

BIO144

DATA

ANALYSIS

IN BIOLOGY



Today...

- Introduction to the course
- A bit about AI assistants
- RStudio server
- You collect some data
- I demonstrate a data analysis
- Review useful knowledge



Overarching goals of the course

Provide a solid foundation for answering biological questions with quantitative data.

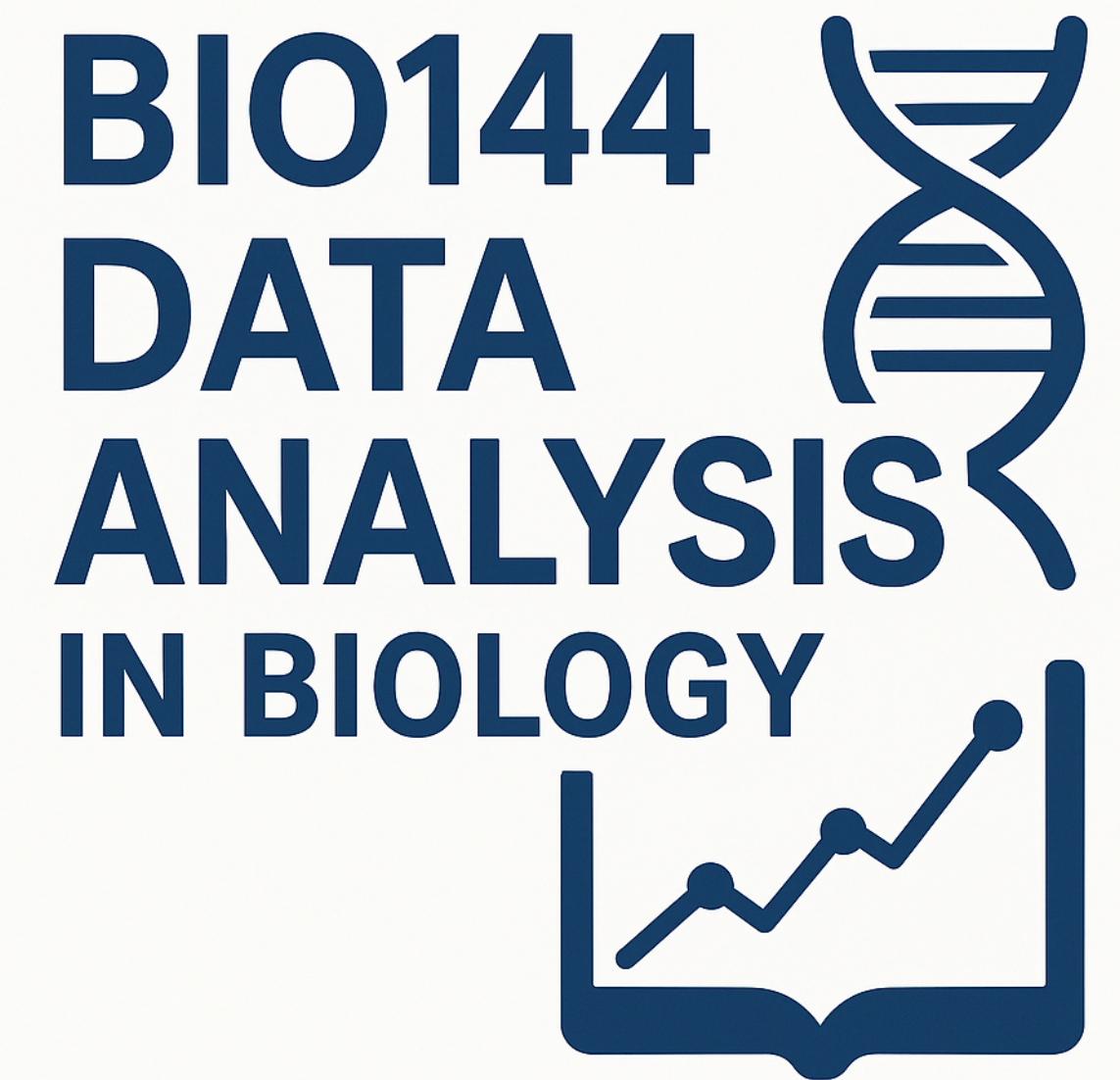
Grow your confidence in your abilities to understand and perform data analysis.

Help you know how to plan research projects so they are easy to analyse.

Excite you about the wonders of data analysis!

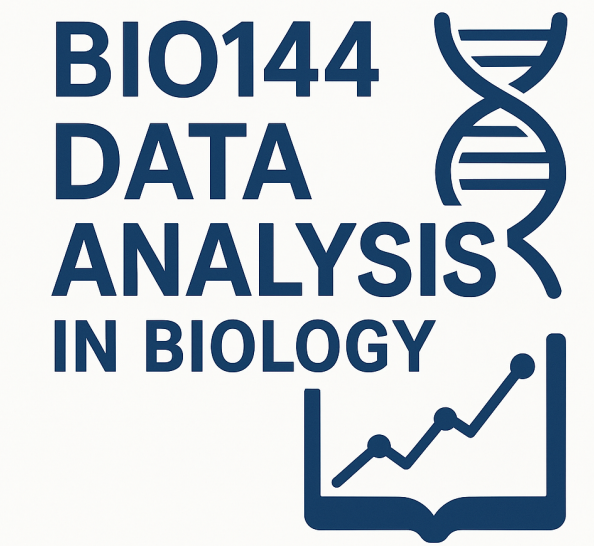
Prepare you for the final exam :)

To have fun!



Why should you care about data analysis?

Only you can answer that question.



Strong feelings of purpose come something being **meaningful to you** and **useful to others**.

Use data to show that a problem exists. Good analysis can make invisible problems visible—and help others take them seriously.

Use data to test solutions. Does an intervention work? Does it work for everyone? Does it have unintended consequences?

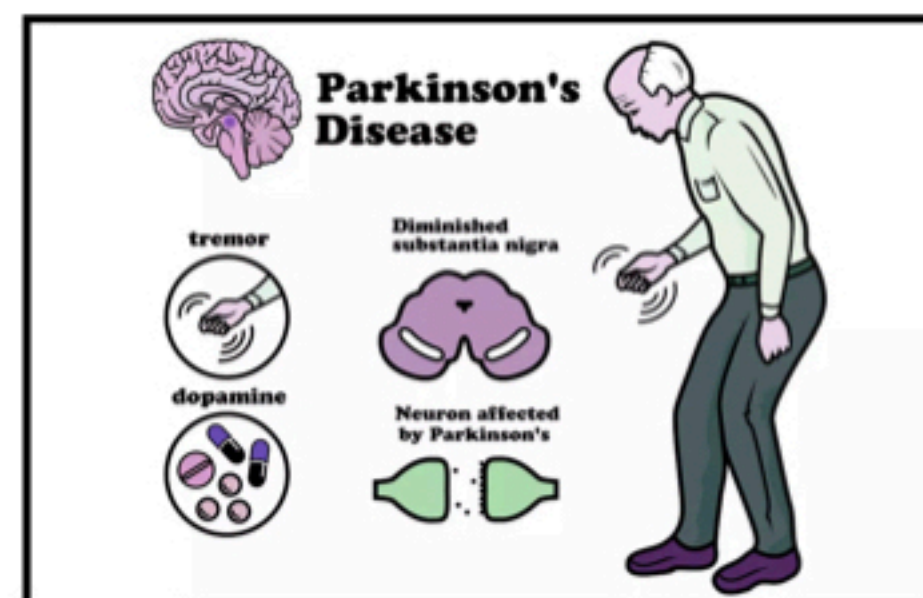
Evaluate evidence that affects people's lives. Showing whether a drug works, doesn't work, or has serious side effects is not abstract, it shapes real decisions.

Question powerful claims. Marketing, policy, and public messaging often rely on selective statistics. You can learn to ask: *What's missing? What's exaggerated?*

Protect others, especially those with less power. From kids targeted by unhealthy products to communities affected by environmental harm, good analysis can act as a form of protection.

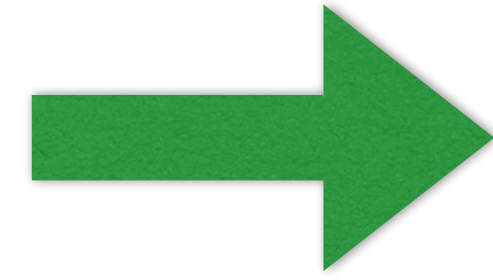
Serve something bigger than yourself. Science, public trust, health, biodiversity, fairness. Data analysis is one way people contribute to their community and society.

Questions you'll work on...

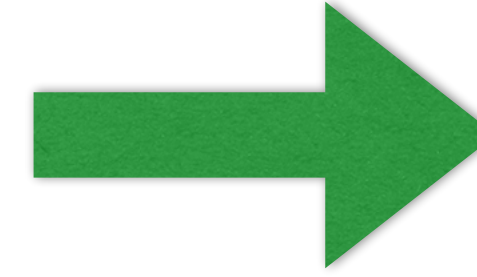




Question
Puzzle
Problem

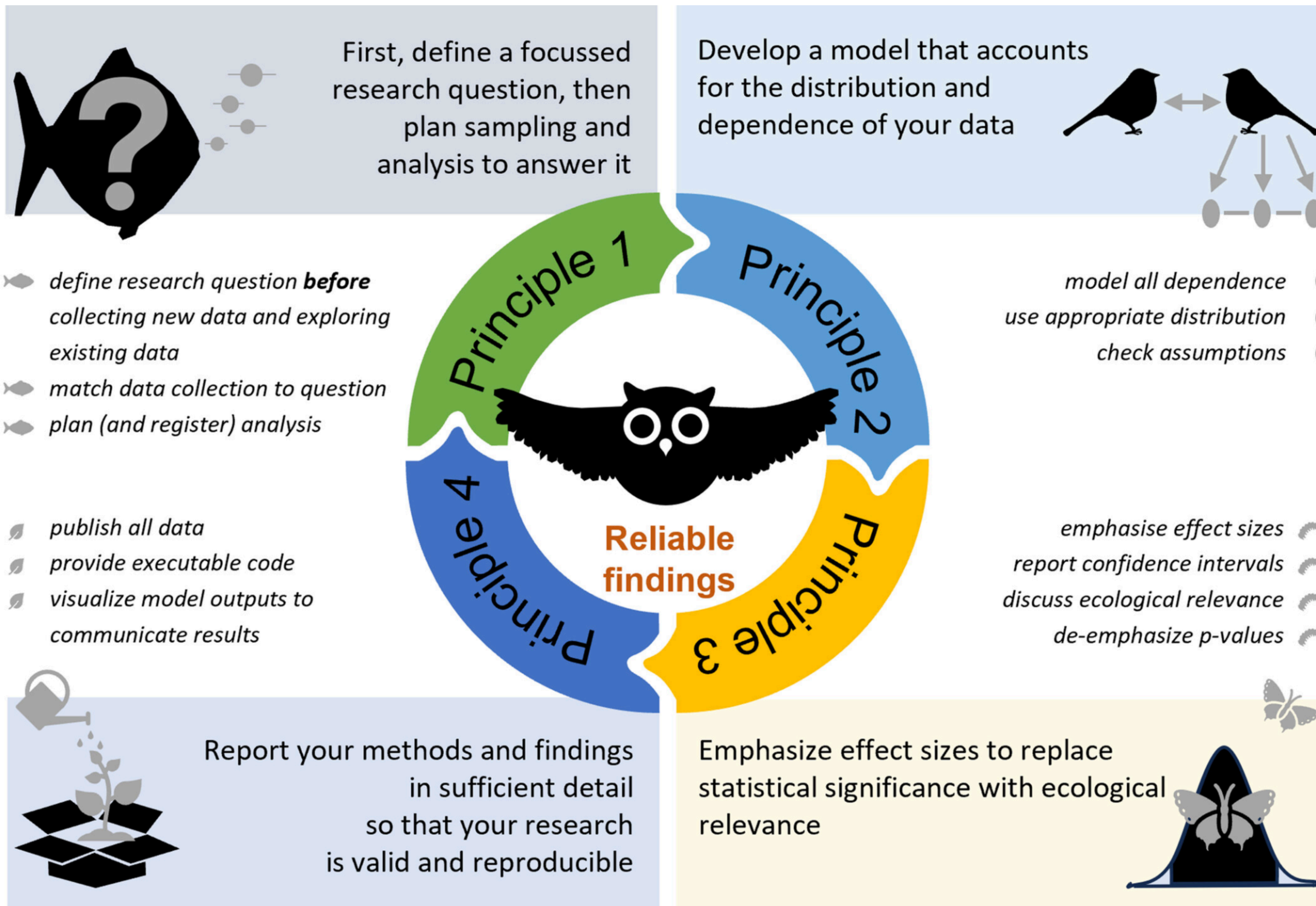


Data
+
Analysis



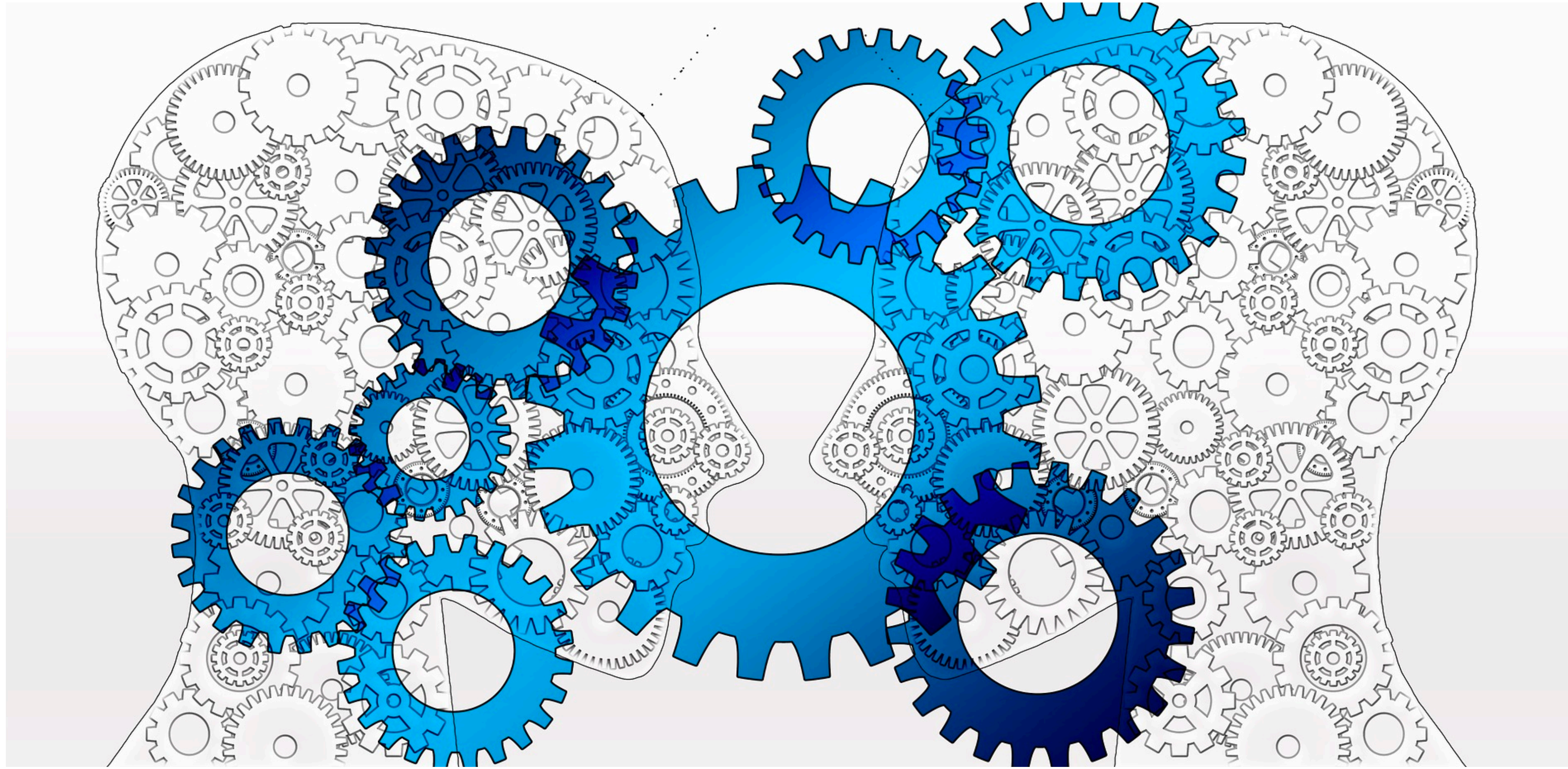
Answer
Solution

But how?
High quality data analysis?



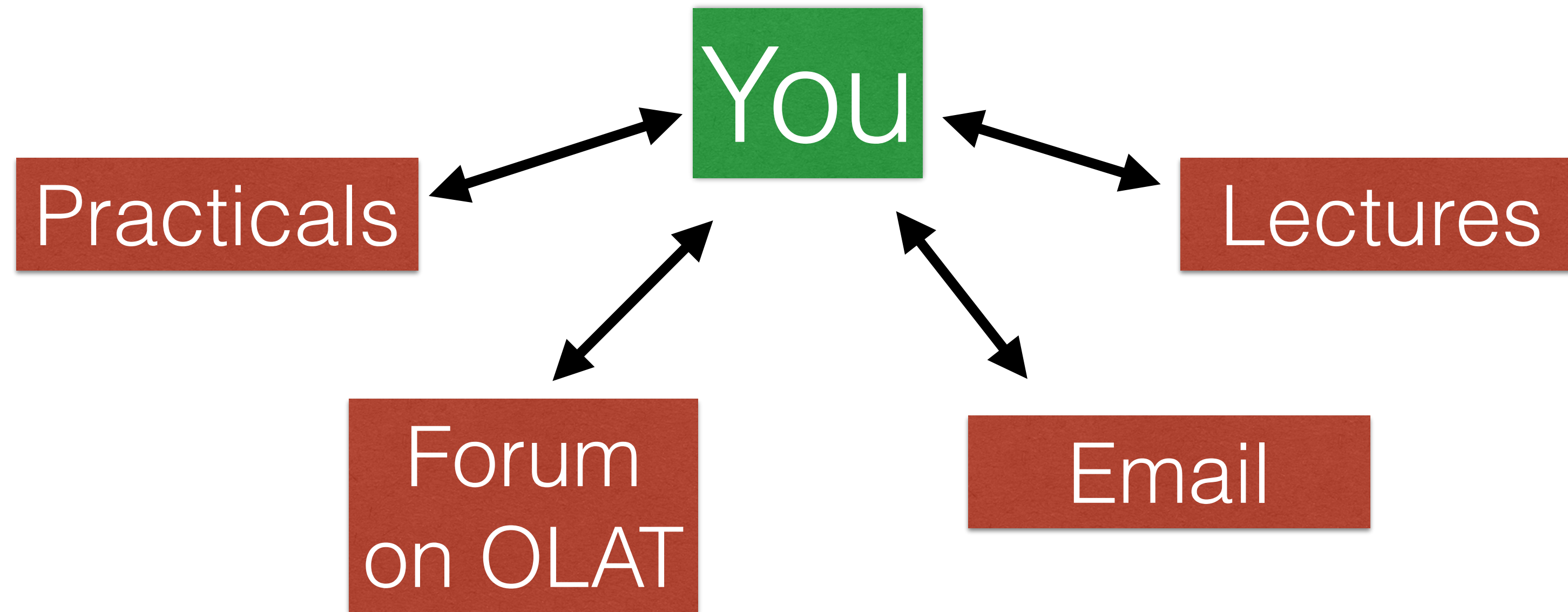
Popovic, G., Mason, T.J., Drobniak, S.M., Marques, T.A., Potts, J., Joo, R., *et al.* (2024). Four principles for improved statistical ecology. *Methods in Ecology and Evolution*, 15, 266-281.





How the course is organised

The team



Prof. Owen Petchey
Director
Instructor



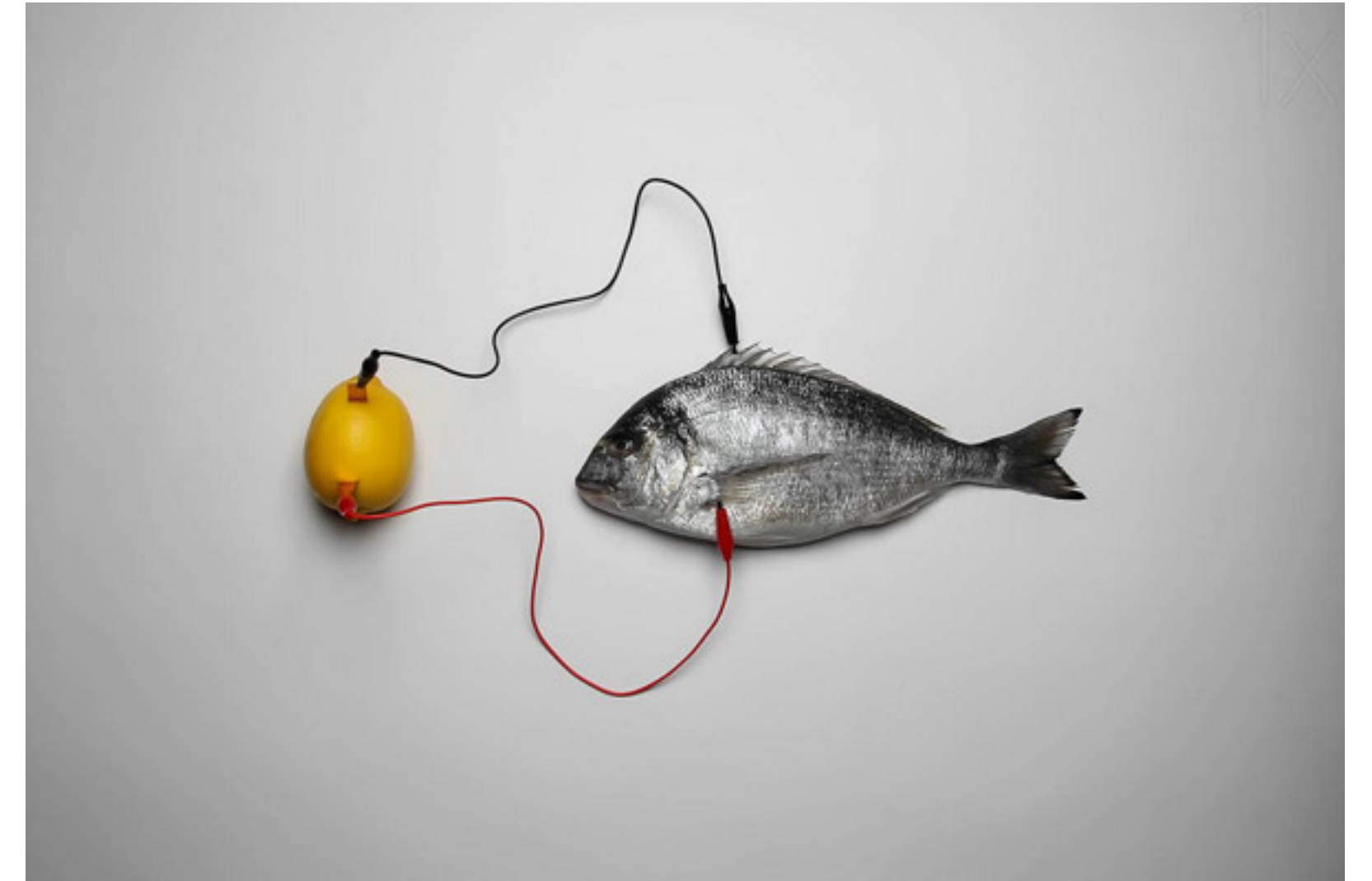
Til Hämmig
Head Teaching Assistant



Dr Frank Pennekamp
Examinations

Numerous
teaching
assistants

The conceptual side



8 - 9:45am Mondays

Lectures (apart from today)

- No powerpoint slides.
- The “course book” contains what you need to know.
- I will talk, draw, code, and ask questions (think, pair, share).
- One lecture by a guest, otherwise me.

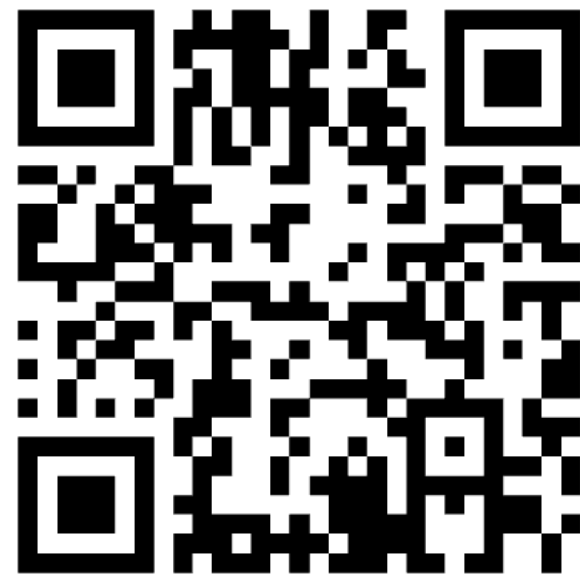


Think, Pair, Share

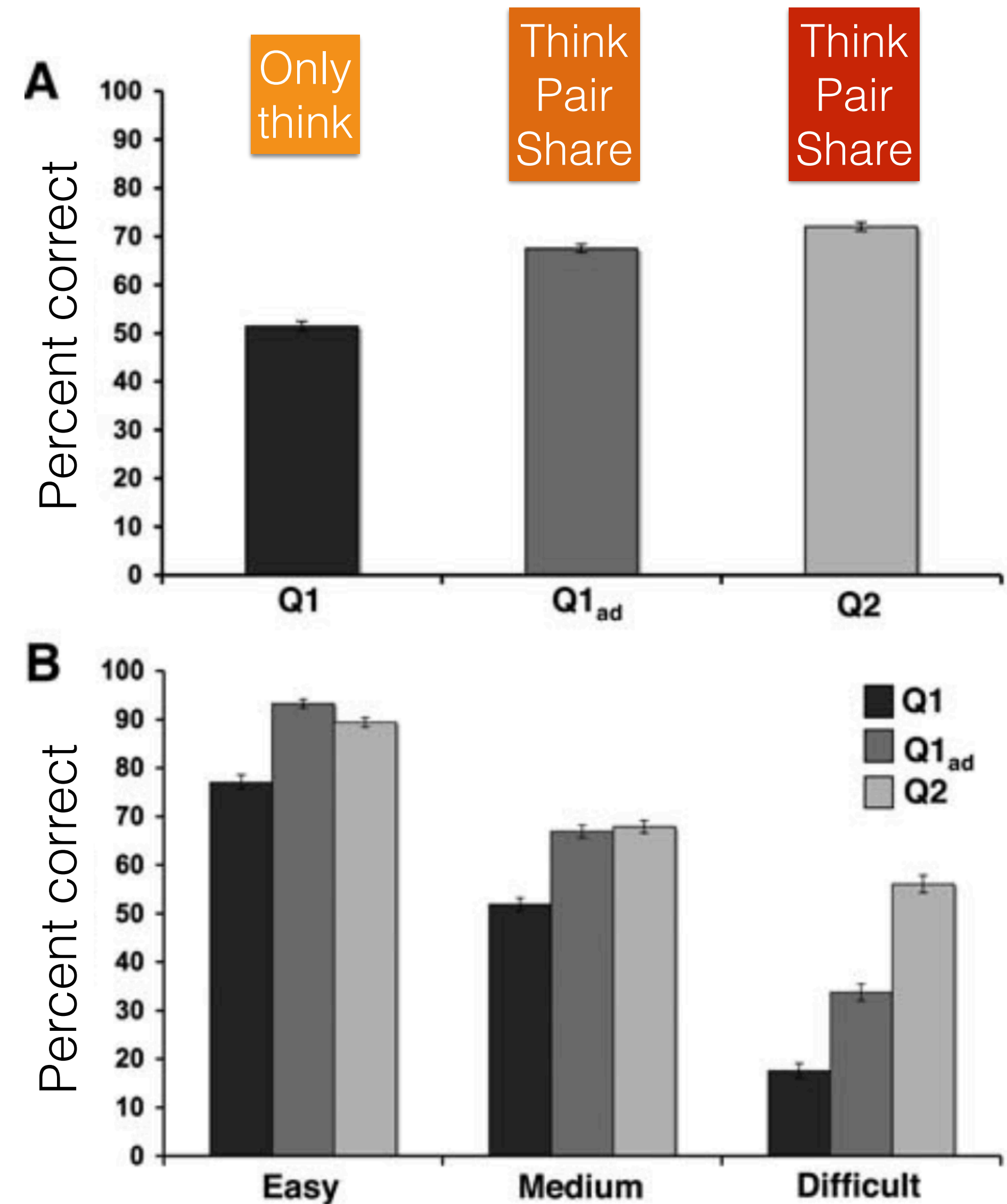


Think, Pair, Share

Our results indicate that peer discussion enhances understanding, even when none of the students in a discussion group originally knows the correct answer.



Smith, M.K., Wood, W.B., Adams, W.K., Wieman, C., Knight, J.K., Guild, N. & Su, T.T. (2009).
Why Peer Discussion Improves Student Performance on In-Class Concept Questions.
Science, 323, 122–124.

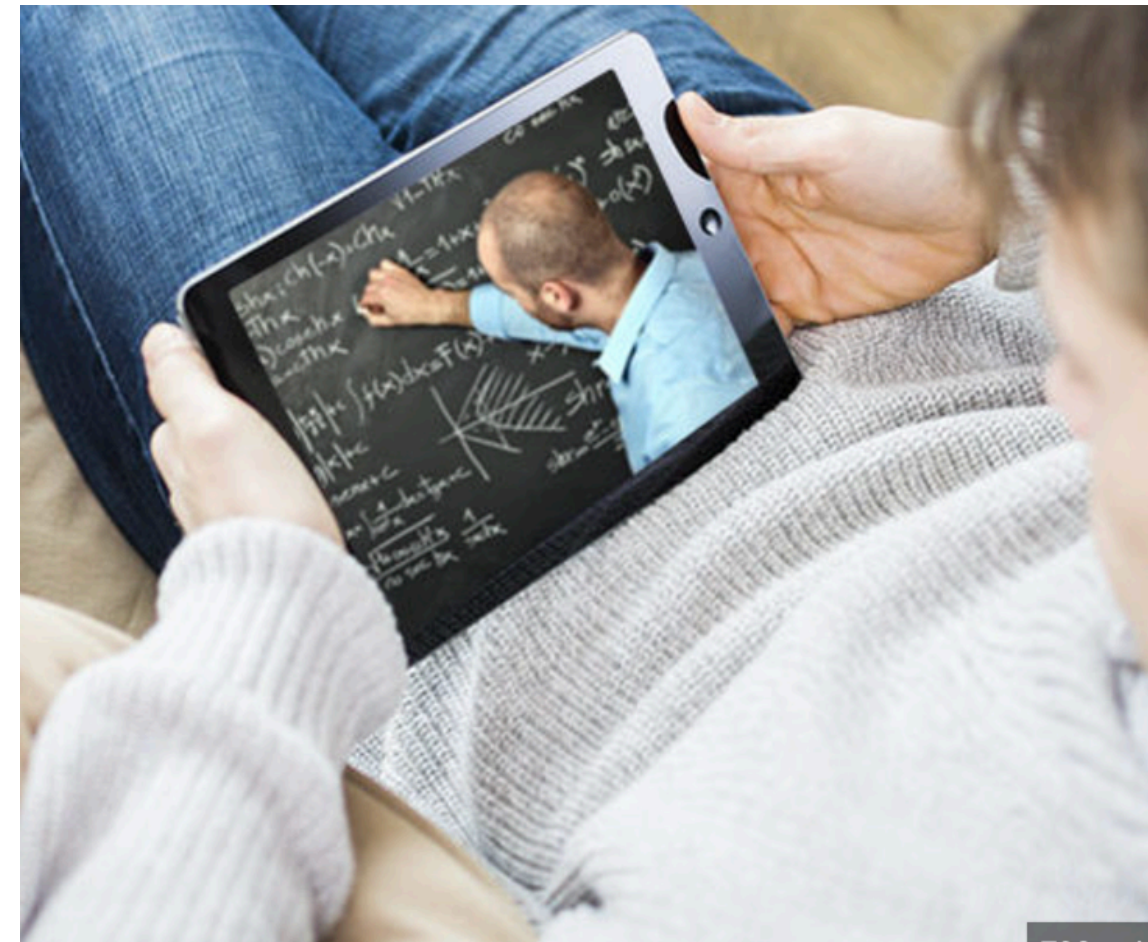


The practical side



1 - 3pm Thurs. or Fri.

Homework



**Sometime that suites you,
best between Monday and Thursday/Friday**

Weekly practice quizzes

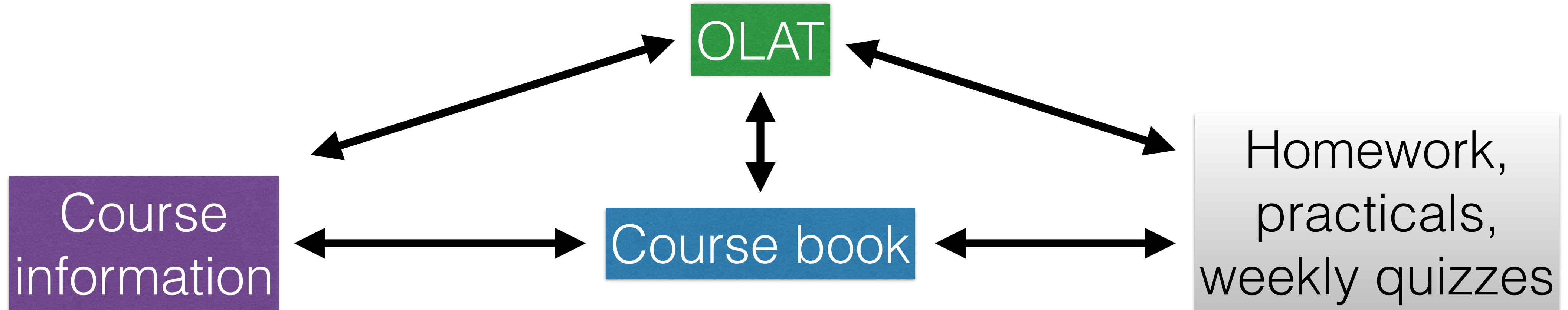


Anytime you like!

***But please please make sure you keep up with the course material,
and reach out to us if you feel like you're falling behind.***

All of that and more is on four websites(!)

<https://lms.uzh.ch/auth/RepositoryEntry/17827987490/CourseNode/112943793799740>



BIO144 Course Information (version 2026)

AUTHOR: Owen Petchey | PUBLISHED: January 6, 2026

Warning
Under construction: This 2026 version of this information is currently being prepared.

Preface
This website contains information about the BIO144 Data Analysis in Biology course at the University of Zurich.

How this website was made
The book was written using a typewriter and R code to produce chapters. Rmarkdown is very useful for me.

This book is a Quarto book. To learn more, visit <https://quarto.org/docs/books/>.

BIO144 Course Book (version 2026)

AUTHOR: Owen Petchey | PUBLISHED: January 3, 2026

Warning
Under construction: This 2026 version of this book is currently being prepared. Some chapters are not yet available, and some content may change before the course starts in February 2026.

Preface
This book contains the content of the course BIO144 Data Analysis in Biology at the University of Zurich. It is intended to be used as a companion to the lectures and practical exercises of the course. All of the required content of the course (i.e., what could be in the final exam) is included in this book. Additional content is included for those who want to learn more.

How to get a copy of this book

Week 1 BIO144 Course Content

Practical
Exercise 1
During the first lecture you collected data on class reaction times. The exercise this week is for you to use R to analyse the dataset. To help, here is a nearly complete and working script of R code. It does, however, contain some deliberate errors. You'll need to find and correct those! It also contains some "???" that you need to replace with appropriate text. The errors get a bit harder as you go through the script. Good luck!

Script file: [reaction_time_analysis.r](#)

Date file: [Link to Reaction Time Dataset](#)

Click on the link, and hopefully the files will be downloaded. If they appear in your browser do 'Save As...' and save them to somewhere safe on your computer. Do not add a file extension. Leave the file extension as they are. Do not open the data file in Excel or any other program.

Get organised by setting up a project folder on your computer and downloading the above files into it.

Then open the script in RStudio and get it working, not forgetting to think about the questions in the script, about assumptions, effect size, biological significance, statistical significance, and so on.

Write a sentence about the finding. Cover statistical and practical significance. Be critical.

If you finish early and are happy with everything so far, and you want a challenge, figure out how to use RMarkdown to create a word document, html file, or pdf of your findings.

Much of the course material is new or newly revised this year. It is not perfect: some sections may be brief, and there may be typos or occasional mistakes. Please let me know when you think you may have experienced issues.



Lots of good feedback, but here is the more critical/constructive:

Feedback

Important to have clear structure in lectures

Some students would have liked a slide deck (like in many lectures)

Homework videos nicer than reading chapters

Mr. Petchey's lectures were extremely slow paced and chaotic, since we switched from script to other stuff and back.

The lectures of Prof. Petchey were always engaging (questions via qr code to the students, showing R, talking about statistics) and it felt very including even if statistics isn't really a strength of a person.

I think course should be restructured to be more slow paced/ lesser load.

For the number of ECTS credits allocated to this module, the workload is far too high (multiple responses like this).

Different units (= weeks have different time requirements)... therefore look early and plan early.

Prof. Petchey talks for hours on end.

Mitigation

Give structure at beginning and end.

Not going to do this. Instead, course book and me guiding you.

Sorry, still some reading to do.

Give structure, check pace, smoother switching.

Continue engaging elements (e.g., think-pair-share).

Fewer topics this year.

Fewer topics this year.

Look ahead, plan ahead early.

Sit back, open popcorn, enjoy :)

US

You



22FS BIO144 Data Analysis in Biology (Preparation)

Course info | Calendar | Participant list | Participant infos | E-Mail | Blog | Wiki | Forum | Documents | Glossary | PARTICIPANT | User role | My course

Course chat | Course search

BIO144

- About the course
- Previous knowledge
- Wiki: FAQ
- Forum**
- Unit 01
- Unit 02
- Unit 03
- Unit 04
- Unit 05

Overview of topics

Open new topic | Archive forum

Search

3 Entries

Type	Discussion topics	Author	Last modified	Marked	New	Posts
Sticky	Hyperlinks in OLAT	Daugaard, Uriah	10/29/2021, 10:20 PM	0	0	1
Sticky	Guidelines for posting code in the forum (read before posting)	Daugaard, Uriah	9/15/2021, 5:08 PM	0	0	1
Sticky	What is this Discussion Forum for? (read before posting)	Daugaard, Uriah	9/15/2021, 4:52 PM	0	0	1



Ask questions / make requests here.



**How hard will this
course be?**

The final examination

- It will involve three or four case-studies in which you have to analyse data and answer questions about it. Embedded in the case studies are numerous conceptual questions.
- Quite a lot of the questions require you to answer by entering numbers that are rounded to the required number of decimal places.
- It will be in-person, on-site, and on your own computers (BYOD exam).
- You may bring notes on both sides of **three** pieces of A4 paper. You may not consult any other sources during the exam.
- You may bring a German-English dictionary (and should allow us to inspect it).
- You will have to use the Safe Exam Browser and RStudio Server/Cloud.
- There will be a mandatory practice exam experience.
- You will get access to a representative examination from a previous year.
- The exam grades are not “curved”, i.e., you are not competing against each other. (We may adjust the score to pass if a particular examination appears very hard compared to previous ones.)

"Failure is an opportunity to grow"

GROWTH MINDSET

"I can learn to do anything I want"

"Challenges help me to grow"

"My effort and attitude determine my abilities"

"Feedback is constructive"

"I am inspired by the success of others"

"I like to try new things"

I and we are here to support you.

Equip you with the knowledge of how to learn more.
And the confidence that you can.

I may be a professor, but I'm also human.
Feel free to address me as you feel comfortable,
Professor Petchey or Owen
(Probably not "dude" or "mate"!)



Our Journey Together: Becoming Great Data Analysts

My Role, Your Role, and How We'll Work Together

- Everyone here can learn to think analytically and statistically.
- I will hold you to high standards, because I believe you can meet them.
- My job is to support you in reaching them.
- Struggle is part of growth, not a sign of failure or deficiency.
- The stress response is your body preparing for a challenging situation. Higher heart rate, faster breaths, tight stomach, your body is giving you more resources to meet the challenge.



“What’s a time when someone believed in you more than you believed in yourself? What happened?”

High standards & Wise Feedback

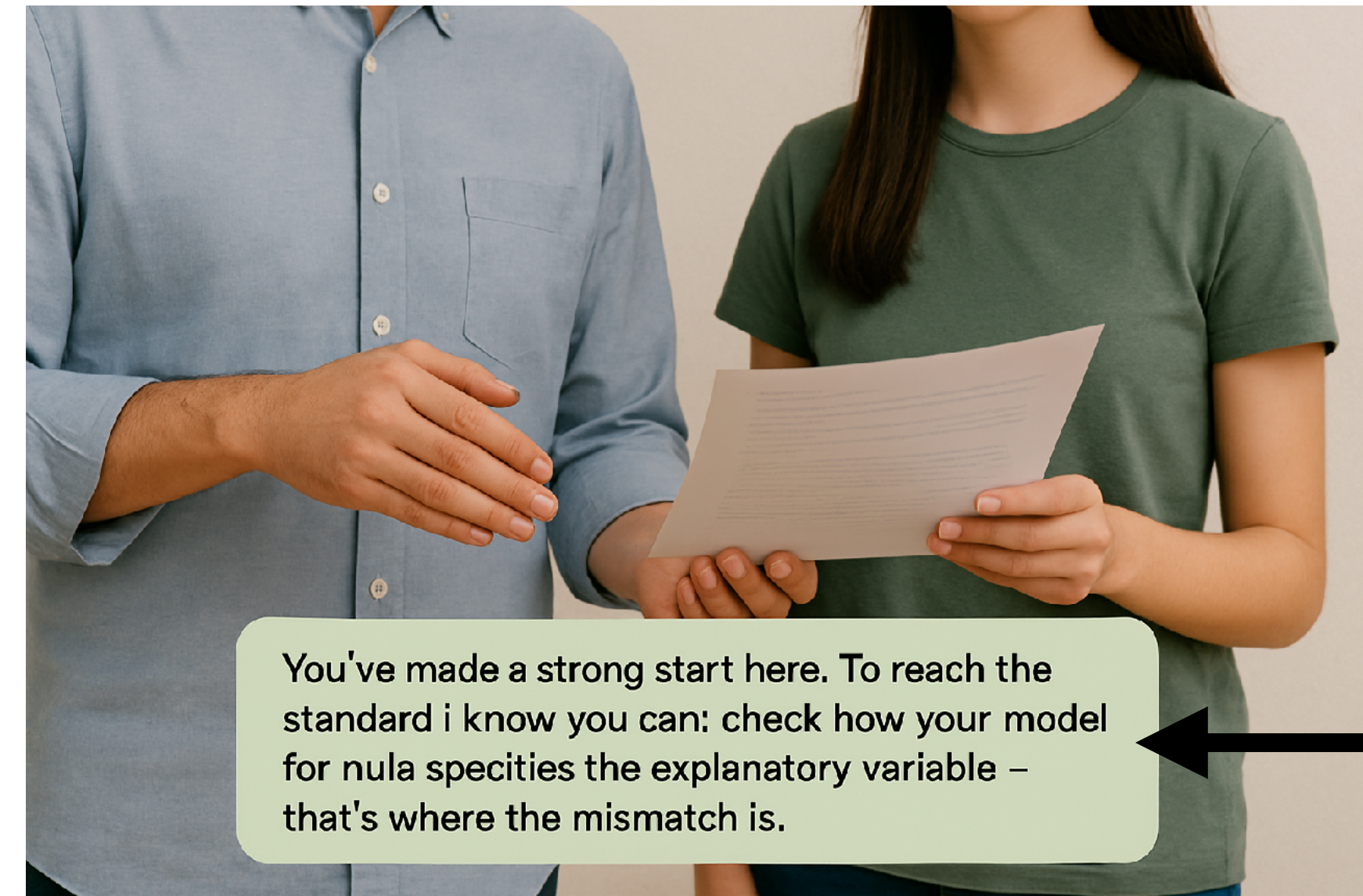
Wise Feedback (Cohen et al., 1999):

“I’m giving you these comments because I have **high standards** and I know that you can **meet them**.”

- Feedback will be **honest and specific**.
- It will come from belief in your **potential**, not criticism of your **person**.

If it ever feels like you are being criticized, try to reframe it as you are being challenged to grow.

This class is a **safe space** for you, and it is a **challenging space** for you.



ChatGPT
WTF?

Think, Pair, Share:
“How comfortable are you receiving critical feedback?”





AI Assistants

Methods in Ecology and Evolution



FORUM | Open Access |

Should we still teach or learn coding? A postgraduate student perspective on the use of large language models for coding in ecology and evolution

This article relates to:

Heather Campbell , Thomas Bluck, Ella Curry, Derrick Harris, Billie Pike, Bethany Wright

First published: 01 October 2024 | <https://doi.org/10.1111/2041-210X.14396>

University of Zurich

Handling Editor: Robert B. O'Hara



“ I always found if I'm trying to research something I just go around websites and pages skipping through and in the end, I've gone around in circles. So that's helpful [about LLMs]. ”

“ It's really helpful with coding for checking what you're doing, but other than that, I haven't used it for anything else. I think it cuts out what I see as important parts of the process of doing a lot of things. If I just ask it something and it gives the answer, then I haven't gone through that process of researching, exploring, and interpreting it myself. ”

“ I've had it get things wrong, for example when I couldn't work out which package I'd used I gave ChatGPT my code and requested it identify the package used for a specific section out of the listed packages I had installed. The response suggested three different packages, none of which I had used. When told these packages were not applicable, ChatGPT acknowledged it was wrong but would not provide an alternative answer. ”

Ask for help understanding...

- What is the mode of a distribution?
- I didn't understand that explanation. Can you give a more easily understood explanation?
- Can you show me the mode graphically?
- Can you please give me an example of getting the mode of a distribution in R?

Ask for help writing code...

- Please use functions from the tidyverse package to write R code.
- Please try to not use square brackets or dollar signs.
- Please write code that will import data, give number of missing values in each variable, and graph all pairwise relationships.
- Please write code that will import a dataset containing two continuous variables, perform linear regression, check model diagnostics, give a summary of the linear regression results, and make a nice visualization including the regression equation, the explanatory power, and the p-value of the slope, the regression line and the confidence and prediction interval around the fitted regression line. Put all this into a quarto file. Do this while I go chill. Thanks.

What can I use an AI assistant for in BIO144?

Think, Pair, Share



Your relationship with AI Assistants

In a strong AI-developer relationship, the AI serves as an empowering tool—a knowledgeable collaborator that complements the developer's skills, accelerates their workflow, and supports their growth as a coder. The developer, in turn, provides context, direction, and critical judgment to extract the maximum value from the AI's capabilities.

In a poor relationship, the developer uses the AI's output without understanding it, leading to bugs, inefficiencies, or misaligned solutions. The developer becomes overly reliant on the AI, copying and pasting solutions without understanding them, which leads to long-term skill stagnation. The developer uses the AI's output without understanding it, leading to bugs, inefficiencies, or misaligned solutions.

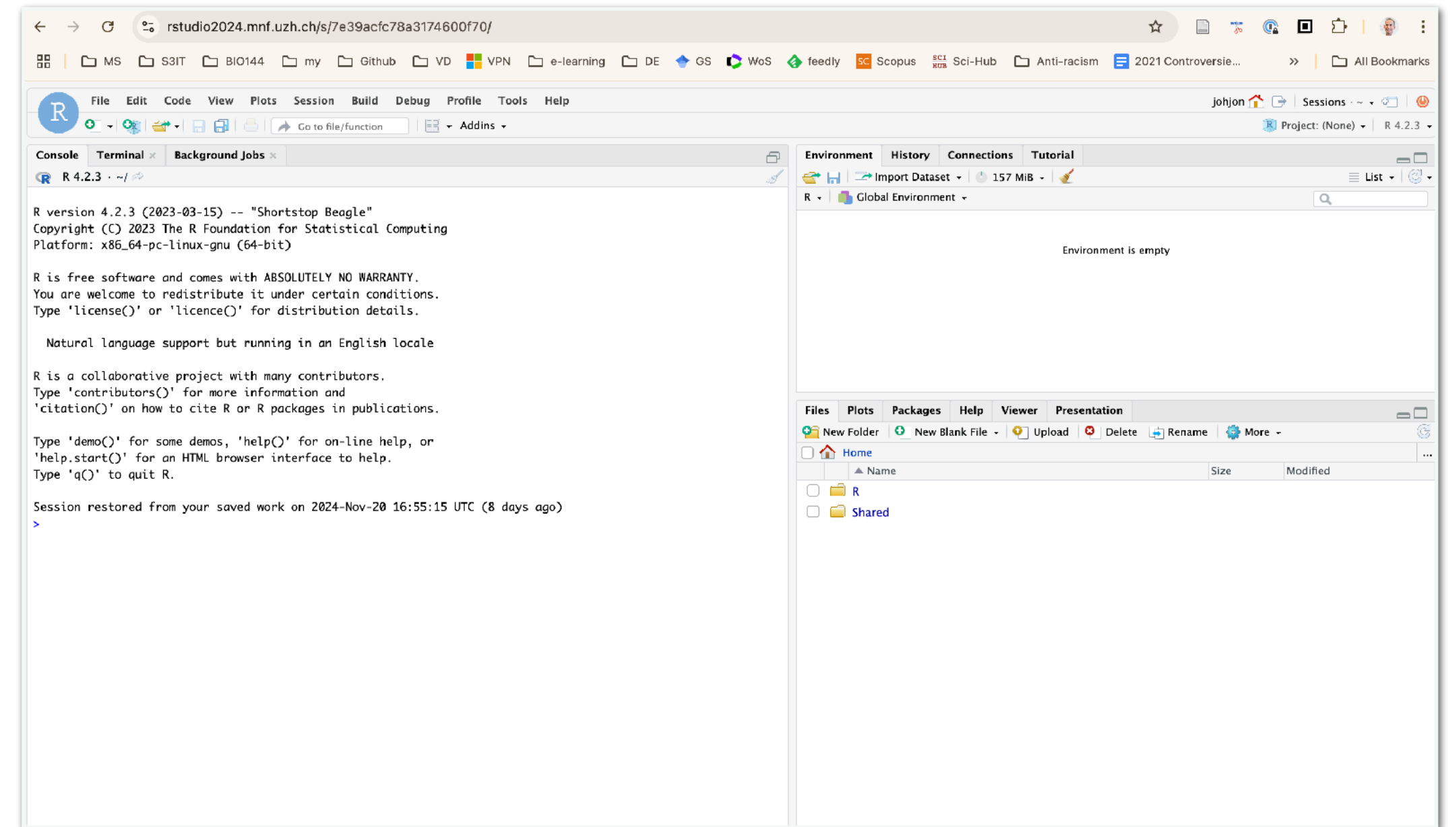
I suggest you try to minimize their use during the course.
(Also try to minimize using the example solutions.)

**Do the work, get the benefit.
(the knowledge and expertise that you contribute to a
good relationship with an AI assistant).**

RStudio Server

You can use RStudio in a web browser—you use an RStudio Server

- **Bad thing:** You have to be connected to the internet
- **Good thing:** Lots of the setup is already done for you
- (**Good thing:** available memory and CPUs can scale)
- **Requirement:** you will have to use RStudio server in the exam.



Users must be registered.
I register everyone in the OLAT course.

Therefore you must be in the OLAT course to get access to RStudio server.

<https://rstudio.mnf.uzh.ch/>

(This is in the course information web page also)

We need some data to work with...
Do some tests and record the resulting data

BIO144 Course Information website (Linked to from OLAT)

the course book.' A red rounded rectangle highlights a section titled 'Lecture' which contains the text: 'In the lecture you'll be asked to record some reaction time data about yourself. Use this google form to record the data: [Link to Reaction Time Data Form](#). Go to this web page for a tool to record your reaction times: [Link to Reaction Time Tool](#). The resulting dataset can be access and or downloaded from this web page: [Link to Reaction Time Dataset](#)'."/>

BIO144 Course Information (version 2026)

Search

Preface
General information
Year specific information
Learning Objectives
FAQs
Discussion Forum
Unit 1
Unit 2
Unit 3
Unit 4
Unit 5
Unit 6
Unit 7
Unit 8
Unit 9
Unit 10
Unit 11
Unit 12

However, just like all weeks, there is:

- A chapter of the course book.
- A “lecture” on Monday.
- Some homework to ideally do between Monday and the practical on Thursday or Friday.
- A practical session on Thursday or Friday.
- A weekly quiz to check your understanding.

Course book

This week we will be covering Chapter 1 of [the course book](#).

Lecture

In the lecture you’ll be asked to record some reaction time data about yourself. Use this google form to record the data: [Link to Reaction Time Data Form](#)

Go to this web page for a tool to record your reaction times: [Link to Reaction Time Tool](#)

The resulting dataset can be access and or downloaded from this web page: [Link to Reaction Time Dataset](#)

Lecture
Homework, Practical, and Weekly Quiz

Edit this page
Report an issue

Live data analysis demonstration

BIO144
Week 1

Its a demonstration...

The idea is to give you a feel of what is involved in data analysis.

You will understand some of the demonstration.

You will not understand some of it.

Keep notes about what you don't understand.

Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate

Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate

The question

- What should be our question?
- As always, there are some influences and some constraints.
- We should ask a question of interest to us, and of some importance.
- And we should be able to collect the data, within our current constraints, necessary to answer the question.
- The question we will address is "***do male and female reaction times of students at the University of Zurich differ?***".
- Why this question? Reaction times are important, safety, sport...

Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate

Expectation

- Quite a lot of work on this already.
- Generally, males tend to have faster reaction times than females. So we expect that to be the same for students at the University of Zurich.

Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate

What graph?

Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

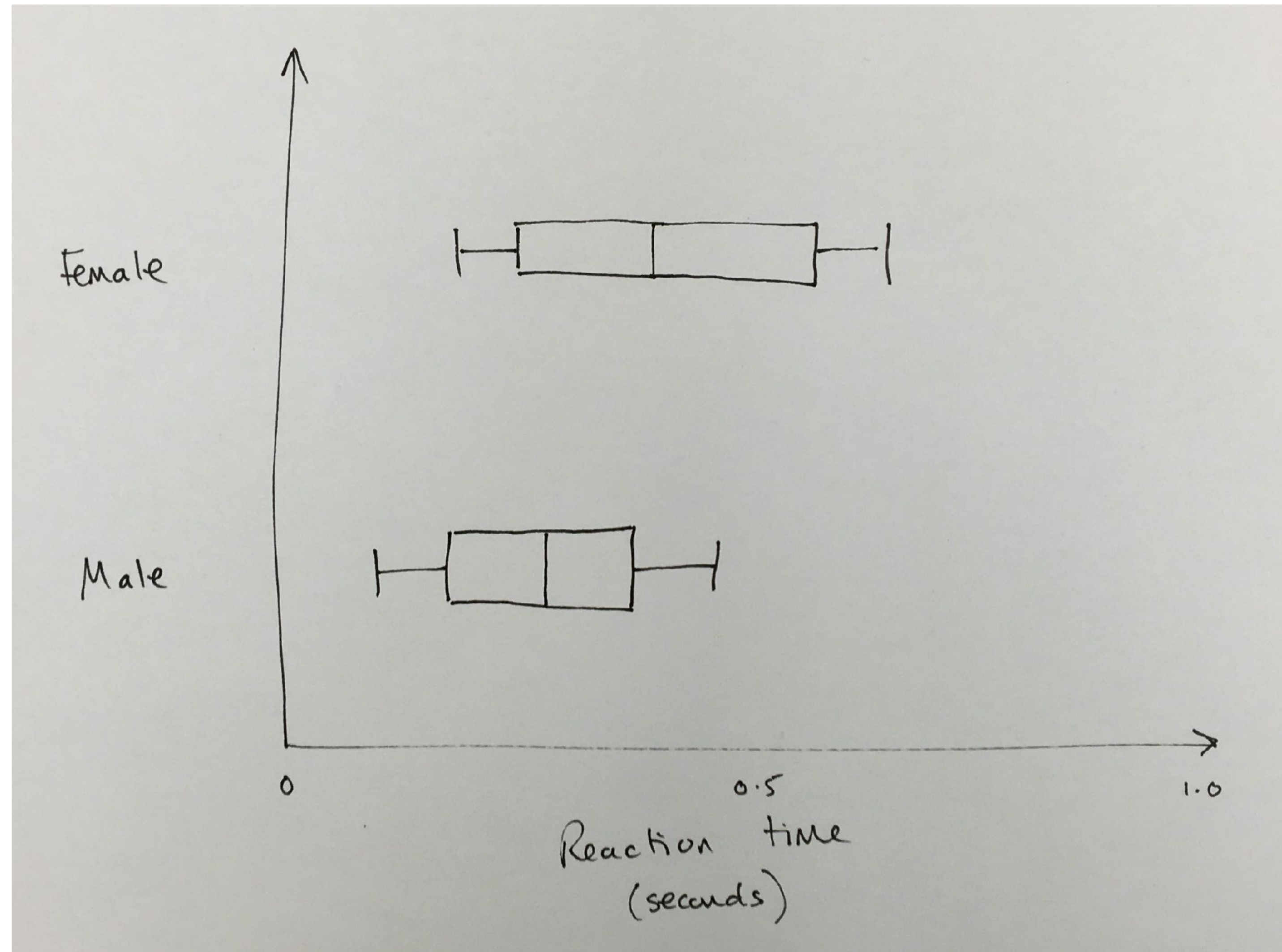
Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate



Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate

What statistical test?

What assumptions?

Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate

Ideally would be random.
Are you a random selection of all
humans?

Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate

Done.

Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate

Since we collected data about you
(humans) we should think
about ethical issues.

Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate

Done.

```
R Console
R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

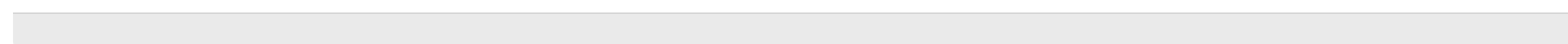
[R.app GUI 1.66 (6996) x86_64-apple-darwin13.4.0]

[Workspace restored from /Users/owenpetchey/.RData]
[History restored from /Users/owenpetchey/.Rapp.history]

>
```



Efficient
Consistent
Repeatable
Reliable
Readable
Robust
Persistent
Sharable
Scalable



Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate

Live in RStudio

Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

Visualise

Statistical test

Critical thinking

Report / communicate

Live data analysis demonstration

The whole data analysis workflow in one hour!!!

Question

Expectation

Planned presentation & analysis

Selection of subjects

How will data be collected?

Ethics / permissions

Data collection

Data wrangling

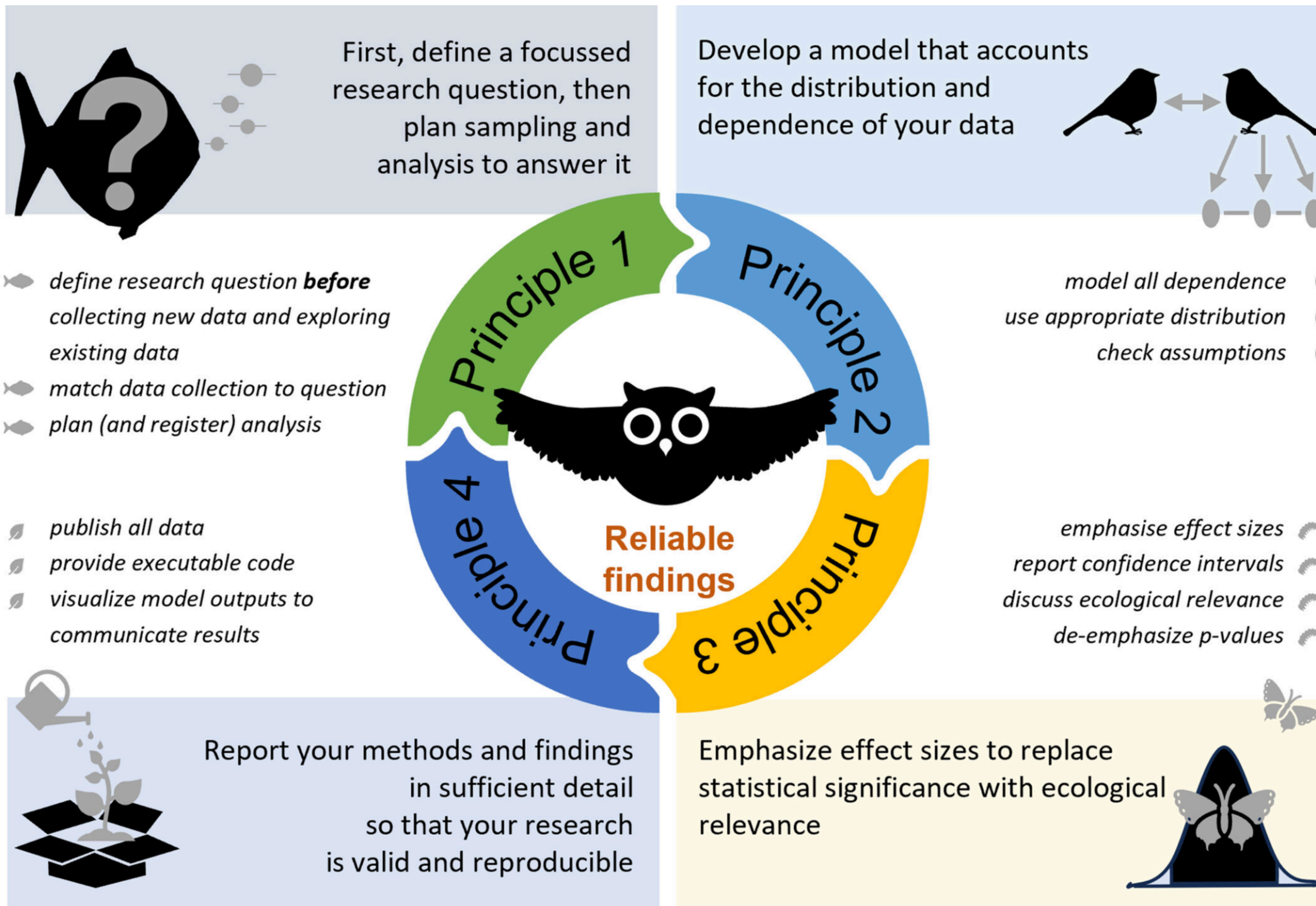
Visualise

Statistical test

Critical thinking

Report / communicate

Live in RStudio



Popovic, G., Mason, T.J., Drobniak, S.M., Marques, T.A., Potts, J., Joo, R., *et al.* (2024). Four principles for improved statistical ecology. *Methods in Ecology and Evolution*, 15, 266-281.



Activate your prior learning!

22FS BIO144 Datenanaly

- ▶ About the course
- ▼ **Previous knowledge**
 - ▶ Introduction
 - ▶ Notation review
 - ▶ Self-tests
 - ▶ Getting R & RStudio
 - ▶ Getting to know R
 - ▶ Help about R and RStudio
 - ▶ What are add-on packages?
- Wiki: FAQ
- Forum
- ▶ Unit 01
- ▶ Unit 02
- ▶ Unit 03
- ▶ Unit 04
- ▶ Unit 05

Introduction

The aims for this section are:

1. To help you know about your knowledge of some of the things you should already know before you start BIO144.
2. To give you a chance to refresh your knowledge about some of the things you don't recall so well.
3. **Important:** Multiple choice questions with little squares in which you need to tick the correct answers can have any number of correct answers. Such questions with little circles have only one correct answer. (See the illustrative examples in the next page.)

Please note that the things covered in this section are not exhaustive, i.e. there might be other things you've previously learned about useful for BIO144.

Your score in the quizzes here doesn't contribute to anything. It's just for you!

Notation review

Self-tests

Getting R & RStudio

Getting to know R

Help about R and RStudio

How to get help about R and RStudio

What are add-on packages?

Use the resources on OLAT

You should have already done this.
If you have not, you still have time

Reviewing what you already experienced / what would be useful for you to already know

Summarising distributions of data

Central tendency: mean, median, mode
Measures of spread: range, quartiles, standard deviation, standard error
Symmetry: skew.
Pointy-ness: kurtosis.
Number of peaks: unimodal, bimodal, multimodal

Statistics:

Expected value
Parameter
Parameter estimate
Variable
Central limit theorem
Hypothesis testing
Degrees of freedom
P-value

Notation

$$\bar{x} \quad (\text{"x bar"})$$
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Types of variable:

numeric, continuous, discrete, categorical, dependent/response, independent/explanatory.

T-test

Linear regression



Difference between a *population* and a *sample*.

Probability: e.g., what is the probability of getting five heads if we toss a fair coin seven times?

Probability distributions: normal, binomial, poisson, exponential.

Frequency distributions, aka, histograms.

Visualisations: box and whisker, scatterplots.

Models of relationships: equation for a straight line.