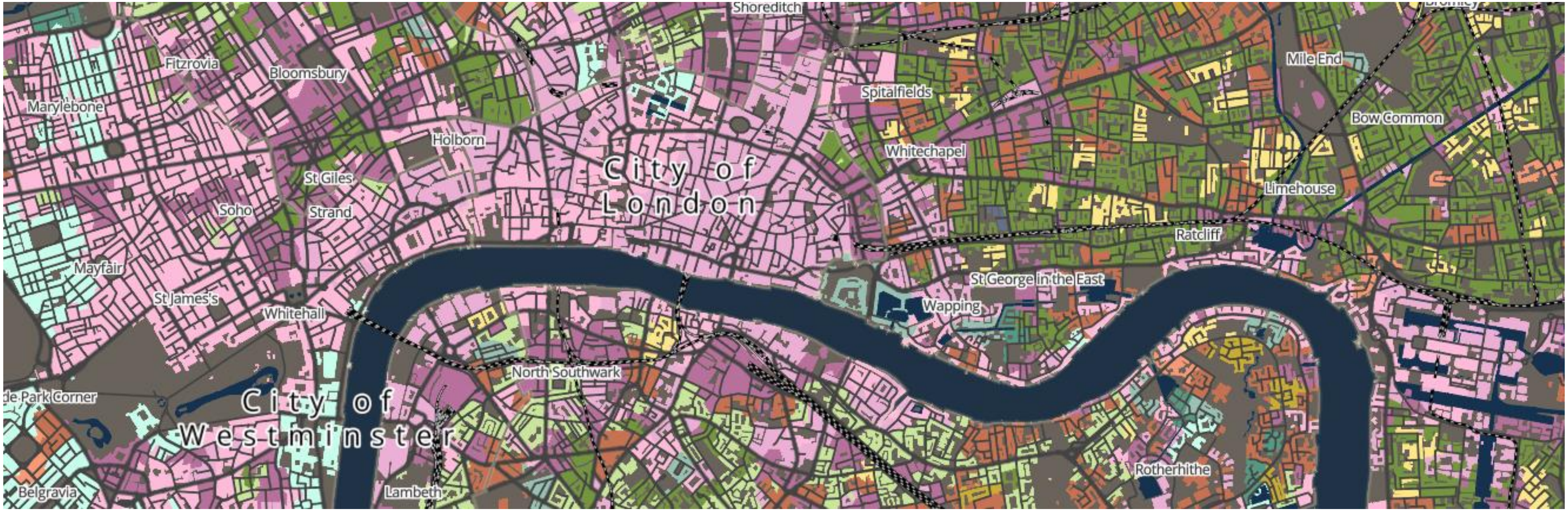
Geocomputation

Spatial Models



Module outline

W1	Reproducible Spatial Analysis	Core Spatial Analysis
W2	Spatial Queries and Geometric Operations	
W3	Point Pattern Analysis	
W4	Spatial Autocorrelation	
W5	Spatial Models	
W6	Raster Data Analysis	Applied Spatial Analysis
W7	Geodemographic Classification	
W8	Accessibility Analysis	
W9	Beyond the Choropleth	Data Visualisation
W10	Complex Visualisations	

This week

- Managing spatial data
- Linear Models
- Spatial Models
- Assessment: Social Atlas

Managing spatial data

- R has the capacity to read, load and store a range of file formats.
- Functions in both the base R library plus a huge host of software-specific packages (e.g. STATA, SPSS) for reading, writing and converting data between different file formats associated with those specific software (e.g. from a SPSS file to a `csv` etc.).
- Base R does not handle the reading, loading, and storing of spatial data.

Managing spatial data

- How do we read in and deal with spatial data?
- GDAL: Geospatial Data Abstraction Library (*reading, writing*)
- GEOS: Geometry Engine Open Source (*spatial operations*)
- PROJ: Cartographic Projection Library (*coordinate transformations)*

GEOS Geometry Engine Open Source

Managing spatial data

- The `sf` (simple features) package facilitates the storage, access and management of geometric objects stored as simple features in R.
- Importantly: `sf` objects are dataframes with a geometry column.

Managing spatial data

```
## Simple feature collection with 100 features and 6 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -84.32385 ymin: 33.88199 xmax: -75.45698 ymax: 36.58965
## epsg (SRID):    4267
## proj4string:    +proj=longlat +datum=NAD27 +no_defs
## precision:      double (default; no precision model)
## First 3 features:
##   BIR74 SID74 NWBIR74 BIR79 SID79 NWBIR79                           geom
## 1  1091     1      10  1364     0      19 MULTIPOLYGON(((-81.47275543...
## 2   487     0      10   542     3      12 MULTIPOLYGON(((-81.23989105...
## 3  3188     5     208  3616     6     260 MULTIPOLYGON(((-80.45634460...
```

Simple feature

Simple feature geometry list-colum (sfc)

Simple feature geometry (sfg)

Managing spatial data

`sf`

- ‘Support for simple features, a standardized way to encode spatial vector data’.
- Fully compliant with the dataframe format (*tidyverse*).

`sp`

- ‘Classes and methods for spatial data’.
- First development in using spatial data in R (2005).
- Not fully compliant with the dataframe format.

Before we start

- Go to www.menti.com
- Use code: 7321 6950

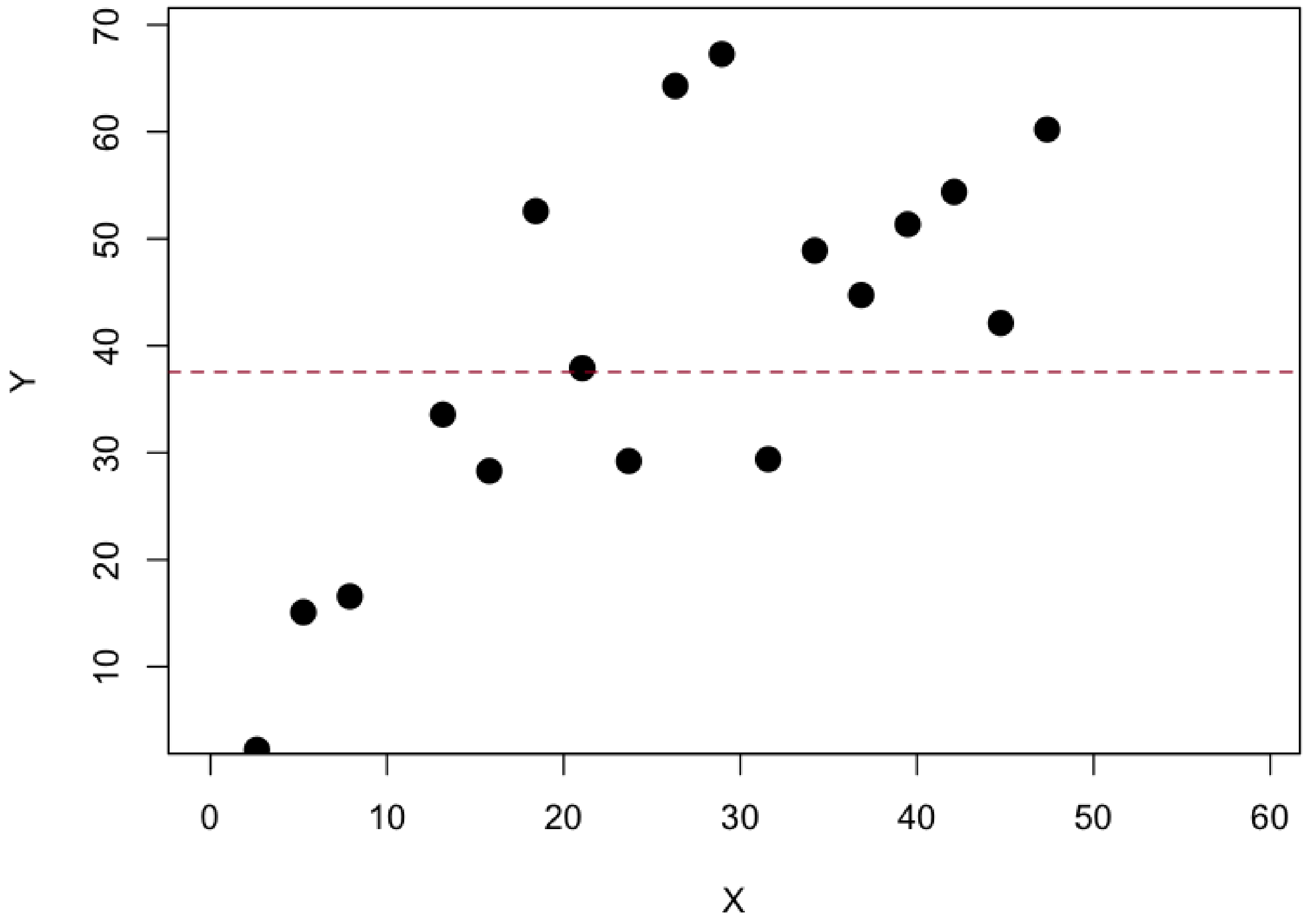
Linear models

- We often want to know to predict how changes in one variable will affect another variable.
- To do this we can use a regression model to examine the relationship between a dependent variable (y) and one or more independent variables (x).

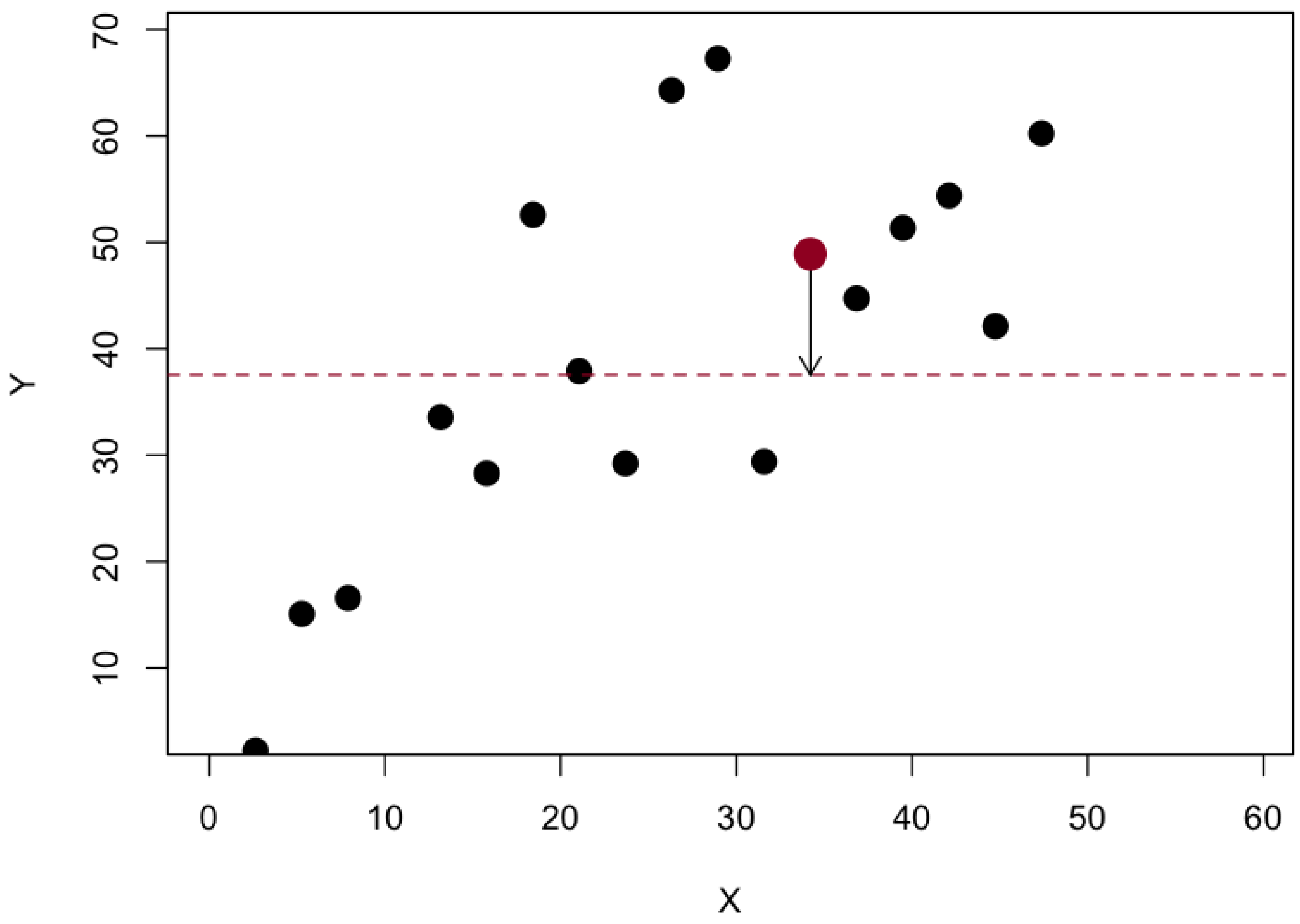
Linear models

- Linear regression uses a line to summarise the relationship between x and y.
- The aim to find the line which best represents the relationships in the data.
- Typically, this line will not pass through every data point meaning we cannot predict y exactly.

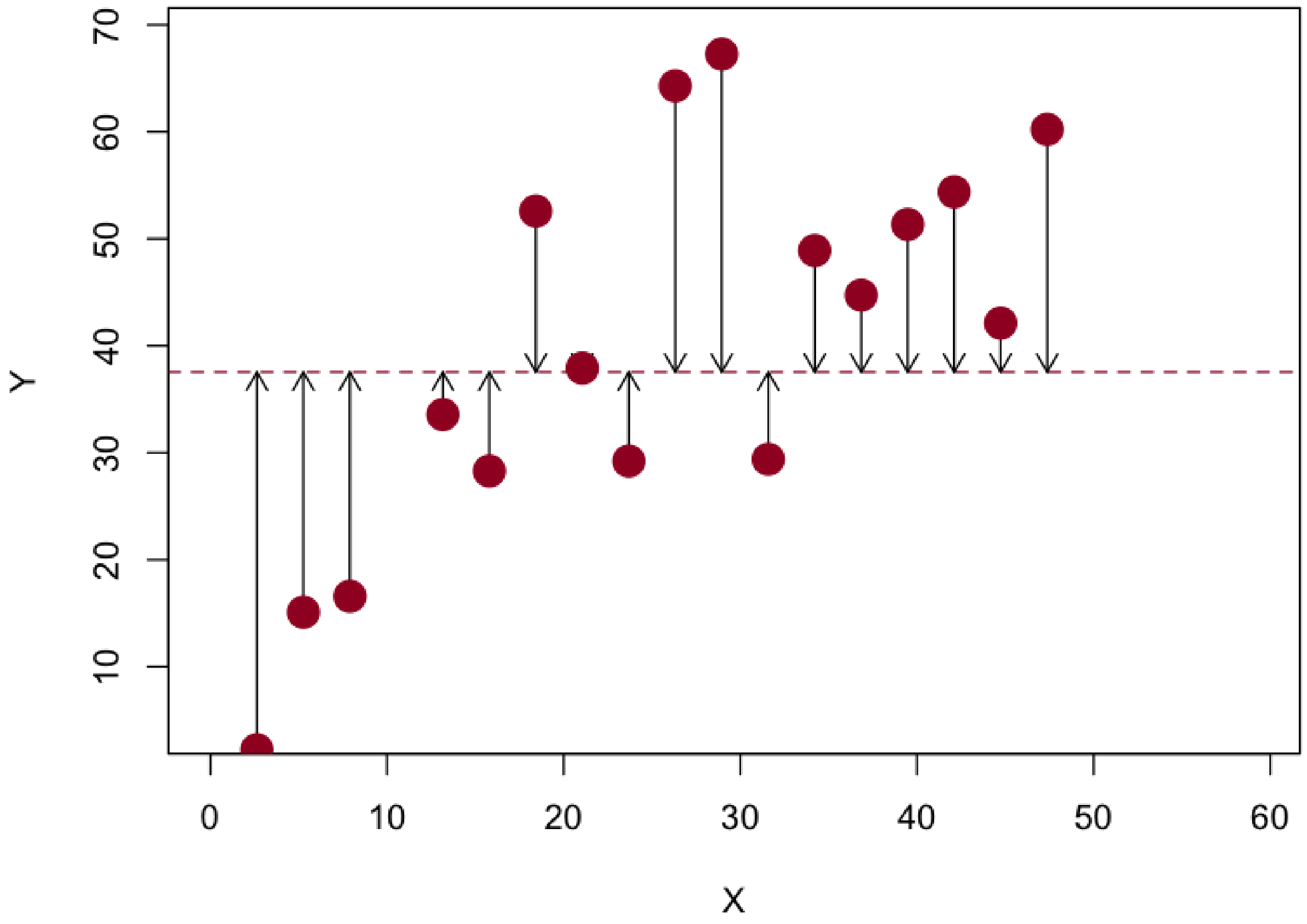
Linear models



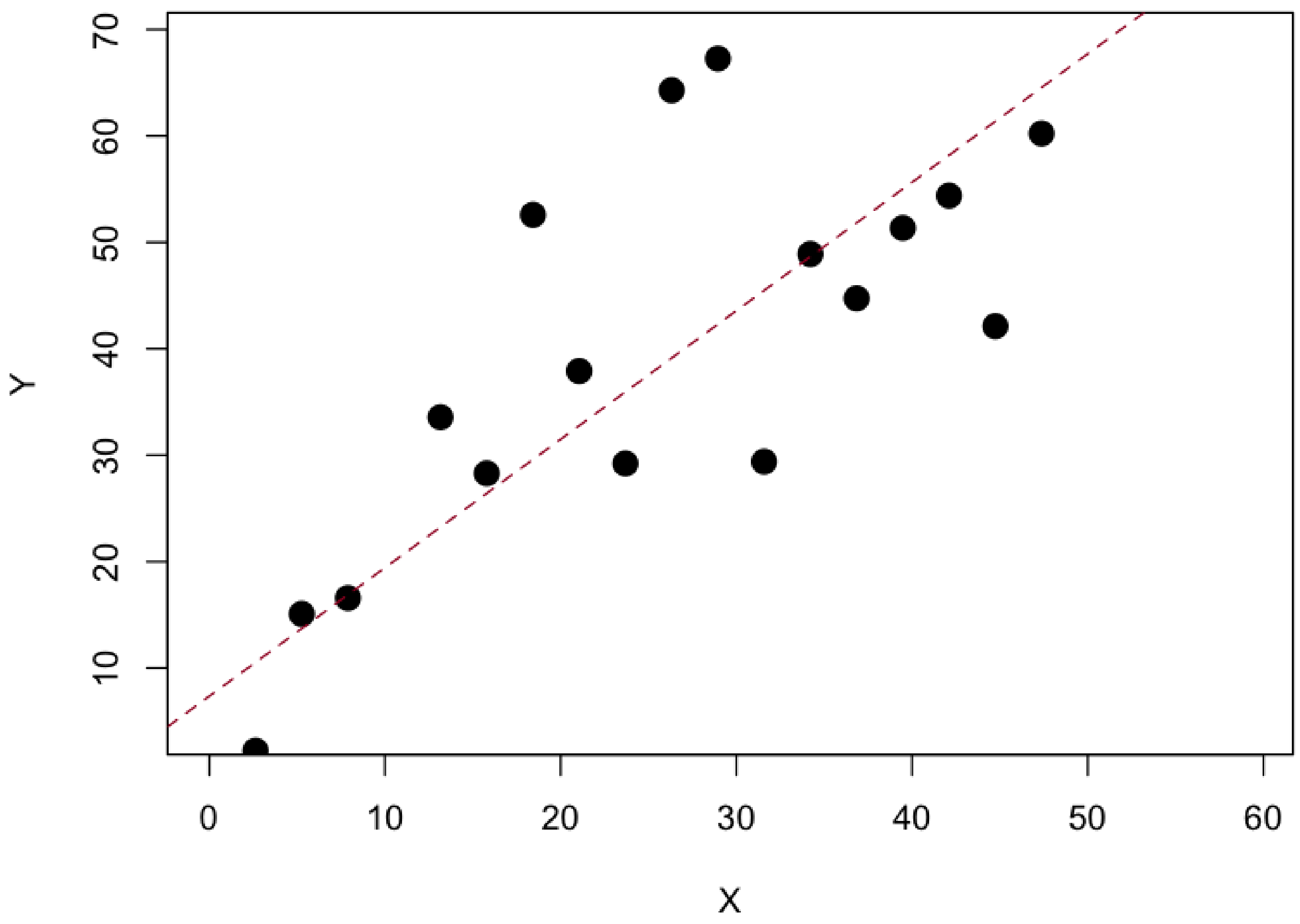
Linear models



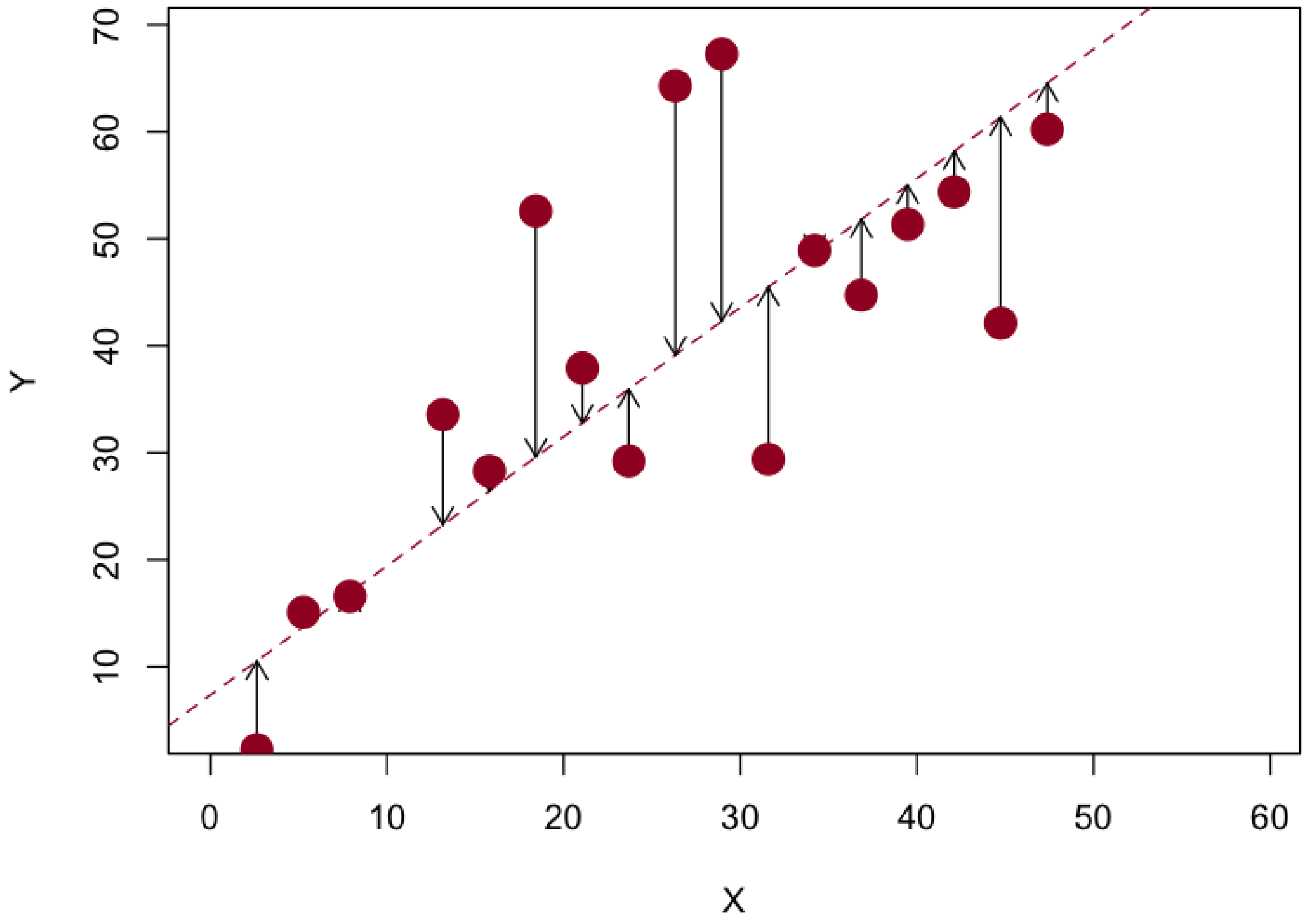
Linear models



Linear models



Linear models



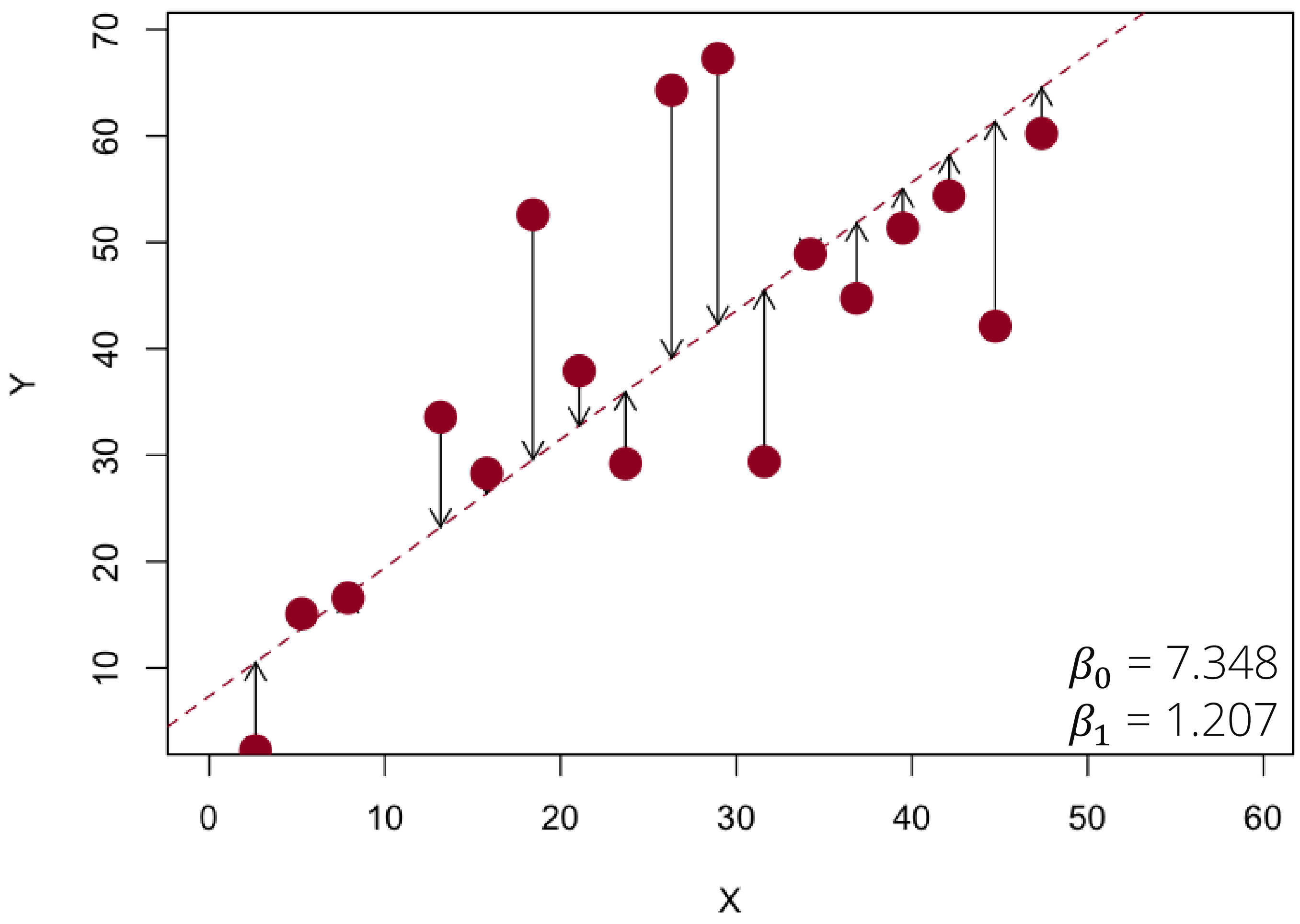
Linear models

- Ordinary Least Squares (OLS) regression:

$$\hat{y} = \beta_0 + \beta_1 x$$

- The $\boldsymbol{\beta}$ terms are coefficients that define the regression line.
- The model estimates these parameters to find the line that gives the smallest sum of squared errors: Ordinary Least Squares (OLS) regression.

Linear models



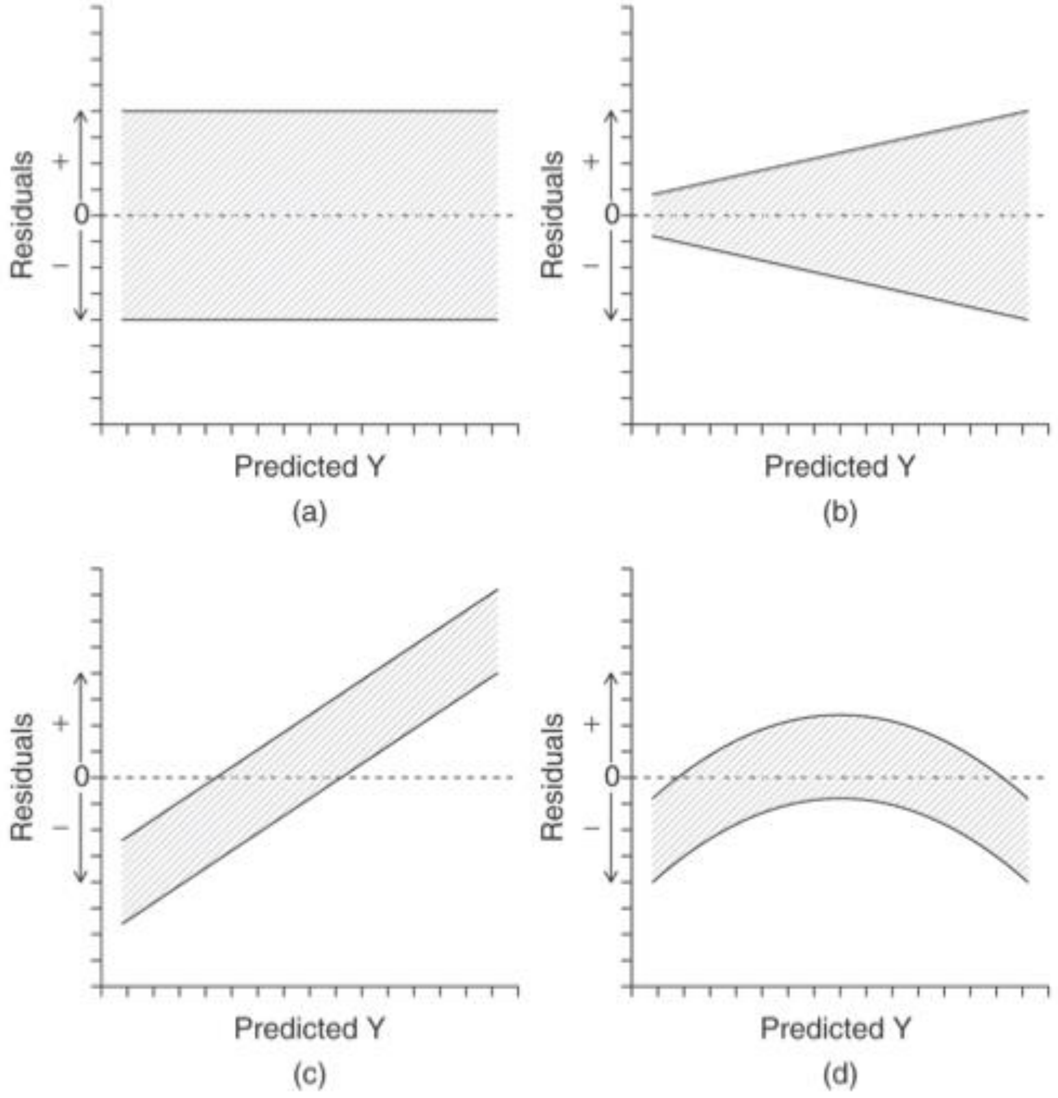
Linear models

Distribution of dependent variable	Example	Suitable model
Continuous	Income	Linear regression
Binary	Employment status	Logistic regression
Binomial	Proportion of homeowners	Logistic regression
Count	Number of crimes	Poisson regression

Linear models

- Important assumptions: homoscedasticity.
- Violating this assumption can lead to inefficient estimates and unreliable results.

Linear models



Linear models

When building a model based on spatial data:

- Map the residuals of the linear model to visually inspect for spatial patterns.
- Calculate Moran’s I statistic on the residuals to assess spatial autocorrelation.
- If spatial autocorrelation is present, fit a spatial linear model to account for it.
- Recalculate Moran’s I statistic on the residuals of the spatial model to confirm that the autocorrelation has been addressed.

Spatial models

A spatial error model adjusts for spatial autocorrelation by adding a spatially lagged error term to the regression equation:

$$y = X\beta + v, v = \lambda W v + \epsilon$$

where $X\beta$ represents the standard regression components, λ is a spatial autoregressive parameter, W represents the spatially weights matrix, and u is a vector of spatially autocorrelated errors.

Spatial models

A spatial lag model incorporates a spatially lagged dependent variable, which is the weighted sum of the dependent variable values in neighboring locations, into the regression equation:

$$y = \rho W y + X\beta + \epsilon$$

where ρ is the spatial autoregressive coefficient, Wy represents the spatially lagged dependent variable, and $X\beta$ represents the standard regression components.

Spatial models

- Both the spatial error and spatial lag models assume that the relationships between variables are the same across the study area, with adjustments made only for spatial dependencies.
- A Lagrange Multiplier Test can be used to make a decision as to which of these two models is most appropriate.
- What about non-stationarity?

Geographically weighted statistics

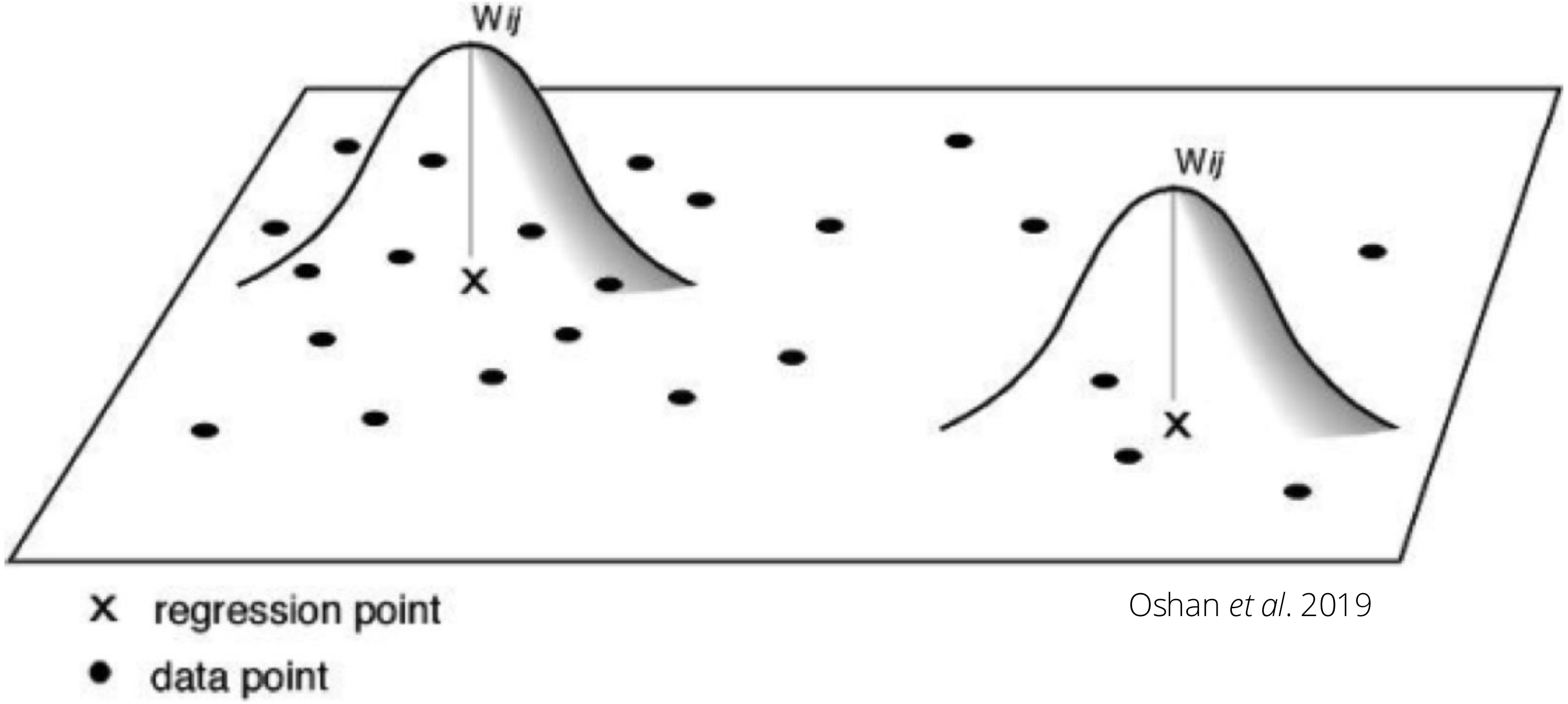
- Unlike traditional global models, which estimate a single set of parameters for the entire study area, geographically weighted statistics allow for parameter estimates that vary across different locations.
- Local means, local standard deviations, local variances.
- Typically uses some kernel function to weigh observations based on their distance from the location of interest.

Geographically weighted statistics

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

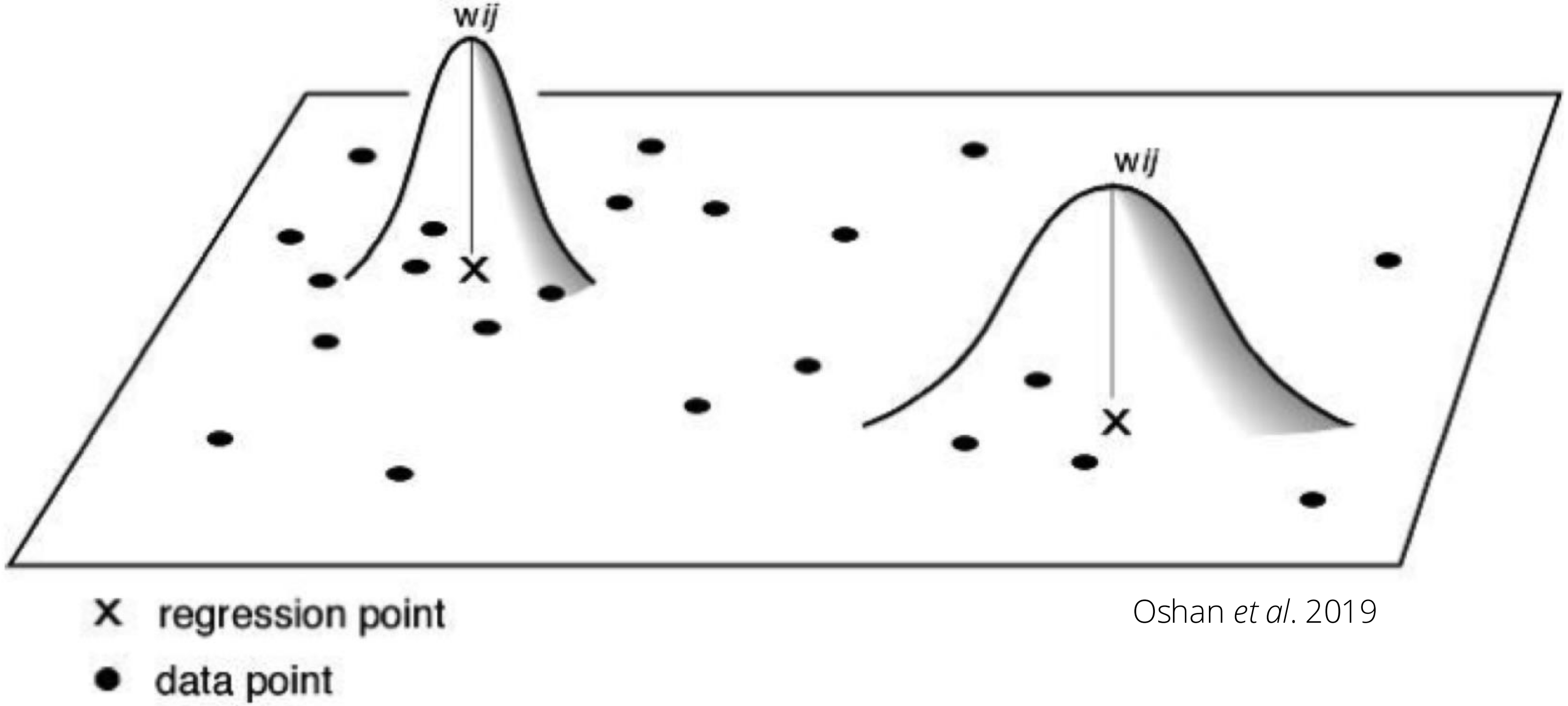
Walter Tobler 1970

Geographically weighted statistics



Oshan *et al*. 2019

Geographically weighted statistics



Oshan *et al*. 2019

Geographically weighted regression

- These ideas can be extended to correlation and regression:
  - Geographically weighted correlation (GWC)
  - Geographically weighted regression (GWR)
- The basic GWR equation is:

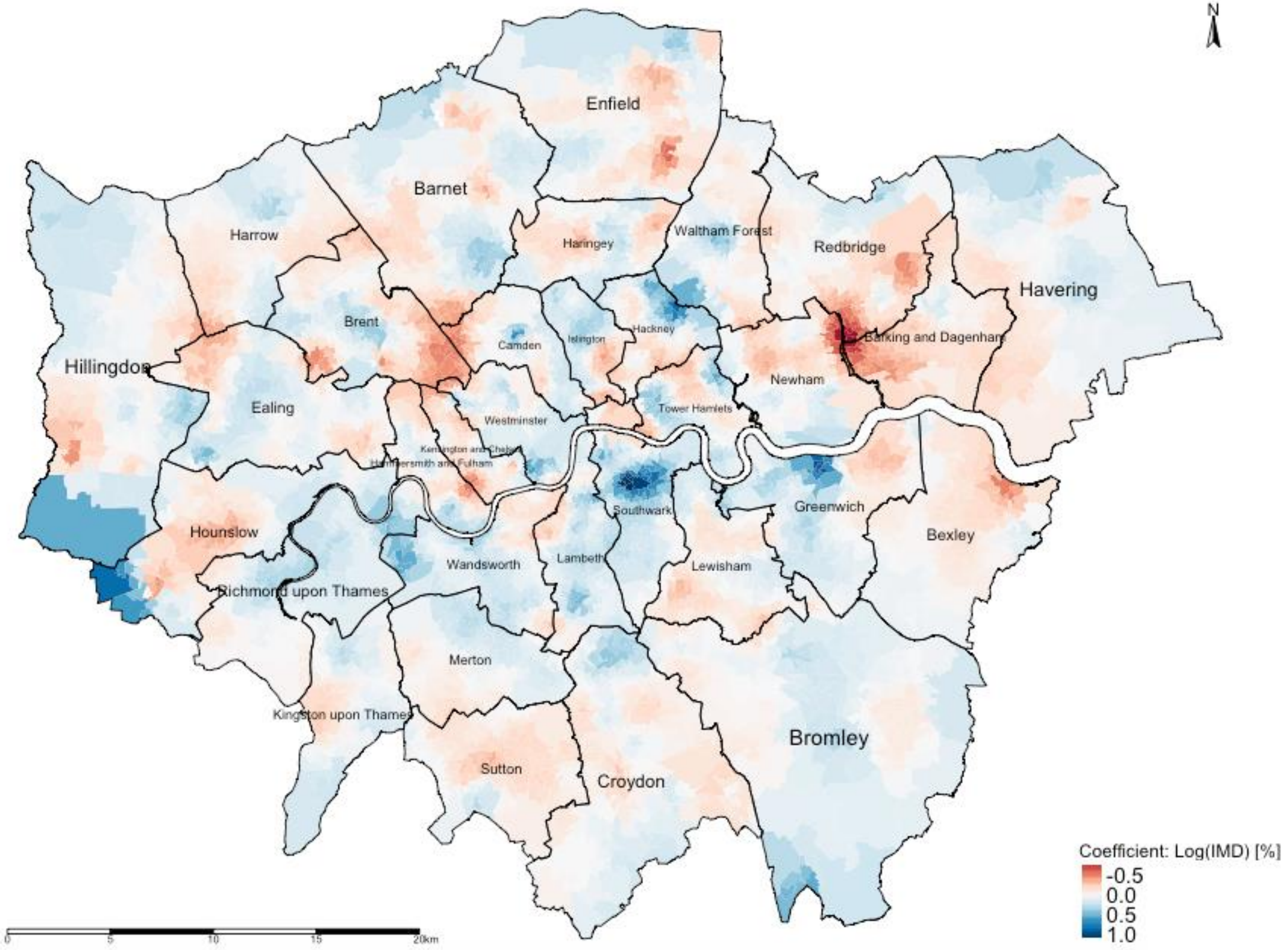
$$y_i = \beta_0(v_i, v_i) + \sum_{k=1}^{p} \beta_k(v_i, v_i) x_{ik} + \epsilon_i$$

where (v_i, v_i) are the coordinates of location i and $\beta k(v_i, v_i)$ are the location-specific coefficients.

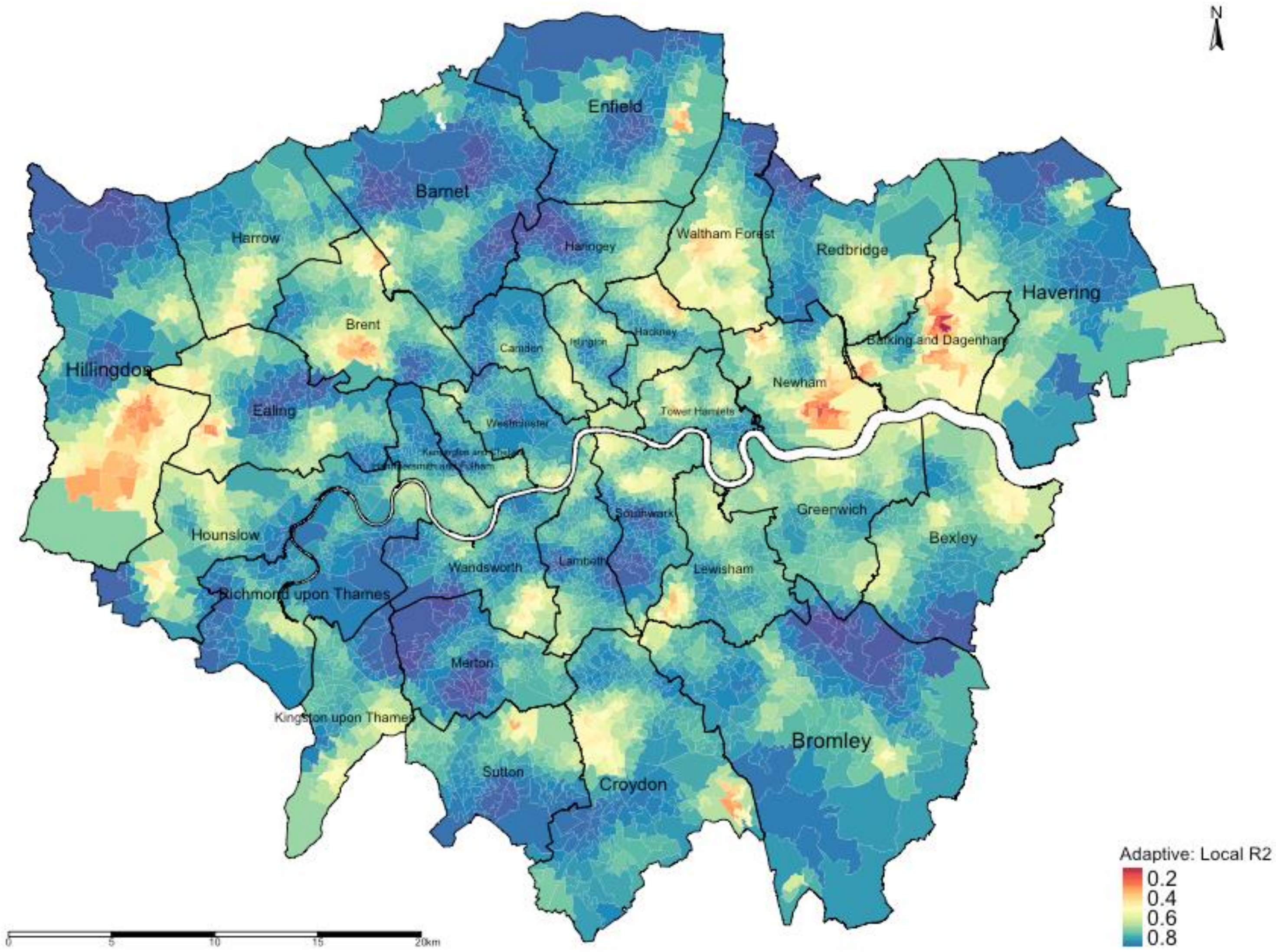
Geographically weighted regression

- Each area has its own set of regression coefficients.
- Each location has its own R^2 value.
- Each area has its own standard errors for the coefficient.
- More recently: bandwidths can vary between different variables.

Geographically weighted regression



Geographically weighted regression



Conclusion

- OLS assumes independent and homoscedastic residuals, but spatial autocorrelation violates these assumptions, leading to biased or inefficient estimates.
- Spatial Error and Spatial Lag models are techniques for handling spatial dependence by explicitly modeling spatial relationships in errors or response variables.
- Relationships between variables can vary across space, challenging the global assumption of uniform coefficients in OLS and spatial models.
- GWR captures spatial heterogeneity by estimating local coefficients, addressing both spatial autocorrelation and non-stationarity to reveal location-specific dynamics.

Assessment

- Social Atlas Coursework Assessment (60%): The first assessment will involve the completion of a spatial analysis project, based on the theory, concepts and application learnt during the module. For this coursework you are required to create a small *social atlas* on a topic or area that interests you. Deadline: April 27, 2026.
- On Moodle: guidance as well as examples from previous years.

Assessment

- You should create a minimum of 4 maps and a maximum of 6.
- You should create a minimum of 2 graphs / charts and a maximum of 4.
- You can choose a specific theme or zoom into a particular area.
- You should aim to utilise a range of different techniques taught in the Geocomputation module to explore your topic – but make sure you apply the techniques in appropriate manner and with the right data types.
- Greater London cannot be used as a case study.

Assessment

Topic: Assessment | GEOG00

moodle.ucl.ac.uk/course/view.php?id=41315&section=3

Module assessment details

Geocomputation is assessed through two separate assignments:

1. **Social Atlas Coursework Assessment (60%).** The first assessment will involve the completion of a 1,500 word spatial analysis project, based on the theory, concepts and application learnt during the module. For this coursework you are required to create a small *social atlas* on a topic or area that interests you.

2. **Exam Assessment (40%).** The second assessment will take the form of a written two-hour exam.

Marking criteria

For the coursework assignment, the marking criteria that will be used can be found in the Geography Coursework Marking Matrix: [Link]

For the exam, the marking criteria that will be used can be found in the Geography Exam Marking Matrix: [Link]

Instructions and examples

GEOG0030 Assessment: Social Atlas PDF

105.1 KB

Social Atlas Example 1 - Low 2:1 PDF

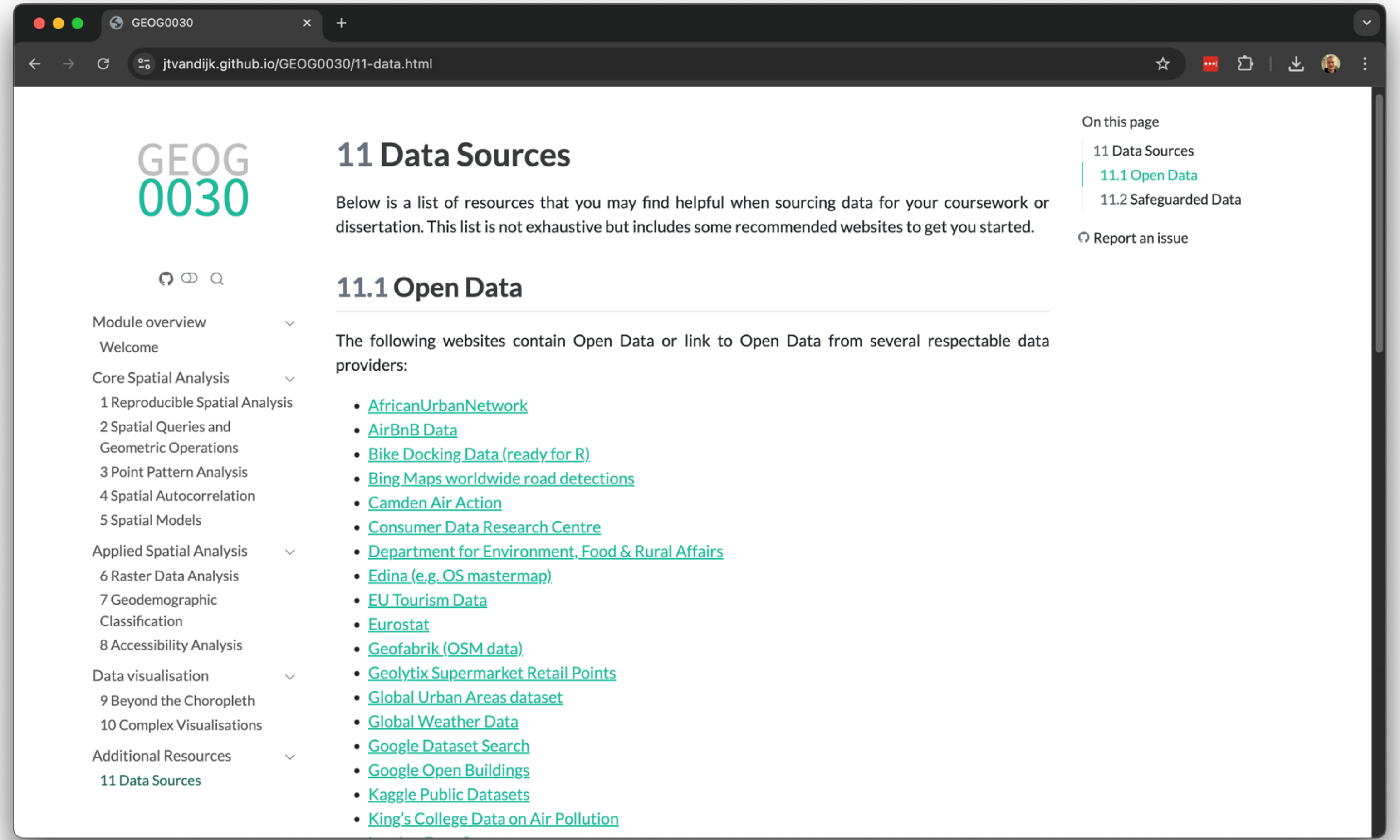
507.9 KB

Social Atlas Example 2 - High 2:1 PDF

2.0 MB

Back to top

Assessment



GEOG0030

jtvandijk.github.io/GEOG0030/11-data.html

GEOG
0030

Module overview
Welcome

Core Spatial Analysis
1 Reproducible Spatial Analysis
2 Spatial Queries and Geometric Operations
3 Point Pattern Analysis
4 Spatial Autocorrelation
5 Spatial Models

Applied Spatial Analysis
6 Raster Data Analysis
7 Geodemographic Classification
8 Accessibility Analysis

Data visualisation
9 Beyond the Choropleth
10 Complex Visualisations

Additional Resources
11 Data Sources

11 Data Sources

Below is a list of resources that you may find helpful when sourcing data for your coursework or dissertation. This list is not exhaustive but includes some recommended websites to get you started.

11.1 Open Data

The following websites contain Open Data or link to Open Data from several respectable data providers:

- AfricanUrbanNetwork
- AirBnB Data
- Bike Docking Data (ready for R)
- Bing Maps worldwide road detections
- Camden Air Action
- Consumer Data Research Centre
- Department for Environment, Food & Rural Affairs
- Edina (e.g. OS mastermap)
- EU Tourism Data
- Eurostat
- Geofabrik (OSM data)
- Geolytix Supermarket Retail Points
- Global Urban Areas dataset
- Global Weather Data
- Google Dataset Search
- Google Open Buildings
- Kaggle Public Datasets
- King's College Data on Air Pollution

On this page
11 Data Sources
11.1 Open Data
11.2 Safeguarded Data

Report an issue

Assessment

TL;DR story of max 1,500 words tied together by 4-6 related maps and 2-4 graphs.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk

Have a good reading week!