

# Data Analysis in Biology

BIO144  
FS 2022



University of  
Zurich<sup>UZH</sup>

BIO  
144

## 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.



# Why is statistical data analysis so relevant for the biological and medical sciences?

Only with knowledge of data and statistical data analysis will it be possible to analyze your data from Bachelor, Master or PhD theses....

- ▶ **Medicine:** What is the effect of a drug? Which factors cause cancer?
- ▶ **Ecology:** What is a suitable habitat for a certain animal? Which resources does it need or prefer?
- ▶ **Evoloutionary biology:** Do highly inbred animals have decreased chances to survive or reproduce?



The “**hottest skill**”  
that got people  
**hired** in 2014?

## **Statistical Analysis**

Source: LinkedIn



0:06 / 3:00





# 6 Reasons To Learn R For Business [2021]

Written by Matt Dancho on December 17, 2020

## 6 Reasons to Learn R for Business

Why R Might Be the Right Choice for You

DS4B Tools: Capability Vs Learning Curve  
R has a longer learning curve but has a massive business capability rating

Business Science  
[www.business-science.io](http://www.business-science.io)

Get Articles in Your Inbox

email

Get Articles

Search for Articles

Search

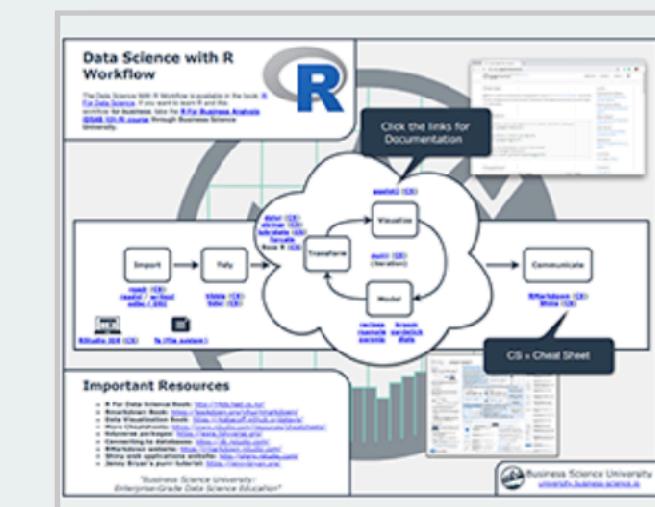


Find Articles By Category

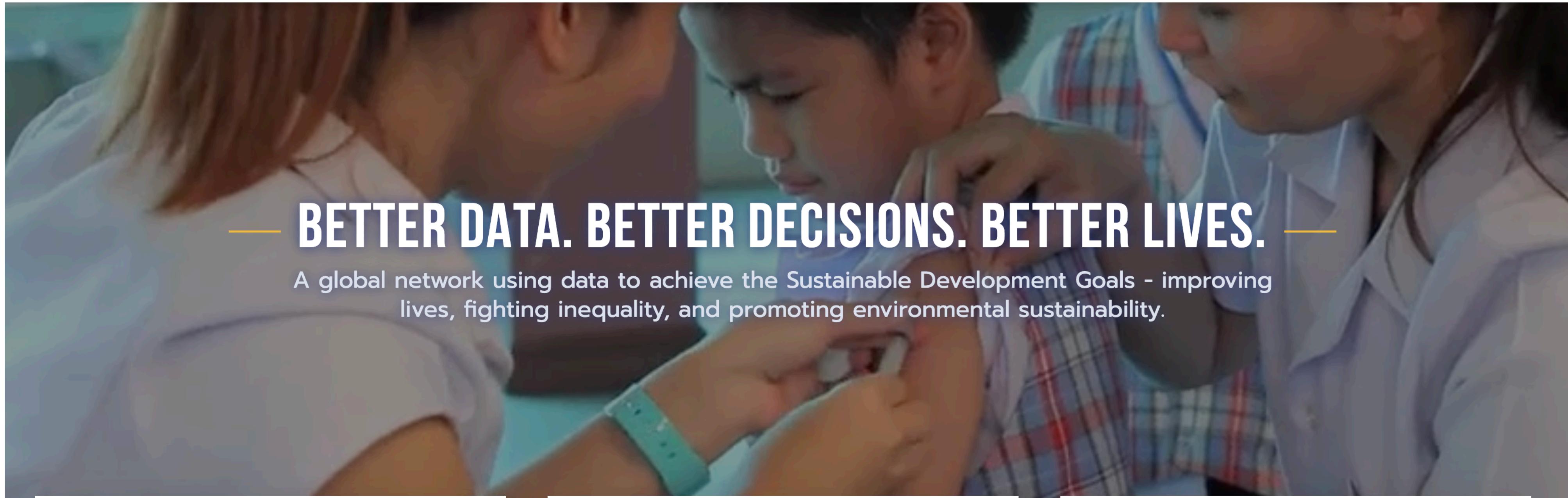


Learning Hub

Download Cheat Sheets



Cheat Sheets



# — BETTER DATA. BETTER DECISIONS. BETTER LIVES. —

A global network using data to achieve the Sustainable Development Goals - improving lives, fighting inequality, and promoting environmental sustainability.

## Who We Are

The Global Partnership for Sustainable Development Data is a global network working together to ensure the new opportunities of the data revolution are used to achieve the Sustainable Development Goals.

[ABOUT US](#)

## Our Community

Our hundreds of partners from governments, the private sector, and civil society organizations are joining forces to take action, galvanize political commitment, build trust, and spur innovation in the booming data ecosystems of the 21st century.

[SEE OUR PARTNERS](#)

## Our Impact

Since our founding in 2015, our network has improved data to monitor and achieve the Sustainable Development Goals, created incentives for new commitments to fund and share data, and enabled knowledge-sharing, bringing partners together to make change.

[LEARN MORE](#)

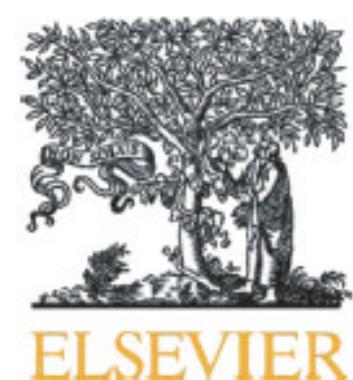
# Examples of insights from data

## Otter (*lutra lutra*)

*Research questions:* What is the preferred habitat by otters? How do otters adapt to human altered landscapes?

*Method:* Study in Austria, 9 Otter were radio-tracked and monitored during 2-3 years.

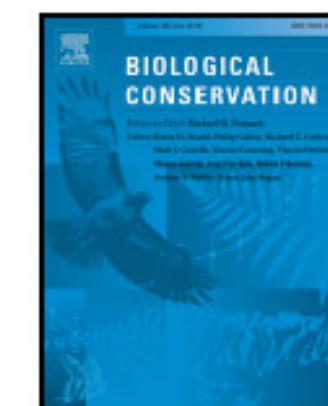
[Biological Conservation 199 \(2016\) 88–95](#)



Contents lists available at [ScienceDirect](#)

Biological Conservation

journal homepage: [www.elsevier.com/locate/bioc](http://www.elsevier.com/locate/bioc)



Flexible habitat selection paves the way for a recovery of otter populations in the European Alps



Irene C. Weinberger <sup>a,\*</sup>, Stefanie Muff <sup>a,b</sup>, Addy de Jongh <sup>c</sup>, Andreas Kranz <sup>d</sup>, Fabio Bontadina <sup>e,f</sup>

<sup>a</sup> Institute of Ecology and Evolutionary Biology, University of Zurich, Winterthurerstr. 190, 8057 Zurich, Switzerland

<sup>b</sup> Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland

<sup>c</sup> Dutch Otterstation Foundation, Spanjaardslaan 136, 8917 AX Leeuwarden, Netherlands

<sup>d</sup> alka-kranz Ingenieurbüro für Wildökologie und Naturschutz, Am Waldgrund 25, 8044 Graz, Austria

<sup>e</sup> SWILD – Urban Ecology & Wildlife Research, Wührstr. 12, 8003 Zurich, Switzerland

<sup>f</sup> Swiss Federal Research Institute WSL, Biodiversity and Conservation Biology, 8903 Birmensdorf, Switzerland

# Mercury (Hg) in the soil

## Wohnzone im Wallis von Quecksilber vergiftet

Vor über vierzig Jahren hatten 3,1 Tonnen Quecksilber einen Abflusskanal nahe der Walliser Gemeinde Visp verschmutzt. Noch heute müssen die Einwohner mit den Folgen leben.



---

### Artikel zum Thema

#### Konvention gegen Quecksilber verabschiedet

Ein neues internationales Abkommen schränkt die Verwendung von Quecksilber in der Industrie ein. Massgeblich daran beteiligt war die Schweiz. [Mehr...](#)

19.01.2013

*Research question:* Is the Hg level in the environment (soil) of people's homes associated to the Hg levels in their bodies (urin, hair)?

*Method:* Measurements of Hg concentrations on people's properties, as well as measurements and survey of children and their mothers living in these properties.

Highly delicate, emotionally charged, political question!

► Schweiz Aktuell, 20. Juni 2016

# Physical activity in children (Splashy study)



University of  
Zurich<sup>UZH</sup>

BIO  
144



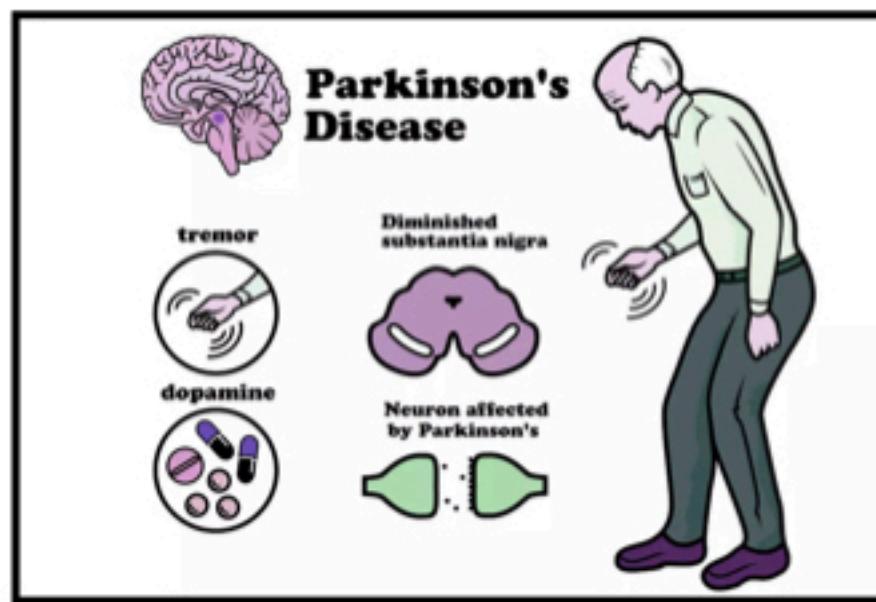
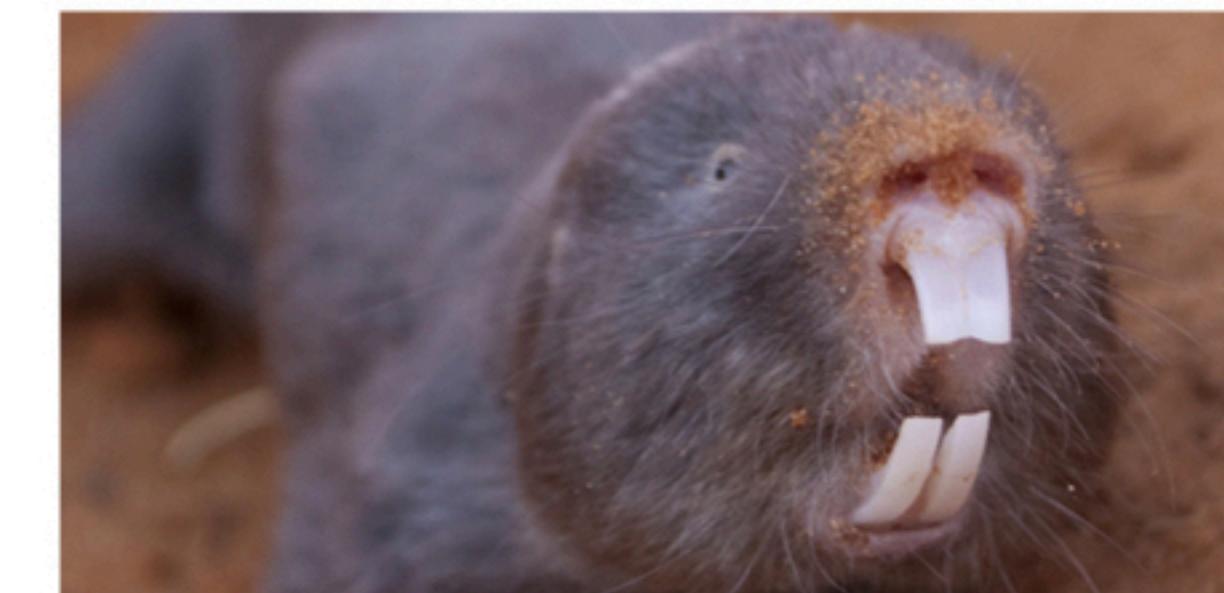
[splashy.ch](http://splashy.ch)

*Research question:* Which factors influence physical activity patterns in children aged 2-6 years?

*Method:* The children had to wear accelerometers for several days. In addition, their parents had to fill in a detailed questionnaire.

Observed variables were, e.g., media consumption, behavior of the parents, age, weight, social structure,...

# Question you will work on



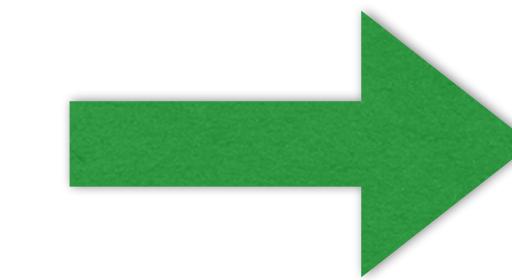
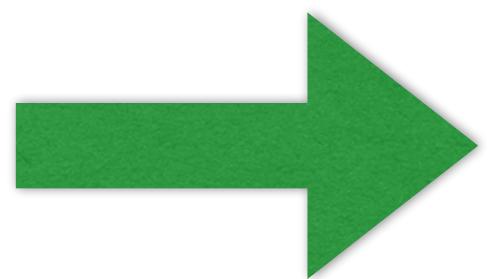
# Producing nonsense with statistics... .

... is too easy ...

The risks of alcohol (by David Spiegelhalter, 23. August 2018)

“Calling bullshit” course (University of Washington)

A profound knowledge of data analysis and statistics protects you from producing nonsense – and helps to detect it.



Question  
Puzzle  
Problem

Data  
+  
Analysis

Answer  
Solution

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

[shorturl.at/dmsS7](http://shorturl.at/dmsS7)



$\bar{x}$ -mean, "sd"

reproducible

objective

robust

efficient

reliable

understandable

significant

Sharable

fun

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

—  
—  
—





# How do we get insights from data... .

... rather than nonsense.

- ▶ Excellent data management practices.
- ▶ Informative graphical visualisations to explore data.
- ▶ Informative numerical summarise to explore data.
- ▶ Appropriate transformations of data.
- ▶ Appropriate statistics tests / models.

Awareness of our “realm”:

- ▶ Description of patterns, including associations (we will do this).
- ▶ Predicting (we will do this).
- ▶ Inferring causation (we will do this, by analysing experiments randomised manipulations).



"Failure is an opportunity to grow"

## GROWTH MINDSET

"I can learn to do anything I want"

"Challenges help me to grow"

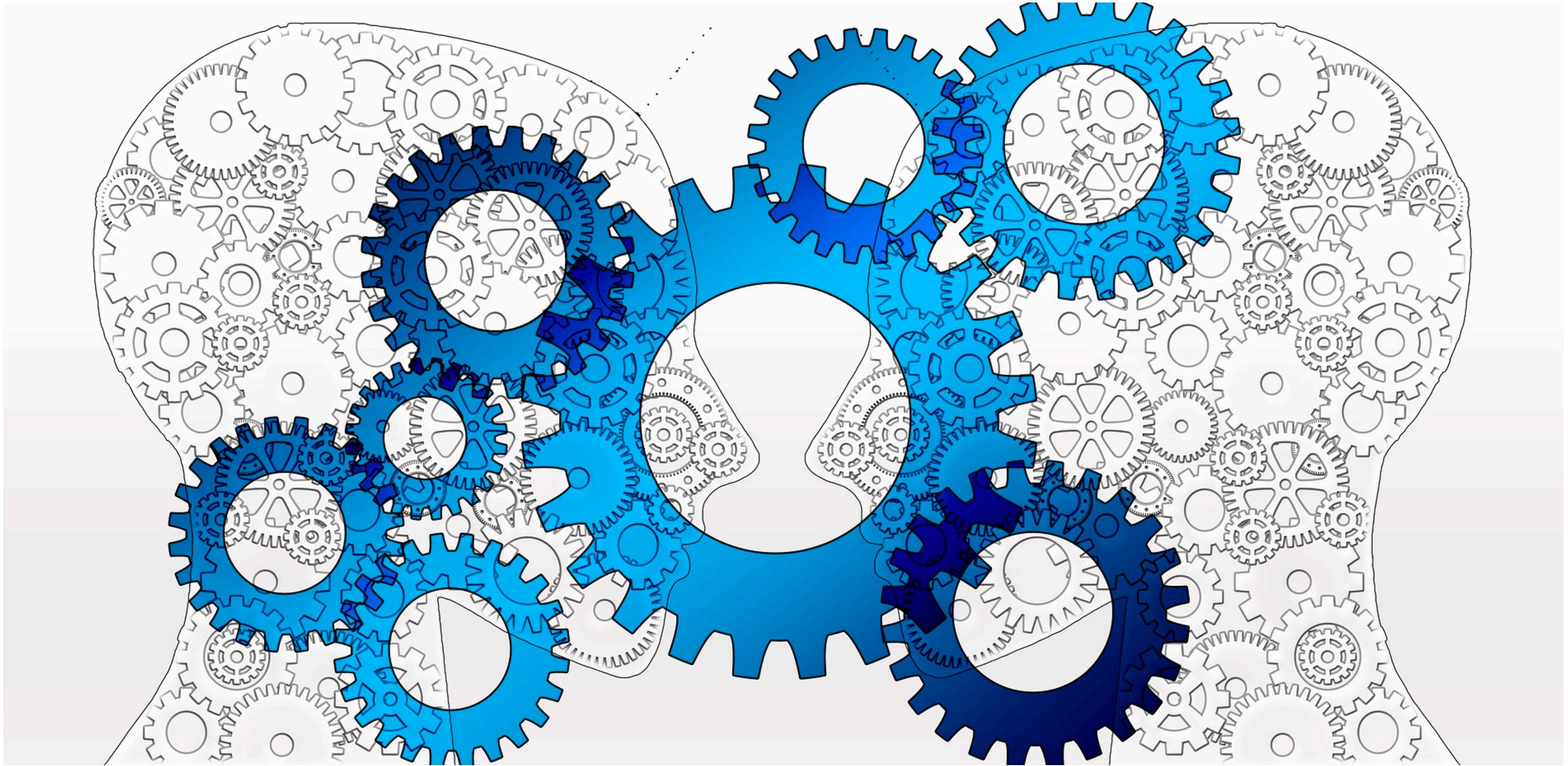
"My effort and attitude determine my abilities"

"Feedback is constructive"

"I am inspired by the success of others"

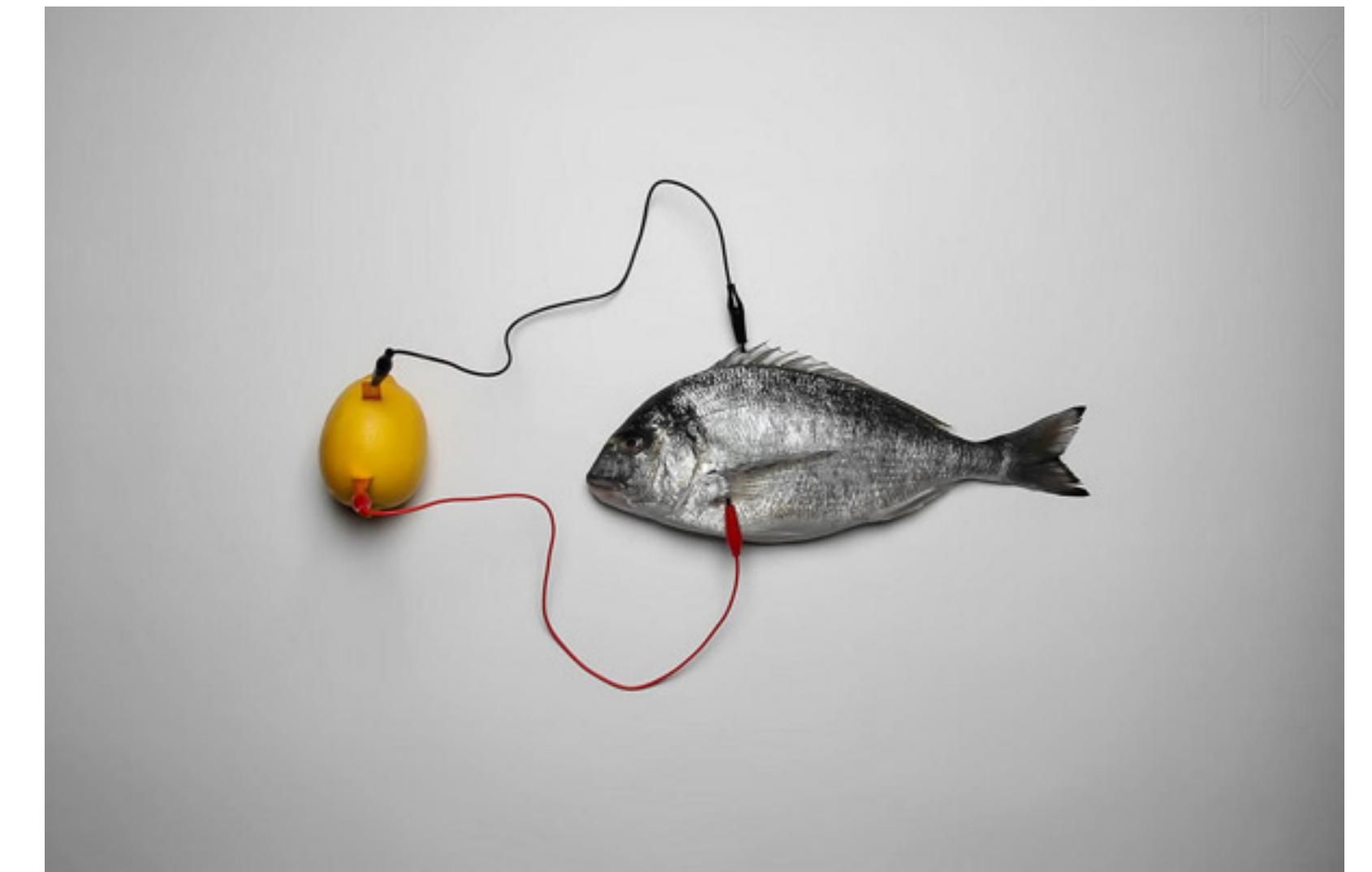
"I like to try new things"

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



# How the course is organised

# The conceptual side



1 - 2:45pm Mondays

# The practical side



1 - 3pm Thurs. or Fri.

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

# Details on OLAT

The screenshot shows the OLAT interface for a course titled "22FS BIO144 Da...". The top navigation bar includes links for Courses, Groups, Authoring, Campus courses, Question bank, and a search bar. Below the navigation is a toolbar with icons for Administration, Status, Course info, Calendar, Participant list, Participant infos, E-Mail, Blog, Wiki, Forum, User role, and My course. A sidebar on the left lists course sections: 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...

Owen sent a welcome email

# 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 RS
  - What are add-on pac
- 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

# Graded assessment questions

The screenshot shows a course navigation bar on the left and an assessment page on the right.

**Course Navigation Bar:**

- BIO144
- About the course
- Previous knowledge
- Wiki: FAQ
- Forum
- Unit 01** (selected)

  - Lecture 1
  - Important info
  - Homework
  - Practical

- Graded Assessment 1** (selected)

  - Assessment 1 (graded)** (selected)

- Unit 02
- Unit 03
- Unit 04

**Assessment 1 (graded) Page Content:**

**Information:**

- This is a graded test.
- You can start the test only once, but once you started it you can suspend it if you need to. This pauses the test and you can leave it. Once you start the test again you will pick it up exactly where you left it.
- For each question you have 2 attempts (unless otherwise specified).
- If not specified otherwise, each question gives 1 point.
- The solutions are given after you submit the test.

Good luck!

There is no score information of this test since you have not taken it yet.  
Maximum number of attempts: 1

Press the start button to begin with your test.  
Results of this test are visible to administrators and tutors of this course.

**Start**



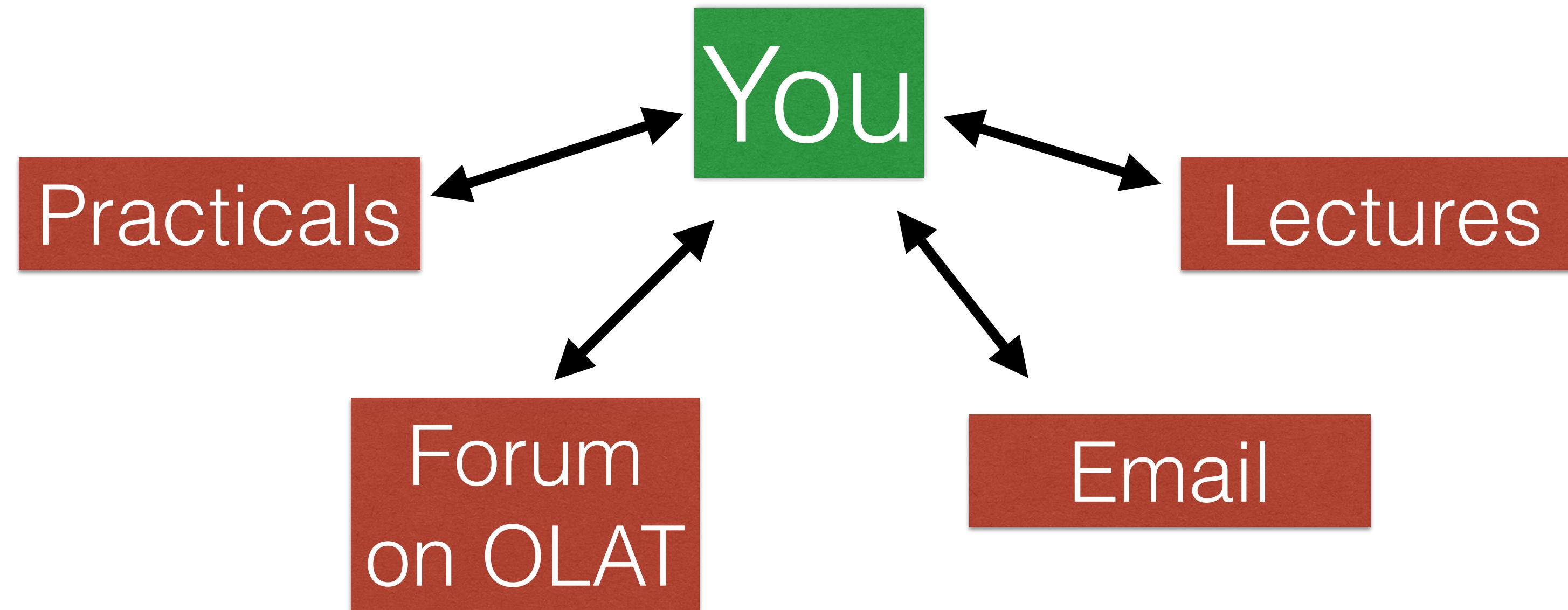
Hyperlink

Some “play” questions

# 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 some general questions also.
- 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 OLAT on your own computers.
- As well as usual examination rules, you are not allowed to make screenshots of the examination questions. Anyone seen doing this will be excluded from the examination.

# Communications and the course team



Dr Erik Willems  
Instructor



Prof. Owen Petchey  
Director  
Instructor



Uriah Daugaard  
Practicals  
Teaching Assistants  
Statistics consultant



Martina Jelic  
Practicals  
Teaching Assistants  
Statistics consultant



Dr Frank Pennekamp  
Assessment  
Examinations



Dr Rainer Krug  
Computing consultant



Anthony  
Sonrel,  
Practicals  
facilitator

Numerous  
teaching  
assistants

# Numerous teaching assistants





## FEEDBACK

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

### Feedback

What is in the course and exam was unclear.

Some variation in amount and type of BC material.

A summary of R commands and corresponding library would be good.

Diverse resources: various books, documents, videos, seem a bit thrown together.

Solution scripts: Please give us solution scripts (promptly). Having to wait a week for help is demotivating (please give solution scripts).

Lack of feedback with IC exercises: some do not have instant-feedback, no way to check if we made the correct analysis or graph (give us a solution script). "Felt a bit alone."

Amount of work: The IC parts were too short, add 30 mins. All the material is a bit too much. BC reading sometimes rather long. Some variation in amount and type of BC material.

Video/podcast quality: in 2020, podcasts from 2019 had to be used. They were of poor quality. In 2021 all lectures will be live and recorded, with good quality.

Critical analysis: would be good to work through a paper together, to assess the data analytic methods used.

Prof. Petchey talks for hours on end.

### Mitigation

**Learning Objectives. Mock exam.**

**Refer to Learning Objectives for what is core.  
Please ask if you are uncertain.**

**Students should make one as they go along.**

**Refer to Learning Objectives for what is core.  
Please ask if you are uncertain.**

**Example solution scripts available.**

**Work on the practical parts at the same time that help is available,  
i.e. during the scheduled practical sessions.**

**Work together in study groups**

**Work steadily, attend live lectures, attend practicals,  
keep up, ask for help.**

**All lectures will be live (synchronous), will be recorded, and will be made available.**

We have not had time to implement this.  
It would require removing other content, so is not a simple change to make.

**Prof. Petchey tries to be more concise.**

**We have other lecturers: Prof. Damien Farine, Dr Hanja Brandl**

You

Us





Courses Groups Authoring Campus courses Question bank 22FS BIO144 Da... 0/181

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

**BIO144**

- ▶ About the course
- ▶ Previous knowledge
- Wiki: FAQ

**Forum**

- ▶ Unit 01
- ▶ Unit 02
- ▶ Unit 03
- ▶ Unit 04
- ▶ Unit 05

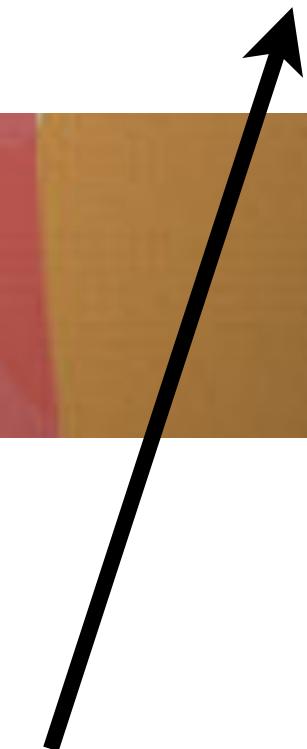
**Overview of topics**

Open new topic Archive forum

Search

3 Entries

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



Ask questions / make requests here.

Short break

# 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

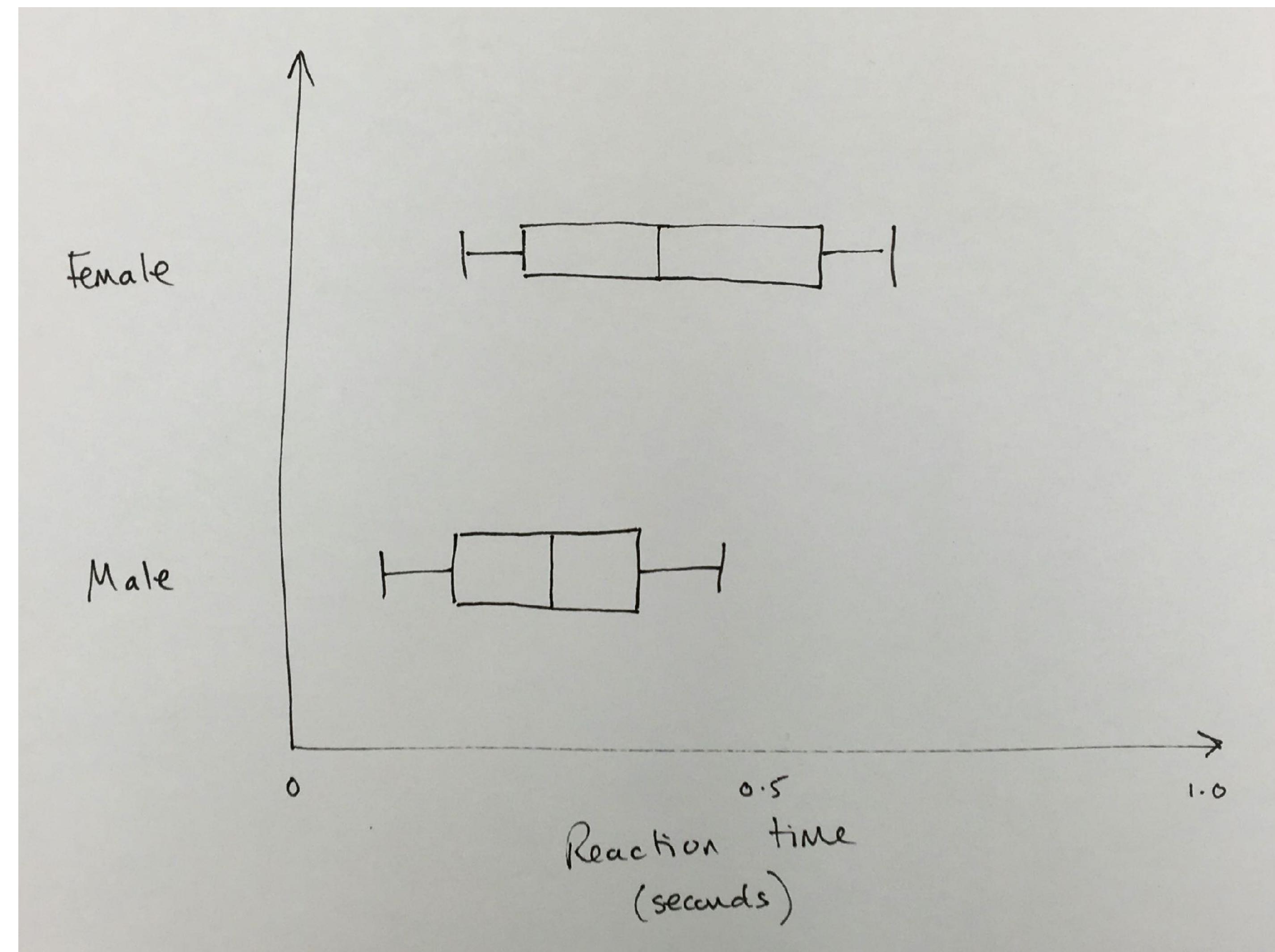
Data wrangling

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

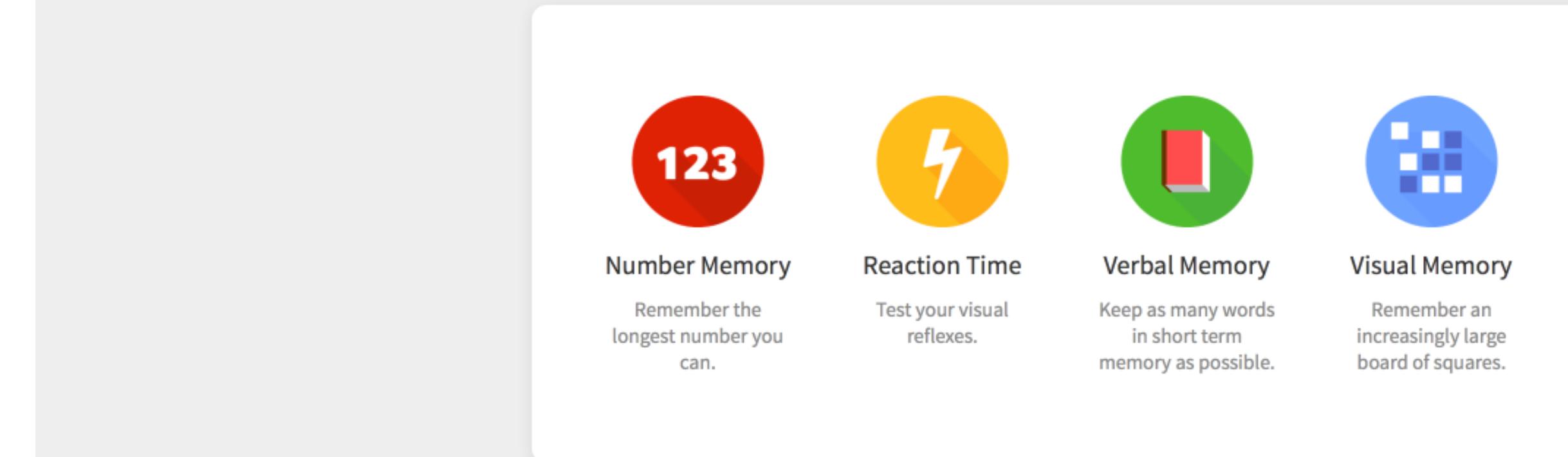
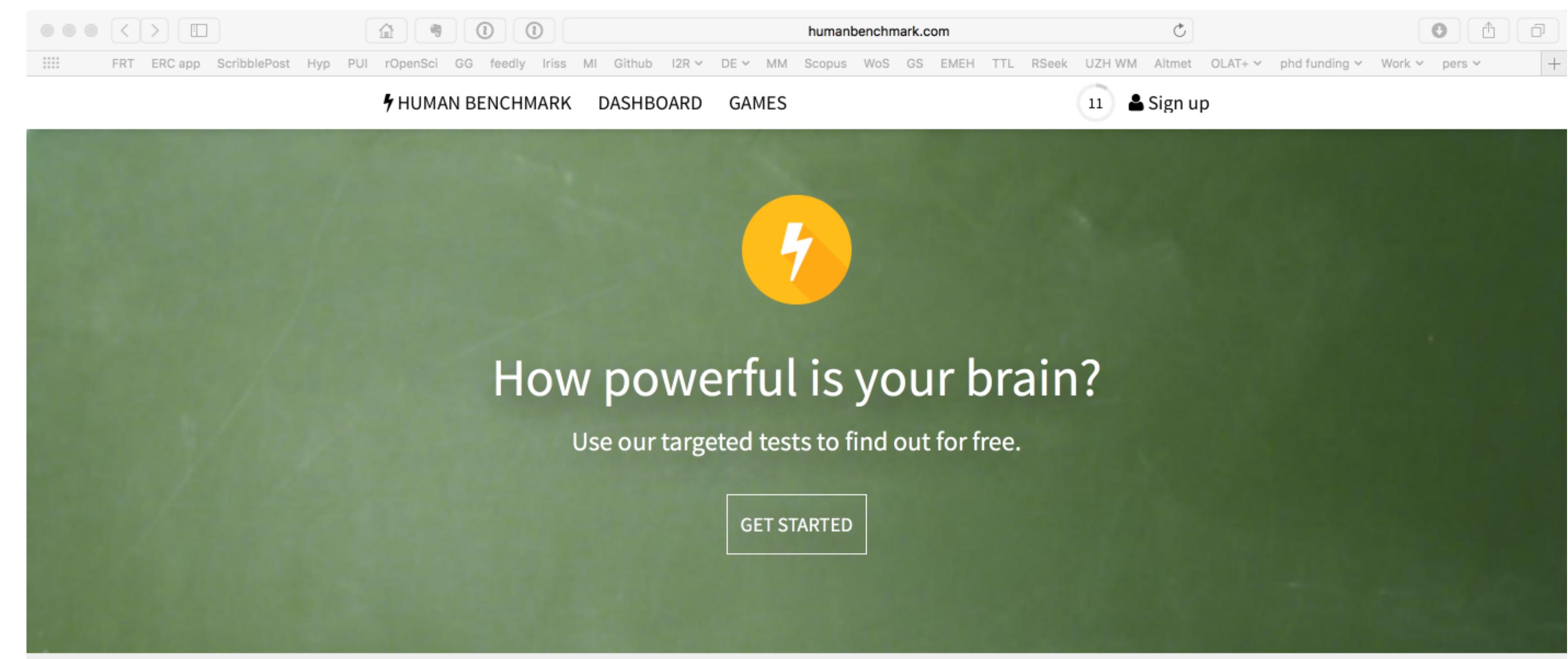
Visualise

Statistical test

Critical thinking

Report / communicate

humanbenchmark.com



# 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/XE88JJXrUdNSD4Vy5>

(Link also in OLAT, Lecture 1 page)

The screenshot shows a Google Forms survey titled "My Human Benchmark results". The survey is intended 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, verbal memory test score, and number memory test score.

**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

Your answer

# 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**

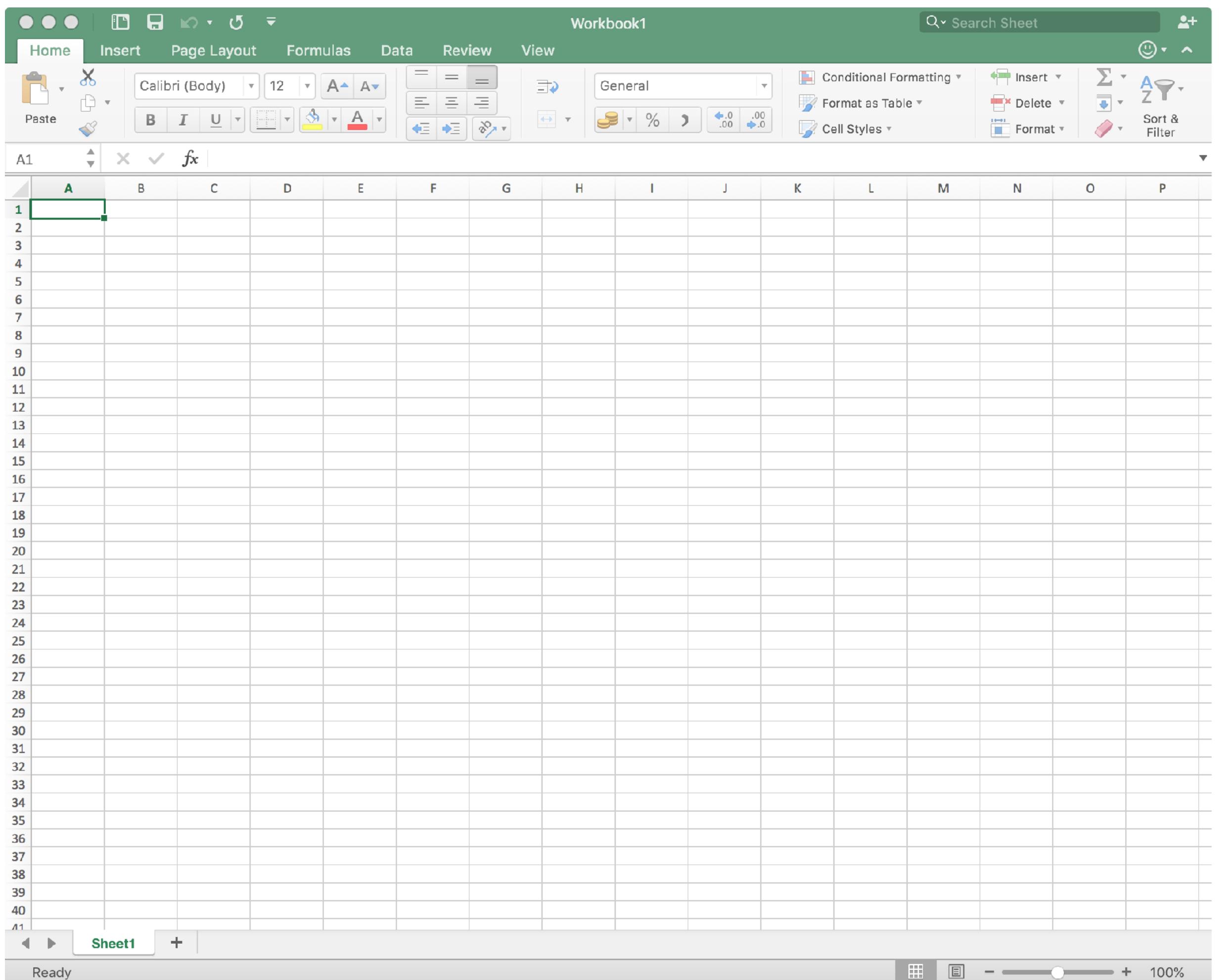
- Data wrangling
- Visualise
- Statistical test
- Critical thinking
- Report / communicate

Check the data in the spreadsheet

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

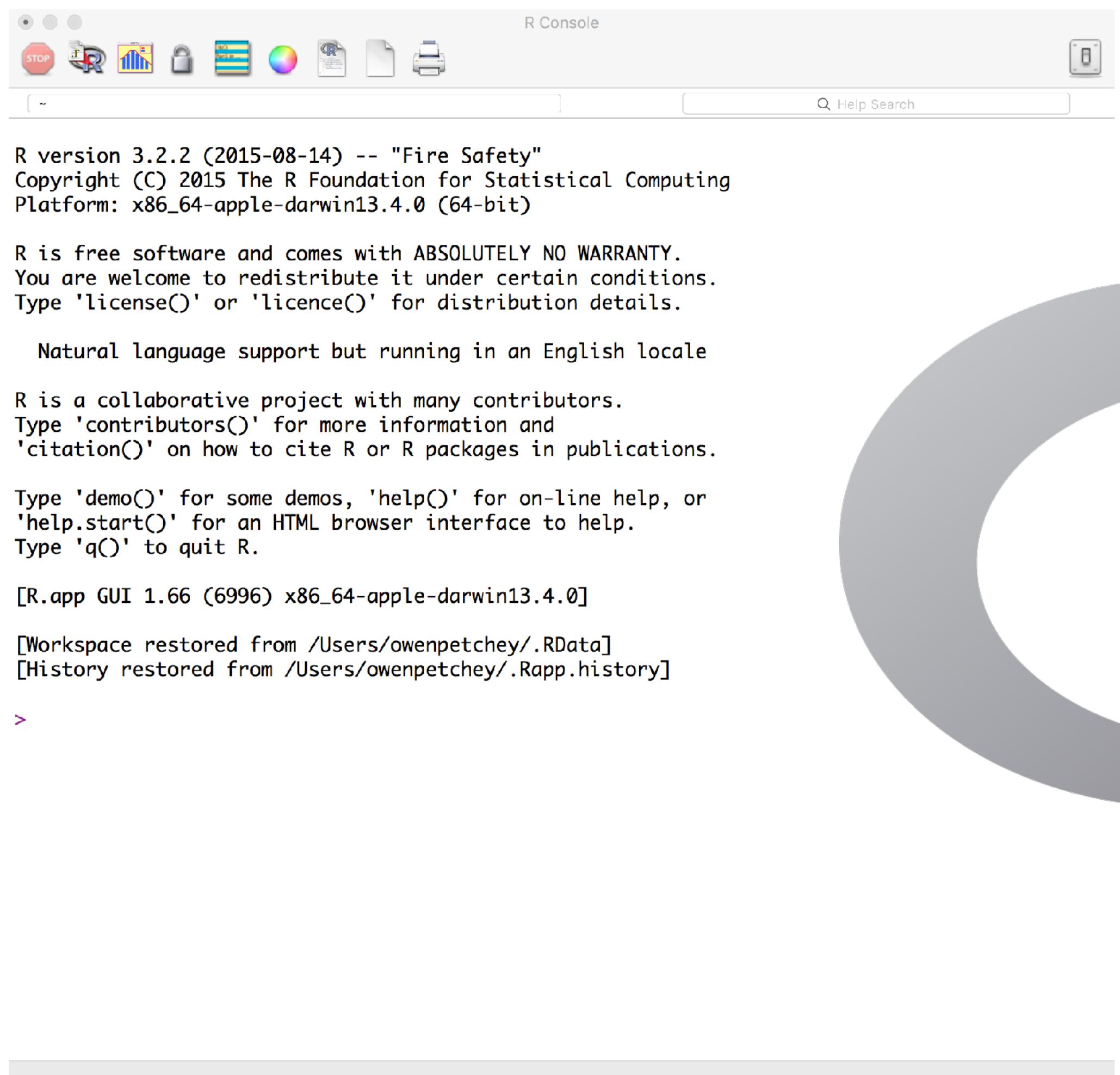
—  
—





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

**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

# 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

