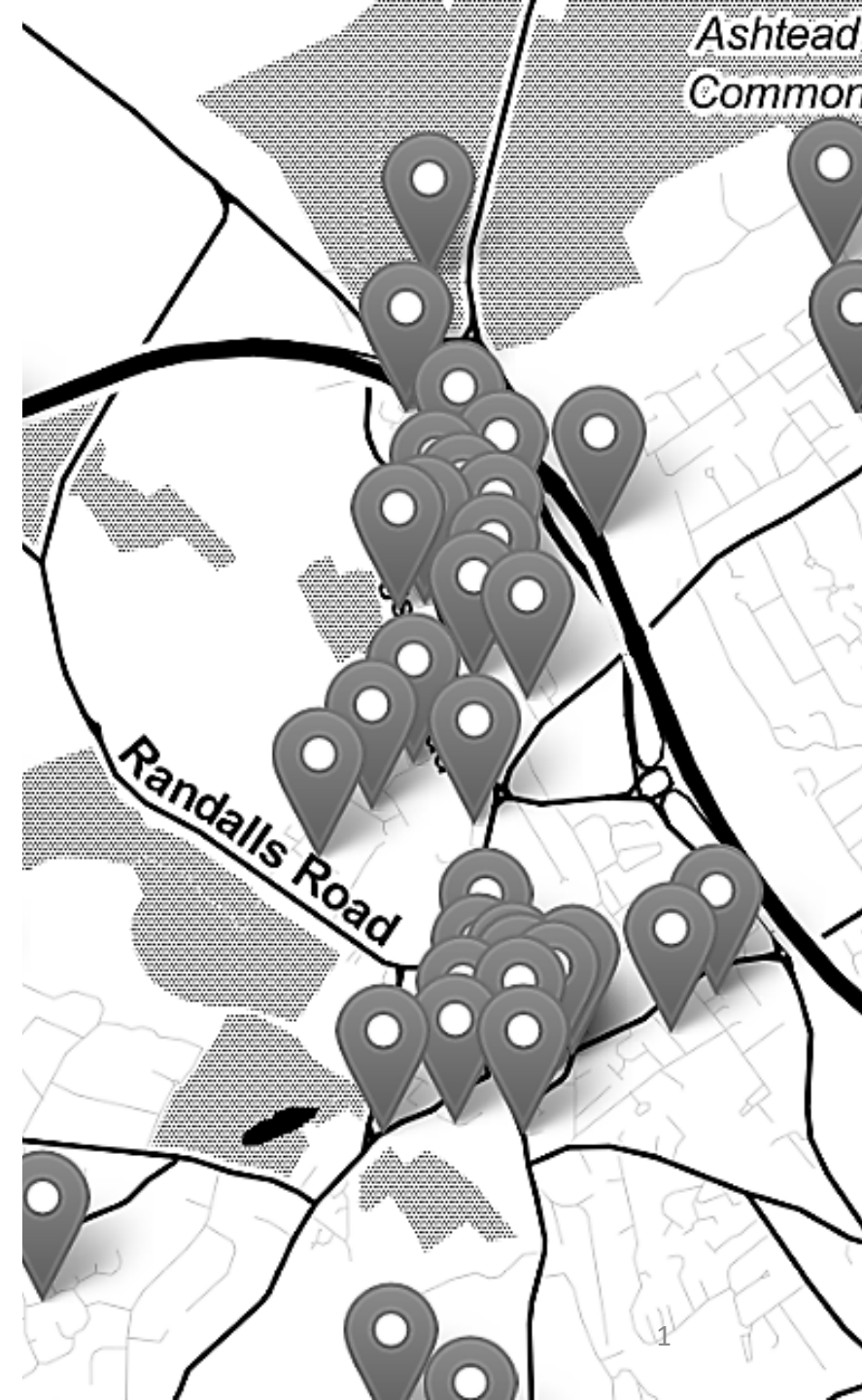# Mapping Crime Data in R:

An Introduction to GIS and Spatial Data (13/02/2024 and 16/02/2024)

Nadia Kennar, Research Associate at the UK Data Service (Cathie Marsh Institute, UoM)
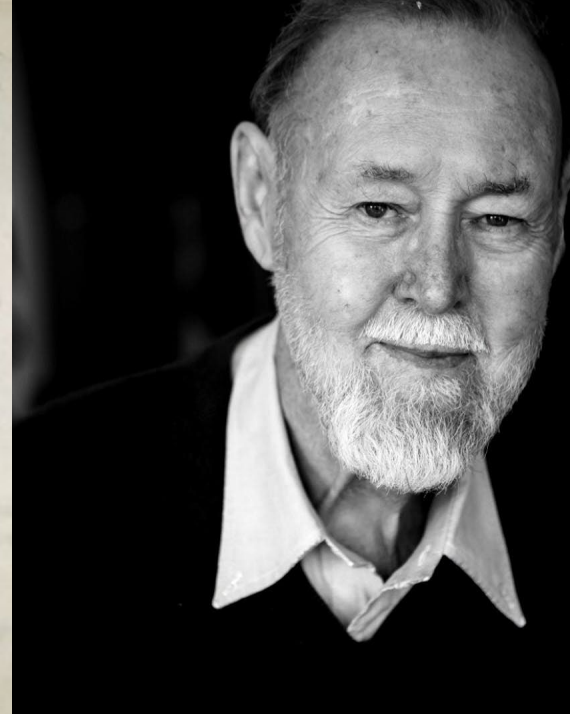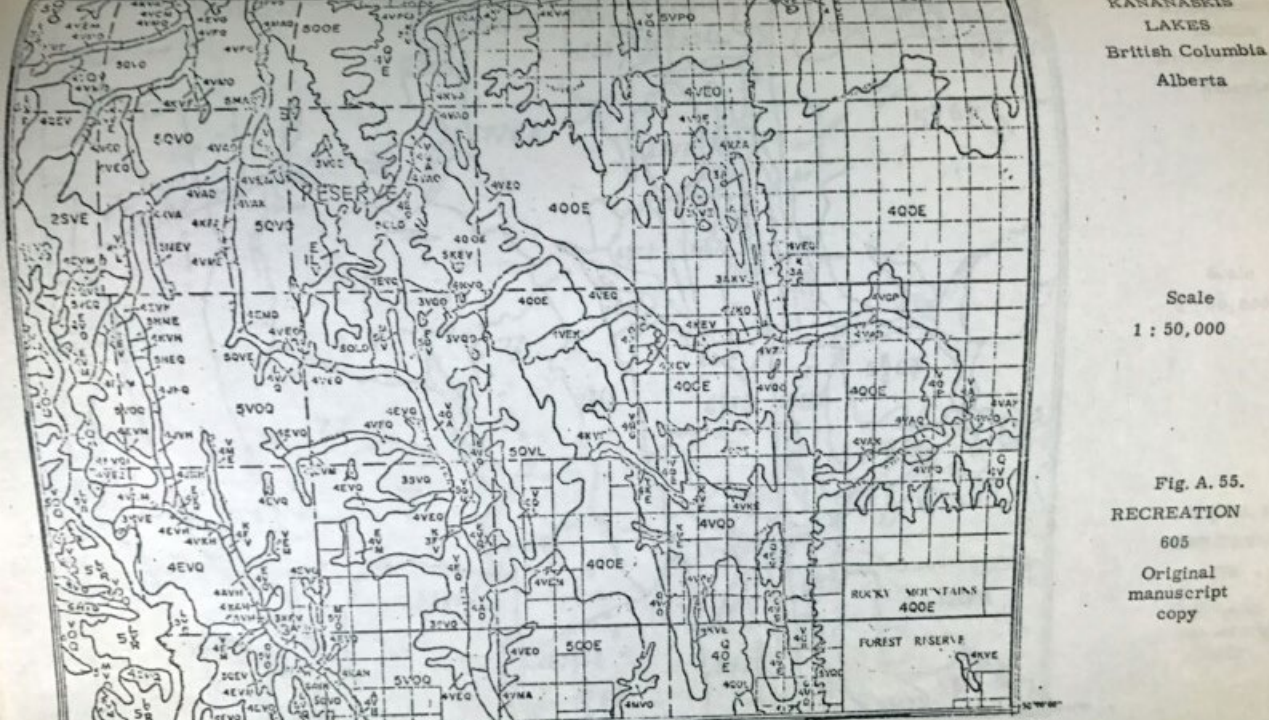
Workshop Content – 13th February 2024

- What is GIS
- Spatial Data vs Non-Spatial Data
- Different types of maps (reference vs thematic)
- Projection Methods and Coordinate Reference Systems
- Spatial Relations (and/or spatial statistics)
    a) Spatial autocorrelation
    b) Spatial interpolation
    c) Hot spot analysis
- Data sources for crime and GIS
- Challenges of Mapping Crime Data

Live Code Demo – 16th of February 2024

- Topic 1 – Descriptive analysis of crime data
- Topic 2 – Shapefiles
- Topic 3 – Combining census data (crime rate)
- Topic 4 – Spatial Relations/statistics
- Extra Topic – Interactive maps via Leaflet Package

KANANASKIS
LAKES
British Columbia
Alberta

Scale
1 : 50,000

Fig. A. 55.
RECREATION
605
Original
manuscript
copy

Dr. Roger F. Tomlinson, OC
1933 - 2014

What is GIS?

- Graphical Information Systems:
- *"computer system for capturing, storing, checking, and displaying data related to positions on Earth's surface"*

City	Score
Cupertino	17.2
Doha	13.5
Dubai,	22.4
Geneva	11.5
Hong Kong	78.3
Luxembourg City	45.3
Mountain View	74.2
San Francisco	12.9
Singapore	56.8
Zurich	45.7

A) interested in identifying the ten cities with the highest average income scores

B) interested in identifying whether these ten cities are geographically clustered

City	Score	Spatial Boundaries
Cupertino	17.2	SW: (28.00, 45.09) NE: (28.09, 45.16)
Doha	13.5	SW: (-60.38, 147.45) NE: (-60.32, 147.52)
Dubai,	22.4	SW: (48.62, 146.83) NE: (48.66, 146.93)
Geneva	11.5	SW: (-81.64, -174.64) NE: (-81.56, -174.58)
Hong Kong	78.3	SW: (69.11, 60.87) NE: (69.18, 60.92)
Luxembourg City	45.3	SW: (37.10, -123.11 NE: (37.19, -123.06)
Mountain View	74.2	SW: (12.87, 132.36) NE: (12.91, 132.46)
San Francisco	12.9	SW: (-47.69, -169.90) NE: (-47.68, -169.87)
Singapore	56.8	SW: (-72.25, -178.46) NE: (-72.23, -178.40)
Zurich	45.7	SW: (-84.93, -163.24) NE: (-84.88, -163.21)

Continued.

- All data in GIS are 'georeferenced' meaning it has both;

    - Attribute (what it is)
    - Location (where it is, a known location)

    -

Georeferenced = is a digital map or image that has been tied to a known earth coordinate system.

What software's are available?

- GeoDa
- **ArcGIS**
- FME
- **QGIS**
- R
- …

What software have/do you use for GIS/Spatial Analysis?

How is GIS used?

1. Identify problems
2. Monitor change
3. Manage and respond to events
4. Perform forecasting
5. Set priorities
6. Understand trends

- Reference: https://www.esri.com/en-us/what-is-gis/overview#liSwitcher

Here are just some of the questions that GIS allows us to explore with crime data.

- Where are the most vulnerable communities located?
- Why do crimes occur in one area and not the other
- How do offenders travel to the crime location?
- Where are there more or less stop and search than we would expect in relation to the distribution of crime?

Reference vs Thematic Maps

- Reference Maps: used to communicate location on more static data points
  - To 'pin point' data on a map
  - Descriptive

- Thematic Maps: used to highlight a spatial relationship
  - To 'study a theme' within a map
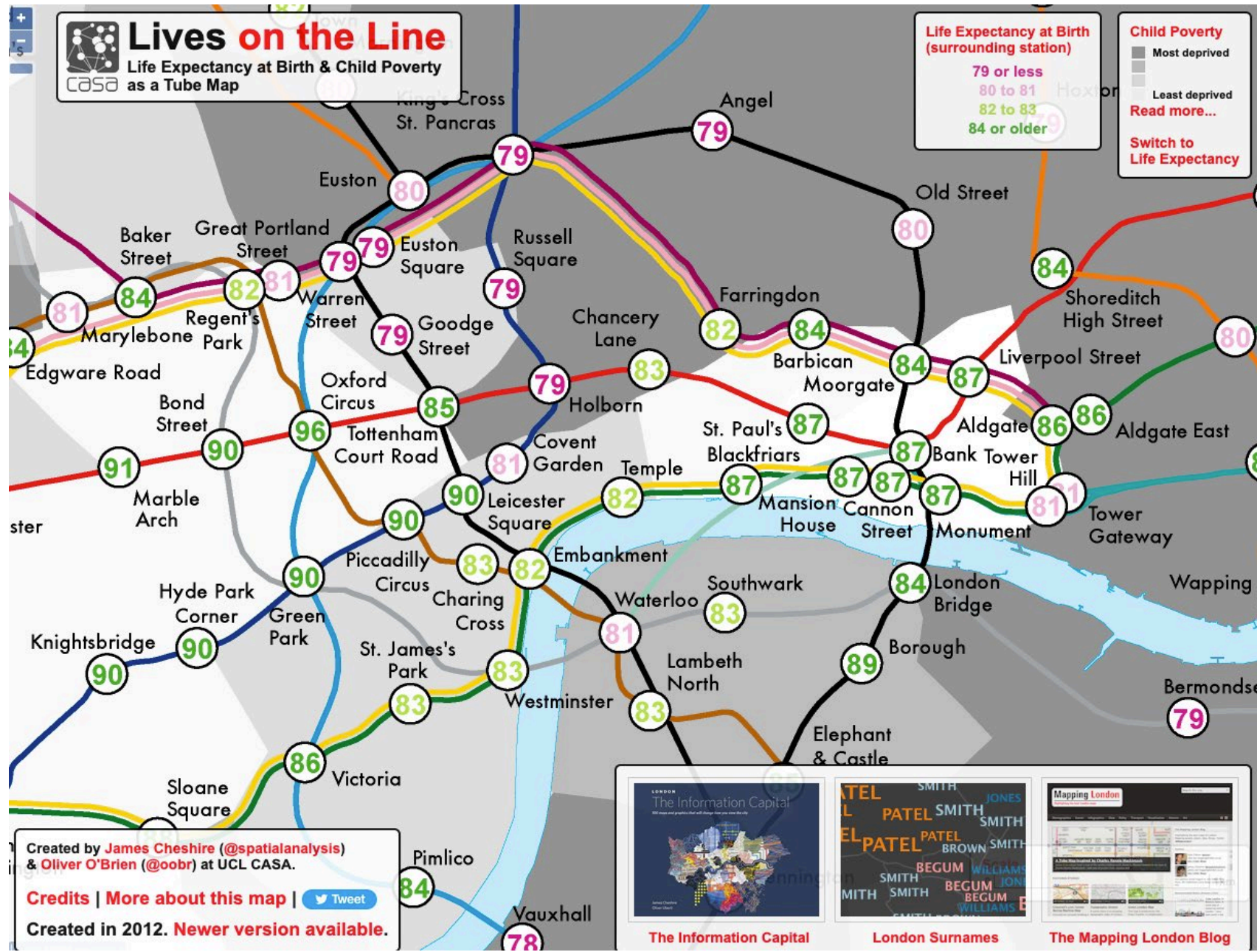 - Explanatory

# Reference vs Thematic Maps Continued…

Vs
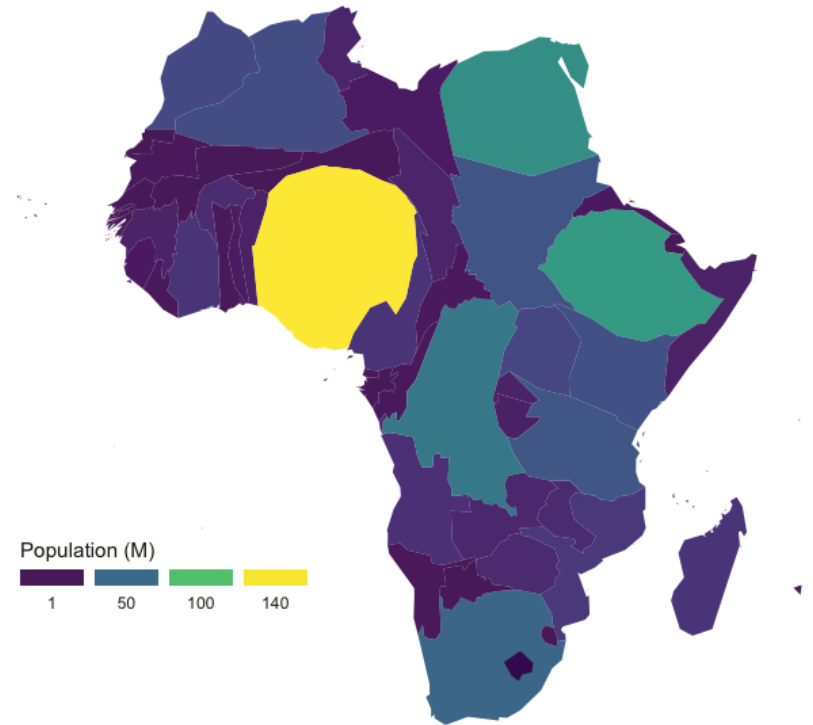
Tube Map Example

Tube Map Example

- How do we class Tube Maps;

- Reference maps because…
  - They show the location of different tube stations, and the location of each tube line

- Thematic maps because…
  - They can be used to predict life expectancy, poverty and median house prices

  - Have a go exploring this!!
  - https://tubecreature.com/#/livesontheline/current/same/U/*/FFTFTF/11.46 9326848406268/-0.1622/51.5142/
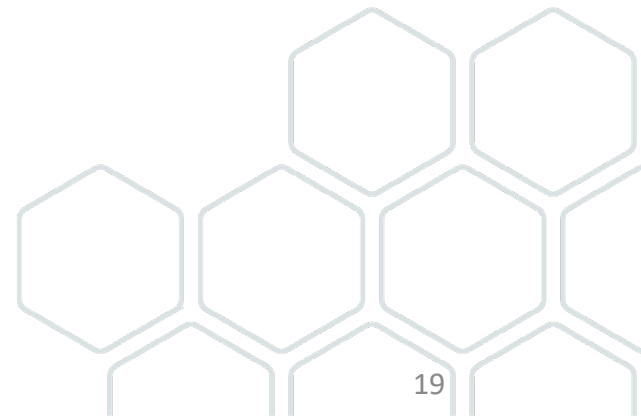
Scenario 1: The visualisation of road networks to improve road safety measures are a type of
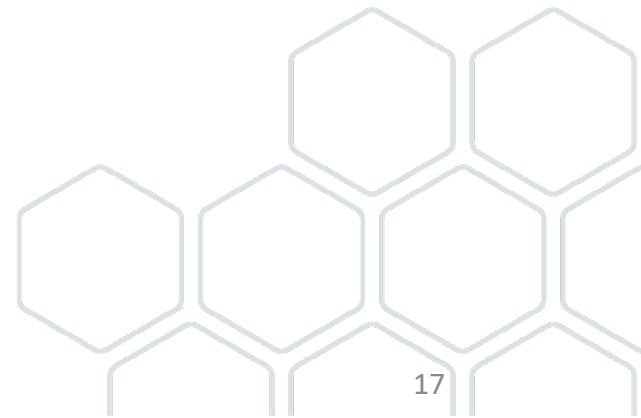
Scenario 2: The visualisation of the earths surface, showing its elevation, is a type of
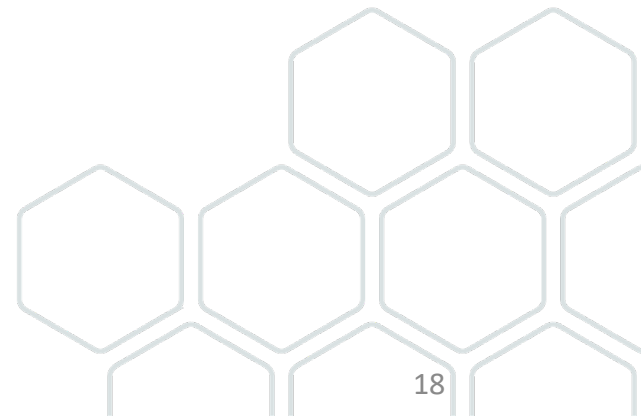
Scenario 3: Navigation tools such as Google maps or City-Mapper can be classed as

To sum up

- All though maps falls broadly into two categories, there are ways in which these types of map overlap or share similarities
- Almost every thematic maps is also a reference map, but not every reference map is a thematic map
- The decision is up to you, it is not entirely necessary to define these in your work but it is important to now what type of map you want to make as these can be affected by the data you have

Can you give examples of any other types of map that share qualities of both thematic and reference maps?
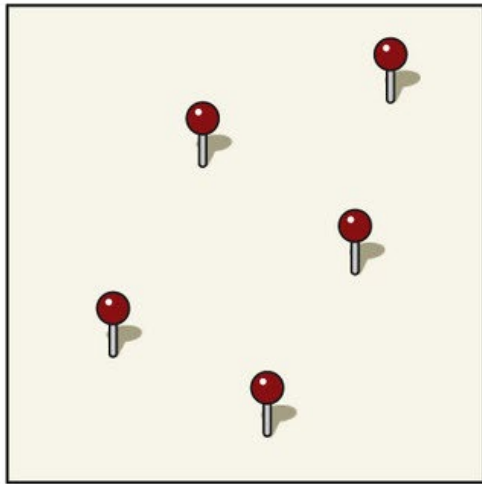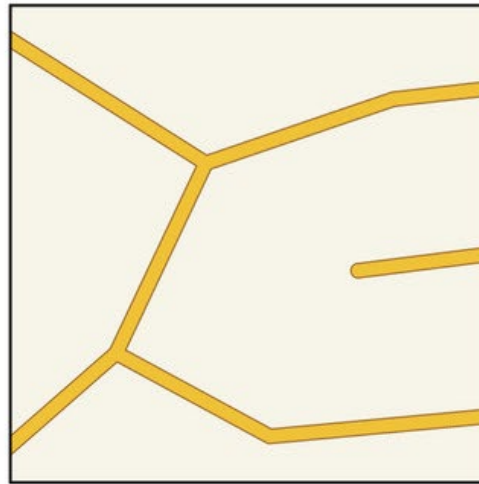
What is Spatial Data?

- Spatial data (or geospatial data) is a data frame that contains information about a specific location, which can be analysed to better understand that location.

- GIS enables this spatial data to be processed and analysed.

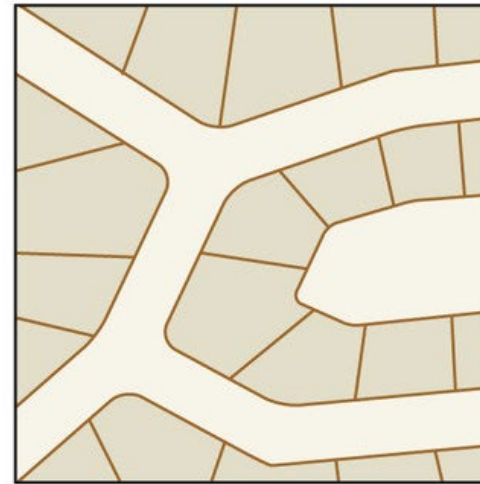- Two types: Vector and Raster

Vector Data

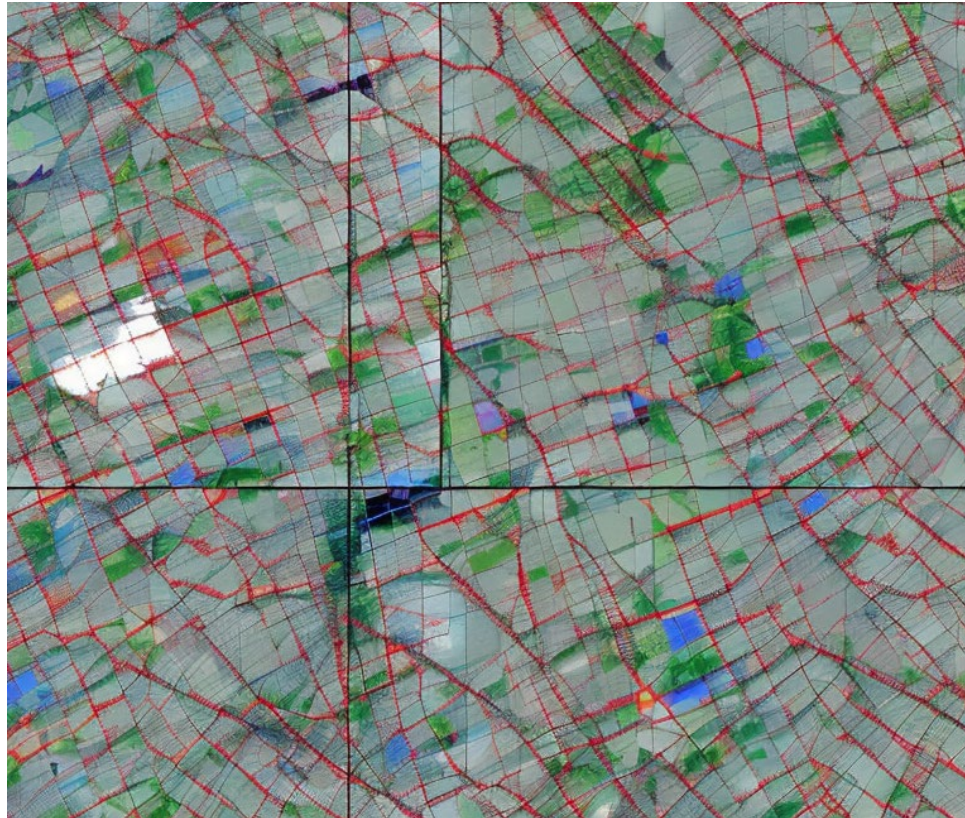- Vector Data (points, lines and polygons)

Points   Lines   Polygons

Raster Data

- Imagery or satellite data that are formed from a grid of pixels.

GIS Data Management

- GeoTiff files are commonly used for raster data
- Shapefile files are commonly used for vector data. Here's a file extension…

File extension	Content
.dbf	Attribute information
.shp	Feature geometry
.shx	Feature geometry index
.aih	Attribute index
.ain	Attribute index
.prj	Coordinate system information
.sbn	Spatial index file
.sbx	Spatial index file

Managing GIS files in R (vector data)

Key Packages:

- sf (Simple Features): A modern approach to managing spatial data, supporting simple features for R.

- rgdal: Provides bindings to the GDAL (Geospatial Data Abstraction Library) for reading and writing a wide range of formats.

Example Code:

```
library(sf)
shapefile <- st_read("path/to/your/file.shp")

shapefile_transformed <-
st_transform(my_shapefile, crs = 4326)

str(shapefile_transformed)
summary(shapefile_transformed)

Plot(shapefile_transformed)
ggplot(data = shapefile_transformed) +
geom_sf(aes(fill = attribute))
```

Modifiable Arial Unit Problem (MAUP)

"MAUP refers to the cartographic representation of data whose attributes are significantly influenced by the spatial scale used"
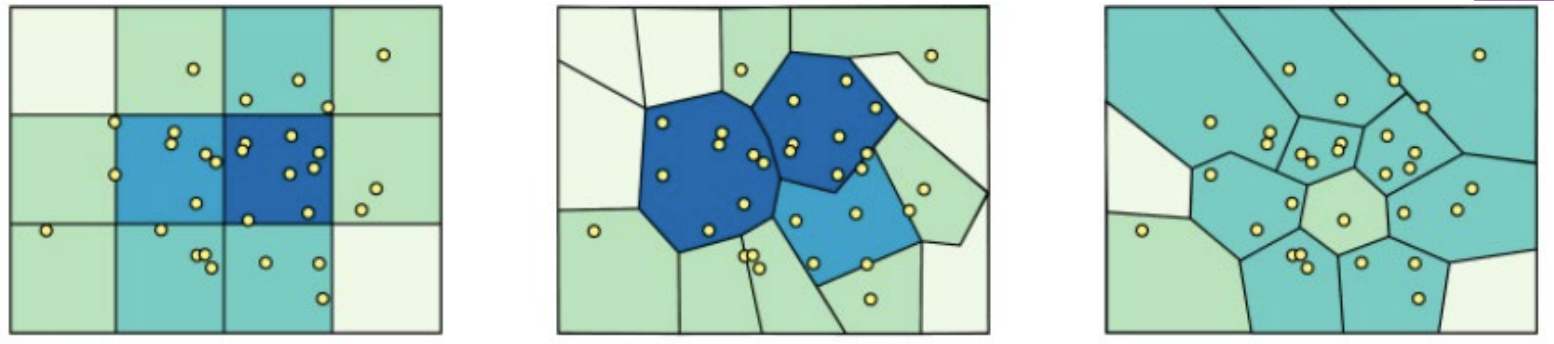
Two key Aspects:

1. Scale Effect: Changing the size of the spatial units (e.g., from neighbourhoods to districts) can alter statistical results, such as means or totals.
2. Zoning Effect: Altering the shape or configuration of spatial units, even if the scale remains constant, can also impact results.

- Scale Effect:
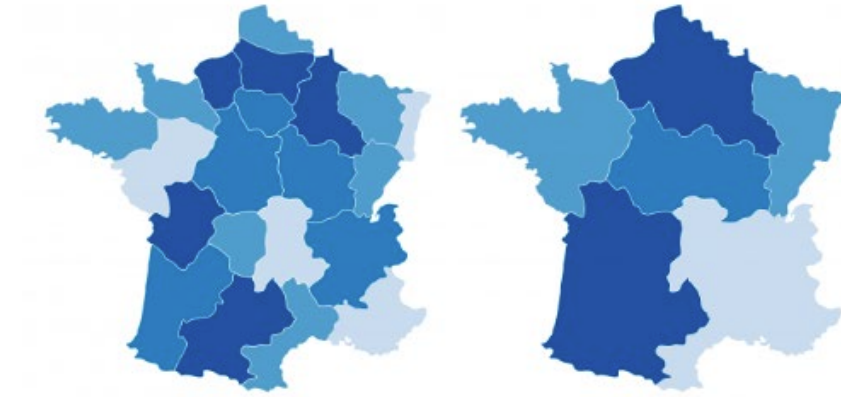
- Zone Effect:

- **Question**: True or False: The scale effect in MAUP refers to changes in statistical results caused by the size of the geographical units used in the analysis.

- **Question**: Which of the following best describes the aggregation effect in MAUP?A) It refers to the impact of data collection methods on analysis outcomes.

- B) It describes how the arrangement and shape of spatial units can alter statistical results.

- C) It is the effect of atmospheric conditions on aerial unit measurements.

- D) It denotes the influence of unit size on data storage requirements.

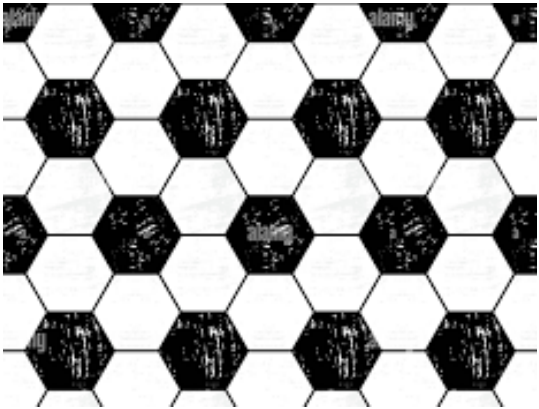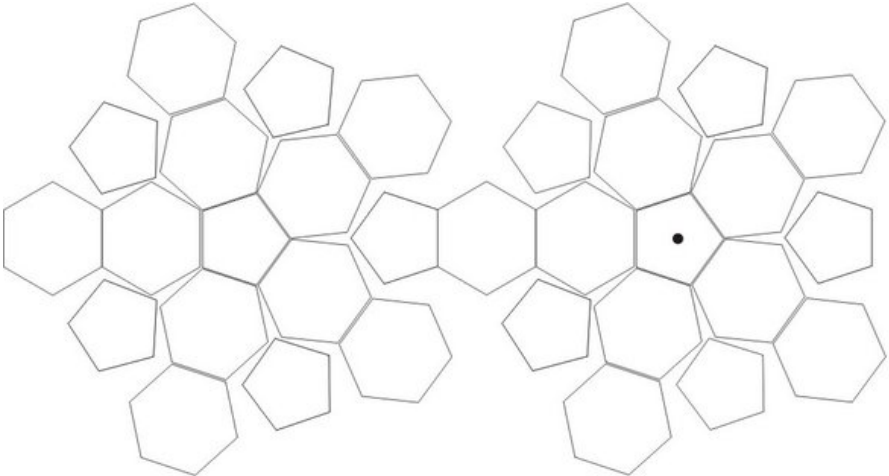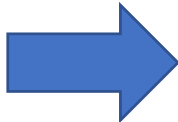- **Question**: What are some potential implications of ignoring the MAUP in crime spatial analysis?

Projection Methods

- Moving from the 3D to the 2D
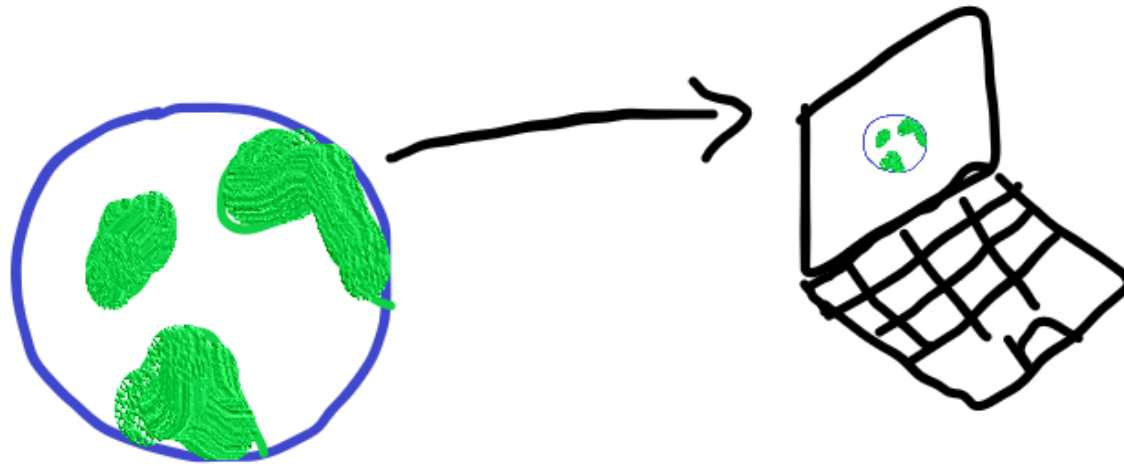
Football Example

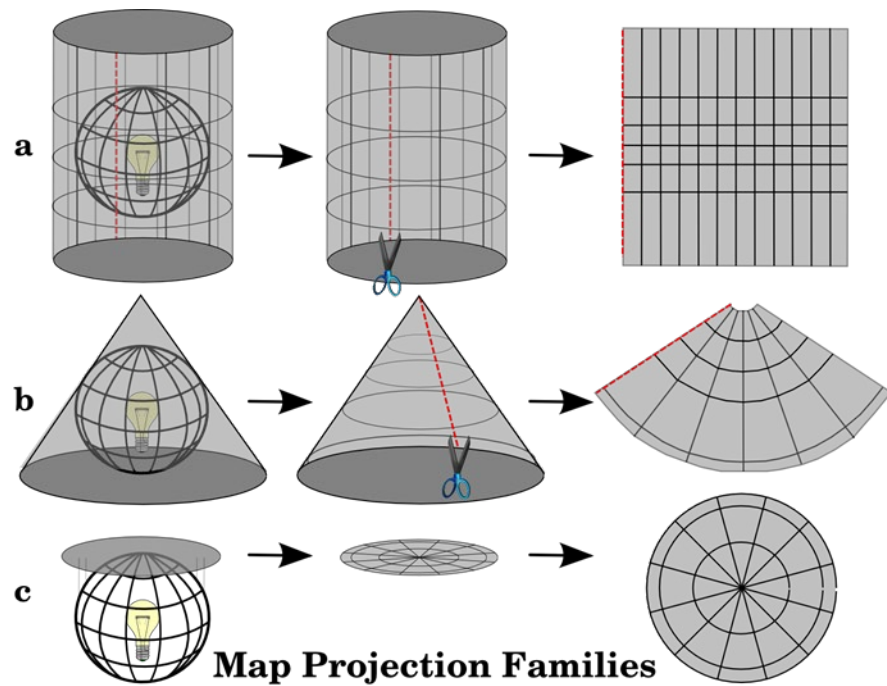# Distortion

The misrepresentation of…

-> Area
-> Shape
-> Distance
-> Direction of points

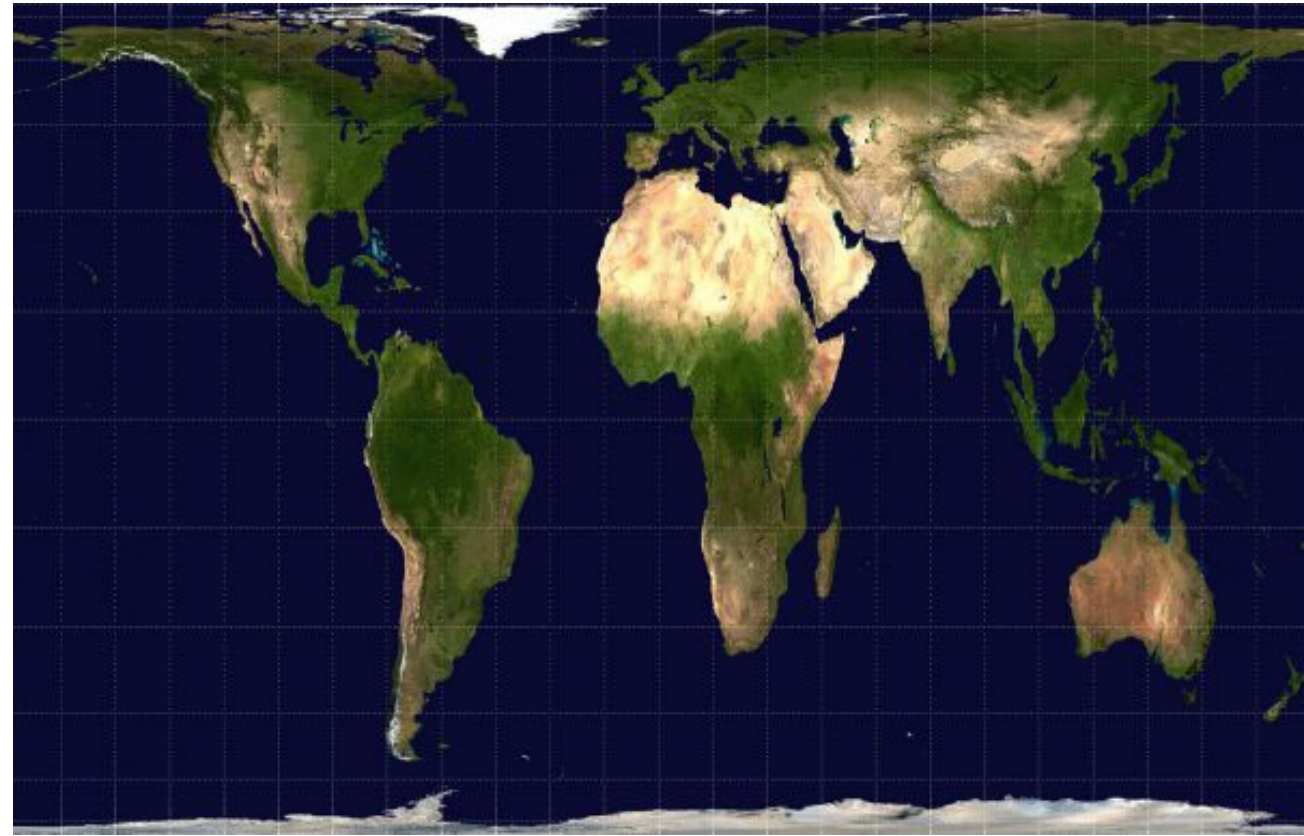# Distortion..

Map Projection Families

Cylindrical
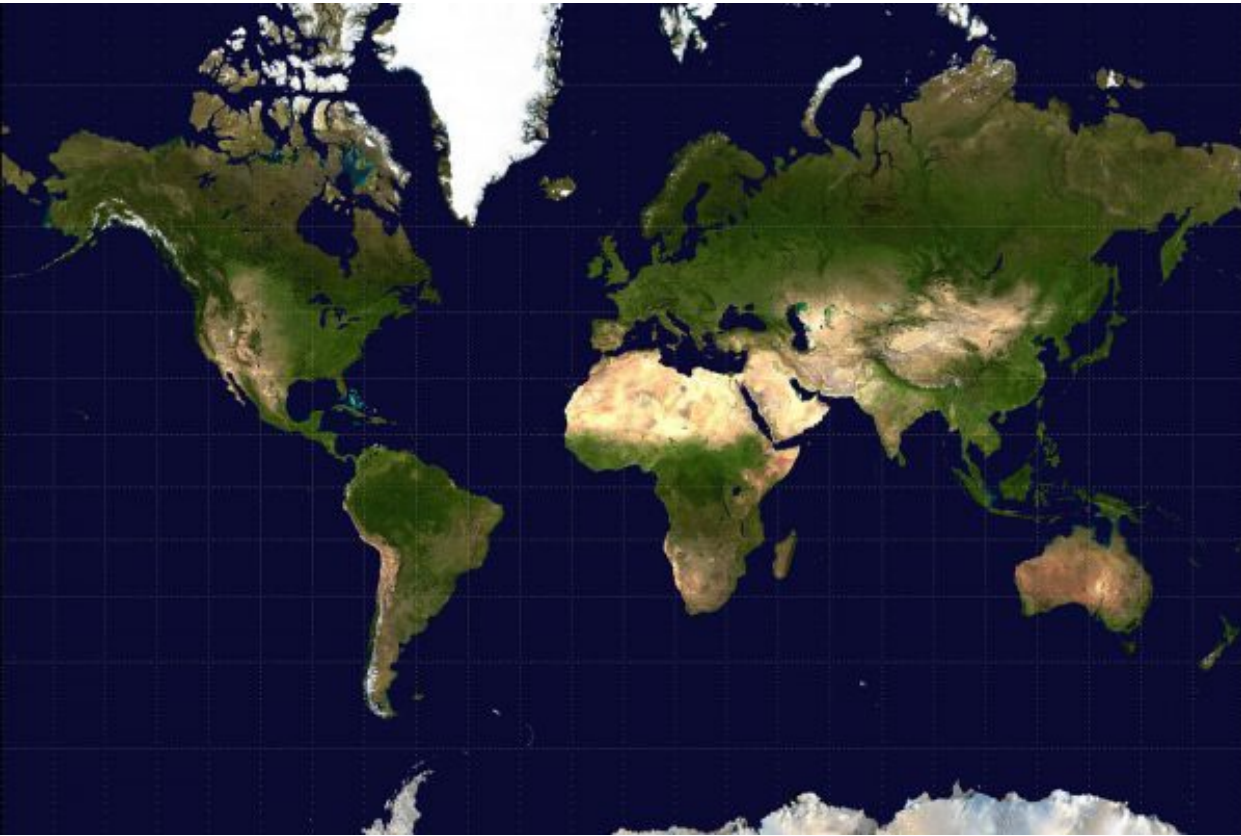
Conical

Planar

Web Mercator vs Gall-Peter Projection

So, how do we actually move from the 3D to the 2D?

(ϕ_o, λ_o)

Origin

(x_o, y_o)

Y

X
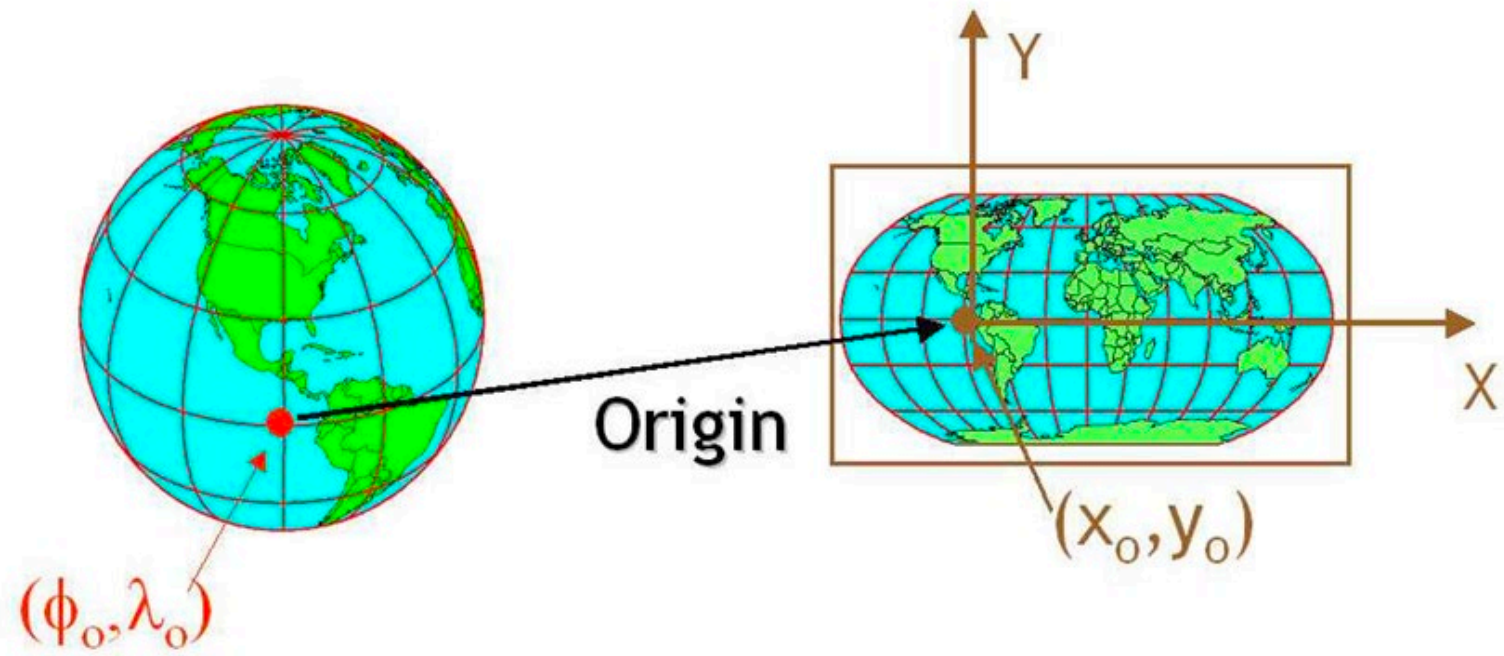
Coordinate Reference Systems (CRS)

- The move from the 3D to the 2D is done with the help of CRS
- Every place on earth is specified by three numbers (i.e. coordinates)
  - Latitude, longitude and Altitude
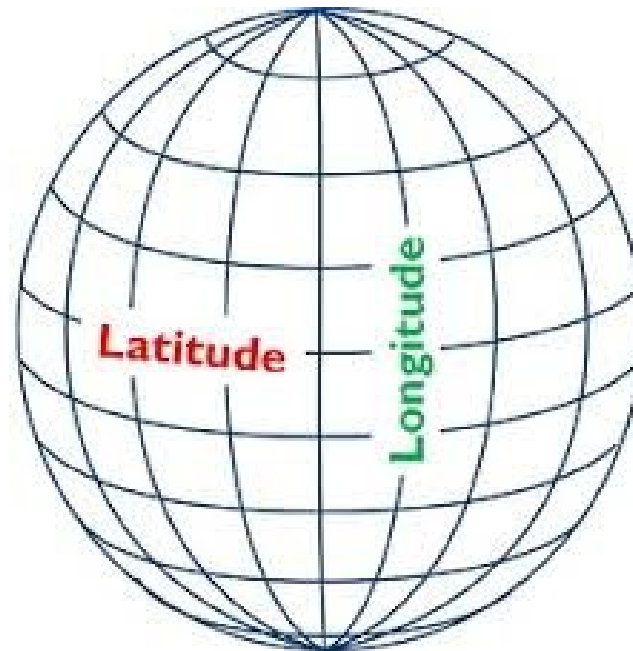
CRS continued…

There are two Main CRS:
Geographic Coordinate System vs **Projected Coordinate Systems**

The 'where'   *The 'how'*

CRS example

- Remember that a PCS, is just a GCS that has been projected !

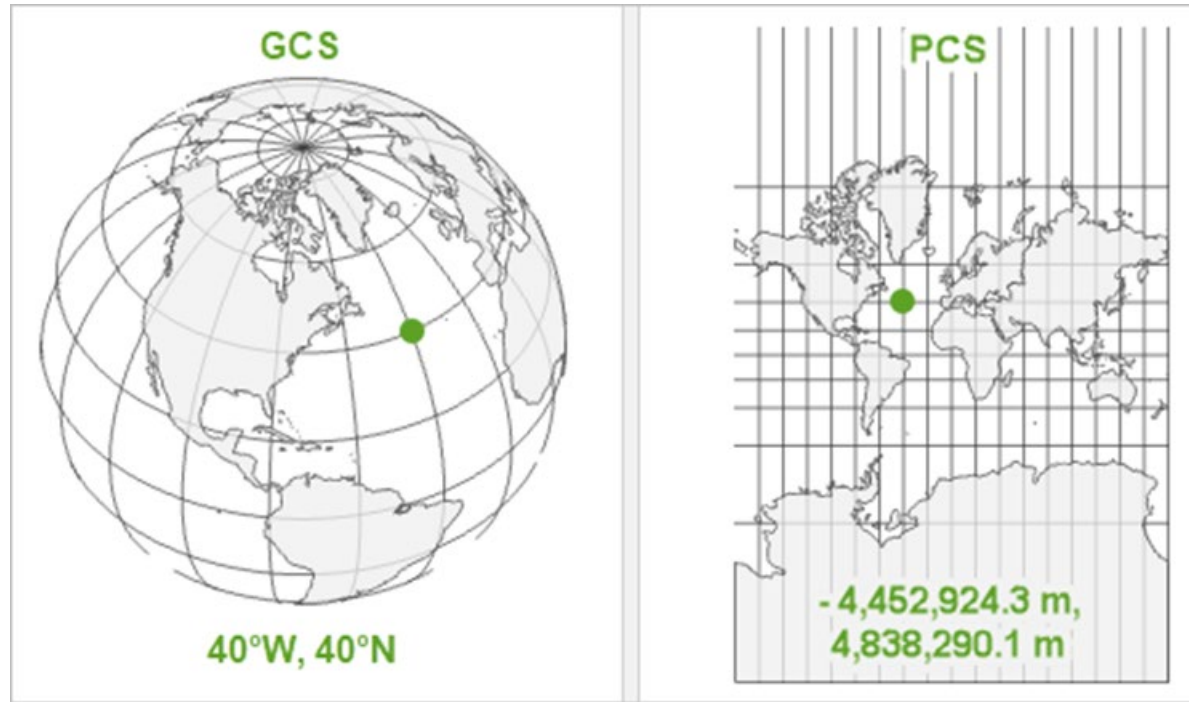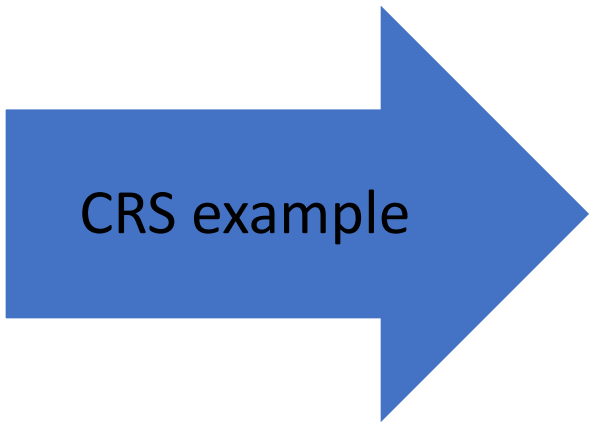## Coordinate System Details

Projected Coordinate System	Fuller (world)
Projection	Fuller
WKID	54050
Authority	Esri
Linear Unit	Meters (1.0)
False Easting	0.0
False Northing	0.0
Option	0.0
Geographic Coordinate System	WGS 1984
WKID	4326
Authority	EPSG
Angular Unit	Degree (0.0174532925199433)
Prime Meridian	Greenwich (0.0)
Datum	D WGS 1984
Spheroid	WGS 1984
Semimajor Axis	6378137.0
Semiminor Axis	6356752.314245179
Inverse Flattening	298.257223563

Tell me how to draw the earth on a flat surface!

Tell me where on the earth the data should draw!

Spatial Relations and Analysis in GIS

- **_Spatial relations_** refer to the ways in which different locations, areas, or objects are situated in relation to each other on the Earth's surface.

- **_Spatial analysis_** refers to studying entities by examining, assessing, evaluating, and modelling spatial data features

a) Point Pattern Analysis
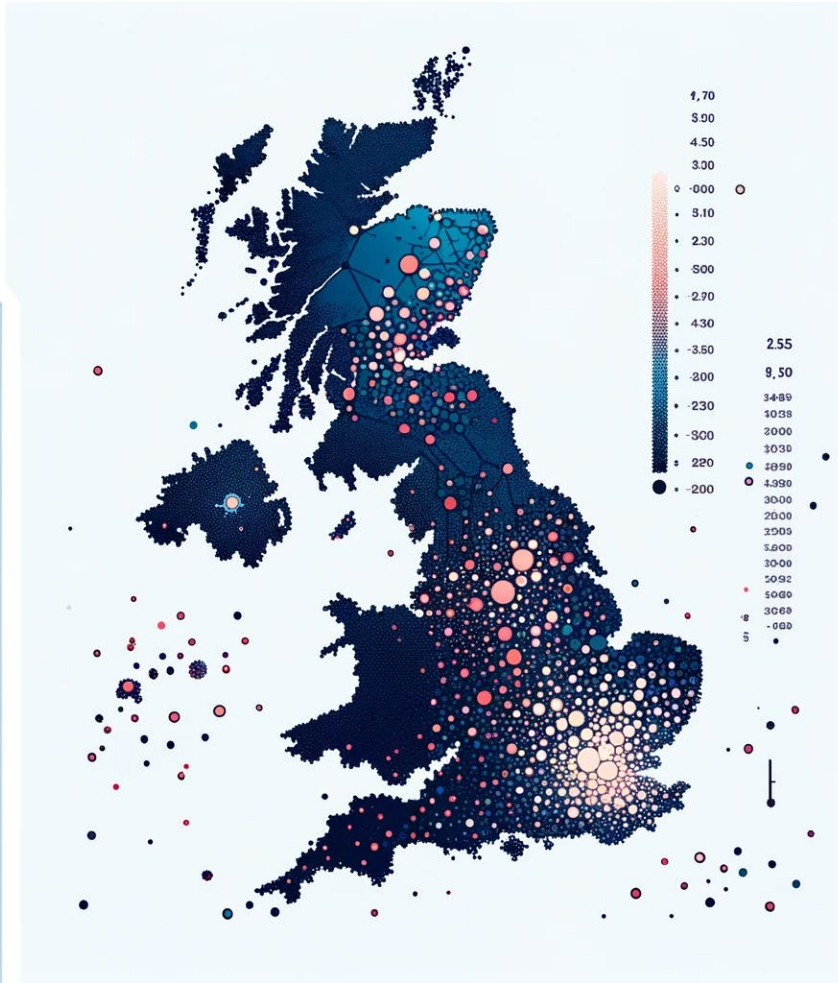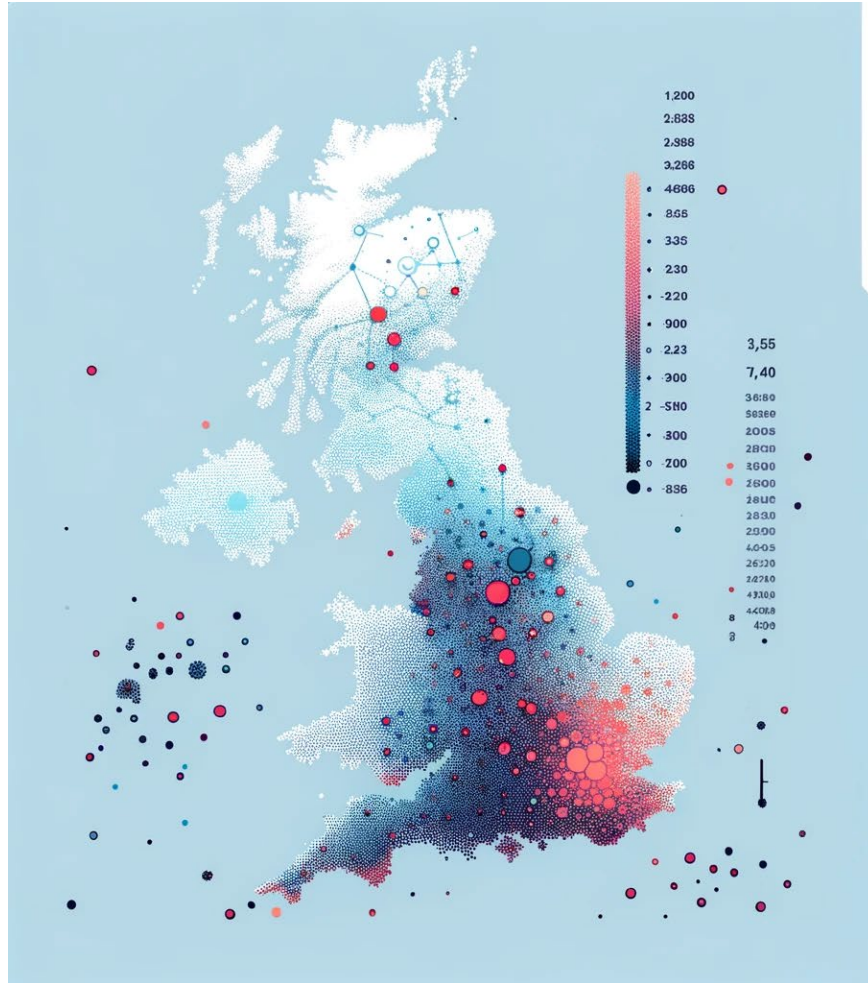
b) Spatial Autocorrelation

This map groups areas with similar crime rates together, using shades of a single color to represent different levels of crime rates, showing patterns of clustering.

c) Spatial Interpolation
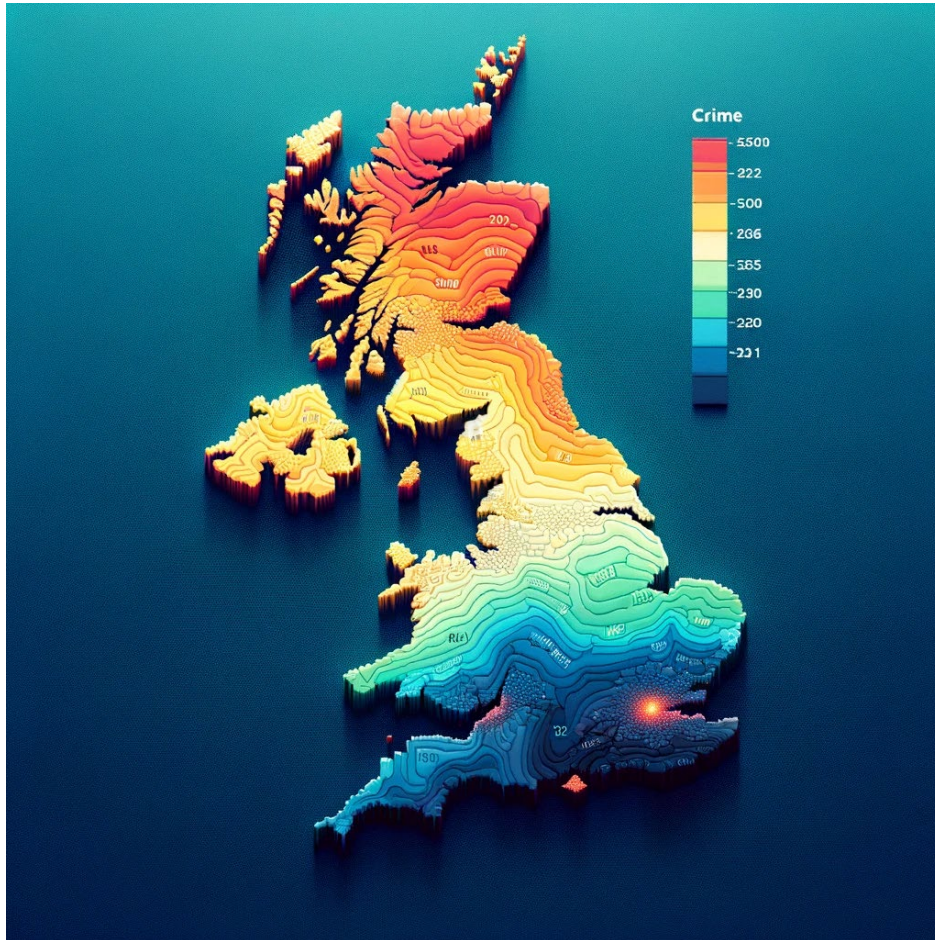
This map uses a gradient of colours to interpolate crime rates between known data points, creating a smooth transition that estimates crime rates in areas with no direct data.

d) Hot spot analysis
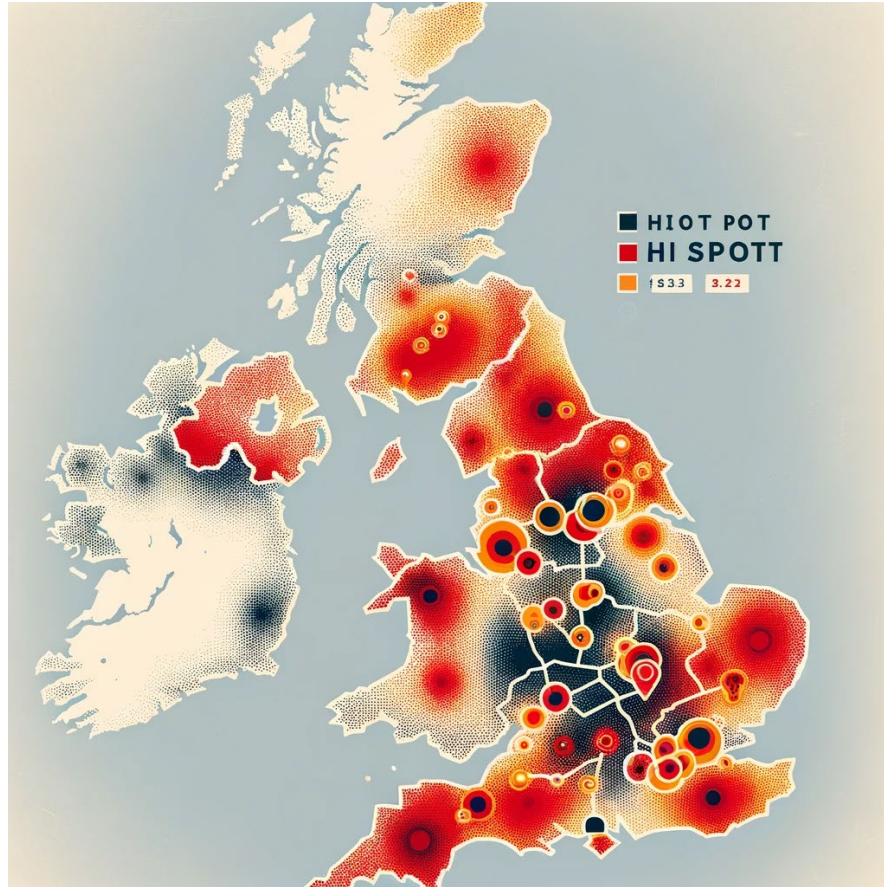
This map highlights areas with significantly higher crime rates in bright colours against a neutral background, indicating geographic concentrations of high crime rates and identifying areas needing immediate attention.

Data Sources for Crime Data

| UK	Non-UK	R Packages	Other
• UK Police Data - Provides street-level crime data, outcomes, and information on the location of the front counter. • Office for National Statistics (ONS) - Offers crime statistics for England and Wales. • Scottish Government Statistics - Publishes crime and justice datasets for Scotland. • Northern Ireland Statistics and Research Agency (NISRA) - Provides crime and justice statistics for Northern Ireland. • London Datastore - Contains datasets specific to London, including the Metropolitan Police Service's crime data. • UK Home Office - Publishes research and statistics on various aspects of the criminal justice system. • Data.gov.uk - A portal for UK government data, including some datasets related to crime and justice.	• FBI's Uniform Crime Reporting (UCR) Program (USA) - Provides a nationwide view of crime based on the submission of statistics by law enforcement agencies. • Bureau of Justice Statistics (BJS) (USA) - Contains criminal victimization data, law enforcement, prosecution, courts, and corrections data. • National Incident-Based Reporting System (NIBRS) (USA) - Offers detailed information about crimes reported to law enforcement. • European Sourcebook of Crime and Criminal Justice Statistics - Compiles criminal justice data from European countries. • INTERPOL's Crime Data - Allows access to international crime statistics. • United Nations Office on Drugs and Crime (UNODC) - Provides global crime and criminal justice statistics.	• crimedata: An R package that provides easy access to cleaned and documented criminal incident data from several U.S. cities. • UKpolice: An R package that uses the data.police.uk API to provide access to UK police data. • opendatatoronto: Although not specific to crime data, this package can access Toronto's Open Data Portal, which includes crime statistics. • crimer: An R package that allows users to access and analyze crime data from the city of Baltimore. • rdflib: While not a crime-specific package, rdflib can be used to parse RDF/XML crime data provided by some organizations.	• Other Resources: • GIS Data Repositories: Such as Esri's Open Data Hub or Harvard's Geospatial Library, which offer geospatial data that can be used in crime mapping. • Social Media and Crowdsourced Data: Platforms like Twitter and Nextdoor, where users may report crime-related incidents. However, this data is unverified and needs careful processing. • Academic Institutions: University crime labs and research centers sometimes share datasets for public use. • Private Data Vendors: Companies like LexisNexis provide crime reports and analytics services, though they often require a subscription.

What are the main challenges of mapping crime data?

- Geomasking and geoprivacy
- The accuracy of police recorded statistics
  - The grey figure of crime
  - Conceptual issues surrounding its definitions of crime types
  - The impacts of seasonality; how has Covid-19 affected police recorded crime statistics

Can you think of anymore?

What are the main challenges of mapping crime data?
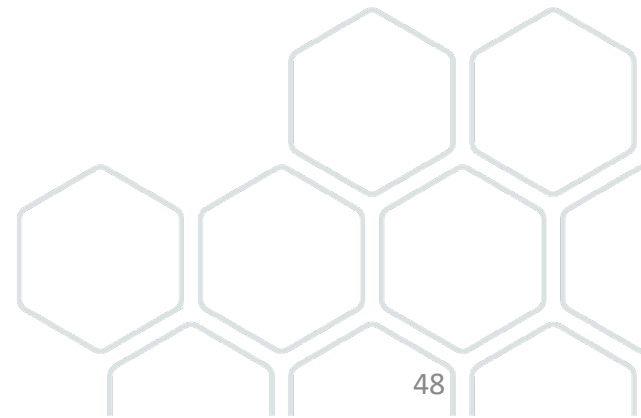
Material for the live code demonstration (08/03/2023)

GitHub:

https://github.com/UKDataServiceOpen/Crime_Data_in_R

(under the Feburary_2024 folder)

Any Questions…

References

- https://www.esri.com/en-us/what-is-gis/overview#liSwitcher

- https://blog.ukdataservice.ac.uk/gis-spatial-data/

- Dermanis (2005) : https://www.researchgate.net/profile/Athanasios-Dermanis/publication/233387161_Coordinates_and_Reference_Systems_in_Greek/links/0912f50a149d9568b7000000/Coordinates-and-Reference-Systems-in-Greek.pdf

- Luc Anselin (2009) : https://link.springer.com/chapter/10.1007/978-3-642-03647-7_5

- Ratcliffe (2009) Spatial and Temporal Challenges: https://link.springer.com/chapter/10.1007/978-0-387-77650-7_2

- Bowers (2001) Mapping and Analysing Crime Data : https://link.springer.com/chapter/10.1007/978-0-387-77650-7_2

- https://core.ac.uk/download/pdf/161890428.pdf

- https://www.ed.ac.uk/files/imports/fileManager/chainey.pdf

- https://www.crimrxiv.com/pub/slkb1v54/release/1

- https://tubecreature.com/#/livesontheline/current/same/U/*/FFTFTF/11.469326848406268/-0.1622/51.5142/

- *https://www.crimrxiv.com/pub/slkb1v54/release/1*

- https://www.sciencedirect.com/topics/earth-and-planetary-sciences/modifiable-areal-unit-problem

- https://engaging-data.com/country-centered-map-projections/

- Images on slide 28, 42, 43, 44 and 45 were generated with the assistance of AI. Source: ChatGPT 4 with Dall-E. Date: 07/01/2024.

UK Data Service

Thank You.

Email: nadia.kennar@manchester.ac.uk

Twitter: @NadiaKennar