ECOL 596

Practical and Reproducible Data Science

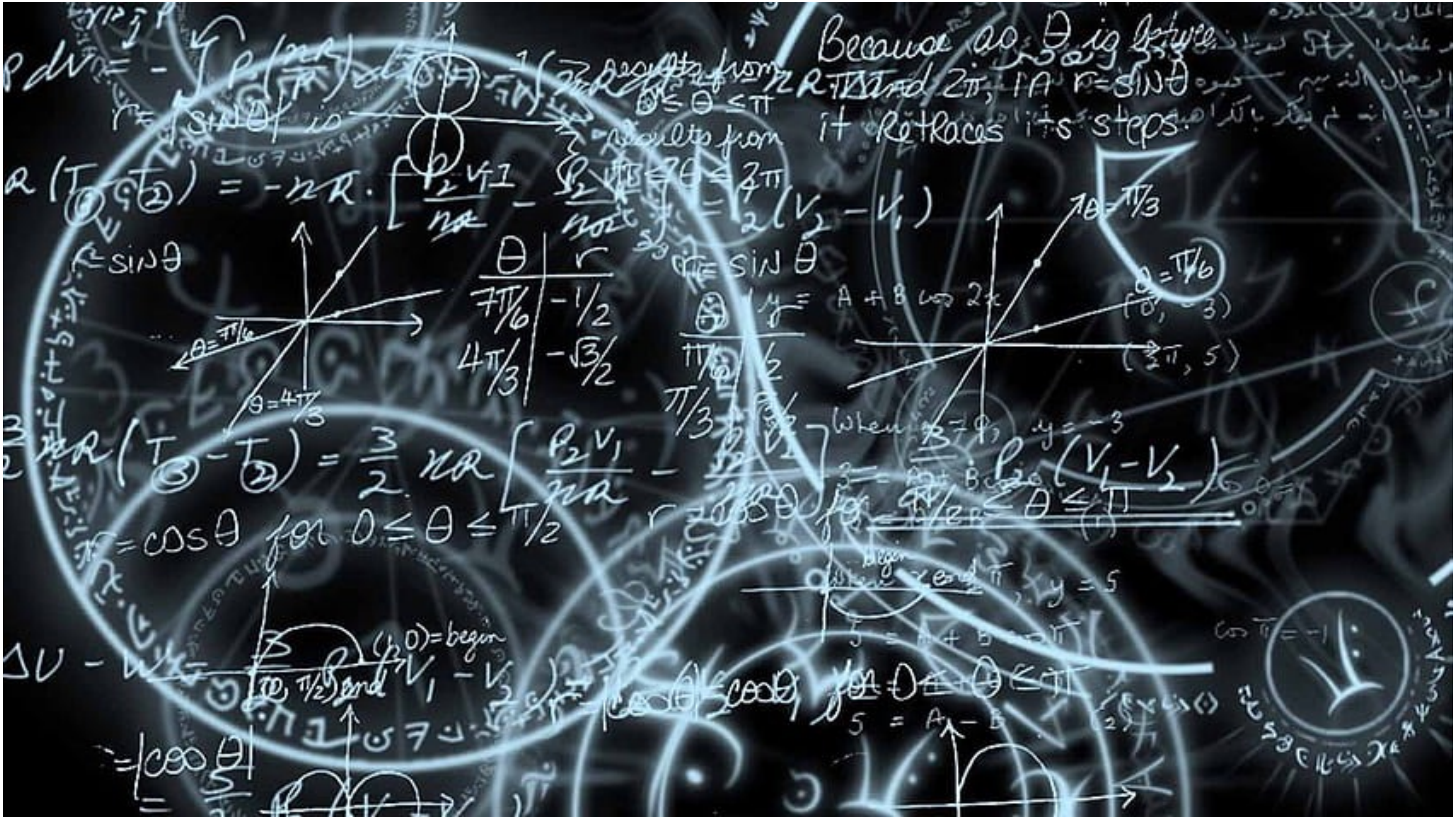
for Ecology and Evolutionary Biology

What is “data science?”

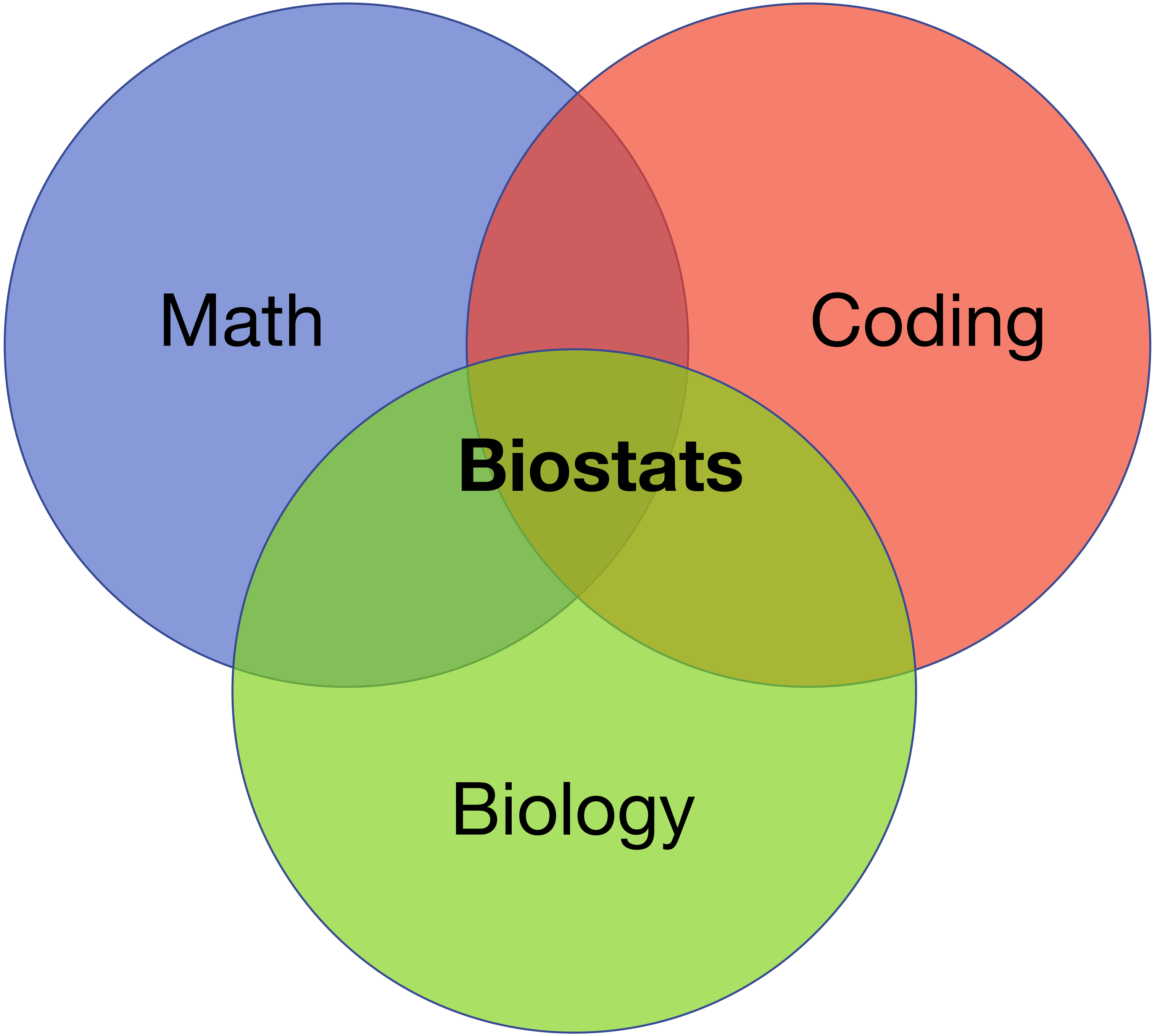
a.k.a. biostatistics, biometry, analysis

ALGORITHM
SOLVING
ANALYSIS
STRUCTURE
PROCESS
DATA SCIENCE

Input Layer
Multiple hidden Layers
Output Layer


Because as θ is between
it Retraces its steps
r = sin θ
θ = π/3
θ = π/6
(2π, 5)
θ | r
7π/6 | -1/2
4π/3 | -√3/2
θ = 4π/3
θ = π/6
y = A + B cos 2x
When
y = -3
y = 5
r = cos θ for 0 ≤ θ ≤ π/2
(0, 0) = begin
(0, π/2) end
cos π = -1
begin
end

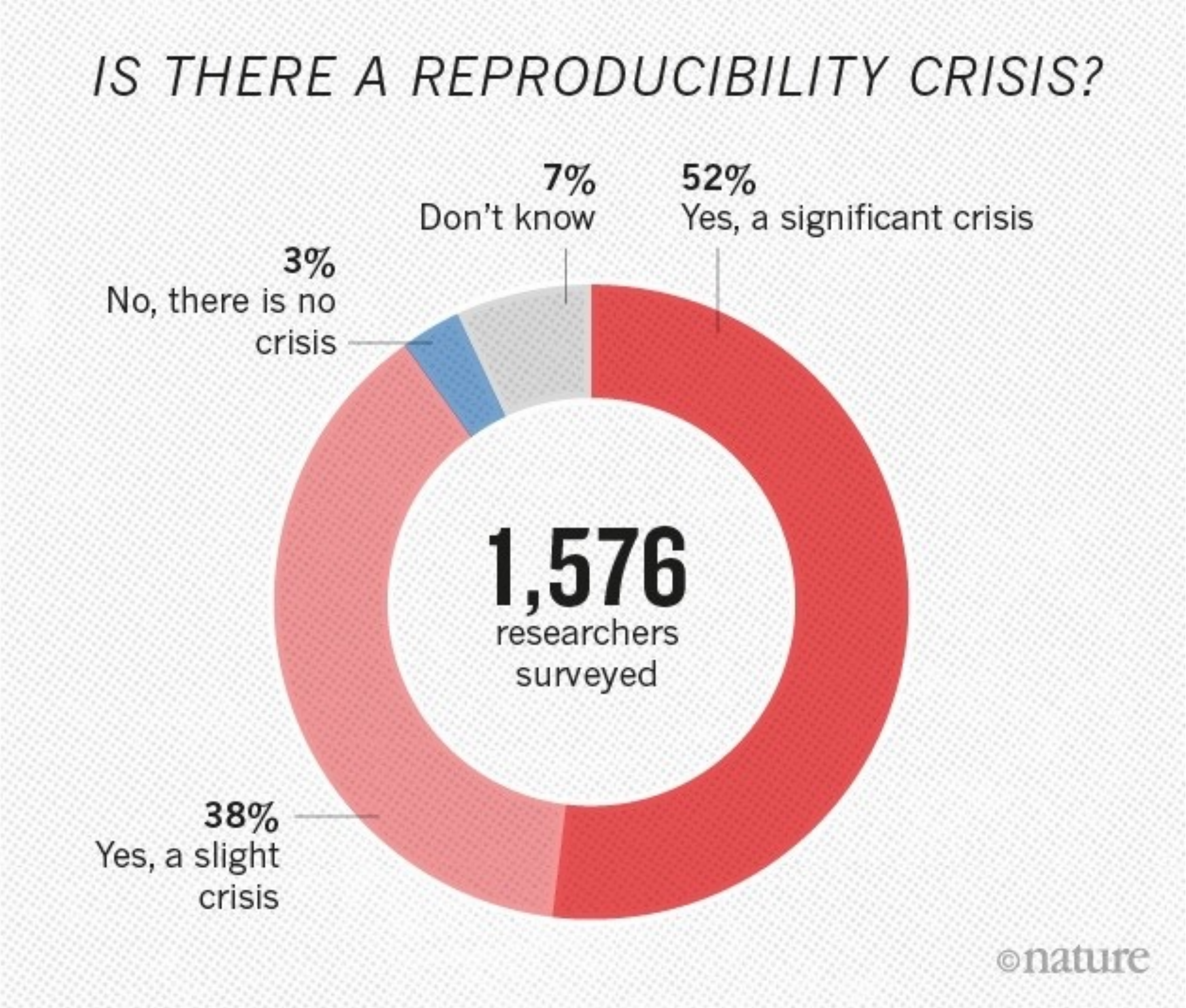
Data science helps us extract the signal from the noise


Math
Coding
Biostats
Biology

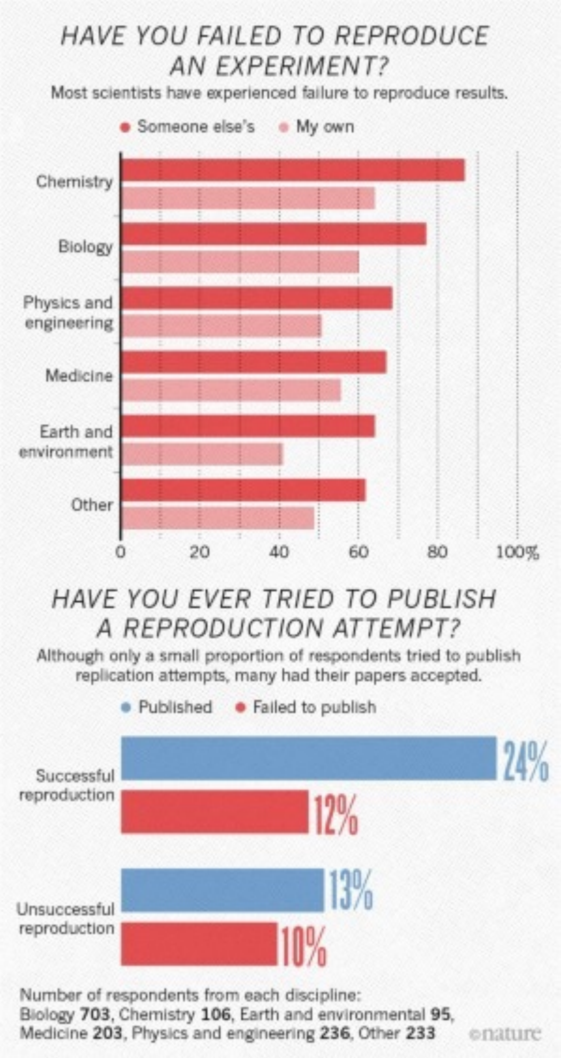
Why do EEB grad students need data science?

- ✓ Design good studies
- Interpret your data
- Test hypotheses
- Extrapolate/predict
- Have confidence in your results

The “reproducibility crisis”



IS THERE A REPRODUCIBILITY CRISIS?



HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.

HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.

Number of respondents from each discipline: Biology **703**, Chemistry **106**, Earth and environmental **95**, Medicine **203**, Physics and engineering **236**, Other **233**

Intentional and unintentional errors

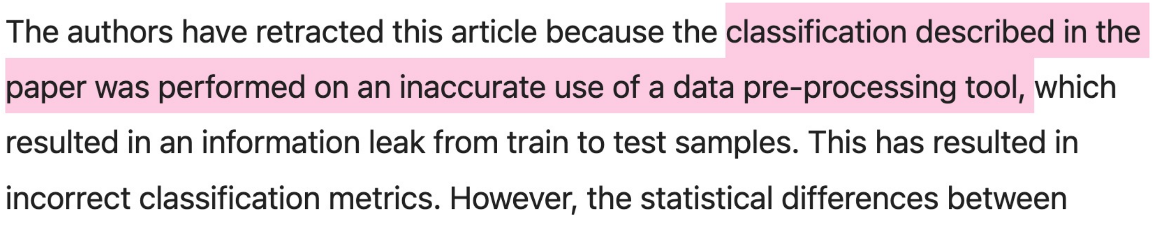
SCIENCEINSIDER | PEOPLE & EVENTS

Embattled spider biologist resigns university post

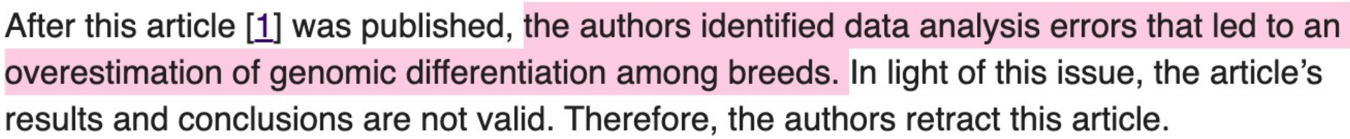
Ecologist Jonathan Pruitt agrees to leave McMaster University, but reasons are still undefined

12 JUL 2022 · 5:30 PM · BY ELIZABETH PENNISI

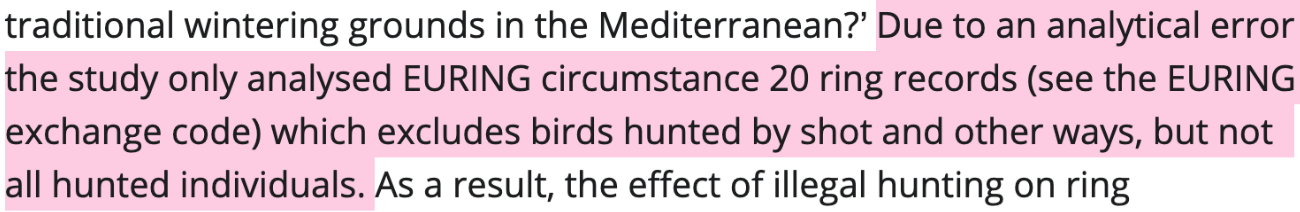
Retraction Watch Database:
1,179 retractions “error in analyses”



The authors have retracted this article because the classification described in the paper was performed on an inaccurate use of a data pre-processing tool, which resulted in an information leak from train to test samples. This has resulted in incorrect classification metrics. However, the statistical differences between



After this article [1] was published, the authors identified data analysis errors that led to an overestimation of genomic differentiation among breeds. In light of this issue, the article’s results and conclusions are not valid. Therefore, the authors retract this article.



traditional wintering grounds in the Mediterranean?’ Due to an analytical error the study only analysed EURING circumstance 20 ring records (see the EURING exchange code) which excludes birds hunted by shot and other ways, but not all hunted individuals. As a result, the effect of illegal hunting on ring

Scope of this class

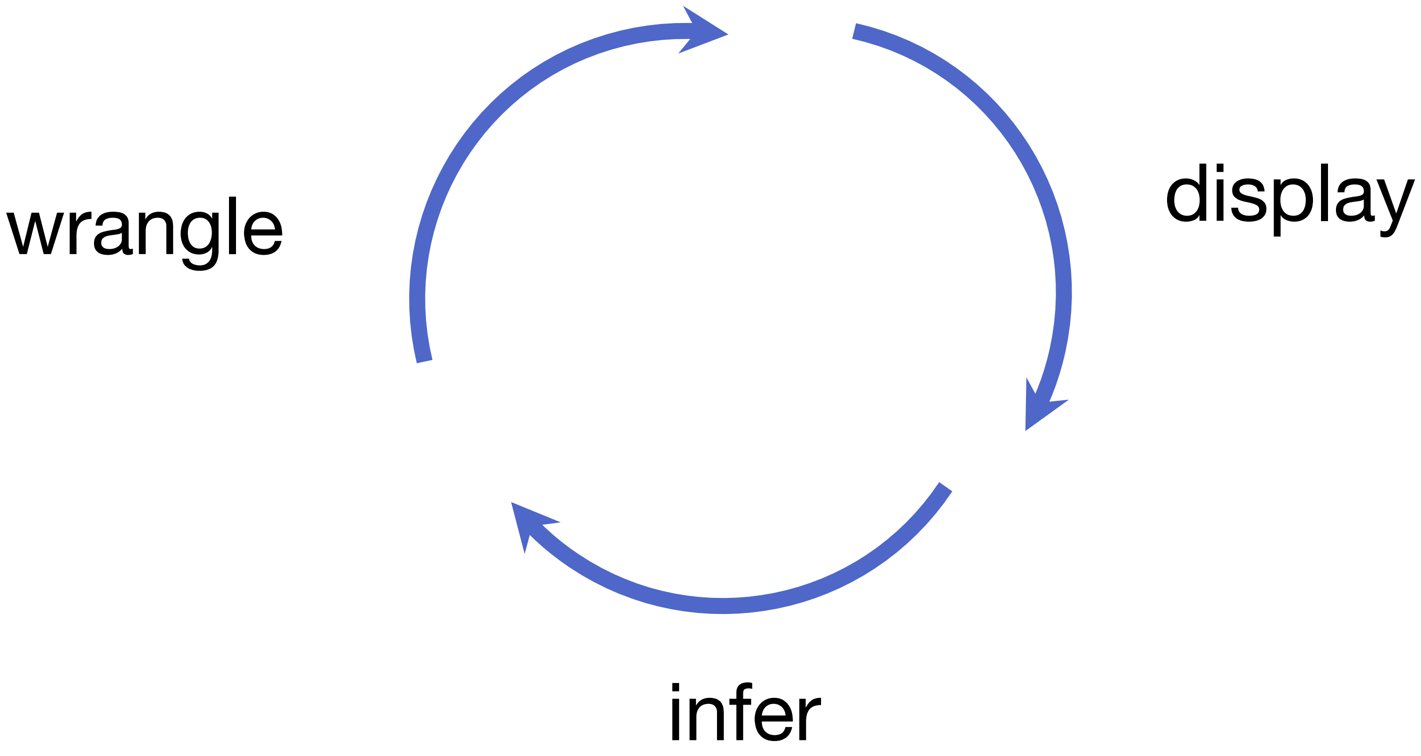
- ✓Good practices for data management
- ✓Coding in R
- ✓Common statistical tests
- ✓**Statistical intuition**
- ✓**How to teach yourself**
- ✓**Topics that will help you**

All of the statistical tests ever

Calculus

Other stats languages

Guiding principle 1: The wheel of analysis

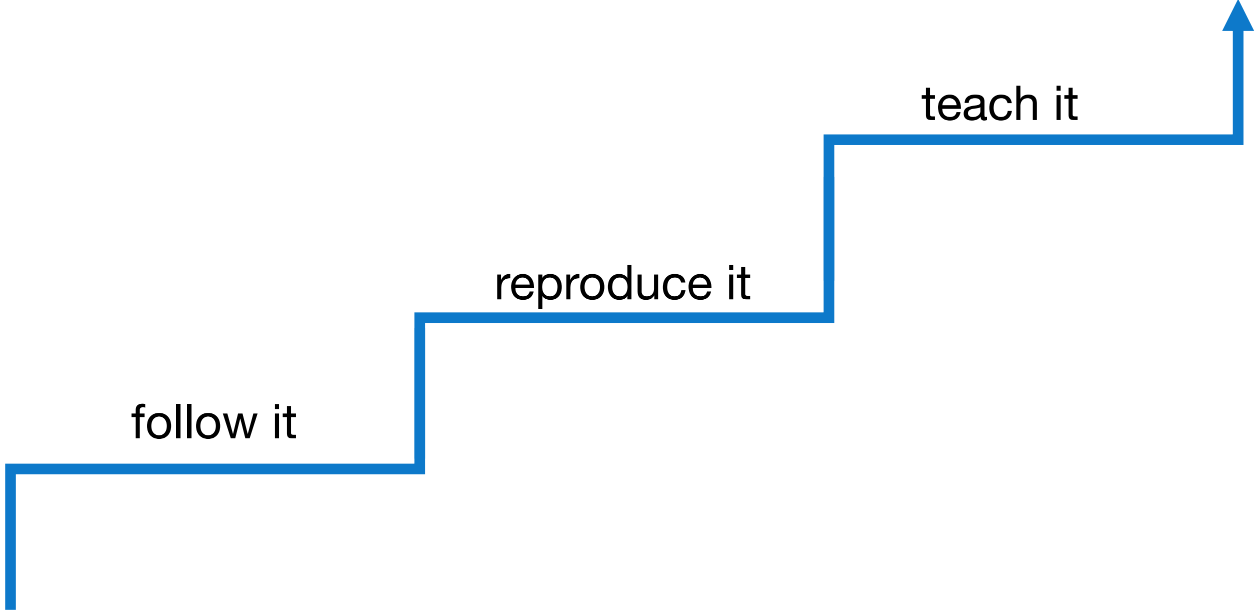


wrangle

display

infer

Guiding principle 2:
You won’t get it until you learn it > 3 times



follow it

reproduce it

teach it