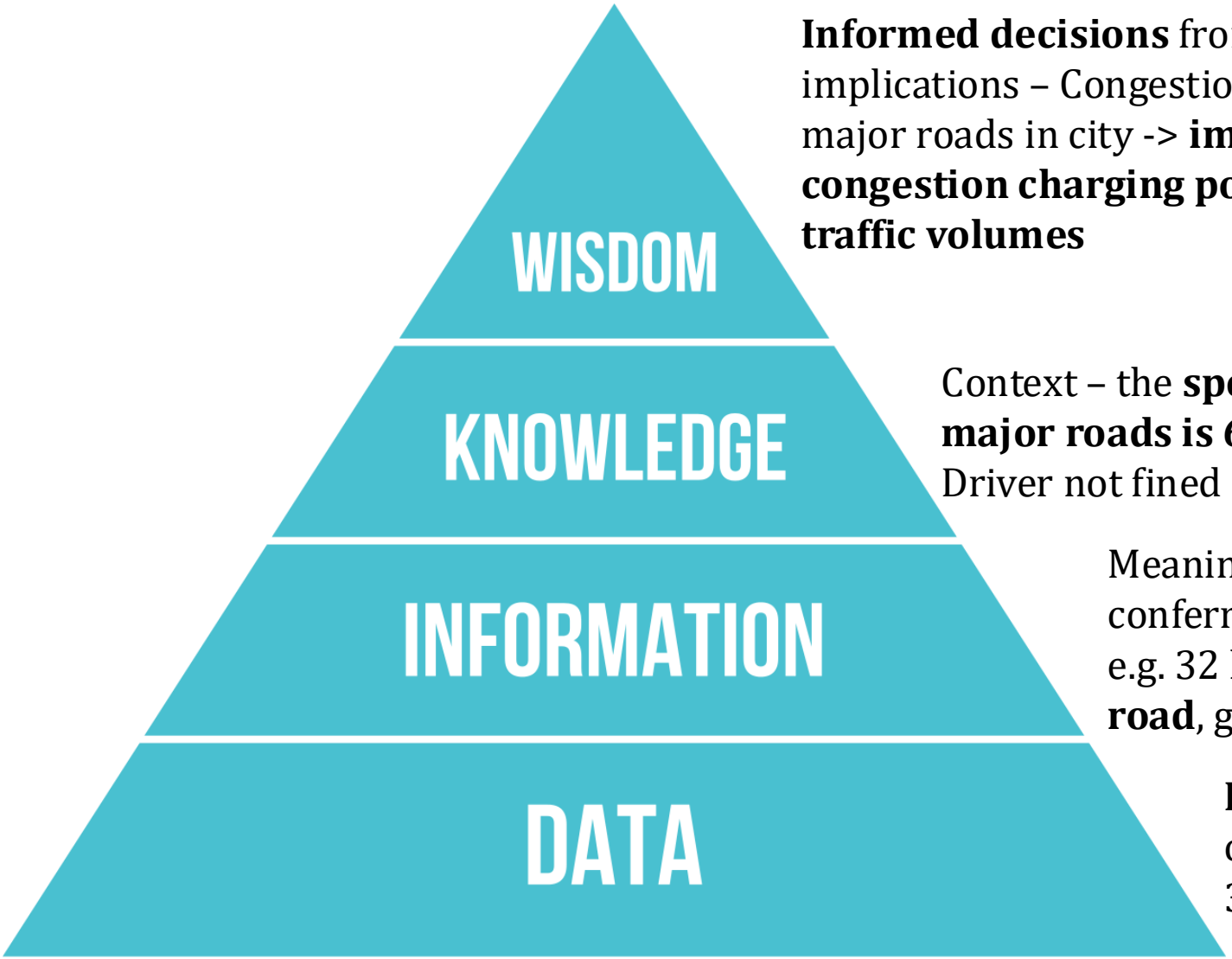
Geospatial Information Systems

Dr. Claire Dooley

(Credit to Adam Dennett & Andy MacLachlan for some content)

Geographic Information/ Spatial Data Science



Informed decisions from wider implications – Congestion is too high along major roads in city -> **implement a new congestion charging policy to reduce traffic volumes**

Context – the **speed limit for major roads is 60kph**. Congestion? Driver not fined for speeding?

Meaning and **meta-data** conferring additional info – e.g. 32 **kph**, red **= major road**, good = **health status**

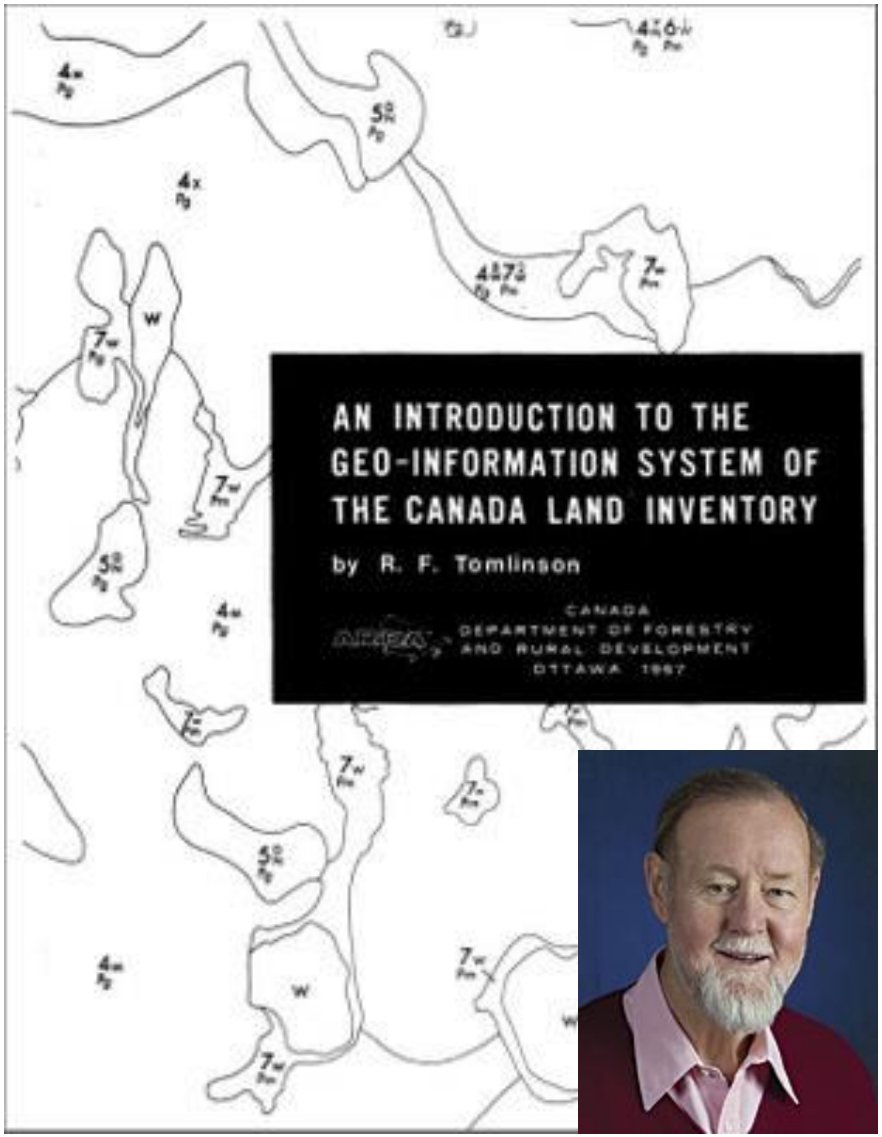
Raw numbers, lacks organisation – e.g. 32, 'red', 'good', etc.

What is a GIS?

- GIS traditionally means Geographic Information Systems
- Geographic Information Systems is different to Geographic Information Science
- We need to separate Geographic Information / Spatial Data ***SCIENCE*** from the tools we use to carry it out
- GI / SD Science is facilitated by the ***information systems*** (technologies – hardware, software + human interaction) we use to:
  - Store (input)
  - Manipulate / Process
  - Distribute
  - Analyse
  - Retrieve (output)

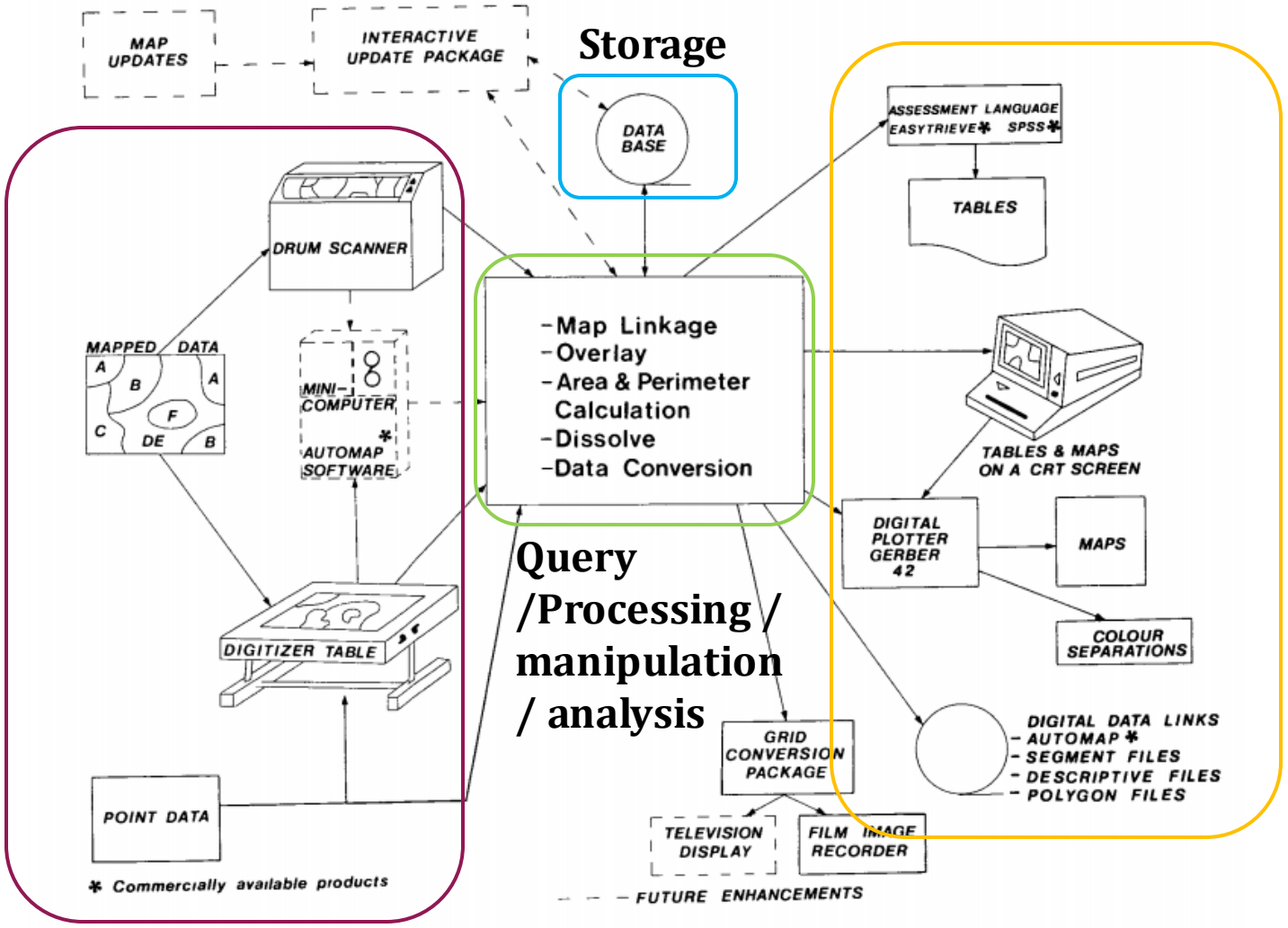
History of GIS

- Invented/developed here at UCL!!
- Roger Tomlinson – completed his PhD in UCL Geography in 1970s
- Credited with developing the **Canada Geographical Information System** (CGIS), the first computerised "geographical information system" in the world in 1960s.



https://www.ucl.ac.uk/made-at-ucl/stories/ucl-phd-student-changes-how-we-look-world

The Canada Geographical Information System



Data Capture / Input

Query /Processing / manipulation / analysis

Display /Output

https://cartogis.org/docs/proceedings/archive/auto-carto-4-vol-1/pdf/an-overview-of-the-canada-geographic-information-system(cgis).pdf

Data Capture / Input

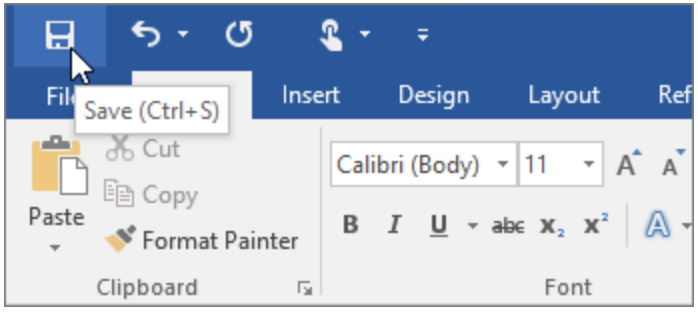
Then

Now

Two-day OpenStreetMap training with Sudanese refugees in Kampala in 2025. Participants explored different open source mapping tools, and learned about basic concepts Geographic Information Systems(GIS) concepts.

Data Storage

Then



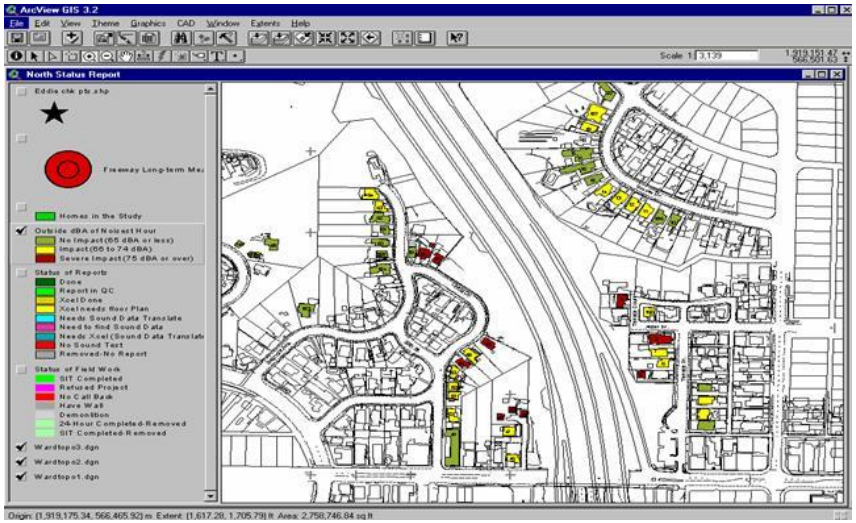
Now

GEOJSON

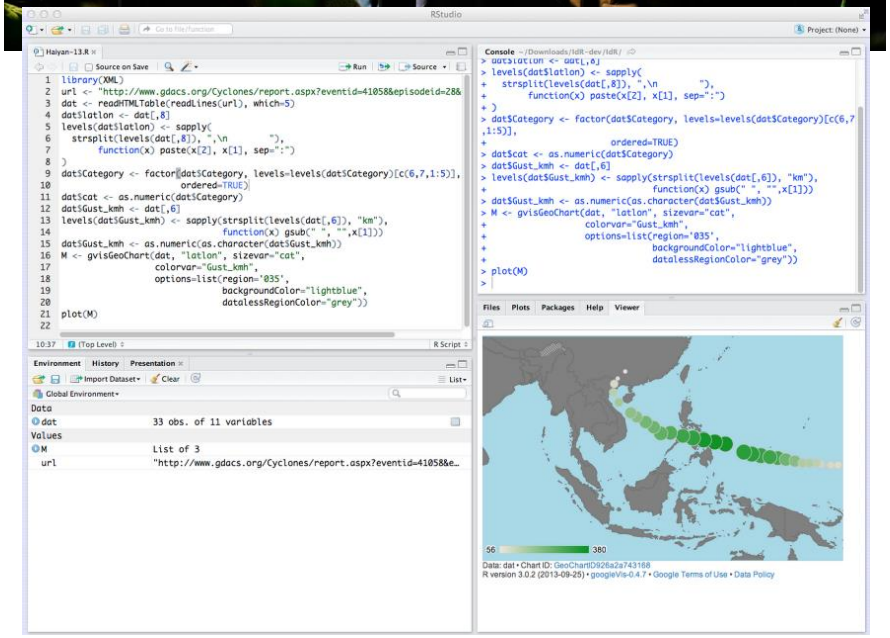
http://switchfromshapefile.org/

Query /Processing / Manipulation / Analysis

Then

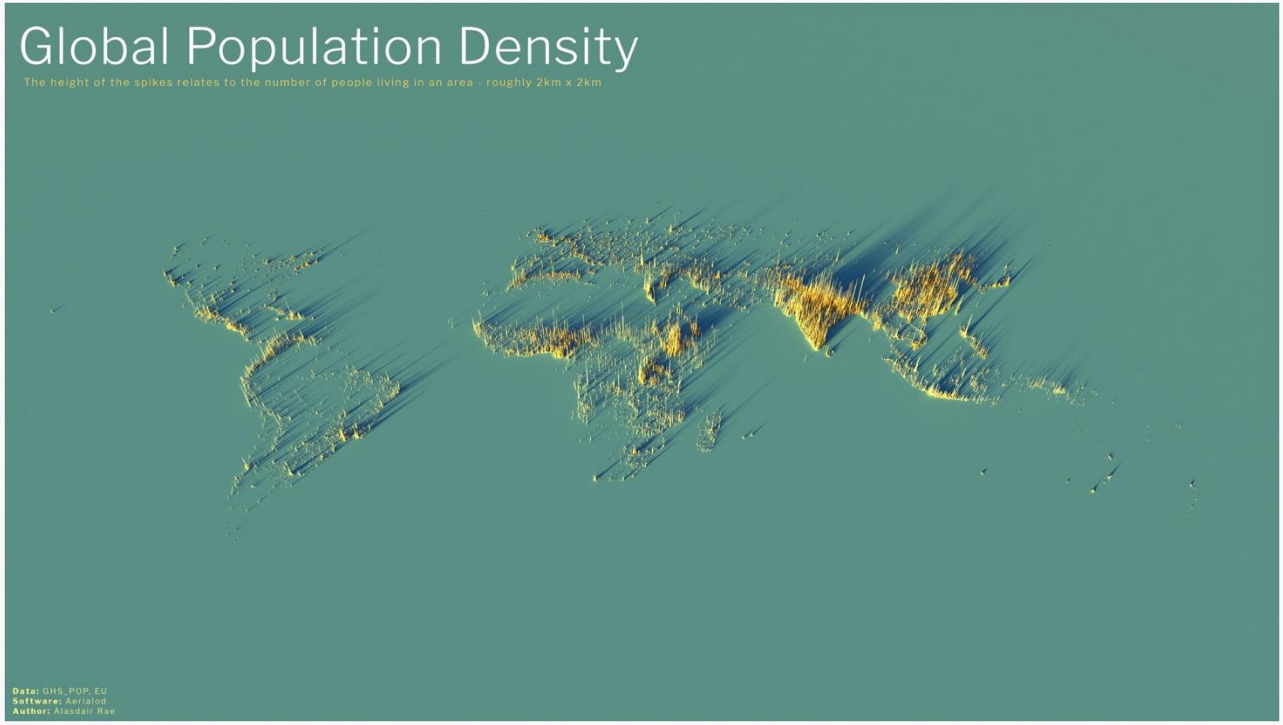


Now

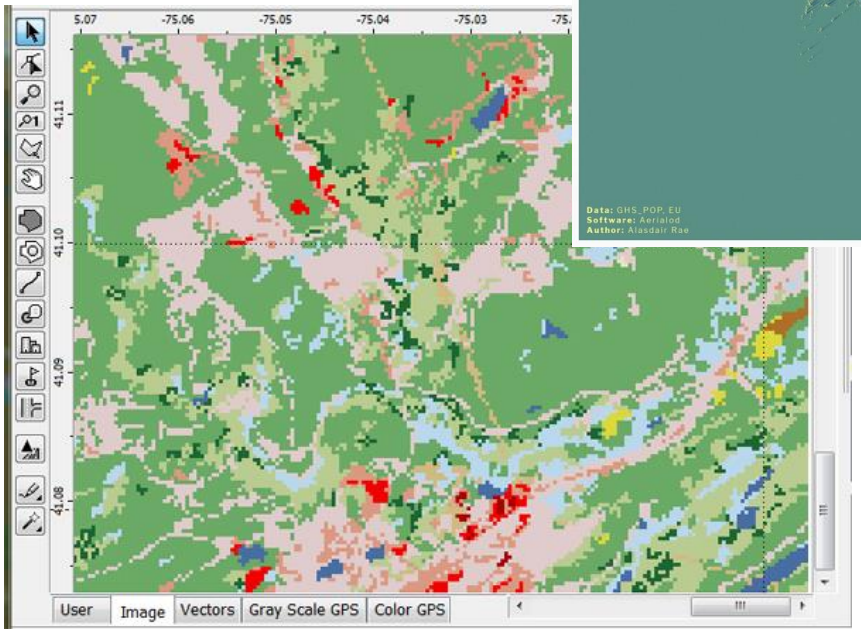


Data Output

Now



Then

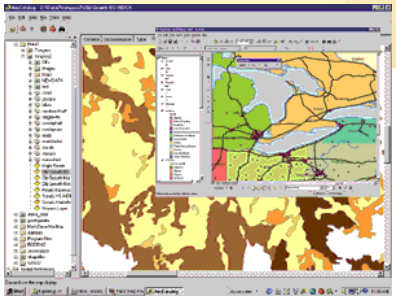


https://www.visualcapitalist.com/cp/3d-mapping-the-worlds-largest-population-densities/

Evolution of commercial GIS products

- From 1982, ESRI cornered the GIS market in terms of data storage, tools for Query /Processing / Manipulation / Analysis and Display / Visualisation
- The Graphical User Interface – GUI – allowed non-specialists to carry out GI Science for the first time

↑ Highlights of the evolution of Esri software from 1982 to present

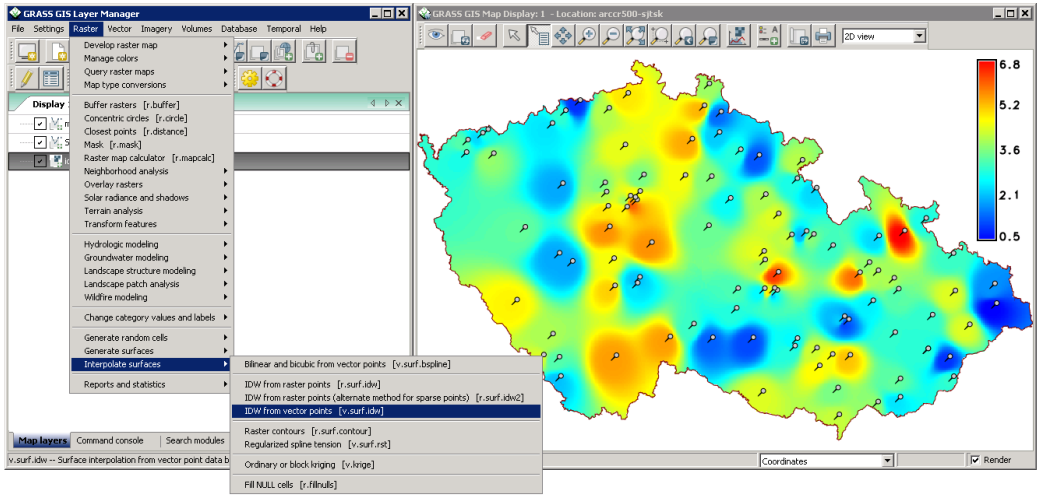


GUI Alternatives to Arc - GRASS

GRASS GIS

Bringing advanced geospatial technologies to the world.

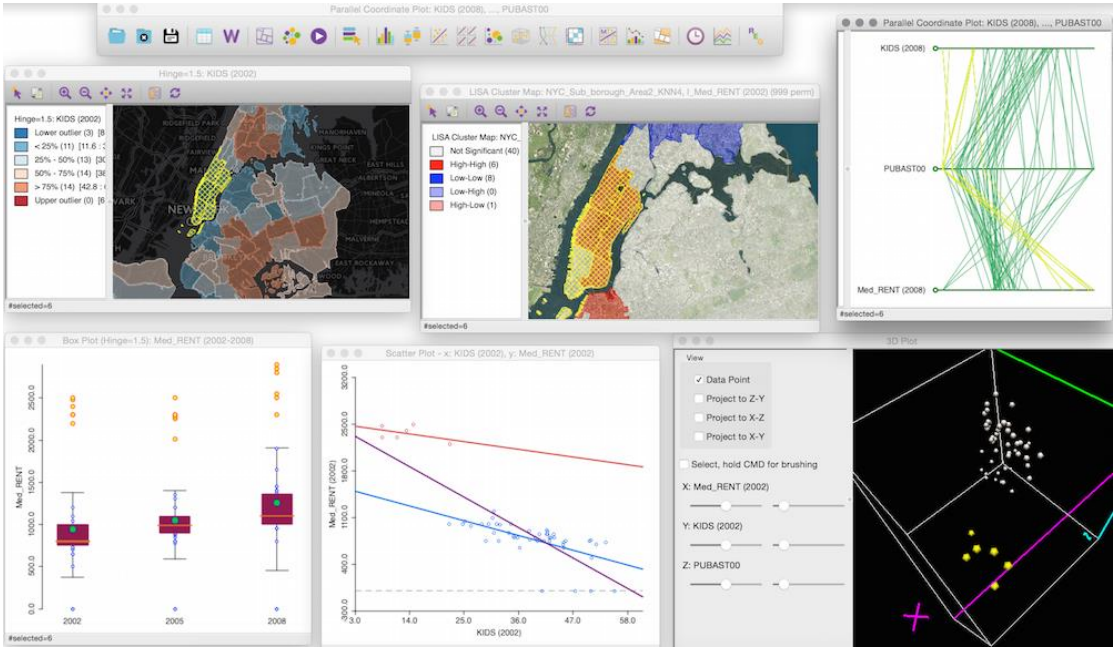
- Under continual development since 1982
- Free and Open Source
- Functions can be called from QGIS
- Not as fully functioned and slick as Arc



https://grass.osgeo.org/

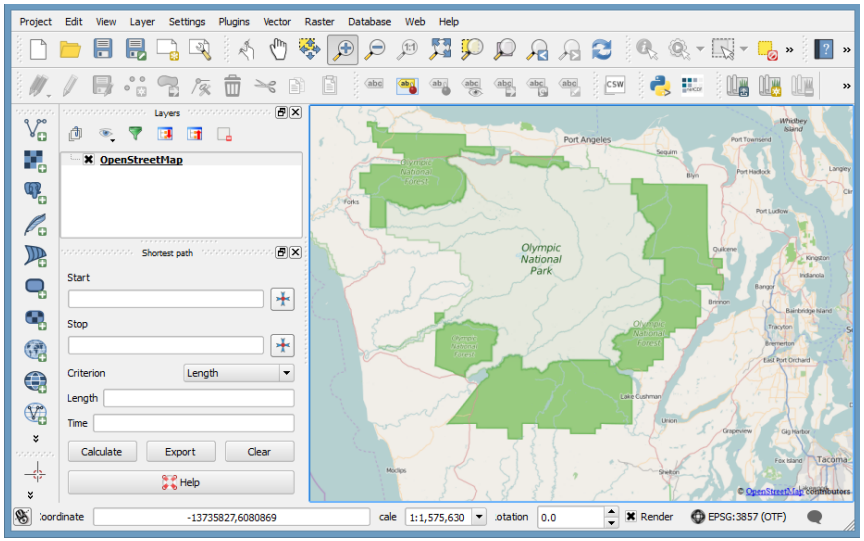
GUI Alternatives to Arc - GeoDa

- Developed at the University of Chicago since 2003
- Particularly good for spatial modelling and visualisation
- Free
- Many functions also in pysal python library



GUI Alternatives to Arc - QGIS

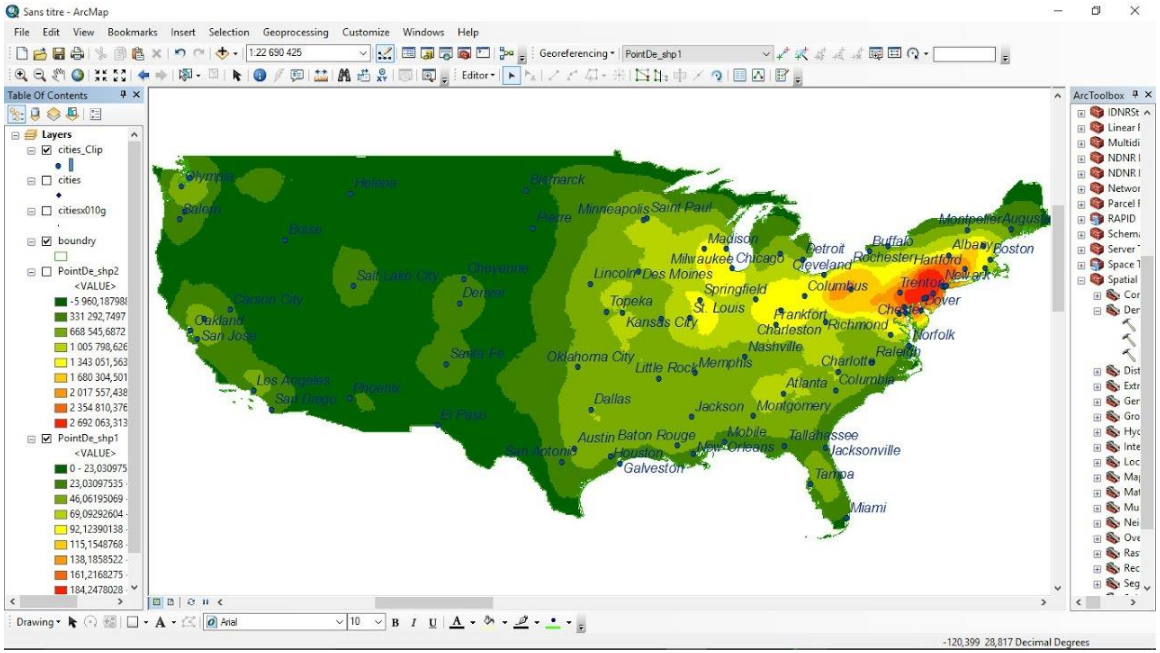
- Under development since 2002
- Free & Open Source – everything is on github!
- Probably best GUI GIS aside from Arc
- Connects to PostGIS database very effectively
- Slick maps with nice default features
- Large library of plugins for analysis



https://www.qgis.org/en/site/

https://github.com/qgis/QGIS

Why is Arc so successful?

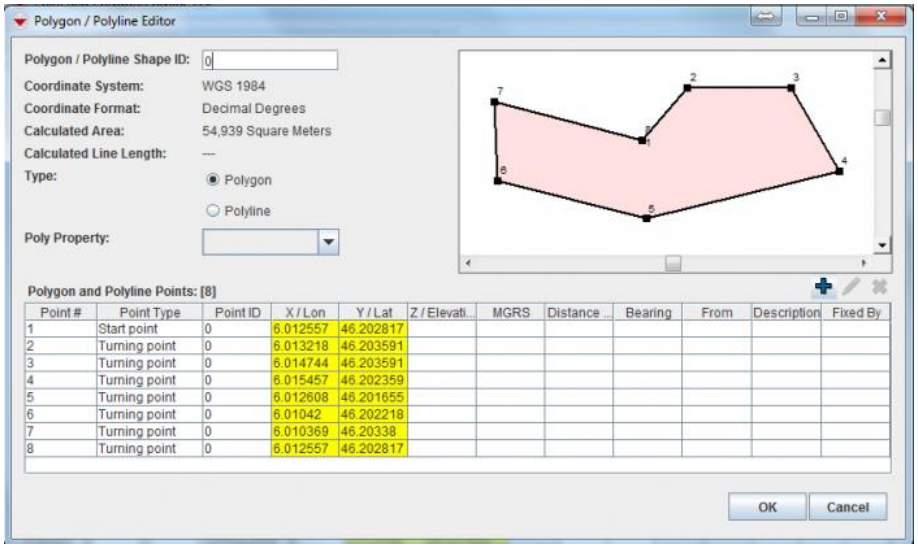


- User friendly
- Continual Development
- User Support
- Big suite of analysis functions / cartographic options + model builder for automation of these processes
- Educational licences
- The Shapefile

EXPENSIVE

Managing Geographic Information

- All vector data formats are essentially the same and comprised of two main elements:
  - **Geometries** - storing coordinates for either points, lines or polygons
  - **Attributes** – storing some other information about the geometries



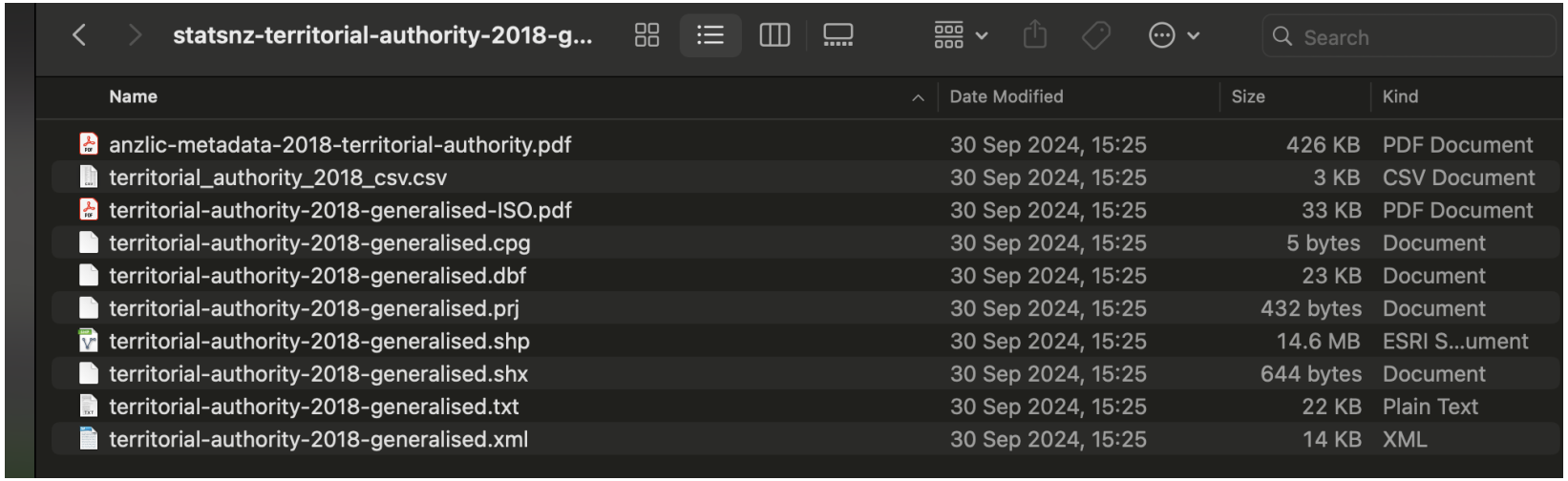
Point #	Point Type	Point ID	X / Lon	Y / Lat	Z / Elevati..	MGRS	Distance ..	Bearing	From	Description	Fixed By
1	Start point	0	6.012557	46.202817							
2	Turning point	0	6.013218	46.203591							
3	Turning point	0	6.014744	46.203591							
4	Turning point	0	6.015457	46.202359							
5	Turning point	0	6.012608	46.201655							
6	Turning point	0	6.01042	46.202218							
7	Turning point	0	6.010369	46.20338							
8	Turning point	0	6.012557	46.202817							

Data Output | Explain | Messages | Notifications

	name character varying	st_astext text
1	Point	POINT(0 0)
2	Linestring	LINESTRING(0 0,1 1,2 1,2 2)
3	Polygon	POLYGON((0 0,1 0,1 1,0 1,0 0))
4	PolygonWithHole	POLYGON((0 0,10 0,10 10,0 10,0 0),(1 1,1 2,2 2,2 1,1 1))
5	Collection	GEOMETRYCOLLECTION(POINT(2 0),POLYGON((0 0,1 0,1 1,0 1,0 0)))

Managing Geographic Information – the shapefile

- Become the de facto spatial data format for storing and sharing (latterly, online) vector data
- Collection of files –
  - .shp - Geometries
  - .shx - Index
  - .dbf – Attributes
  - + some others!



shapefiles

The Good

- Shapefile is by far the most widely supported format in existing software packages.
- For many use cases, it is *good enough*.
  - Index files (e.g. *.shx) enable good reading performance.
  - It is relatively efficient in terms of file size. The resulting file, even un-zipped, is relatively small compared to some other (mostly text-based) formats.

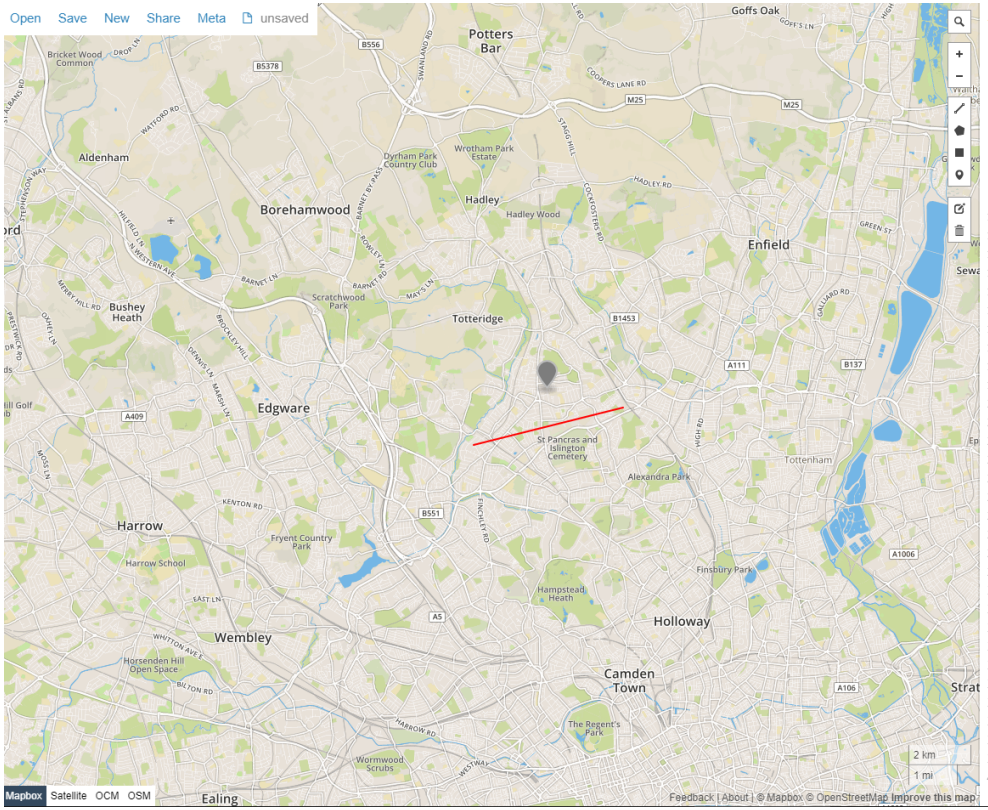
The Bad

- It's a multifile format.
- Attribute names are limited to 10 characters.
- Only 255 attributes. The DBF file does not allow you to store more then 255 attribute fields.
- Limited data types. Data types are limited to float, integer, date and text with a maximum 254 characters.
- It's limited to 2GB of file size. Although some tools are able to surpass this limit, they can never exceed 4GB of data.
- Single geometry type per file. There is no way to save mixed geometry features.

Managing Geographic Information – Alternatives to .shp: GeoJSON

- GeoJSON (mark-up similar to XML)
- Simple – particularly good for web applications
- Doesn't support every data type – e.g. dates

http://geojson.io/

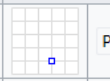
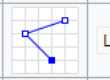
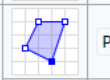
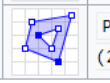


Managing Geographic Information - Standards

- The Open Geospatial Consortium
- Governing standards around spatial data since 1992
- Defined the 'well-known-text' (WKT) format for representing vector geometries

OGC®
Making location count.

Geometry primitives (2D)

Type	Examples	
Point		POINT (30 10)
LineString		LINESTRING (30 10, 10 30, 40 40)
Polygon		POLYGON ((30 10, 40 40, 20 40, 10 20, 30 10))
		POLYGON ((35 10, 45 45, 15 40, 10 20, 35 10), (20 30, 35 35, 30 20, 20 30))

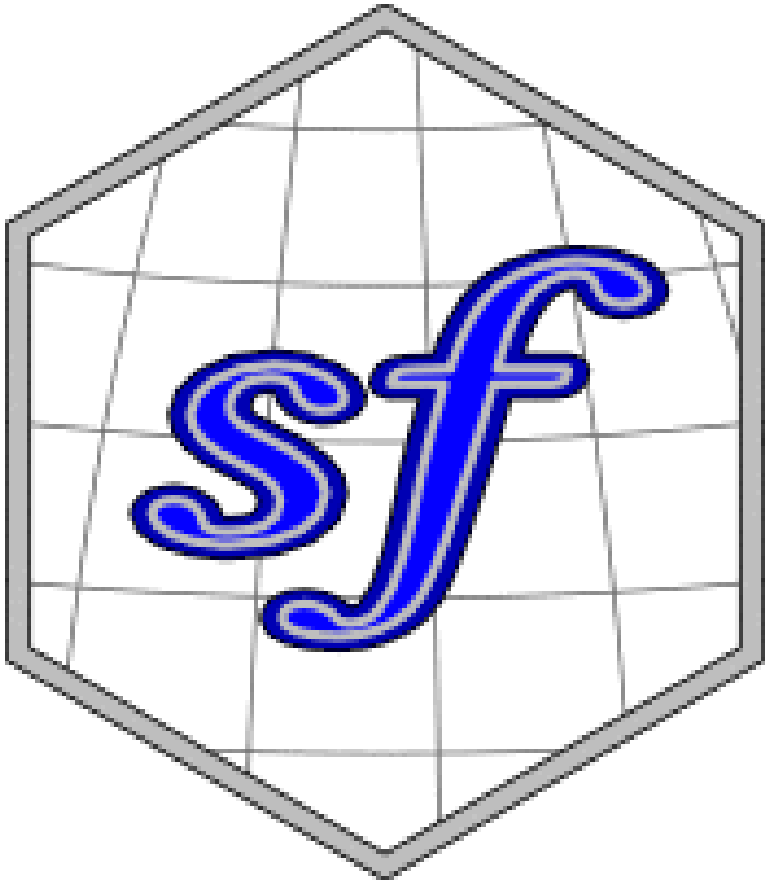
Multipart geometries (2D)

Type	Examples	
MultiPoint		MULTIPOINT ((10 40), (40 30), (20 20), (30 10))
		MULTIPOINT (10 40, 40 30, 20 20, 30 10)
MultiLineString		MULTILINESTRING ((10 10, 20 20, 10 40), (40 40, 30 30, 40 20, 30 10))
MultiPolygon		MULTIPOLYGON (((30 20, 45 40, 10 40, 30 20)), ((15 5, 40 10, 10 20, 5 10, 15 5)))
		MULTIPOLYGON (((40 40, 20 45, 45 30, 40 40)), ((20 35, 10 30, 10 10, 30 5, 45 20, 20 35), (30 20, 20 15, 20 25, 30 20)))
GeometryCollection		GEOMETRYCOLLECTION (POINT (40 10), LINESTRING (10 10, 20 20, 10 40), POLYGON ((40 40, 20 45, 45 30, 40 40)))

Managing Geographic Information - Standards

Geometry Type	WKT Example	Description
Point	POINT(30 10)	A single point with x coordinate 30 and y coordinate 10
LineString	LINESTRING(30 10, 10 30, 40 40)	A line connecting three points with x and y coordinates (30 10), (10 30), and (40 40)
Polygon	POLYGON((30 10, 40 40, 20 40, 10 20, 30 10))	A polygon with five vertices and an interior ring with x and y coordinates (30 10), (40 40), (20 40), (10 20), and (30 10)
MultiPoint	MULTIPOINT((10 40), (40 30), (20 20), (30 10))	A collection of four points with x and y coordinates (10 40), (40 30), (20 20), and (30 10)
MultiLineString	MULTILINESTRING((10 10, 20 20, 10 40), (40 40, 30 30, 40 20, 30 10))	A collection of two line strings, each connecting multiple points
MultiPolygon	MULTIPOLYGON(((30 20, 45 40, 10 40, 30 20)),((15 5, 40 10, 10 20, 5 10, 15 5)))	A collection of two polygons
GeometryCollection	GEOMETRYCOLLECTION(POINT(10 40), LINESTRING(30 10, 10 30, 40 40), POLYGON((30 10, 40 40, 20 40, 10 20, 30 10)))	A collection of point, linestring, and polygon geometries
Point ZM	POINT ZM (1 2 3 4)	A point with x, y, z, and m values of 1, 2, 3, and 4, respectively
Point M	POINT M (1 2 3)	A point with x, y, and m values of 1, 2, and 3, respectively

Managing Geographic Information – Alternatives to .shp: Simple Features



Simple Features - OGC standard that specifies a **common storage and access model** for 2D geometries.

Simple Features implemented in:

- The 'sf' package in R
- MySQL Spatial Extensions
- PostGIS extension for PostgreSQL
- SpatiaLite extension for SQLite
- Oracle Spatial
- Microsoft SQL Server since version 2008, with significant additions in the 2012 version.

Managing Geographic Information – Alternatives to .shp: others

- GeoPackage
  - SQLite Database file/container - .gpkg
  - The SQLite - open source and multi-platform
  - SQLite stores data in a single file

https://www.geopackage.org/

- PostGIS
  - Free and Open Source
  - Powers some of the world's most famous platforms (Facebook, Instagram, OpenStreetMap etc.)
  - Perfect if you want to do some serious heavy-lifting big-data processing & spatial queries
  - Difficult to master!

Managing Geographic Information – sf in R

- The Simple Features (sf) package facilitates the storage, access and management of geometric objects stored as simple features, in R
- sf arrived in around 2018
- sf objects are data tables with a geometry column containing WKT geometries at the end


```
## 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)

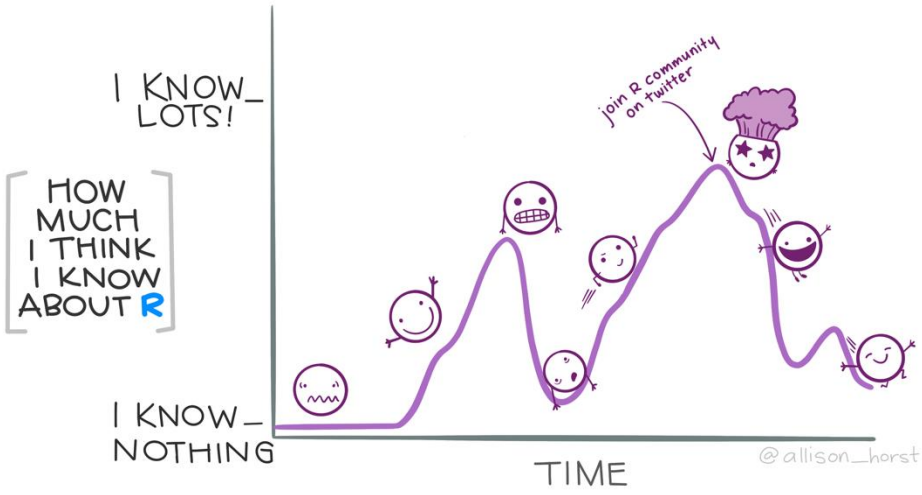
Simple feature geometry (sfg)

https://r-spatial.github.io/sf/index.html

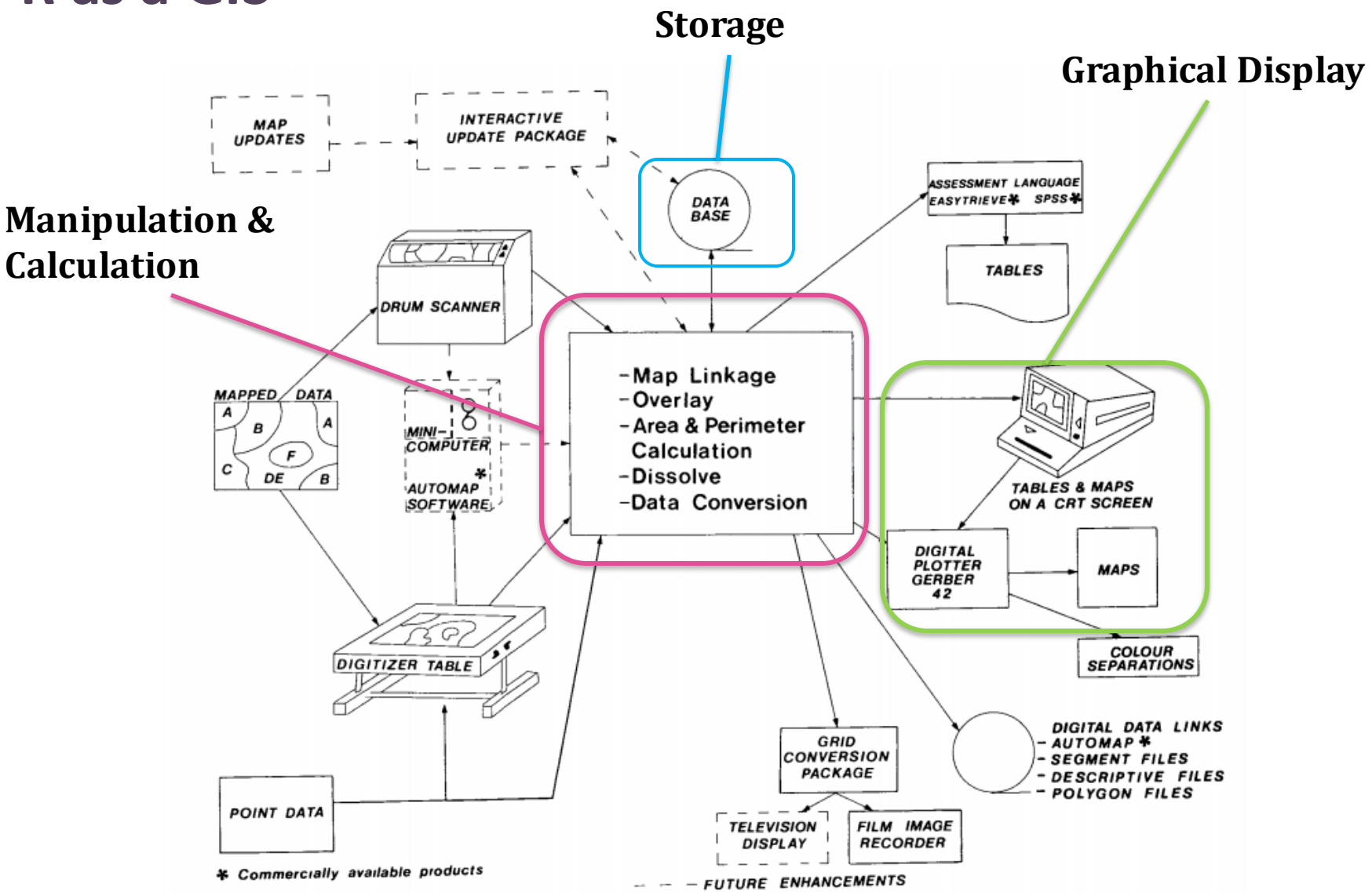
R as a GIS

https://www.r-project.org

- "R is a language and environment for statistical computing and graphics."
- "R is an integrated suite of software facilities for **data manipulation, calculation and graphical display**"



R as a GIS

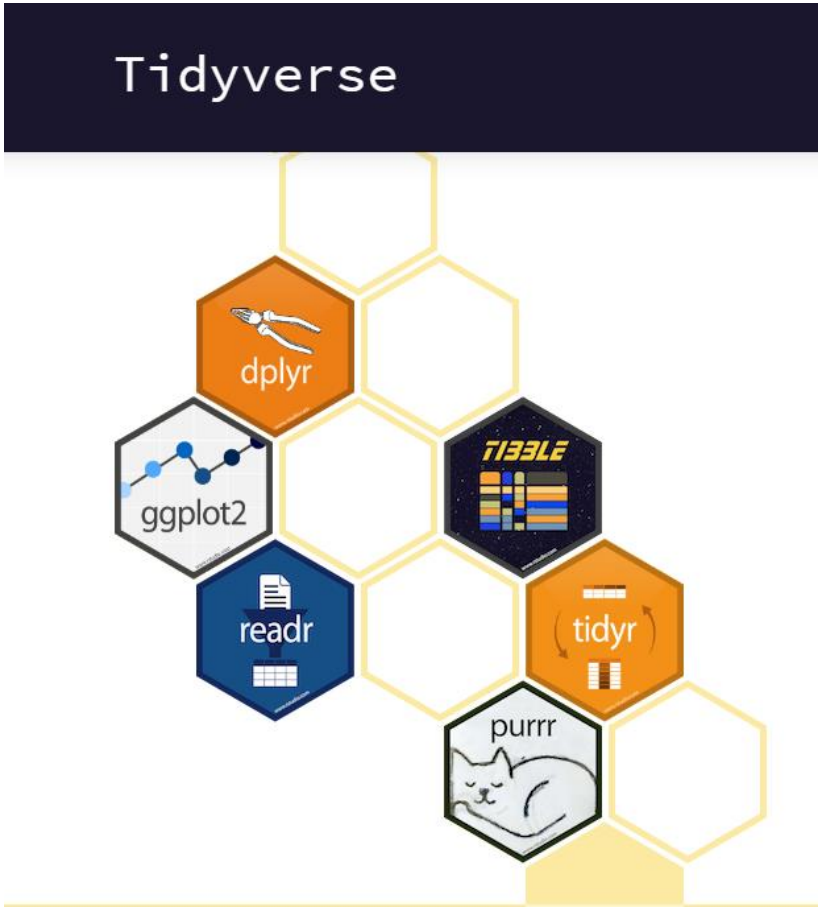

Storage
Graphical Display
Manipulation & Calculation
MAP UPDATES
INTERACTIVE UPDATE PACKAGE
DATA BASE
ASSESSMENT LANGUAGE
EASYTRIEVE* SPSS*
TABLES
DRUM SCANNER
MAPPED DATA
A
B
A
C
F
DE
B
MINI-COMPUTER
AUTOMAP SOFTWARE
-Map Linkage
-Overlay
-Area & Perimeter Calculation
-Dissolve
-Data Conversion
TABLES & MAPS ON A CRT SCREEN
DIGITAL PLOTTER GERBER 42
MAPS
COLOUR SEPARATIONS
DIGITIZER TABLE
GRID CONVERSION PACKAGE
DIGITAL DATA LINKS
- AUTOMAP *
- SEGMENT FILES
- DESCRIPTIVE FILES
- POLYGON FILES
POINT DATA
TELEVISION DISPLAY
FILM IMAGE RECORDER
* Commercially available products
FUTURE ENHANCEMENTS

R as a GIS – data management

- Data management facilitated through connection to a range of file formats (almost any you can think of)
- Plus a huge host of software packages for reading, writing and converting data held within these files into a format R can handle

R as a GIS – data manipulation & processing

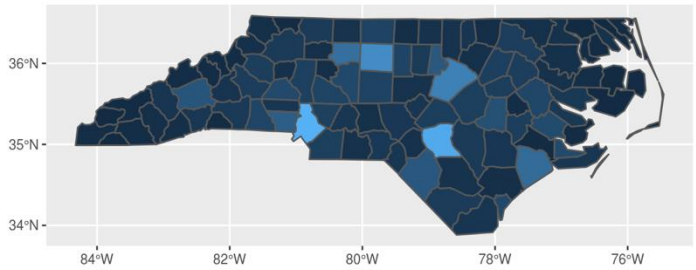
- R has always had packages for manipulating, processing and visualising data
- Packages are pieces of software designed for specific tasks within the R environment
- Recently, some of the best have been brought together under the 'tidyverse' banner
- "The tidyverse is a collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures."
- Tidyverse packages cover much of the data manipulation, processing and visualisation capabilities required of a GIS



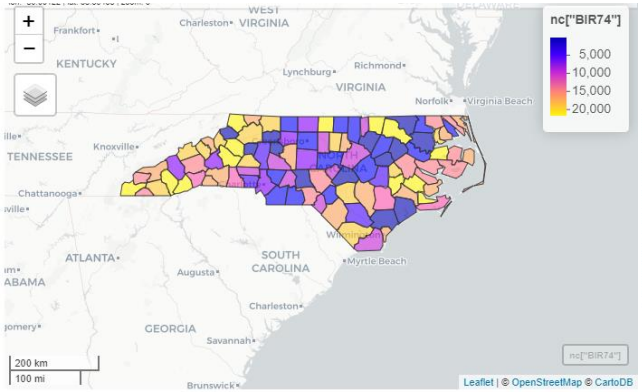
https://www.tidyverse.org

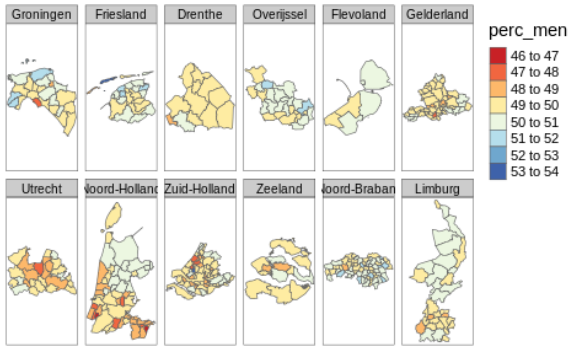
R as a GIS – data visualisation, display and output

- R has a variety of packages that facilitate visualisation of geometric data
- Packages like **ggplot2** and **tmap** follow the tidyverse principles



ggplot2



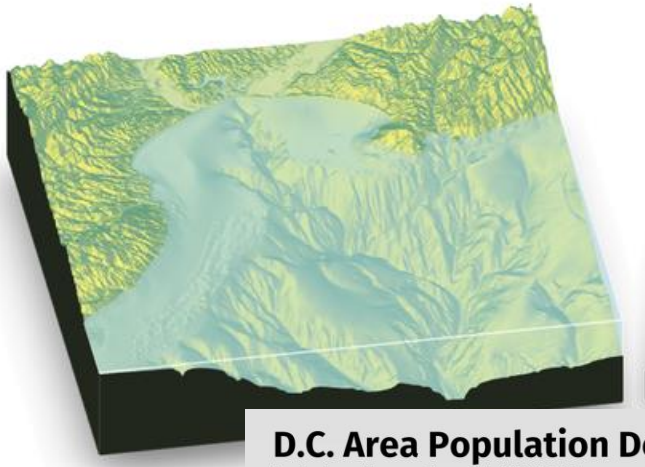


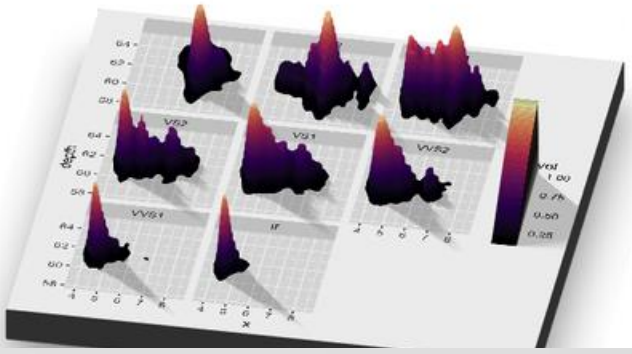
tmap

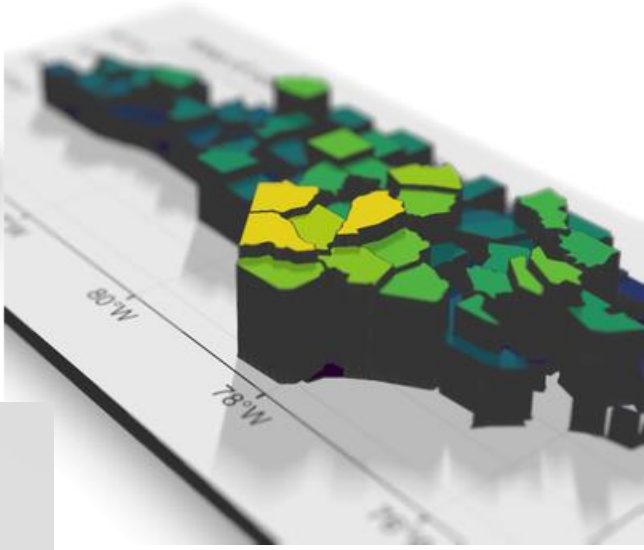
Leaflet

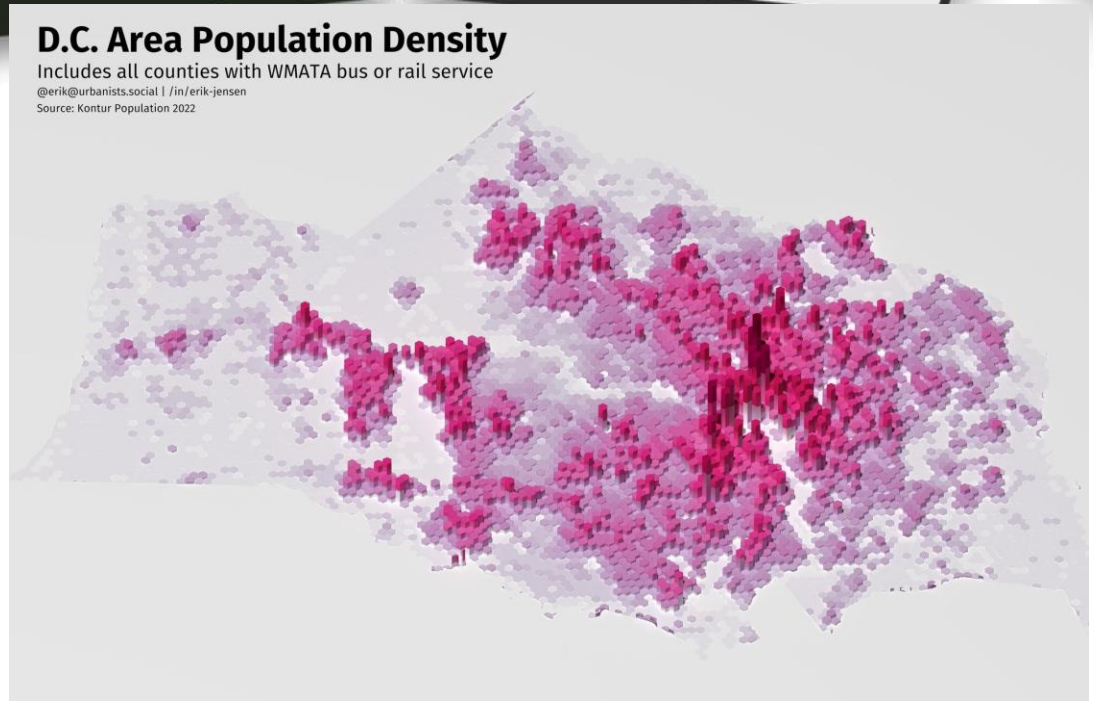
https://r-spatial.github.io/sf/articles/sf5.html

R as a GIS – data visualisation, display and output









R as a GIS – Dissemination and Sharing

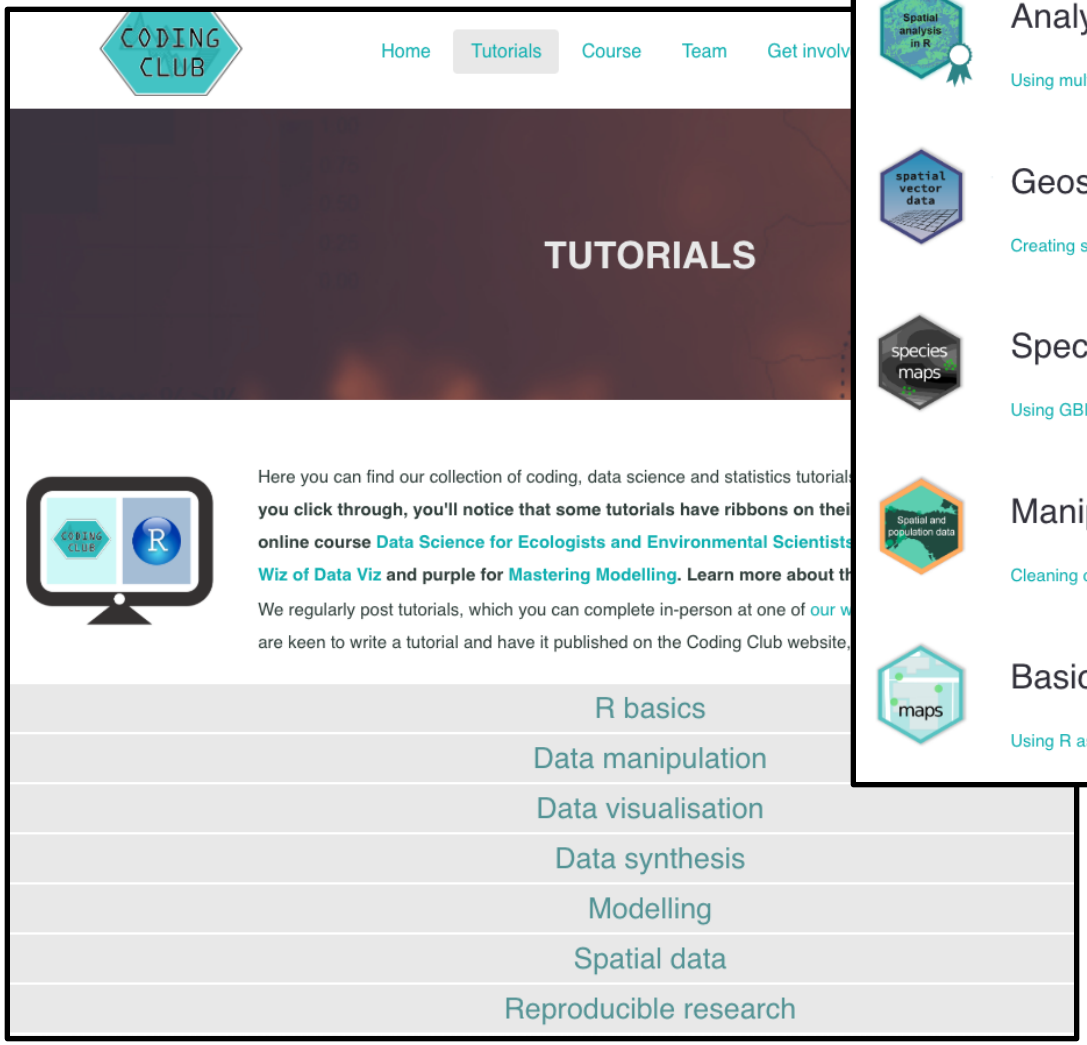
- R Markdown allows you to document the full end-to-end process of reading, cleaning, analysing, visualising and the publishing research
- From the same code, HTML web pages, PDF documents, word documents, LaTeX documents, dashboards and presentations can all be created.
- Shiny allows fully-fledged interactive web-applications to be created: https://shiny.posit.co/

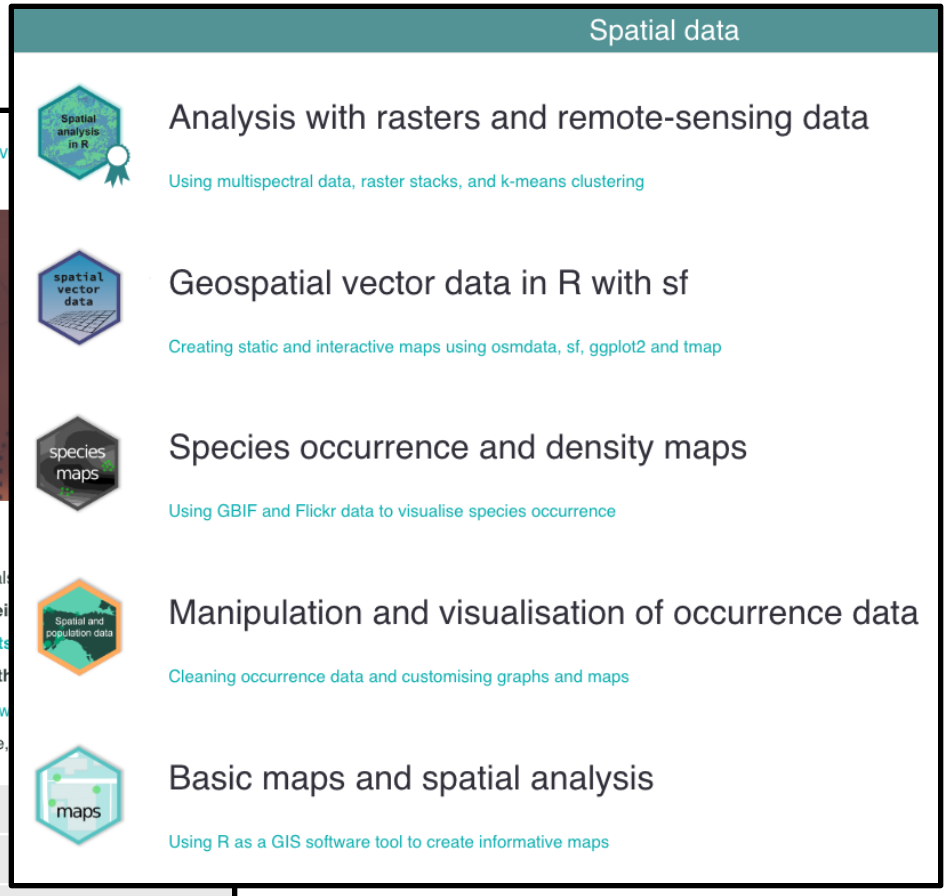
quarto®

Why use R as a GIS?

- Seamless integration of data
  - Reading / writing
  - Cleaning
  - Manipulating
  - Analysis
  - Visualisation
  - Documentation
  - Dissemination
- Through integrating all of these, reproducible science is facilitated (more of which in future weeks)
- It's FREE!!
- Constantly evolving
- Has an almost endless collection of spatial analysis, statistics and visualisation packages which will cover all of your needs.

R as a GIS

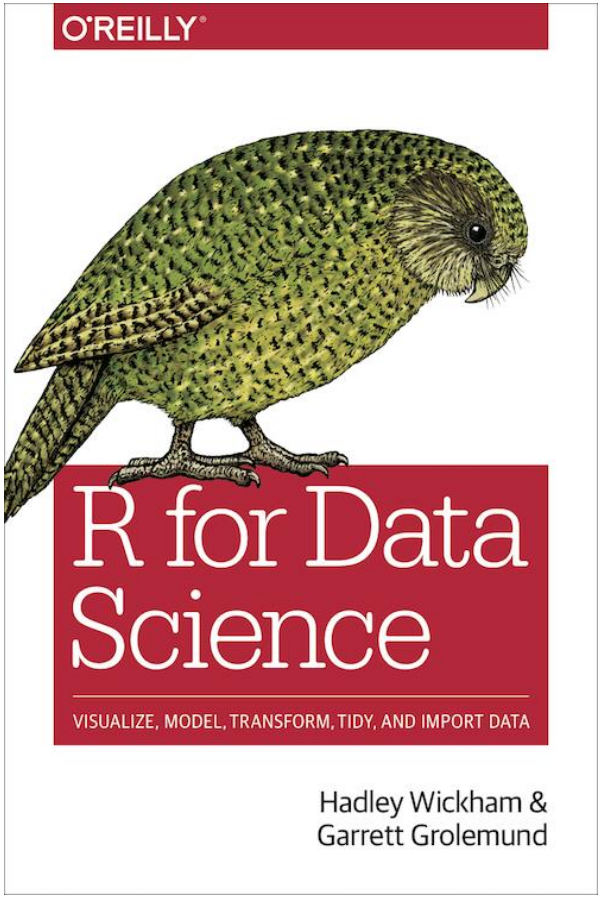




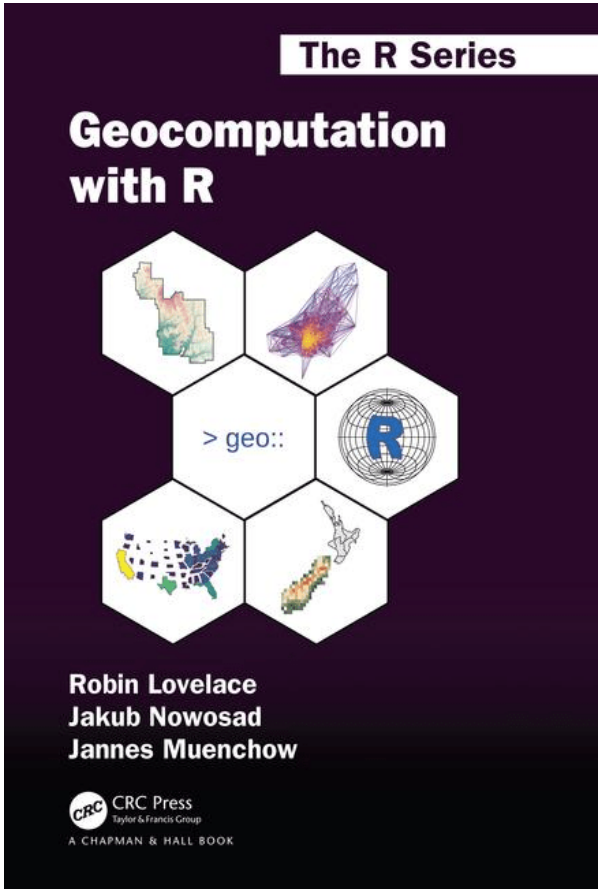
https://ourcodingclub.github.io/tutorials.html

Authors:
Gergana Daskalova
Maude Grenier
Boyan Karabaliev
John Godlee

R as a GIS



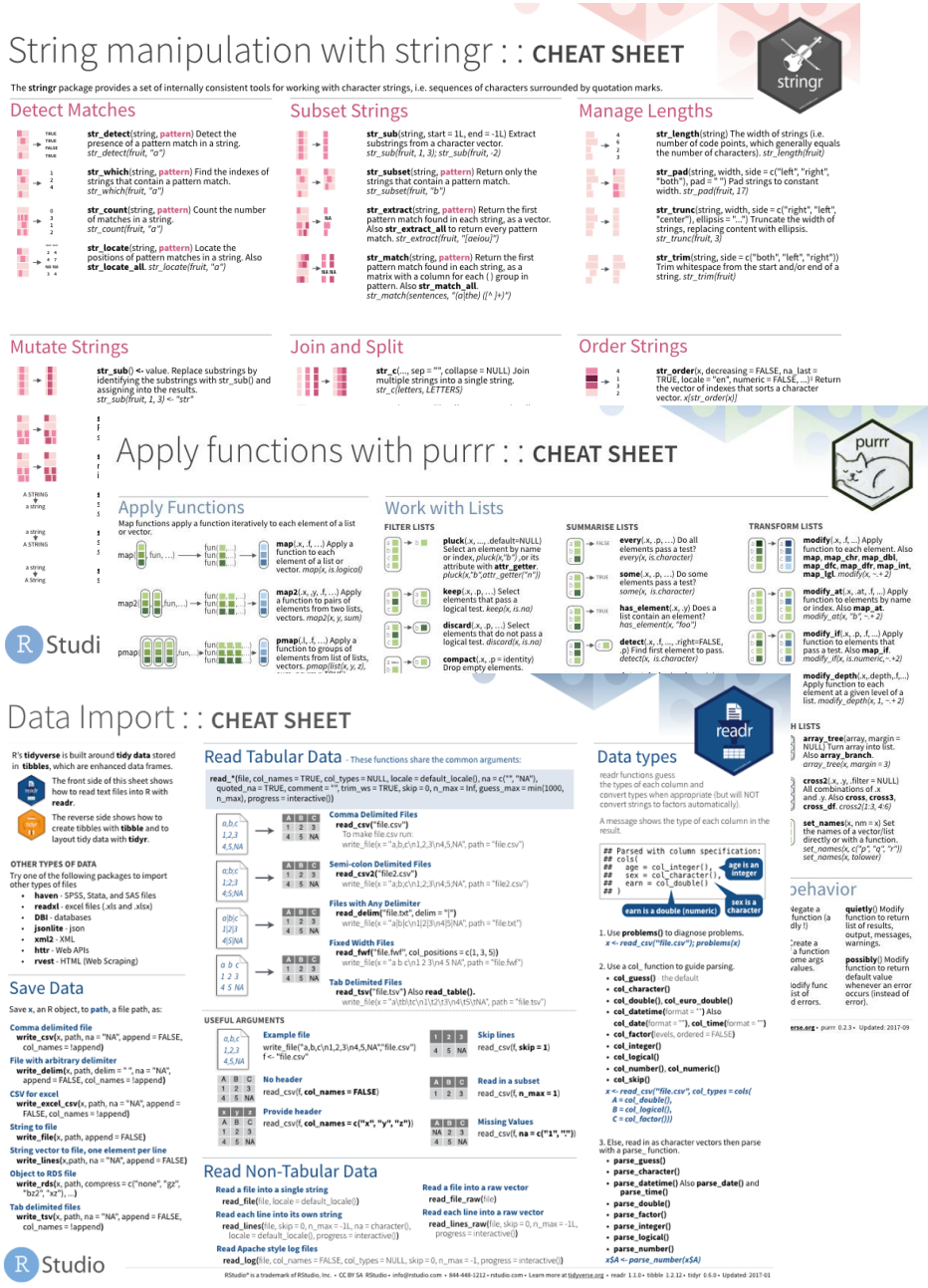
https://r4ds.had.co.nz/



https://geocompr.robinlovelace.net/

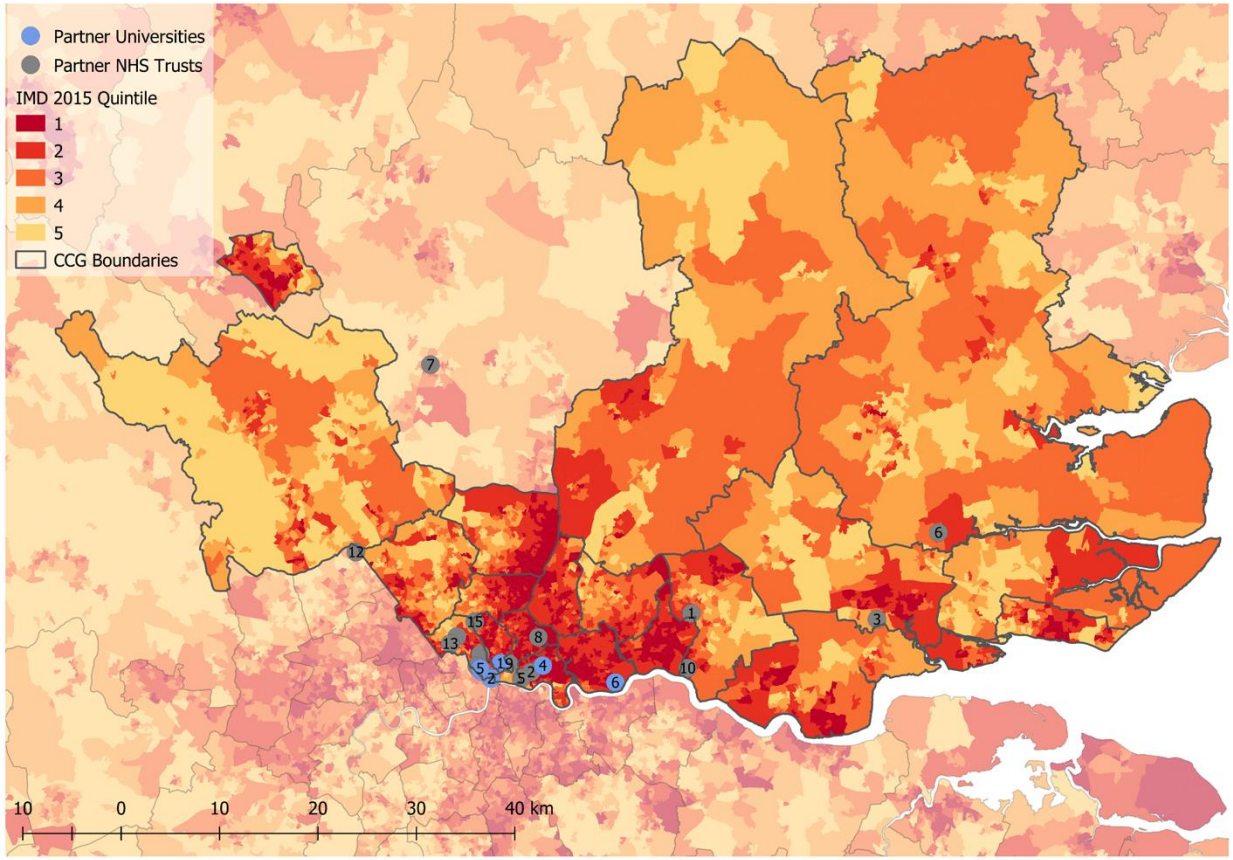
R Studio Cheatsheets

- And virtually everything you will need to know about data manipulation and analysis will be found via the brilliant and growing selection of R Studio Cheatsheets
- https://rstudio.com/resources/cheatsheets/



Why use anything else as a GIS?

- Cartography = graphical representations of geographic areas
- For fine-tuning cartographic outputs for print, sometimes better to use software specifically designed to be a GIS… QGIS
- Sometimes it's just easier to click a few buttons rather than write elaborate code!



Why use anything else as a GIS?

andrewmaclachlan.github.io/CASA0005repo/qgis-maps.html

Relaunch to update

All Bookmarks

Code

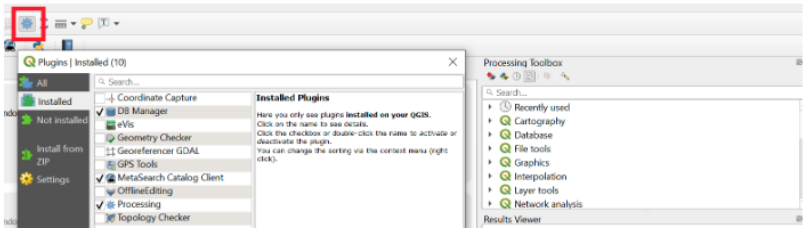
QGIS Maps

This uses the same data from the mapping practical - airbnbs and hotels (OSM) in London.

Ok, now we're going to produce our map in QIGS.

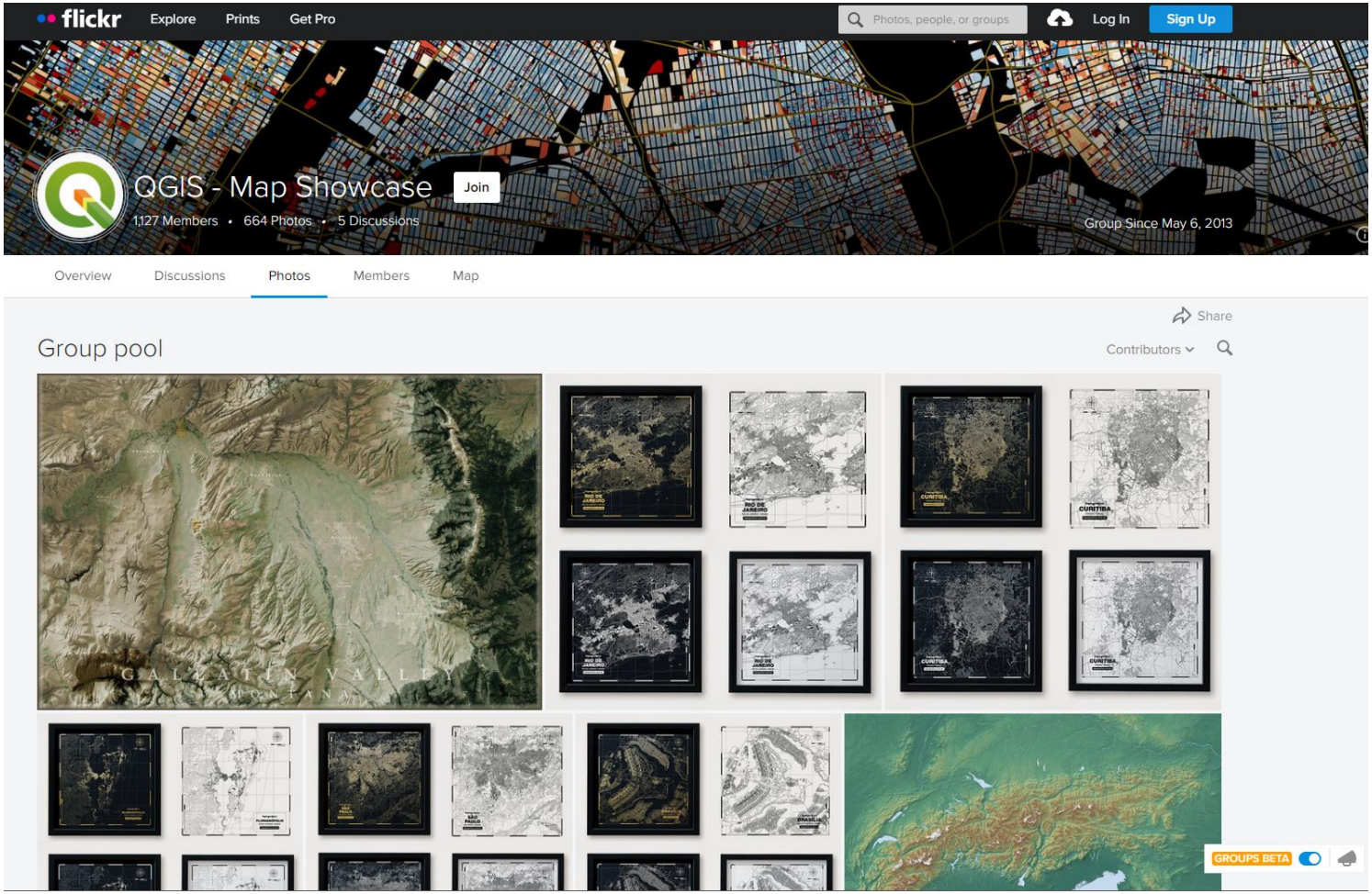
8.20 Load data

1. Load QGIS, Open and Save a new project (Project > New)
2. Right click on `GeoPackage` and create a new database to store our data in `.gpkg`
3. Load our data layers: London boroughs and OSM data (OSM data should be the `gis_osm_pois_a_free_1 polygon layer`).
4. Make sure the processing toolbox is active...go Plugins > Manage and Install Plugins > Installed (left side of the box that opens), Processing should be ticked....then select the cog that is in the toolbar — within the sqaure box in the image below.



Why use anything else as a GIS?

Why use anything else as a GIS?



https://www.flickr.com/groups/2244553@N22/pool/with/50355460063/

Conclusions

- In many ways Geospatial / Geographic Information Systems have evolved considerably since their invention in the 1970s
- However, at their core, they still carry out exactly the same functions of data input, storage, processing/manipulation, analysis and visualisation that they did 50 years ago
- Software and data have evolved. Analysis has become more sophisticated
- R was developed as a piece of statistical software, but being open source and developed by a huge community of R users means that it is now one of the most comprehensive GIS solutions available today
- With R & QGIS, you can do everything!!

Thanks!

- Any Questions?