



Introduction to GIS

Raphaëlle ROFFO

Sciences Po - GETEC Masters
Fall Semester 2021





Session 5

Dealing with vector data:
Table joins, Geoprocessing &
Common issues troubleshooting

Today's plan

1. Session 4 recap
2. Tutorial 4 debriefing / questions
3. Navigating data types
4. Geoprocessing: buffer, clip, intersect, merge, dissolve



Session 4 Recap & Tutorial Debriefing

Session 4 learning objectives

After Session 4 you should now be comfortable with:

- **Cartographic design principles**
- General rules for improving your maps' **accessibility**
- The various **layout elements** you must always include in a map
- What **choropleth** maps are
- How to define **class breaks** (number of classes, break values) - see [this article](#) if you're still confused.
- What type of **colour ramp** is more appropriate for your data

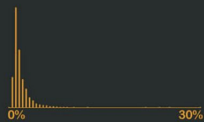


more on the matter

MAPPING THE TRUTH

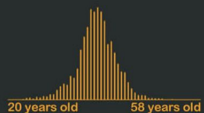
map readers, scope the range breaks; map makers, understand options and tradeoffs; everybody, classification matters. here are examples of two different datasets each classified by three different, common, methods -with way different results...

% MULTI-ETHNIC per county, u.s. census bureau



REALLY SKEWED!

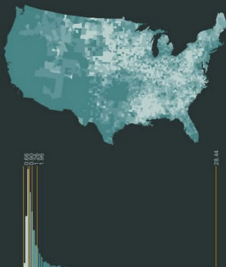
AVERAGE AGE per county, u.s. census bureau



SUPER NORMAL!

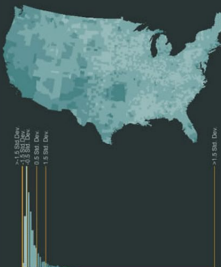
QUANTILE

unevenly spaced, evenly filled buckets



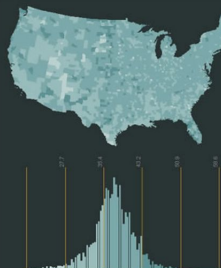
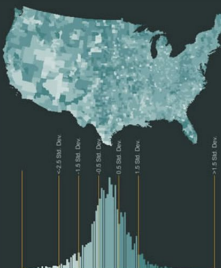
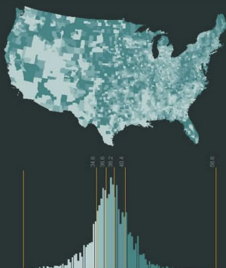
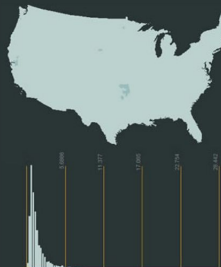
STANDARD DEVIATION

evenly spaced (to a stat geek), unevenly filled buckets



EQUAL INTERVAL

evenly spaced, unevenly filled buckets



© John Nelson, IDV Solutions

Will always depict variability, even if there is very little variability in the data.

Results in a reliably lively map but can be misleading (and the legend may seem arbitrary).

Consider this method if the data is highly clumped but you still need to tease out visual variation.

Useful for comparing map elements against a baseline average (as such, you'd want a "diverging" color scheme, but whatever).

Tends to tease out visual variation well even with clumped data, but is bad news for bi-modal data.

Good luck explaining the legend.

Useful for "getting" the map quickly and easy, though there may not be much to get.

Will invariably result in a visually bland map unless the data are really flatly distributed (which is rare).

Usually a poor option for social data, but works alright for environmental variables like temperature which are readily thought of in, say, chunks of ten.

30 Days Map Challenge

(The “Inktober” of GIS)

Find the rules for the 2021 edition here:

<https://github.com/tjukanovt/30DayMapChallenge>

And check out #30dayMapChallenge on Twitter to see how people contributed!

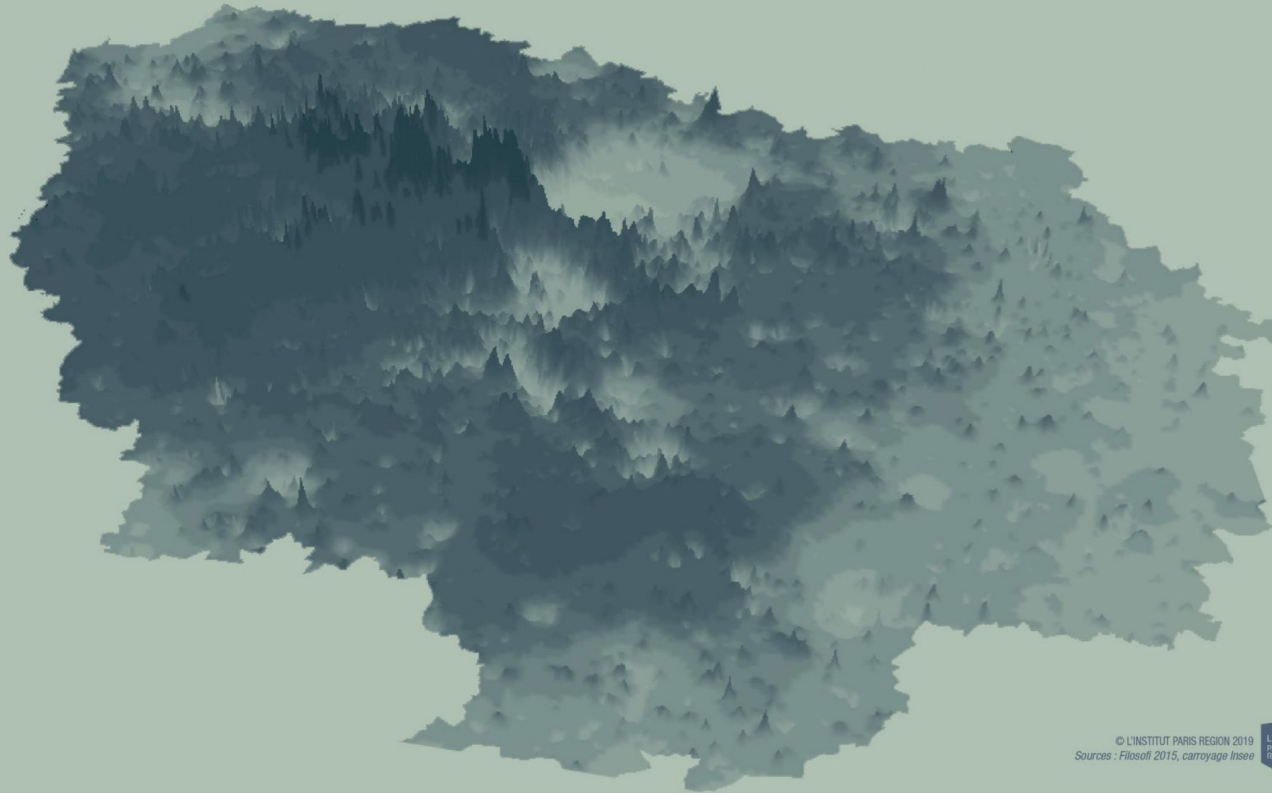
<https://twitter.com/hashtag/30daymapchallenge>

PICS ET VALLÉES DES REVENUS

LE RELIEF FRANCILIEN

Voici une carte en relief qui dessine une autre géographie de l'espace francilien. Elle nous invite à visualiser les espaces où se concentrent les hauts revenus tels des pics de montagne, et les plus bas tels des creux de vallée.

Au-delà des simples contours de ce paysage, ce sont les pentes qui nous interpellent : des espaces de transition abruptes parfois difficiles à franchir.





Tutorial debriefing & Questions

Tutorial debriefing

The Tutorial looked at styling and symbology. You should now be comfortable:

- Loading data from OpenStreetMap
- Accessing vector symbology settings for your points, lines, polygons
- Creating rule-based symbology
- Adding and styling labels
- Setting scale-dependent visibility
- Saving spatial bookmarks

Additionally, in Tutorial 5 you will get to build a choropleth map and use the print layout composer to export your map.

Questions?

Don't wait for the coursework deadline to ask questions!

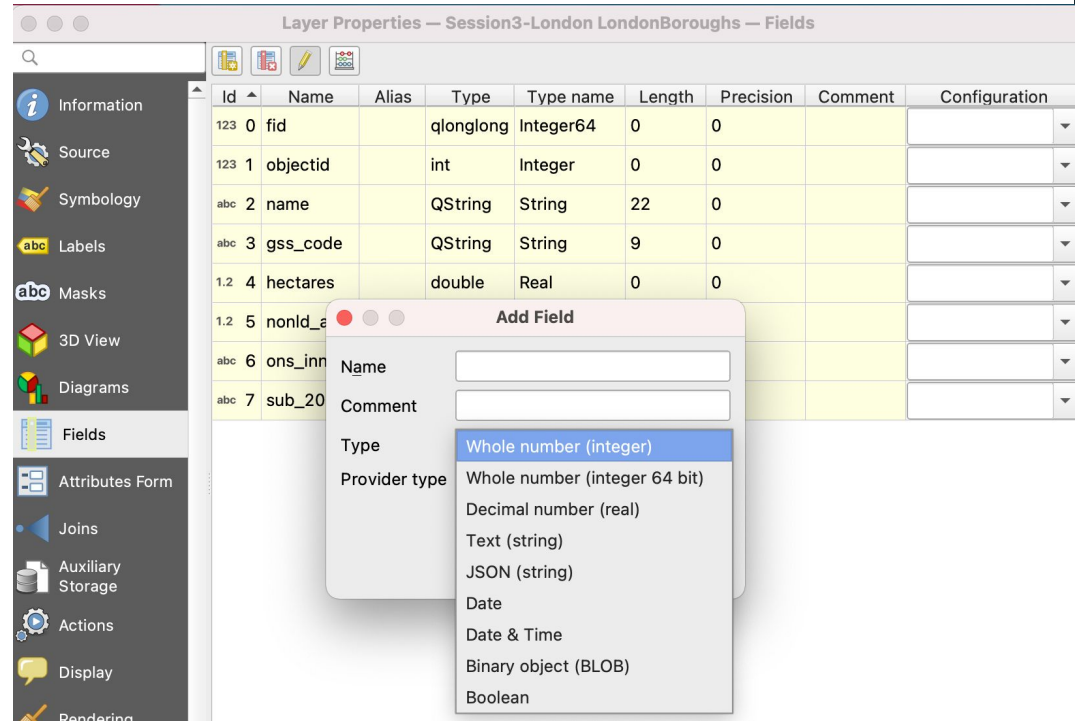
On Slack you've asked about: QGIS crashing unexpectedly, expressions, rendering and layer order, values not loading in the symbology panel, etc.



Navigating data types

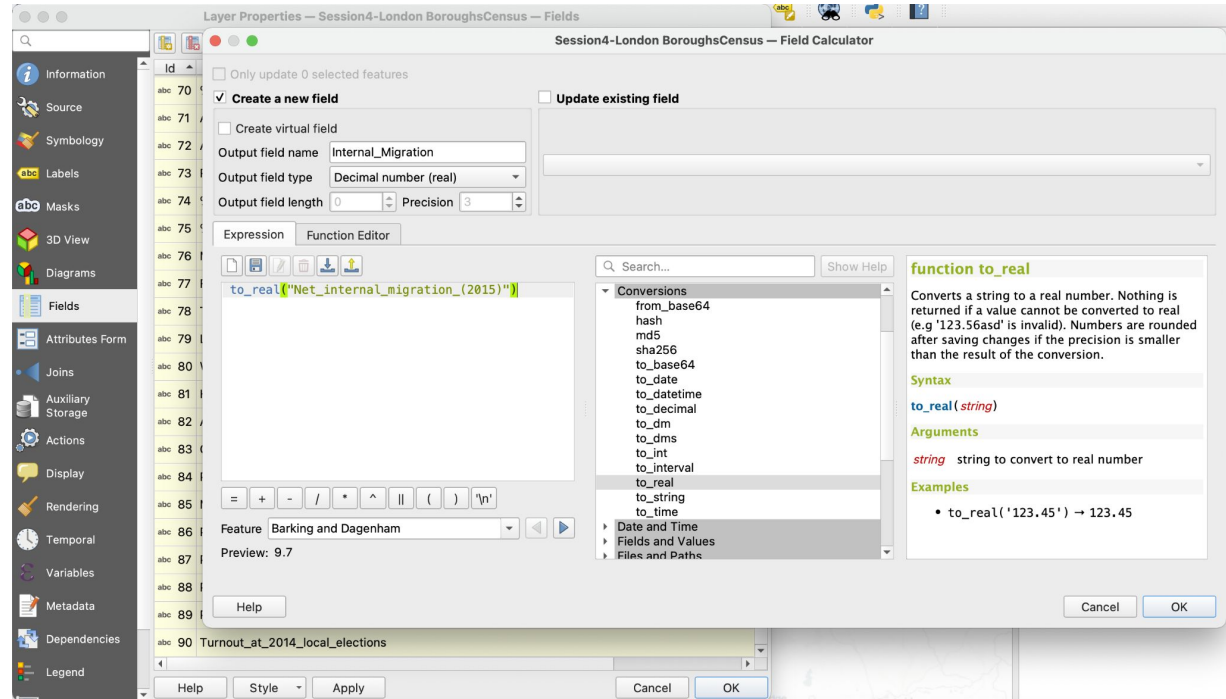
Data types in QGIS

- Integer (whole numbers)
- Real number (decimal numbers)
- String (text)
- Date / time
- Binary / boolean (1 or 0)



You may need to edit the data type!

Typically, to be able to use graduated symbology on numerical values that are currently saved as text.





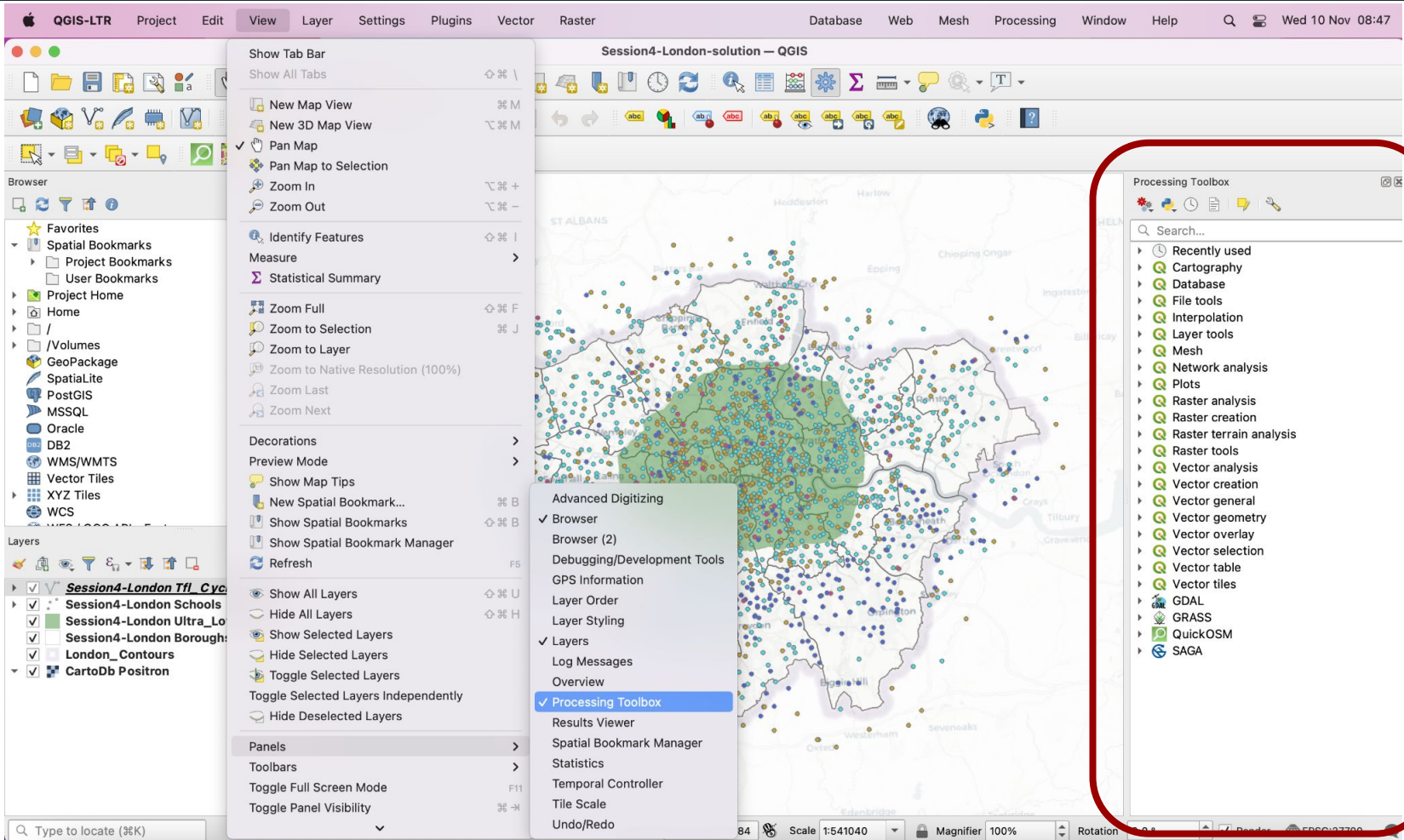
Geoprocessing

Geoprocessing

Geoprocessing tools are operations and tools that allow you to manipulate one or many geographic datasets, and that result in new data (one or many datasets).

In QGIS, these tools can be found in your processing toolbox.

In this introduction course we will only cover the most commonly used vector geoprocessing tools: buffer, clip, difference, merge, dissolve, intersect.

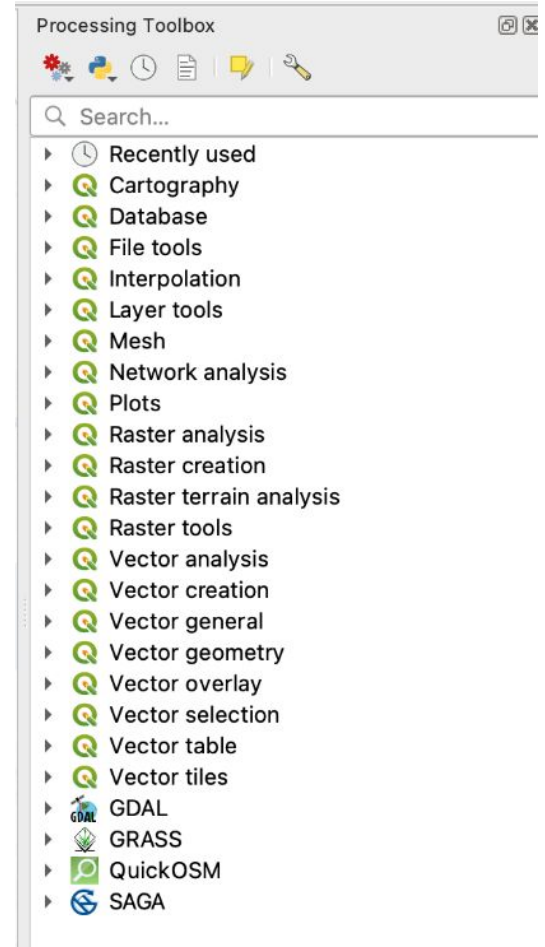


The Processing toolbox

All the geoprocessing tools can be retrieved through the processing toolbox.

You can search by name, or navigate by category.

The tools you've last used will be available to you in the *Recently used* category.

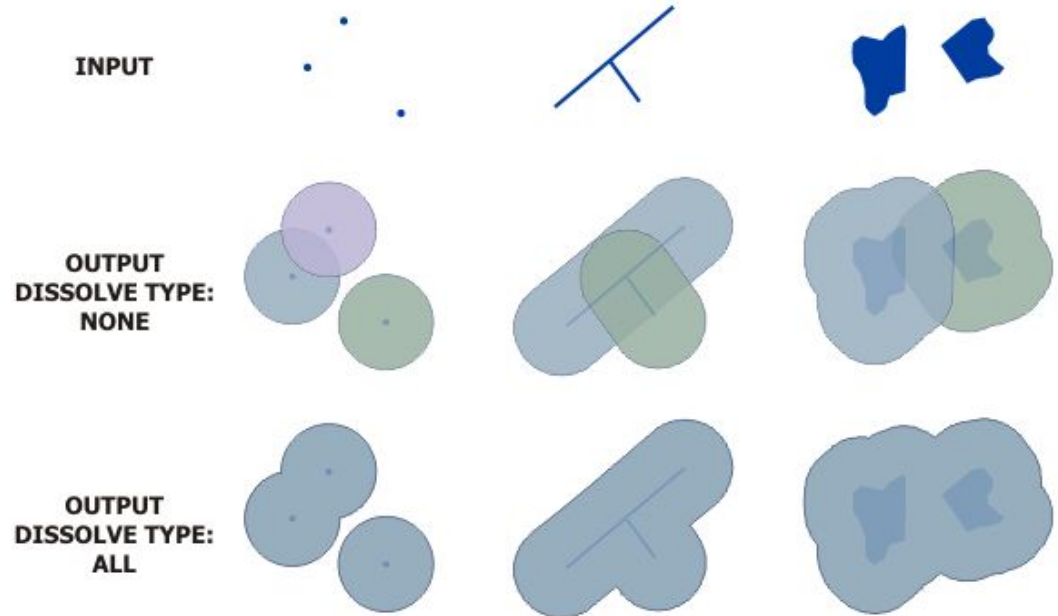


Buffering

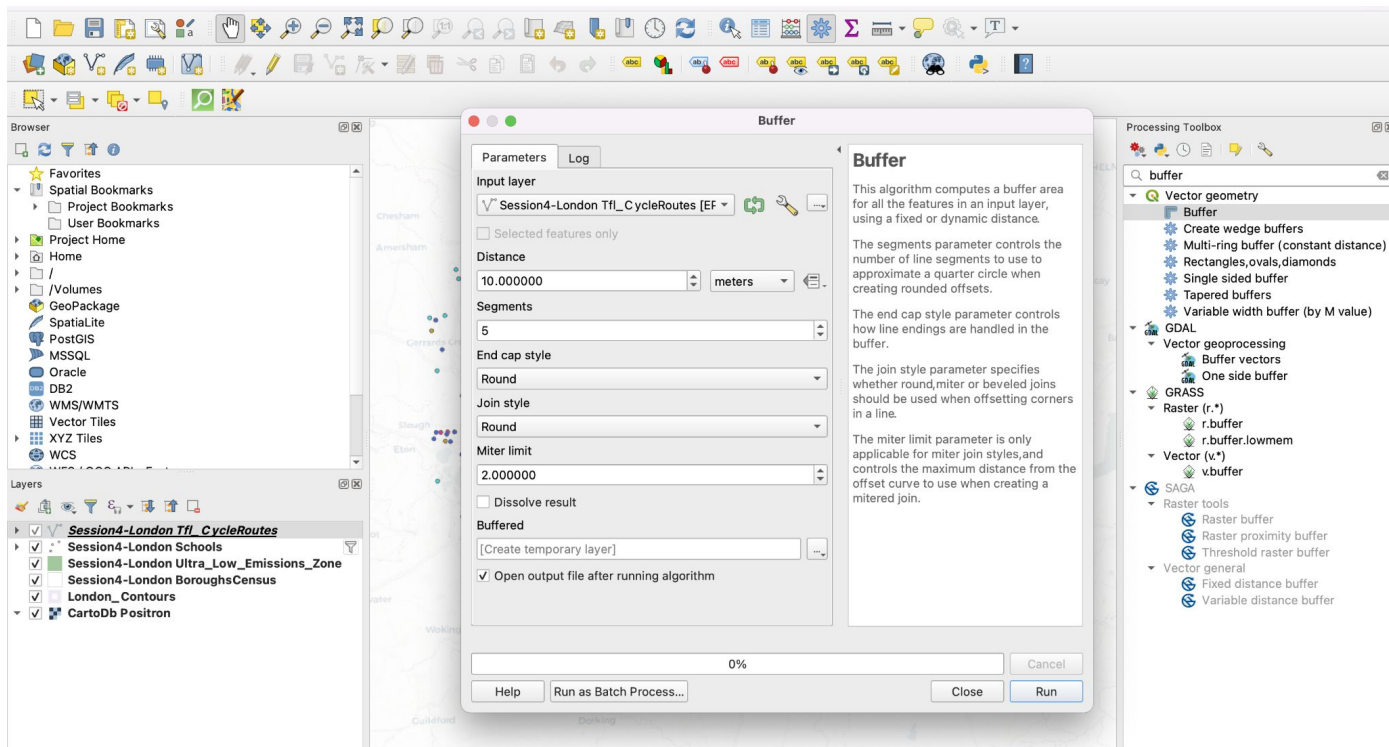
Take an input vector and specify a distance.

→ The tool creates buffer polygons around your input features.

Example use: find out all areas that fall within 1km of a school.



Buffering



Clipping

Take some target features (your input) and overlay a polygon like a “cookie-cutter” on top; you only keep the target features that fall within the clip polygon.

→ The boundaries of the clip are imposed on your target features and the rest is discarded.

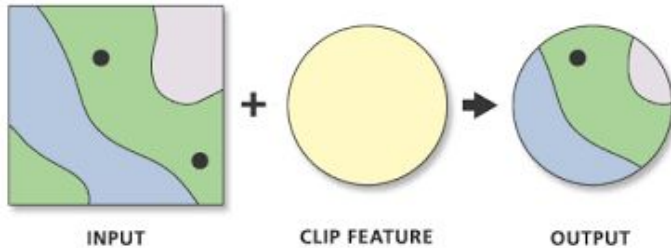
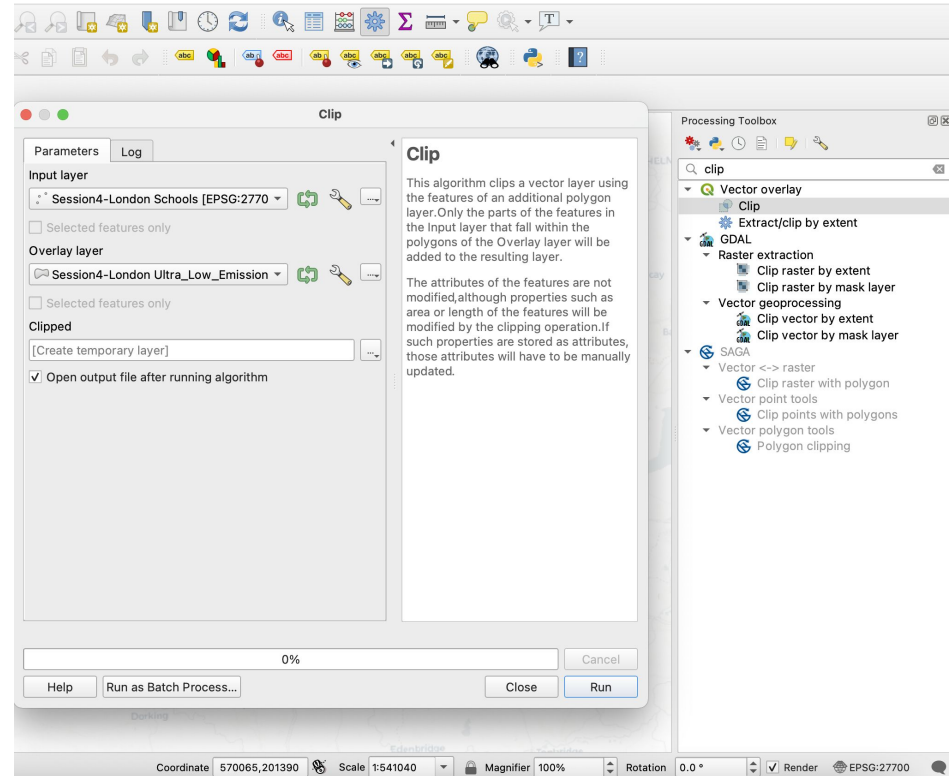


Image credit: [ESRI](#)



Difference (Erase)

You can think of it as a reverse clip!

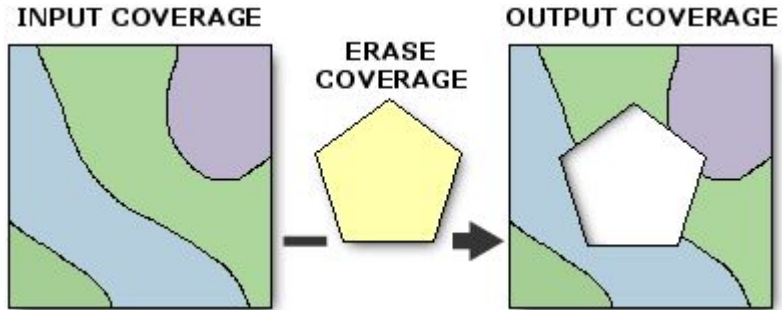
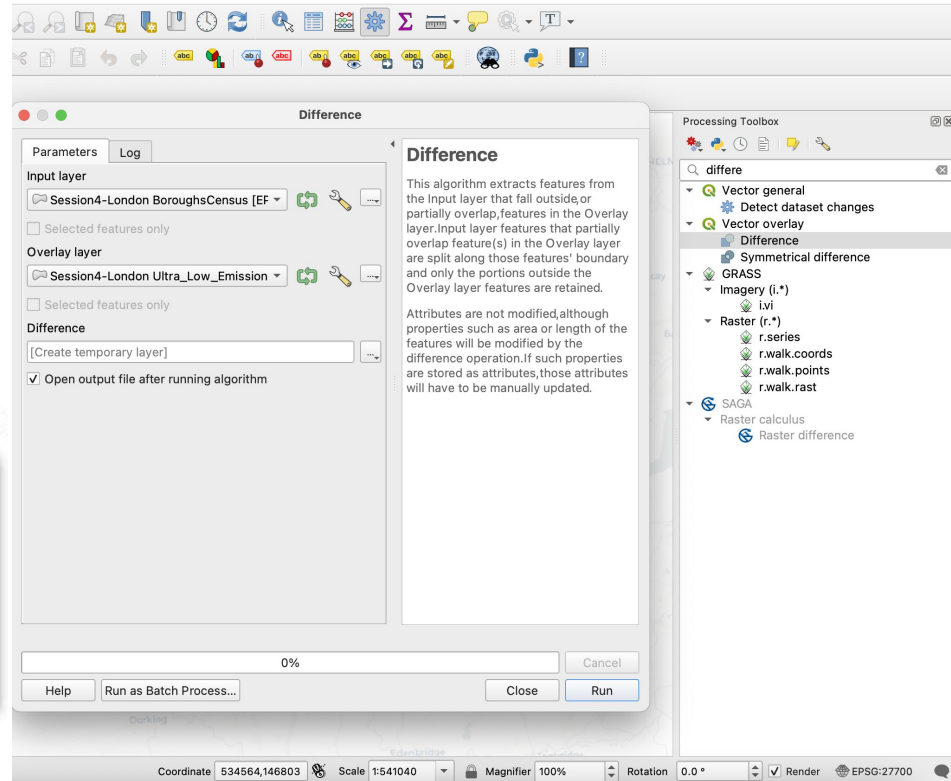


Image credit: [ESRI](#)



Intersect

Input two vector layers; the output will be all features - or portions of features - that overlap in both layers.

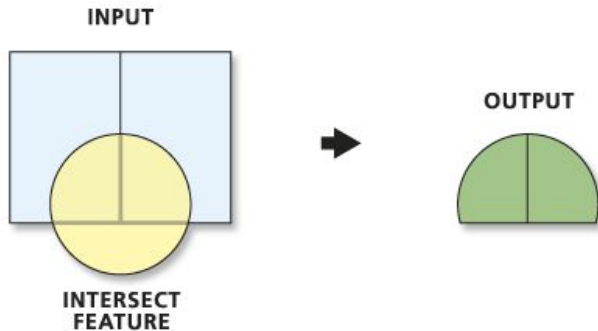
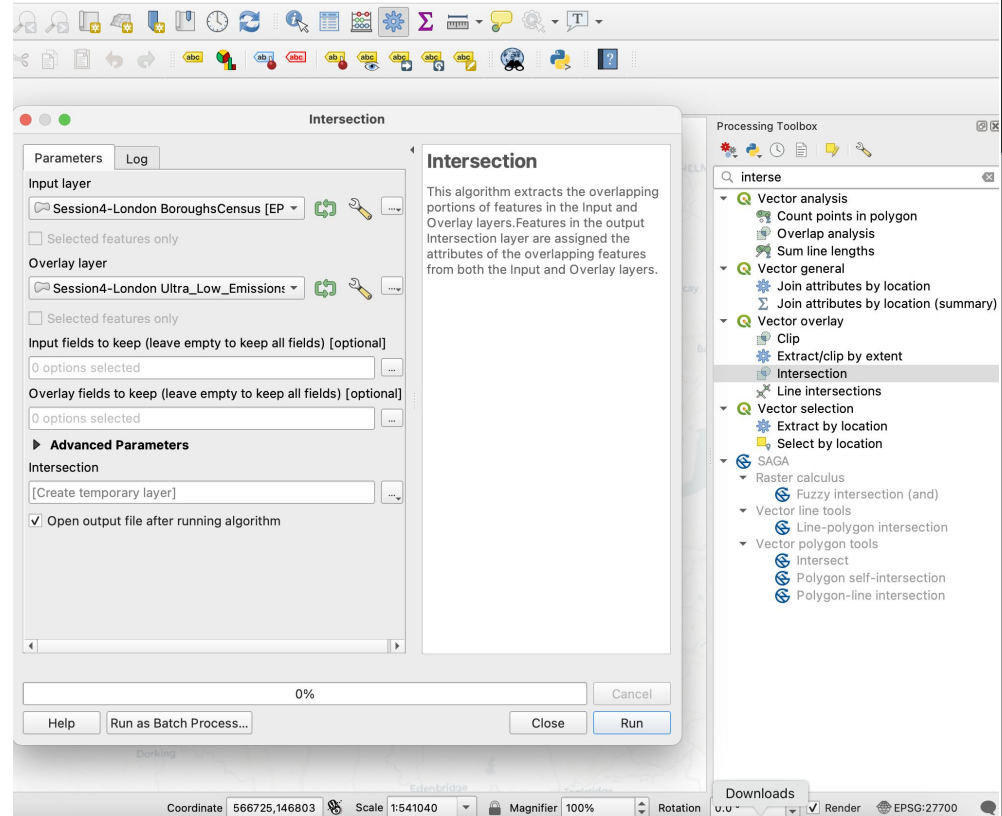


Image credit: [ESRI](#)



Symmetrical Difference

You can think of it as a reverse intersect!
The output layer will contain all features that are only present in **one and not both** the input layers.

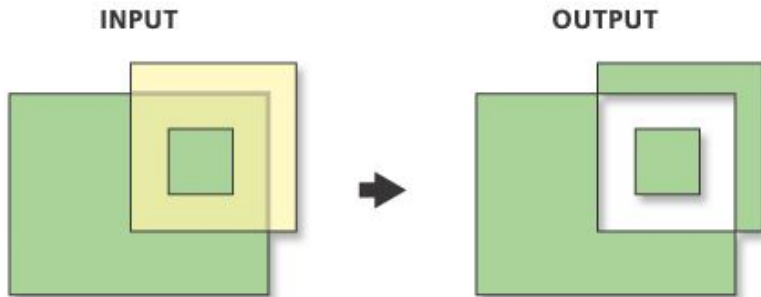
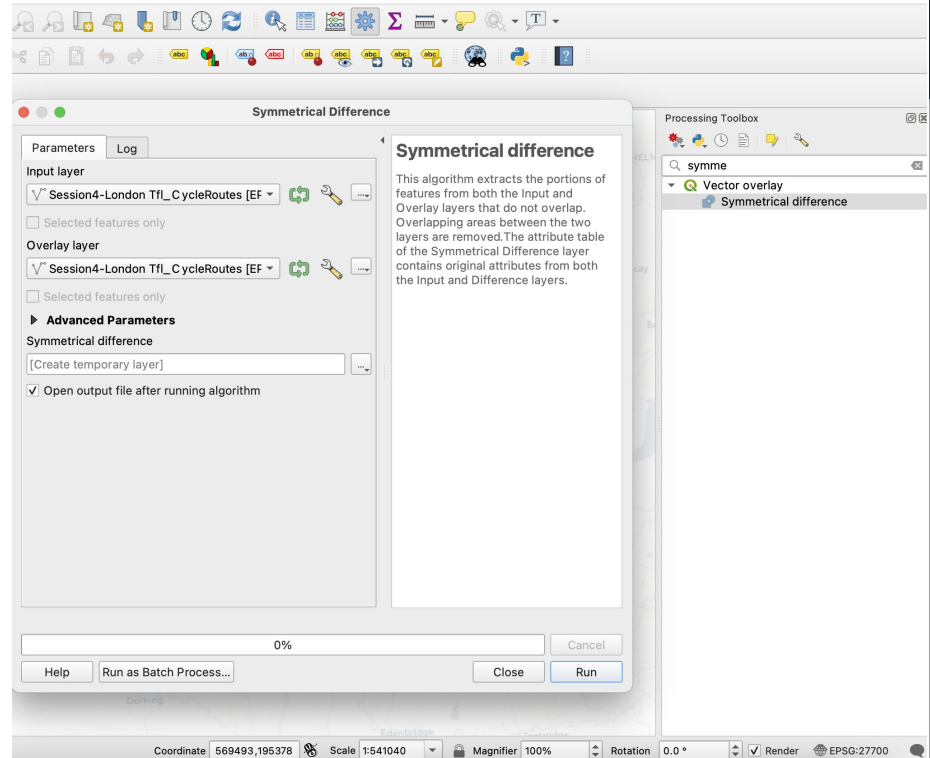


Image credit: [ESRI](#)



Merging

Combine multiple vector layers of the same data type (either all points, all lines or all polygons) into a single, new output dataset.

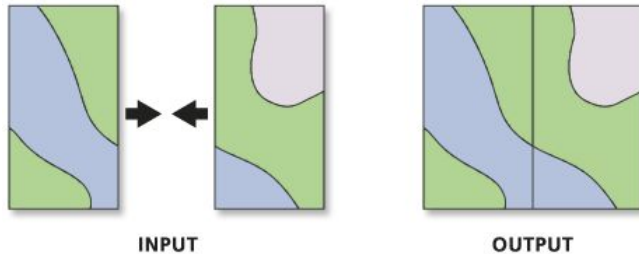
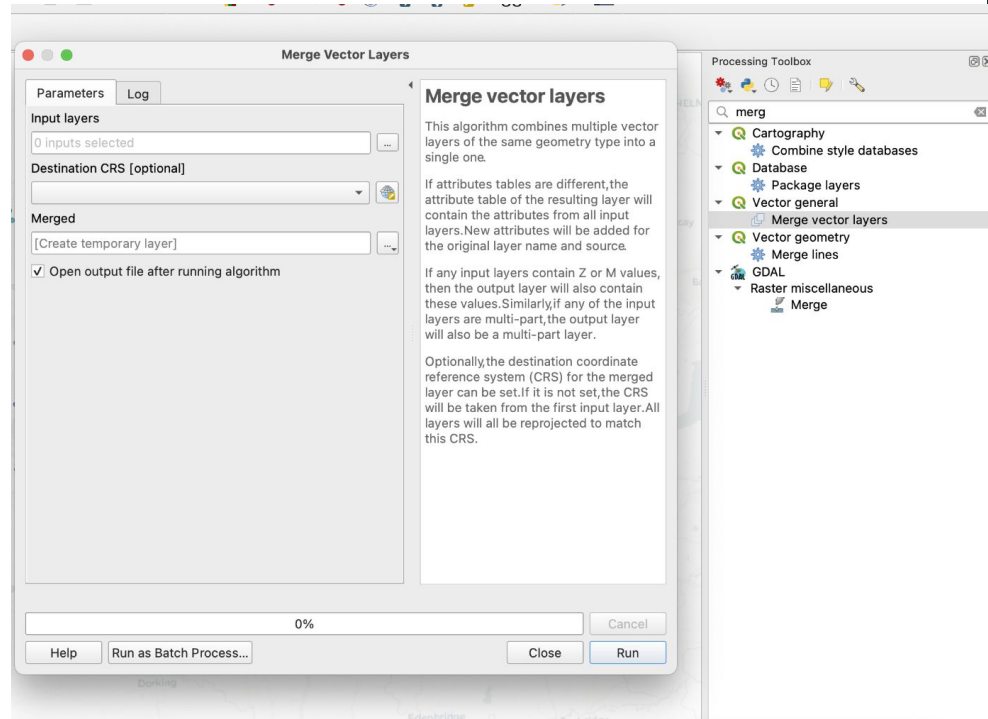


Image credit: [ESRI](https://www.esri.com/)



Dissolving

Take a vector layer and choose and merge adjacent polygons, lines, or regions that have the same value for a specified field.

For example, you can “dissolve” all counties within a province, or all wards within a borough.

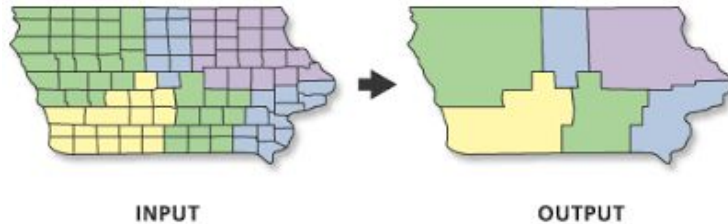
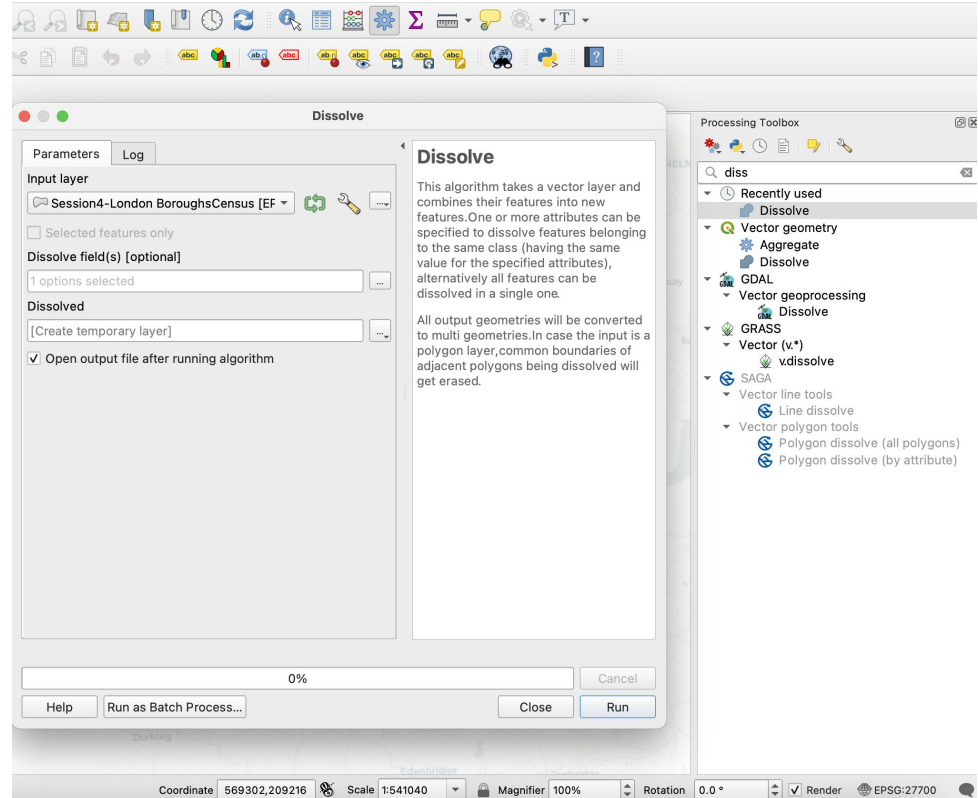


Image credit: [ESRI](https://www.esri.com/)

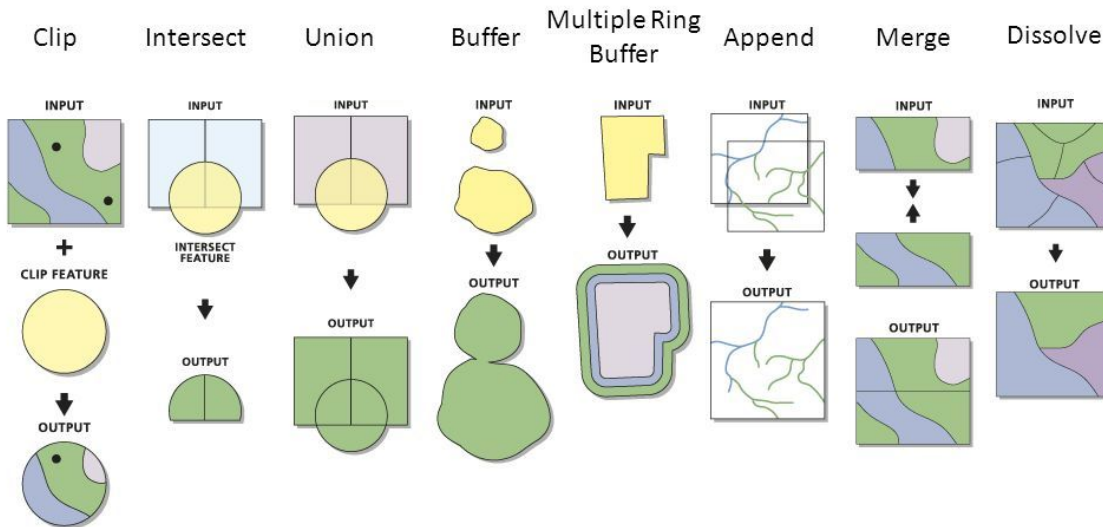


Summary

Geoprocessing

Many more operations are available to you; in doubt, your best allies are:

- the [QGIS documentation](#),
- [GIS Stack exchange](#) on Stack Overflow,
- ...and of course your favourite search engine!





Tutorial

Homework

Note: The course material has a [new home](#) :)

1. Do the [Session 5 tutorial](#)
2. Make sure you have a topic for your final coursework!
(see next slides)
3. Do use Slack if you have questions (#help), and more importantly to start helping each others out too!
Explaining to someone else is a great way to test your own understanding.



Coursework

Final Coursework

The final coursework is a **map production exercise**. Find a research question, carry out a simple map production workflow, and write a report summarizing your findings. You will be assessed on your capacity to:

- Frame your research question in a clear and concise manner, and ensure a few maps can provide interesting insights,
- Identify datasets that are relevant to answering your question (*technically here you may want to work backwards and use the data you already have to find your research question*),
- When appropriate, use table joins to “enrich” your vector data,
- Carefully choose your symbology, and ensure your map is accessible and colour-blind safe,
- Create map exports complete with all key cartographic elements (title, legend, data source etc).
- Analyse and interpret the patterns that emerge from your maps, explaining what this might mean in terms of policy or research outlooks.

Final Coursework

Some practical points:

- Deadline: Wednesday 24th November, 23.59 Paris time. *If you're late, minus 1 point for each day behind the deadline.*
- Work in groups of 2-3 students
- Work on a **European city** of your choice (*cities in the UK and other non-EU european countries are ok*)
- The report must be 3 pages minimum, 5 pages *maximum*
- You are encouraged to use the data provided for the tutorials, however if you want to challenge yourselves, you will get **+2 points bonus for working with data you have sourced yourselves.**

Final coursework: Proposed outline

You are strongly encouraged to follow this outline:

1. Executive summary (maximum ½ page, bullet points are fine)
2. Introduction / Problem / Context
3. Data sources in a table
4. High-level methodology. Keep it very short but use precise terminology. Explain your symbology choices including your choice of class breaks if you built a choropleth.
5. A minimum of 2 maps. Careful, you only have 5 pages maximum in this report so these maps must be relevant to answering your policy question
6. Analysis of the findings (half a page)
7. Conclusion / next steps for policy makers or future research outlook (half a page)

In this exercise, concision and precision are key!