

HIGH QUALITY MAPS WITH R AND GGPLOT

Simon
Hailstone

WHAT AM I HERE TO TALK ABOUT?

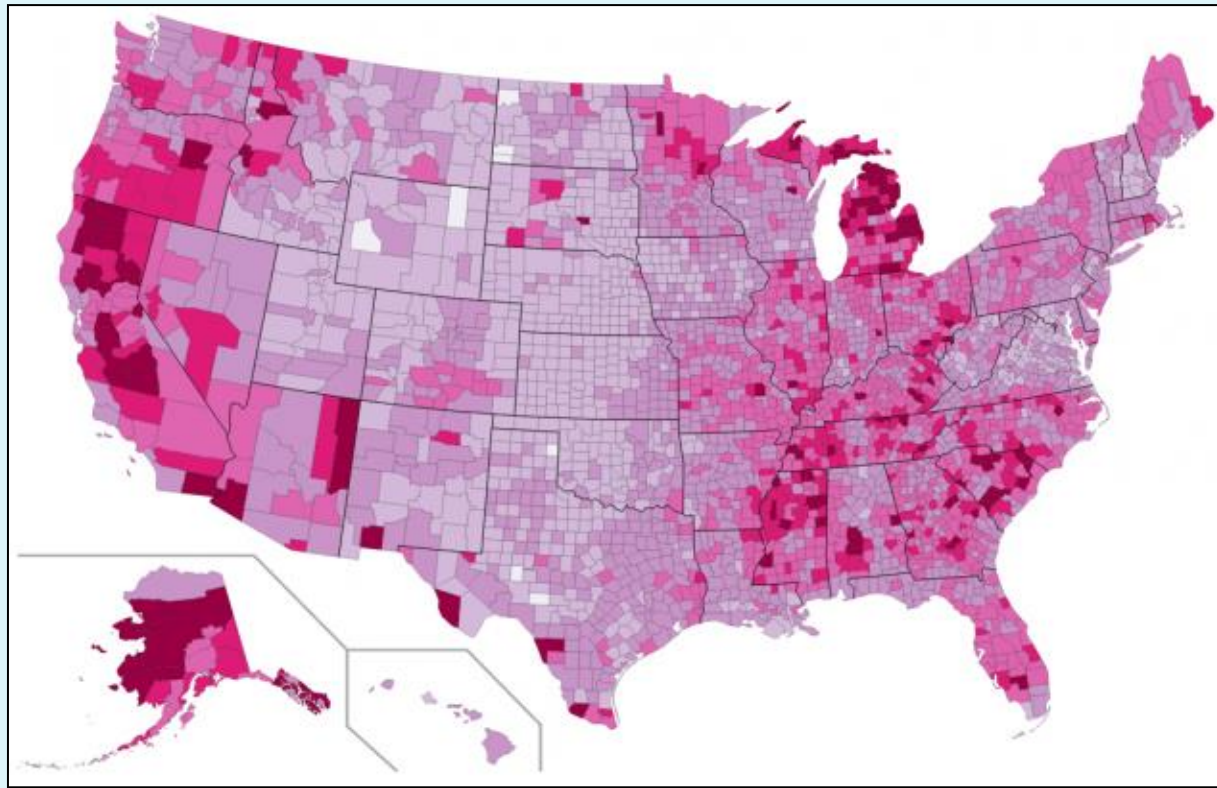
- Where to get data to map
- Where to get shapefiles
- How to geocode the easy way
- How to combine all of this in ggplot
- Pros and Cons



WHY USE R FOR MAPS?

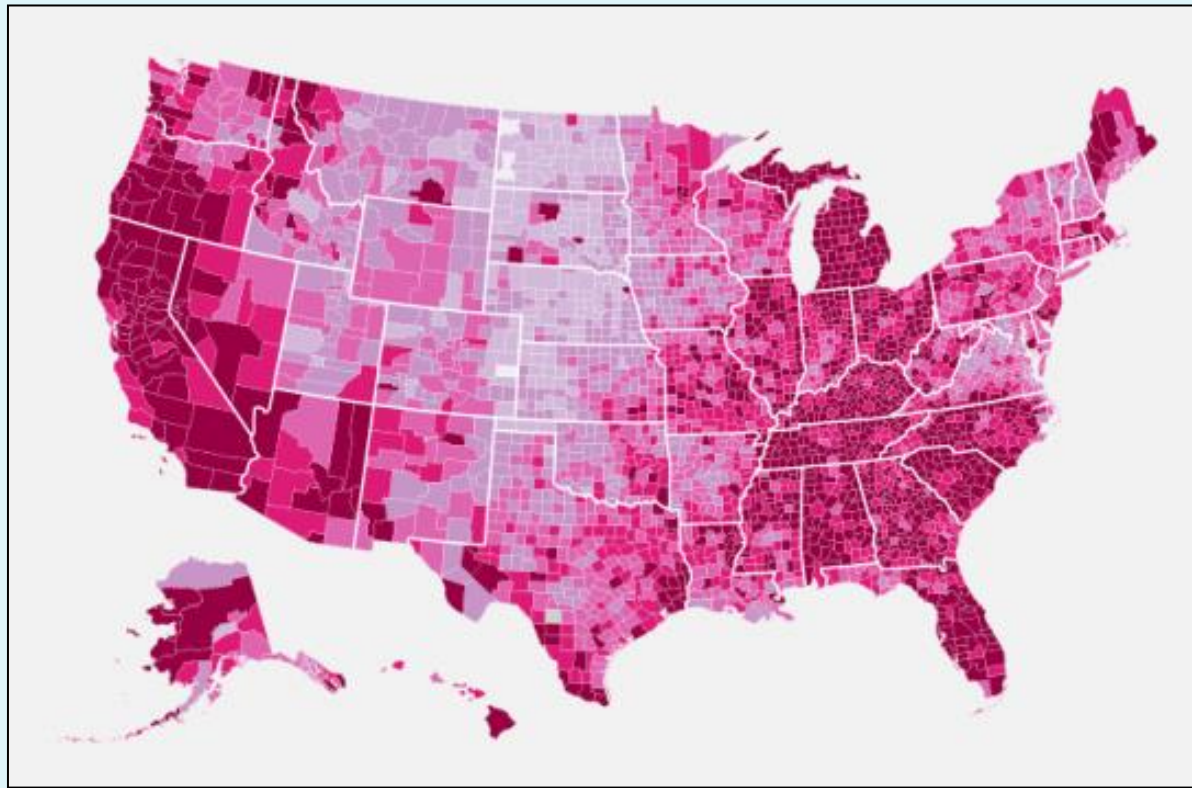
- Business Objects Reporting System → Limited charting capabilities and no maps!
- R can be used to run through chunks of data and generate image files
- R can dynamically name the files it creates
- Business Objects can then concatenate hyperlinks together

THE BENCHMARK



<http://flowingdata.com/2009/11/12/how-to-make-a-us-county-thematic-map-using-free-tools/>

THE BENCHMARK



<http://www.thisisthegreenroom.com/2009/choropleths-in-r/>

DATA SOURCES

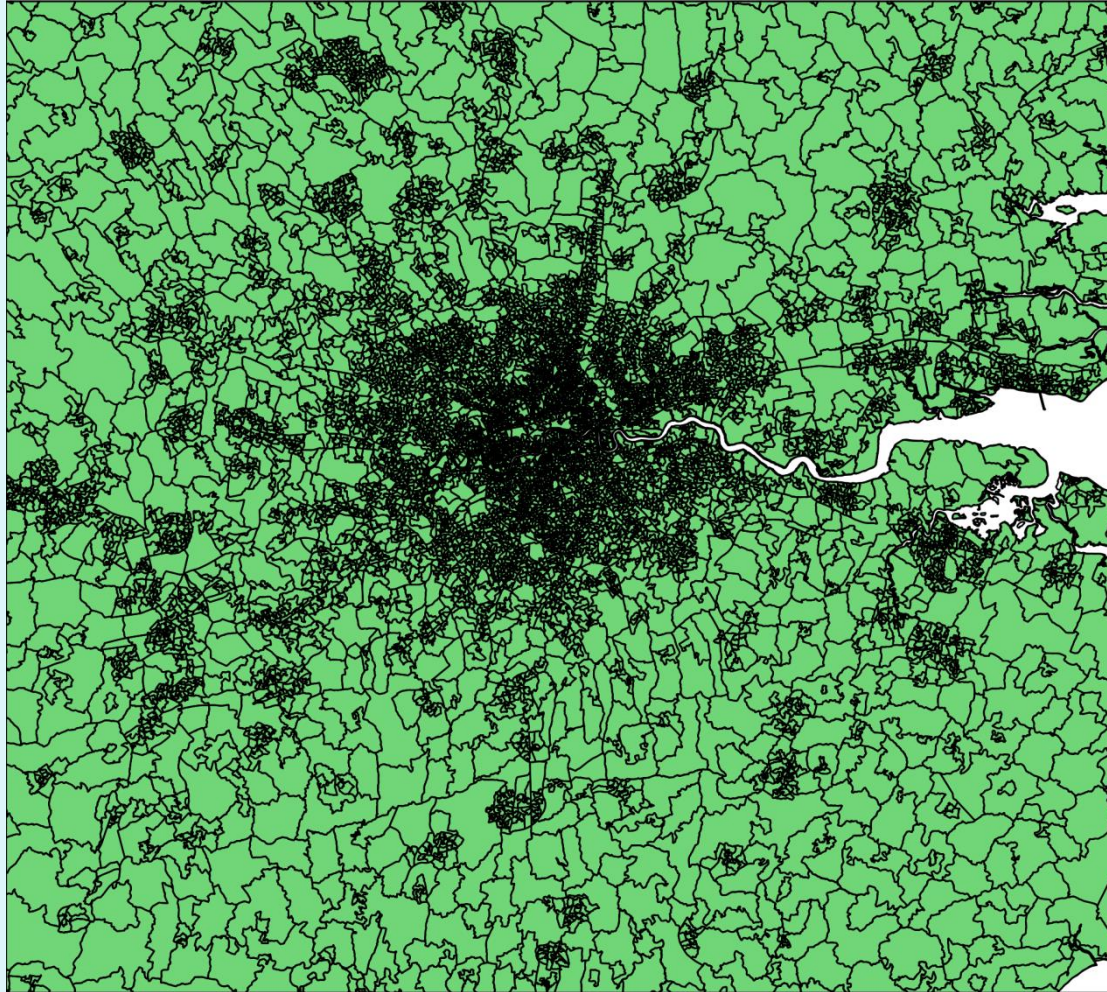
FIND SOMETHING INTERESTING

- Lots of statistics in the UK are released with a geographic dimension
- Plenty of examples over at data.london.gov.uk/datastore and also at the Office of National Statistics (ONS)
- Look for the terms MSOA and LSOA! This allows really easy matching of data to UK geography

LSOA AND MSOA

- We are very lucky in the UK to have the ONS who maintain geographic sub-divisions of the country
- The whole country is split into census output areas. These are then aggregated to form larger areas such as wards and regions.
- LSOA and MSOA are common geographic areas for which statistics are released

LSOA AND MSOA



FIND SOMETHING INTERESTING

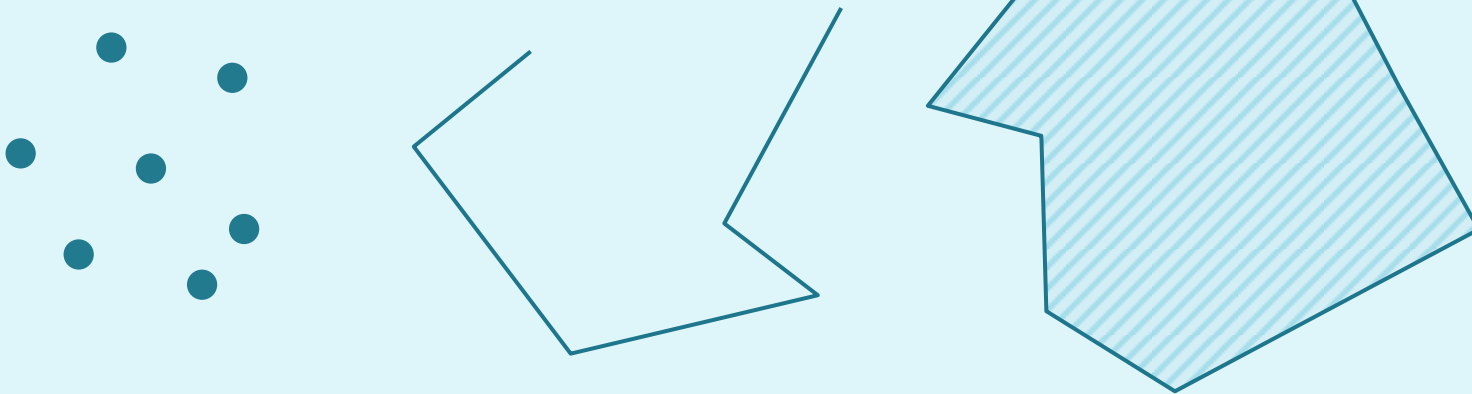
- Ambulance Service Incidents from London datastore website
- LSOA level dataset for a couple of interesting domains
 - Binge Drinking
 - Assault
 - Deprivation
 - Population
- Also identified A&E departments and sizes to plot as reference points: <http://www.england.nhs.uk/statistics>



SHAPEFILES

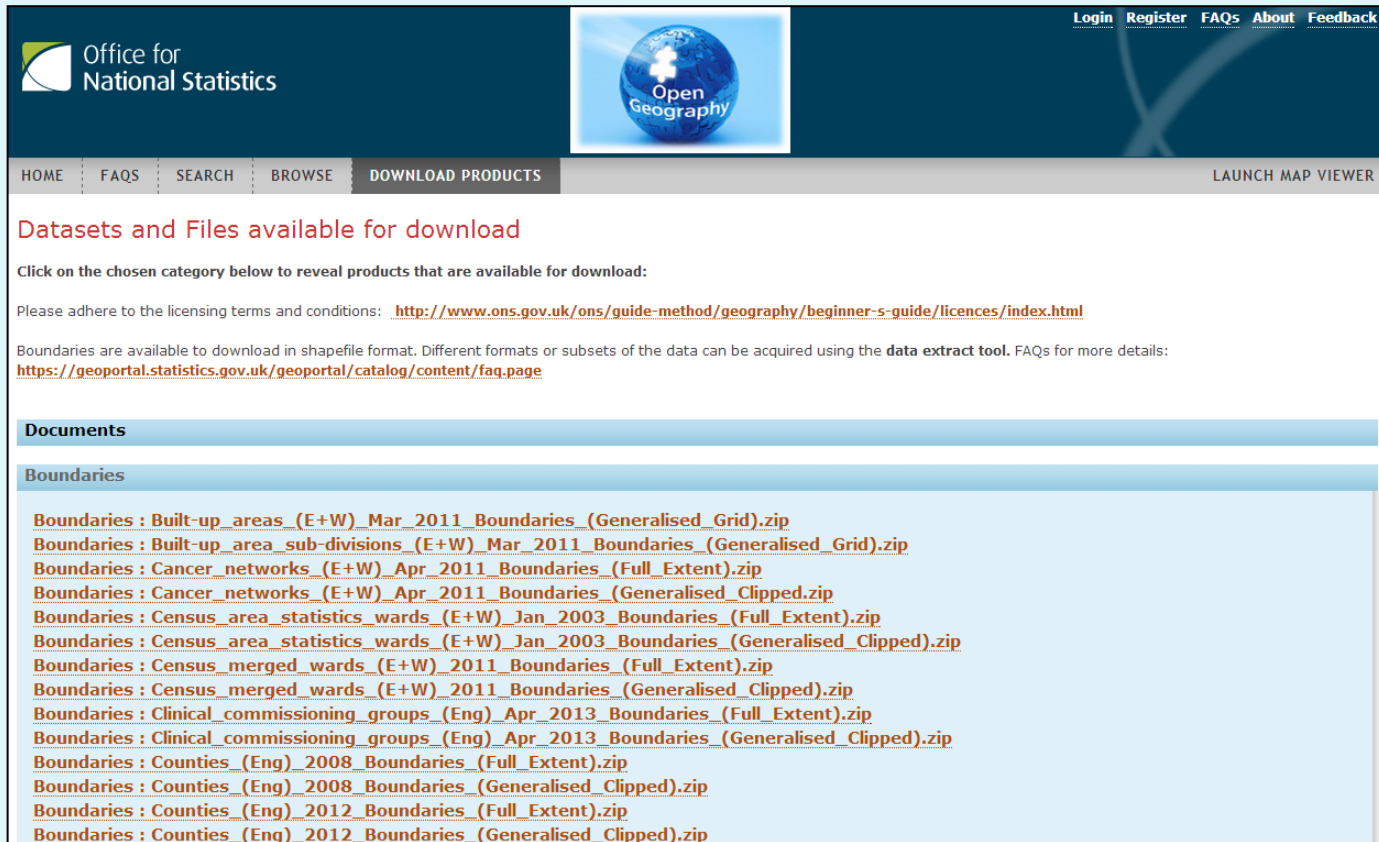
WHAT IS A SHAPEFILE?

- Widely used file type for saving geographical features
- Vector based:
 - Point
 - Polyline
 - Polygon



WHERE CAN I GET SHAPEFILES?

- Once again, ONS to the rescue: geoportal.statistics.gov.uk



The screenshot shows the Office for National Statistics (ONS) Geoportal website. The header includes the ONS logo, a globe icon with 'Open Geography' text, and navigation links: Login, Register, FAQs, About, Feedback. Below the header is a navigation bar with links: HOME, FAQs, SEARCH, BROWSE, DOWNLOAD PRODUCTS (highlighted), and LAUNCH MAP VIEWER.

Datasets and Files available for download

Click on the chosen category below to reveal products that are available for download:

Please adhere to the licensing terms and conditions: <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/licences/index.html>

Boundaries are available to download in shapefile format. Different formats or subsets of the data can be acquired using the **data extract tool**. FAQs for more details: <https://geoportal.statistics.gov.uk/geoportal/catalog/content/faq.page>

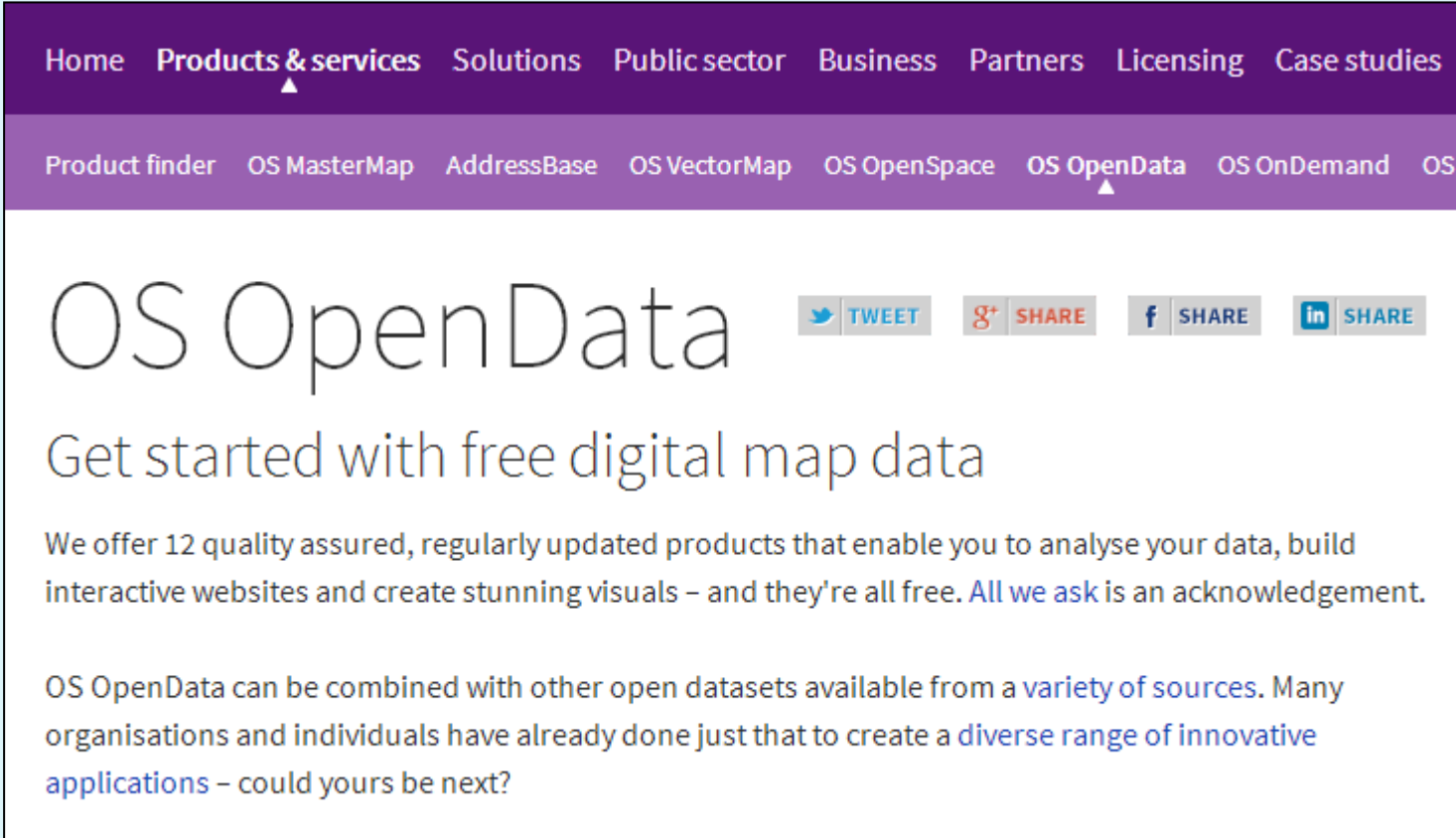
Documents

Boundaries

- [Boundaries : Built-up_areas_\(E+W\)_Mar_2011_Boundaries_\(Generalised_Grid\).zip](#)
- [Boundaries : Built-up_area_sub-divisions_\(E+W\)_Mar_2011_Boundaries_\(Generalised_Grid\).zip](#)
- [Boundaries : Cancer_networks_\(E+W\)_Apr_2011_Boundaries_\(Full_Extent\).zip](#)
- [Boundaries : Cancer_networks_\(E+W\)_Apr_2011_Boundaries_\(Generalised_Clipped\).zip](#)
- [Boundaries : Census_area_statistics_wards_\(E+W\)_Jan_2003_Boundaries_\(Full_Extent\).zip](#)
- [Boundaries : Census_area_statistics_wards_\(E+W\)_Jan_2003_Boundaries_\(Generalised_Clipped\).zip](#)
- [Boundaries : Census_merged_wards_\(E+W\)_2011_Boundaries_\(Full_Extent\).zip](#)
- [Boundaries : Census_merged_wards_\(E+W\)_2011_Boundaries_\(Generalised_Clipped\).zip](#)
- [Boundaries : Clinical_commissioning_groups_\(Eng\)_Apr_2013_Boundaries_\(Full_Extent\).zip](#)
- [Boundaries : Clinical_commissioning_groups_\(Eng\)_Apr_2013_Boundaries_\(Generalised_Clipped\).zip](#)
- [Boundaries : Counties_\(Eng\)_2008_Boundaries_\(Full_Extent\).zip](#)
- [Boundaries : Counties_\(Eng\)_2008_Boundaries_\(Generalised_Clipped\).zip](#)
- [Boundaries : Counties_\(Eng\)_2012_Boundaries_\(Full_Extent\).zip](#)
- [Boundaries : Counties_\(Eng\)_2012_Boundaries_\(Generalised_Clipped\).zip](#)

WHERE CAN I GET SHAPEFILES?

- Ordnance Survey: <http://www.ordnancesurvey.co.uk/business-and-government/products/opendata-products.html>



The screenshot shows the top of the Ordnance Survey OS OpenData website. It features a dark purple navigation bar with links: Home, Products & services (with a dropdown arrow), Solutions, Public sector, Business, Partners, Licensing, and Case studies. Below this is a lighter purple bar with links: Product finder, OS MasterMap, AddressBase, OS VectorMap, OS OpenSpace, OS OpenData (with a dropdown arrow), OS OnDemand, and OS. The main content area has a white background with the heading 'OS OpenData' in a large, thin font. To the right of the heading are four social media share buttons: Twitter (TWEET), Google+ (SHARE), Facebook (SHARE), and LinkedIn (SHARE). Below the heading is the subheading 'Get started with free digital map data'. The main text reads: 'We offer 12 quality assured, regularly updated products that enable you to analyse your data, build interactive websites and create stunning visuals – and they're all free. All we ask is an acknowledgement.' At the bottom, it says: 'OS OpenData can be combined with other open datasets available from a variety of sources. Many organisations and individuals have already done just that to create a diverse range of innovative applications – could yours be next?'

Home Products & services Solutions Public sector Business Partners Licensing Case studies

Product finder OS MasterMap AddressBase OS VectorMap OS OpenSpace OS OpenData OS OnDemand OS

OS OpenData

TWEET SHARE SHARE SHARE


Get started with free digital map data

We offer 12 quality assured, regularly updated products that enable you to analyse your data, build interactive websites and create stunning visuals – and they're all free. All we ask is an acknowledgement.

OS OpenData can be combined with other open datasets available from a variety of sources. Many organisations and individuals have already done just that to create a diverse range of innovative applications – could yours be next?

WHERE CAN I GET SHAPEFILES?

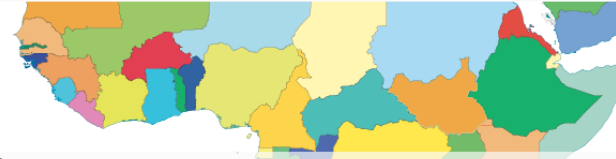
■ Natural Earth: naturalearthdata.com



Natural Earth

Free vector and raster map data at
1:10m, 1:50m, and 1:110m scales

[Home](#) [Features](#) [Downloads](#) [Blog](#) [Forums](#) [Corrections](#) [About](#)




Map Gallery

Admin 0 – Countries


Natural Earth is a public domain map dataset available at 1:10m, 1:50m, and 1:110 million scales. Featuring tightly integrated vector and raster data, with Natural Earth you can make a variety of visually pleasing, well-crafted maps with cartography or GIS software.

Natural Earth was built through a collaboration of many [volunteers](#) and is supported by [NACIS](#) (North American Cartographic Information Society), and is free for use in any type of project (see our [Terms of Use](#) page for more information).



Convenience

Natural Earth solves a problem: finding suitable data for making small-scale maps. In a time when the web is awash in geospatial data, cartographers are forced to waste time sifting through confusing tangles of poorly attributed data to make clean, legible maps. Because your time is valuable, Natural Earth data comes ready-to-use.



Neatness Counts

The carefully generalized linework maintains consistent, recognizable geographic shapes at 1:10m, 1:50m, and 1:110m scales. Natural Earth was built from the ground up so you will find that all data layers align precisely with one another. For example, where rivers and country borders are one and the same, the lines are coincident.

COUNTRYNAM	SCALERANK	FEATURECLA	SOVEREIGNTY
Alghanistan	1.000000000000	Countries	Alghanistan
Aland	3.000000000000	Countries	Finland
Albania	1.000000000000	Countries	Albania
Algeria	1.000000000000	Countries	Algeria

GIS Attributes

Natural Earth, however, is more than just a collection of pretty lines. The data attributes are equally important for mapmaking. Most data contain embedded feature names, which are ranked by relative importance. Other attributes facilitate faster map production, such as width attributes assigned to river segments for creating tapers.

WHERE CAN I GET SHAPEFILES?

■ Open Streetmap: openstreetmap.org

The screenshot displays the OpenStreetMap website interface. At the top, there are navigation links: "Edit", "History", "Export", "GPS Traces", "User Diaries", "Copyright", "Help", "About", "Log In", and "Sign Up". A search bar is located on the left with the text "Where am I?" and a "Go" button. Below the search bar, the "Export" panel is visible, showing a bounding box with coordinates: 51.51251, -0.08373, 51.50970, and -0.07874. A link "Manually select a different area" is provided. Below the coordinates, the "Licence" section states: "OpenStreetMap data is licensed under the [Open Data Commons Open Database License \(ODbL\)](#)". A blue "Export" button is present. Further down, a note says: "If the above export fails, please consider using one of the sources listed below:". Several alternative sources are listed with links: "Overpass API", "Planet OSM", "Geofabrik Downloads", "Metro Extracts", and "Other Sources". The main map area shows a street grid in London, with labels for streets like Lime Street, Fenchurch Avenue, Fenchurch Street, London Street, and Great Tower Street. Various landmarks and buildings are visible, including the Guild Church of Saint Margaret Patten, All Hallows Staining (defunct), and several restaurants like Pret A Manger and China Delight. The map includes a scale bar at the bottom left and a copyright notice at the bottom right: "© OpenStreetMap contributors" and "Make a Donation".

HOW DO I GET SHAPEFILES INTO R?

- The maptools package provides the function readShapeSpatial which allows easy importing of shapefiles:

```
readShapeSpatial("London_CCG.shp",  
IDvar="PCO12NM")
```

- If your shapefiles are really complex try simplifying them using gSimplify in the rgeos package.
- Simplifying shapefiles can give bad results so consider filtering to the bare minimum first!

GEOCODING

GEOCODING THE EASY WAY!

- Geocoding is the act of adding geographic information to data.
- In most cases this involves using postcodes to attach coordinates
- This can be a bit of a pain
 - Large postcode tables (2.5 million records)
 - Inconsistent postcode formats in different systems
 - Maintenance: Need to keep postcode tables up to date
- I don't have postcodes for the A&E departments!

GEOCODING THE EASY WAY!

- Luckily R (with a little help from Google) makes it easy if you only have a small amount of data to geocode!

- `library("ggmap")`

```
AAE$Address <- paste0(AAE$Name, ", LONDON, UK")
```

```
geocode(AAE$Address)
```

GEOCODING THE EASY WAY!

Address	lon	lat
OXLEAS NHS FOUNDATION TRUST,LONDON,UK	0.0654641	51.46894
HOUNSLOW AND RICHMOND COMMUNITY HEALTHCARE NHS TRUST,LONDON,UK	-0.1254872	51.50852
NORTH EAST LONDON NHS FOUNDATION TRUST,LONDON,UK	-0.1356832	51.52462
MOORFIELDS EYE HOSPITAL NHS FOUNDATION TRUST,LONDON,UK	-0.0894717	51.52754
ROYAL FREE LONDON NHS FOUNDATION TRUST,LONDON,UK	-0.1661535	51.55387
THE WHITTINGTON HOSPITAL NHS TRUST,LONDON,UK	-0.1254872	51.50852
EALING HOSPITAL NHS TRUST,LONDON,UK	-0.3465190	51.50740
LEWISHAM HEALTHCARE NHS TRUST,LONDON,UK	-0.0167340	51.45308
KINGSTON HOSPITAL NHS TRUST,LONDON,UK	-0.2820644	51.41422
CHELSEA AND WESTMINSTER HOSPITAL NHS FOUNDATION TRUST,LONDON,UK	-0.1824880	51.48471
UNIVERSITY COLLEGE LONDON HOSPITALS NHS FOUNDATION TRUST,LONDON,UK	-0.1356832	51.52462
HOMERTON UNIVERSITY HOSPITAL NHS FOUNDATION TRUST,LONDON,UK	-0.1356832	51.52462
CROYDON HEALTH SERVICES NHS TRUST,LONDON,UK	-0.1254872	51.50852
THE HILLINGDON HOSPITALS NHS FOUNDATION TRUST,LONDON,UK	-0.1356832	51.52462
WEST MIDDLESEX UNIVERSITY HOSPITAL NHS TRUST,LONDON,UK	-0.3261690	51.47408
EPSOM AND ST HELIER UNIVERSITY HOSPITALS NHS TRUST,LONDON,UK	-0.1254872	51.50852
ST GEORGE'S HEALTHCARE NHS TRUST,LONDON,UK	-0.1745438	51.51792
KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST,LONDON,UK	-0.0935120	51.46792
NORTH MIDDLESEX UNIVERSITY HOSPITAL NHS TRUST,LONDON,UK	-0.3261690	51.47408
BARNET AND CHASE FARM HOSPITALS NHS TRUST,LONDON,UK	-0.1254872	51.50852
GUY'S AND ST THOMAS' NHS FOUNDATION TRUST,LONDON,UK	-0.0867684	51.50346
CENTRAL LONDON COMMUNITY HEALTHCARE NHS TRUST,LONDON,UK	-0.1375024	51.49766
NORTH WEST LONDON HOSPITALS NHS TRUST,LONDON,UK	-0.0459010	51.55082
BARKING, HAVERING AND REDBRIDGE UNIVERSITY HOSPITALS NHS TRUST,LONDON,UK	-0.1254872	51.50852
SOUTH LONDON HEALTHCARE NHS TRUST,LONDON,UK	-0.0459010	51.55082
IMPERIAL COLLEGE HEALTHCARE NHS TRUST,LONDON,UK	-0.1745438	51.51792
BARTS HEALTH NHS TRUST,LONDON,UK	-0.1254872	51.50852

GGPLOT

PULLING THINGS TOGETHER IN GGPLOT

- Before you can use a shapefile with ggplot, the fortify command needs to be used:

```
fortify(msoa, region="MSOA04CD")
```

- This converts a shapefile to a dataframe
- This can be quite time consuming
 - keep the number of features as low as possible
 - simplify if it still takes a long time.

PULLING THINGS TOGETHER IN GGPLOT

- ggplot works using layers which allows fine tuned control of a lot of graphical features
- Shapefile polygons can be coloured based on data
- Because of ggplot's layering abilities, additional layers can be added:
 - CCG borders
 - A&E departments

PULLING THINGS TOGETHER IN GGPLOT

- `geom_polygon` function is used to plot shapefiles
- `coord_map` is used to set the projection of the map (mercator was used in this instance)
- `theme_bw` was used to minimise additional graphical elements. Those remaining were removed using ggplot's theme options:

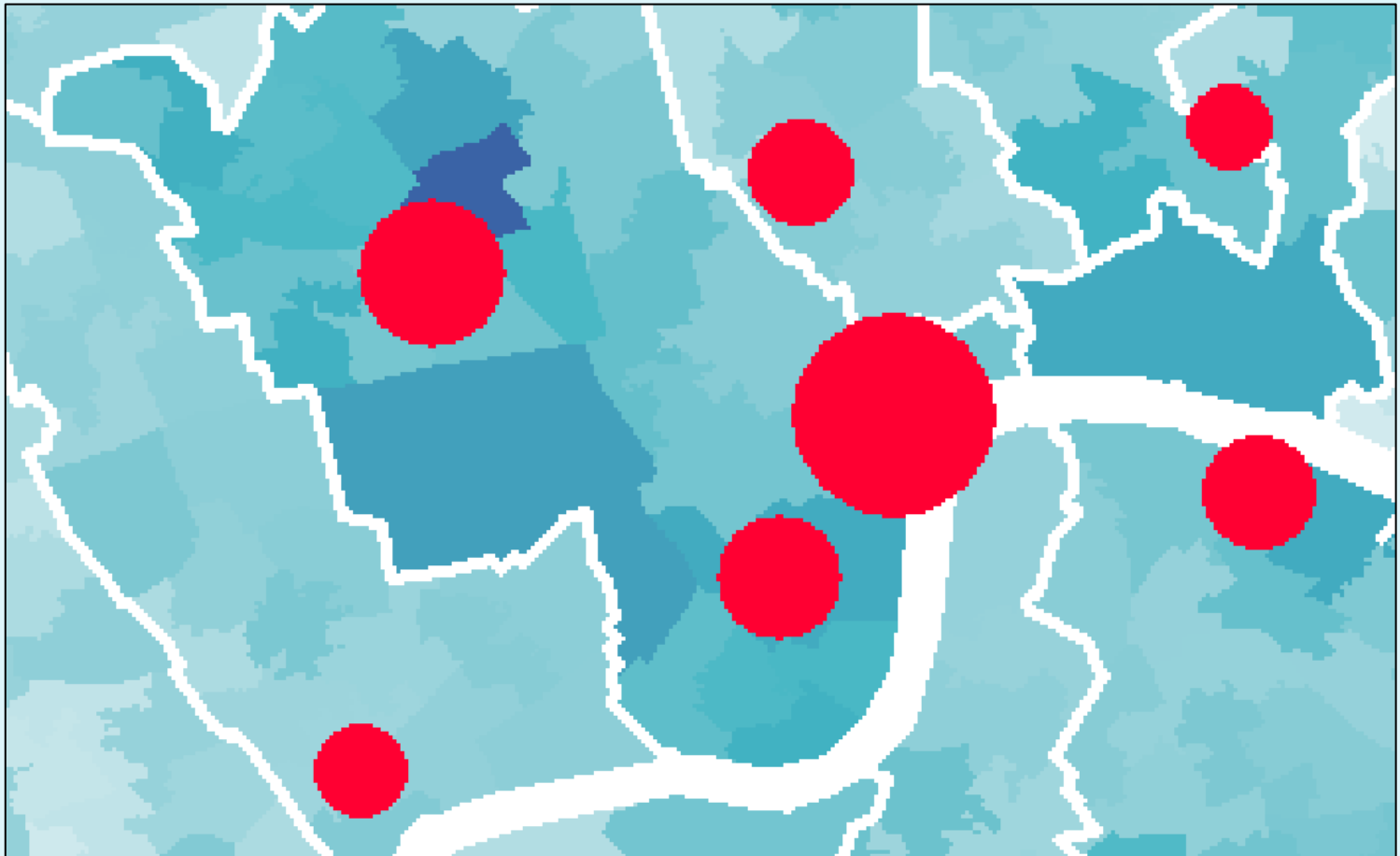
```
theme(  
  legend.position="bottom",  
  legend.title=element_blank(),  
  legend.key = element_rect(linetype= 0),  
  axis.text=element_blank(),  
  axis.title=element_blank(),  
  axis.ticks=element_blank(),  
  panel.border=element_blank(),  
  panel.grid=element_blank()  
)
```

IMPROVING OUTPUT QUALITY

- Important to use strokes in ggplot.
- Shapes surrounded by a stroke line give a far superior graphical finish.
- The Cairo package is also used to improve graphical output with anti-aliasing:
- `ggsave(file="Population.png", plot=p, width=200, height=200, units="mm", type="cairo-png")`

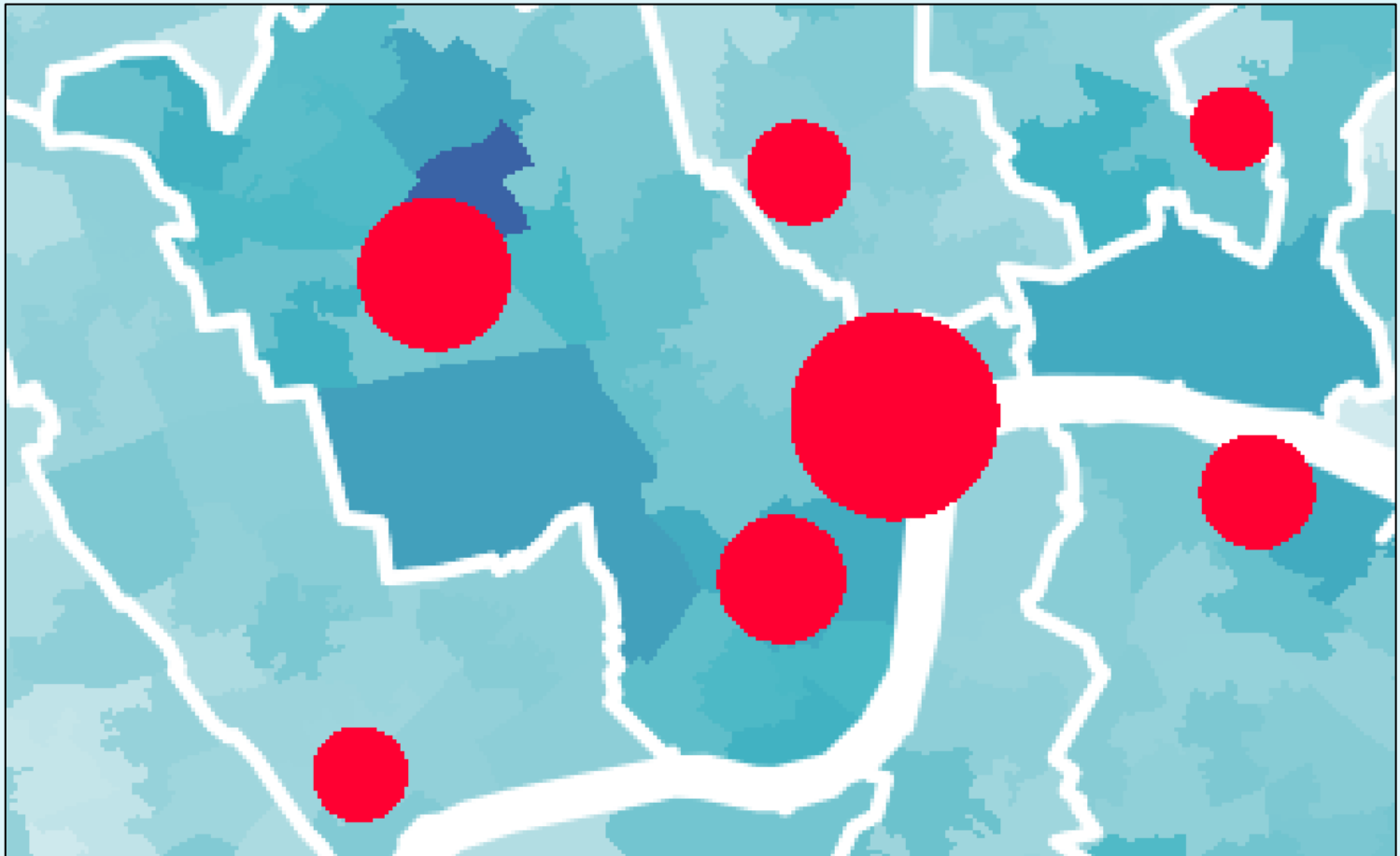
OUTPUT QUALITY

- Default output!



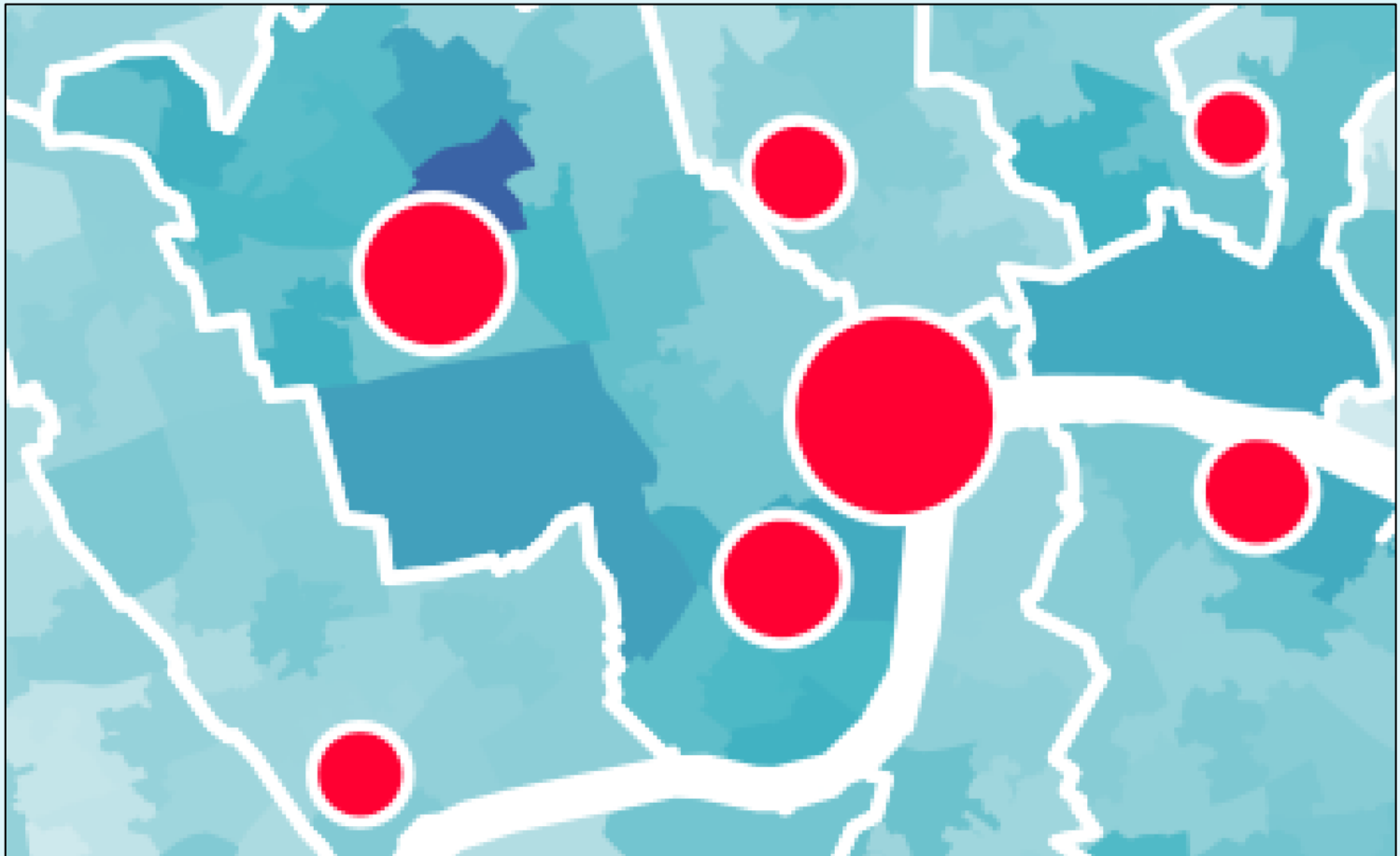
OUTPUT QUALITY

- With Cairo package:

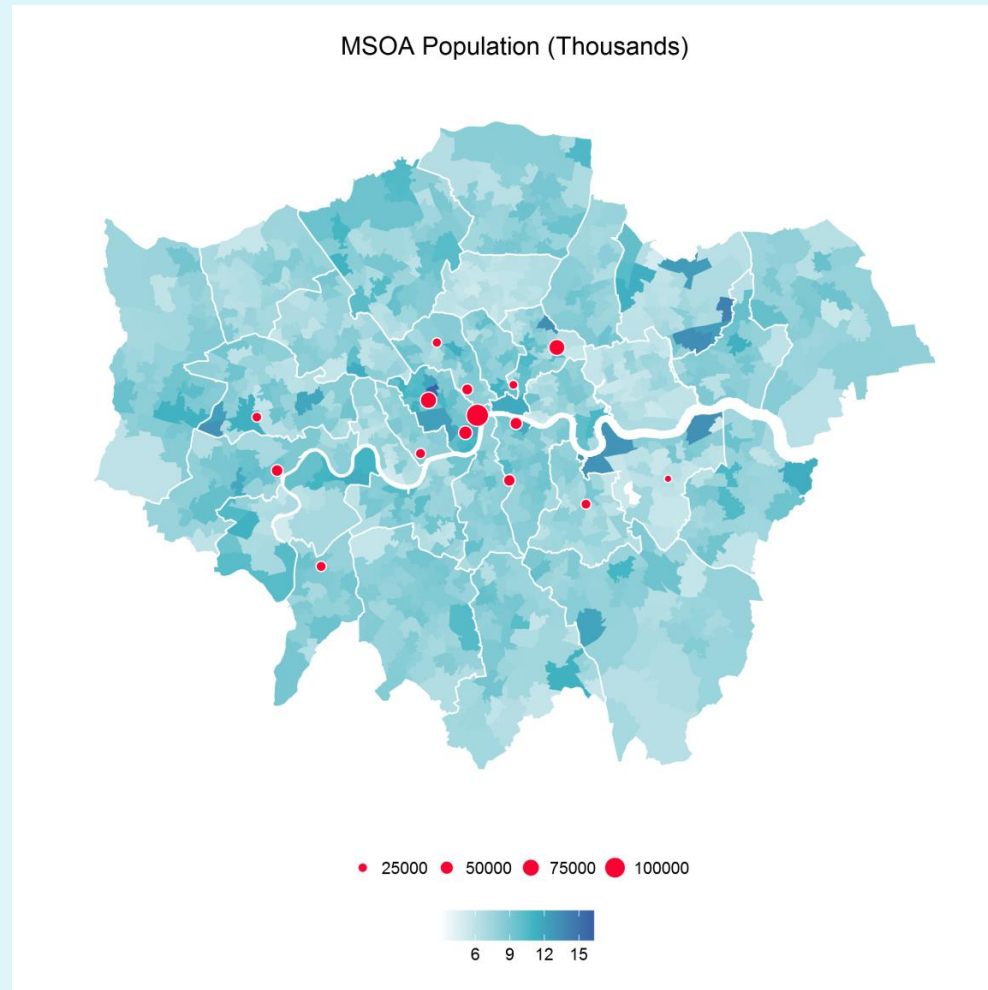


OUTPUT QUALITY

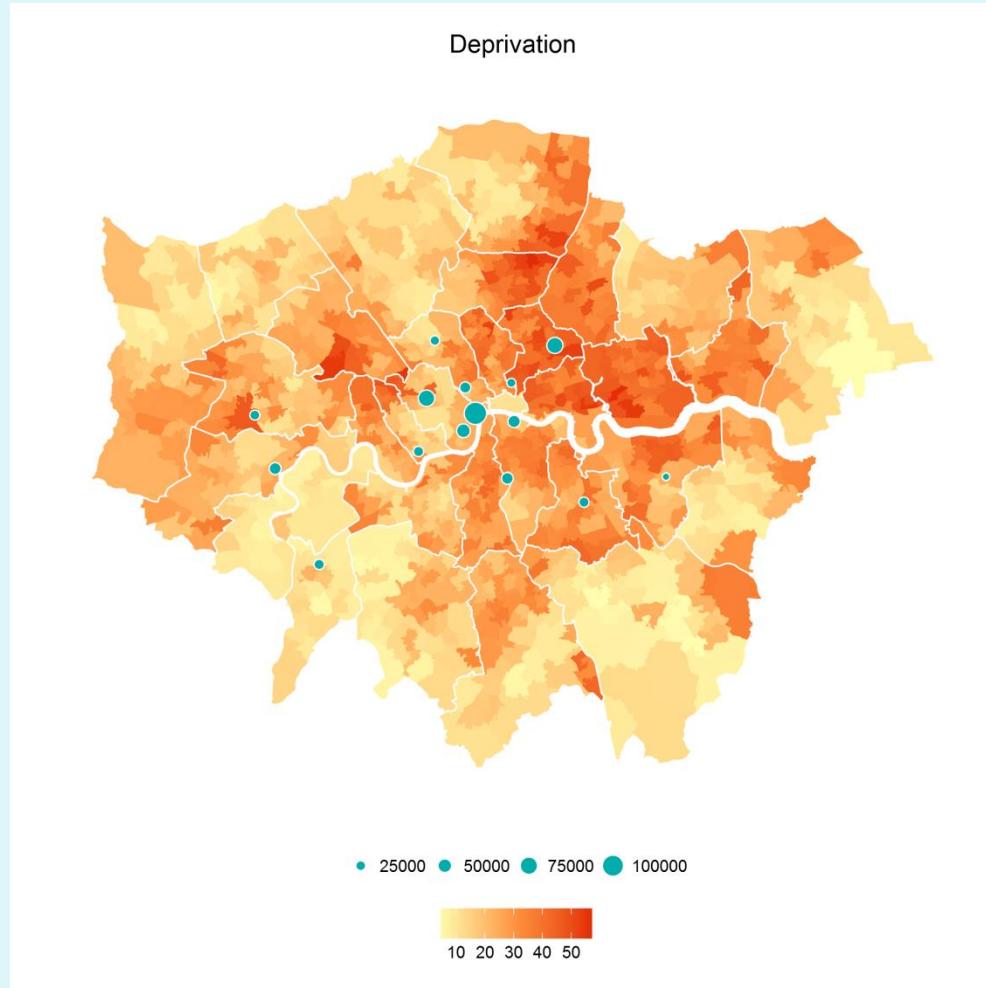
- With Cairo package and strokes added



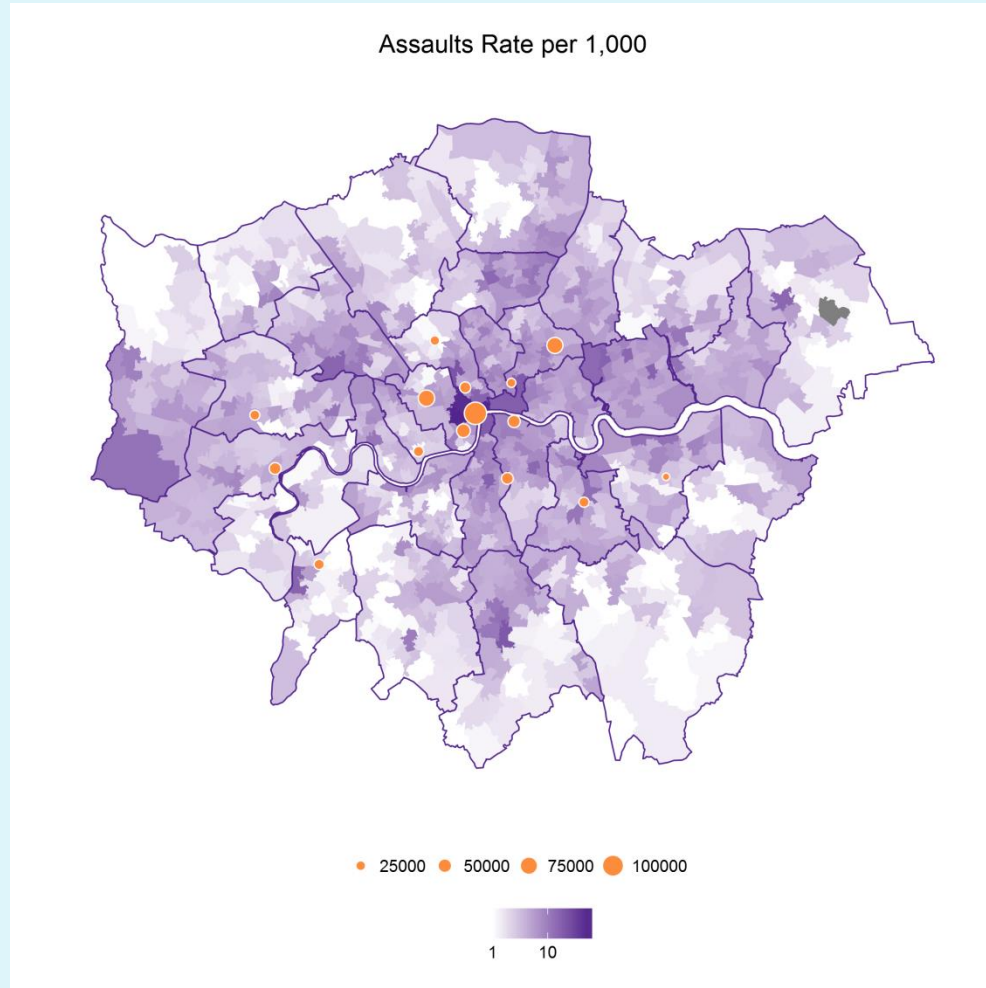
FINAL OUTPUT: POPULATION



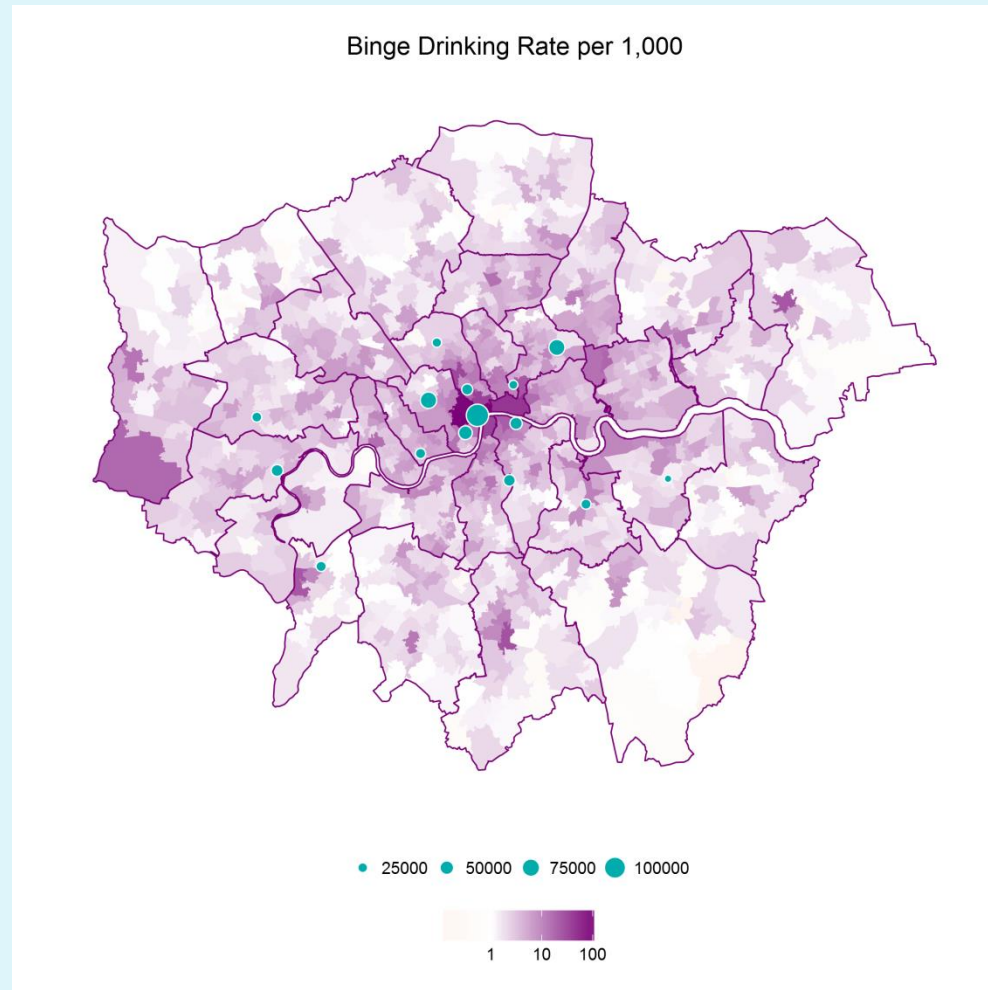
FINAL OUTPUT: DEPRIVATION



FINAL OUTPUT: ASSAULTS



FINAL OUTPUT: BINGE DRINKING



PROS & CONS

PROS & CONS

■ Pros

- The usual!
 - Re-usable → Automation
 - Shareable
 - Transparent code
 - Flexible
 - Precise control
 - Really nice output images!

■ Cons

- Labels and text formatting in general
- Processing time → Not as fast as specialised GIS packages
 - Although much of the processing only needs to be done as a one off
- Not as user friendly for single bits of analysis → QGIS still wins here

QUESTIONS