Outbreak Investigation Module: Mapping in R

Introduction to case study

Amy Mikhail

Friday 09 December 2022

Objectives / session structure

1. Creating geographic maps: what you need to know (refresher)
2. Example maps
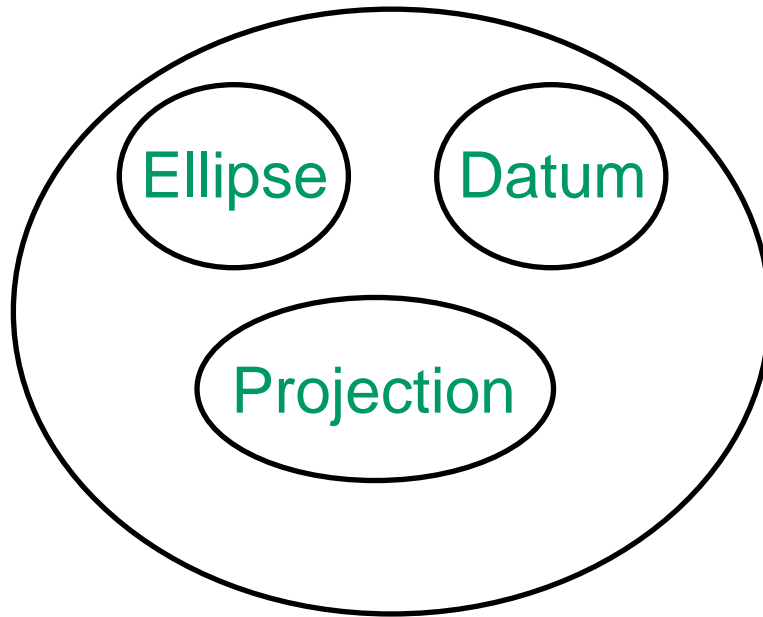3. Introduction to mapping case study

Mapping terminology refresher

Some key terminology: a refresher

- CRS
- Ellipse
- Datum
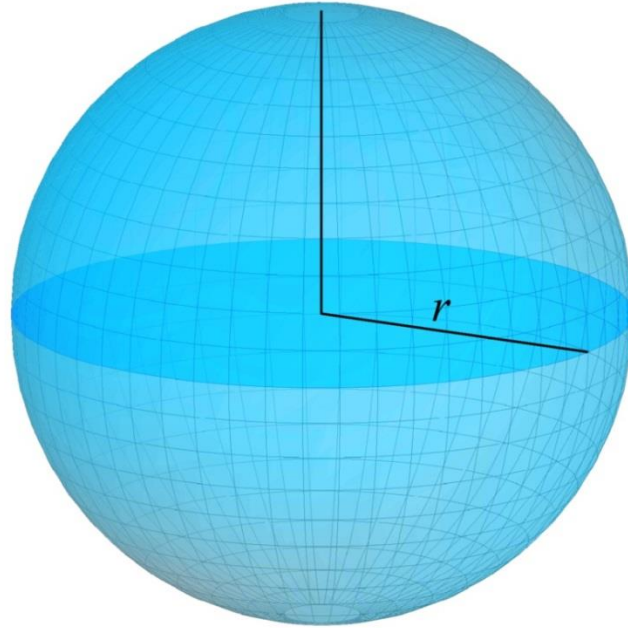- Projection
- EPSG

Terminology

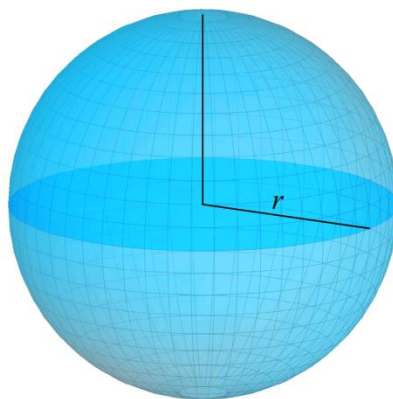CRS = Coordinate Reference System

Terminology

Ellipse

Datum
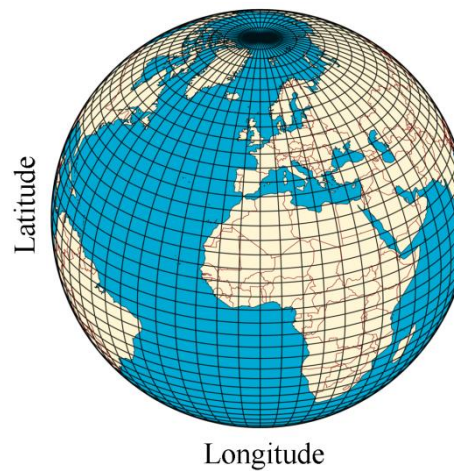
Terminology
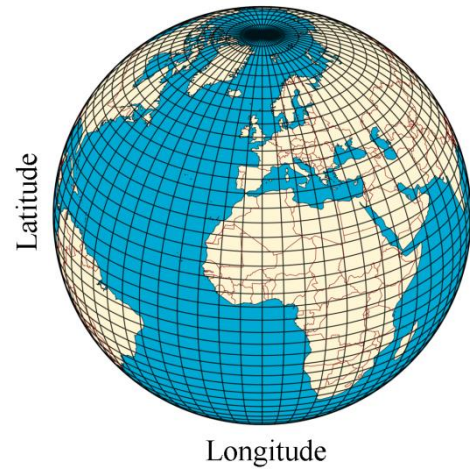

Ellipse   Datum   3D Globe
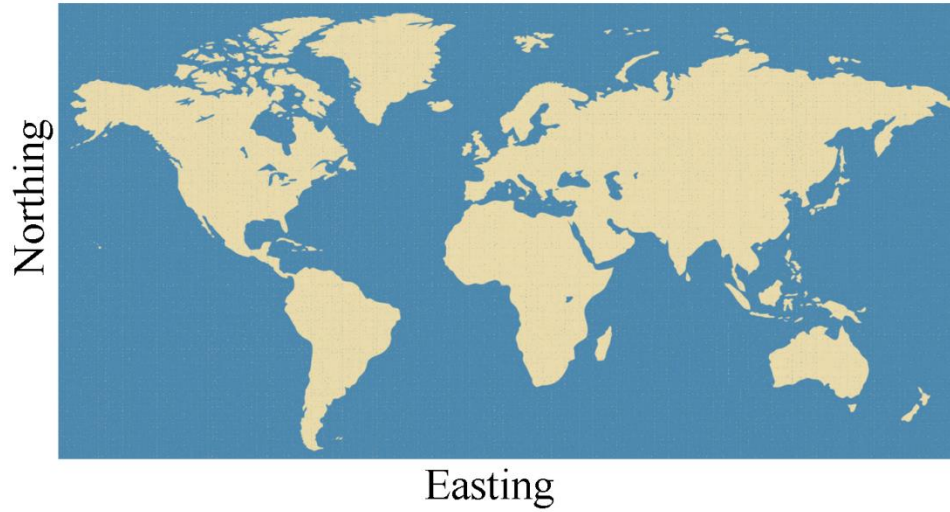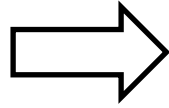
Terminology

3D Globe   2D map

Terminology

EPSG   =

Terminology

WGS84 (EPSG: 4326)
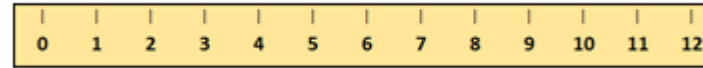
OSGB 1936 (EPSG: 27700)
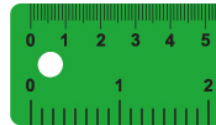
NAD27 (EPSG: 4267)
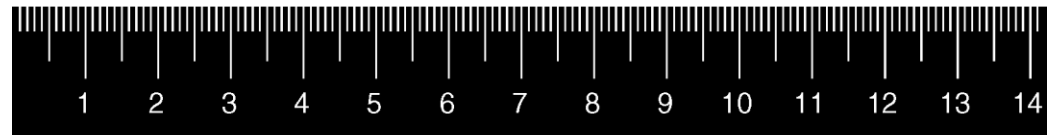
Maps in R: some examples

Why map in R?

- Repeatable

- Link to data sources

- Ability to annotate

- Integration between visualisation and analysis

- Convert between geographic reference types

Challenges of mapping in R

1. Many different approaches & packages – need familiarity with R

2. Certain mapping tasks need connection to the internet:
   - Converting postcodes and addresses to a suitable CRS
   - Defining the edges of your map
   - Importing static background map tiles
   - Importing shape files (unless stored locally)

Example 1: static point map of cases
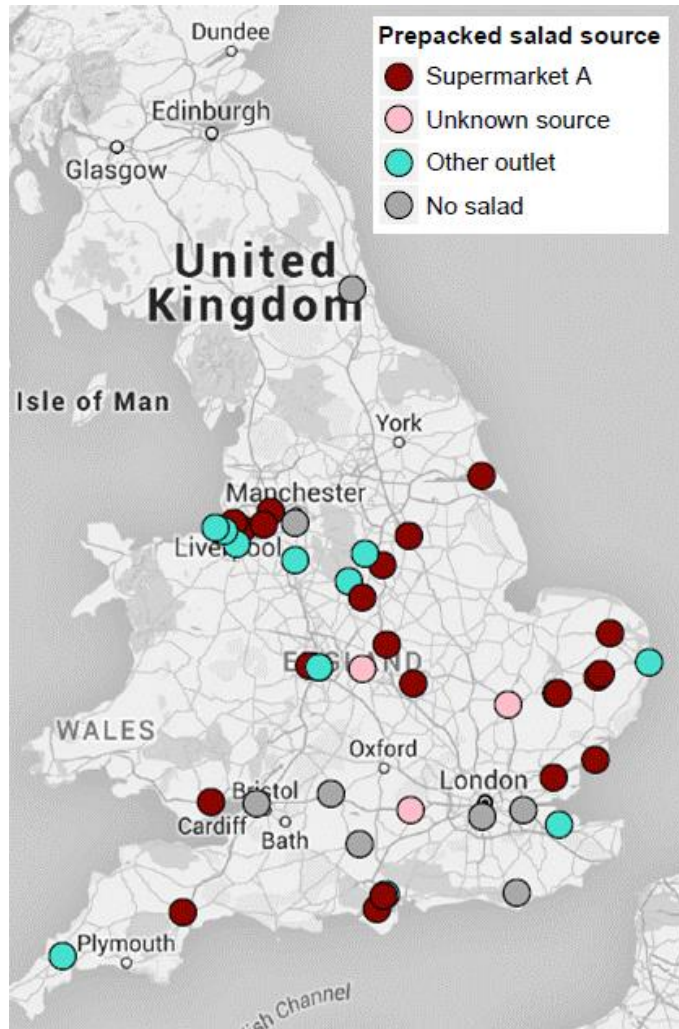
- Map outbreak cases

- No need for shape file

- Can overlay points on OpenStreetMap

- Access to map options via ggmap

- Map is secure (on your local machine)

- Can stratify by exposure etc.

Example 2: case map with shape files

- Map outbreak cases
- Shape files: coordinates for polygons that define specific areas or regions
- Overlay points on specific geographic areas (e.g. health regions)
- Limit the map to a specific geographic area (region or country)

Example 3: choropleth map of incidence

Polio vaccination (IPV) coverage in infants aged 12 months, London, UK 2014
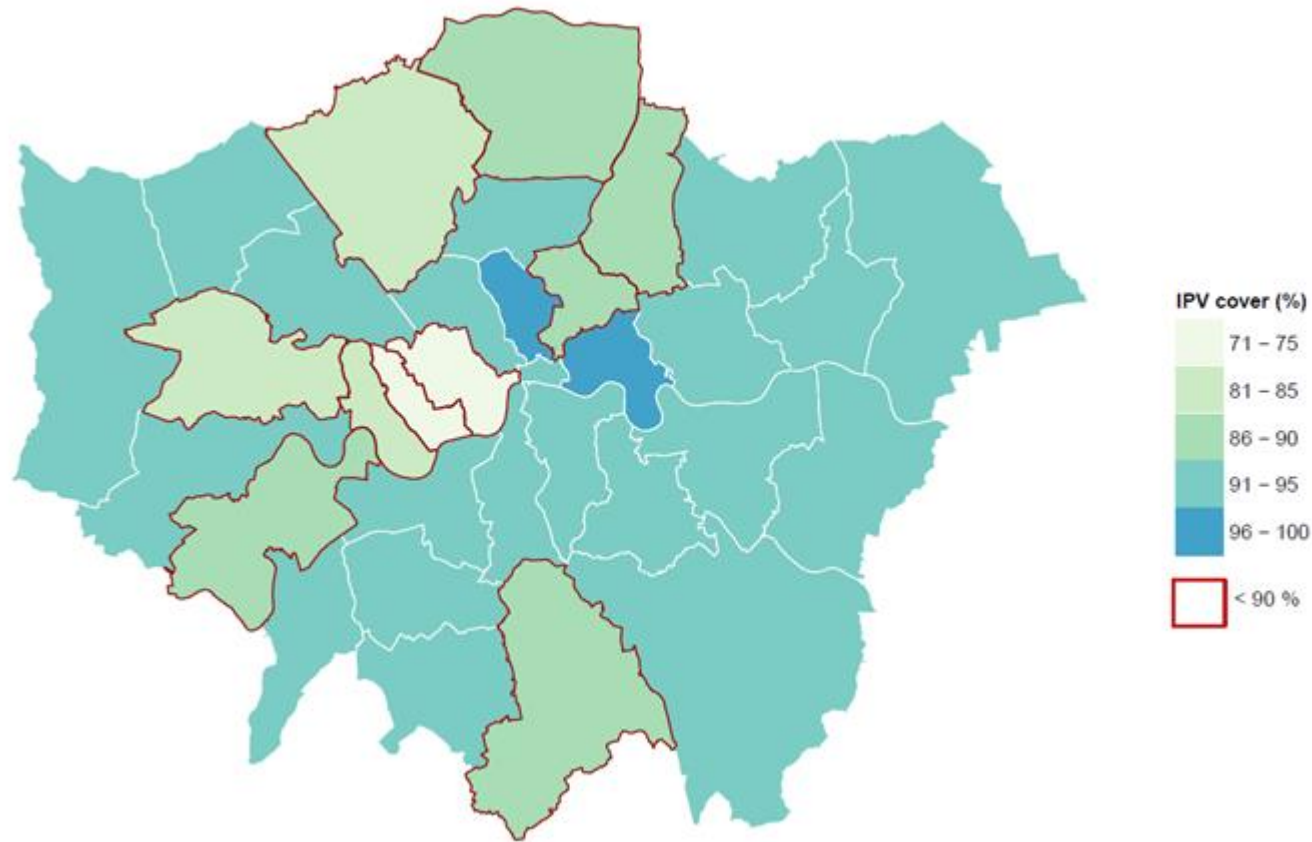
Example 4: map of case distribution over time
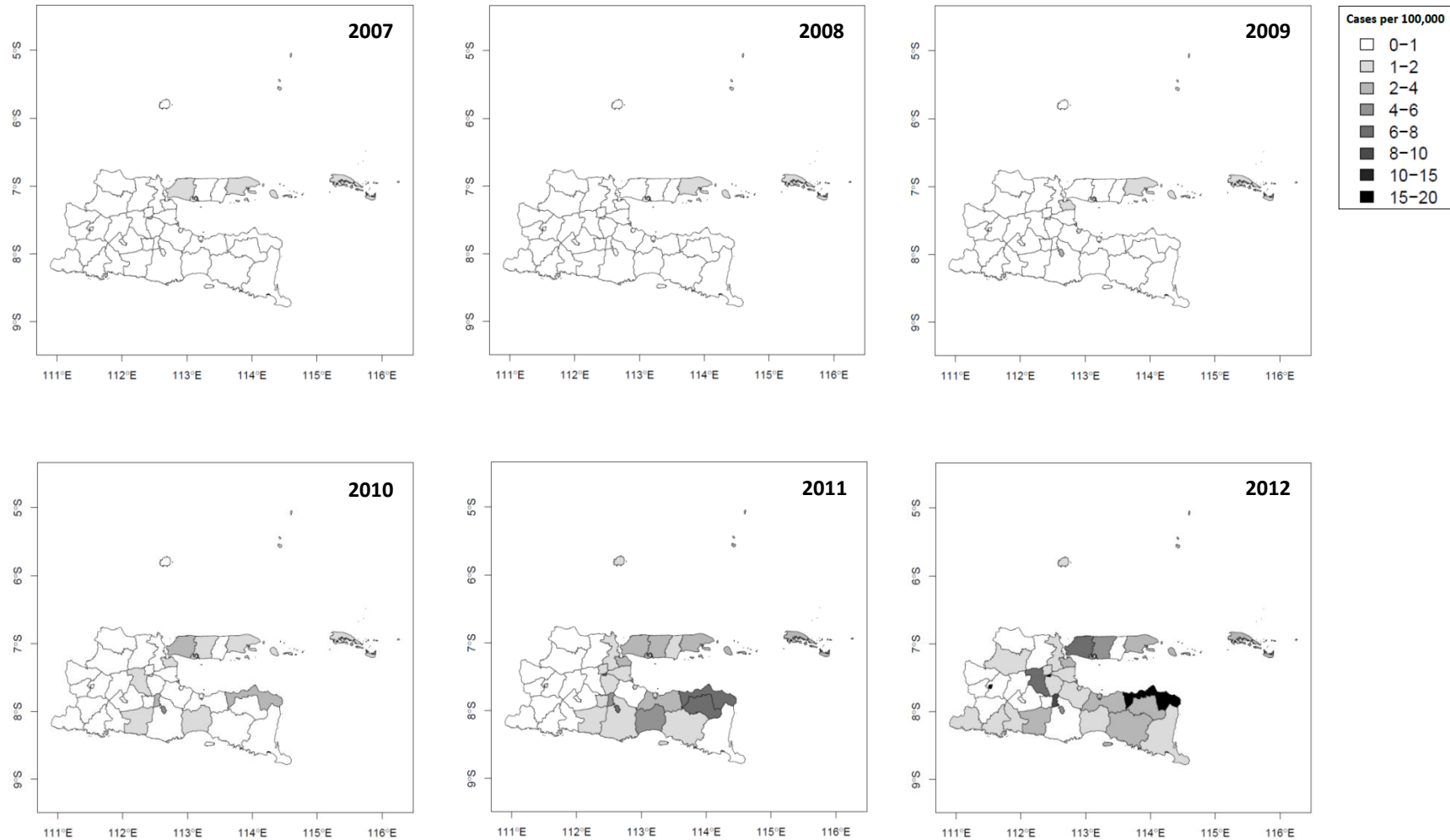
Diphtheria incidence per 100,000 population in East Java, Indonesia, 2007 - 2012

Example 5: contour map with hot-spots
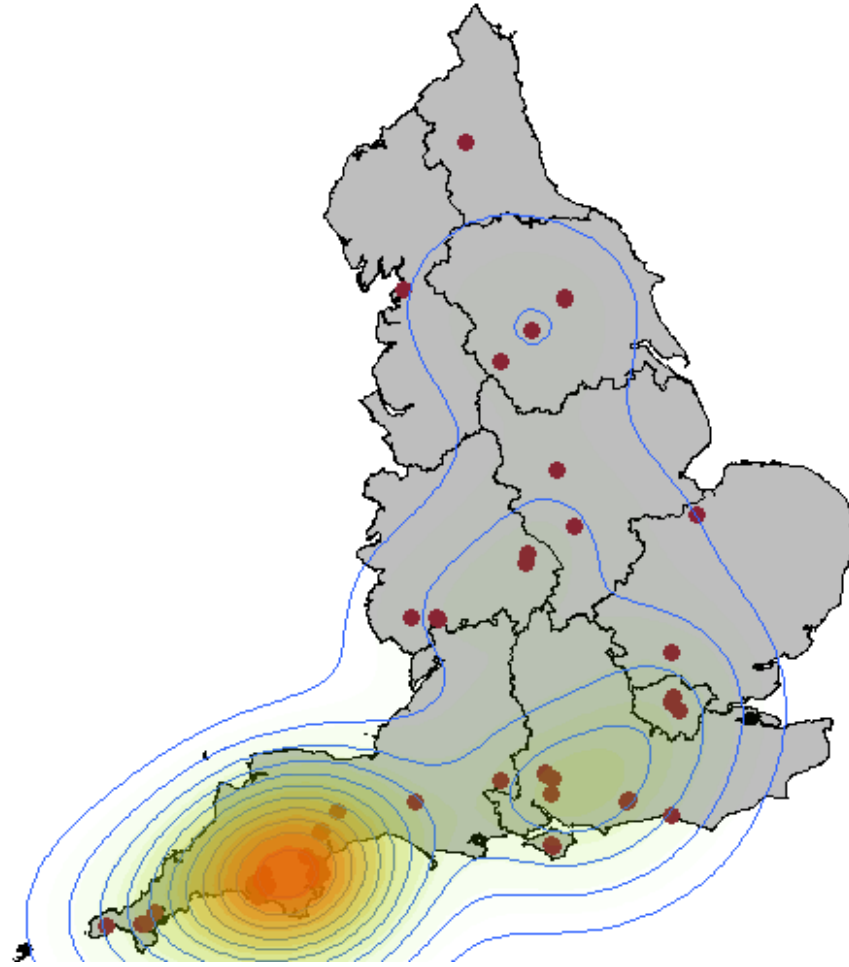
Example 6: map with spatial statistics (Satscan)

cluster	start	end	observed	expected	p.value
1	14 Sep 2015	25 Oct 2015	20	7.15	0.581
2	14 Sep 2015	25 Oct 2015	82	53.59	0.628
3	21 Sep 2015	25 Oct 2015	5	0.48	0.733
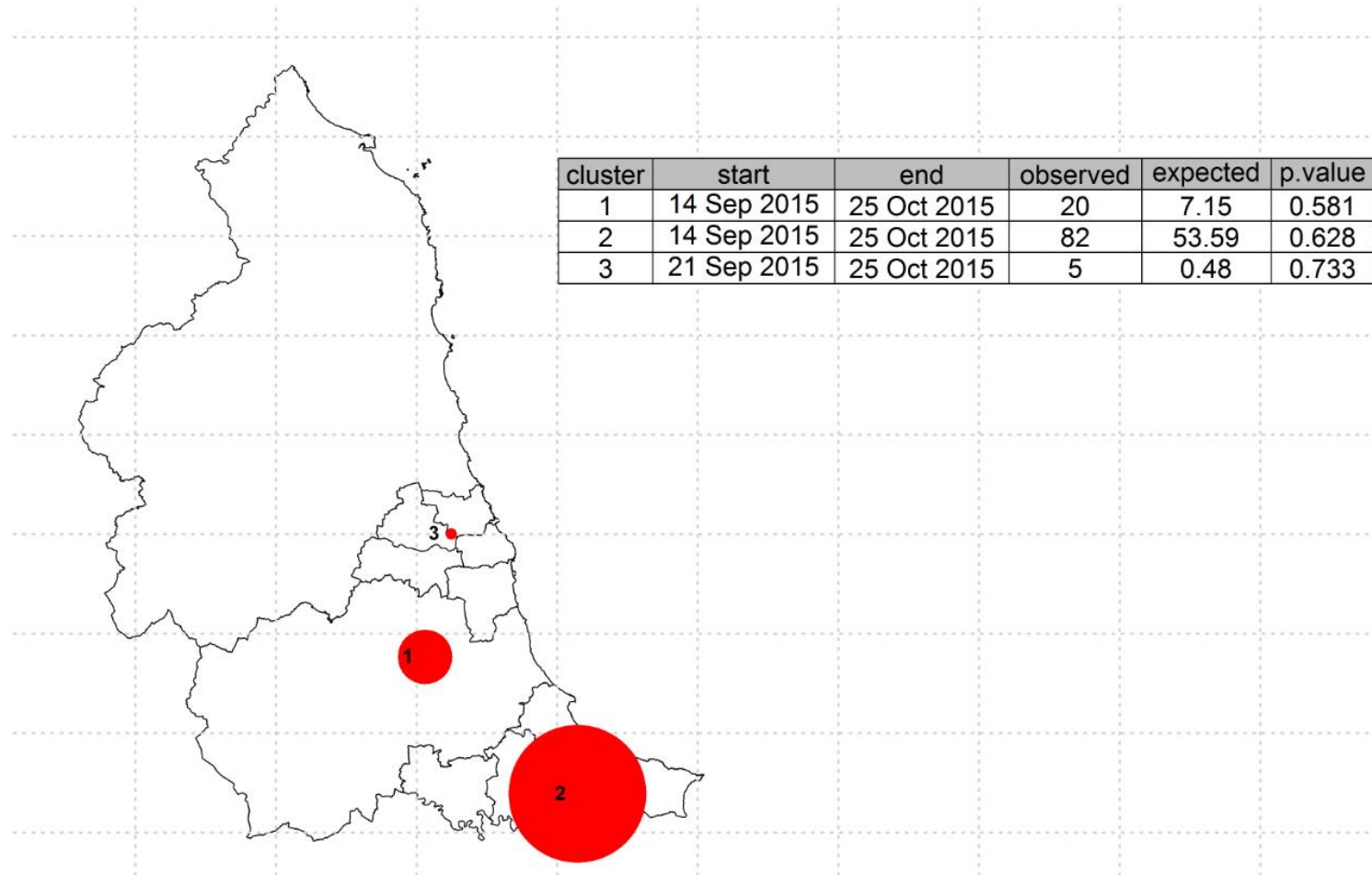
Mapping in R: case study

Introduction

R mapping case study: materials

Software:

- <u>R</u> version 4.2.0 or greater
- <u>RStudio</u> 2021.09.1 build 372 or similar

R packages:

- <u>Data wrangling</u>: `here, tidyverse, ggplot2, htmlwidgets`
- <u>Mapping</u>: `tidygeocoder, ggmap, sf, osmdata, scales, leaflet`

Data:

- CaseStudy_RDM_anon_data_coords.csv (outbreak case line list)
- PHEC_population.csv (population for health regions)
- En_PHE_Centre.shp (shape files for health regions)

R mapping case study: folder structure

All material is provided in the Case study pack:

- <u>data folder</u> (contains three raw data files)

- <u>guide folder</u> (contains this presentation and the html practical guide for this session)

- <u>shapefiles folder</u> (contains shape files for health regions in England)

In addition the following are provided in the root of the Mapping folder:

- `.here`

- `Mapping.RPROJ`

- `Mapping_R_template.RMD`

R mapping case study: getting started

1. Save the **Mapping** folder somewhere on your computer.

2. Open the **Mapping** folder and double click on **Mapping.RPROJ** (this will open RStudio)

3. In the **guide** sub-folder, open **Mapping_R_guide.html** (this will open in your browser)

4. From your RStudio console, go to:

    - Files tab
    - Click on Mapping_R_template.R to open it within Rstudio

5. Work through the html guide, modifying and adding to code in the R template file as you go.

R mapping case study: setting the scene

Initial notification of the outbreak:

- Outbreak of Shiga toxin-producing *Escherichia coli* (STEC)
- Serotype O157:H7, Phage type 21/28, *stx 2a*
- 5 cases identified in autumn 2014
- All exposed to unpasteurised milk from a farm in South West England
- Cattle from the farm were infected with the same STEC strain

Subsequent information:

- Case isolates belonged to a 5-SNP whole genome sequencing (WGS) cluster
- 4 further cases identified by WGS as also falling within this cluster
- Wider phylogeny (25-SNP level) identified 59 linked cases in total
- Unclear if these additional cases had the same exposure

R mapping case study: geospatial investigation

Because of the uncertainty around exposures for the additionally identified cases, you have been asked to investigate the spatial distribution of these cases, specifically:

1. Are pre-2014 cases also clustered near the farm in the South West?

2. Are all cases from the outbreak year (2014) clustered near the farm in the South West?

3. If not, does their spatial distribution pattern indicate an alternative source for this strain of STEC?

4. Are phylogenetically linked cases from other outbreaks also linked to the South West region?

Mapping in R: case study

Conclusions

Mapping in R: case study conclusions

- Further investigations revealed that some cases with no travel history to the South West of England, had bought raw milk from the farm via a mobile van, which made deliveries over a wide area.

- Despite an initially dispersed pattern of distribution, corrections for case and population density revealed significant geospatial clustering in the South West region, with a hot-spot around the implicated farm.

- Looking at the geographic dispersal of cases over time revealed that the outbreak strain had been circulating in this area for some years prior to notification of this event.

- Public health measures were implemented to address the risk posed by raw drinking milk in the area and more widely across England.

Mapping in R: case study conclusions

Fig. 4. Map showing the single point of exposure for cases falling within the outbreak cluster between 2009 and 2015. RDM, Raw cows' drinking milk.

Reference:

Butcher H, Elson R, Chattaway MA, Featherstone CA, Willis C, Jorgensen F, Dallman T, Jenkins C, McLauchlin J, Beck C, & Harrison S (2016). Whole genome sequencing improved case ascertainment in an outbreak of Shiga toxin-producing Escherichia coli O157 associated with raw drinking milk. Epidemiology and Infection, 144(13), 2812-2823.

https://doi.org/10.1017/S0950268816000509