

# Data Analysis in Biology

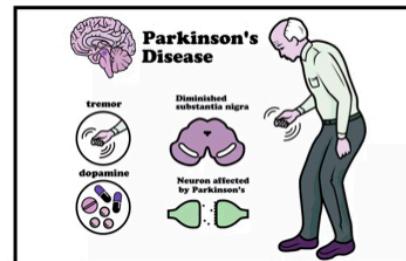
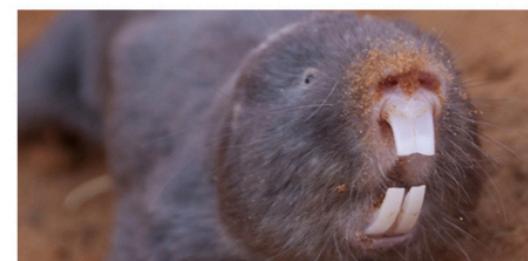
BIO144  
FS 2025

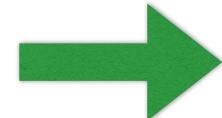
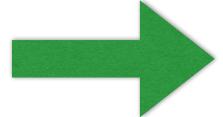


## Overarching goals of the course

- ▶ Provide a solid foundation for answering biological questions with quantitative data.
- ▶ Help students to understand the language of a statistician.
- ▶ Ability to understand and interpret results in research articles.
- ▶ Give the students a challenging, engaging, and enjoyable learning experience.

# Questions you'll work on...





Question  
Puzzle  
Problem

Data  
+  
Analysis

Answer  
Solution

But how?

What words would you use to describe a  
***high quality data analysis?***

reproducible

objective

robust

efficient

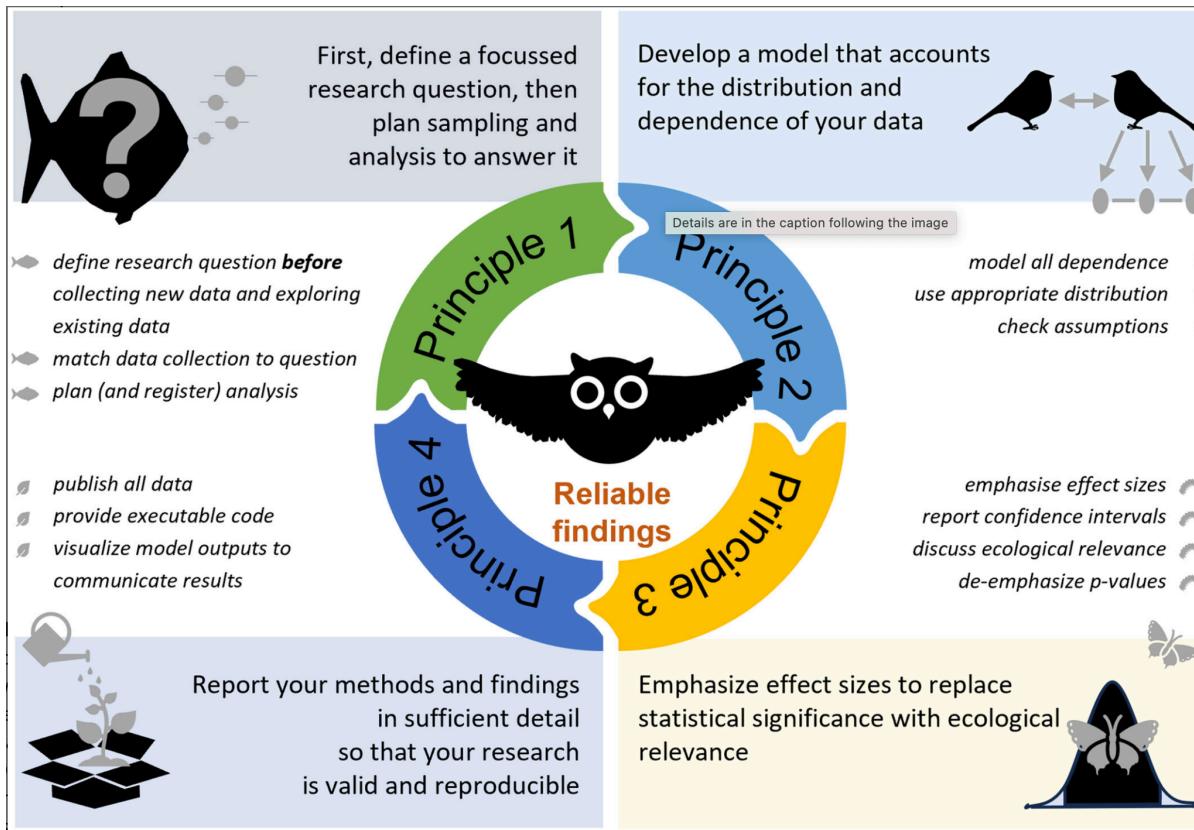
reliable

understandable

significant

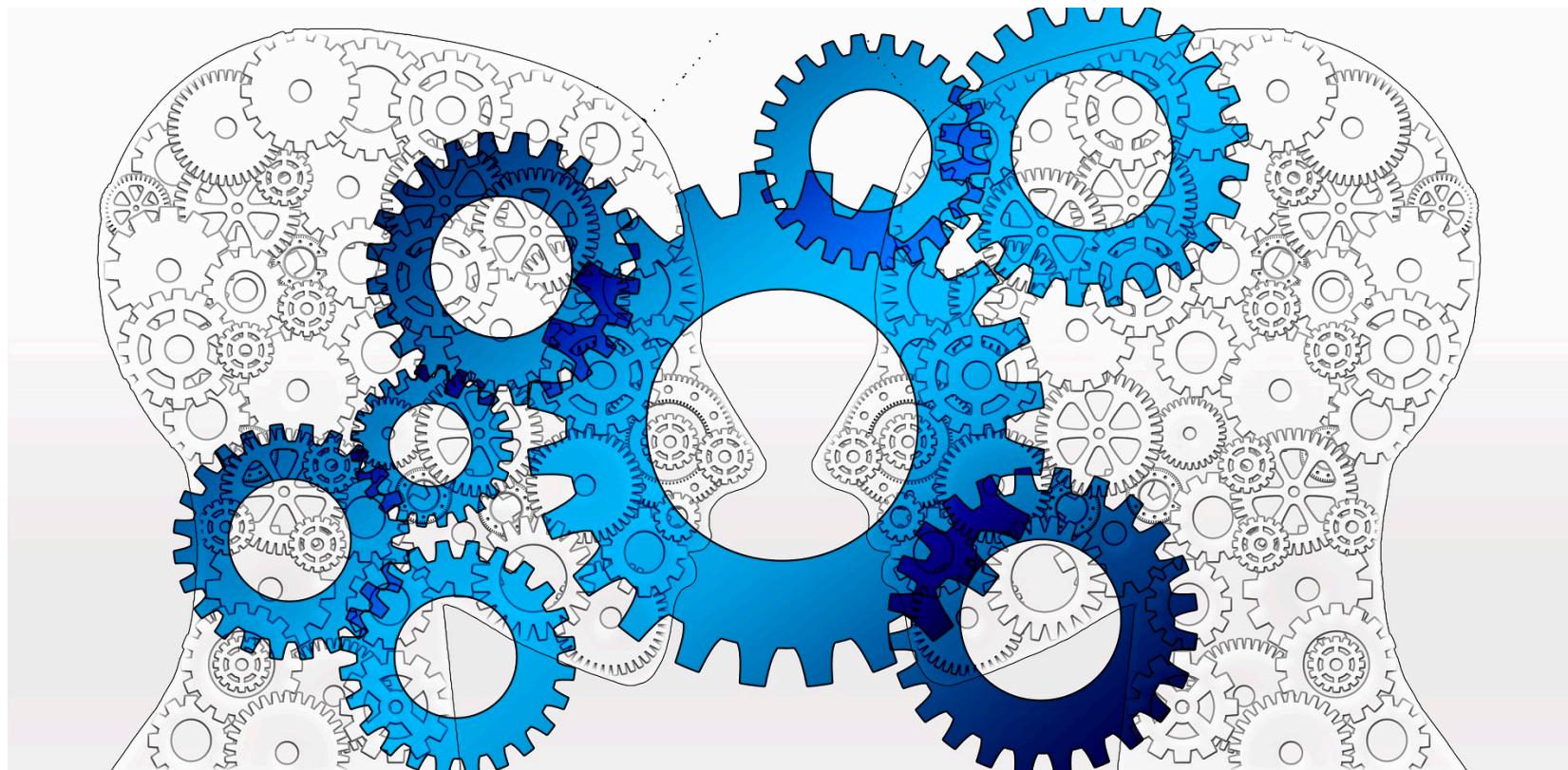
shareable

fun



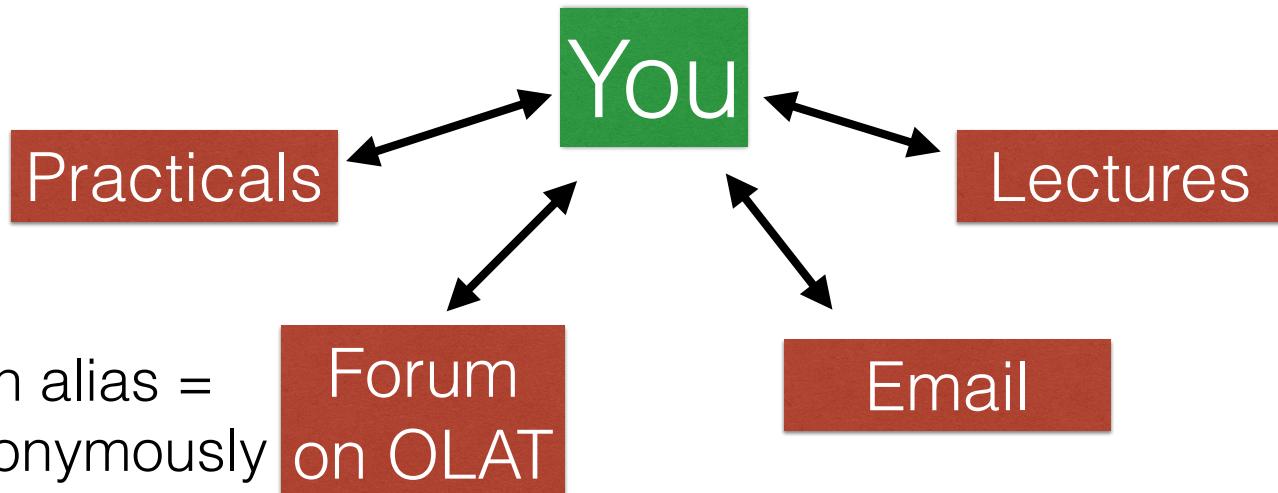
Popovic, G., Mason, T.J., Drobniak, S.M., Marques, T.A., Potts, J., Joo, R., et al. (2024). Four principles for improved statistical ecology. *Methods in Ecology and Evolution*, 15, 266-281.





# How the course is organised

# The team



Dr Erik Willems  
Instructor



Prof. Owen Petchey  
Director  
Instructor



Martina Jelic  
Head Teaching Assistant

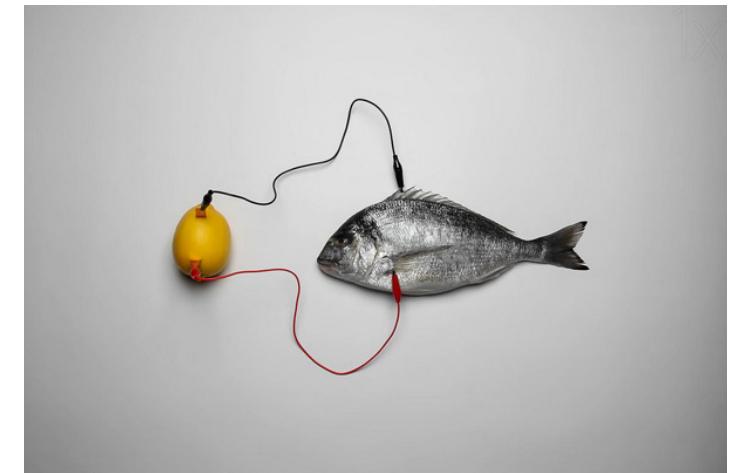


Dr Frank Pennekamp  
Examinations

Numerous  
teaching  
assistants

|   |          |
|---|----------|
| <b>Units 1 and 2 - Introduction; all about data</b>   | <b>2</b> |
| <b>Unit 3 - Linear Regression Part 1</b>  | <b>2</b> |
| <b>Unit 4 - Linear regression part 2, and multiple regresion</b>                            | <b>3</b> |
| <b>Unit 5 - Binary/categorical explanatory variables, and interactions</b>                  | <b>3</b> |
| <b>Unit 6 - ANOVA</b>   | <b>4</b> |
| <b>Unit 7 - ANCOVA &amp; Matrix alegbra</b>   | <b>4</b> |
| <b>Unit 8 - Model selection</b>   | <b>4</b> |
| <b>Unit 9 - Interpretation, causality, and cautionary notes</b>                             | <b>5</b> |
| <b>Unit 10 - Analysing count data</b>   | <b>5</b> |
| <b>Unit 11 - Analysing binary data</b>  | <b>5</b> |
| <b>Unit 12 - Measurement error; repeated measures and random effects; recap and outlook</b> | <b>6</b> |

# The conceptual side



1 - 2:45pm Mondays

# The practical side



1 - 3pm Thurs. or Fri.

# Homework



Sometime between Monday and Thursday/Friday

# Weekly practice quizzes



**Anytime you like, but we'll get in touch  
if you seem to fall behind**

# The final examination

- It will involve three or four case-studies in which you have to analyses 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 two pieces of A4 paper. You may not consult any other sources during the exam.
- You will have to use the Safe Exam Browser and RStudio Server.
- You will get access to a representative examination from a previous year.

# All of that and more is on OLAT

The screenshot shows the OLAT course preparation interface for the course '22FS BIO144 Data Analysis in Biology (Preparation)'. The top navigation bar includes links for Courses, Groups, Authoring, Campus courses, Question bank, and a search bar. The main content area is titled 'PREPARATION' and contains sections for Administration, Status, Course info, Calendar, Participant list, Participant infos, E-Mail, Blog, Wiki, Forum, User role, and My course. Below this, there is a toolbar with icons for Documents, Glossary, Course chat, and Course search. On the left, a sidebar lists course units: BIO144, About the course, Previous knowledge, Wiki: FAQ, Forum, Unit 01, Unit 02, Unit 03, Unit 04, Unit 05, Unit 06, and Unit 07. The main content area displays a list of preparation items:

- Aims and learning outcome
- Course schedule
- Preparing for the course
- Expected workload
- Bring your own laptop
- What to do each week
- Course texts
- Getting datasets
- Attendance

Learning objectives  
Schedule  
Weekly structure / activities  
Assessment  
Getting help  
Giving feedback  
Attendance  
Etc...

**"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"**

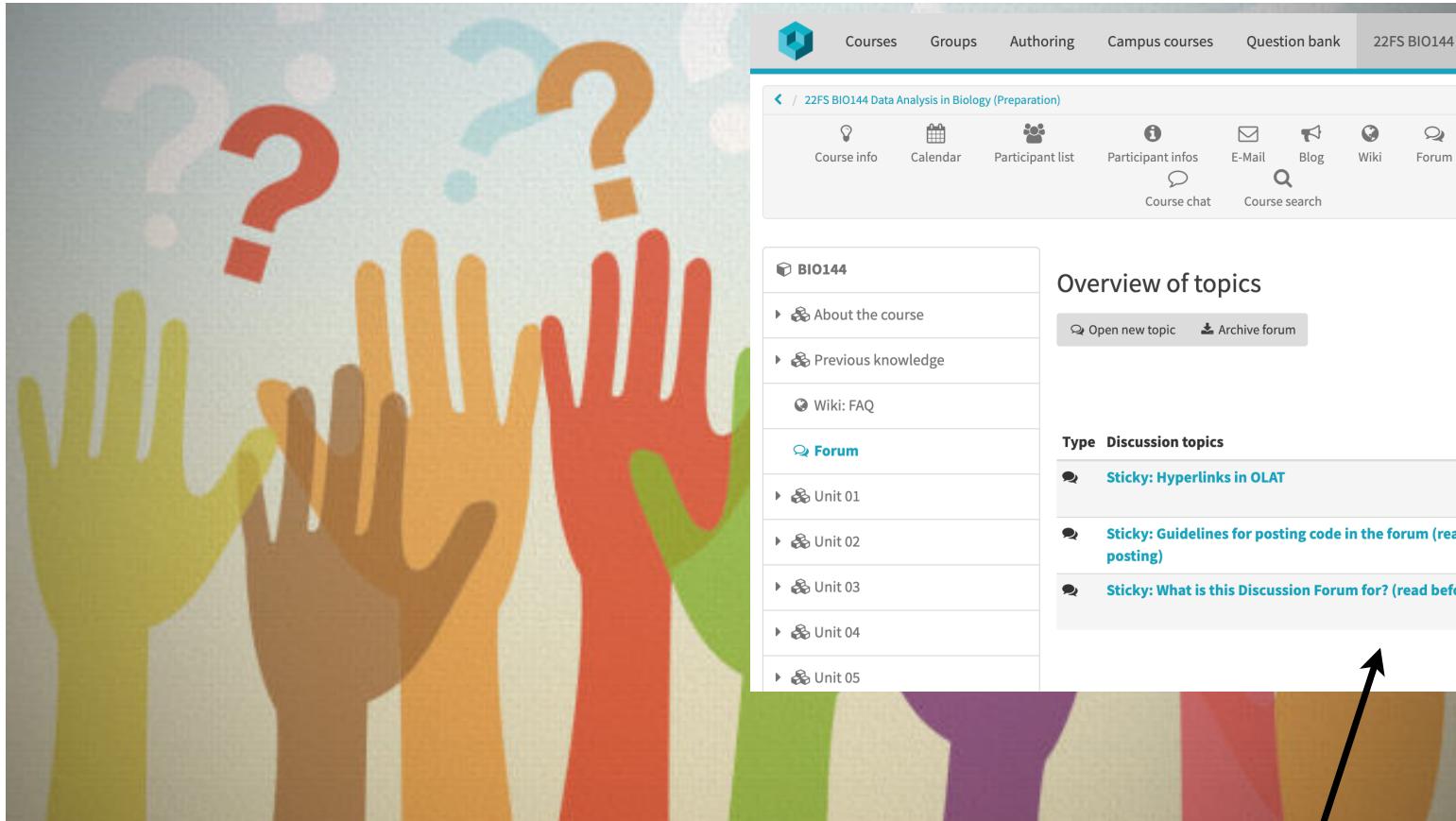
**"I like to try new things"**

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"!)







Courses Groups Authoring Campus courses Question bank 22FS BIO144 Da... X

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

Modifications OFF

| 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.



# AI Assistants

**Methods in Ecology and Evolution** BRITISH ECOLOGICAL SOCIETY

FORUM | Open Access | CC BY

**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. ”

# Methods in Ecology and Evolution



FORUM | [Open Access](#) |

## ChatGPT is likely reducing opportunity for support, friendship and learned kindness in research

[This article relates to:](#) ▾

Joseph Millard , Alec P. Christie, Lynn V. Dicks, Justin E. Isip, Thomas F. Johnson, Grace Skinner, Rebecca Spake



Cooper et al. (2024) write that ‘students may be happier to ask chatbots for help than to ask a human instructor, out of fear that the latter will judge them harshly for not knowing the answers to simple questions’. Sadly, this is likely true for many students.

# 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.

# What can I use an AI assistant for in BIO144?



# 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.

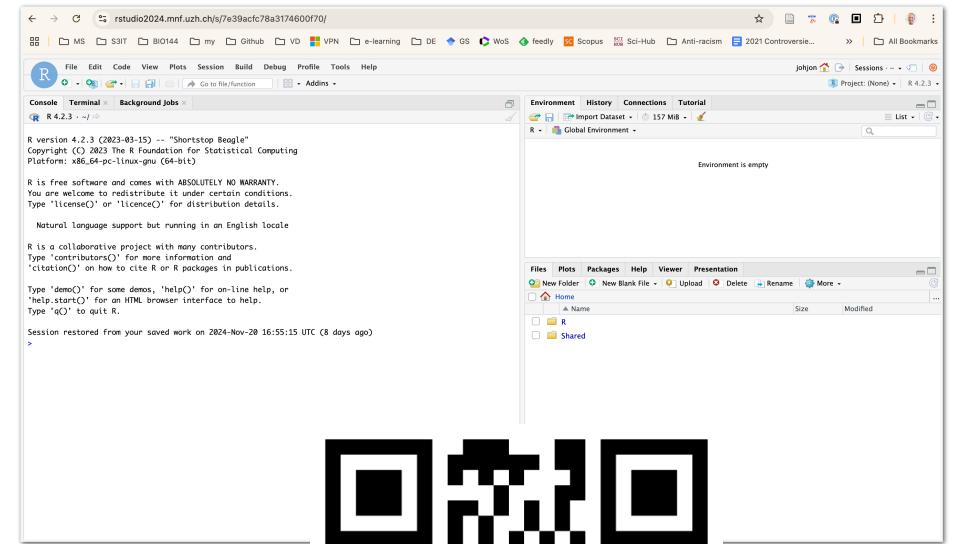
I suggest you try to minimize their use during the course.  
(Also try to minimize using the example solutions.)

**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.

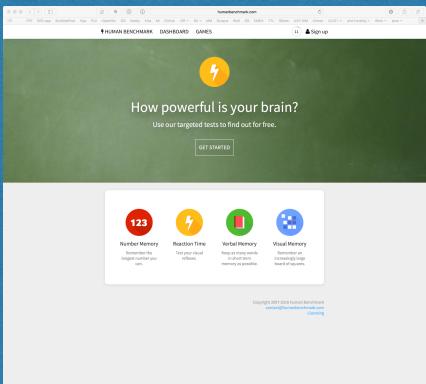


<https://rstudio2024.mnf.uzh.ch/>

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

***Do the tests here***

humanbenchmark.com



***Record the results here***

A screenshot of a Google Form titled "My Human Benchmark results". The form is designed for data entry. It includes fields for gender (radio buttons for Female, Male, or Other), average reaction time (text input field), verbal memory score (text input field), number memory score (text input field), and visual memory score (text input field). There is also a "SUBMIT" button at the bottom. A note at the bottom of the form states: "Never submit passwords through Google Forms."

Link is also in a Forum post on OLAT

# Activate your prior learning!

The screenshot shows a navigation menu on the left and several content sections on the right.

**Navigation Menu:**

- 22FS BIO144 Datenanalyse
- About the course
- Previous knowledge** (selected)
- 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

**Content Sections:**

- Introduction**

The aims for this section are:

  - To help you know about your knowledge of some of the things you should already know before you start BIO144.
  - To give you a chance to refresh your knowledge about some of the things you don't recall so well.
  - 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

$\bar{x}$  ("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.

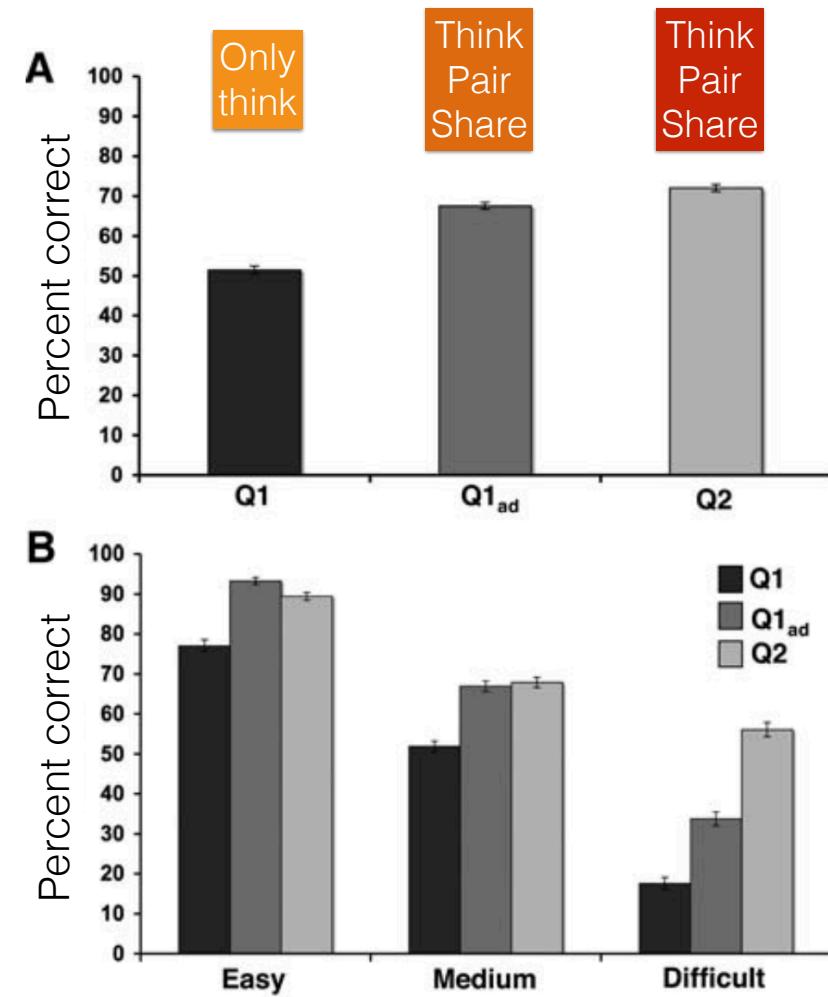
Models of relationships: equation for a straight line.

# 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.

# 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.
- Given that you know this pattern, and you are the subjects, its interesting to see if you women can buck the trend, perhaps by trying especially hard. Though now the men know you might do this, it probably won't work!

# 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

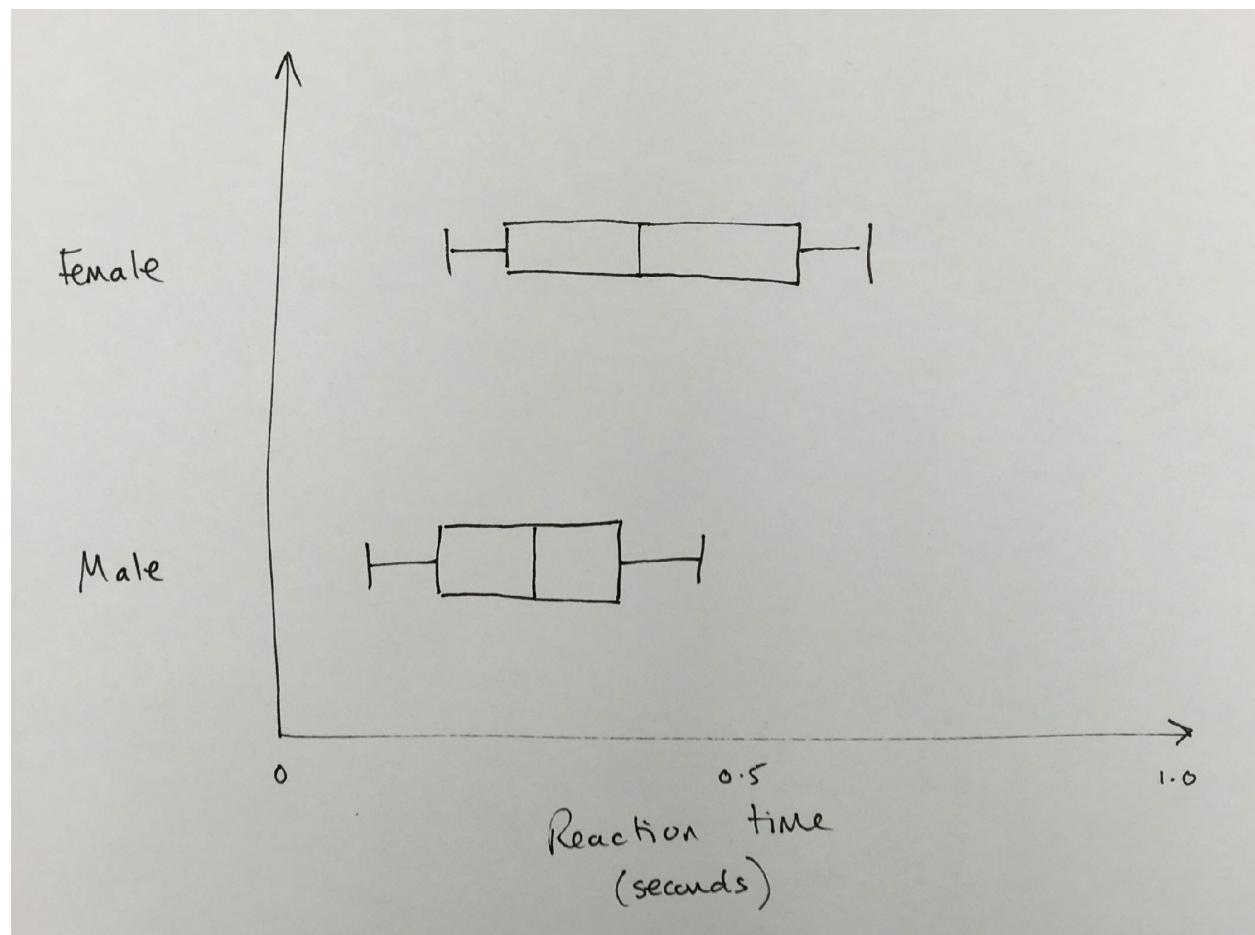
Data wrangling

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

# 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

Make up a unique ID code for yourself.  
It should not be anything that could identify you.  
Keep it safe.

# 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

humanbenchmark.com

The screenshot shows the Human Benchmark website. At the top, there is a navigation bar with various links like FRT, ERC app, ScribblePost, Hyp, PUI, rOpenSci, GG, feedly, Iriss, MI, Github, I2R, DE, MM, Scopus, WoS, GS, EMEH, TTL, RSeek, UZH WM, Altmet, OLAT+, phd funding, Work, and pers. Below the navigation bar, there is a sign-up button and a notification circle with 11 notifications.

The main header features a lightning bolt icon and the text "HUMAN BENCHMARK DASHBOARD GAMES". The main message on the page is "How powerful is your brain? Use our targeted tests to find out for free." with a "GET STARTED" button.

A pop-up window in the center contains four memory-related tests:

- Number Memory**: Represented by a red circle with the number "123". Description: "Remember the longest number you can."
- Reaction Time**: Represented by a yellow circle with a lightning bolt icon. Description: "Test your visual reflexes."
- Verbal Memory**: Represented by a green circle with a book icon. Description: "Keep as many words in short term memory as possible."
- Visual Memory**: Represented by a blue circle with a grid of squares icon. Description: "Remember an increasingly large board of squares."

At the bottom of the page, there is copyright information: "Copyright 2007-2016 Human Benchmark" and links to "contact@humanbenchmark.com" and "Licensing".

# 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

<https://forms.gle/o9kJh9dBjbJHXCTv6>

Link is also in a Forum post on OLAT

The screenshot shows a Google Forms survey titled "My Human Benchmark results". The survey is designed for a live data analysis demonstration in BIO144, Data Analysis in Biology. It includes fields for entering a unique ID code, gender (Female, Male, Other), average reaction time in seconds, and scores from Verbal and Number Memory tests. The form is set against a purple background.

My Human Benchmark results

For live data analysis demonstration, BIO144, Data Analysis in Biology

\*Required

Please enter the unique ID code you gave yourself. \*

Your answer \_\_\_\_\_

What is your gender? \*

Female

Male

Other: \_\_\_\_\_

Please enter your average reaction time in seconds (e.g., 0.326). \*

Your answer \_\_\_\_\_

Please enter your score on the Verbal Memory test. \*

Your answer \_\_\_\_\_

Please enter your score on the Number Memory test



# 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

Check the data in the spreadsheet

(The hyperlink might only work for Owen.)

reproducible

objective

robust

efficient

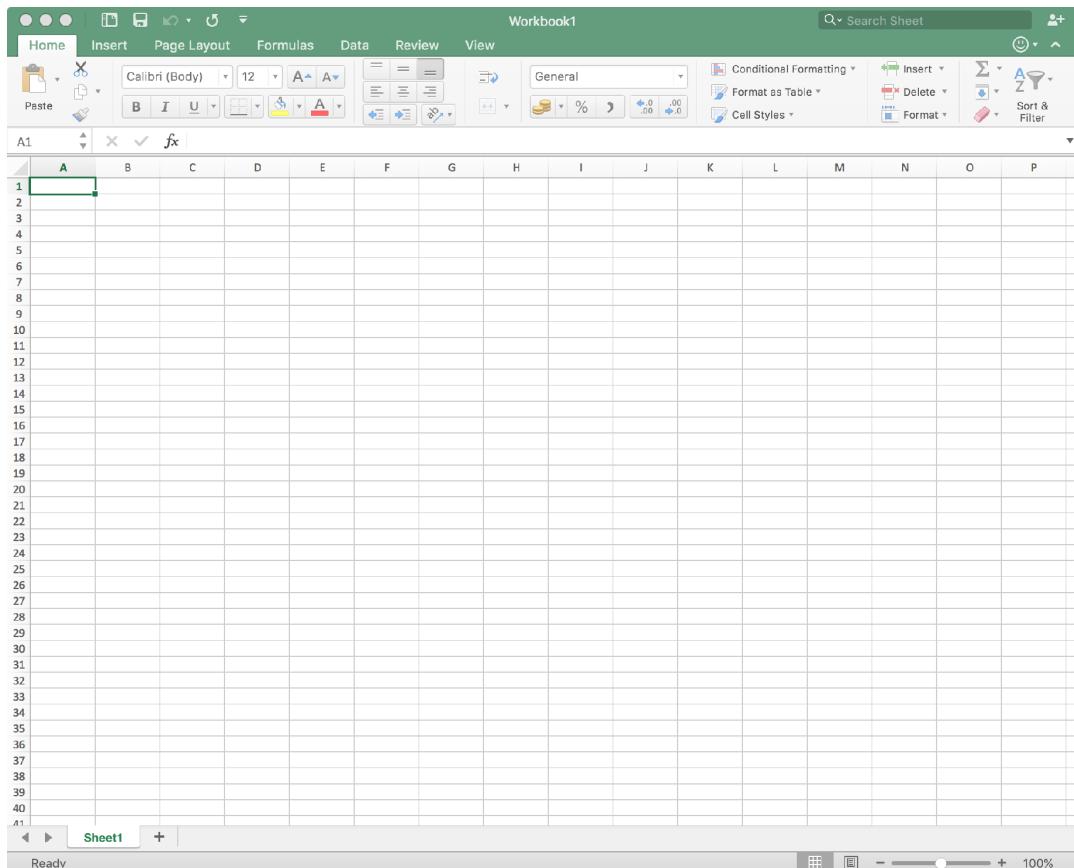
reliable

understandable

significant

shareable

fun



**Efficient**  
**Consistent**  
**Repeatable**  
**Reliable**  
**Readable**  
**Robust**  
**Persistent**  
**Sharable**  
**Scalable**

**Efficient**  
**Consistent**  
**Repeatable**  
**Reliable**  
**Readable**  
**Robust**  
**Persistent**  
**Sharable**  
**Scalable**



```
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]

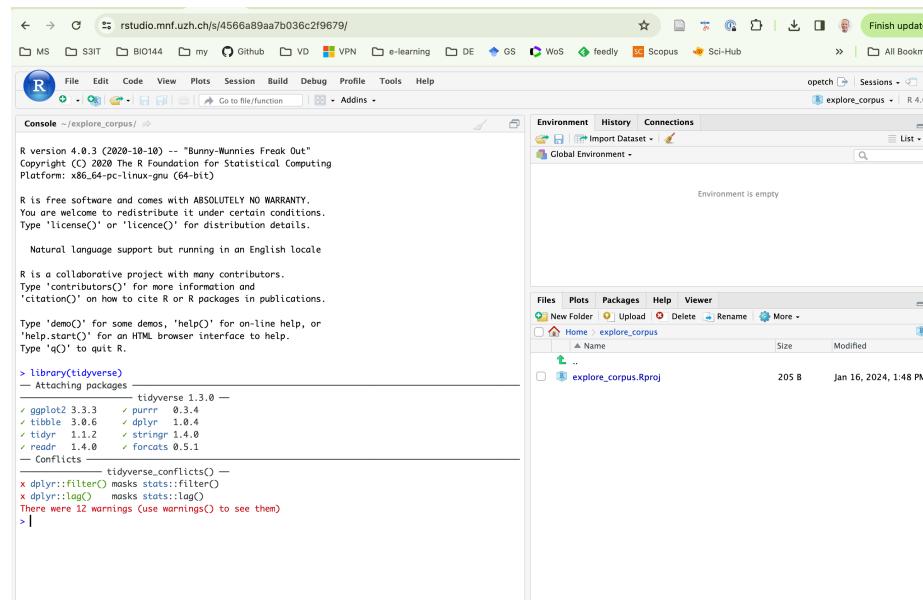
>
```

Some computers have great trouble installing and running R / Rstudio

If this seems to be you, then quickly switch to

*RStudio Cloud*

Ask a TA about this in the first practical



# 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

*With the help of ChatGPT*

# 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

*With the help of ChatGPT*

# 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

