel, D. (2019). A course on geographic data science. *Journal of Open Source Education*, *2*(16), 42.

# Recapitulation

**ESDA** is a family of techniques to explore and spatially interrogate data

Main function: characterise **spatial autocorrelation**, which can be explored:

- **Globally** (e.g. Moran Plot, Moran ' s I)
- **Locally** (e.g. LISAs)

# For next class..

**Finish** Labs to practice programming

**Complete** Homework and review your peers' work

**Check** Assignment contents and due date

**See** "To do before class" for next lecture (~ 1 hour of self study)

**Read** paper for **Discussion** session before next week (~ 1 hour)

**Post** questions on the **Discussion** forum on Brightspace