

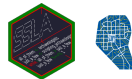
Can you code this?

W4 CAA Scripting Languages Hackathon I

Clemens Schmid¹, Martin Hinz², Carolin Tietze³

¹ Römisch-Germanisches Zentralmuseum Leibniz-Forschungsinstitut für Archäologie: Mainz, Rheinland-Pfalz, clemens@nevrone.de; ² Institut für Archäologische Wissenschaften, Universität Bern, martin.hinz@iaw.unibe.ch; ³ Institut für Klassische Altertumskunde, Christian-Albrechts-Universität zu Kiel, ctietze1991@gmail.com

April 23, 2019



- ▶ A **repository** with all information and data for this workshop is available at https://github.com/sslarch/caa2019_hackathon
- ▶ 2-4 **groups** are formed on site according to framework preference and skill levels (*Unconference style*).
- ▶ We have until **4pm** to work together. Breaks can be taken as you wish, but a joint lunch break at 12:30pm might be a good idea.
- ▶ It is not necessary to complete all possible tasks. Work as far as you can go. Most tasks are independent of each other, so you can skip boring ones. This is **NOT** an exam.
- ▶ All results must be submitted in one **reproducible report** with all code and plots. This can be rendered from IPython Notebook, Rmarkdown, Latex, etc. or compiled manually. Ideally the report is submitted as a Pull Request to the hackathon repository on github. The file(s) should be added to the *reports* directory.
- ▶ The organizers of this workshop are available for questions and advice. They are able to assist you with problems as far as they are familiar with your toolset.

The data for this exercise — Michelsberg| --- are taken from the R package archdata| (Carlson/Roth 2018). It is a features by types table of abundance data on vessel types in archaeological features of the Younger Neolithic Michelsberg Culture from Belgium, France and Germany by Birgit Höhn (2002). The 109 observations/lines represent individual features (pits, ditches, etc.) from sites. For each feature we have information about 42 variables/columns. These include identifiers, classified pottery type counts, phase attribution and site coordinates.

CAA 2019 Check Object Integrity - 47th CAA - 23-27 April 2019

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ 4 ↺ 🔍 ↻

1. Create a table with information about the sites (`site_name|`) and the amount of features (`feature_name|`) documented per sites. Based on this table: Who many sites are represented in the table by more than one feature?
2. Create a table with information about the total material occurrence: How many individual objects of `to3|`, `f4|`, ..., `t1a|` do we have in total across all sites? The table should not just contain the shortend type ids (e.g. `to3|`), but also a human readable type name (eg. Pot/vessel type 3|). Plot an ordered bar chart based on this table.
3. How much artefacts are documented per MBK phase (`mbk_phases|`) in total? Visualise the grouped counts with a time series plot. `mbk_phases|` is an ordinaly scaled time variable. Reduce your data selection to the 6 sites with the most features (cf. Task 1) and draw the same time series for each of these variables (ideally in a plot matrix).
4. Plot a spatial map of sites based on the coordinate information. The map can be interactive. Add a meaningful background layer. The size and colour of the site markers should reflect the occurence of Carinated bowl type| (`kw3|`) at the site.
5. Run a Correspondence Analysis (CA) of features and pottery type variables (`to3|`, `f4|`, ..., `t1a|`). Remove very rare types and features with almost no finds beforehand. Prepare two result scatter plots -- 2D: Dimension1 Dimension2 and point colour according to the MBK phase (`mbk_phases|`). 3D: Dimension1 Dimension2 Dimension3.
6. Run a Correspondence Analysis (CA) of sites (!) and pottery type variables. Use only binary presence-absence information for the CA, not object count. Plot the resulting rank order along Dimension1 or along a fitted principal curve on a spatial map.
7. Calculate cultural distance between features (Euclidian distance, Chi-square distance, ...) based on the pottery type variables and visualize the distance network.
8. Make your own task! What else comes to mind with this kind of data?