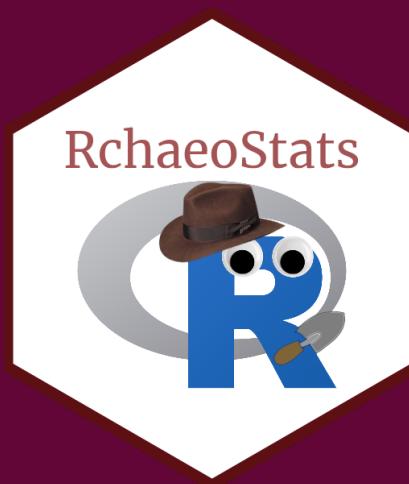


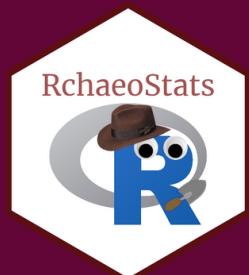
RchaeoStats

Open, online teaching materials for learning statistics with
R



What is it?

rchaeology.github.io/RchaeoStats



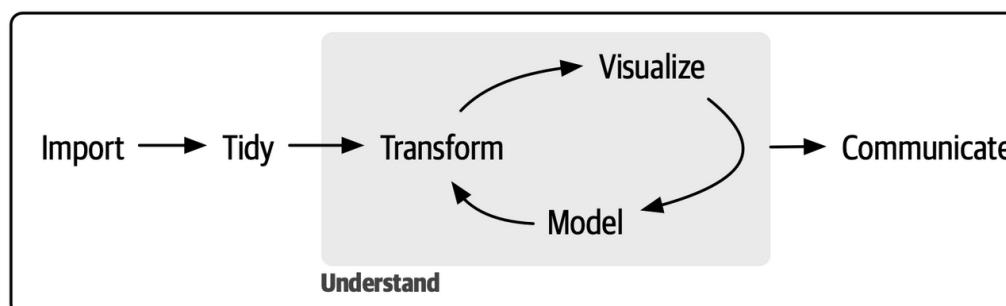
- Course Overview
- Assignments
- Materials
 - Core
 - 00 - Getting Started
 - 01 - R Basics
 - 02 - An Example Workflow
 - 03 - Getting Organised
 - 04 - Cleaning Data
- EDA
- 05 - Visualising Data
- 06 - Transforming Data
- Extensions
- 08 - Communicating Results
- Slides

R and Statistics for Archaeologists

Course overview

The teaching materials on this site are intended for teaching Statistics and R programming workshops to archaeologists.

This course follows the [tidyverse philosophy](#) of understanding and analysing your data through a repeated cycle of transforming, visualising, and modelling.



Program

Image taken from the [R4DS book](#)

Upcoming and past workshops can be found here:

Course modules

On this page

- [Course overview](#)
- [Course modules](#)
- [Course outcomes](#)



Motivation

Asked to give workshop Deutsches Archäologisches Institut

Many (not all) of the archaeological teaching materials are focused on specific applications and University course materials

(nice comprehensive list: [The didactic map of computational archaeology](#))

Something I've wanted to do for a while!

Motivation

R seems to be popular in Archaeology

Table 4: Platforms and programming languages used by open archaeology projects

Platform	n	p
R	200	68.5%
Python	43	14.7%
QGIS	15	5.1%
Mobile app	7	2.4%
MATLAB	6	2.1%
ArcGIS	3	1.0%
LibreOffice Calc	3	1.0%
Microsoft Excel	3	1.0%
Blender	2	0.7%
Open Data Kit	2	0.7%
Other	8	2.7%

Inspiration

Good enough practices in scientific computing (Wilson et al. 2017)

- data management
- project organisation
- collaboration (project portability)

R for Data Science textbook

Data Carpentry - R for Social Scientists

Teaching philosophy

Not meant to produce programmers and statisticians

Meant to enable researchers to confidently and
reproducibly do their work in R

Programming concepts sparsely sprinkled throughout (as
needed)

Teaching philosophy

Code along

- can't learn a programming language without doing

“Time to first plot” - **Mine Cetinkaya-Rundel**

- early ‘wins’
- keep learners motivated

Frequent formative assessments (a.k.a exercises)

- helps manage the pace
- feedback on comprehension (workshops)

Teaching philosophy

tidyverse ecosystem mixed with some base R

- gentlest entry to R (opinion!)
- best-suited for data visualisations and analyses (fact-ish...)
- best illustrations (fact!)

Teaching philosophy

tidyverse ecosystem mixed with some base R

- gentlest entry to R (fact!)
- best-suited for data wrangling and analyses (factish...)
- best illustrations (fact!)

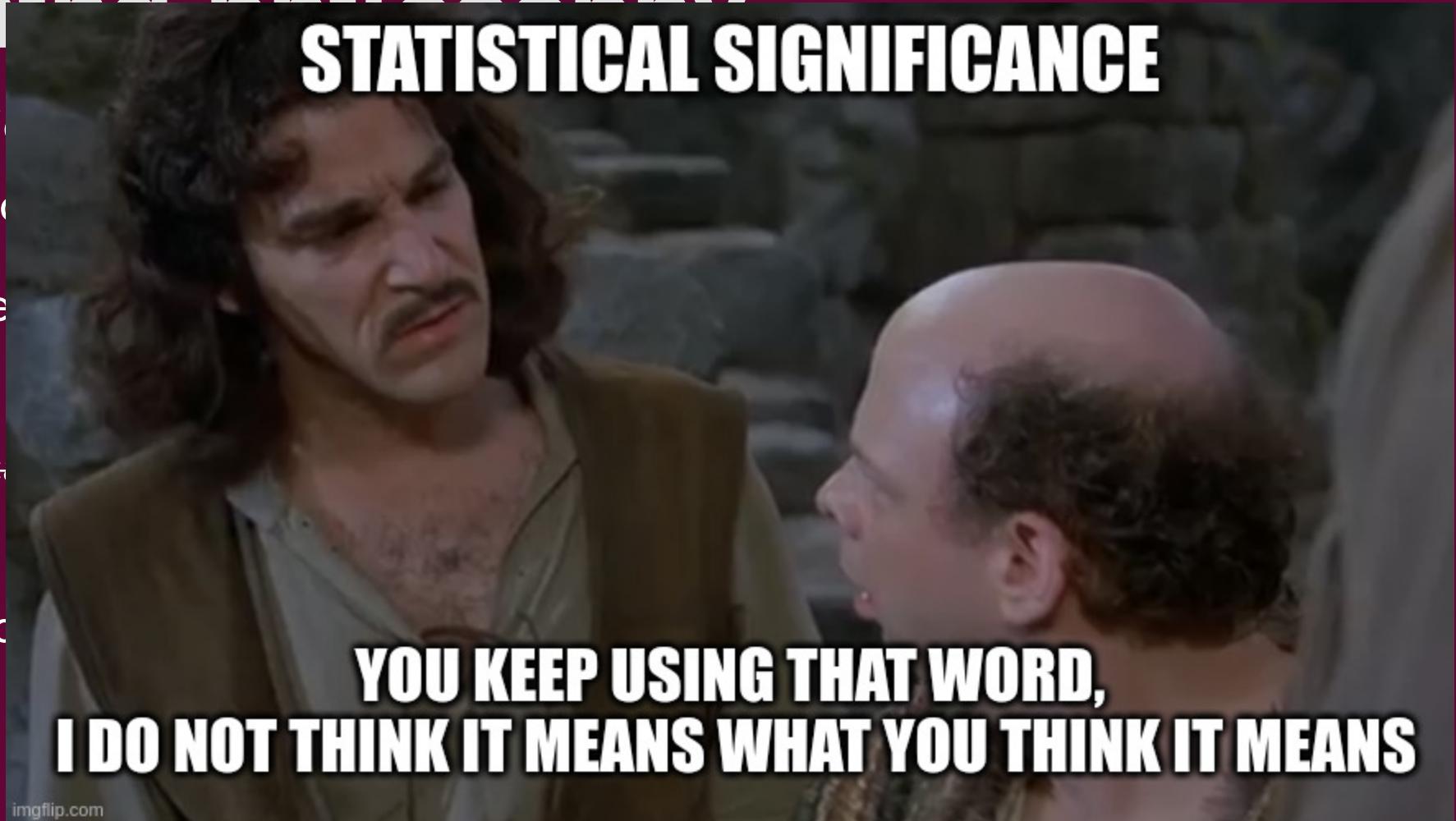


Teaching philosophy **STATISTICAL SIGNIFICANCE**

EDA > statistics

Teach statistics

- general linear models
- p-values...
 - statistical significance
 - NHST
 - continuous vs discrete
- confidence intervals
- effect size
- point/parameter estimates
- practical implications (actual significance)



Teaching philosophy

Live coding during workshops

- mirror learners' environment (RStudio with no customisation)
- what, how, why, repeat.
- explain errors (both intentional and unintentional)

Notes from workshops are hosted on  GitHub and generated during the workshop with `gitautopush`
Central repo with new branch for each workshop

Modularity

Make the materials as flexible and customisable as possible:

- duration of course/workshop
- target audience of workshop (beginners, intermediate, advanced)
- field of research (bioarchaeology, neolithic, paleobotany, etc.)

Modularity

Current modules

Basics

- R basics
- example workflow
- Project organisation
- Data cleaning

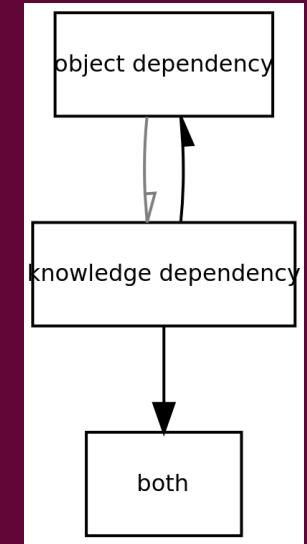
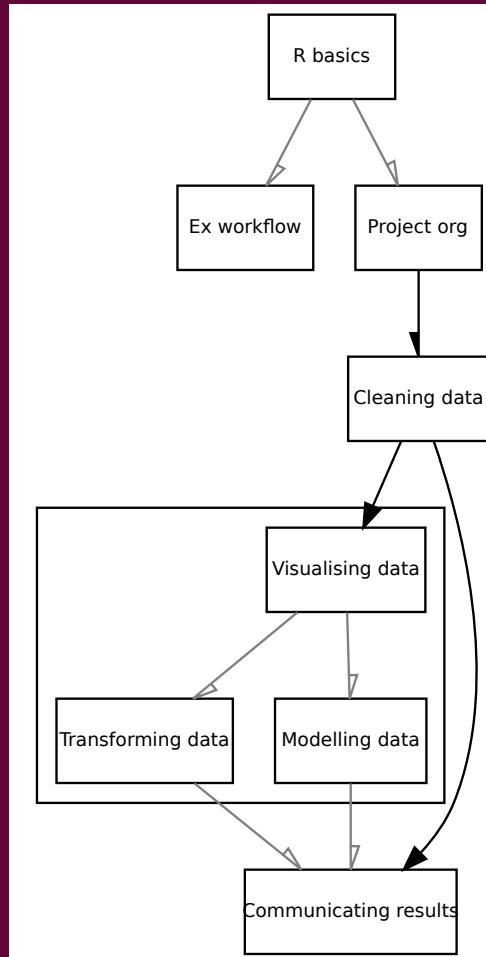
Exploratory Data Analysis (EDA)

- Visualising data
- Transforming data

Communication

- Quarto document

Modularity



Modularity

Beginners

Brief workshop (ca. 2 hours)

- R basics + Example workflow

1-day workshop (max. 6 hours)

- Basics

2-day workshop

- Basics + EDA
- assignment 1

4-day workshop

- Basics + EDA + modelling + communicating

Advanced users

1-day workshop (max. 6 hours)

- cleaning + modelling (or topic-specific)

2-day workshop

- cleaning + modelling + topic-specific
- assignment 1 + 2

Materials

Slides

- brief context for R(Studio) and Quarto
- statistical concepts

Code-along materials

- step-by-step code with explanations
- used by the instructor while teaching
- can also be used for self-study

Materials

sheep-data.csv

Sheep astraguli data Mediterranean Iron Age.

The contribution of Mediterranean connectivity to morphological variability in Iron Age sheep of the Eastern Mediterranean Sierra A. Harding, Angelos Hadjikoumis, Shyama Vermeersch, Roee Shafir, Nimrod Marom. bioRxiv 2022.12.24.521859; <https://doi.org/10.1101/2022.12.24.521859>

Accessed from: nmar79. (2023). nmar79/Med_Sheep_Astragals: v0.1 (v0.1). Zenodo. <https://doi.org/10.5281/zenodo.10276147> (sheep_specinfo_20230824.csv)

Modifications: removes missing values and variables that can be calculated from existing variables

Materials

`mortuary-data.xlsx`

Burial data from northeastern Taiwan ranging from the Iron Age through the European colonization period.

Li-Ying Wang & Ben Marwick, (2021). Compendium of R code and data for “A Bayesian networks approach to infer social changes from burials in northeastern Taiwan during the European colonization period”. Online at <https://osf.io/xga6n/>

Accessed from: <https://osf.io/zem9p> (Kiwulan_Burials.xlsx - burials sheet)

Modifications: removes some variables that need cleaning (reduce cleaning complexity)

Assignments

Assignment 1: Case Study

Finding, importing, cleaning, and exploring/analysing a dataset.

1.1 Finding, importing, and cleaning

1.2 Exploring

1.3a Modelling (if modelling module is taught)

1.3b Communicating (if communicating module is taught)

Assignments

Assignment 2: Peer feedback

Can someone else reproduce your analysis?

Participants are paired up and need to reproduce each other's work

Each participant produces a CODECHECK-style report

- problems encountered
- steps to solve problems

Feedback is incorporated into own project

If Git/GitHub is taught, this will be done via Git

Technical details

Site is built with Quarto

R environment captured with the  **renv** R package

Hosted on GitHub Pages

- accessible via rchaeology.github.io/RchaeoStats
- automated with GitHub Actions

Source code at github.com/rchaeology/RchaeoStats

- releases archived on Zenodo
- versioning with CalVer

F - Materials archived on Zenodo DOI 10.5281/zenodo.13983363

A - Accessible online at rchaeology.github.io/RchaeoStats

I - Quarto and R (dependencies captured with  **renv**)

R - Open, permissive licenses for materials  and data

Roadmap

Development of additional modules

EDA

- correlations

Modelling Data (in progress)

Communication

- Quarto manuscripts (with extensions), presentations
- Quarto parametrised reports

Better integration between the statistics and coding

To **tidymodels** or not to **tidymodels**?

Research-specific modules

- age-at-death and sex estimations
- tidy dental data

Version control and collaboration

- ⚡ Git and GitHub

What now?

Need community to contribute topic-specific modules
(other contributions are welcome)

- dendrochronology
- radiocarbon dating (and others)
- more

More iterations to improve existing materials

Acknowledgements



DEUTSCHES
ARCHÄOLOGISCHES
INSTITUT

Also thanks to early test subjects at Österreichische
Archäologische Institut

Thank you!

Join us at UnArchaeology.nl 7th (and 8th) November in Leiden!