

Report on the training school atRium at Brno

September 16-20, 2014

Sara Krubeck

The summer school offered many insights and detailed practical training in two categories of data analyses within the programming environment R, i.e. querying linked open data (LOD) and using as well as visualising spatial data for analyses.

I participated in the training to become methodically equipped in spatial analyses for my PhD project and to gain deeper insights into possibilities of R coding (the latter of which I learned and applied previously during my Master's in Kiel, Germany). Due to the complementary and well structured training sessions, I was able to get to know many new R packages and functions, and I also developed a different perspective on coding than before, when I was mostly using base R. Furthermore, I learned for the first time about ontologies, and the ARIADNE research infrastructure in particular. I believe this may prove very valuable in my future research – a door has been opened.

The training programme consisted of several sessions per day. On the first day, Monday, an introduction to R was given by Peter Tkáč, starting from basic mathematical operations and ending with various ways to plot our practice data (on Bronze Age cups) with the ggplot2 package.

The next day, we then entered the world of linked open data, which was incredibly well explained by Petr Pajdla. In this context, we learned to know Ariadne the Greek mythological figure and ARIADNE the open source data network. The sessions on Tuesday span from theoretical knowledge on Getty AAT Subjects, URIs and RDF to the practical application of code in triple format (first directly in different serializations and then by sparql in R) in order to query the data from DBpedia, Wikidata and, last but not least, ARIADNE.

On Wednesday, Giacomo Bilotti introduced us to the nature and application of spatial data, using both the sf and terra packages for vector and raster data, respectively. After learning how to subset and join the spatial datasets for different uses, we started creating and customizing maps and mapping our data points. Two very useful functions from Giacomo's code are st_buffer() and, not just for spatial analyses, grepl(). After the sessions on Wednesday, the middle of the training week was celebrated with a keynote talk by Michael Kempf and Margaux Depaermentier on combined stable isotope and spatial analyses in a case study of the Neolithic Carpathian Basin. This was followed by a social get-together enabling us participants to engage in networking and discussing the impact of the training sessions on our individual research project. It was quite enlightening.

The atRium summer school proceeded on Thursday with more aspects on spatial analysis, this time instructed by Michael Kempf, and focused on the tool of complete spatial randomness (CSR) as a comparative background for conducting point pattern analyses (PPA). We applied Ripley's K for inhomogeneous data (Kinhom() function) and discussed the resulting density graphs. With the rhohat() function we then included a covariate from underlying spatial raster data to assess the explanatory level of this covariate for the point pattern, which was then visualised as an intensity graph. For both analytical steps, we also learned to modify the kernel radius and thus adapt the model to our hypothetical research question.

The final day of the training school, Friday, started with an overview and reminder of important aspects in working with R scripts, e.g. to construct the code in a way that is adjusted for sharing it with other researchers, for instance in defining and pointing to the working directory and in calling functions by :: to avoid confusion. The group of participants then split up in those working with their own respective datasets and a small group of people going over the sparql queries on ARIADNE data with Petr Pajdla once more. I joined the latter group. We constructed and refined a query on coins, discussing every triple of subject, property and object in detail to be able to produce similar code on our own, based on the complex AO-Cat Ontology. In the end, we also mapped the distribution of our selection of artefacts. This repeated session on sparql and the ARIADNE ontology was very useful to consolidate the large amount of new knowledge gained on Tuesday.

I am very grateful to all of the training school teachers and organisers for realising this productive and helpful week of intensive training in R!