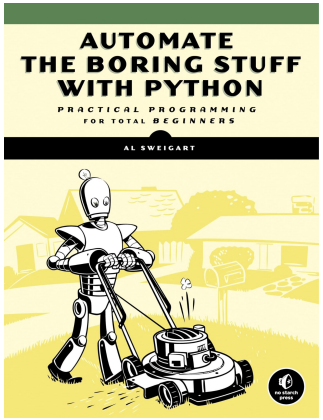
Digging with Python

Making Archaeology More Powerful, One Script at a Time

Why This Talk?



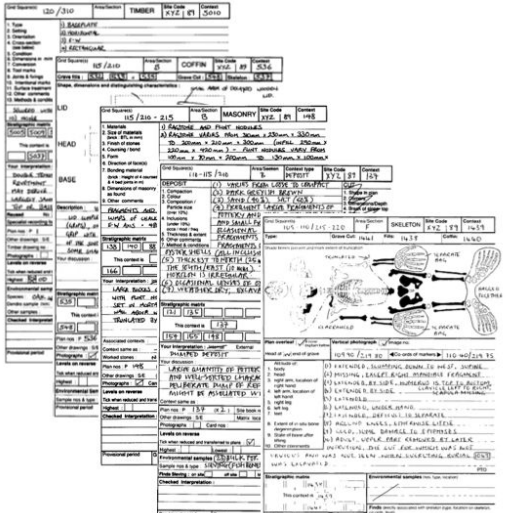
- Bsc in Archaeology
- 6 years of field archaeologist experience
- Quantitative Archaeology, statistics + R
- Bsc in Data Science, MSc in Advanced Computing
- 3 years of experience in Data

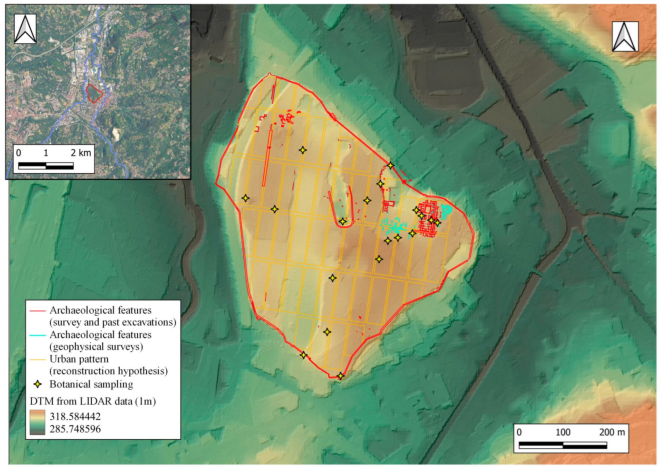
Most archaeologists don't think of themselves as coders. And many coders don't realise how much archaeology *is* data.

"Python isn't just for tech companies. It's for storytellers, researchers, and curators – and yes, for archaeologists. **It's designed to automate tedious tasks, data analysis, file processing, AI..."**

Archaeology: A Data-Driven Discipline

- A lot of archaeology data recording and processing:
  - Surveys
  - Site maps
  - Samples
  - Context sheets
  - Logs
  - Paperwork
  - Articles
- A lot of processing, compilation, formatting
- Manual work

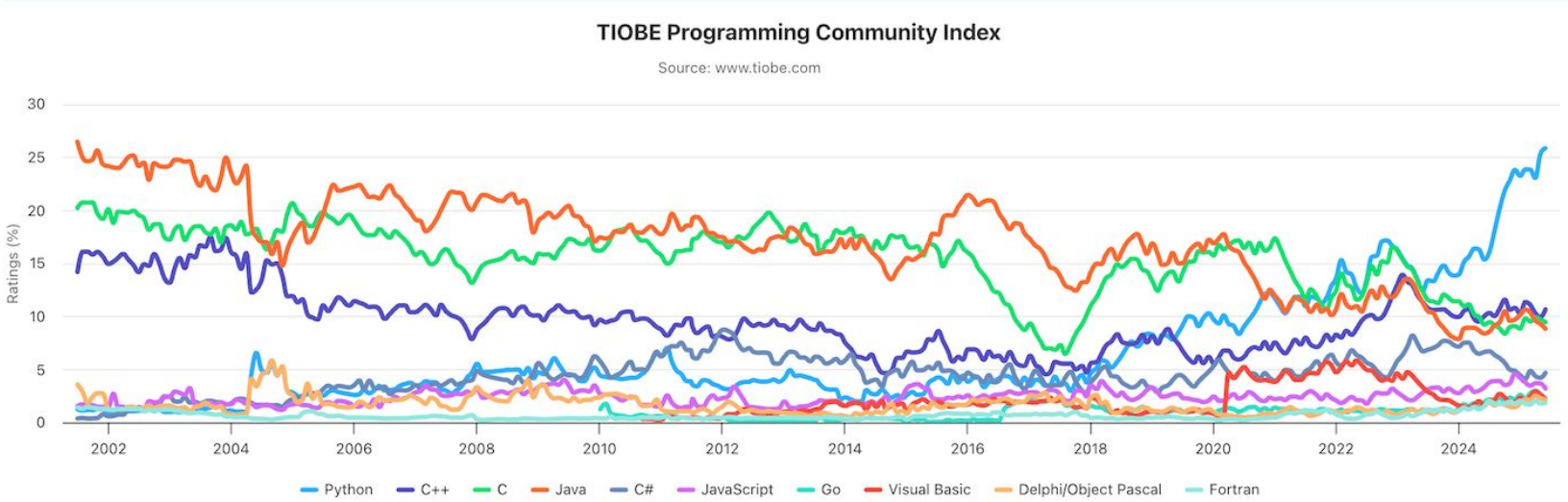


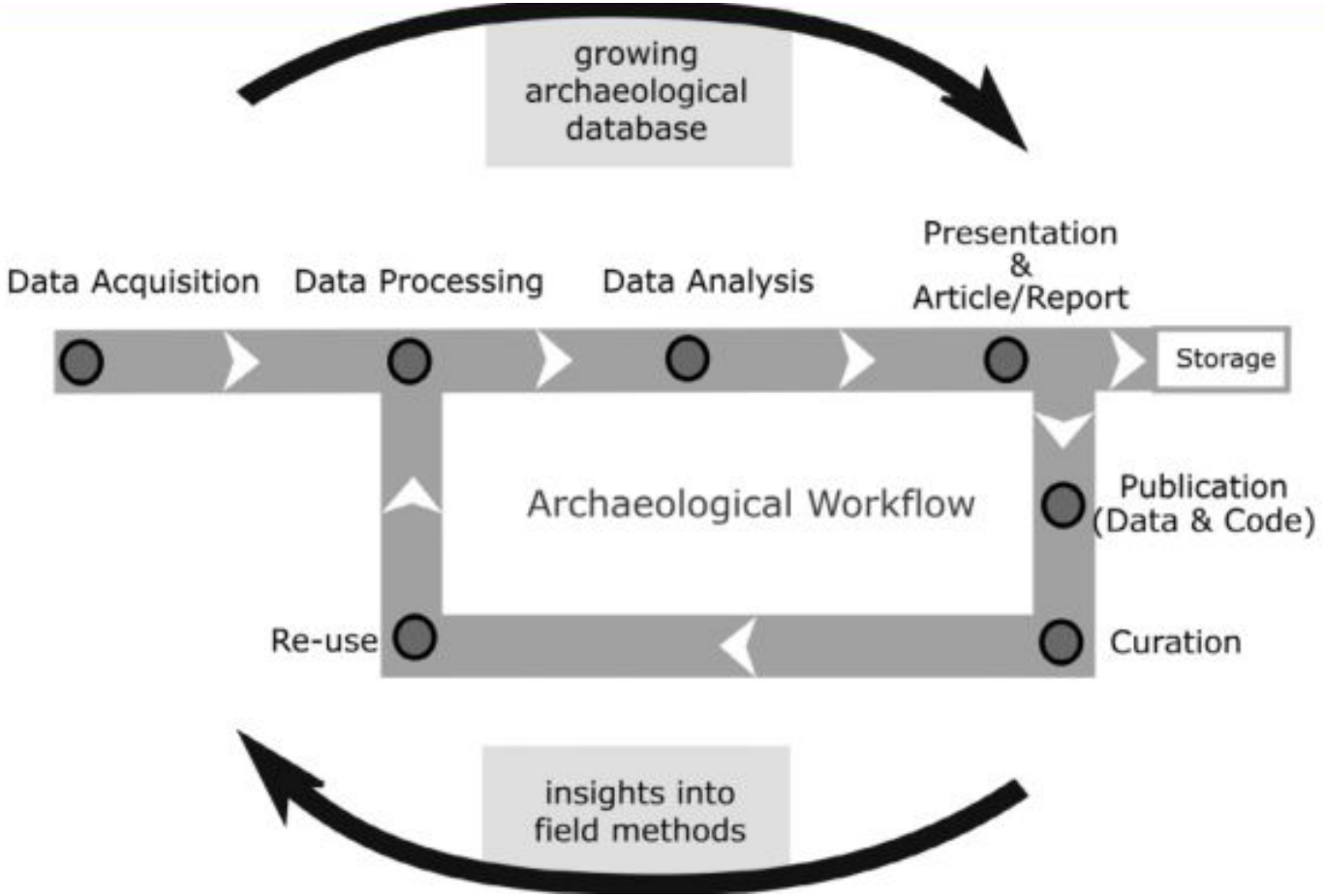


But... why Python?

- Simple syntaxis
- Beginner “friendly” (almost)
- Wide community
- Multiple fields and tools

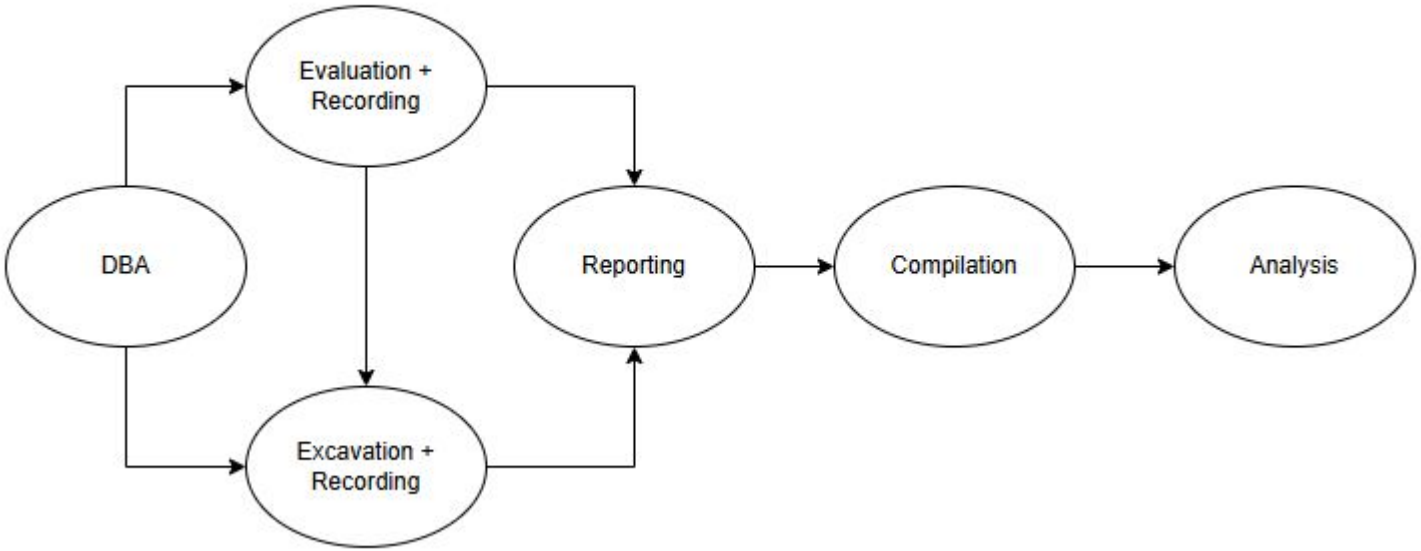
- Broad integration
- Multi-platform
- Interpreted (no compilation needed).





Gillings, M., Hacıgüzeller, P. and Lock, G. eds., 2020. *Archaeological spatial analysis: a methodological guide*. Routledge.

Archaeological process (simplified)



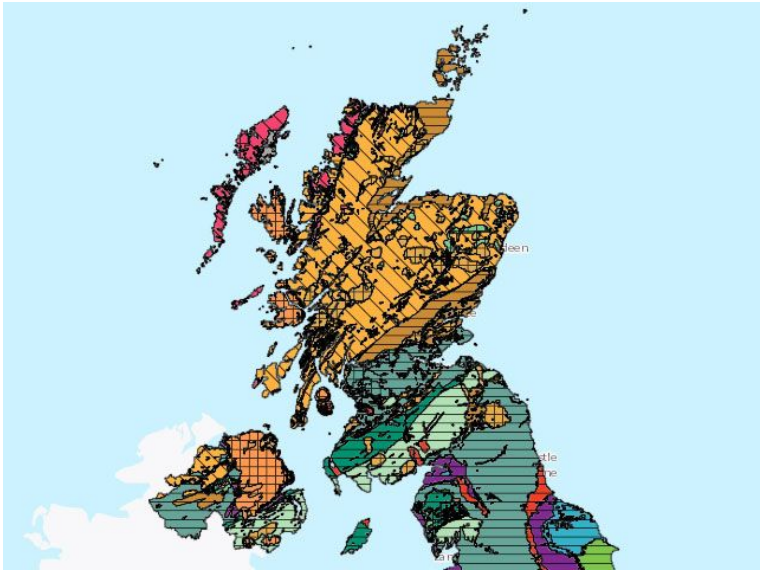
DBA: Nearby CANMORE locations

- Given the coordinates of interest (for example, site outline boundaries), this script compiles the locations in CANMORE DB in a given radius, and save them in a CSV file.
- Runtime: 9 seconds
- Possible future features:
  - Web scraping of Canmore URL
  - Plotting of locations around the coordinates of interest
  - Text summary of near heritage and archaeological elements

(Script: short program written in an interpreted language that automates tasks)

DBA: Geology of the site

- Given the coordinates of interest (for example, site outline boundaries), checks the geological information on BGS Geology 625K Database.
- Runtime: 7 seconds
- Possible future features:
  - Use of more detailed databases
  - Text summary of geology of the area



Evaluation: Trench Placer

- Given the outline of the site, pct of area to be evaluated, width of the trenches and the default trench length, this script can calculate the number of trenches needed, the blank space between them, and place all of them on the site map.
- Runtime: 23 seconds
- Possible improvements:
  - Different axis
  - Improve behaviour near to the site outline


Cathcart Street
Bank Street
Terrace Road
Bank Street
Well Park
Lyndoch Street
Station Avenue
Station Avenue
Regent Street
Terrace Road
Greenock Central
Bogle Street

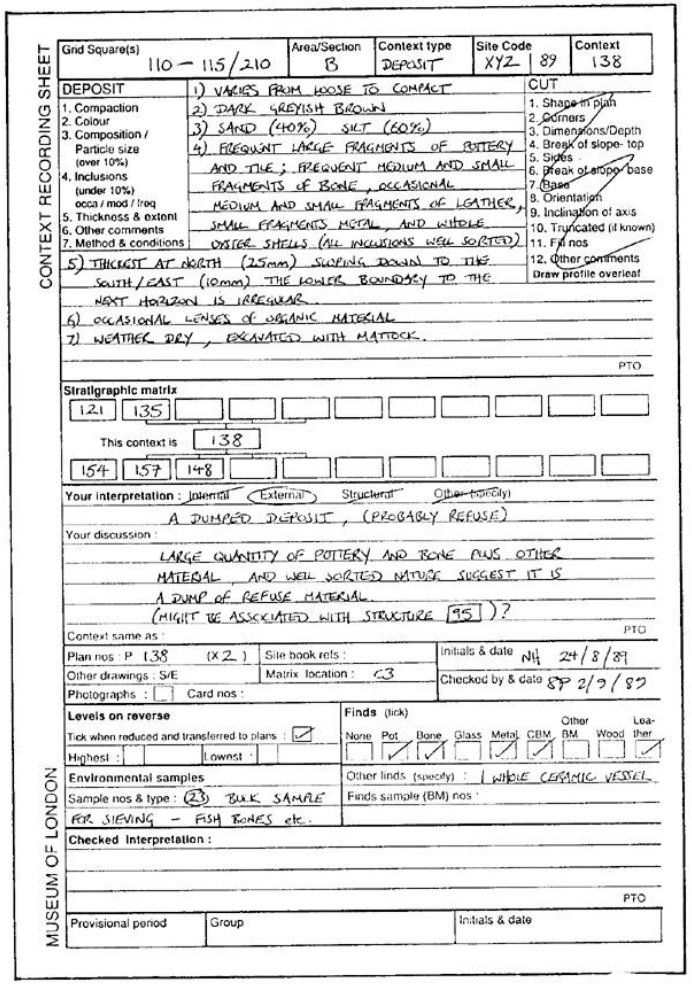

Trenches within Projected Site Polygon
676100
227950
228000
228050
228100
228150

Recording: Multiple spreadsheets joiner

- A simple script to join multiple spreadsheets, for example for multiple users and devices, on an unique file, checking if the file format matches for all cases.
- For records with some missing information, they are not included on the final file, but in a different one, to join the rest of the records once the gaps have been manually filled.
- Runtime: 3 seconds for 15 files and 1500 records

Recording: Automatic Context sheet transcription

- Using a local LLM model (Chatgpt style, but in your device), takes the image of a context sheet, and extract the details to include in a spreadsheet
- Stochastic parrot!!
- Model used: gemma3:4b using Ollama
- Runtime: 30 secs



CONTEXT RECORDING SHEET

Grid Square(s)	Area/Section	Context type	Site Code	Context
110 – 115/210	B	DEPOSIT	XYZ 89	138

DEPOSIT

1. Compaction
2. Colour
3. Composition / Particle size (over 10%)
4. Inclusions (under 10%) occa / mod / freq
5. Thickness & extent
6. Other comments
7. Method & conditions

1) VARIES FROM LOOSE TO COMPACT
2) DARK GREYISH BROWN
3) SAND (40%) SILT (60%)
4) FREQUENT LARGE FRAGMENTS OF POTTERY AND TILE; FREQUENT MEDIUM AND SMALL FRAGMENTS OF BONE, OCCASIONAL MEDIUM AND SMALL FRAGMENTS OF LEATHER, SMALL FRAGMENTS METAL, AND WHOLE OYSTER SHELLS (ALL INCLUSIONS WELL SORTED)
5) THICKEST AT NORTH (25mm) SLOPING DOWN TO THE SOUTH/EAST (10mm) THE LOWER BOUNDARY TO THE NEXT HORIZON IS IRREGULAR.
6) OCCASIONAL LENSES OF ORGANIC MATERIAL
7) WEATHER DRY, EXCAVATED WITH MATTOCK.

PTO

CUT

1. Shape in plan
2. Corners
3. Dimensions/Depth
4. Break of slope- top
5. Sides
6. Break of slope- base
7. Base
8. Orientation
9. Inclination of axis
10. Truncated (if known)
11. Fill nos
12. Other comments

Draw profile overleaf

Stratigraphic matrix

121 | 135 | | | | | | | | | |

This context is 138

154 | 157 | 148 | | | | | | | | |

Your interpretation : Internal / External / Structural / Other (specify)

A DUMPED DEPOSIT, (PROBABLY REFUSE)

Your discussion :

LARGE QUANTITY OF POTTERY AND BONE PLUS OTHER MATERIAL, AND WELL SORTED MATRIX SUGGEST IT IS A DUMP OF REFUSE MATERIAL. (MIGHT BE ASSOCIATED WITH STRUCTURE 95)?

PTO

Context same as :

Plan nos : P 138 (X2)	Site book refs :	Initials & date NH 24/8/89
Other drawings : S/E	Matrix location : C3	Checked by & date BP 2/9/89
Photographs :	Card nos :	

Levels on reverse

Tick when reduced and transferred to plans : ☑

Highest : Lowest :

Finds (tick)

None	Pot	Bone	Glass	Metal	CBM	Other BM	Wood	Leather
	✓	✓		✓	✓			✓

Environmental samples

Sample nos & type : (23) BULK SAMPLE FOR SIEVING – FISH BONES etc.

Other finds (specify) : WHOLE CERAMIC VESSEL

Finds sample (BM) nos :

Checked Interpretation :

PTO

Provisional period	Group	Initials & date

MUSEUM OF LONDON


```
{
  "context_number": "138",
  "type": "deposit",
  "context_above": "135",
  "context_below": "148",
  "colour": "dark greyish brown",
  "composition": "sand silt",
  "description": "loose dark greyish brown sand silt with frequent black fragments of pottery and white frequent shelly material",
  "interpretation": "lower silty loam deposit"
}
```

Recording: Multiple images processor

- Getting all the images in a folder, and the photo registry spreadsheet, with the title of the images to be included in the final report, this script will generate a Word document, with the needed images, in the right dimensions, including the given title and the details.
- Other tasks that it could be automatically sort photos into folders by trench or date based on file names or metadata

Image 8: View of the wall.

Orientation: W, Area: 1021

Image 9: Detail of the wall next to the entry.

Orientation: NES, Area: 1020

Image 1: General view of the exterior wall.

Orientation: NWS, Area: 1012

Image 6: Detail of the wall next to the entry.

Orientation: NWS, Area: 1012

Compilation of multiple tables in a SQL DB

Bray, P. _et al._ (2024) ‘A Catalogue of British Bronze Age Axeheads’. Archaeology Data Service. Available at: https://doi.org/10.5284/1122315.

This file provides a dataset of database data from a dataset list of roughly 8000 Bronze Age British axeheads, alongside associated elemental analyses, isotopic measurements and radiocarbon dates. It integrates several major existing data collection efforts and published catalogues, whilst also providing a basic typology.

- Original excel file has 3 tables: main | radiocarbon | elemental | isotopic
- Each table has the same column ArtefactID that can be used for joins
- Saving in a relational database, allows us to create complex queries in a much easier way:

❖ What is the average lead content per site?

```
execute_query("""SELECT m.SiteName
                , avg(e.Pb) as mean_Pb
                FROM main m
                JOIN elemental e ON m.ArtefactID = e.ArtefactID
                GROUP BY m.SiteName;""")
```

	SiteName	mean_Pb
0	Abbotsbury, 8 West Street	0.26000
1	Abdie (parish)	0.00000
2	Aberdovey and Machynlleth (between)	0.00000
3	Abergele (field SE of, Berth-topic)	0.00000
4	Abernethy	1.45000
..	...	...
756	unprovenanced, West Wales?	0.00000
757	unprovenanced, Wiltshire (or Malborough)?	0.00000
758	unprovenanced, Yorkshire	1.88900
759	unprovenanced, Yorkshire Moors	0.00000
760	unprovenanced, Yorkshire?	0.01626

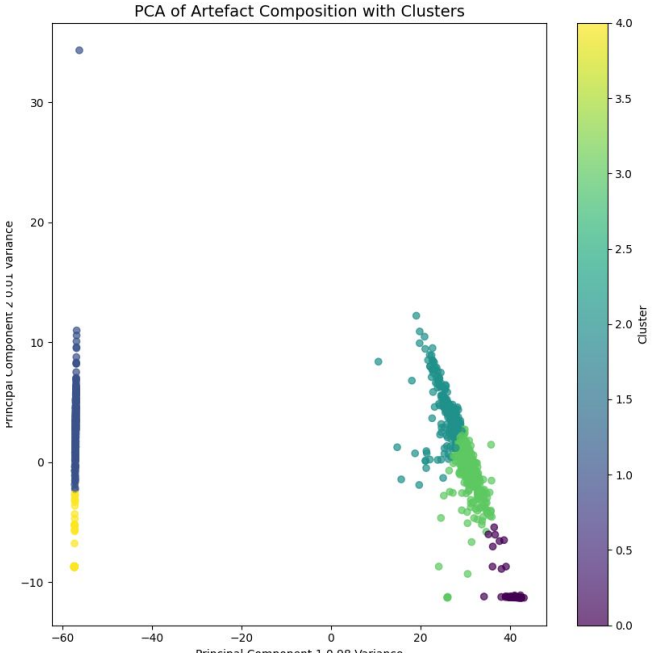
[761 rows x 2 columns]

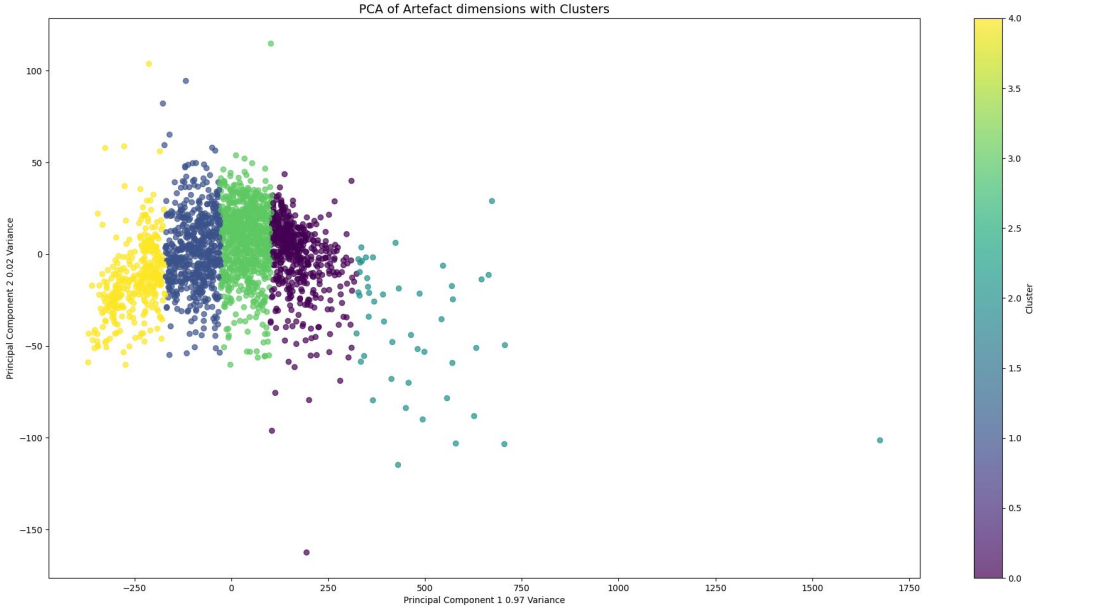
❖ What is the average Pb208/Pb206 ratio for each object type?

```
execute_query("""SELECT m.Object
                , avg(i.Pb208_Pb206) as mean_Pb208_Pb206
                FROM main m
                JOIN isotopic i ON m.ArtefactID = i.ArtefactID
                GROUP BY m.Object;""")
```

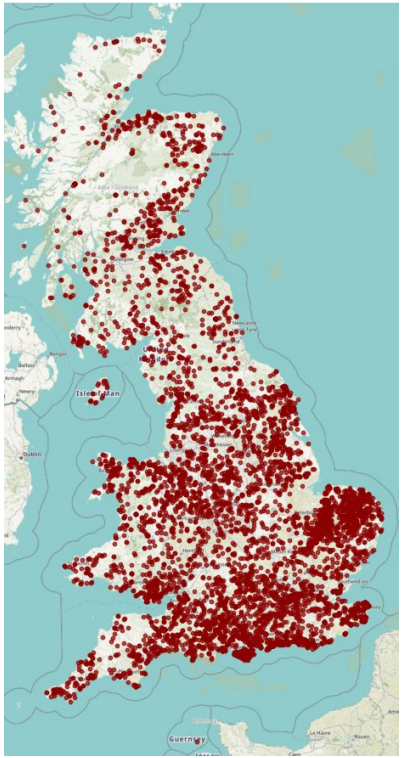
	Object	mean_Pb208_Pb206
0	adze-chisel (palstave)	2.109500
1	axehead (flanged)	2.086864
2	axehead (flanged/palstave)	2.087870
3	axehead (flat)	2.052618
4	axehead (flat/flanged)	2.094059
5	axehead (palstave)	2.097950
6	axehead mould (palstave)	2.084795

Analysis: PCA + Kmeans Clustering

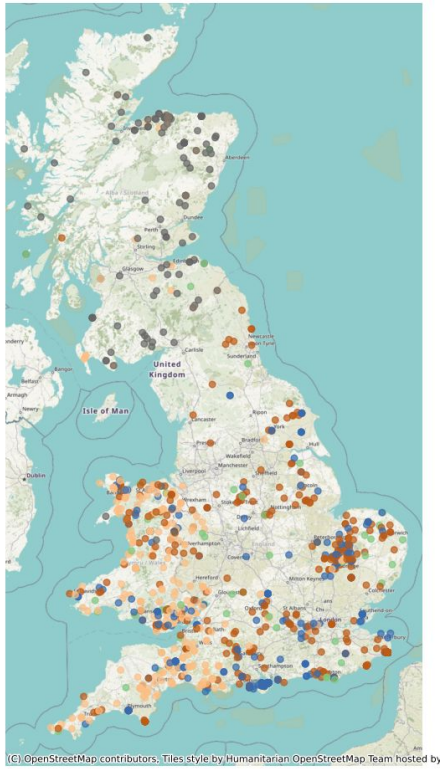




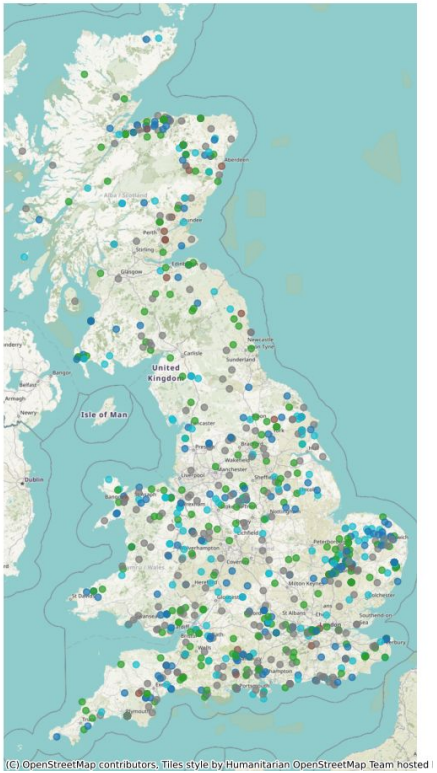
Analysis: Maps



PCA of Artefact Composition with Clusters on Map



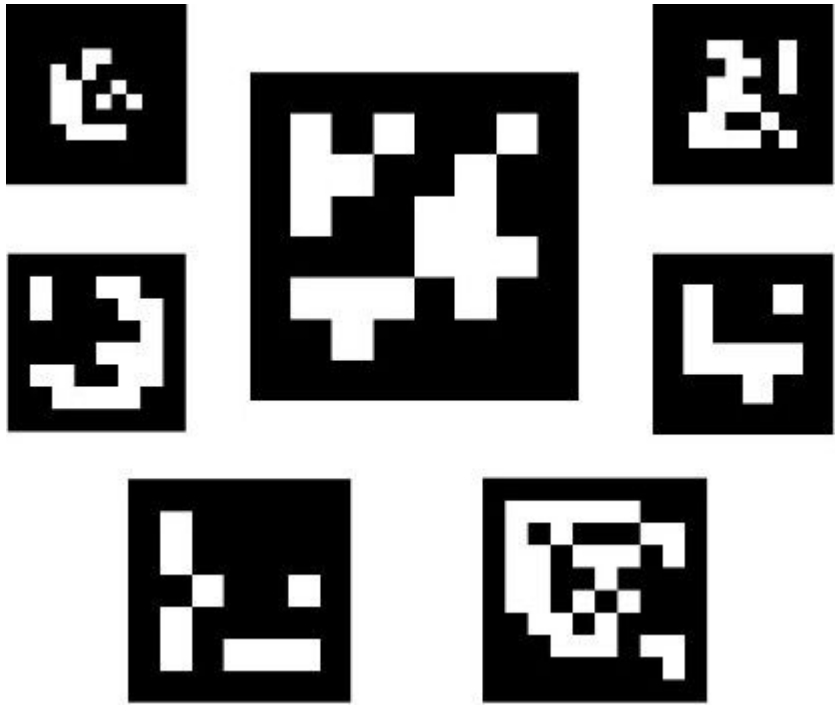
PCA of Artefact dimensions with Clusters on Map



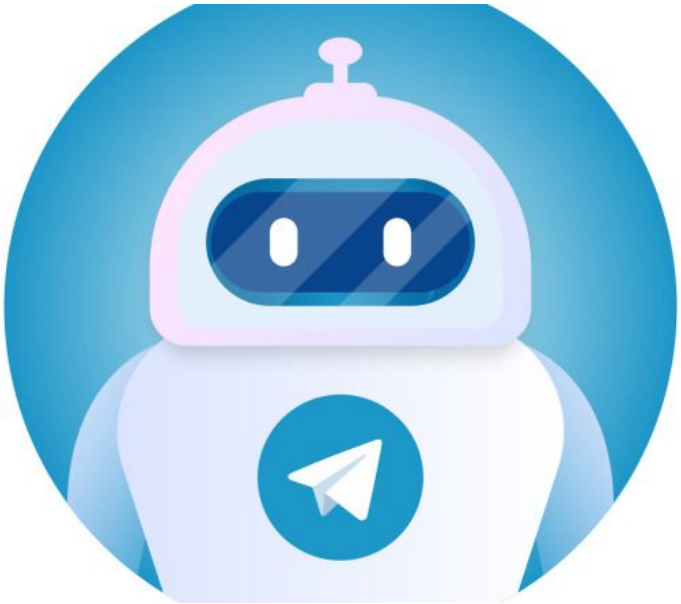
Other possible ideas

Automatic image labelling

- Use of Aruco labels to identify elements on the imagen (like the photo board)
- Include them into scales or poles, for automatic distortion correction and scaling



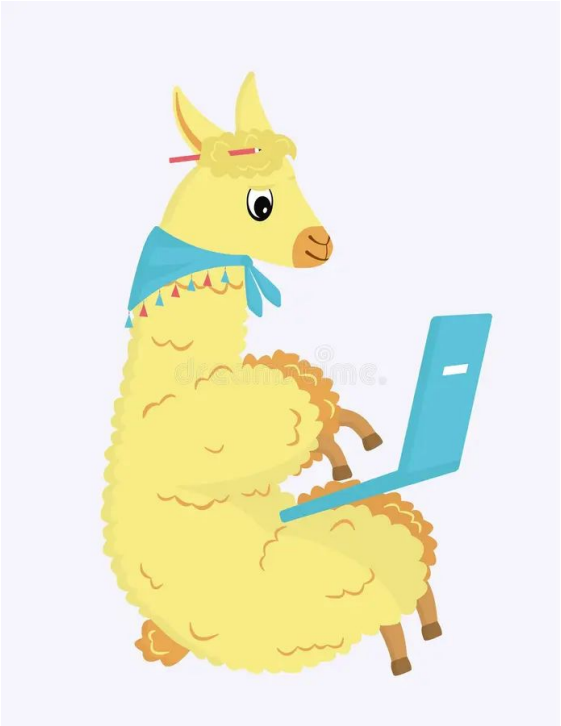
Telegram channels and bots for field recording



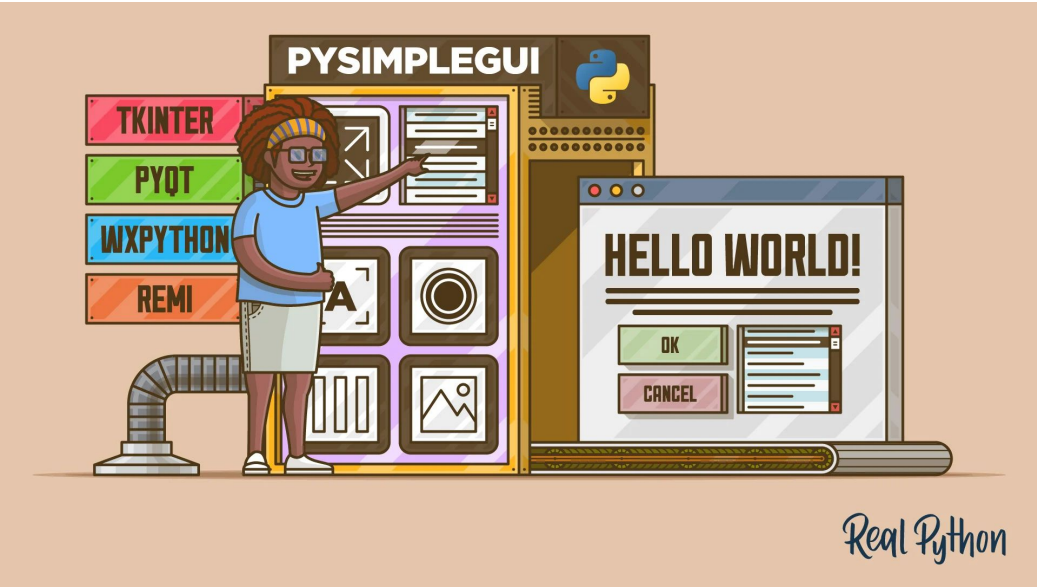
- Use of telegram channels for field staff to send data directly to the office.
- Building of bots for guided context recording
- Backend automatic data processing

Use of LLM for reporting and DB writing

- Automatic workflows for reporting writing using Local or Cloud based LLMs



Build of GUIs and Web-GUIs



- Multiple libraries
- Some limitations
- Web applications can be used remotely (but internet connection is needed)

Creation of open source community tools

- Open science
- Collaboration and standardisation
- Bridging the gap between academic and commercial archaeology

Join the conversation in the Python Glasgow Room in [Matrix]

https://www.linkedin.com/in/jfpalomeque/