Geocomputation

Complex Visualisations

UCL

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

- Some pointers for the exam.
- Package and code management.
- Tidy data.
- Grammar of graphics.

Before we start

- Go to www.menti.com
- Use code: 6825 6964

Exam

Some pointers for the exam

Some pointers for the exam

- Two-hour timed exam scheduled for May 29th.
- Two essays (1,000-word limit per question), aim to spend one hour on each.
- Six questions to choose from a range of topics covered in the lectures.
- No questions on coding or programming: focus on theory, underlying principles, methods and applications.
- Q&A Session on April 29th.

How to prepare

- Refer to previous exams to get an idea of the types of questions.
- Strategically read the relevant chapters in the textbook (Longley *et al.* 2018) and more theoretical papers.
- Strategically read the more applied papers, make notes of relevant examples or applications of methods.

Example question #1

- Critically reflect on the strengths and weaknesses of Kernel Density Estimation in identifying spatial clusters. In your answer make use of Cheshire and Longley (2011).

Example question #2

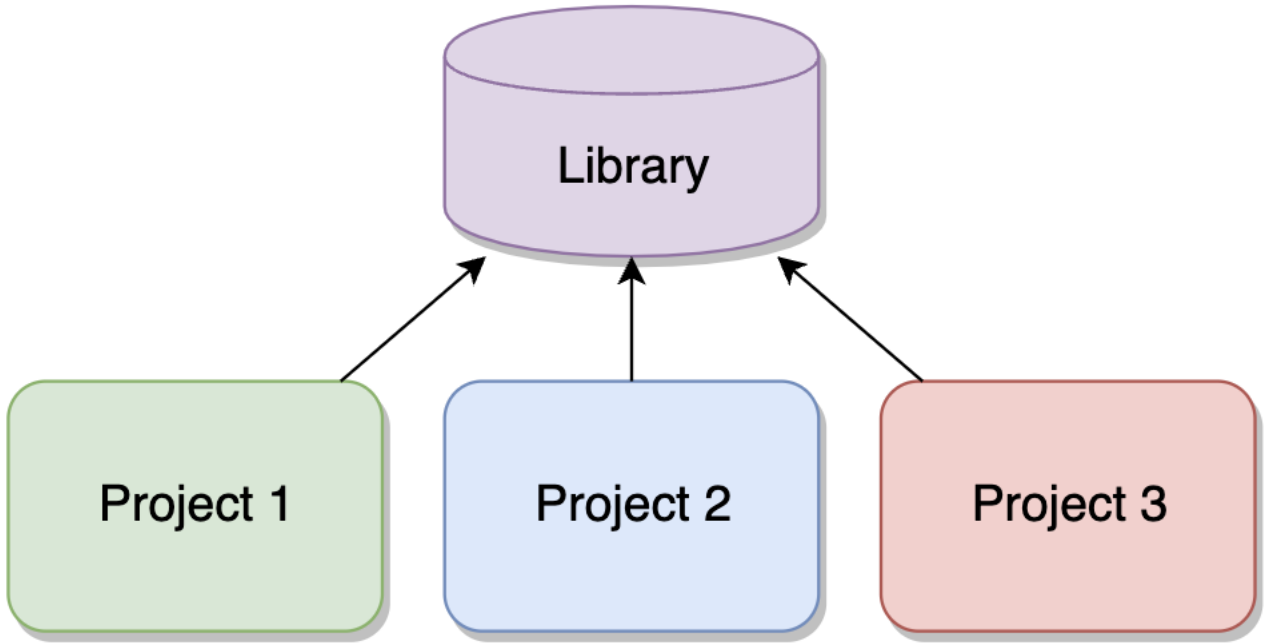
- The NHS wants to improve their ambulance response times and they have asked you to assess their current response times in London. Outline the data that you would request, the analytical steps you might take to complete this task and the final maps you could produce in response.

Package and data management

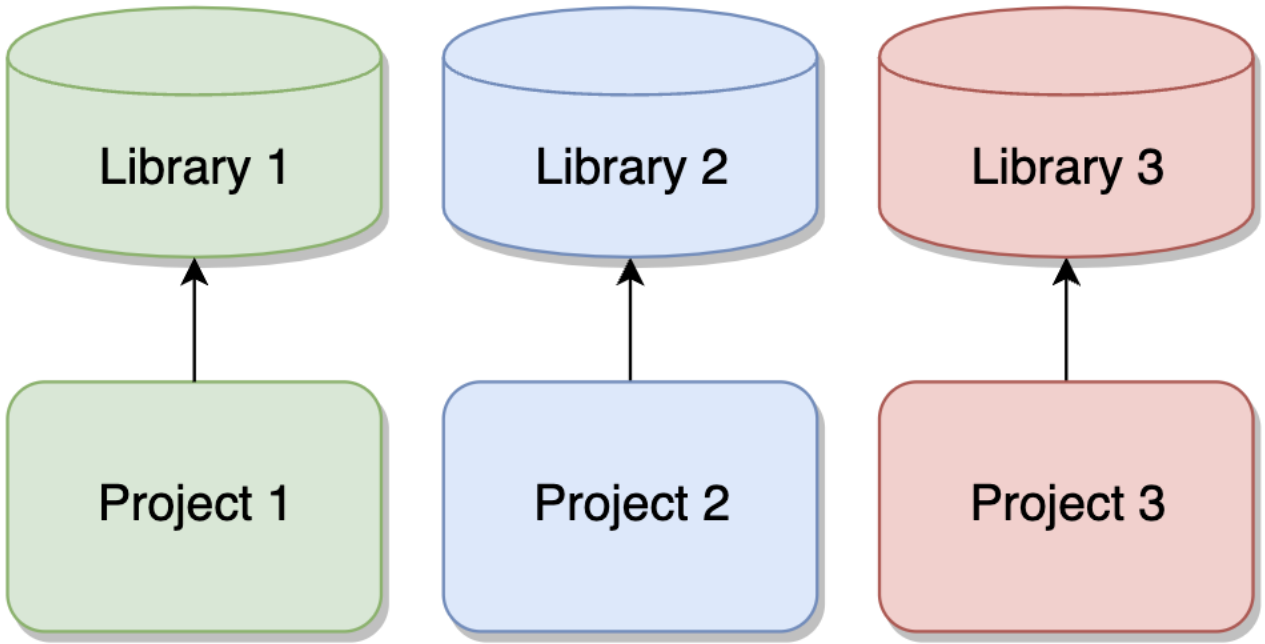
Package management

- Package management is the process of handling the many and varied dependencies and artifacts for your servers, applications, and developers.
- Toolkit used to manage project-local libraries.
- Combination with version management through Git.

Package management



Package management



Code management

- Version control: The practice of tracking and managing changes to software code.
- Example: Git (GitHub, GitLab).
- “Track changes" on steroids.
- Beyond the scope of this module but you can still set up your own small version control system (”_v0.1”, “v_0.2”, “v_1.0”).

Tidy data

Managing data

- Wickham 2014. 80 percent of your time goes to data cleaning and preparation (‘data wrangling’).
- Tidy data refers to the structure and organisation of your data set.
- The idea boils down to three principles.
- Brought together in the tidyverse.

Tidy data

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

Each variable must have its own column

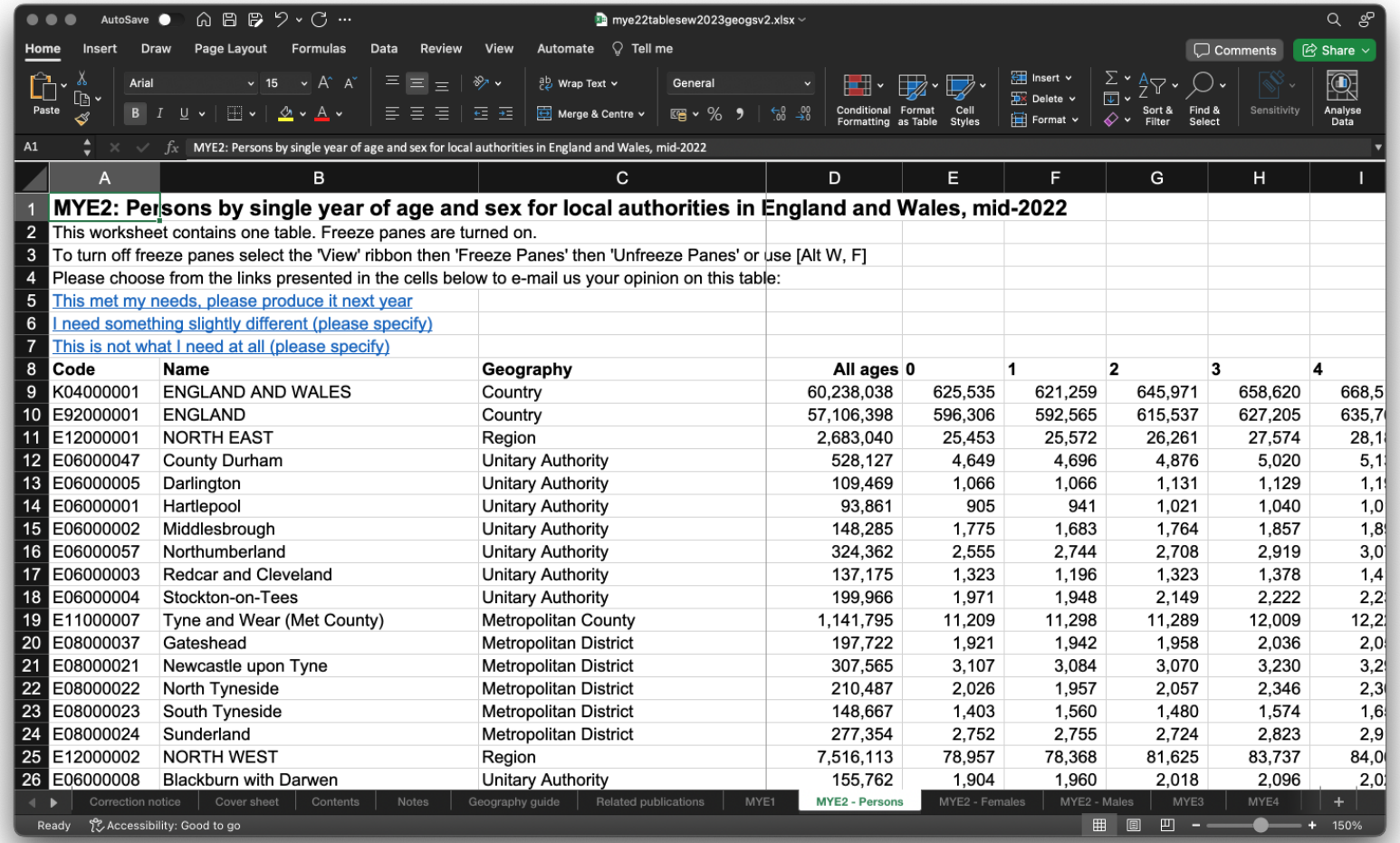
Tidy data

Each observation must have its own row

Tidy data

Each value must have its own cell

Tidy data



AutoSave mye22tablesew2023geogsv2.xlsx

Home Insert Draw Page Layout Formulas Data Review View Automate Tell me Comments Share

A1 MYE2: Persons by single year of age and sex for local authorities in England and Wales, mid-2022

	A	B	C	D	E	F	G	H	I
1	**MYE2: Persons by single year of age and sex for local authorities in England and Wales, mid-2022**								
2	This worksheet contains one table. Freeze panes are turned on.								
3	To turn off freeze panes select the 'View' ribbon then 'Freeze Panes' then 'Unfreeze Panes' or use [Alt W, F]								
4	Please choose from the links presented in the cells below to e-mail us your opinion on this table:								
5	This met my needs, please produce it next year								
6	I need something slightly different (please specify)								
7	This is not what I need at all (please specify)								
8	**Code**	**Name**	**Geography**	**All ages**	**0**	**1**	**2**	**3**	**4**
9	K04000001	ENGLAND AND WALES	Country	60,238,038	625,535	621,259	645,971	658,620	668,5
10	E92000001	ENGLAND	Country	57,106,398	596,306	592,565	615,537	627,205	635,7
11	E12000001	NORTH EAST	Region	2,683,040	25,453	25,572	26,261	27,574	28,1
12	E06000047	County Durham	Unitary Authority	528,127	4,649	4,696	4,876	5,020	5,1
13	E06000005	Darlington	Unitary Authority	109,469	1,066	1,066	1,131	1,129	1,1
14	E06000001	Hartlepool	Unitary Authority	93,861	905	941	1,021	1,040	1,0
15	E06000002	Middlesbrough	Unitary Authority	148,285	1,775	1,683	1,764	1,857	1,8
16	E06000057	Northumberland	Unitary Authority	324,362	2,555	2,744	2,708	2,919	3,0
17	E06000003	Redcar and Cleveland	Unitary Authority	137,175	1,323	1,196	1,323	1,378	1,4
18	E06000004	Stockton-on-Tees	Unitary Authority	199,966	1,971	1,948	2,149	2,222	2,2
19	E11000007	Tyne and Wear (Met County)	Metropolitan County	1,141,795	11,209	11,298	11,289	12,009	12,2
20	E08000037	Gateshead	Metropolitan District	197,722	1,921	1,942	1,958	2,036	2,0
21	E08000021	Newcastle upon Tyne	Metropolitan District	307,565	3,107	3,084	3,070	3,230	3,2
22	E08000022	North Tyneside	Metropolitan District	210,487	2,026	1,957	2,057	2,346	2,3
23	E08000023	South Tyneside	Metropolitan District	148,667	1,403	1,560	1,480	1,574	1,6
24	E08000024	Sunderland	Metropolitan District	277,354	2,752	2,755	2,724	2,823	2,9
25	E12000002	NORTH WEST	Region	7,516,113	78,957	78,368	81,625	83,737	84,0
26	E06000008	Blackburn with Darwen	Unitary Authority	155,762	1,904	1,960	2,018	2,096	2,0

Correction notice | Cover sheet | Contents | Notes | Geography guide | Related publications | MYE1 | MYE2 - Persons | MYE2 - Females | MYE2 - Males | MYE3 | MYE4

Ready Accessibility: Good to go 150%

Common errors

- Column headers are values rather than variable names.
- Multiple variables are stored in one column.
- Variables are stored in both rows and columns.
- Multiple observational units are stored in the same column.
- A single observation is stored in multiple tables.

Tidy ?

country	year	type	count
Afghanistan	2019	cases	745
Afghanistan	2019	population	19 987 071
Afghanistan	2020	cases	2 666
Afghanistan	2020	population	20 595 360
Brazil	2019	cases	3,7737
Brazil	2019	population	172 006 362
Brazil	2020	cases	80 488
Brazil	2020	population	174 504 898
China	2019	cases	212 258
China	2019	population	1 272 915 272
China	2020	cases	213 766
China	2020	population	1 280 428 583

Tidy ?

country	year	rate
Afghanistan	2019	745 / 19,987,071
Afghanistan	2020	2,666 / 20,595,360
Brazil	2019	3,7737 / 172,006,362
Brazil	2020	80,488 / 174,504,898
China	2019	212,258 / 1,272,915,272
China	2020	213,766 / 1,280,428,583

Tidy ?

Cases

country	2019	2020
Afghanistan	745	2 666
Brazil	3,7737	80 488
China	212 258	213 766

Population

country	2019	2020
Afghanistan	19 987 071	20 595 360
Brazil	172 006 362	174 504 898
China	1 272 915 272	1 280 428 583

Tidy ?

country	year	cases	population
Afghanistan	2019	745	19 987 071
Afghanistan	2020	2 666	20 595 360
Brazil	2019	3,7737	172 006 362
Brazil	2020	80 488	174 504 898
China	2019	212 258	1 272 915 272
China	2020	213 766	1 280 428 583

Grammar of graphics

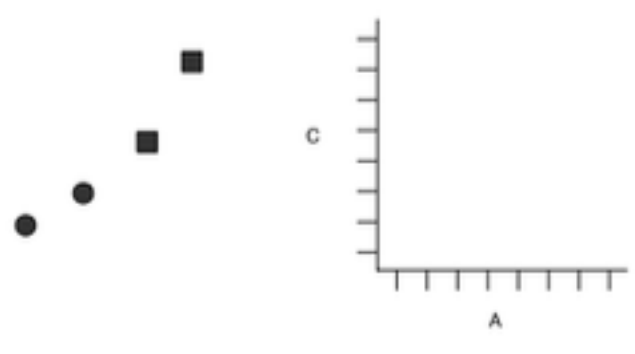
Grammar of graphics

- Graphics are constructed by layering multiple elements of data.
- Values in a dataset serve as aesthetics: attributes that can be visually represented in a graphic.
- Data, scales, coordinate systems, and plot annotations are layered on top of these aesthetics to create the final graphic.

Grammar of graphics

Dataframe

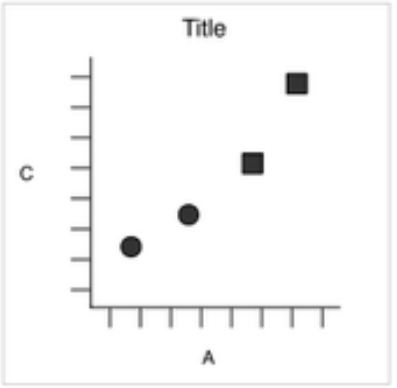
x	*y*	Shape
2	4	a
1	1	a
4	15	b
9	80	b



Dataframe values

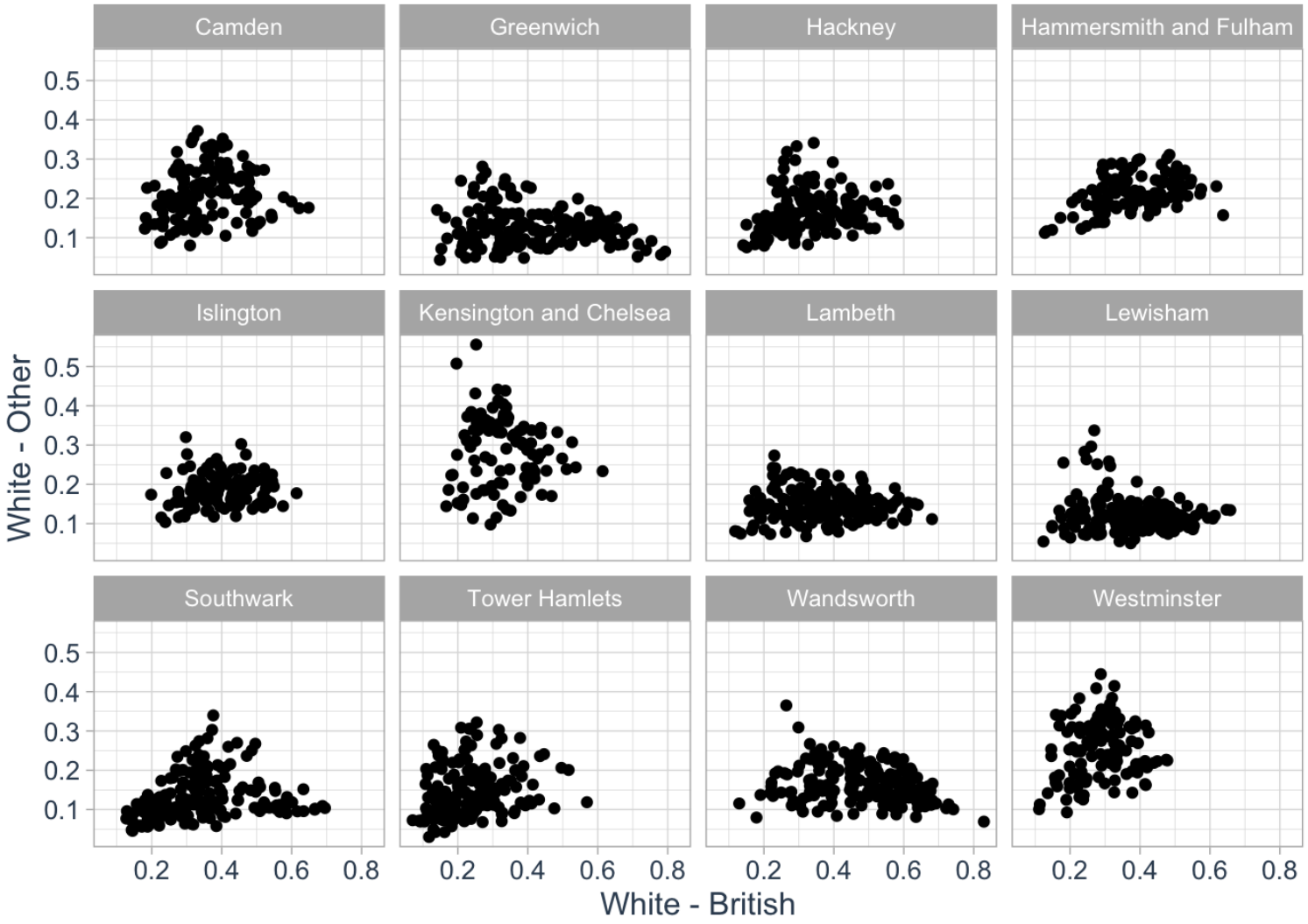
Dataframe scale

Dataframe annotations

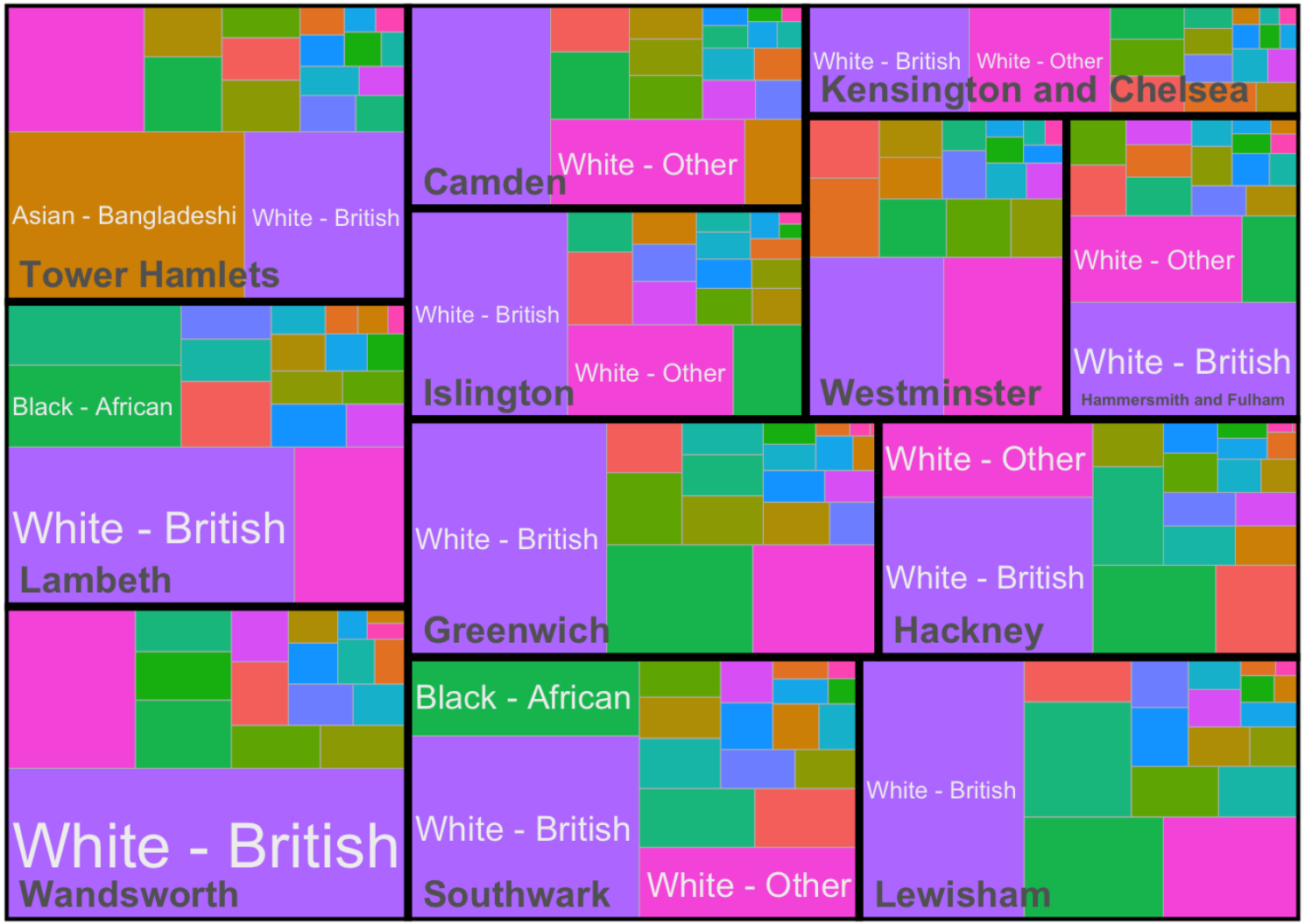


Final Graphic

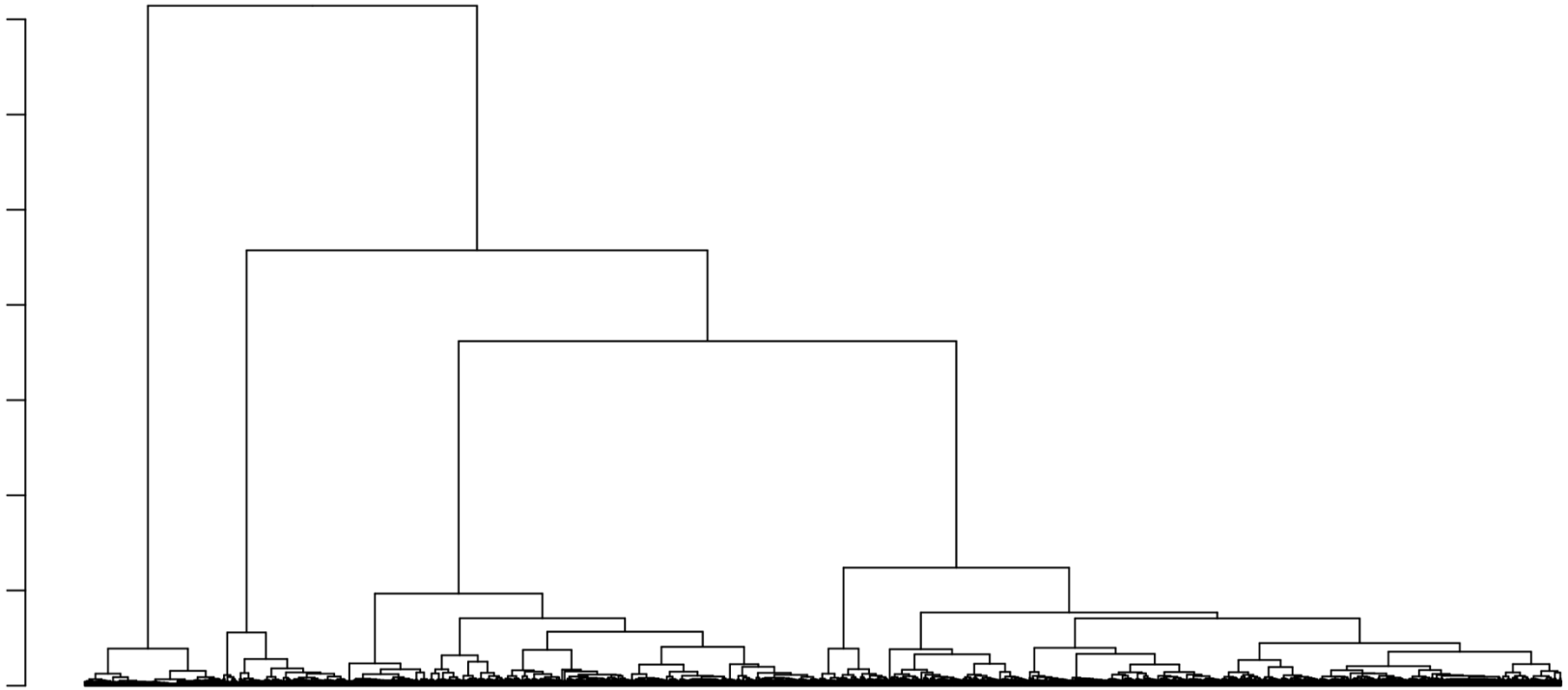
Grammar of graphics



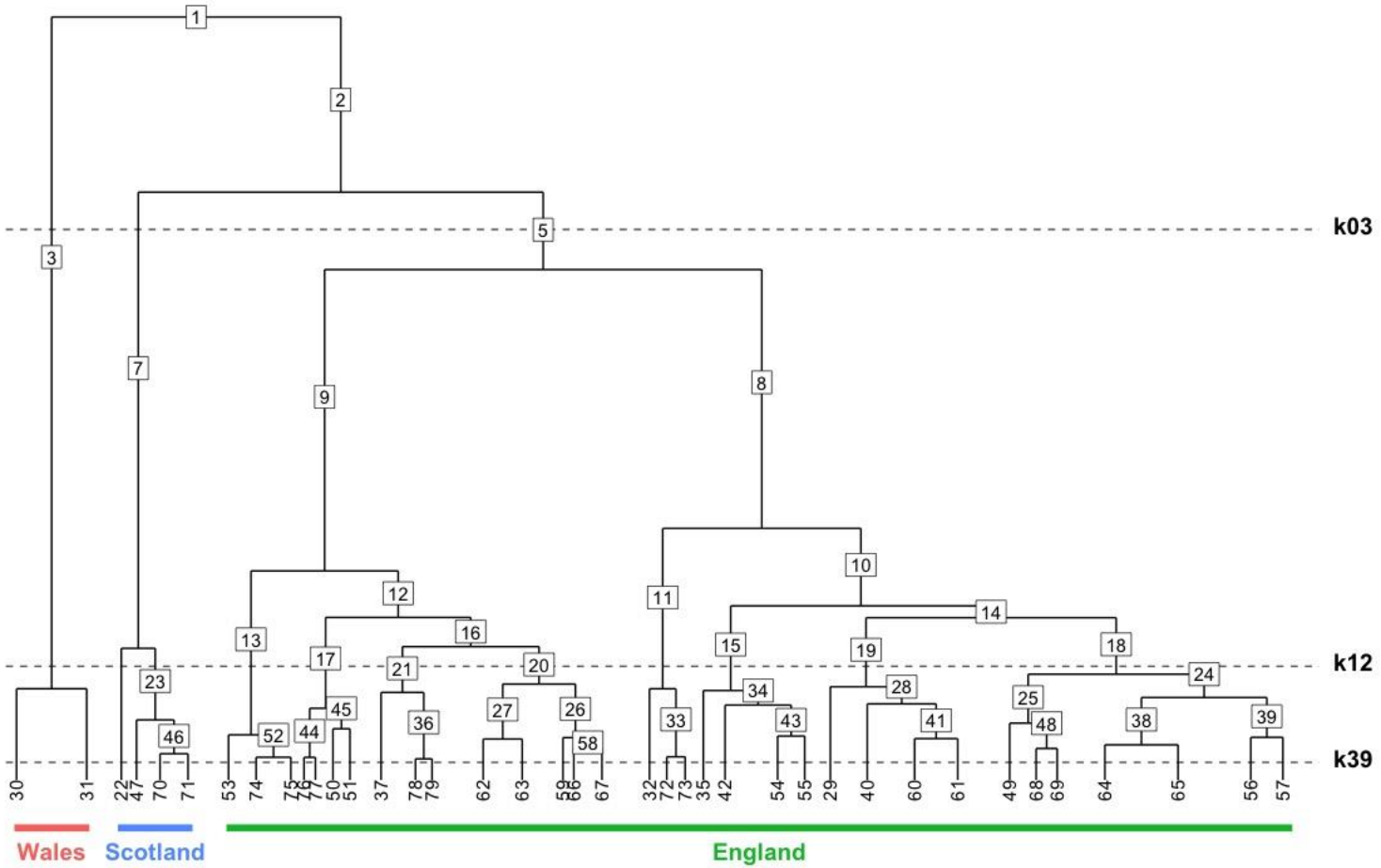
Grammar of graphics



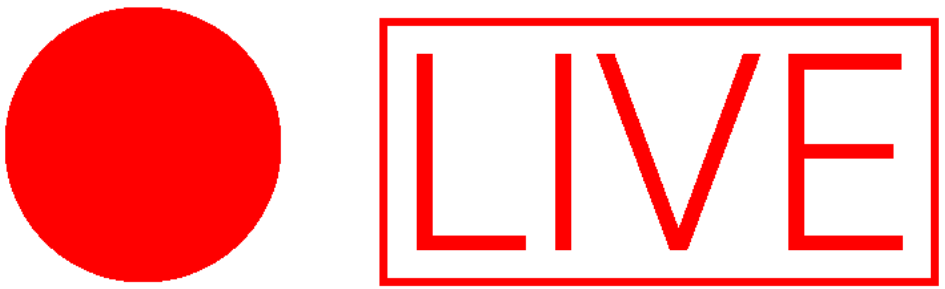
Grammar of graphics



Grammar of graphics



RStudio



Everything we covered

Everything we covered

1) Geocomputation as a GIS 2.0: working with geographic data in a computational way, focusing on code, reproducibility and modularity.
2) Spatial queries and geometric operations: the core of spatial analysis.
3) Working with point event data: special attention to clustering and visualisation of these using DBSCAN and Kernel Density Estimation.
4) The First Law of Geography in action: measuring spatial autocorrelation.
5) Dealing with spatial autocorrelation and non-stationarity: spatial models

Everything we covered

6) Dealing with raster data: spatial data interpolation.
7) Geodemographics: analysis of people by where they live.
8) Measuring accessibility: working with a digital network.
9) Positioning the map: Coordinate Reference Systems.
10) Complex Visualisations: using ggplot2 to build a graphic layer by layer.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk

Thank you!