

ARCH 3XX: Archaeological Data Science with R

Scott Stetkiewicz

Fall 20XX

Course Description

Archaeological investigations can produce vast amounts of data. From excavation plans and artifact catalogs to survey readouts and architectural records, the range of available data types and potential treatments of these forms of evidence necessitate a heavy reliance on computers. As such, archaeological analyses in the 21st century require a firm grounding in the basic principles of computational data management and quantitative analytical methods. Moreover, as funding bodies increasingly look to web-based outputs as a means of extending research impact beyond academia, the ability to efficiently, aesthetically and interactively present results has become critical to ensuring a project's long term success.

This course takes a hands-on approach to introducing the open-source programming language R to students, using it to probe and visualize complex archaeological data. We will explore the ways in which different types of data are stored, how R accesses and manipulates this information, how to build scripts for basic statistical/geospatial analyses and methods of deploying dynamic interfaces to effectively communicate results. Course readings will focus on how data science methods can be inappropriately applied to research questions, and how to combine multiple strands of analysis into synthetic archaeological assessments. While a background in computer programming/coding would be beneficial, this course is designed for students with no previous computing experience and is therefore not a requirement.

Class Structure

This course will meet twice weekly for 75 minutes per session. In the first meeting (traditional classroom setting) we will discuss assigned readings and explore common methods of managing data in archaeological projects. The second session (the lab practical) will be held in computer lab facilities, allowing us to work together as we become familiar with different packages in R to build our scripts for the final semester project.

Course Goals

In this course, you will:

- Learn the fundamentals of archaeologically-oriented data science techniques in R
- Apply statistical and GIS-based approaches to the analysis of archaeological datasets
- Critically evaluate the merits and shortcomings of common analytical methods
- Deploy fully-functional, interactive computer applications for archaeological analysis

Course Objectives

After successfully completing this course, you will be able to:

- Analyze archaeological data within a computational setting
- Create dynamic visualization platforms to communicate analytical results
- Use insights from iterative output stages to inform research directions
- Present results in an informative and interactive format

Required Texts

All course materials are available online or will be provided in .pdf form:

Drennan, R.D. 2010. Statistics for archaeologists. Springer.

James, G., D. Witten., T. Hastie. & R. Tibshirani. 2013. An introduction to statistical learning. Springer.

Reinhart, A. 2015. Statistics done wrong: The woefully complete guide. No starch press.

Verzani, J. 2014. Using R for introductory statistics. Chapman and Hall/CRC.

Wickham, H. & G. Grolemund. 2016. R for data science: import, tidy, transform, visualize, and model data. O'Reilly Media, Inc.

Assignments and Grading

Attendance – 10%

Participation – 10%

Response Essays:

Good programming does not occur within a vacuum; rather, it forms a vital step in the wider investigative *chaîne opératoire*. Thus, the suitability of statistical tests you choose to employ and directions your research takes must be properly contextualized within the framework of the final project. Four short response essays (~1000 words) will be assigned over the course of the semester to encourage critical reflection on themes of research design and the best data science practices within archaeology. Three of these will be selected at random for marking.

4 short essays (3 graded) – 30%

Semester Project:

Over the first few weeks, toy datasets will be used to highlight course themes and general good practice in the R ecosystem; however, the final project will revolve around constructing a dashboard app based on actual archaeological data. To achieve this, students will be encouraged to liaise with members of staff early in the semester who either have ongoing research projects or extant databases in need of analysis. Online data repositories are also available, though students should make every effort to establish connections with staff. These collaborations will provide a tangible connection between course objectives and actual research priorities to students, while also underscoring the vagaries of real world data analysis.

Project proposal – 5%

Static dashboard sections – $5\% \times 3 = 15\%$

Final dynamic project dashboard – 30%

Class Schedule

Week 01, 08/15 - 08/19: Introduction to Archaeological Data Science

Class Topic: What is data science?

Lab Topic: The RStudio interface, part I

Week 02, 08/22 - 08/26: R Basics

Class Topic: Objects, vectors & other data types

Lab Topic: The RStudio interface, part II

Week 03, 08/29 - 09/02: Archaeological data types

Class Topic: What types of data does archaeology produce?

Lab Topic: Import/export functionality

Assignment: Short Response Essay #1 Due

Week 04, 09/05 - 09/09: Analysis of archaeological data

Class Topic: How do archaeologists analyze data?

Lab Topic: `lm` and plotting functions

Week 05, 09/12 - 09/16: Project Design

Class Topic: Data management strategies and workflow

Lab Topic: Github and `rmarkdown`

Week 06, 09/19 - 09/23: The grammar of graphics and `ggplot2`

Class Topic: What are aesthetics? How to layer calls

Lab Topic: `ggplot2`

Assignment: Project Proposals Due

Week 07, 09/26 - 09/30: Data wrangling in the tidyverse, part I

Class Topic: Cleaning and “tidying” data

Lab Topic: `tidyr` and `dplyr`

Assignment: Short Response Essay #2 Due

Week 08, 10/03 - 10/07: Data wrangling in the tidyverse, part II

Class Topic: “Magrittr” and the concept of %>%

Lab Topic: tidyr and dplyr

Assignment: Static Dashboard Script #1 Due

Week 09, 10/10 - 10/14: Using R for GIS, part I

Class Topic: Shapefiles

Lab Topic: leaflet

Week 10, 10/17 - 10/21: Using R for GIS, part II

Class Topic: Raster layering

Lab Topic: rgdal

Assignment: Short Response Essay #3 Due

Week 11, 10/24 - 10/28: Building scripts

Class Topic: Debugging, error checks and unit testing

Lab Topic: RStudio break points

Assignment: Static Dashboard Script #2 Due

Week 12, 10/31 - 11/04: shiny, part I

Class Topic: Fundamentals of reactivity

Lab Topic: shiny

Assignment: Short Response Essay #4 Due

Week 13, 11/07 - 11/11: shiny, part I

Class Topic: Parameterization

Lab Topic: shiny

Assignment: Static Dashboard Script #3 Due

Week 14, 11/14 - 11/18: shinydashboard

Class Topic: Dashboard structures

Lab Topic: shinydashboard

Week 15, 11/21 - 11/25: App Deployment and Web Hosting

Class Topic: Venues for data, and who has access?

Lab Topic: shinyapps.io