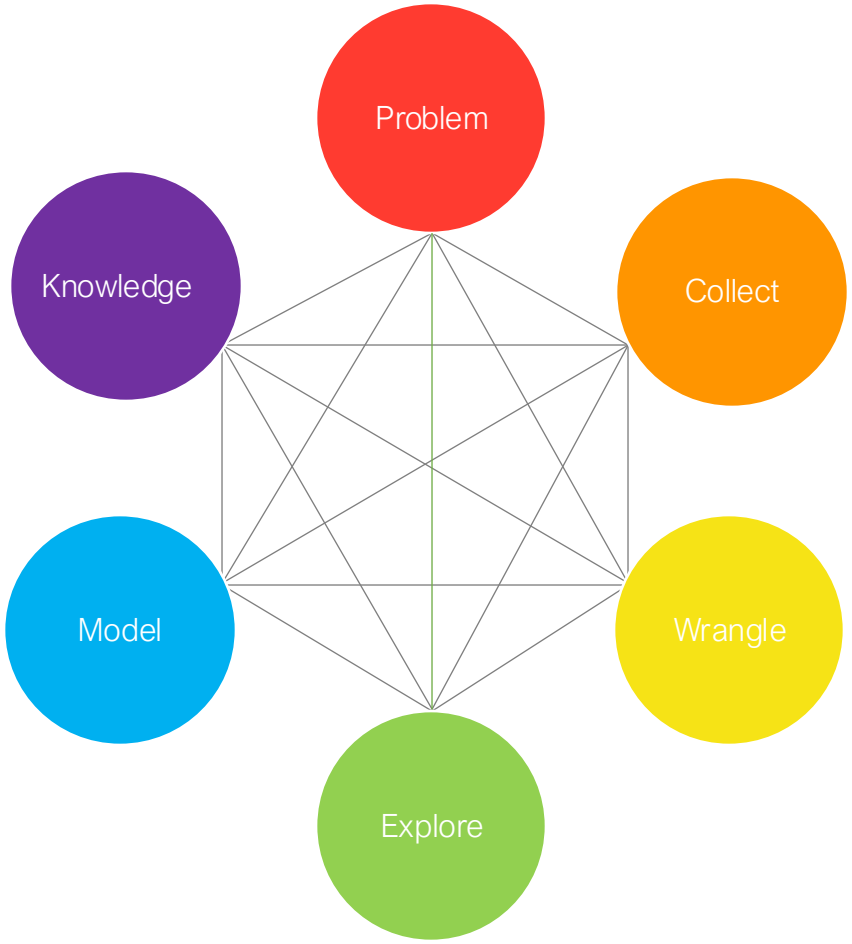
GEOG0114: PRINCIPLES OF SPATIAL ANALYSIS

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

Contents

- Timetable & Assessment
- What will we learn in GEOG0114?
- The Beginning: Introduction to Spatial Analysis & Data Science



Timetable and Assessment

Timetable

Lecture: North-West Wing, Room G17

(Monday: 01:00pm-02:00pm)

Computer practical sessions*: North-West Wing, Room G07

(Thursday: 03:00pm-05:00pm)

*IMPORTANT NOTE: Please bring your own laptops with you to the computer practicals on Thursday

Coursework & Assessment

- What is the assessment?
  - Spatial data analysis project which is worth 100%
  - It is a report (max 3,000 words) + code (R).

- You can check the coursework out right now and start thinking about the topic for your project.

- The deadline for the coursework is in the 1st week of **January (Term 2)**. 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/pixeled 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, 8 and 9): Spatial Models e.g., Lag and Error Models, GWRs, Spatial Risk Models.

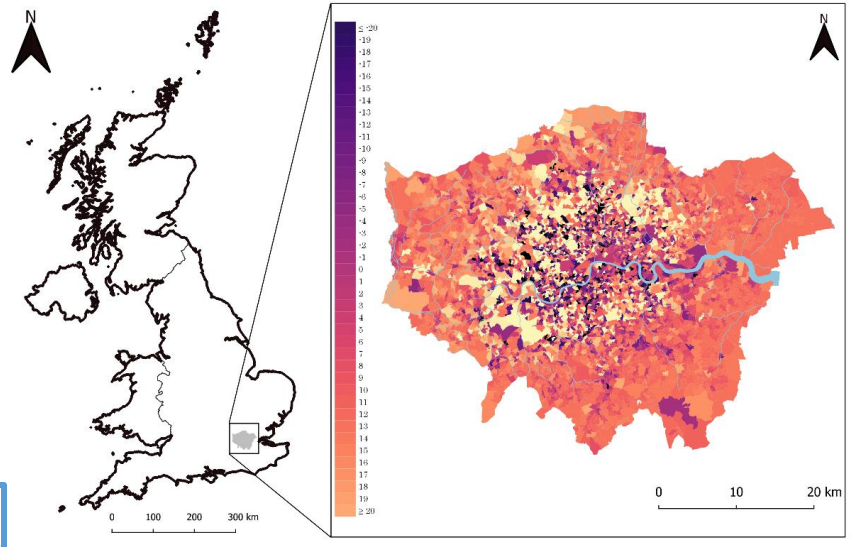
Part Four (Week 10): Revision

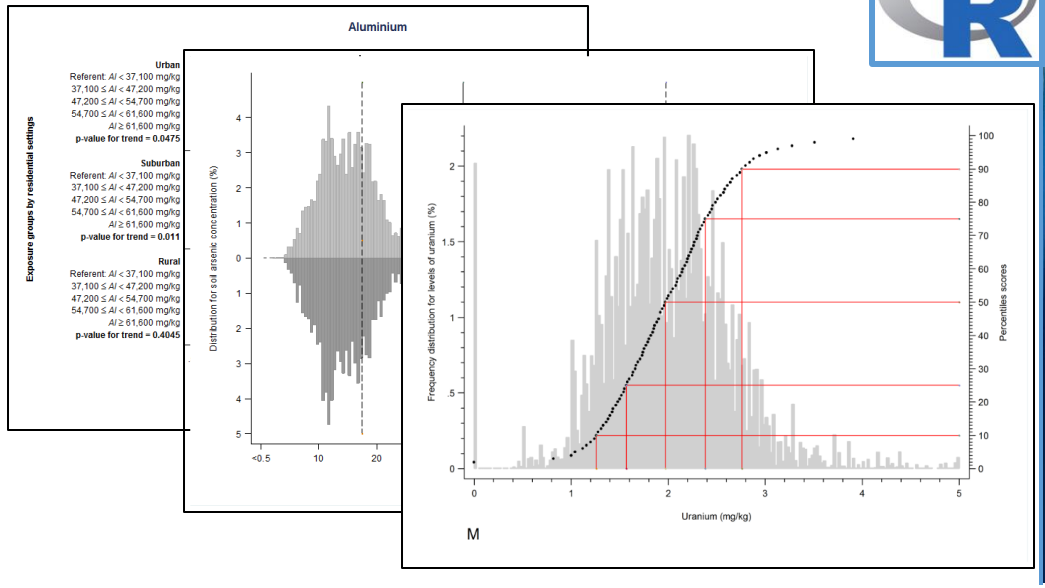
- Basic principles & visualisation
- Geoprocessing
- Various spatial analytical techniques
- Spatial models

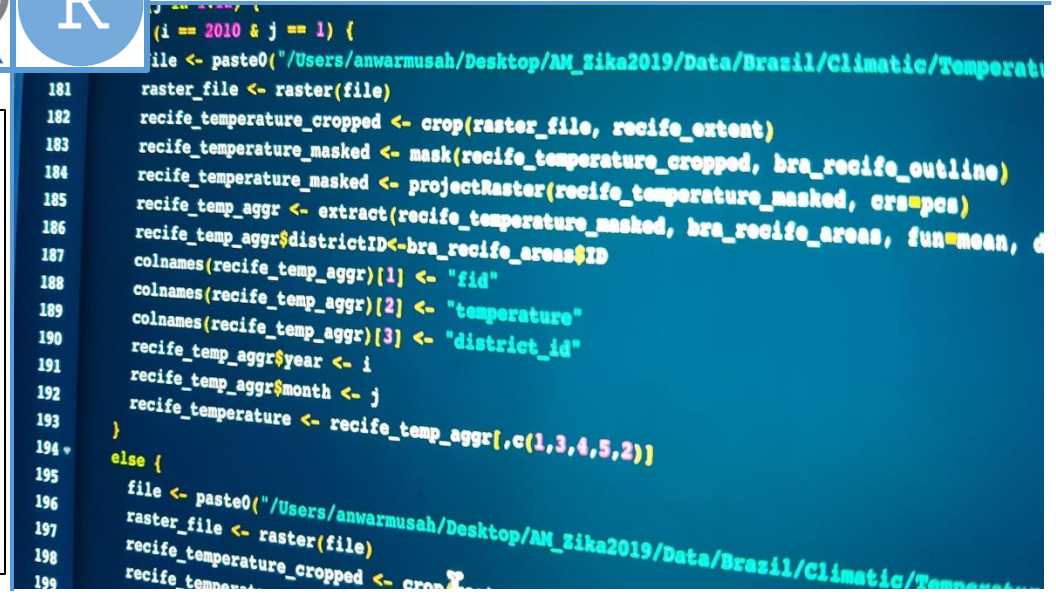
Breakdown of what will we be learning

Week	Weekly Topics
1	Spatial Analysis for Data Science
2	Graphical Representation of Spatial Data
3	Spatial Autocorrelation
4	Suitability Mapping: Part I (Analytical Hierarchy Process)
5	Suitability Mapping: Part II (Ecological Niche Models)
	Reading Week
6	Geostatistics using Kriging
7	Spatial Models: Part I (Spatial Lag & Error Models)
8	Spatial Models: Part II (Geographically Weighted Regression)
9	Spatial Models: Part III (Bayesian Risk Models)
10	Revision

We will be using RStudio throughout the course

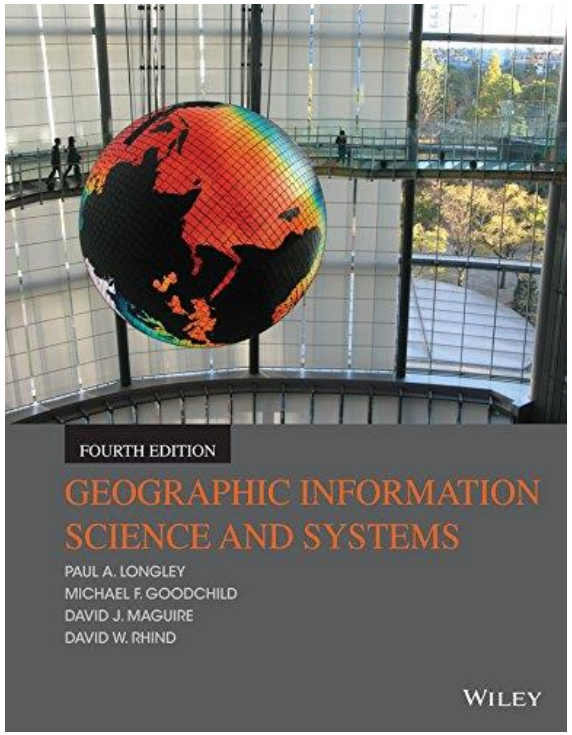


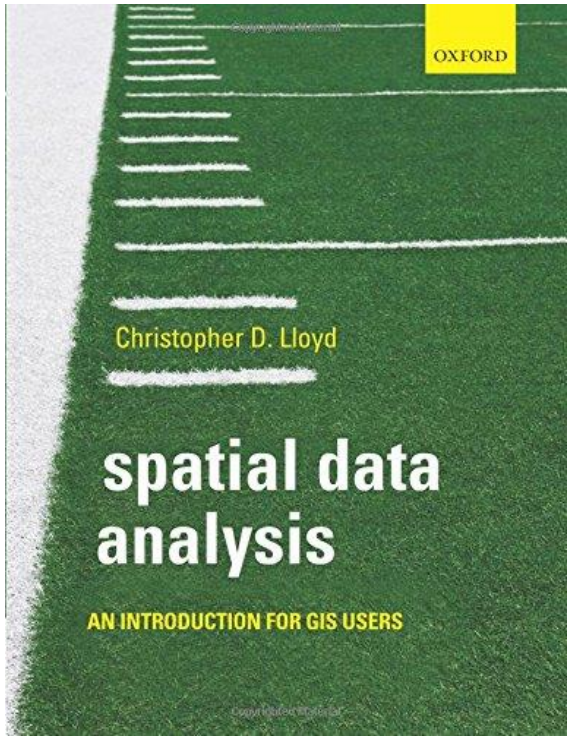




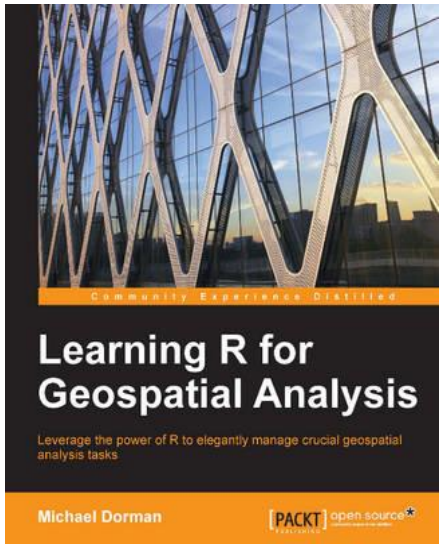
```
(i == 2010 & j == 1) {
ile <- paste0("/Users/anwarmusah/Desktop/AM_Zika2019/Data/Brazil/Climatic/Temperat
  raster_file <- raster(file)
  recife_temperature_cropped <- crop(raster_file, recife_extent)
  recife_temperature_masked <- mask(recife_temperature_cropped, bra_recife_outline)
  recife_temperature_masked <- projectRaster(recife_temperature_masked, crs=pcs)
  recife_temp_aggr <- extract(recife_temperature_masked, bra_recife_areas, fun=mean,
  recife_temp_aggr$districtID<-bra_recife_areas$ID
  colnames(recife_temp_aggr)[1] <- "fid"
  colnames(recife_temp_aggr)[2] <- "temperature"
  colnames(recife_temp_aggr)[3] <- "district_id"
  recife_temp_aggr$year <- i
  recife_temp_aggr$month <- j
  recife_temperature <- recife_temp_aggr[,c(1,3,4,5,2)]
}
else {
  file <- paste0("/Users/anwarmusah/Desktop/AM_Zika2019/Data/Brazil/Climatic/Temperature
  raster_file <- raster(file)
  recife_temperature_cropped <- crop
```

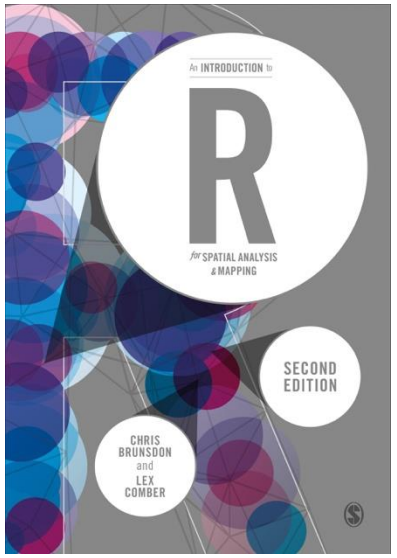
Recommended Books for GEOG0114

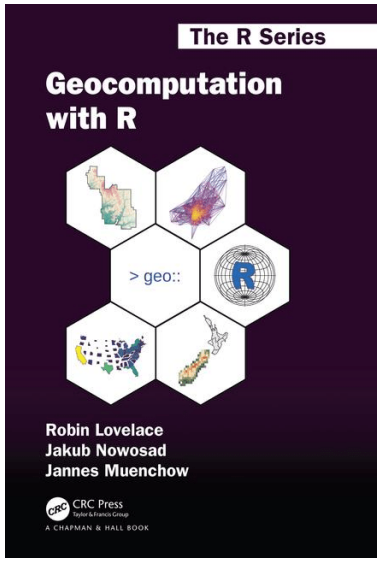




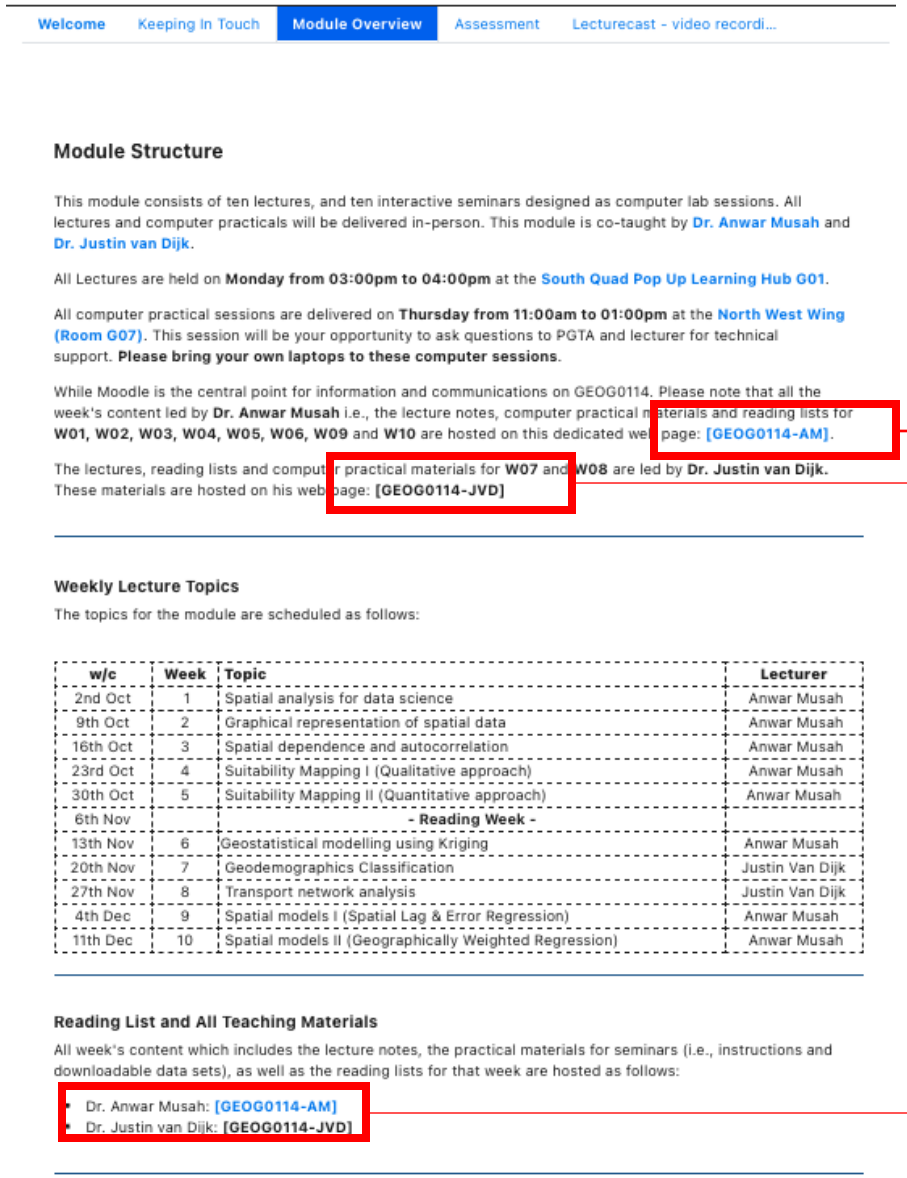
High recommendation for the mastery of basic theory and principles of spatial analysis







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



Welcome | Keeping In Touch | **Module Overview** | Assessment | Lecturecast - video recordi...

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. This module is co-taught by Dr. Anwar Musah and Dr. Justin van Dijk.

All Lectures are held on **Monday from 03:00pm to 04:00pm** at the South Quad Pop Up Learning Hub G01.

All computer practical sessions are delivered on **Thursday from 11:00am to 01:00pm** at the North West Wing (Room G07). This session will be your opportunity to ask questions to PGTA and lecturer for technical support. **Please bring your own laptops to these computer sessions**.

While Moodle is the central point for information and communications on GEOG0114. Please note that all the week's content led by **Dr. Anwar Musah** i.e., the lecture notes, computer practical materials and reading lists for **W01, W02, W03, W04, W05, W06, W09** and **W10** are hosted on this dedicated web page: **[GEOG0114-AM]**.

The lectures, reading lists and computer practical materials for **W07** and **W08** are led by **Dr. Justin van Dijk.** These materials are hosted on his web page: **[GEOG0114-JVD]**

Weekly Lecture Topics

The topics for the module are scheduled as follows:

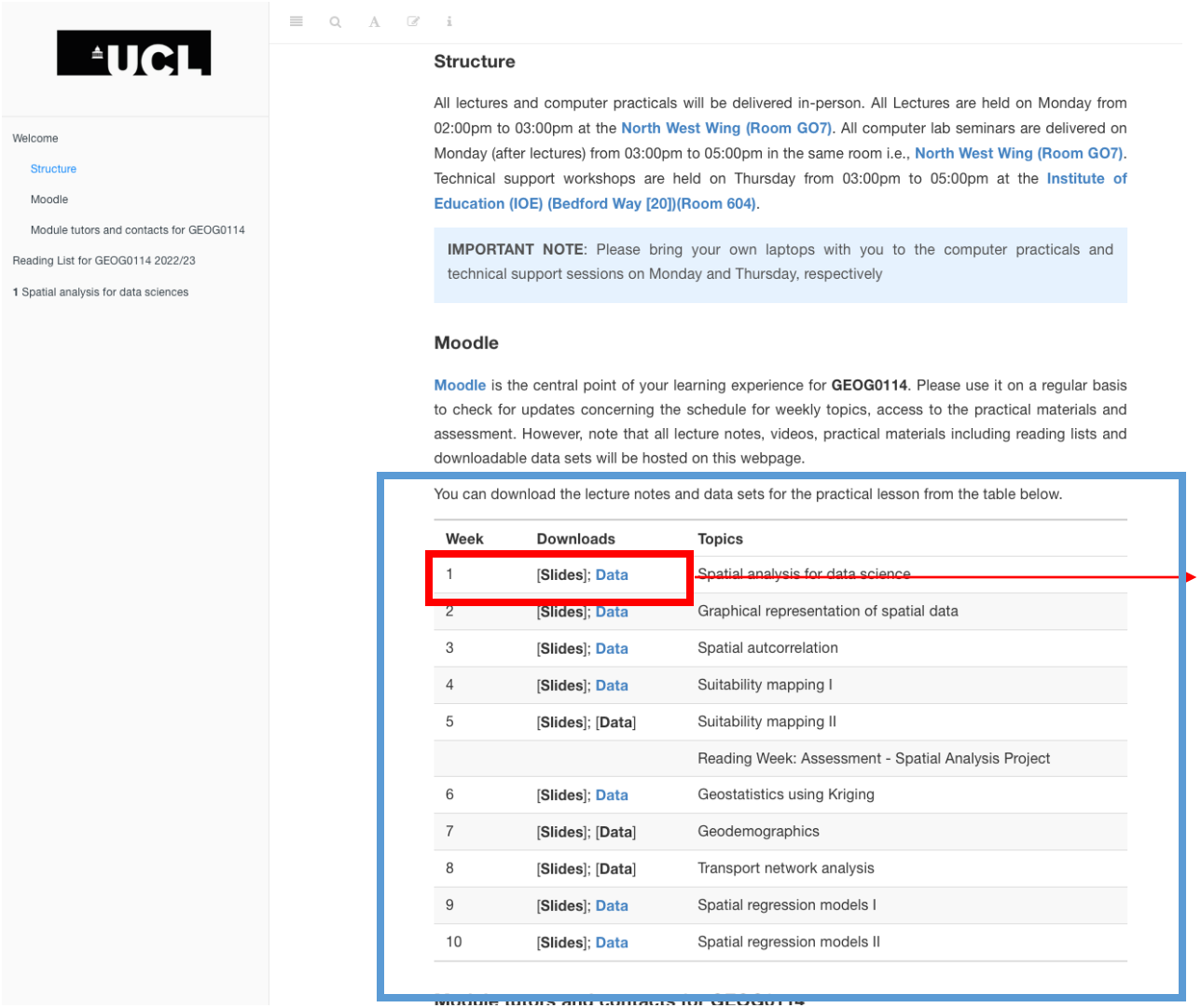
w/c	Week	Topic	Lecturer
2nd Oct	1	Spatial analysis for data science	Anwar Musah
9th Oct	2	Graphical representation of spatial data	Anwar Musah
16th Oct	3	Spatial dependence and autocorrelation	Anwar Musah
23rd Oct	4	Suitability Mapping I (Qualitative approach)	Anwar Musah
30th Oct	5	Suitability Mapping II (Quantitative approach)	Anwar Musah
6th Nov		**- Reading Week -**	
13th Nov	6	Geostatistical modelling using Kriging	Anwar Musah
20th Nov	7	Geodemographics Classification	Justin Van Dijk
27th Nov	8	Transport network analysis	Justin Van Dijk
4th Dec	9	Spatial models I (Spatial Lag & Error Regression)	Anwar Musah
11th 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 as follows:

- Dr. Anwar Musah: **[GEOG0114-AM]**
- Dr. Justin van Dijk: **[GEOG0114-JVD]**

- 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 by the end of Thursdays before next week's Monday's and Thursday's lectures & practical session, respectively.



UCL

Welcome
- Structure
- Moodle
- Module tutors and contacts for GEOG0114

Reading List for GEOG0114 2022/23

1 Spatial analysis for data sciences

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

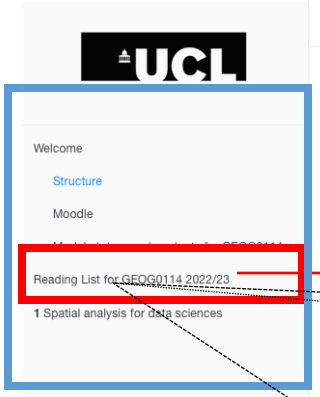
Moodle

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 autcorrelation
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

- Download the lecture and practical materials for that week



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

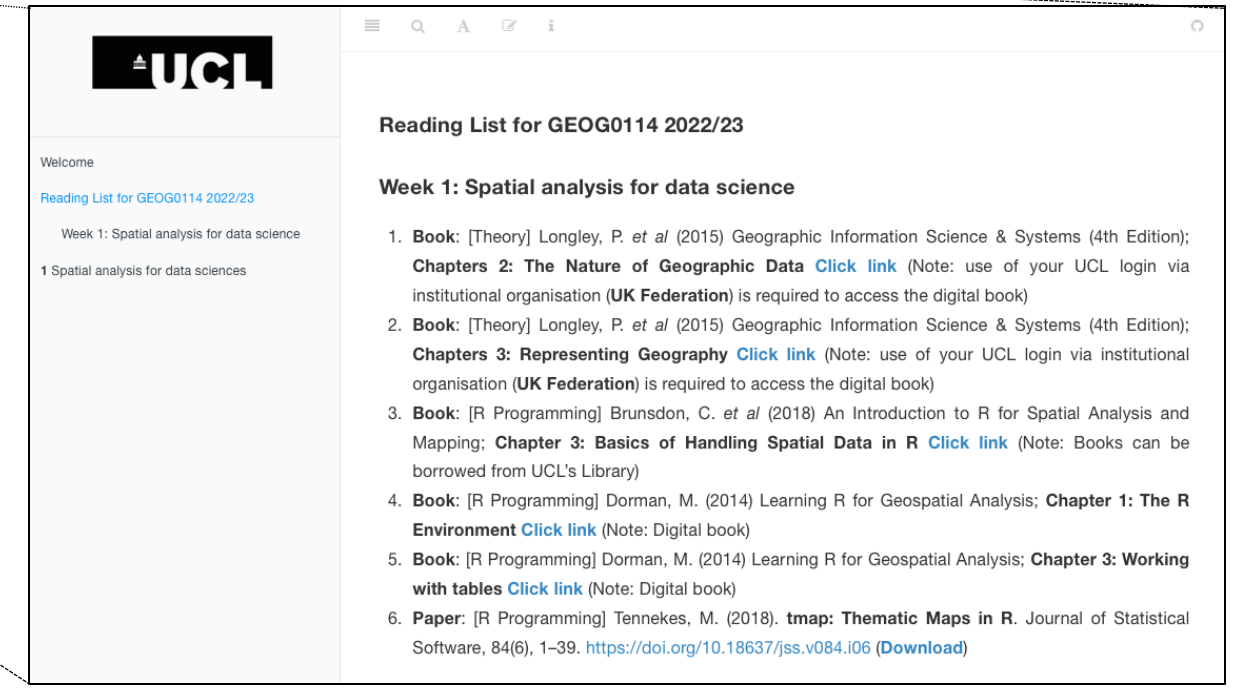
Moodle is the central point of your learning experience for **GEOG0114**. Please use it on a regular basis to check for updates concerning the schedule for weekly topics, access to the practical materials and assessment. However, note that all lecture notes, videos, practical materials including reading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

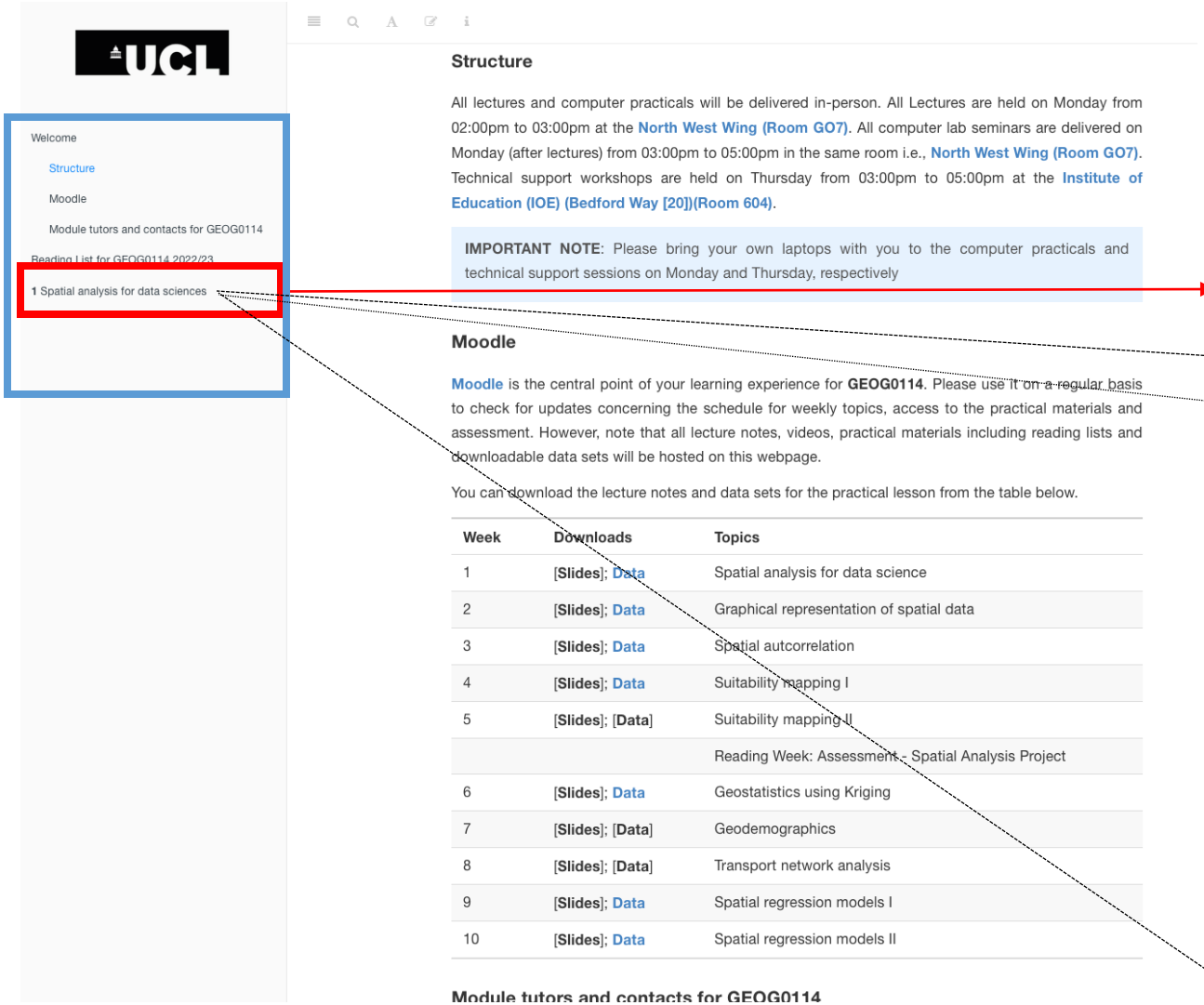
You can download the lecture notes and data sets for the practical lesson from the table below.

Week	Downloads	Topics
1	[**Slides**]; **Data**	Spatial analysis for data science
2	[**Slides**]; **Data**	Graphical representation of spatial data
3	[**Slides**]; **Data**	Spatial autcorrelation
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

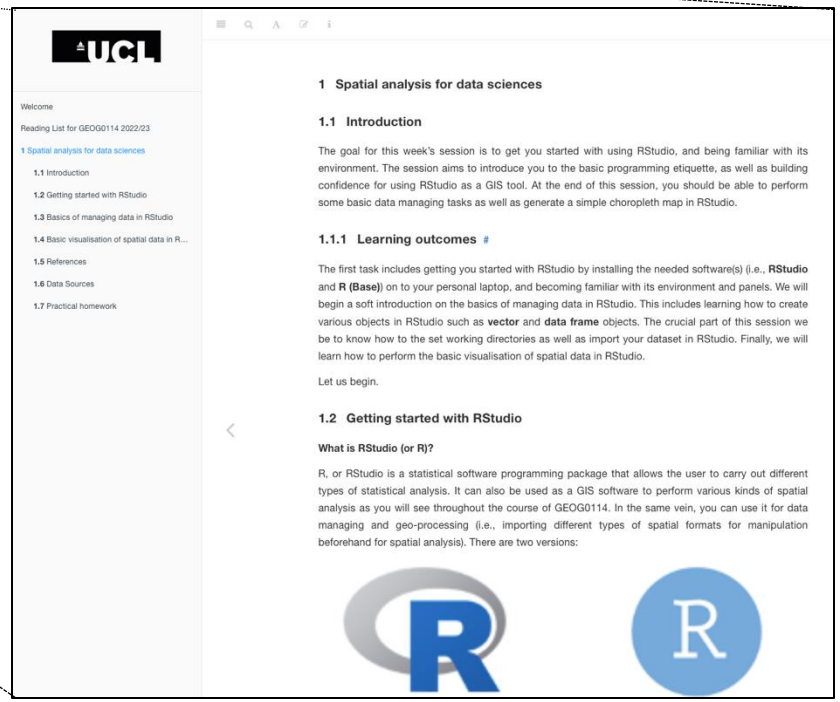
Module tutors and contacts for GEOG0114

- Access to the reading list for GEOG0114 2023/24





- Access to the chapter that actual computer practicals



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 assertation that nearby geographical observations (or objects) are somewhat associated in someway in space.

- The field is interdisciplinary as it brings to 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/gridded** outcomes, **interactions** between objects in space, but also how **an objects has an impact on other nearby objects**.

Traditional statistics versus spatial statistics [1]

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

Traditional statistics versus spatial statistics [2]

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)

First Law of Geography [1]

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:**

- **Spatial Autocorrelation**
- **Distance Decay**
- **Spatial Spillover**

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

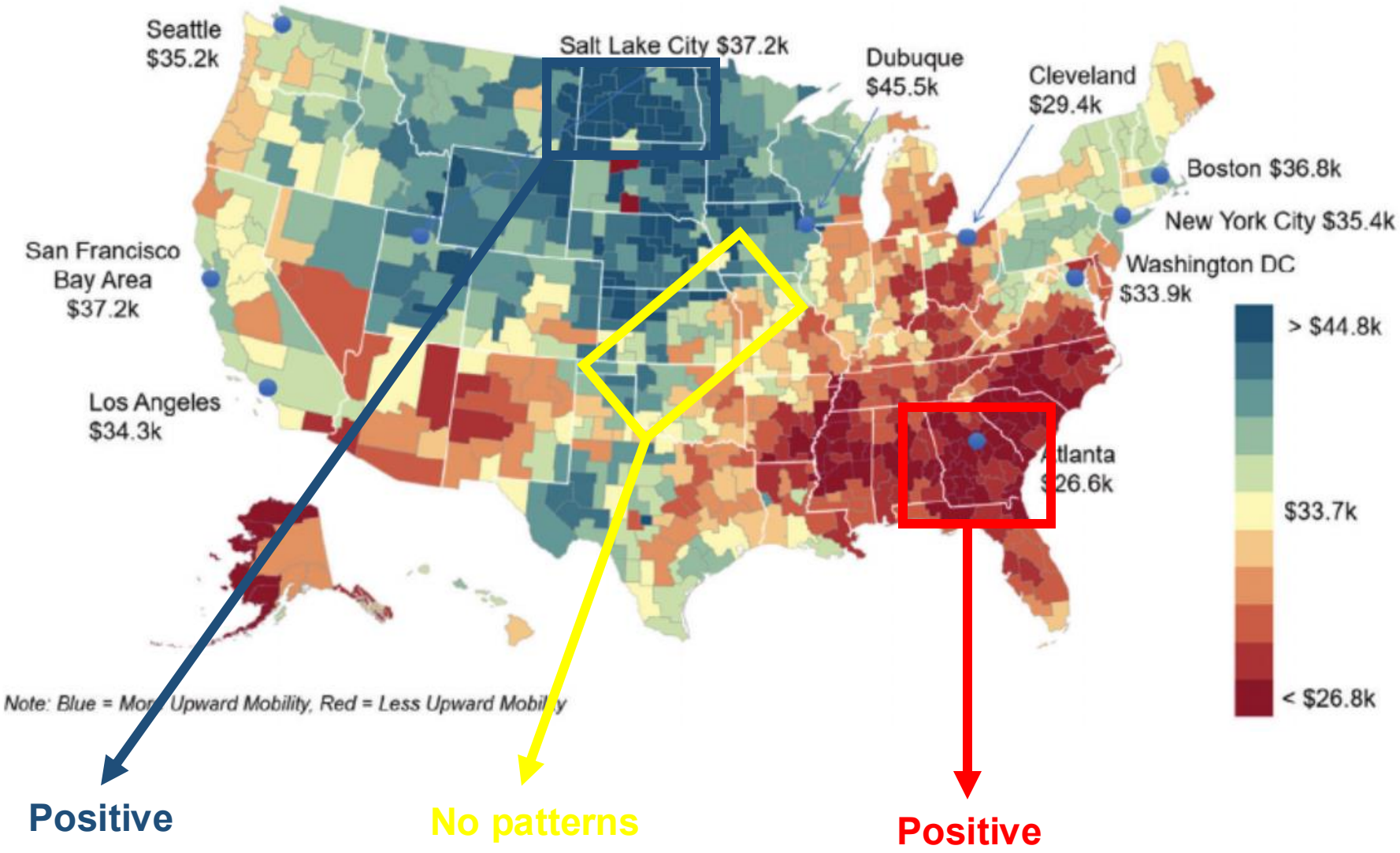
"This simply refers to whether (or not) similar values cluster together over geographic space"

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)**

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



Spatial dependence and autocorrelation, particularly in areal data, will be covered in-depth in Week 3

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."

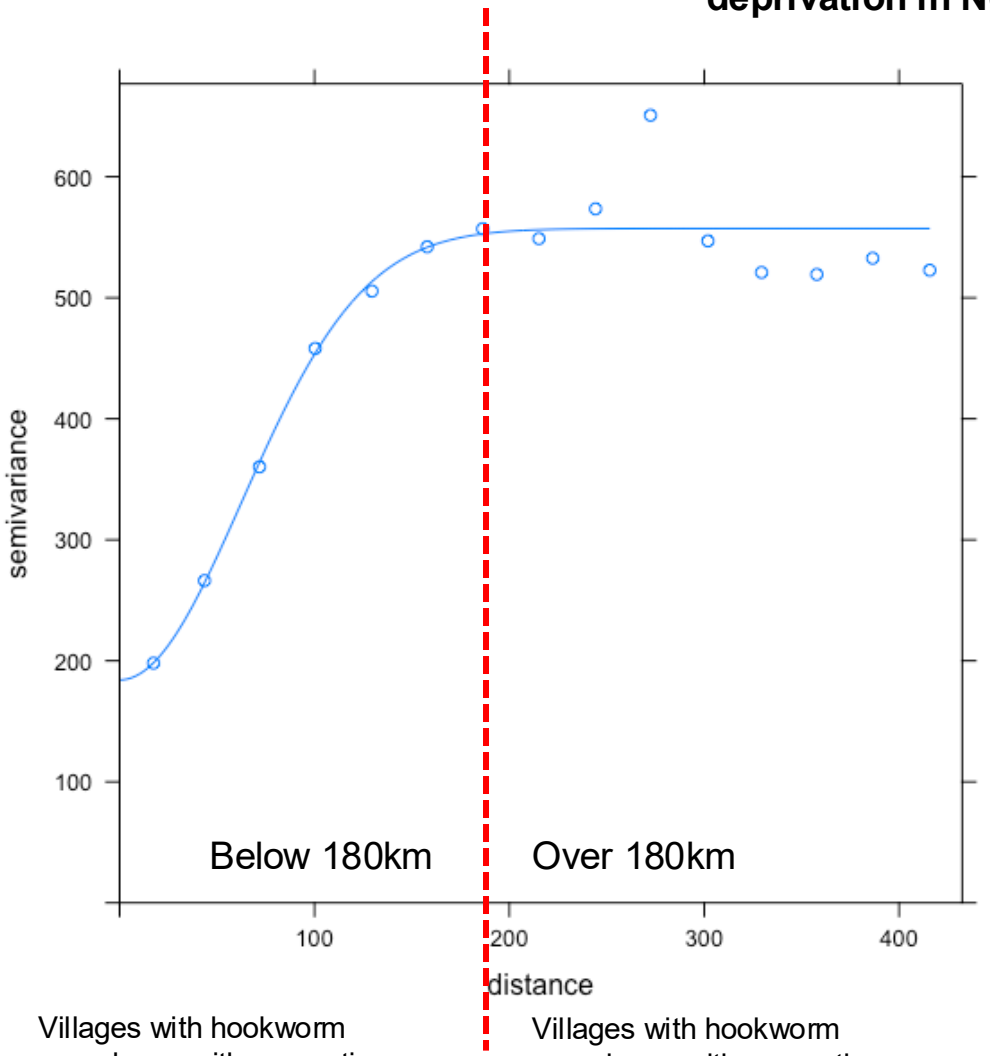
In short, what distance decay (in the context of Tobler's theory) says**:**

- 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

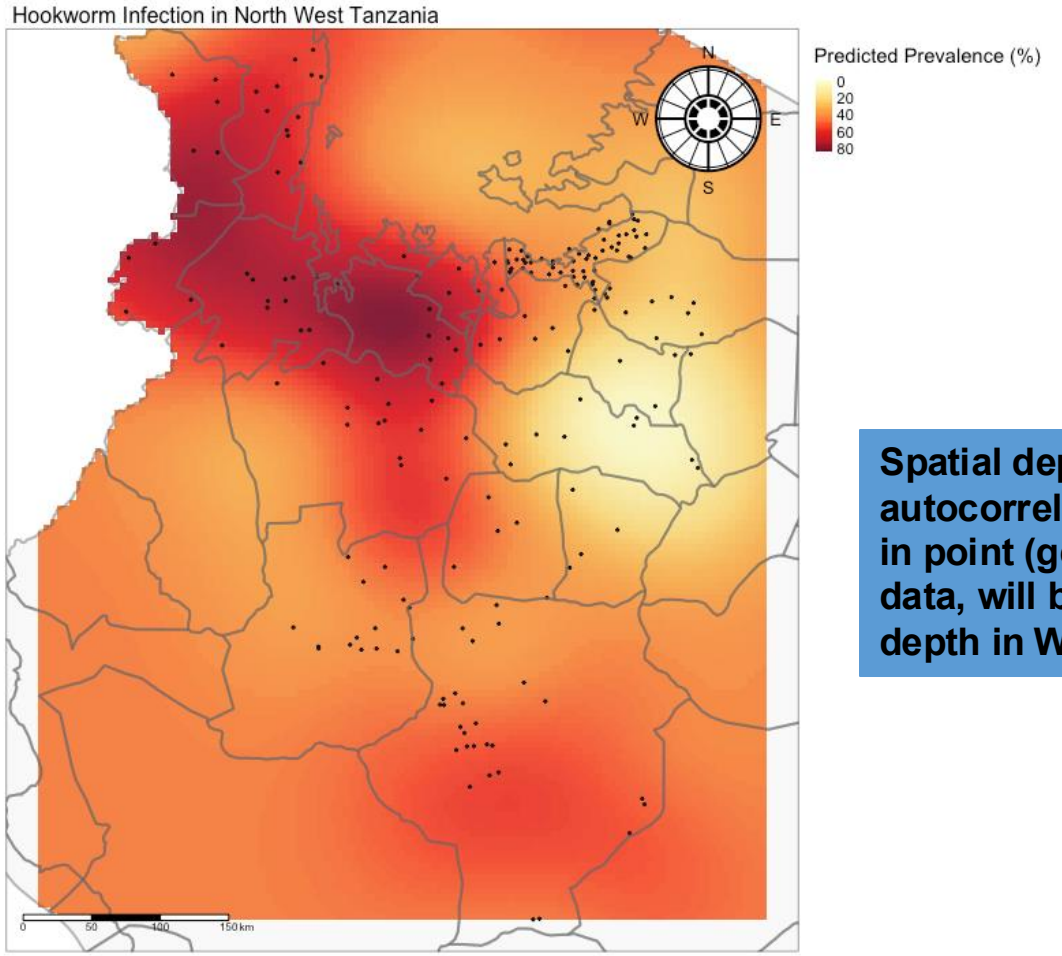
Using Kriging to spatially predict areas with intense hookworm infection associated with socioeconomic deprivation in Northwestern Tanzania



Villages with hookworm prevalence with separation distance below 180km have similar prevalence

Villages with hookworm prevalence with separation distance 180km and over don't have related prevalence

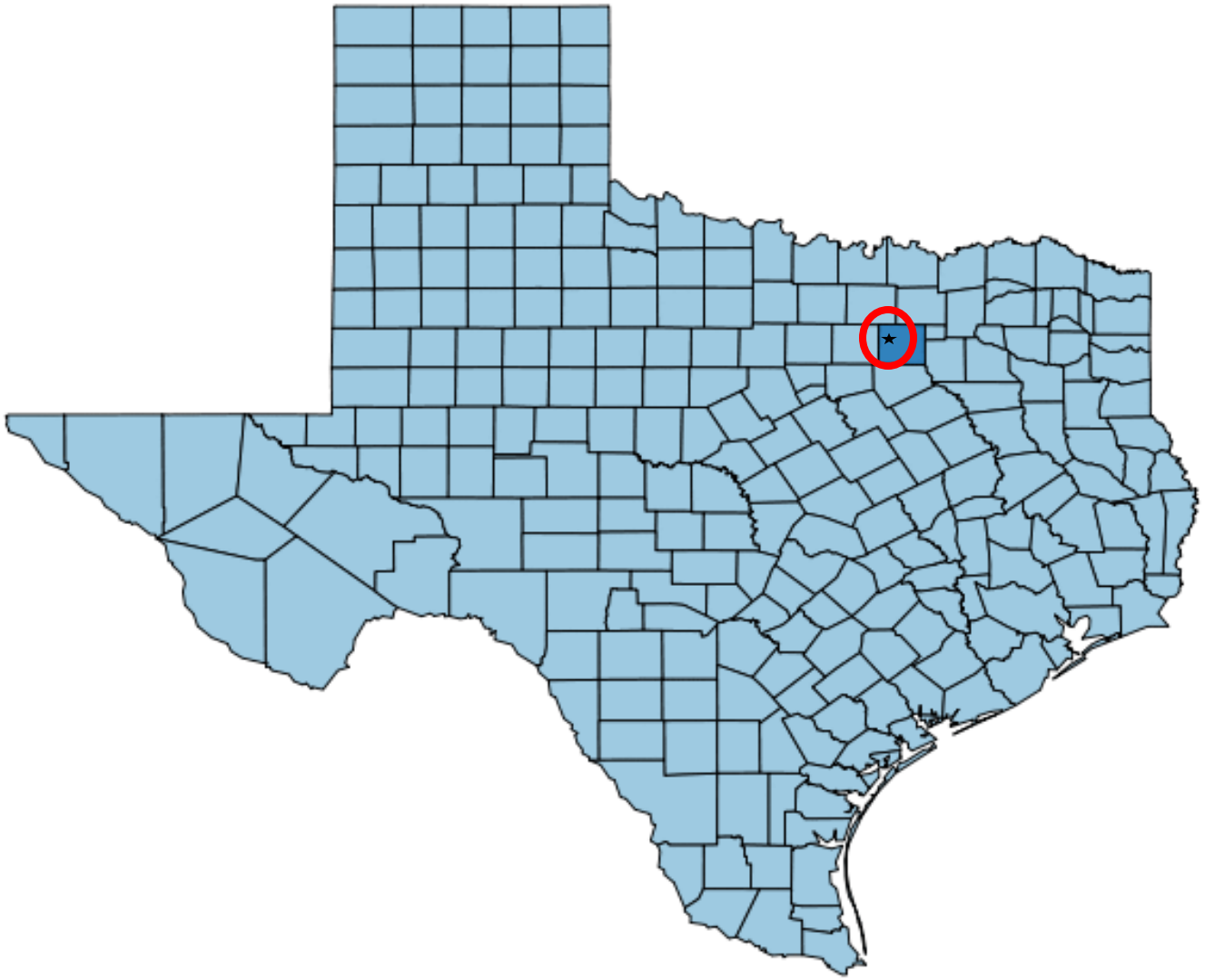
We used the concept of distance decay, or in this 'separation distance' to account for spatial dependence between our survey points



Spatial dependence and autocorrelation, particularly in point (geo-statistical) data, will be covered in-depth in Week 6

We have survey points (of villages) reporting prevalence of hookworm in Northwestern Tanzania. We assumed there some spatial dependence in the prevalence of hookworm in these locations to predict prevalence where there are no information (or points). We use this information on form the left panel to build our model for making geostatistical predictions.

Spatial Spillover: "…where you are in geographic space matters" an event in one location can someway have an impact on other events in neighboring areas

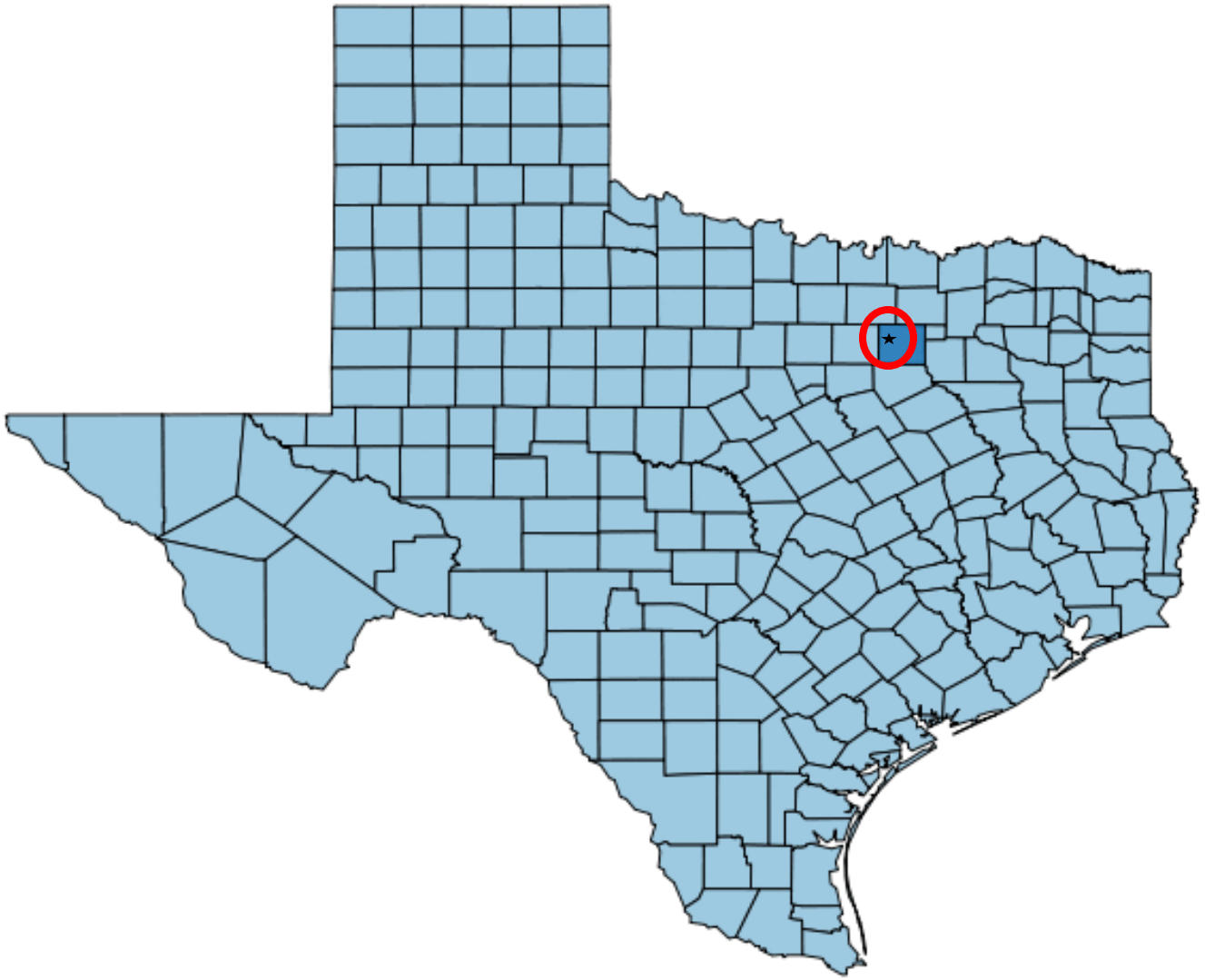


In Spatial Statistics, especially in Spatial Regression, Geostatistics – we often try to account for **spatial spillover** effects in our models

There are four types of spillovers:

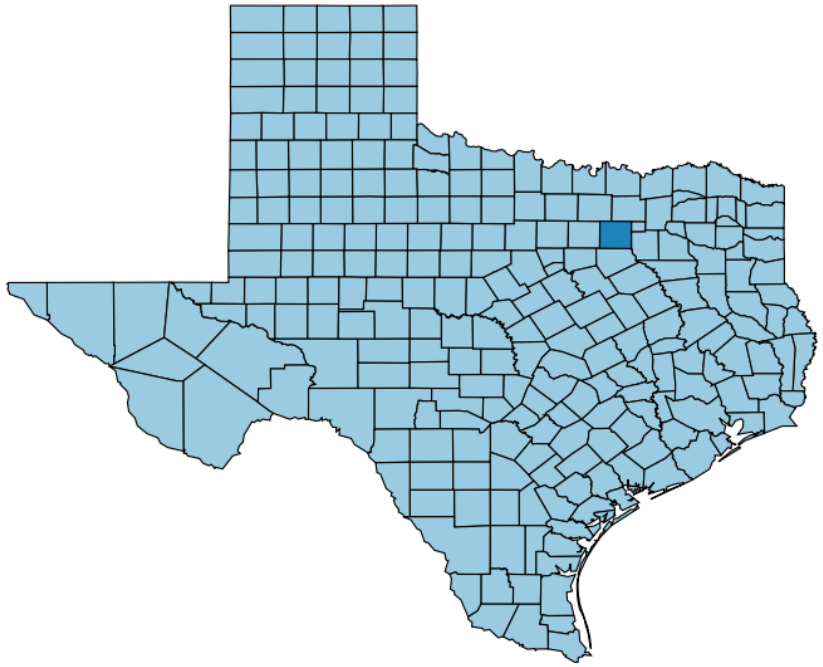
- No spillover
- Local spillover
- Global spillover
- Rippling spillovers

Spatial Spillover: "…where you are in geographic space matters" an event in one location can someway have an impact on other events in neighboring areas



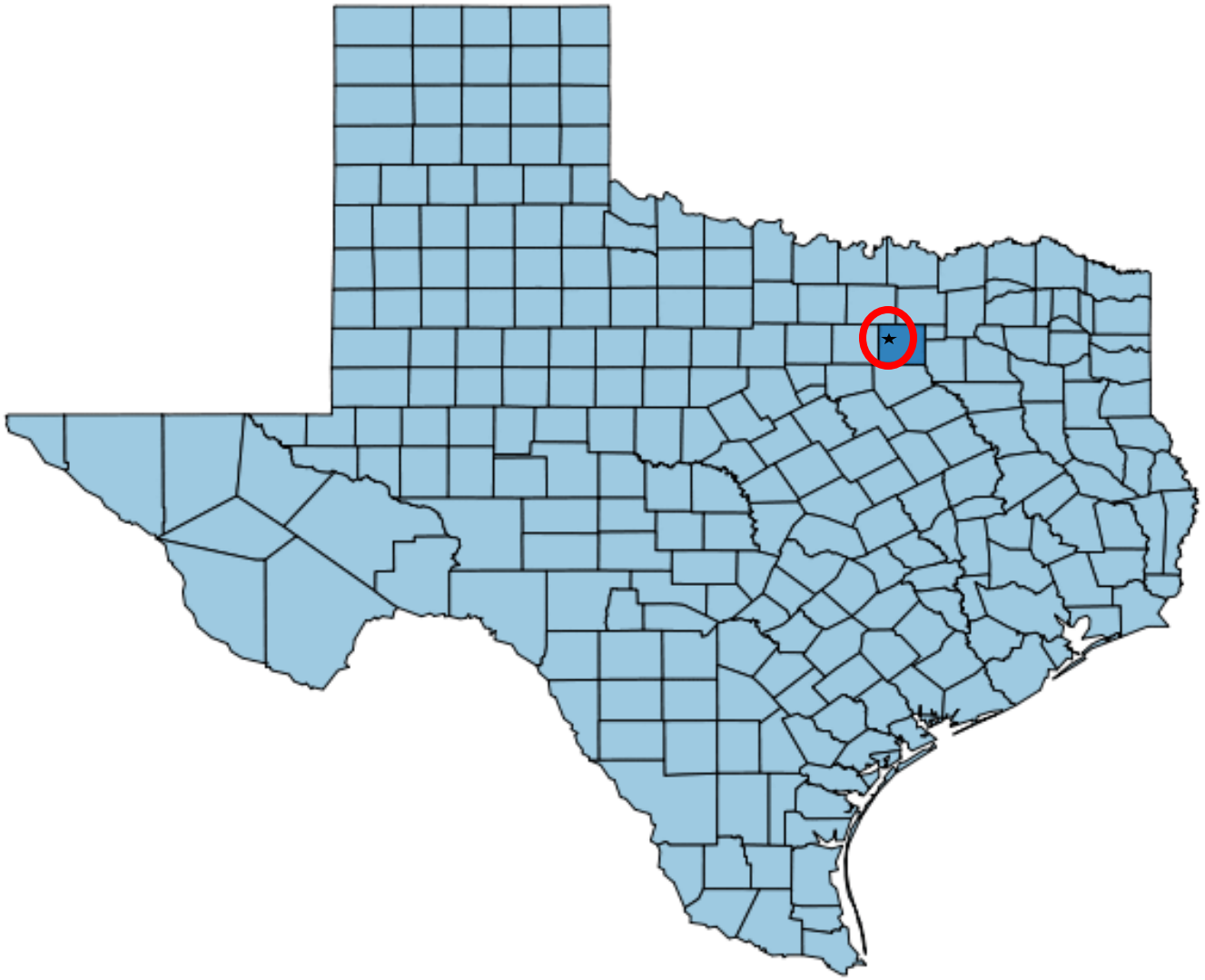
In Spatial Statistics, especially in Spatial Regression, Geostatistics – we often try to account for **spatial spillover** effects in our models

No spillover (independence)



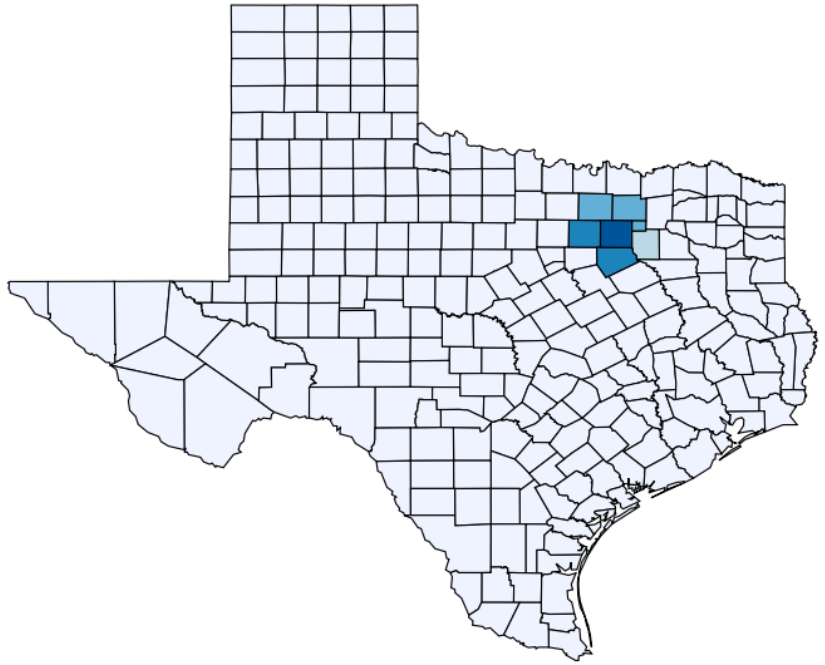
A change in unemployment in that dark blue area will only have an impact on crime rates within its own area, but does not influence crime rates in neighboring areas

Spatial Spillover: "…where you are in geographic space matters" an event in one location can someway have an impact on other events in neighboring areas



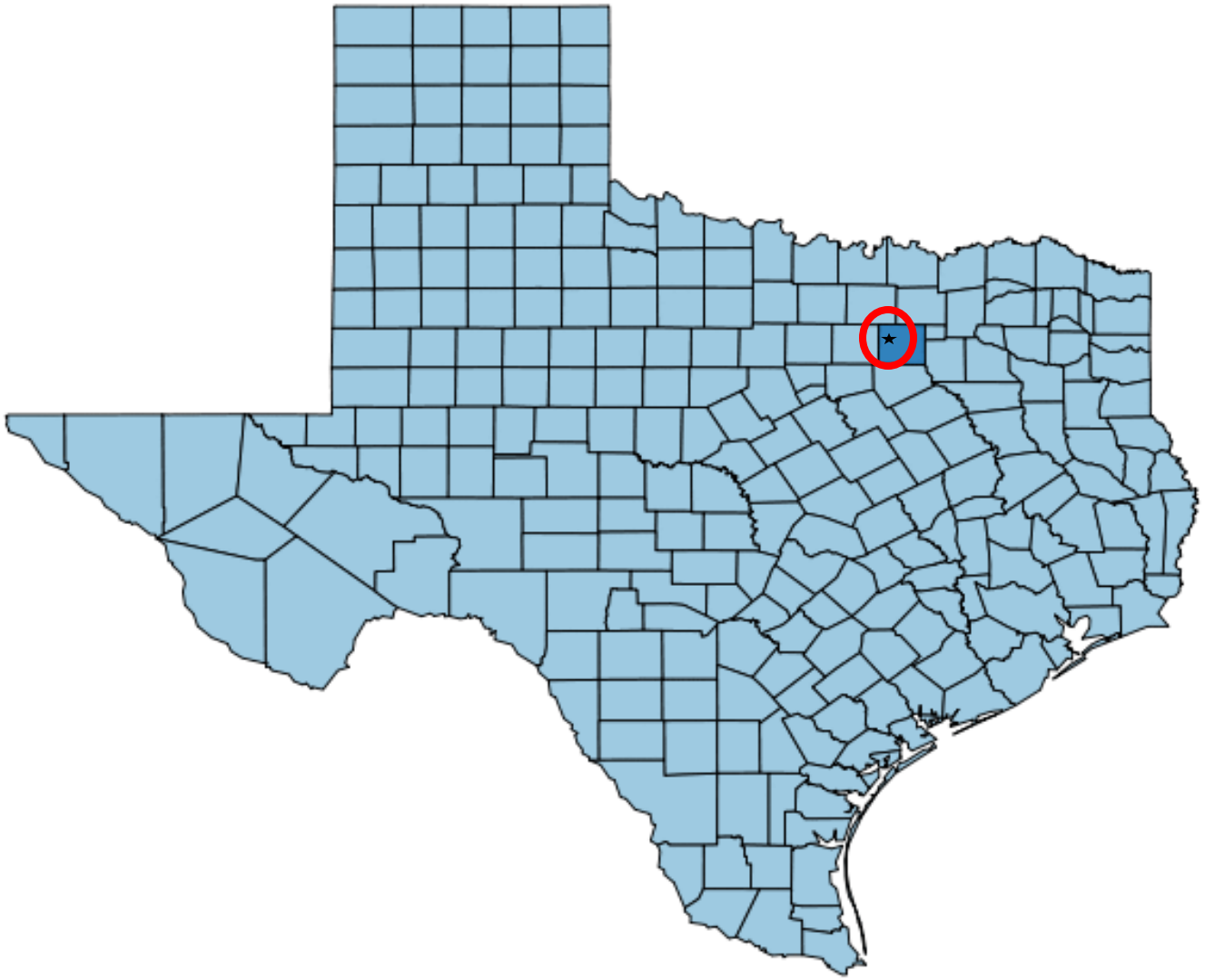
In Spatial Statistics, especially in Spatial Regression, Geostatistics – we often try to account for **spatial spillover** effects in our models

Local spillover (dependence)



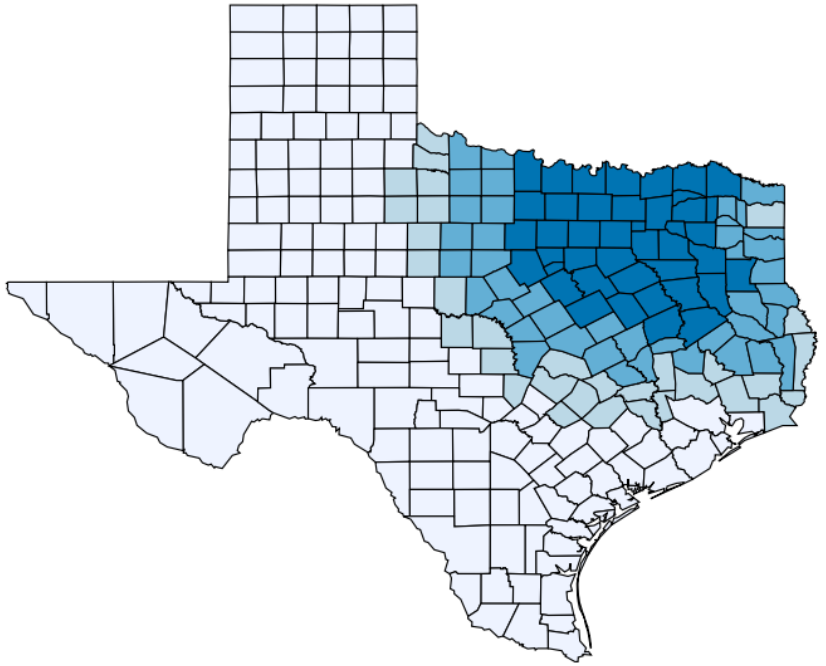
A change in unemployment in that dark blue area (in the center) will not only have an impact on crime rates within its own area, but it will also have a direct influence on crime rates in neighboring areas only.

Spatial Spillover: "…where you are in geographic space matters" an event in one location can someway have an impact on other events in neighboring areas



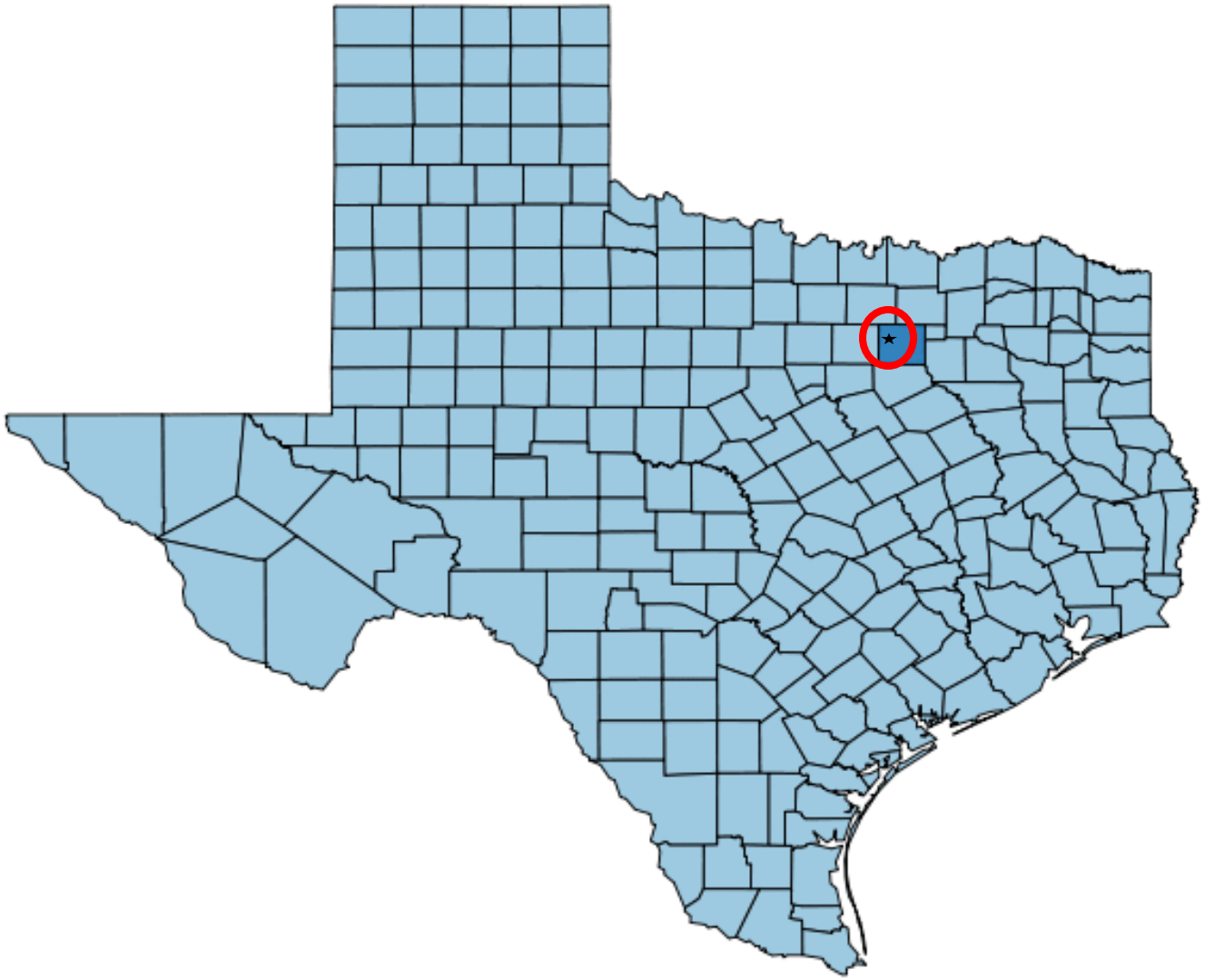
In Spatial Statistics, especially in Spatial Regression, Geostatistics – we often try to account for **spatial spillover** effects in our models

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 someway 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.

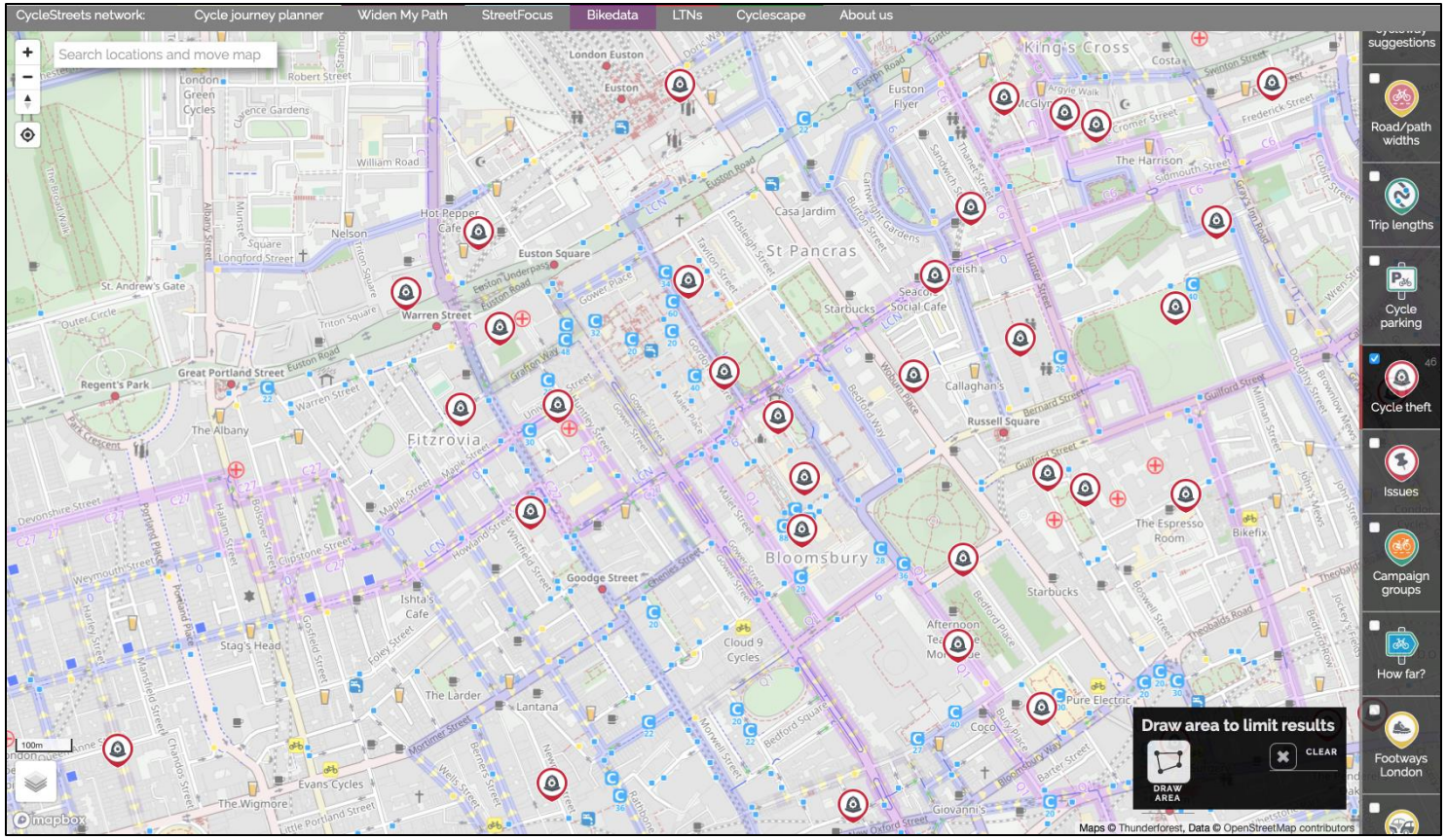
e.g.,

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

Spatial dependence in spillovers, particularly in areal data from a modelling approach, will be covered in-depth in Week 9

What is Spatial Data & its Features?

Point Pattern Data (PPD)



Represents point locations of bicycle thefts in Central London area
Source: BikeData.CycleStreets network https://bikedata.cyclestreets.net/

Key Characteristics

The main interest is the occurrences of an event at a points (or points). These events occur at "random" at any given geographic space and time.

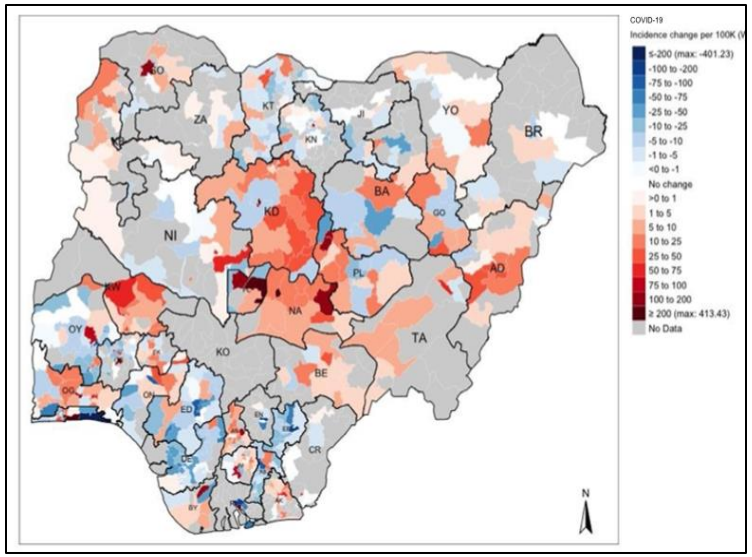
Examples of PPDs events (or outcomes):

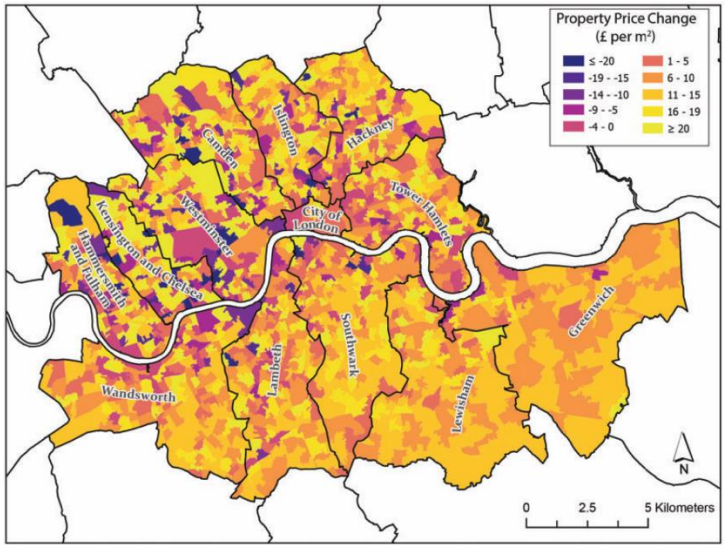
- Point locations of burglaries
- Riots
- Locations of car collisions etc.,
- Locations of where an adult tree needs to be replanted

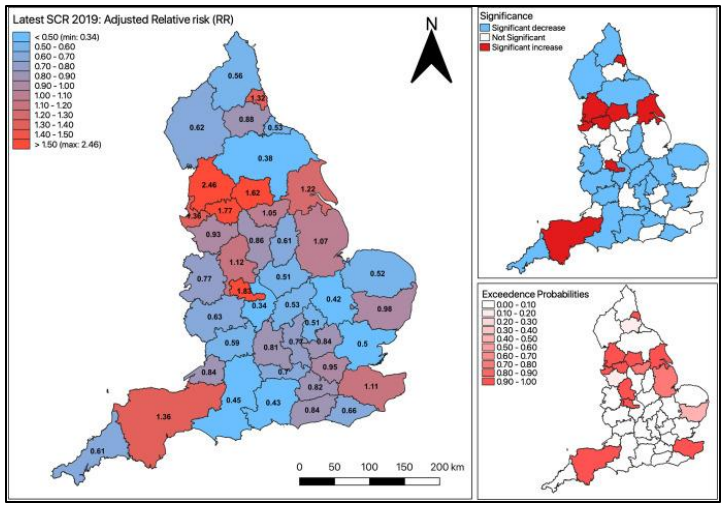
Some PPDs events may carry additional information that may describe the occurrence of an observed event (or outcome):

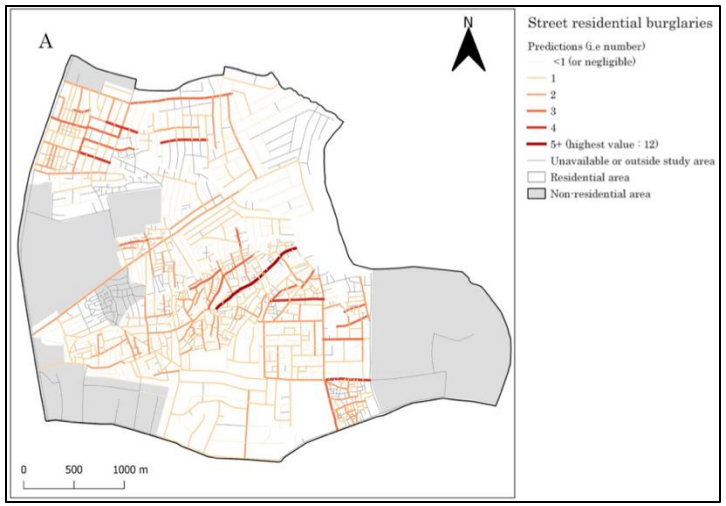
- Burglary: Type of premise that was burgled, time of day the burglary occurred etc.,
- Car collision: type of road, weather condition etc.

Aggregated Data









Key Characteristics

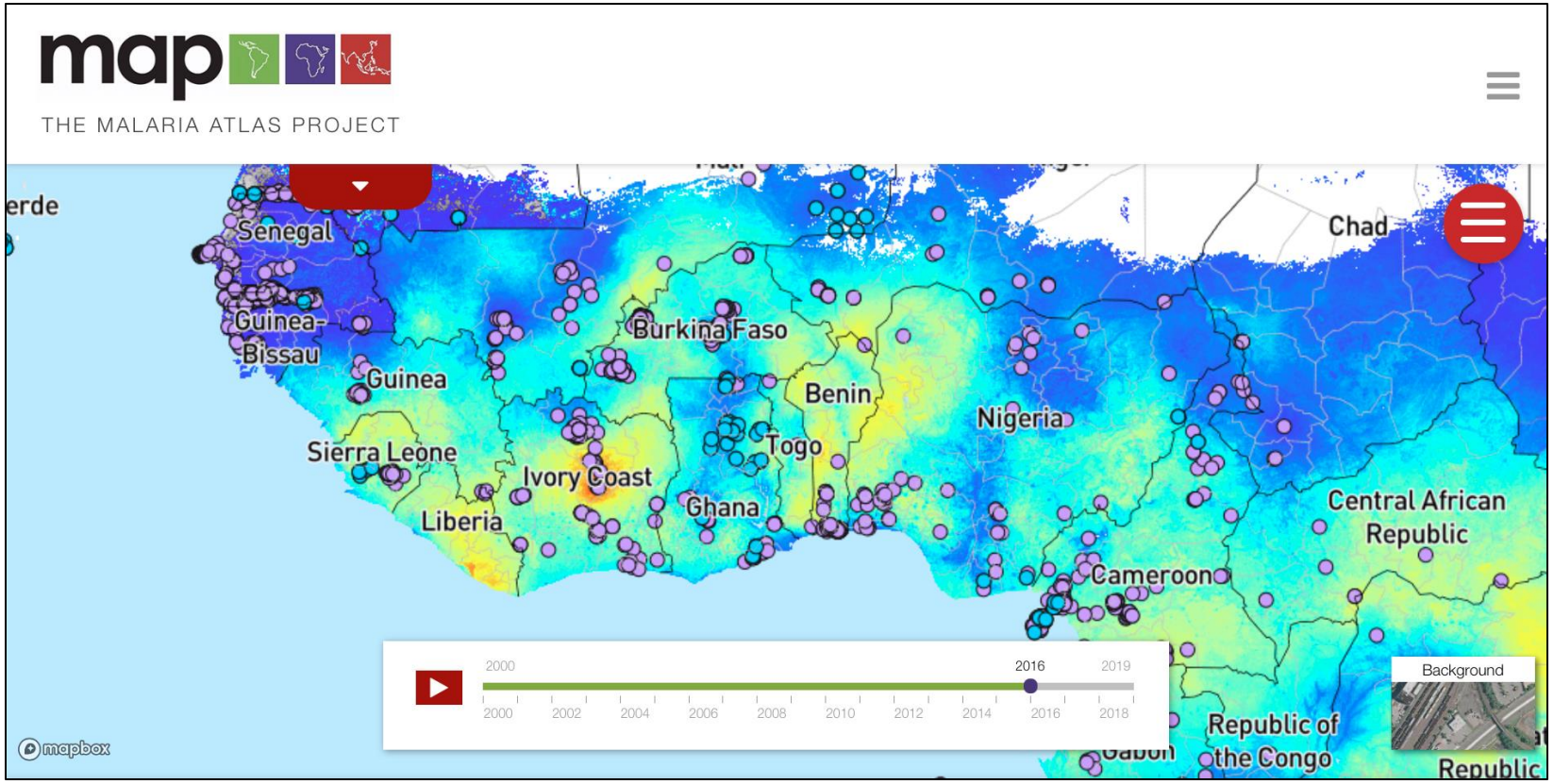
The main interest the quantity of interest defined for line segments, areas, or regions.

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

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

Sources:
Musah, A., *et al*. (2020): https://doi.org/10.1016/j.apgeog.2019.102126
Todd, J., *et al*. (2021): https://doi.org/10.1177/23998083211001836
Li, L., *et al*. (2022): https://doi.org/10.1016/j.apgeog.2022.102718
Elimian, K., *et al*. (2022): http://dx.doi.org/10.1136/bmjopen-2022-063703

Geostatistical Data



Sampled points are surveys on prevalence of malaria, which were used to make survey predication of prevalence at unsampled areas in Sub-Saharan Africa

Source: The Malaria Atlas Project https://malariaatlas.org

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 "**sampled**" points with continuous data values:

- Land surface elevation
- Diffusion of ambient air pollutants
- Environmental suitability for breeding habitats of mosquitoes
- Community surveys pertained to disease burden

Non-spatial context & data structure:

	ATTRIBUTE			
	Variable 1	**Variable 2**	$\cdots$	**Variable n**
Entity 1	$attribute_{11}$	$attribute_{12}$	$\cdots$	$attribute_{1n}$
Entity 2	$attribute_{21}$	$attribute_{22}$	$\cdots$	$attribute_{2n}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Entity m	$attribute_{m1}$	$attribute_{m2}$	$\cdots$	$attribute_{mn}$

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 the 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	$attribute_{m1}$	$attribute_{m2}$	⋯	$attribute_{mn}$

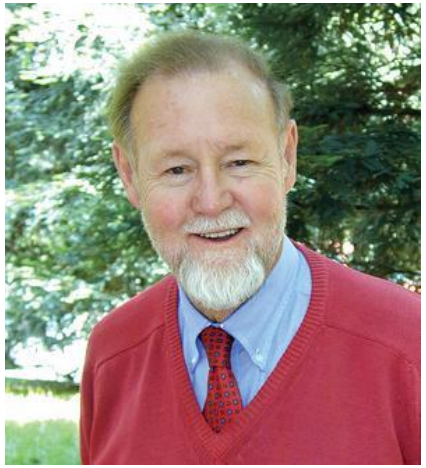
In this example, what defines the entity's geographic location are **X, Y GPS coordinates.**

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


30


GDAL


OGC®
Making location count.


Arc
ESRI GIS


ArcGIS® Pro


OSGeo
Project


CARTO


PROJ.4


gv
SIG


OPEN JUMP
Jump Unified Mapping Platform


R Studio®


30


GDAL


OGC®
Making location count.


Arc
ESRI
GIS


ArcGIS® Pro


OSGeo
Project


CARTO


gv
SIG


PROJ.4


OPEN JUMP
Jump Unified Mapping Platform


R Studio®

R (Standard)

RStudio (Best)

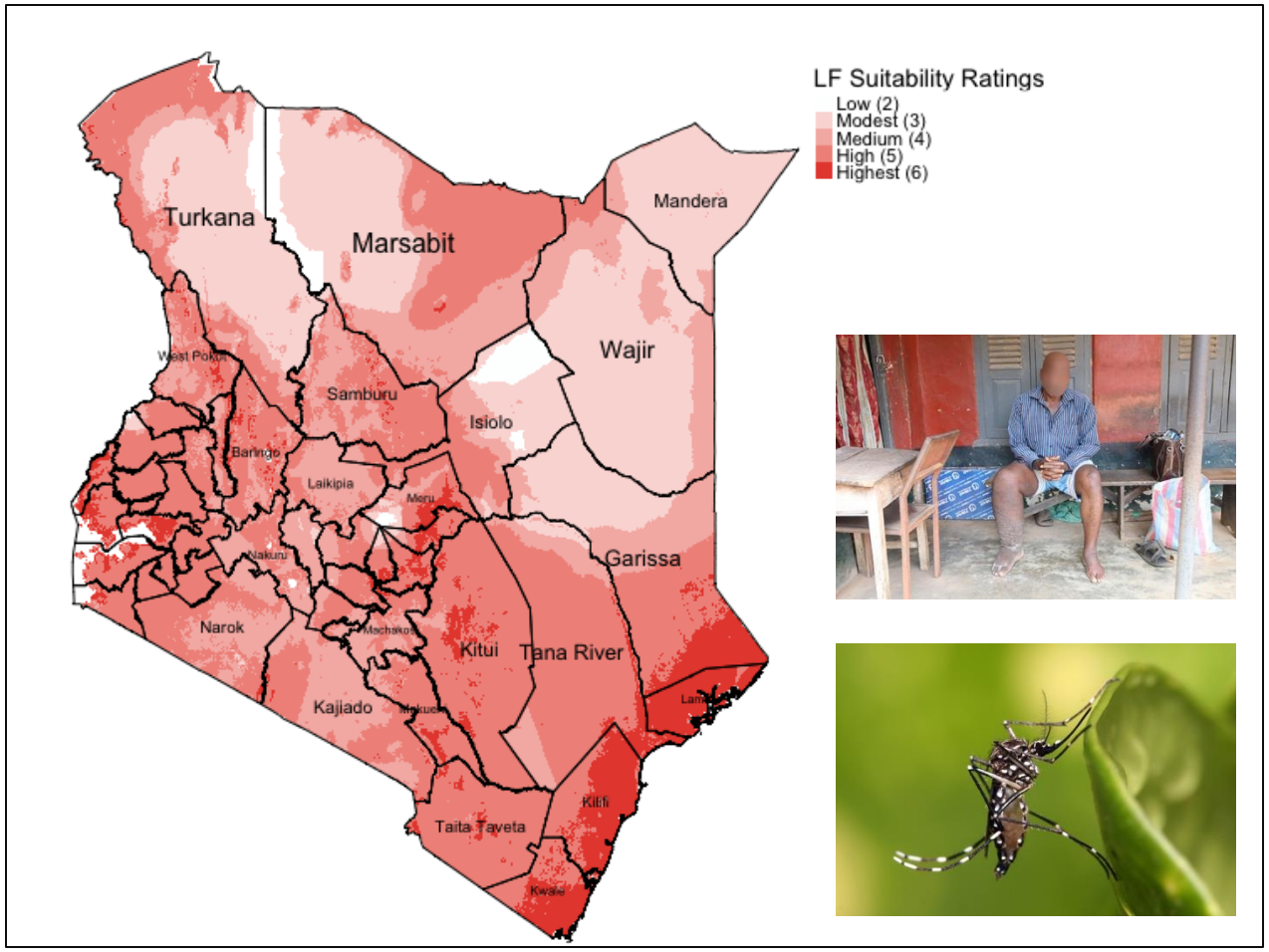
There are two version of the software: 1.) R, and 2.) RStudio; The second is much preferred as its straightforward and intuitive.

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.

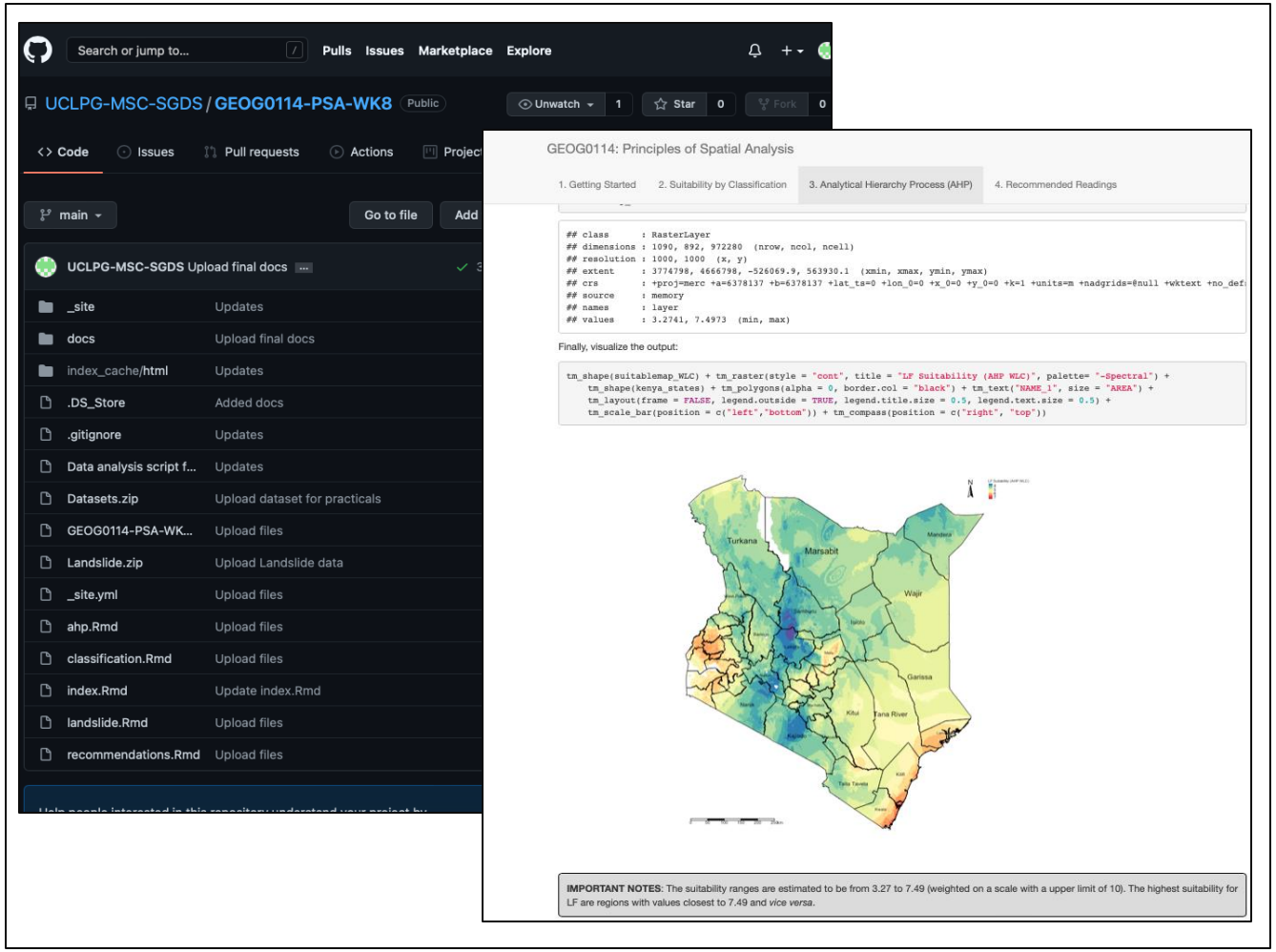
Sources:
1. Global Atlas for Helminths Infection (http://www.thiswormyworld.org)
2. ESPEN (https://espen.afro.who.int)

… 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.,)

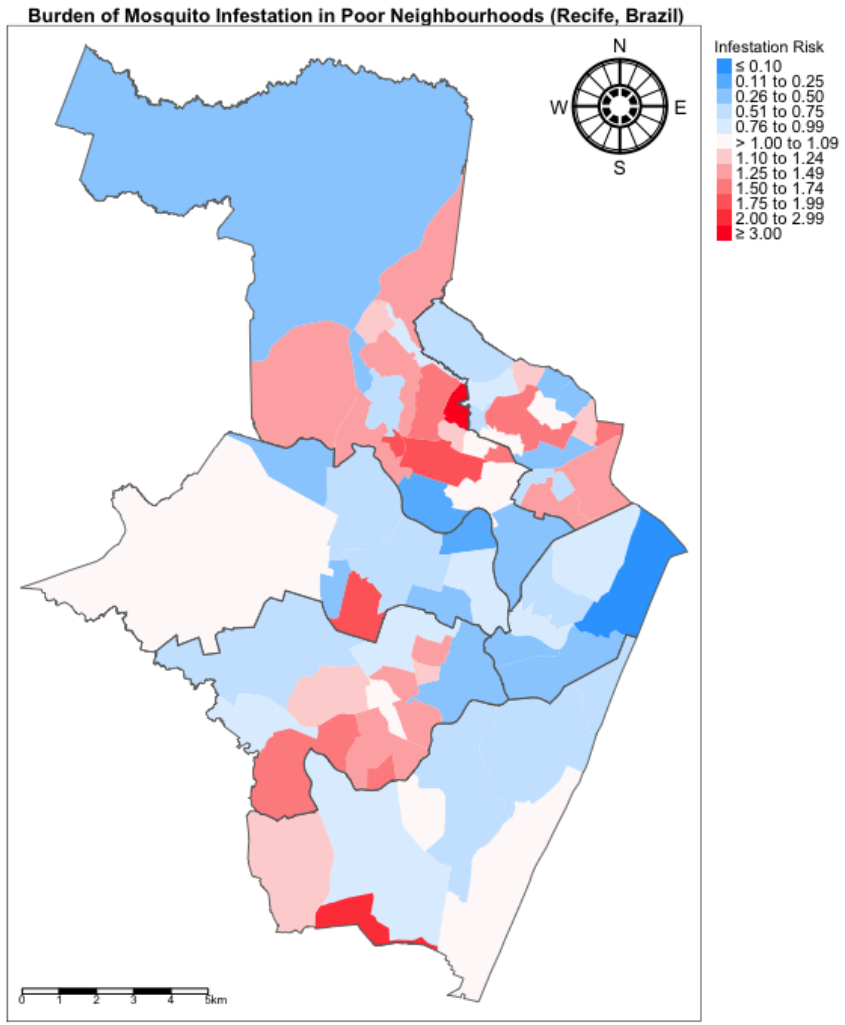
You can literally pull-off some really creative stuff like generating websites, accessing tools via APIs etc.



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.

Sources:
1. GitHub (https://github.com)

Example of a basic code structure in RStudio




```
# comment: activate packages for performing GIS in R
library("sf")
library("tmap")

# comment: add neighbourhood shapefile w/mosquito infestation data using read_sf()
recife.neighbourhoods <- read_sf("Recife_neighb_epsg3857_fixed.shp")
recife.healthzone <- read_sf("Recife_regions_epsg3857_fixed.shp")

# comment: assigning labels for the risk estimate legends
RiskCategorylist <- c("\u2264 0.10", "0.11 to 0.25", "0.26 to 0.50", "0.51 to 0.75",
"0.76 to 0.99",">1.00 to 1.09", "1.10 to 1.24", "1.25 to 1.49", "1.50 to 1.74", "1.75
to 1.99", "2.00 to 2.99", "\u2265 3.00")

# comment: generating the divergent color scheme from Blues to Red spectrum
RRPalette <- c("#33a6fe", "#65bafe", "#98cffe", "#cbe6fe", "#dfeffe", "#fef9f9",
"#fed5d5", "#feb1b1", "#fe8e8e", "#fe6a6a", "#fe4646", "#fe2424", "#fe0000")

# comment: map of risk of infestation
tm_shape(recife.neighbourhoods) +
  tm_fill("RelativeRiskCat",
          style = "cat",
          title = "Infestation Risk",
          palette = RRPalette,
          labels = RiskCategorylist) +
tm_shape(recife.healthzone) +
  tm_polygons(alpha = 0, border.alpha = 0.90) +
  tm_layout(frame = TRUE,
            main.title = "Mosquito Infestation in Neighbourhoods (Brazil)",
            main.title.size = 0.8,
            main.title.position = 0.02,
            main.title.fontface = 2,
            legend.outside = TRUE,
            legend.outside.position = "right",
            legend.title.size = 0.8,
            legend.text.size = 0.7) +
  tm_scale_bar(position = c("left", "bottom")) +
  tm_compass(type = "radar", show.labels = 2, position = c("right", "top"))
```

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 three key words: **Spatial autocorrelations** & **Distance decay** effect, and 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?