

GEOG0114: PRINCIPLES OF SPATIAL ANALYSIS

WEEK 1: INTRODUCTION TO SPATIAL ANALYSIS & DATA SCIENCE

Dr Anwar Musah (a.musah@ucl.ac.uk)

Lecturer in Social and Geographic Data Science
UCL Geography

Welcome to GEOG0114

Meet the Team



Anwar Musah
Lecturer in Social & Geographic
Data Science



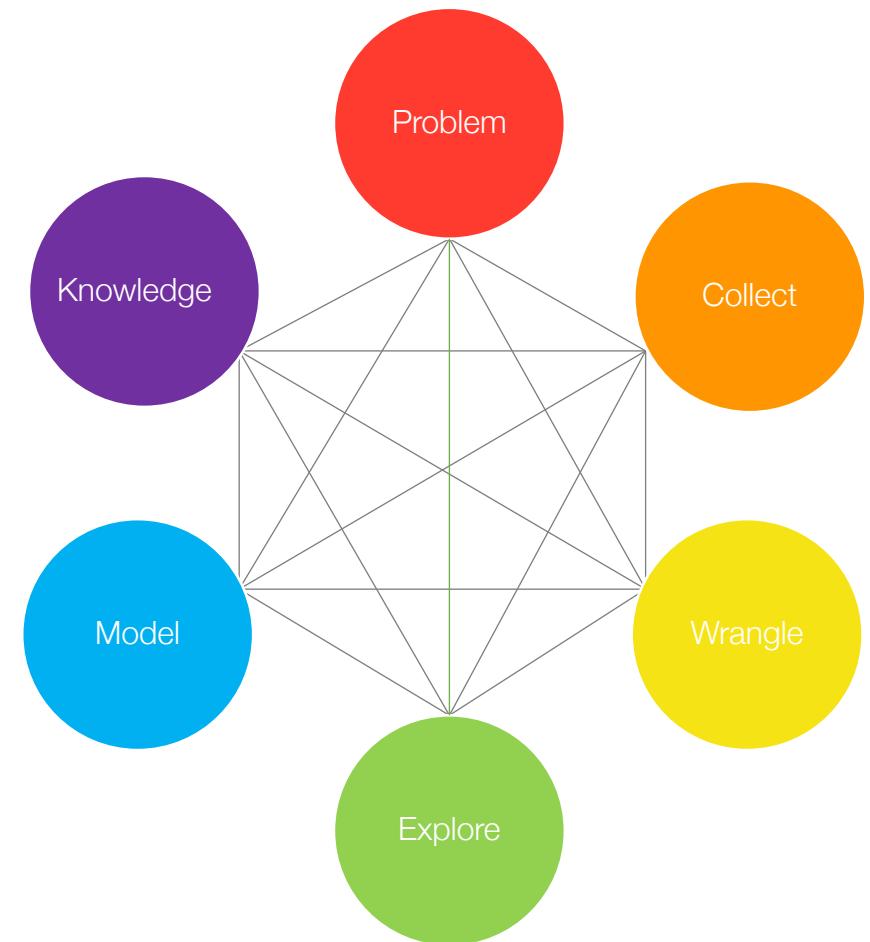
Justin van Dijk
Lecturer in Social & Geographic
Data Science



Shun-Chan Tsai
Post Graduate Teaching Assistant
PhD in UCL Geography

Contents

- Timetable & Assessment
- What will we learn?
- The Beginning: Introduction to Spatial Analysis & Data Science



Timetable and Assessment

Timetable

Lecture: North-West Wing, Room GO7
(Monday: 02:00pm-03:00pm)

Computer practical sessions*: North-West Wing, Room GO7
(Monday: 03:00pm-05:00pm)

Technical support sessions*: Institute of Education (IoE), Bedford Way 20, Room 604, (Thursday: 03:00pm-05:00pm) [Drop-in]

***IMPORTANT NOTE:** Please bring your own laptops with you to the computer practicals and technical support sessions on Monday and Thursday, respectively

Coursework & Assessment

- What is the assessment?
 - Spatial data analysis project which is worth 100%
 - It is a report (max 3,000 words) + code (R).
- The coursework will be released on Reading Week (after Week 5)
- The deadline for the coursework is **Monday 9th January**. The submission instructions will be available on Moodle.

What will we learn?

What will we learn?

- To introduce you to key principles of spatial statistics
- Provide an introduction and knowledge of methods for exploring various types of spatial data (i.e., point, areal, line segments and gridded/pixelated data)
- You will know how to adopt various spatial analytical techniques for testing out hypothesis, and for addressing problems related to social phenomena and its spatial components.
- You will learn to how to apply various families of spatial models (e.g., geographic weighted regressions, Global & Local Moran's I, Kriging and many more) for making spatial predictions and studying patterns of associations between risk factors and outcomes
- You will gain programming skills for carry out data managing, geoprocessing and analysis of spatial data using the software package R/RStudio. You will gain expert knowledge on how to use R/RStudio as a GIS software to perform high-level map visualization

Structure of course

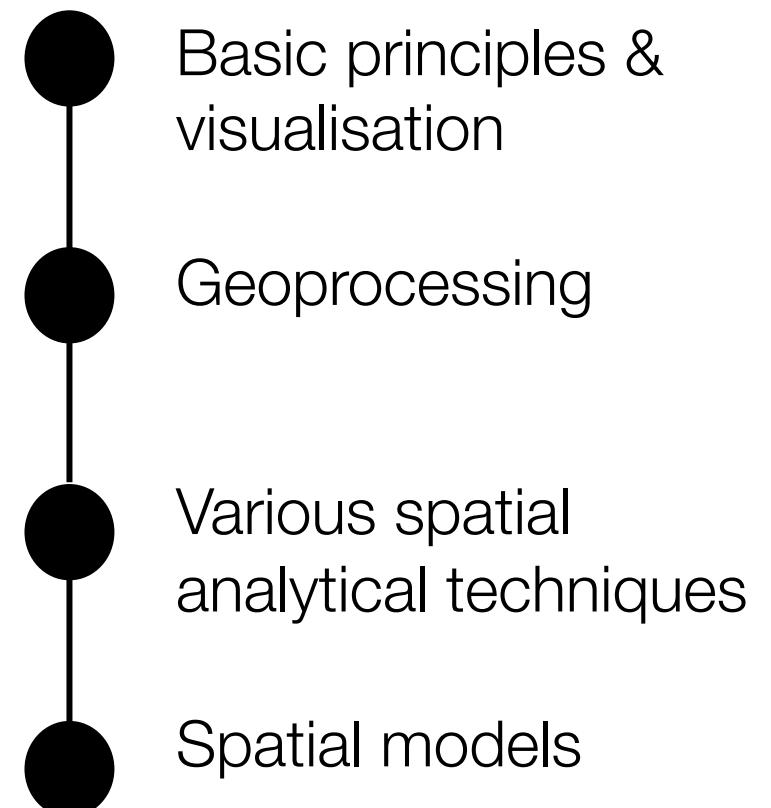
The first part of the module provides foundational knowledge for performing geospatial data analysis with various types of spatial data. In the second part of the module, we will focus on key concepts and methods for dealing gridded data using Raster-based analysis. The third part of the module, we will introduce you to more specialized techniques with real-world applications; and lastly, we will various spatial modelling techniques for evidence-based research.

Part One (Week 1-3): Foundation and theory (i.e., representation of spatial data, geoprocessing and spatial autocorrelation)

Part Two (Week 4-6): Raster-based analysis (i.e., Suitability mapping e.g., AHPs and Niches; and geostatistical modeling using Kriging)

Part Three (Week 7 and 8): Specialized techniques (e.g., Geodemographics and Transport Network Analysis)

Part Four (Week 9 and 10) : Spatial regression-based models (e.g., Spatial lag and error models, and Geographically weighted regression models)

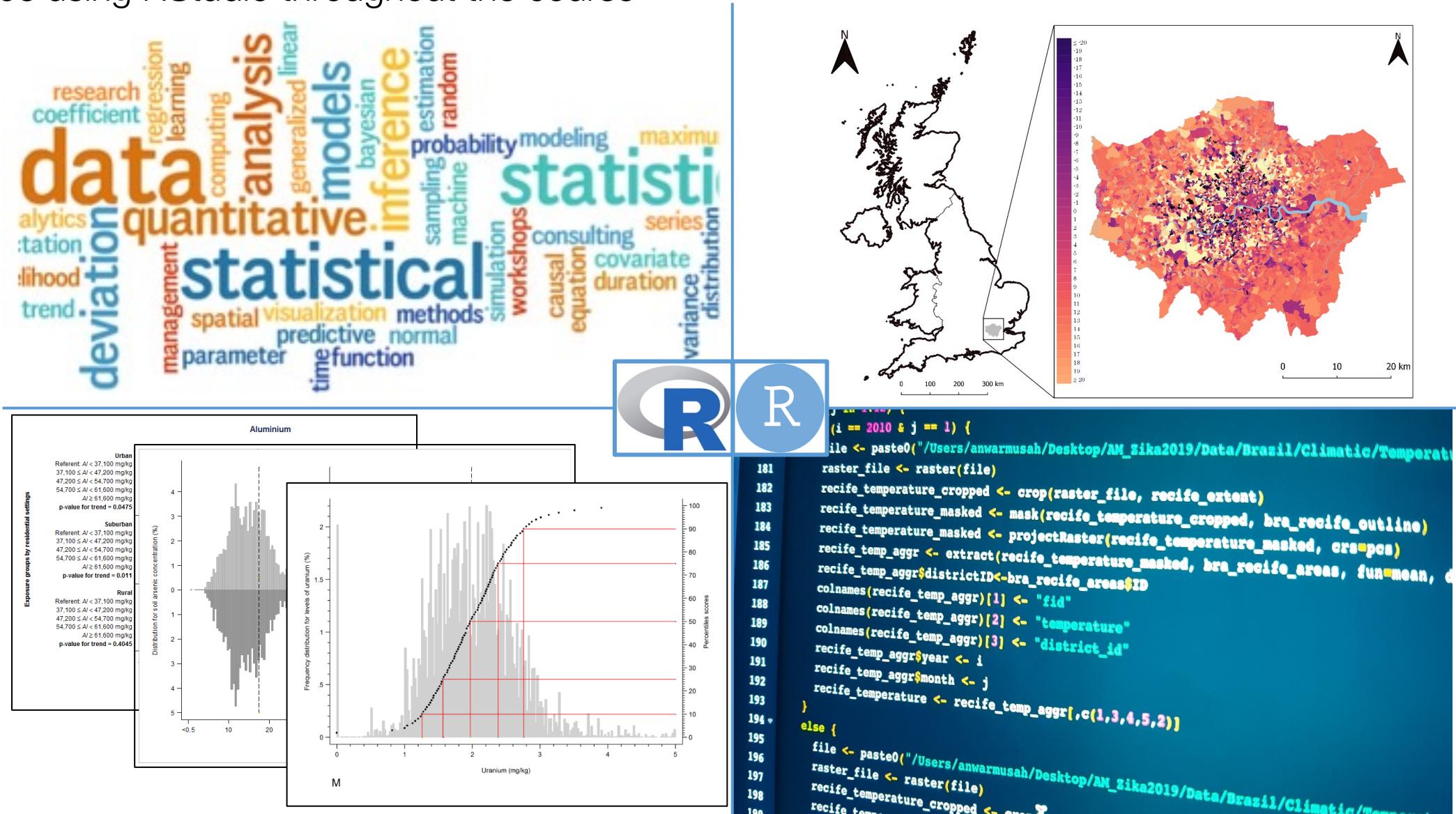


Breakdown of what will we be learning

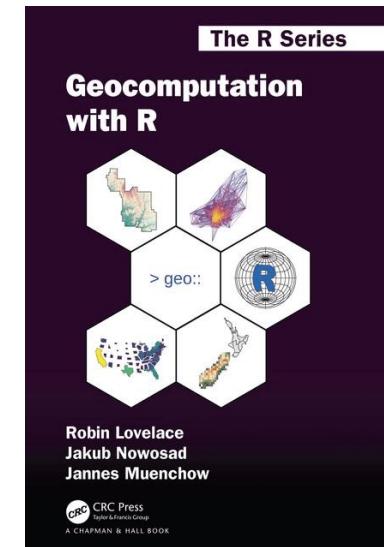
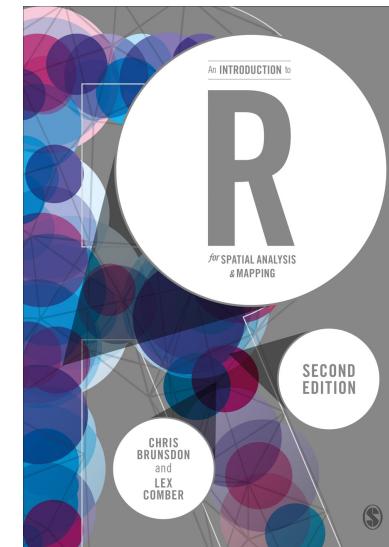
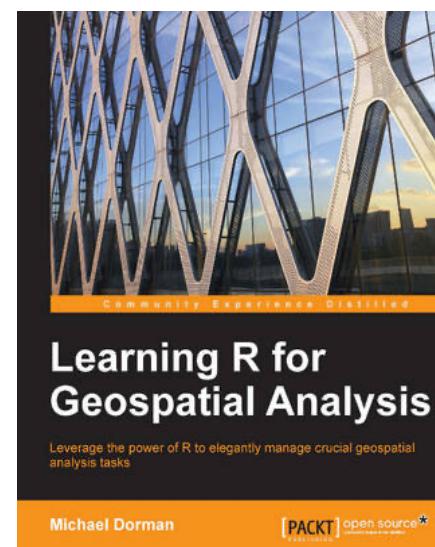
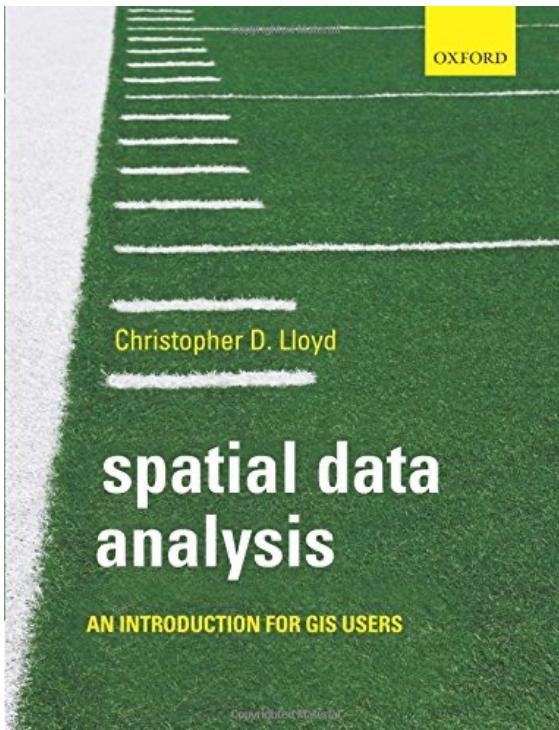
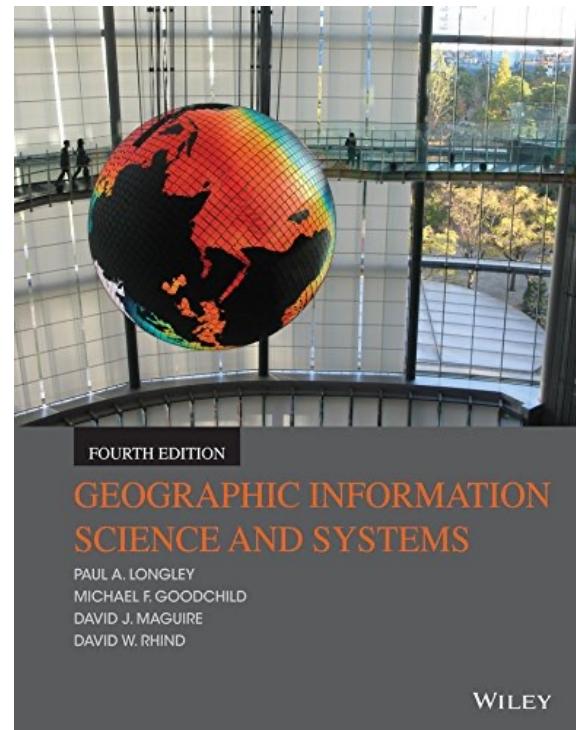
Week	Date	Lecturer	PGTA	Weekly Topics
1	03/10	AM	SCT	Spatial Analysis for Data Science
2	10/10	AM	SCT	Graphical Representation of Spatial Data
3	17/10	AM	SCT	Spatial Autocorrelation
4	24/10	AM	SCT	Suitability Mapping: Part I (Qualitative approach)
5	31/10	AM	SCT	Suitability Mapping: Part II (Quantitative approach)
Reading Week (Release of Assessment)				
6	14/11	AM	SCT	Geostatistics using Kriging
7	21/11	JVD	SCT	Geodemographics
8	28/11	JVD	SCT	Transport Network Analysis
9	05/12	AM	SCT	Spatial Models: Part I (Spatial Lag & Error Models)
10	12/12	AM	SCT	Spatial Models: Part II (Geographically Weighted Regression)

AM: Anwar Musah | JVD: Justin Van Dijk | SCT: Shun-Chan (Dan) Tsai

We will be using RStudio throughout the course



Recommended Books for GEOG0114



High recommendation for the coding experience and execution of spatial analysis in R

High recommendation for the mastery of basic theory and principles of spatial analysis

Home Events My courses This course Staff Help Student Help Services

Welcome Keeping In Touch Module Overview Assessment Lecturecast - video recordings of classes

Module Structure

This module consists of ten lectures, and ten interactive seminars designed as computer lab sessions. All lectures and computer practicals will be delivered in-person. All Lectures are held on **Monday from 02:00pm to 03:00pm** at the **North West Wing (Room G07)**. All computer lab seminars are delivered on Monday (after lectures) from 03:00pm to 05:00pm in the same room i.e., **North West Wing (Room G07)**. Technical support workshops are held on **Thursday from 03:00pm to 05:00pm** at the **Institute of Education (IOE) (Bedford Way [20])(Room 604)**.

While Moodle is the central point for information and communications on GEOG0114; please note that all the week's content including the lecture notes, the practical materials for seminars (i.e., instructions and downloadable data sets), as well as the reading lists for that week are hosted on this [GEOG0114 GitHub website](#).

Weekly Topics

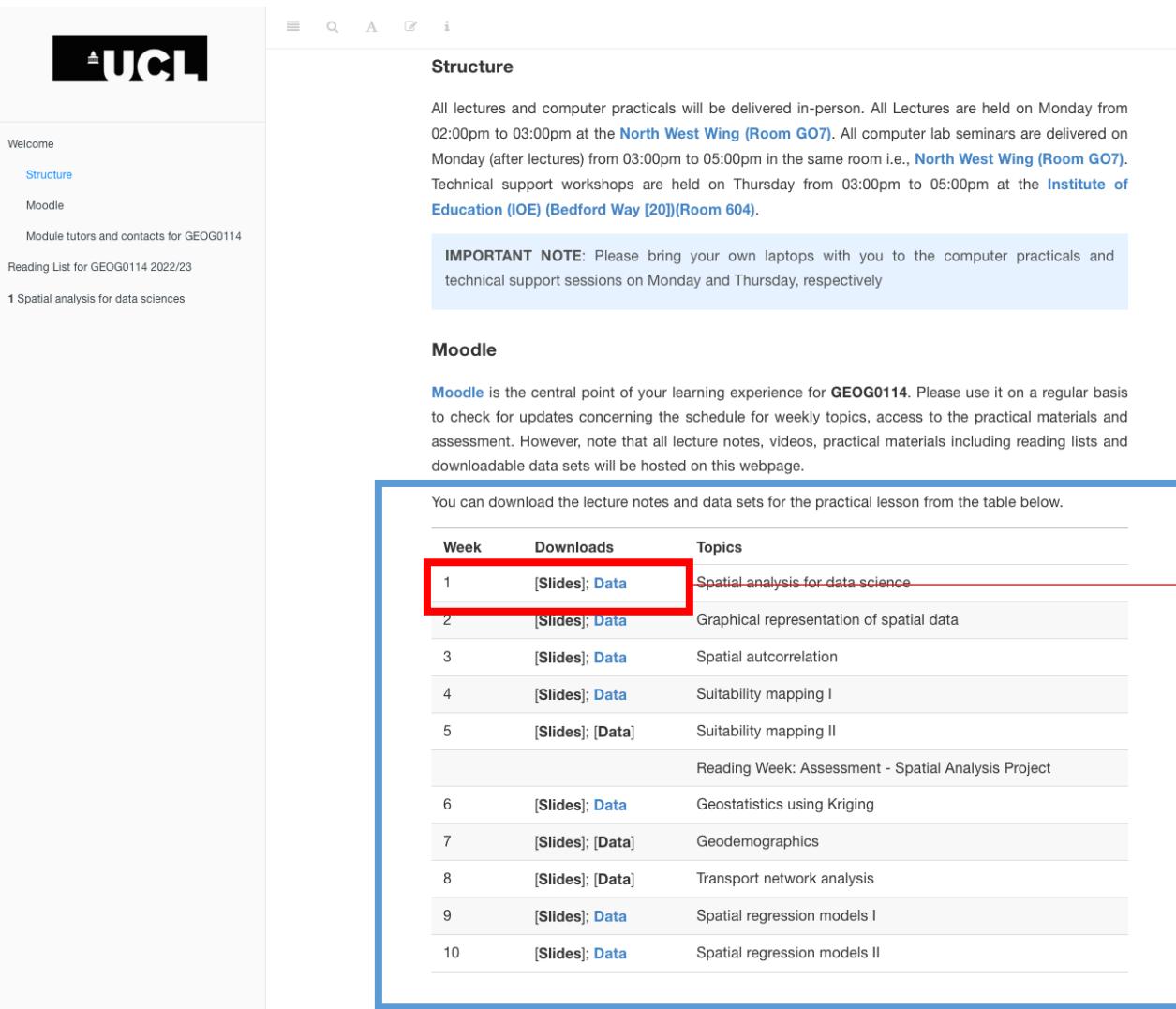
The topics for the module are scheduled as follows:

w/c	Week	Topic	Lecturer
3rd Oct	1	Spatial analysis for data science	Anwar Musah
10th Oct	2	Graphical representation of spatial data	Anwar Musah
17th Oct	3	Spatial autocorrelation	Anwar Musah
24th Oct	4	Suitability Mapping I (Qualitative approach)	Anwar Musah
31st Oct	5	Suitability Mapping II (Quantitative approach)	Anwar Musah
7th Nov		- Reading Week -	
14th Nov	6	Geostatistics using Kriging	Anwar Musah
21st Nov	7	Geodemographics	Justin Van Dijk
28th Nov	8	Transport network analysis	Justin Van Dijk
5th Dec	9	Spatial models I (Spatial Lag & Error Regression)	Anwar Musah
12th Dec	10	Spatial models II (Geographically Weighted Regression)	Anwar Musah

Reading List and All Teaching Materials

All week's content which includes the lecture notes, the practical materials for seminars (i.e., instructions and downloadable data sets), as well as the reading lists for that week are hosted on this [GEOG0114 GitHub website](#).

- Moodle for GEOG0114 is central point for course and assignment information [<https://moodle.ucl.ac.uk/>]
- Module Overview** section is where you can see the timetable and access **GEOG0114 GitHub website**
- Please click on any of these links to be teleported to the **GEOG0114 GitHub website**.
- Here is where the practical and lecture note materials are hosted. Please download as soon as you gain access.
- The new lecture notes and practical materials will be released on Thursdays before next week's Monday's practical session



The screenshot shows the 'Structure' section of the GEOG0114 module page. It includes a navigation bar with links for Welcome, Structure (which is active), Moodle, Module tutors and contacts for GEOG0114, Reading List for GEOG0114 2022/23, and Spatial analysis for data sciences. The main content area starts with a heading 'Structure' and a paragraph about lecture and practical delivery. It includes an 'IMPORTANT NOTE' box about bringing laptops. Below is a 'Moodle' section with a paragraph about Moodle being the central learning point. A table lists weekly download links and topics. At the bottom is a 'Moodle tutors and contacts for GEOG0114' link.

Structure

All lectures and computer practicals will be delivered in-person. All Lectures are held on Monday from 02:00pm to 03:00pm at the [North West Wing \(Room G07\)](#). All computer lab seminars are delivered on Monday (after lectures) from 03:00pm to 05:00pm in the same room i.e., [North West Wing \(Room G07\)](#). Technical support workshops are held on Thursday from 03:00pm to 05:00pm at the [Institute of Education \(IOE\) \(Bedford Way \[20\]\) \(Room 604\)](#).

IMPORTANT NOTE: Please bring your own laptops with you to the computer practicals and technical support sessions on Monday and Thursday, respectively

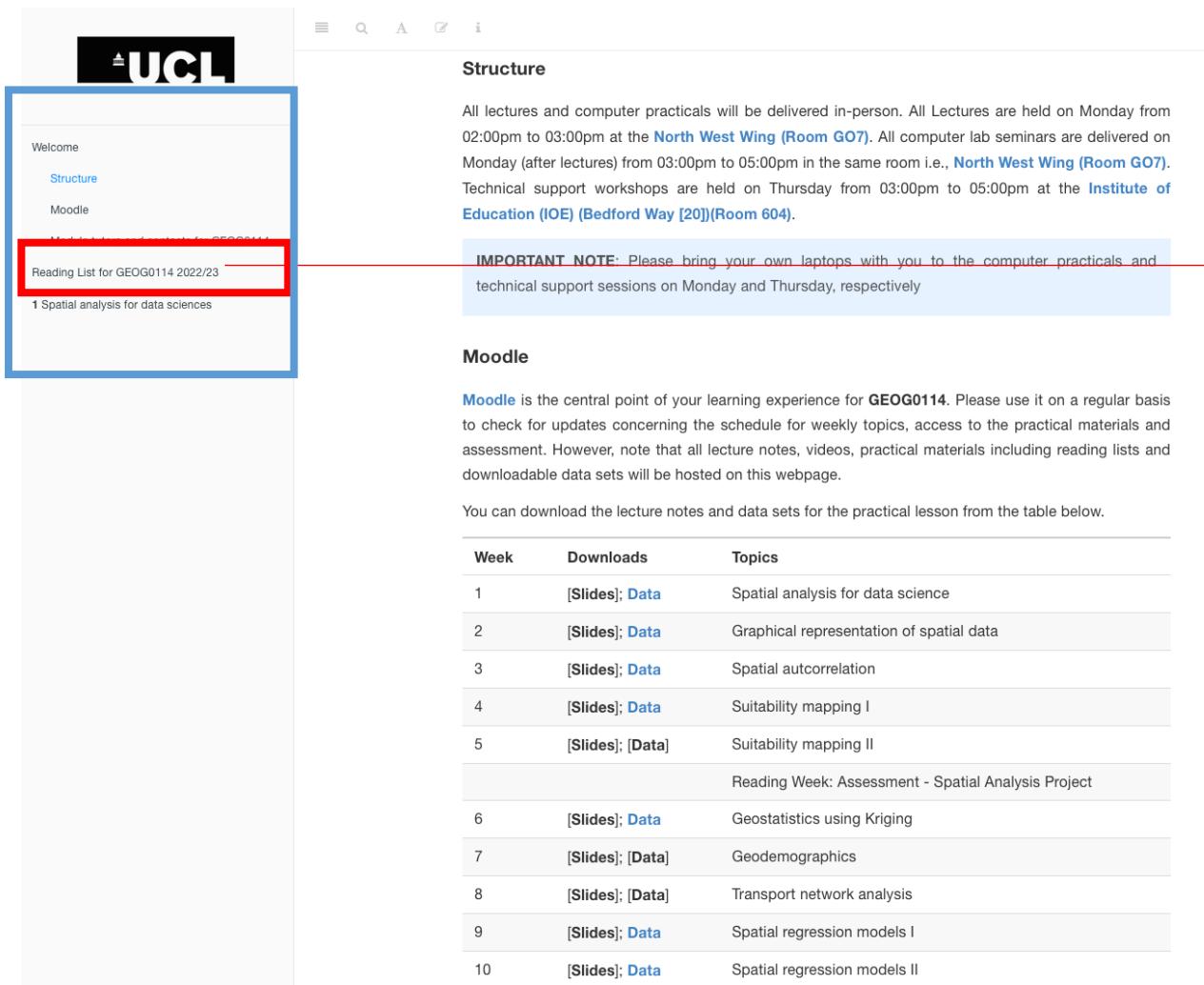
Moodle

Moodle is the central point of your learning experience for **GEOG0114**. Please use it on a regular basis to check for updates concerning the schedule for weekly topics, access to the practical materials and assessment. However, note that all lecture notes, videos, practical materials including reading lists and downloadable data sets will be hosted on this webpage.

Week	Downloads	Topics
1	[Slides]; Data	Spatial analysis for data science
2	[Slides]; Data	Graphical representation of spatial data
3	[Slides]; Data	Spatial autocorrelation
4	[Slides]; Data	Suitability mapping I
5	[Slides]; Data	Suitability mapping II
		Reading Week: Assessment - Spatial Analysis Project
6	[Slides]; Data	Geostatistics using Kriging
7	[Slides]; Data	Geodemographics
8	[Slides]; Data	Transport network analysis
9	[Slides]; Data	Spatial regression models I
10	[Slides]; Data	Spatial regression models II

[Moodle tutors and contacts for GEOG0114](#)

- Download the lecture and practical materials for that week



The screenshot shows the 'Structure' section of the GEOG0114 module page. It includes a sidebar with links for Welcome, Structure, Moodle, and a red-highlighted 'Reading List for GEOG0114 2022/23'. The main content discusses lecture times and locations, and includes an 'IMPORTANT NOTE' about laptops and technical support sessions. Below this is a 'Moodle' section describing its role in the learning experience, followed by a table of weekly download links and topics.

Structure

All lectures and computer practicals will be delivered in-person. All Lectures are held on Monday from 02:00pm to 03:00pm at the [North West Wing \(Room GO7\)](#). All computer lab seminars are delivered on Monday (after lectures) from 03:00pm to 05:00pm in the same room i.e., [North West Wing \(Room GO7\)](#). Technical support workshops are held on Thursday from 03:00pm to 05:00pm at the [Institute of Education \(IOE\) \(Bedford Way \[20\]\) \(Room 604\)](#).

IMPORTANT NOTE: Please bring your own laptops with you to the computer practicals and technical support sessions on Monday and Thursday, respectively

Moodle

Moodle is the central point of your learning experience for **GEOG0114**. Please use it on a regular basis to check for updates concerning the schedule for weekly topics, access to the practical materials and assessment. However, note that all lecture notes, videos, practical materials including reading lists and downloadable data sets will be hosted on this webpage.

You can download the lecture notes and data sets for the practical lesson from the table below.

Week	Downloads	Topics
1	[Slides]; Data	Spatial analysis for data science
2	[Slides]; Data	Graphical representation of spatial data
3	[Slides]; Data	Spatial autocorrelation
4	[Slides]; Data	Suitability mapping I
5	[Slides]; [Data]	Suitability mapping II
		Reading Week: Assessment - Spatial Analysis Project
6	[Slides]; Data	Geostatistics using Kriging
7	[Slides]; [Data]	Geodemographics
8	[Slides]; [Data]	Transport network analysis
9	[Slides]; Data	Spatial regression models I
10	[Slides]; Data	Spatial regression models II

- Access to the reading list for GEOG0114 2022/23

Structure

All lectures and computer practicals will be delivered in-person. All Lectures are held on Monday from 02:00pm to 03:00pm at the [North West Wing \(Room G07\)](#). All computer lab seminars are delivered on Monday (after lectures) from 03:00pm to 05:00pm in the same room i.e., [North West Wing \(Room G07\)](#). Technical support workshops are held on Thursday from 03:00pm to 05:00pm at the [Institute of Education \(IOE\) \(Bedford Way \[20\]\) \(Room 604\)](#).

IMPORTANT NOTE: Please bring your own laptops with you to the computer practicals and technical support sessions on Monday and Thursday, respectively

Moodle

Moodle is the central point of your learning experience for **GEOG0114**. Please use it on a regular basis to check for updates concerning the schedule for weekly topics, access to the practical materials and assessment. However, note that all lecture notes, videos, practical materials including reading lists and downloadable data sets will be hosted on this webpage.

You can download the lecture notes and data sets for the practical lesson from the table below.

Week	Downloads	Topics
1	[Slides]; Data	Spatial analysis for data science
2	[Slides]; Data	Graphical representation of spatial data
3	[Slides]; Data	Spatial autocorrelation
4	[Slides]; Data	Suitability mapping I
5	[Slides]; Data	Suitability mapping II
		Reading Week: Assessment - Spatial Analysis Project
6	[Slides]; Data	Geostatistics using Kriging
7	[Slides]; Data	Geodemographics
8	[Slides]; Data	Transport network analysis
9	[Slides]; Data	Spatial regression models I
10	[Slides]; Data	Spatial regression models II

Module tutors and contacts for GEOG0114

- Access to the chapter that actual computer practicals

1 Spatial analysis for data sciences

1.1 Introduction

The goal for this week's session is to get you started with using RStudio, and being familiar with its environment. The session aims to introduce you to the basic programming etiquette, as well as building confidence for using RStudio as a GIS tool. At the end of this session, you should be able to perform some basic data managing tasks as well as generate a simple choropleth map in RStudio.

1.1.1 Learning outcomes #

The first task includes getting you started with RStudio by installing the needed software(s) (i.e., [RStudio](#) and [R \(Base\)](#)) on to your personal laptop, and becoming familiar with its environment and panels. We will begin a soft introduction on the basics of managing data in RStudio. This includes learning how to create various objects in RStudio such as [vector](#) and [data frame](#) objects. The crucial part of this session we be to know how to the set working directories as well as import your dataset in RStudio. Finally, we will learn how to perform the basic visualisation of spatial data in RStudio.

Let us begin.

1.2 Getting started with RStudio

What is RStudio (or R)?

R, or RStudio is a statistical software programming package that allows the user to carry out different types of statistical analysis. It can also be used as a GIS software to perform various kinds of spatial analysis as you will see throughout the course of GEOG0114. In the same vein, you can use it for data managing and geo-processing (i.e., importing different types of spatial formats for manipulation beforehand for spatial analysis). There are two versions:




The Beginning: Introduction to Spatial Analysis and Data Science

What is Spatial Statistics (or Analysis)?

Definition:

The field of Spatial Statistics (or Analysis) is built on the assertion that nearby geographical observations (or objects) are somewhat associated in some way in space.

- The field is interdisciplinary as it brings together theories from **statistics, geography, computer science** and integrates **evidence-based research** methodologies (i.e., study designs)

- Usage for describing spatial distribution of **areal/point outcomes, interactions** between objects in space, but also how **an object has an impact to other nearby objects** – and how such impact **decays with distance**.

How does traditional statistics differ from spatial analysis?

The benchmark in traditional statistical theory is the “**common**” assumption of **independence** about observations which states that:

“the value of an observation does not influence or affect the value of other observations”

Crude examples:

- Assessing the Body Mass Index (BMI) of new enrolled students of GEOG0114
- Average test scores from Year 10 mathematics classes from each school

The field of spatial statistics is based on assumption of dependence (or non-independence) of observations

Spatial statistical theory of assumption of **dependence (or non-independence)** states that:

“...closer observations or nearby units in space are in someway closely related (i.e., similar in characteristics or interact more) than those that are distant from each other...”

The assertion is the “**cornerstone**” of spatial statistics. It is called:

The First Law of Geography (Tobler, 1979)



Waldo Rudolph Tobler (1930 – 2018)

Tobler's First Law of Geography on “everything is related to everything else, but near things are more related to distant things”

This first law is the foundation of the fundamental concepts of spatial **dependence**. It is integrated to most of the families of models in spatial statistics

Very important concepts to keep in mind when you think about **spatial dependence**:

- **Distance decay**
- **Spatial autocorrelation**

What is **Distance decay** in the context of Tobler's theory?

“This refers to the distance between two events (or objects) and their degree of interaction in space as distance vary.”

In short, what distance decay (in the context of Tobler's theory) says:

- It dictates how objects interact. The larger the distance between two events in a space – the less is their interaction & *vice versa*

Examples of where Distance decay manifests in studies cities and urban space:

- As distance from the focal point of a city center increases – high population density, taller buildings, and accessibility to multiple modes of transport decreases
- Housing marketing – house prices decreases, and in turn, residential mobility increases from expensive to affordable areas

What is **Spatial autocorrelation** in the context of Tobler's theory?

"This simply refers to whether (or not) similar values cluster together over geographic space"

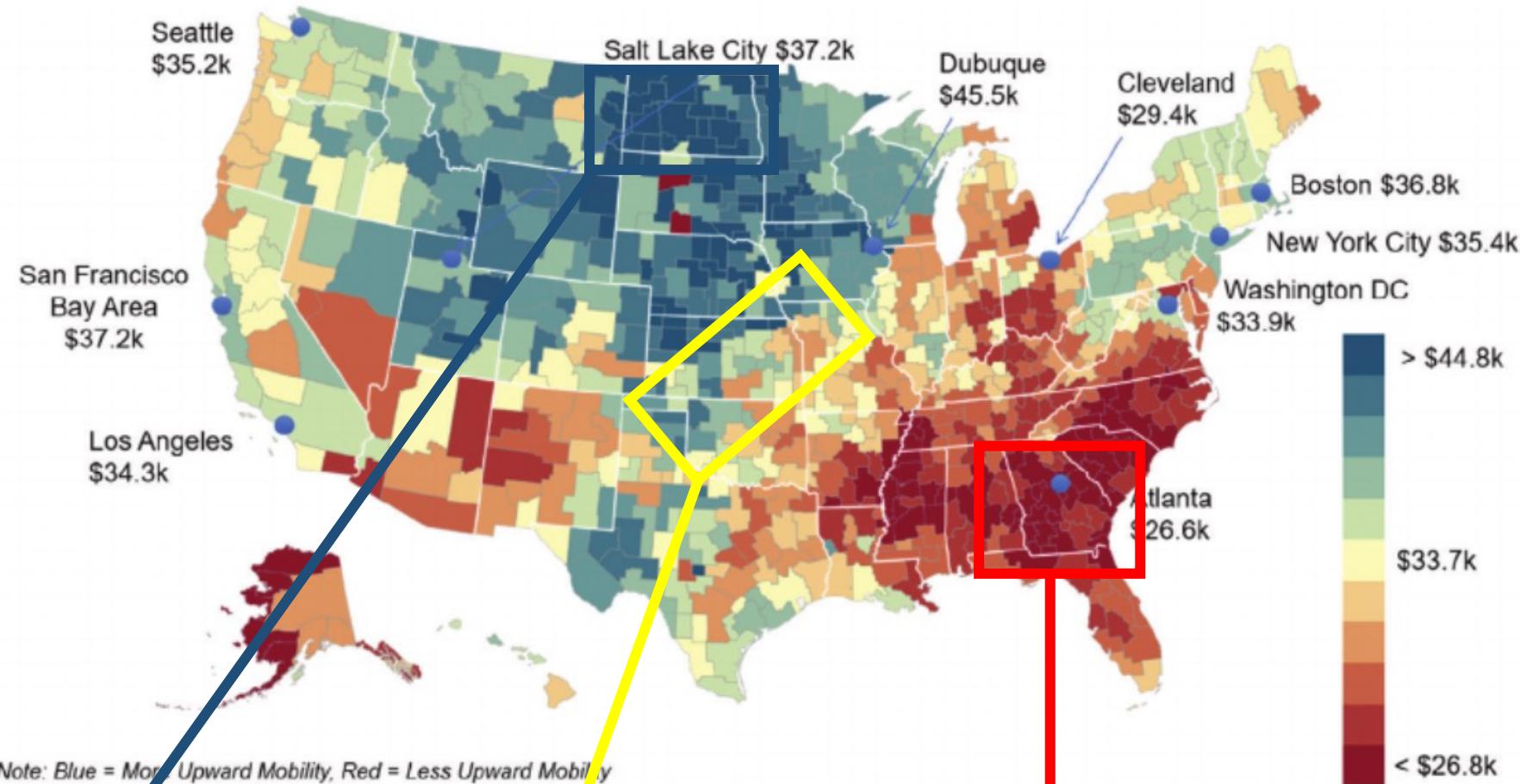
Etymology ("Auto" means self; and "correlation" means the degree of relative correspondence)

Similar values that cluster together are said to have **positive spatial autocorrelation (& spatially dependent)**

Random patterns or values that cluster together are said to have **no spatial autocorrelation (and thus no dependence)**

We use a phenomenon called "**Spatial Spillovers**" to explain patterns of autocorrelation

The Geography of Upward Mobility in the United States: Average Household Income for Children with Parents Earning \$27,000 (25th percentile)



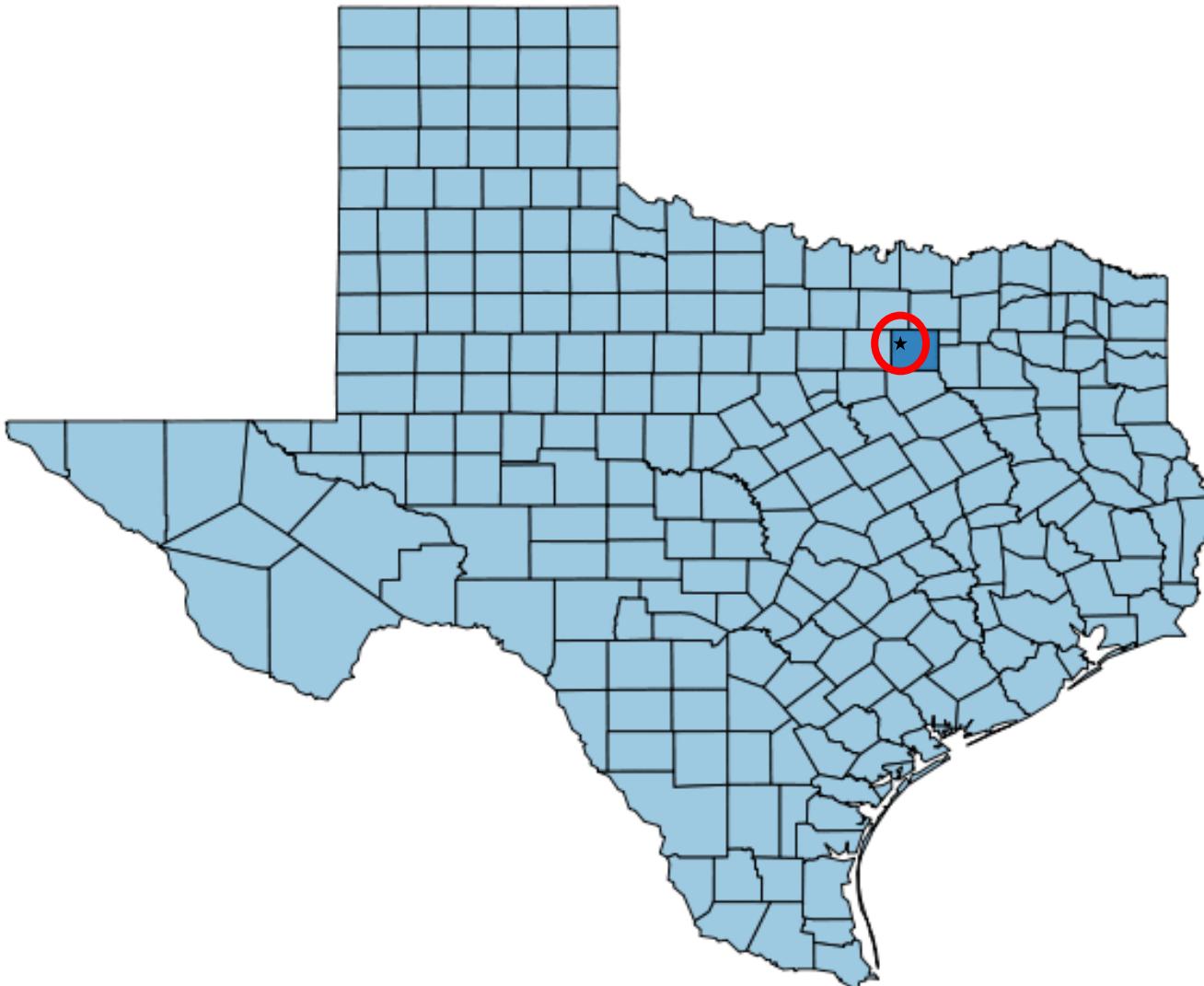
Positive

No patterns

Positive

Spatial dependence and autocorrelation will be covered in-depth in WK3

Spatial Spillover: “...where you are in geographic space matters” an event in one location can somehow have an impact on other events in neighboring areas

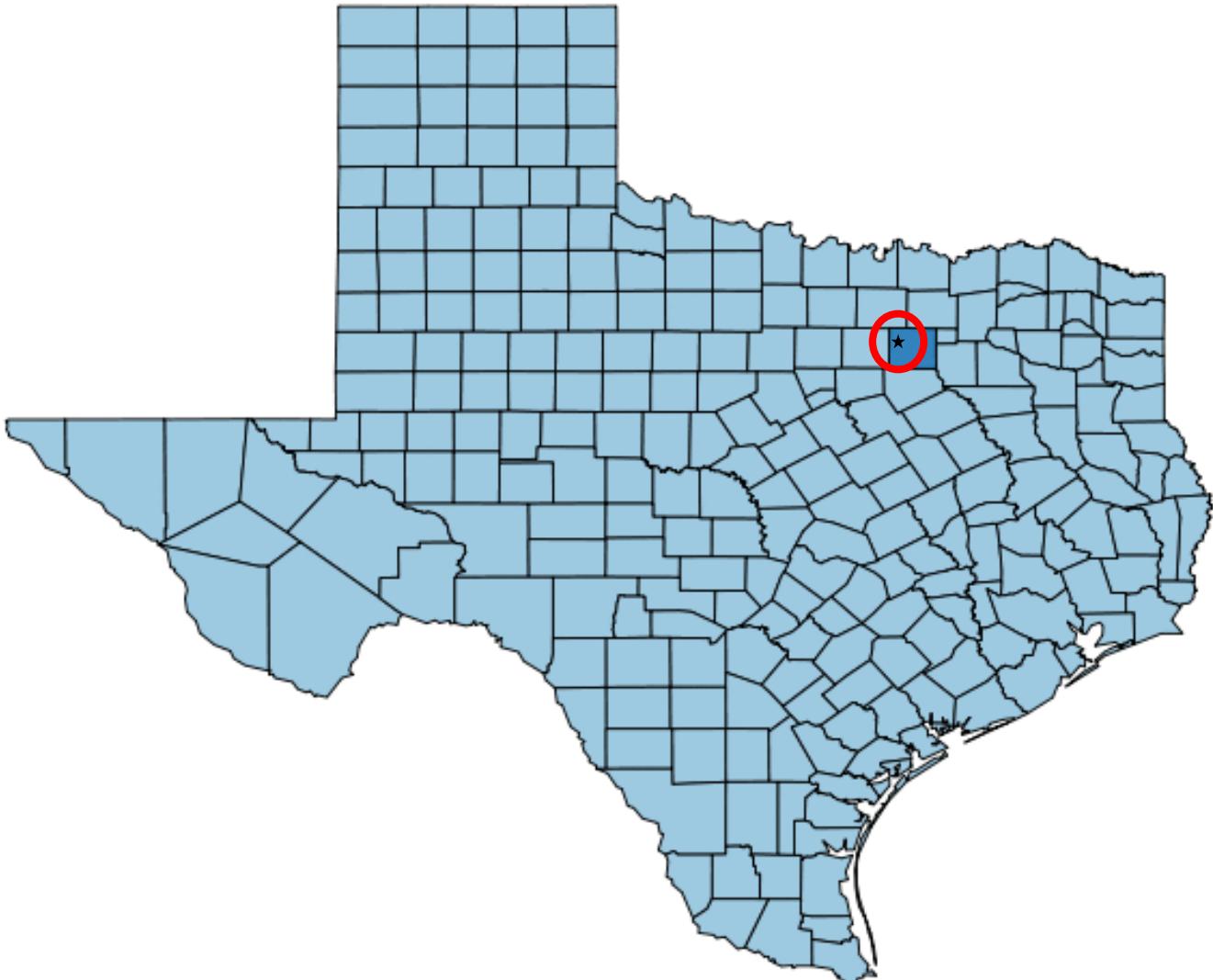


In Spatial Statistics, especially in Spatial Regression, Geostatistics – we often try to account for **spatial spillover** effects in our models

There are four types of spillovers:

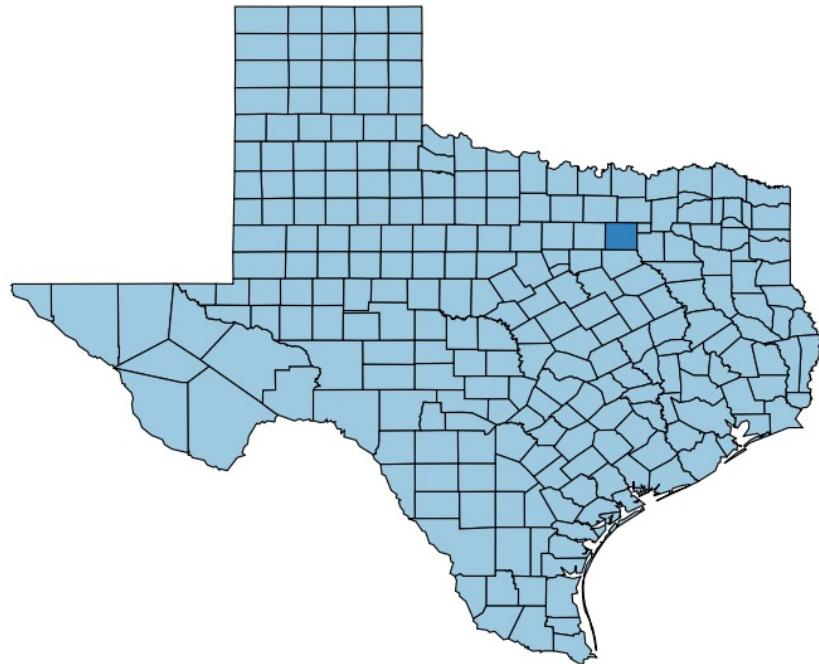
- No spillover
- Local spillover
- Global spillover
- Rippling spillovers

Spatial Spillover: “...where you are in geographic space matters” an event in one location can somehow have an impact on other events in neighboring areas



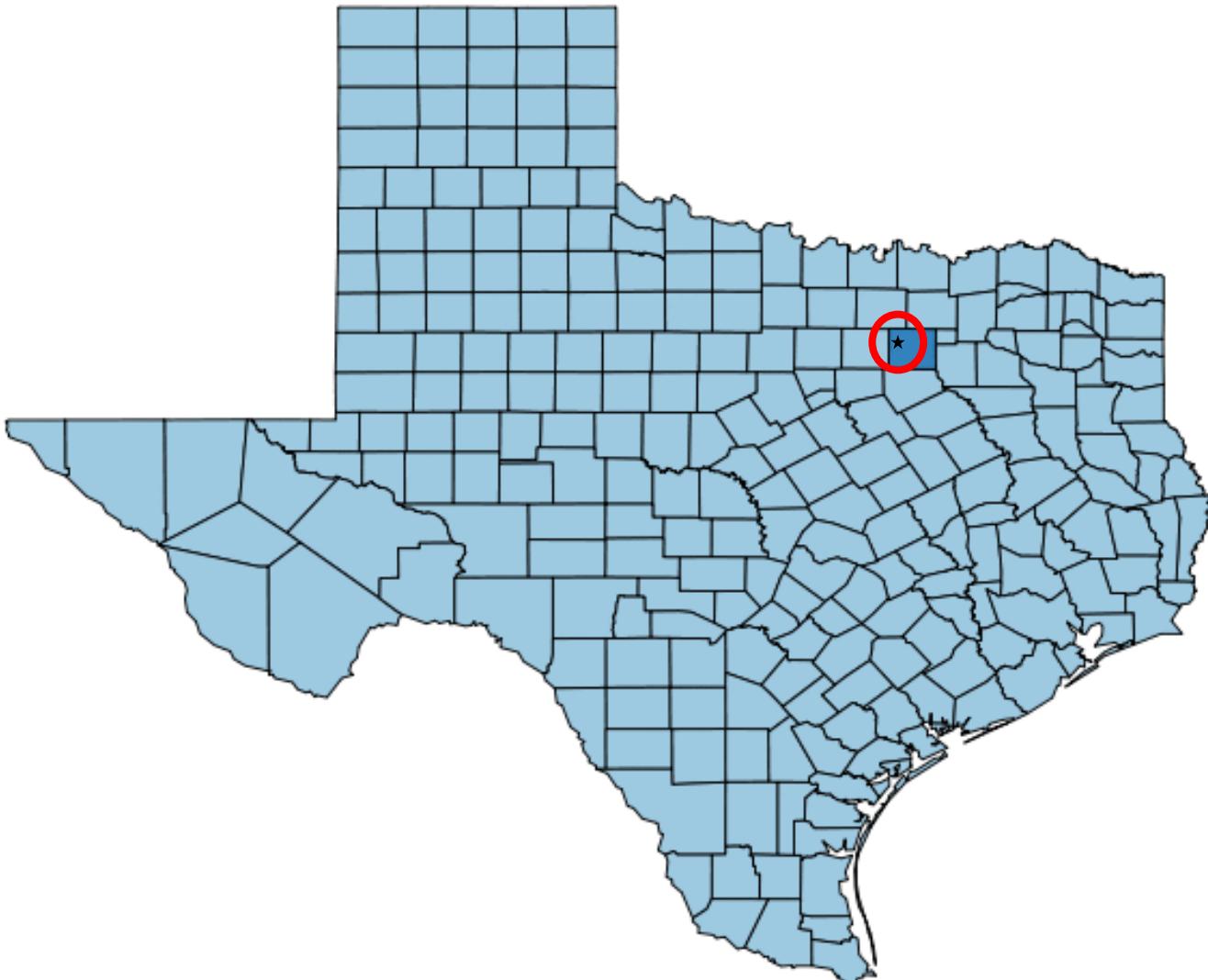
In Spatial Statistics, especially in Spatial Regression, Geostatistics – we often try to account for **spatial spillover** effects in our models

No spillover (independence)



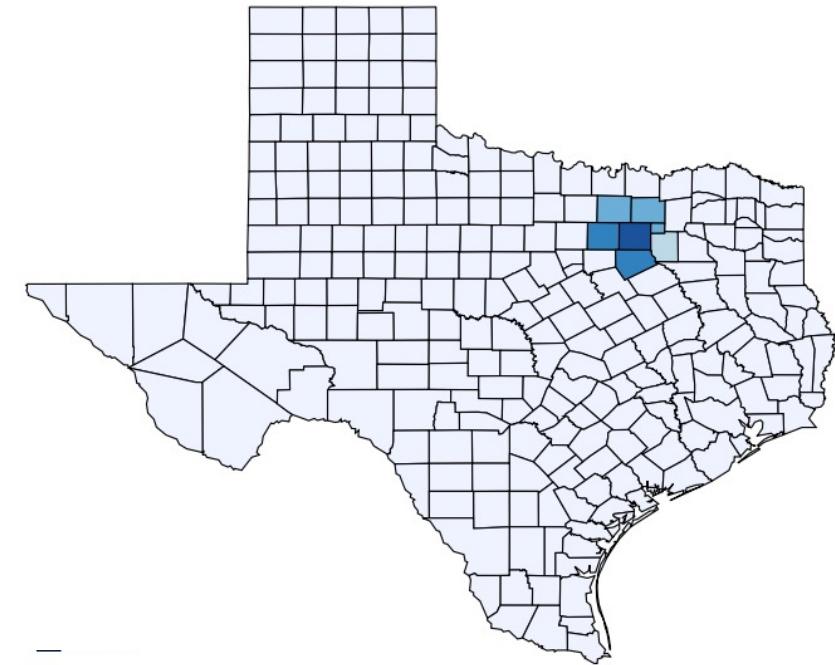
A change in unemployment in that dark blue area will only have an impact on crime rates within its own area, but does not influence crime rates in neighboring areas

Spatial Spillover: “...where you are in geographic space matters” an event in one location can somehow have an impact on other events in neighboring areas



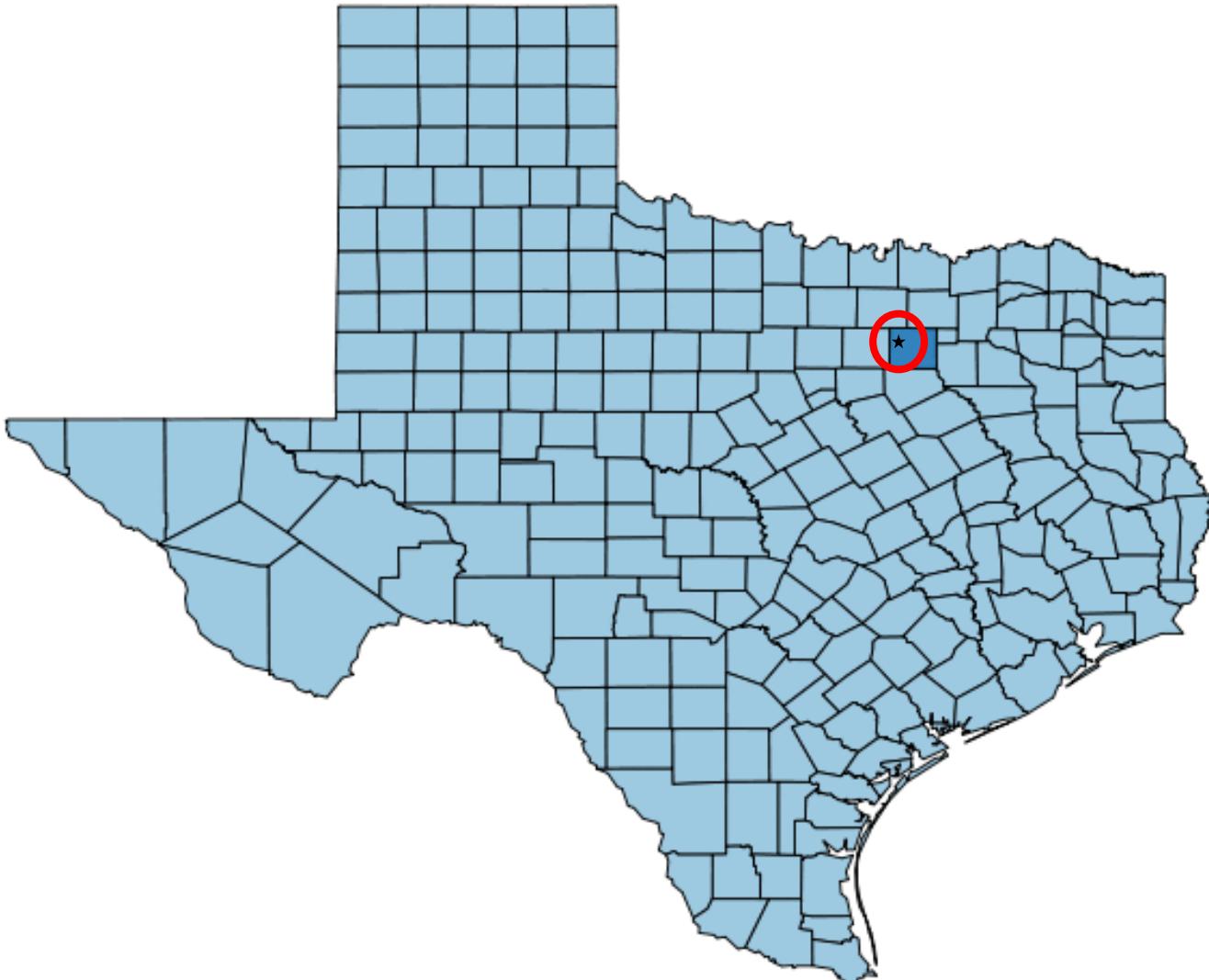
In Spatial Statistics, especially in Spatial Regression, Geostatistics – we often try to account for **spatial spillover** effects in our models

Local spillover (dependence)



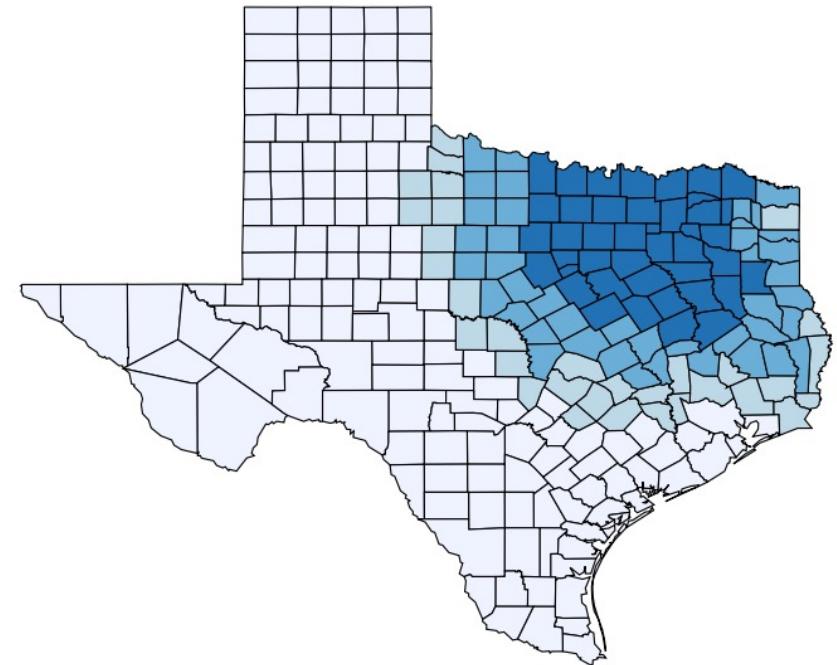
A change in unemployment in that dark blue area (in the center) will not only have an impact on crime rates within its own area, but it will also have a direct influence on crime rates in neighboring areas only.

Spatial Spillover: “...where you are in geographic space matters” an event in one location can somehow have an impact on other events in neighboring areas



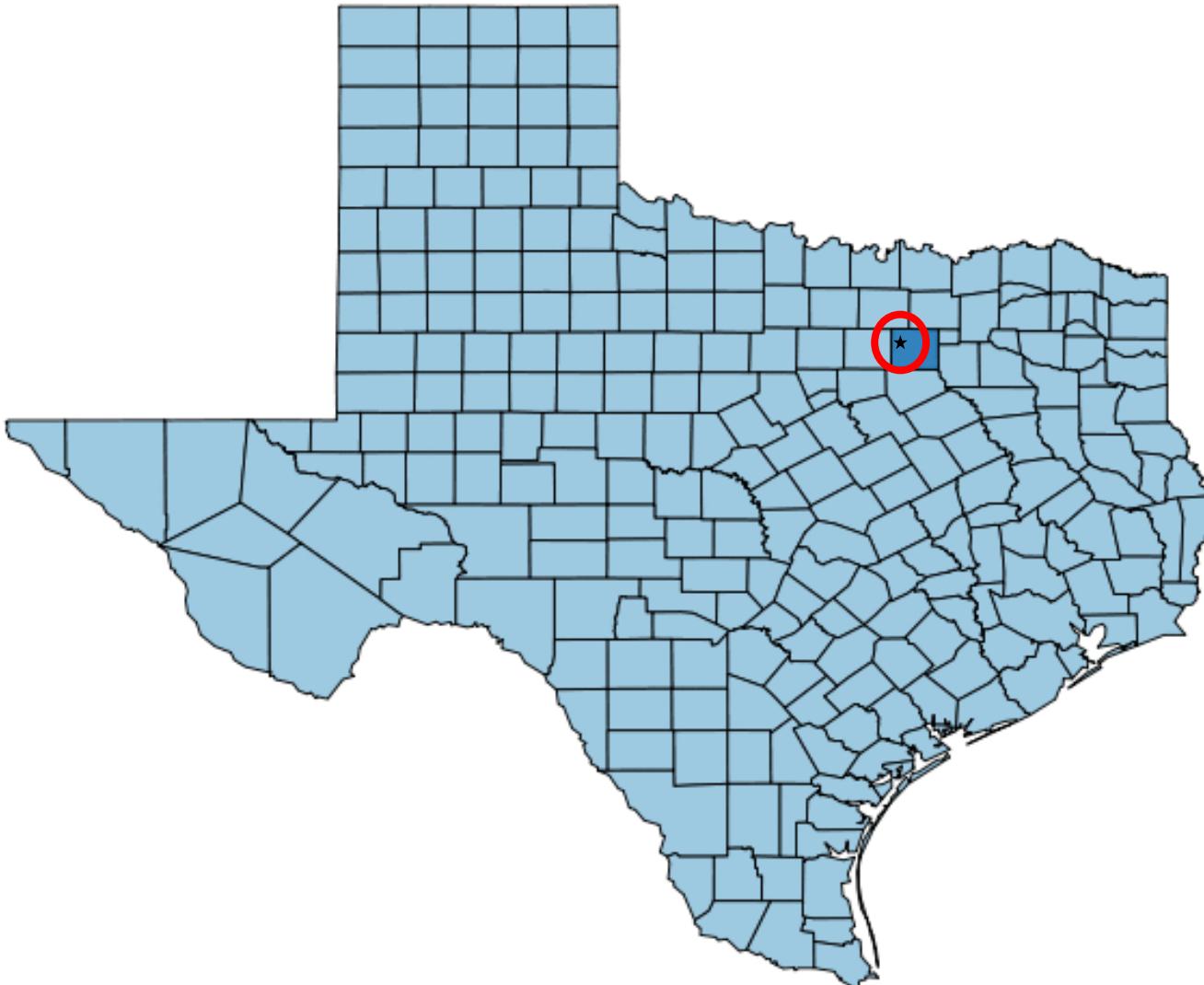
In Spatial Statistics, especially in Spatial Regression, Geostatistics – we often try to account for **spatial spillover** effects in our models

Global spillover (dependence)



A change in unemployment in that dark blue area (in the center) will not only have an impact on crime rates within its own area, but it will also have a wider influence on crime rates beyond its direct neighbors.

Spatial Spillover: “...where you are in geographic space matters” an event in one location can somehow have an impact on other events in neighboring areas



In Spatial Statistics, especially in Spatial Regression, Geostatistics – we often try to account for **spatial spillover** effects in our models

Rippling spillover (dependence)

Where there is a focal point for an event and its influence may have a rippling (or trickle down) effect across space, triggering other events, which then diminishes with time and distance.

e.g.,

- Natural disasters - an earthquake and building destruction.
- Disease spread and outbreaks
- World financial markets and crashes

What is Spatial Data & its Features?

Point Pattern Data (PPD)



Key Characteristics

- The main interest is in the location of occurrences of these event points

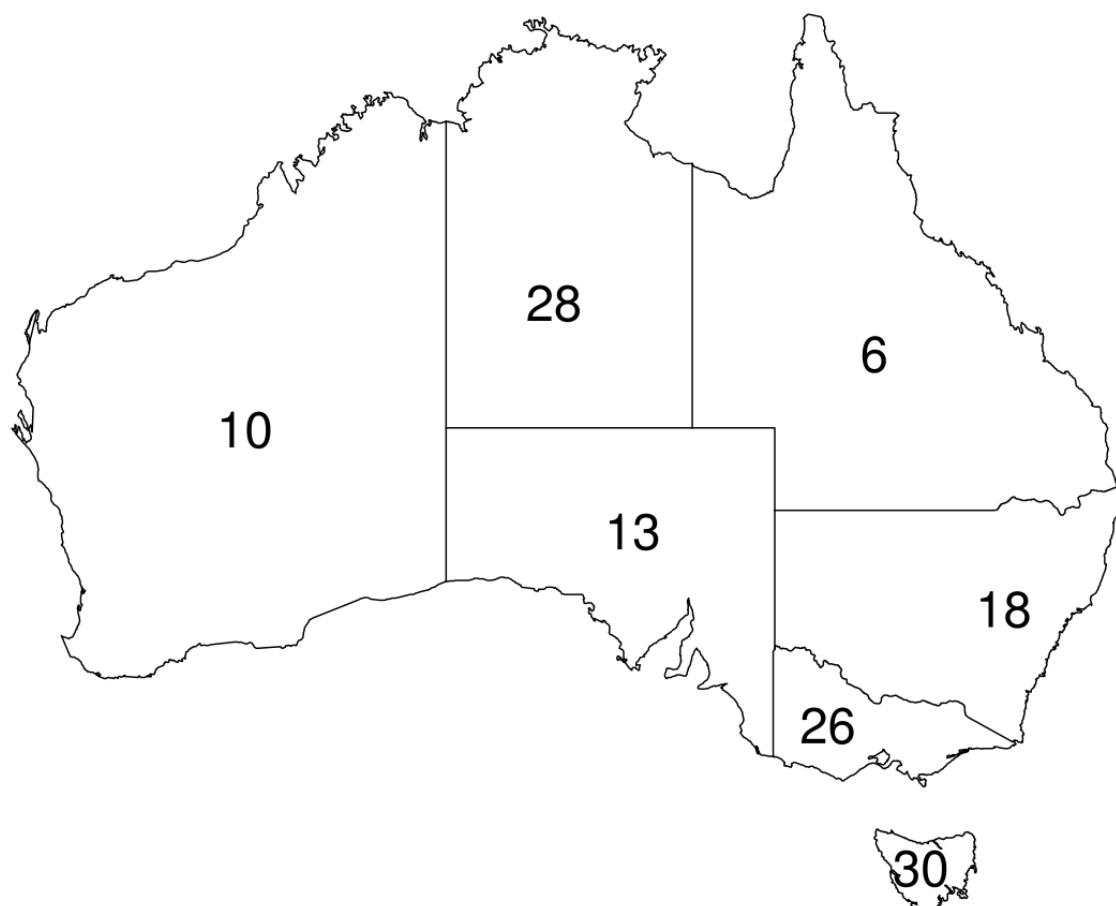
Examples of PPDs events (or outcomes):

- Point locations of burglaries
- Riots
- Locations of car collisions etc.,
- Locations of where an adult tree needs to be replanted

Some PPDs events may carry additional information that may describe the occurrence of an observed event (or outcome):

- Burglary: Type of premise that was burgled, time of day the burglary occurred etc.,
- Car collision: type of road, weather condition etc.

Areal/Regional Data



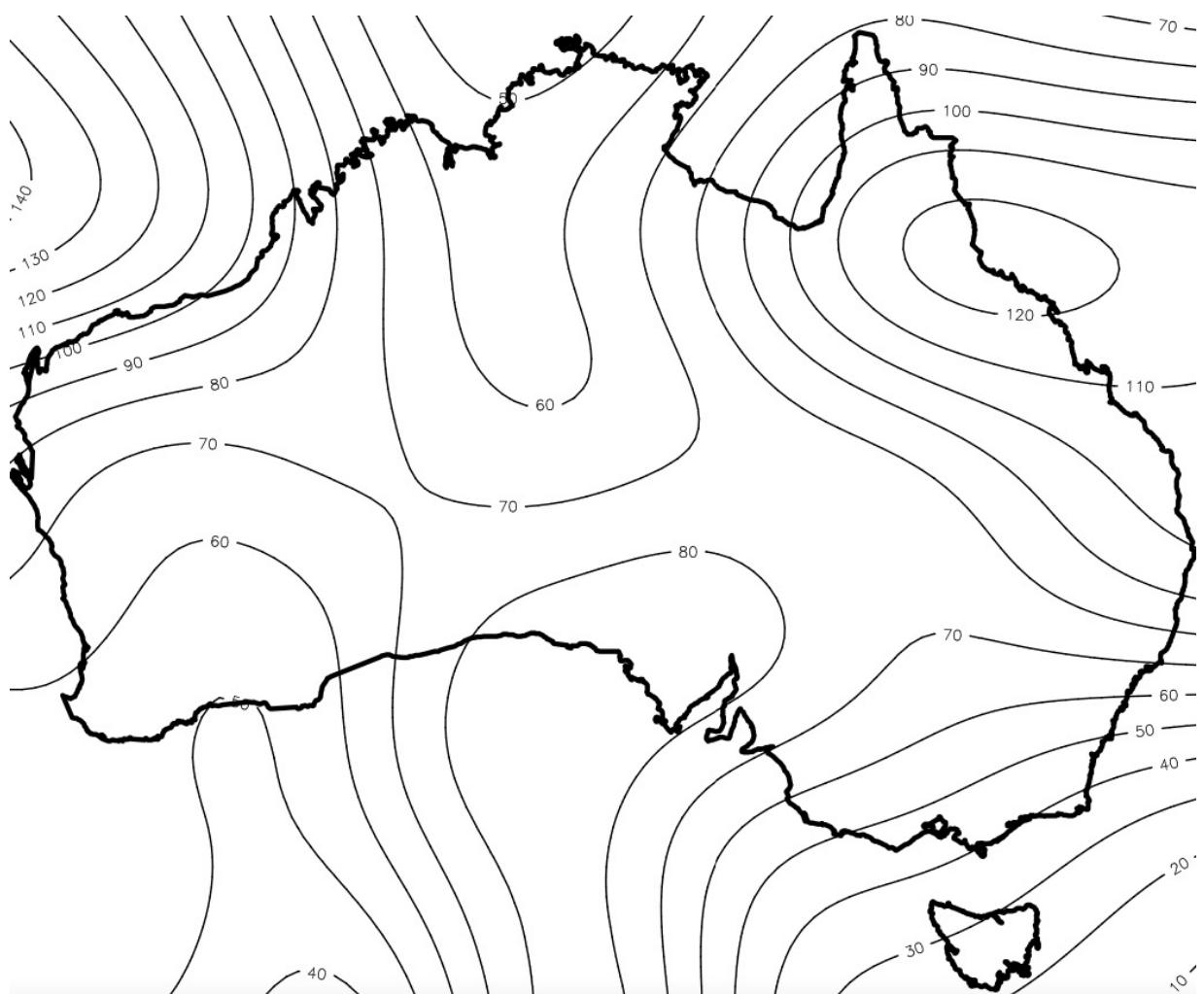
Key Characteristics

- The main interest the quantity of interest defined for areas, or regions.

Events (or outcomes) that are aggregated measures to areas:

- Prevalence of a disease in areas
- Population density in a county
- Regional unemployment rates

Geostatistical Data



Key Characteristics

- The quantity of interest has a value at any location across a given area.
- These are values over a grid/raster

Events are usual statistical prediction determined from points with continuous data values:

- Land surface elevation
- Diffusion of ambient air pollutants
- Environmental suitability for breeding habitats of mosquitoes

Non-spatial context & data structure:

	ATTRIBUTE			
	Variable 1	Variable 2	...	Variable n
Entity 1	$attribute_{11}$	$attribute_{12}$...	$attribute_{1n}$
Entity 2	$attribute_{21}$	$attribute_{22}$...	$attribute_{2n}$
:	:	:	..	:
Entity m	$attribute_{m1}$	$attribute_{m2}$...	$attribute_{mn}$

To apply some spatial analysis to data – you must have some variable that defines the entity's geographic location. This can be **GPS coordinates, spatially referenced geometries** of an area

NOTES: It is not enough to have just the name of the area(s). It must be some geometric entry!

- Attributes that defines an entity's location are typically excluded from the analysis
- The conventional statistical methods, that assumes independence, are used for analyzing such dataset
- Results churned from this dataset are completely independent from “**spatial arrangement**” of the entities.

Spatial context & data structure:

	Geographical Coordinate		ATTRIBUTE			
	X	Y	Variable 1	Variable 2	...	Variable n
Entity 1	X_1	Y_1	$attribute_{11}$	$attribute_{12}$...	$attribute_{1n}$
Entity 2	X_2	Y_2	$attribute_{21}$	$attribute_{22}$...	$attribute_{2n}$
:	:	:	:	:	..	:
Entity m	X_m	Y_m	$attribute_{m1}$	$attribute_{m2}$...	$attribute_{mn}$

In this example, what defines the entity's geographic location are **X, Y GPS coordinates**.

Definition of an entity's location are not limited to coordinates, you can have spatially reference areas with their associated boundaries with geometries.

This instance illustrates an example of geostatistical data.

- Attributes that defines an entity's location are typically explicitly incorporated into analysis
- Spatial statistical methods, that assumes dependence, are used for analyzing such geographically referenced dataset
- Results churned from this dataset are completely dependent from “**spatial arrangement**” of the entities.

Geographical Information Systems (GIS)



Roger F. Tomlinson (1933 – 2014)

He is an English geographer and computer scientist, who is revered as the “**Father of GIS**”

- His PhD research titled: “*The application of electronic computing methods and techniques to the storage, compilation, and assessment of mapped data*” (1962, @UCL)

- Conceived the idea of analyzing multiple layers of spatial data within a single environment

- **Geographical Information Systems (GIS)**, which is now a software, grown into a billion-dollar industry



PROJ.4





PROJ.4





R (Standard)

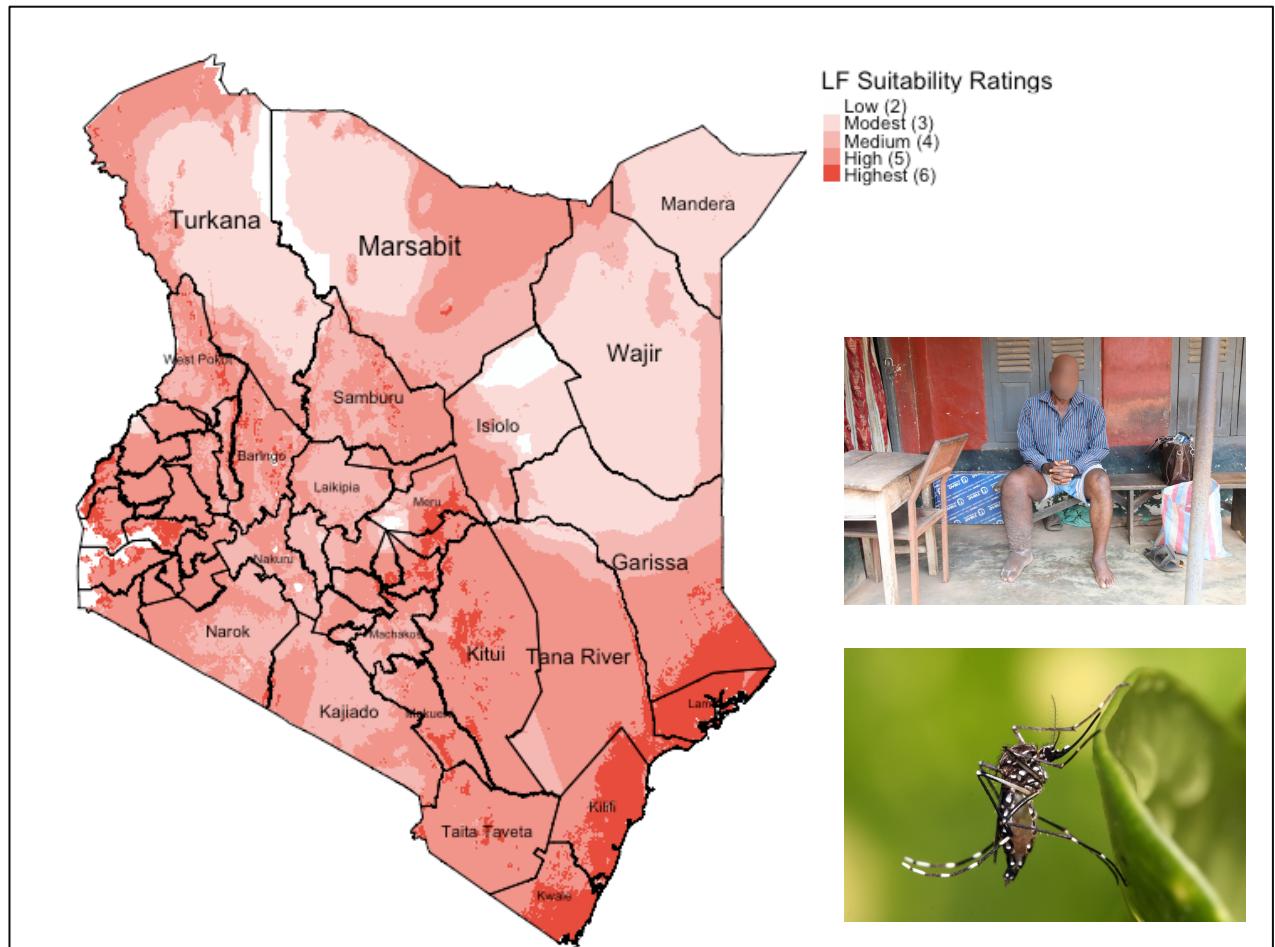


RStudio (Best)

There are two versions of the software: 1.) R, and 2.) **RStudio**; The second is much preferred as its straightforward and intuitive.

Why are we teaching RStudio?

1. Flexible and provides access to powerful packages for analysis
2. Impressive graphs, visualizations and maps
3. Excellent statistical capabilities too



Example: Map generated in R to illustrate areas that are environmentally suitable for the spread of neglected tropical disease called 'Lymphatic Filariasis (LF)' in Kenya.

Sources:

1. Global Atlas for Helminths Infection (<http://www.thiswormyworld.org>)
2. ESPEN (<https://espen.afro.who.int>)

... and why learn how to code in RStudio?

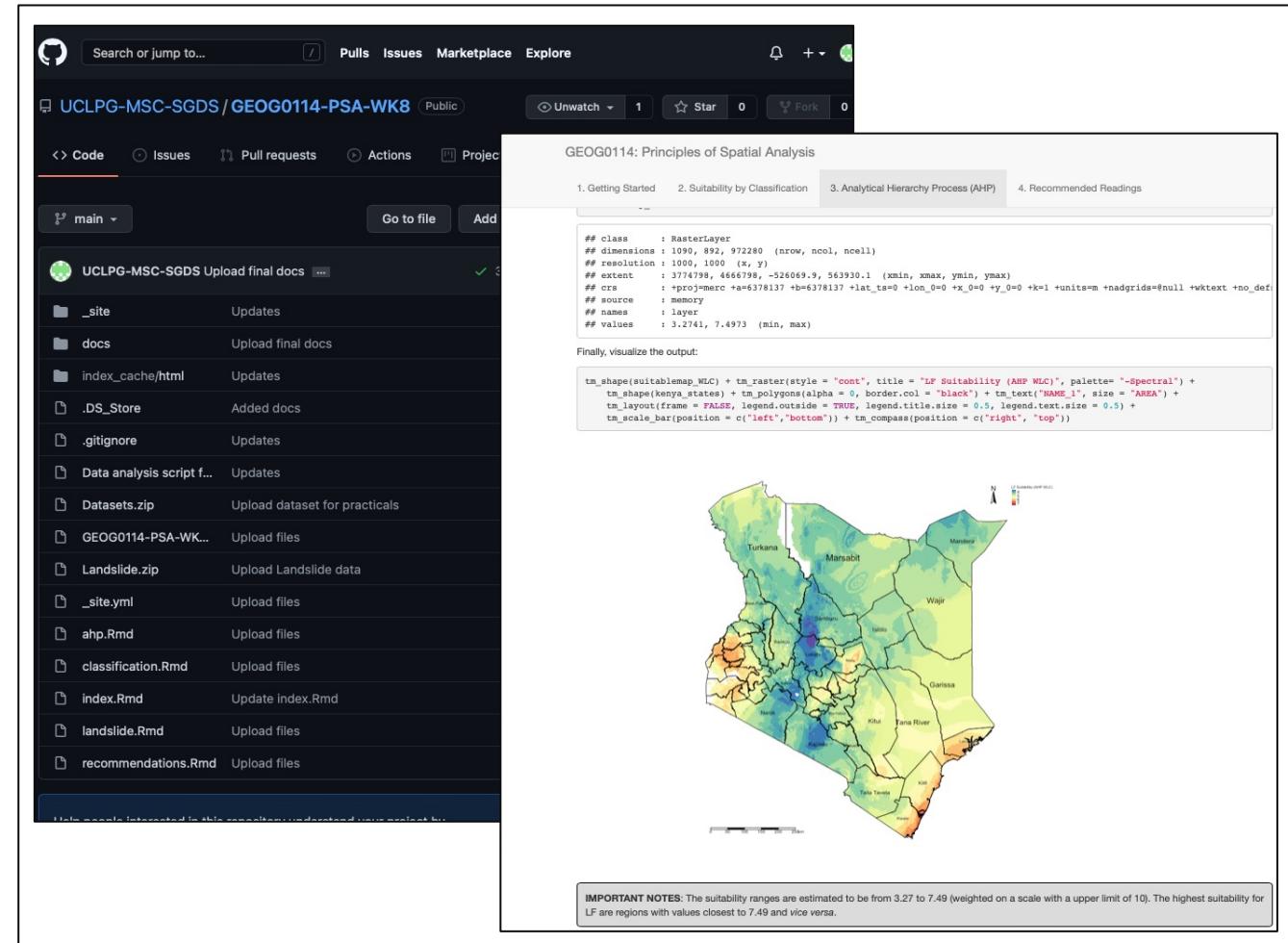
1. Efficiency

- Automated tasks and data managing
- Can recycle & reuse code scripts for new projects

2. Fosters good scientific practice

- Transparency and replication (AKA reproducible research)
- Creates log so anyone can follow in your footstep (i.e., github, gitlab etc.,)

You can literally pull-off some really creative stuff like generating websites, accessing tools via APIs etc.



Example: Working in RStudio and synchronising it with GitHub to not only use as a cloud back-up, but to generate a website through RStudio and GitHub for teaching MSc Students.

Summary

The take home message is:

- Spatial statistics is grounded on spatial dependence, drawn from **Tobler's 1st Law of Geography**
- Fundamental concepts for explaining spatial dependence – are two key words: **Spatial autocorrelations & Distance decay** effect. The former key word can be explained by Spatial Spillovers.
- Spatial data and data structure – three types: areal/regional data, point patterns and geostatistical
- Two important people to commit to memory: Dr Waldo Tobler (1st Law of Geography) and Dr Roger Tomlinson (Father of GIS)
- Best GIS packages are usual the ones that are FREE (i.e., R/RStudio for its packages that enable GIS functionality, and QGIS)

Any questions?