

Data Analysis in Biology

BIO144
FS 2022



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?



Now also in Zürich!

- who?** genders underrepresented in the R community.
R beginners to experts are welcome!
- why?** encouraging, inspiring and empowering
R users from minority genders.
- how?** regular events where we learn more about R
and grow as a community.

R help forum on Slack to promote exchange
and mutual help among our members.

**Ladies, register for our first event on Nov 15th
on MeetUp! We can't wait to get to know you!**





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

Get Articles in Your Inbox

email

Get Articles

Search for Articles

Search

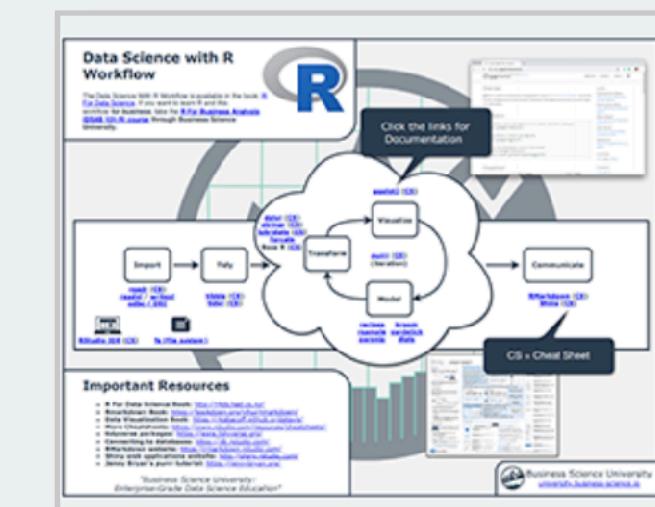


Find Articles By Category



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.

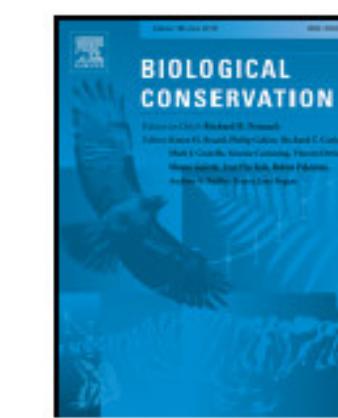
Biological Conservation 199 (2016) 88–95



Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/bioc



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



CrossMark

Irene C. Weinberger ^{a,*}, Stefanie Muff ^{a,b}, Addy de Jongh ^c, Andreas Kranz ^d, Fabio Bontadina ^{e,f}

^a Institute of Ecology and Evolutionary Biology, University of Zurich, Winterthurerstr. 190, 8057 Zurich, Switzerland

^b Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland

^c Dutch Otterstation Foundation, Spanjaardslaan 136, 8917 AX Leeuwarden, Netherlands

^d alka-kranz Ingenieurbüro für Wildökologie und Naturschutz, Am Waldgrund 25, 8044 Graz, Austria

^e SWILD – Urban Ecology & Wildlife Research, Wührstr. 12, 8003 Zurich, Switzerland

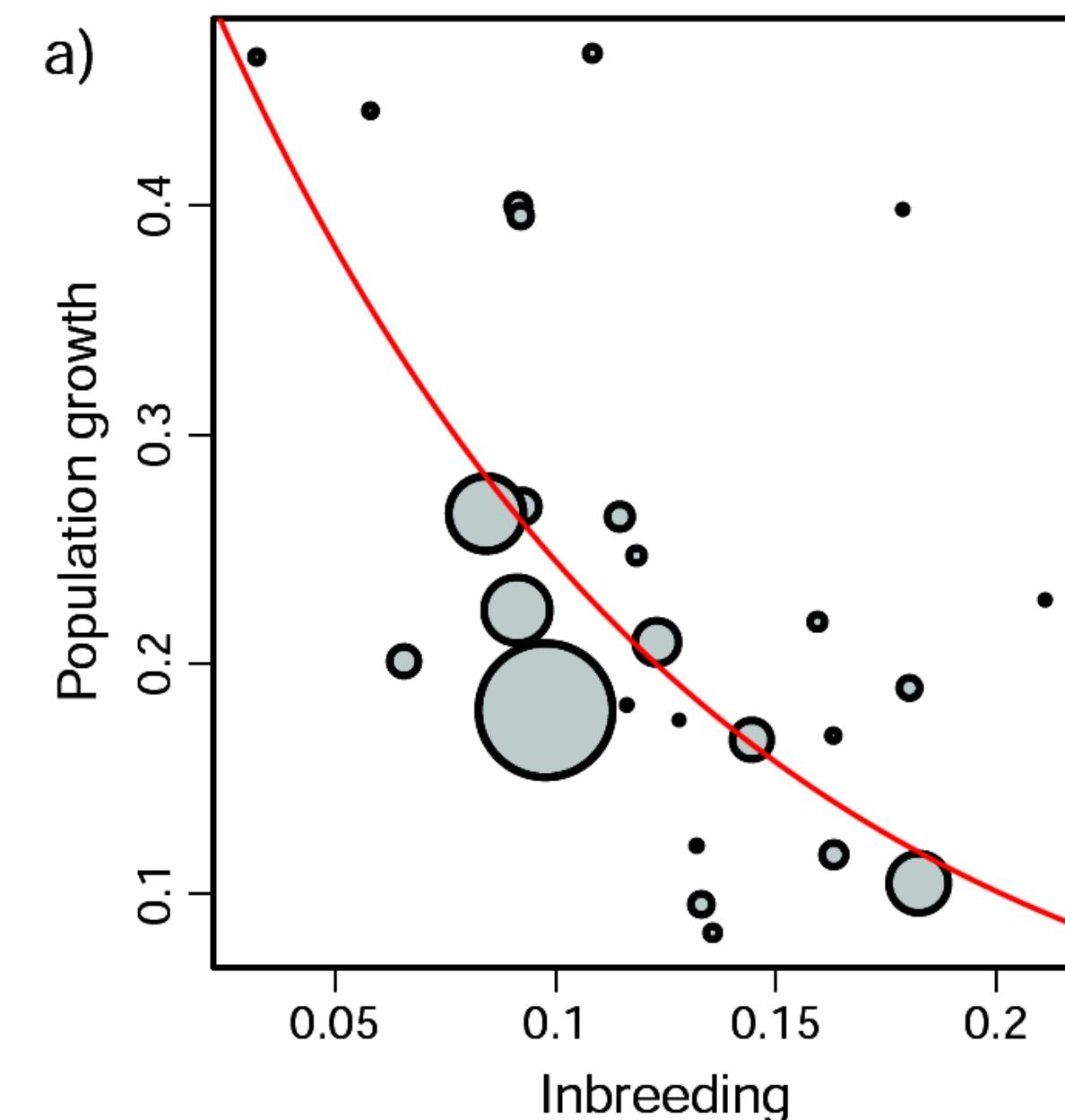
^f Swiss Federal Research Institute WSL, Biodiversity and Conservation Biology, 8903 Birmensdorf, Switzerland

<http://www.prolutra.ch/>

Inbreeding in Alpine ibex

Research question: Does inbreeding in Alpine ibex populations have a negative effect on long-term population growth? Inbreeding depression!

Methods: Genetic information from blood samples allow to quantify the level of inbreeding in each ibex population. In addition, long-term monitoring of population sizes and harvest rates.



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^{UZH}

BIO
144



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,...

Statistics in the news (April 2016)

NZZ am Sonntag 3. April 2016 | **Wissen** | 61

Überschätzte Statistiken

Daten-Analysen entscheiden heute darüber, ob ein Medikament als wirksam gilt. Bloss verstehen viele Forscher die Bedeutung dieser Berechnungen gar nicht. **Von Patrick Imhasly**

Karriere machen können nicht nur Menschen, sondern auch statistische Größen. Das gilt besonders für den sogenannten p-Wert, mit dem jeder Mittelschüler und jede Studentin in Kontakt kommt, vor allem aber jeder, der im weitesten Sinn etwas mit Statistik zu tun hat. Inzwischen ist der p-Wert indes auf die schiefe Bahn geraten. Denn was der Vater der modernen Statistik, der britische Genetiker Ronald Fisher, 1920 als eine Art informelles Sensorium für die Aussagekraft von Daten entwickelte, ist in der Praxis oftmals zu einem simplen Lackmust verkommen.

Ergibt die statistische Analyse von Daten einen $p < 0,05$ (5 Prozent) oder noch besser $< 0,01$ (1 Prozent), gelten diese als signifikant – den Daten wird dann automatisch Beweiskraft zugestanden. Das entscheidet etwa darüber, ob ein neues Medikament als wirksam eingestuft wird oder ob ein Forstner seine Studie in einem angesehenen Fachblatt publizieren kann. «Der p-Wert war aber nie dazu gedacht, wissenschaftliches Denken außer Kraft zu setzen», hat sich Ron Wasserstein, der Direktor der Amerikanischen Statistischen Vereinigung (ASA), jüngst öffentlich beklagt.

Wissenschaftler verwenden den p-Wert immer häufiger, ohne zu verstehen, was er bedeutet – das fördert schlechte Forschung und untergräbt die Glaubwürdigkeit der Wissenschaft. Der Mediziner und Epidemiologe John Ioannidis von der Universität Stanford sprach in einem Kommentar von «Drogen-abhängigen»: «Der falsche Gebrauch des p-Wertes ist dermaßen einfach und erfolgt so automatisiert, dass manche süchtig werden.

«Studien führen heute zu dermassen vielen Daten, dass man allen Unfug testen kann und so zu Hunderten von p-Werten kommt.»

werden danach – vor allem wenn sie mit Forschungsgeldern und Publikationen belohnt werden.» Angesichts der Missstände sah sich die ASA jetzt veranlasst, zum ersten Mal in ihrer fast 180-jährigen Geschichte Empfehlungen zu veröffentlichen, wie man mit einer statistischen Größe vernünftig umgeht.

Wider die Null-Hypothese

«Der p-Wert sagt nicht das aus, was man gemeinhin von ihm erwartet», erklärt der Berner Epidemiologe Peter Jüni, der seit kurzem am Applied Health Research Centre der Universität Toronto tätig ist. Das bedeutet: Der p-Wert misst nicht die Wahrscheinlichkeit, ob eine bestimmte Hypothese zutrifft, und auch nicht, ob ein bestimmtes Resultat zufällig zustande gekommen ist, wie die ASA festhält. Vielmehr dienen p-Werte dazu, die sogenannte Null-Hypothese zu testen und möglichst zu verworfen.

Die Hypothese in einer Patientenstudie könnte zum Beispiel lauten, dass ein Medikament A gegen Nierenprobleme besser wirkt als ein Medikament B. Die Null-Hypothese besagt dann genau das Gegenteil davon, nämlich dass das Medikament A nicht besser wirkt als das Medikament B. Beim Test berechnet der Forstner im Prinzip, wie gross die Wahrscheinlichkeit für das Auftreten eines tatsächlich festgestellten oder noch grösseren Unterschieds zwischen den beiden untersuchten Medikamenten ist – unter der Annahme, dass die Null-Hypothese stimmt. Diese Wahrscheinlichkeit ist der p-Wert, und je geringer er ist, desto weniger spricht für die Null-Hypothese. Ein p-Wert von 0,05 bedeutet, dass das festgestellte (oder ein noch extremeres) Resultat unter den Bedingungen der Null-Hypothese mit einer Wahrscheinlichkeit von lediglich 5 Prozent zustande kommen kann – und nicht, dass eine bestimmte Hypothese mit einer Sicherheit von 95 Prozent wahr ist.

Über die eigentlich untersuchte Hypothese kann der p-Wert nur indirekt etwas aussagen, weil er über zwei Ecken gedacht ist. Der omnöse Wert liefert also keine ultimativen Beweise für einen postulierten Unterschied oder Zusammenhang. «Der p-Wert ist eine bedingte und nicht eine absolute Wahrscheinlichkeit», erklärt Peter Jüni. «Doch genau das verstehen viele Forstner nicht, und es interessiert sie auch nicht.»

Hinzu kommt, dass die Signifikanzgrenzen von 5 Prozent bzw. 1 Prozent historisch entstanden und keineswegs klar definierte Werte sind. Ronald Fisher, der Erfinder des p-Werts, überließ es jedenfalls ausdrücklich der Interpretation des einzelnen Forstners, ab welcher Grösse ein p-Wert in einer Untersuchung Aussagekraft haben soll. «Trotzdem haben sich die willkürliche gewählten Signifikanzgrenzen ins Gehirn von Generationen von Forstnern gebrannt», sagt Leonhard Held vom Institut für Epidemiologie, Biostatistik und Prävention der Universität Zürich.

Die britischen Statistiker Jonathan Sterne und George Davey Smith haben schon vor 15 Jahren im «British Medical Journal» dazu aufgerufen, die Resultate von medizinischen Studien nicht mehr als «signifikant» oder «nichtsignifikant» darzustellen, sondern im Kontext der gesamten Untersuchung und anhand anderer Ergebnisse zu interpretieren. Genützt hat das herzlich wenig, wie das Team von John Ioannidis in einer soeben erschienenen Studie festgestellt hat. Demnach sind in den vergangenen 25 Jahren in der biomedizinischen Forschung immer mehr Studien erschienen, die p-Werte angeben, die zudem immer klarer signifikant ausfallen. Gleichzeitig werden Zusatzinformationen zu den festgestellten Effekten immer seltener («JAMA», Bd. 315, S. 1141).

5% aufgerufen, die Resultate von medizinischen Studien nicht mehr als «signifikant» oder «nichtsignifikant» darzustellen, sondern im Kontext der gesamten Untersuchung und anhand anderer Ergebnisse zu interpretieren. Genützt hat das herzlich wenig, wie das Team von John Ioannidis in einer soeben erschienenen Studie festgestellt hat. Demnach sind in den vergangenen 25 Jahren in der biomedizinischen Forschung immer mehr Studien erschienen, die p-Werte angeben, die zudem immer klarer signifikant ausfallen. Gleichzeitig werden Zusatzinformationen zu den festgestellten Effekten immer seltener («JAMA», Bd. 315, S. 1141).

Es gibt Alternativen

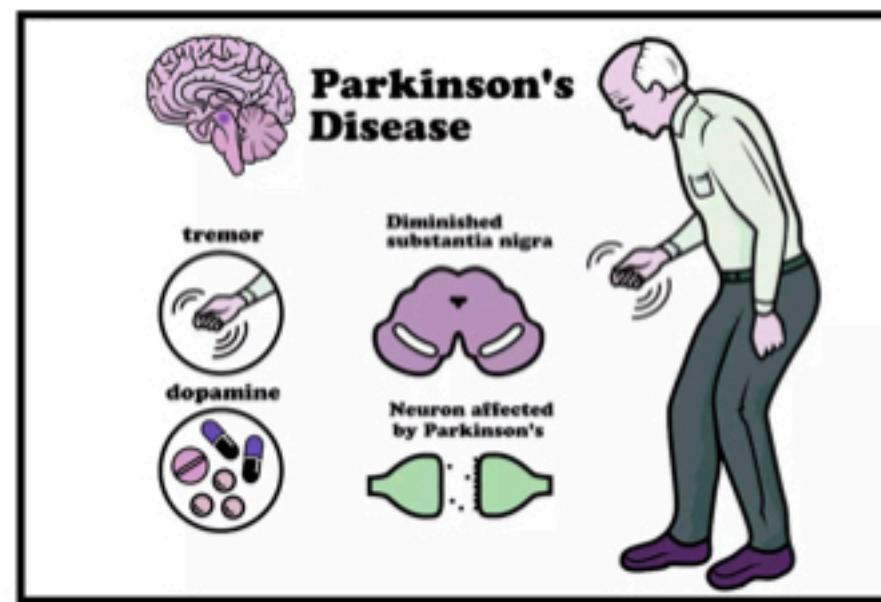
Das Problem ist dabei nicht nur, dass der p-Wert ein trügerisch einfaches statistisches Instrument ist. «Studien führen heute zu dermassen vielen Daten, dass man allen Unfug testen kann und so zu Hunderten von p-Werten kommt», erklärt Leonhard Held. «Der eine oder andere fällt dann bestimmt signifikant aus, auch wenn kein Effekt vorhanden ist.» Leonhard Held und Peter Jüni verlassen sich deshalb bei der Planung und Auswertung von Studien schon lange nicht mehr nur auf die p-Werte.

Jüni empfiehlt, die Resultate von Studien mindestens mit Vertrauensintervallen zu versehen, die spezifische Aussagen über die Unsicherheit einer Schätzung machen. Und Held erforscht alternative Massen für statistische Evidenz – zum Beispiel Bayes-Faktoren. Mit ihrer Hilfe lässt sich die Wahrscheinlichkeit einer Hypothese anhand der Daten anpassen, statt dass diese wie beim p-Wert nach einem Schwarz-Weiss-Schema angenommen oder abgelehnt wird.

CATTY MAES

Daten werden meist von Leuten analysiert, die nicht dafür ausgebildet sind.

Question you will work on



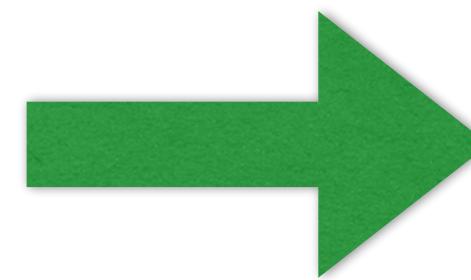
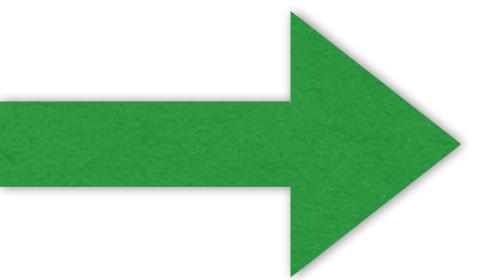
Producing nonsense with statistics... .

... 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. See for example:



Question
Puzzle
Problem

Data
+
Analysis

Answer
Solution



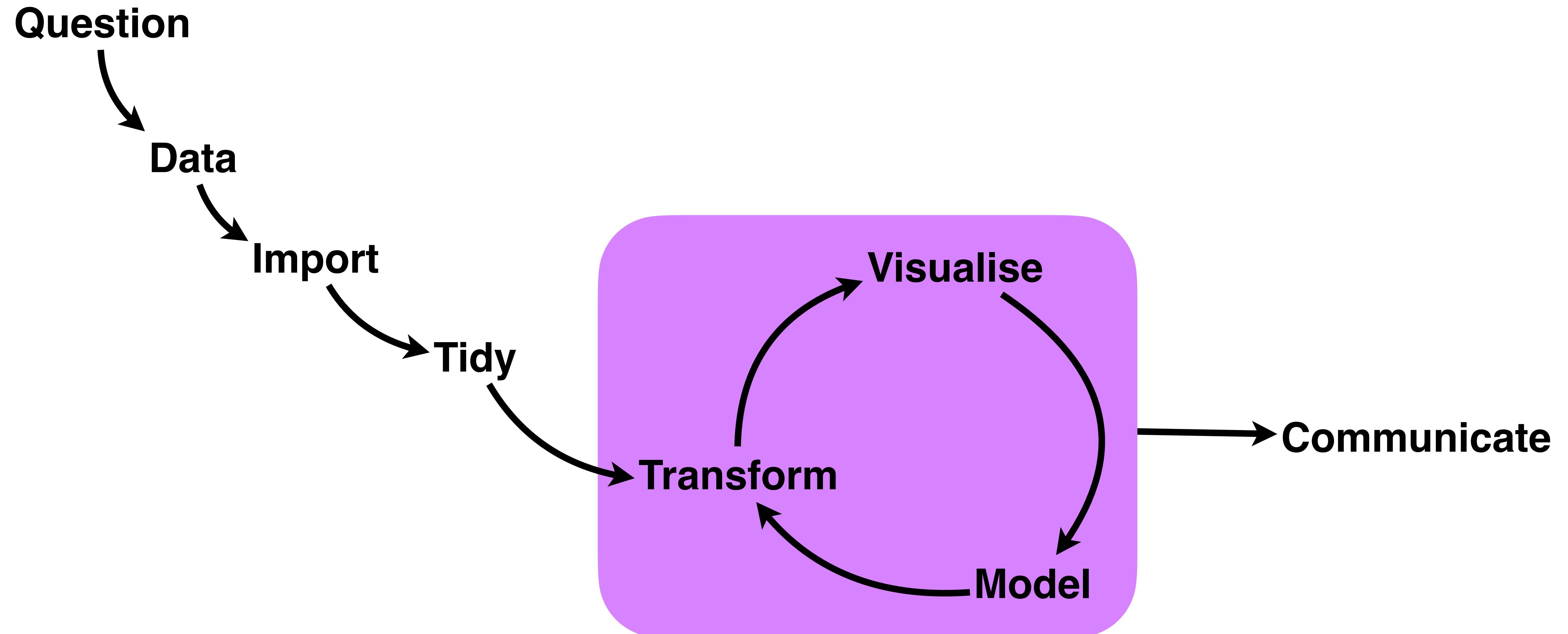
How do we get insights from data... .

... 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).



Efficient
Consistent
Repeatable
Reliable
Readable
Robust
Persistent
Sharable
Scalable

—
—
—





"Failure is an opportunity to grow"

GROWTH MINDSET

"I can learn to do anything I want"

"Challenges help me to grow"

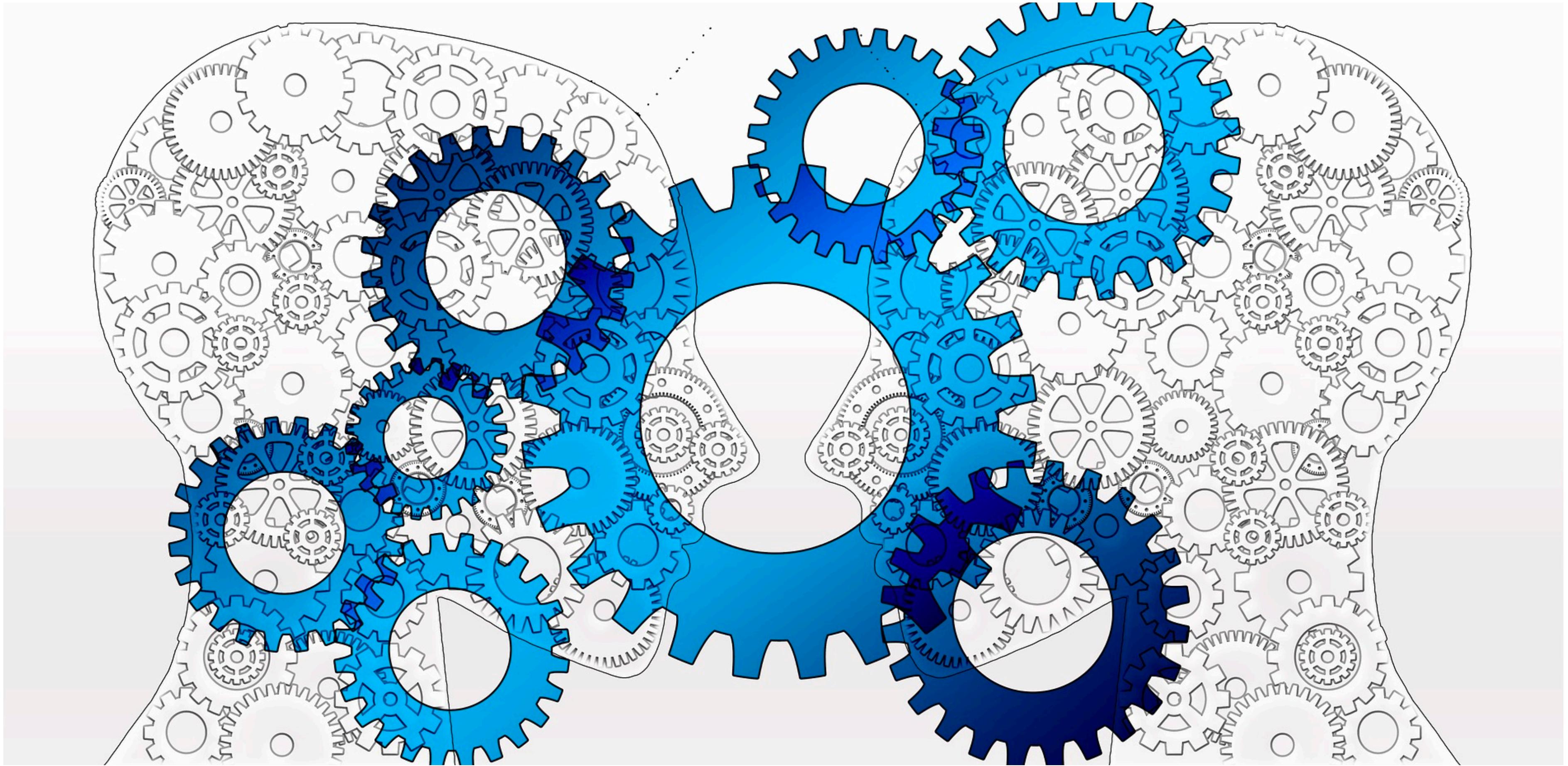
"My effort and attitude determine my abilities"

"Feedback is constructive"

"I am inspired by the success of others"

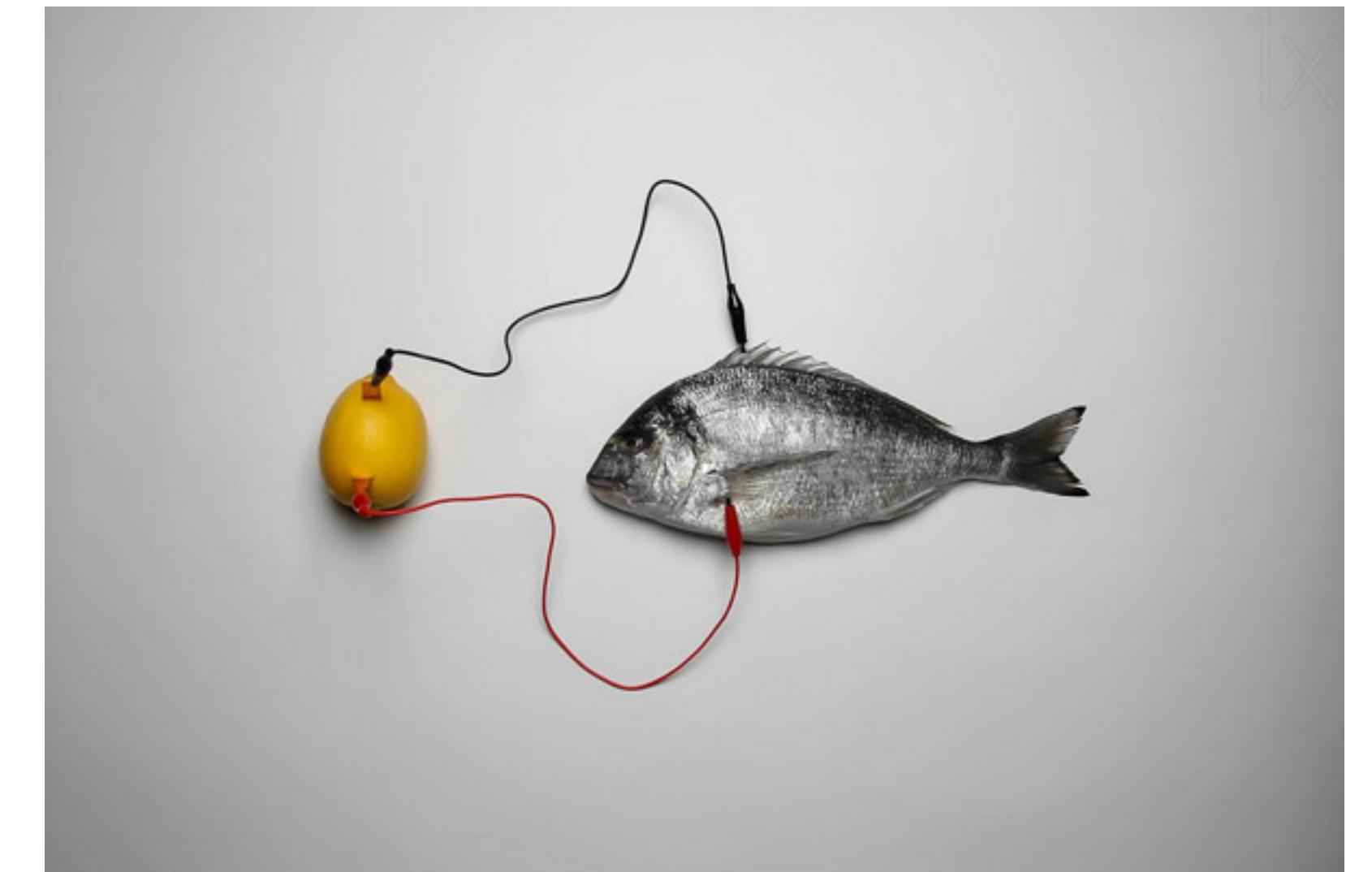
"I like to try new things"

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



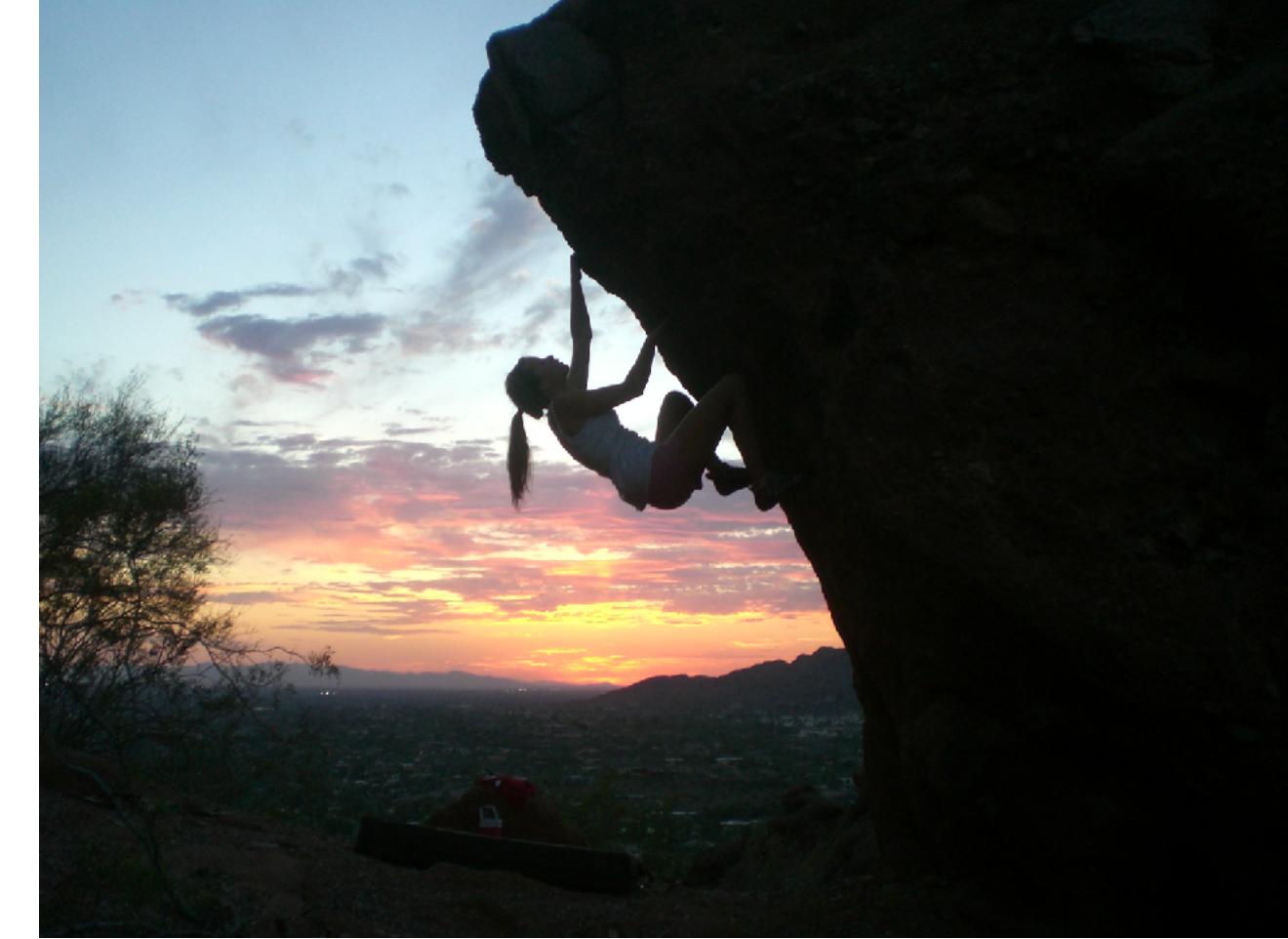
How the course is organised

The conceptual side



1 - 2:45pm Mondays

The practical side



1 - 3pm Thurs. or Fri.

Schedule (12 lecture units + 2 self-study weeks)

Unit 1 Introduction and outlook

Unit 2 No lecture

Unit 3 Simple linear regression

Unit 4 Residual analysis, model validation

Unit 5 Multiple linear regression

Unit 6 ANOVA

Unit 7 ANCOVA; Matrix Algebra

Unit 8 Model selection

Unit 9 Interpretation of results, causality

Unit 10 Count data (Poisson regression)

Unit 11 Binary Data (logistic regression)

Unit 12 Measurement error, random effects, selected topics

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, Unit 07, and Attendance. 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

Online learning until further notice

Lectures:

- by ZOOM, recorded, and available to watch later
(be patient with me for making them available).
- Mics off, *cameras on is better.*
Speak up if you have a question.
(I likely won't see a raised hand or ZOOM chat comment).
- Google doc is another way for comments,
questions, jokes.
(show link and doc)

This can be very monotonous:

- reading
- watching videos
- doing online exercises

It would be really nice
for us to see each other
(cameras on would be
very nice!)

And likely makes for
better learning:

such as the value of nonverbal cues in communication (Miller, [1988](#)), improved instructor effectiveness (Mottet, [2000](#)), and building instructor–student and student–student relationships (Falloon, [2011](#); Mottet, [2000](#)).

Online learning until further notice

Practicals:

- You call in to us to ask questions.
- I and a team of TAs will be waiting in online rooms.
- Online rooms will be listed in a google doc (show link and doc)
- You will very likely need to share your screen, perhaps even give control.

“I really appreciate the effort of making online video calls available, but in this case I didn't really know what exactly to ask.”
- A student, FS 2020

What worked well last year:

- Some students called in together, pairs, threes, fours (they organised this).
- Some students compiled a list of questions to ask.

Remember:

- We love to talk with you!
- For us, you are the main thing that makes the course interesting!!!
- You can call in to talk about anything... you may have a specific question, but “I just don't get it” and “my head feels like its gonna explode” are also great ways to start a conversation.
- **TAs have been trained in how to help.**



Online learning until further notice

This can be rather monotonous

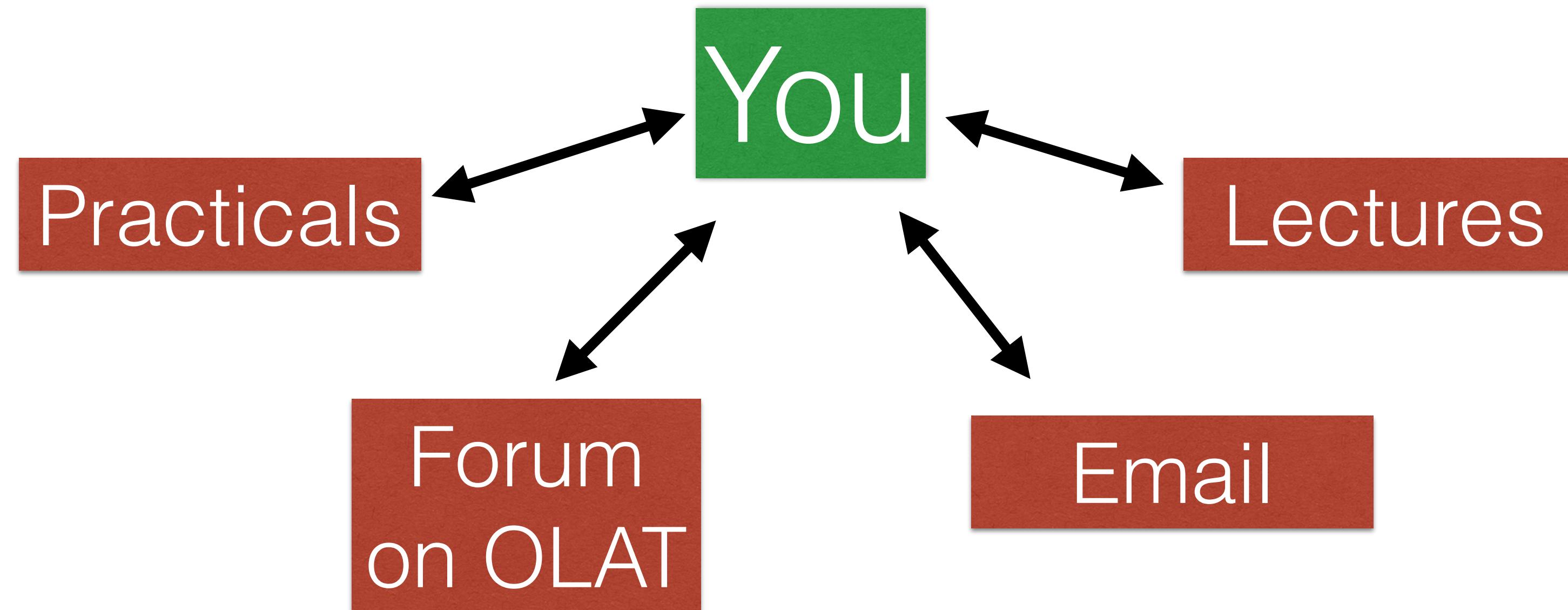
reading
watching videos
doing online exercises

Low diversity in your environment.

We will return to on-site, in-person, when we can.

*We strongly advise you to form working groups and do practicals together. (And call in to TAs together also.)
(And in lectures even—side comms.)*

Communications and the course team



Dr Hanja Brandl
Lecturer



Prof. Damien Farine
Lecturer



Prof. Owen Petchey
Director
Instructor



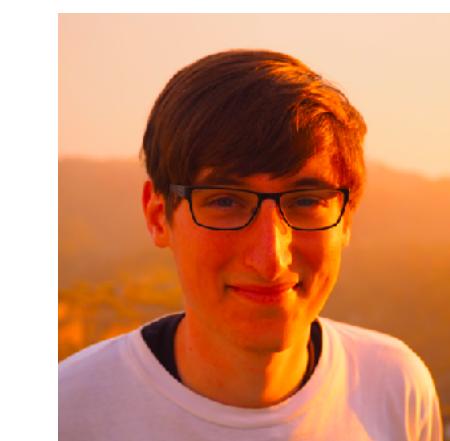
Uriah Daugaard
Practicals
Teaching Assistants
Statistics consultant



Dr Frank Pennekamp
Assessment
Examinations



Dr Rainer Krug
Computing consultant



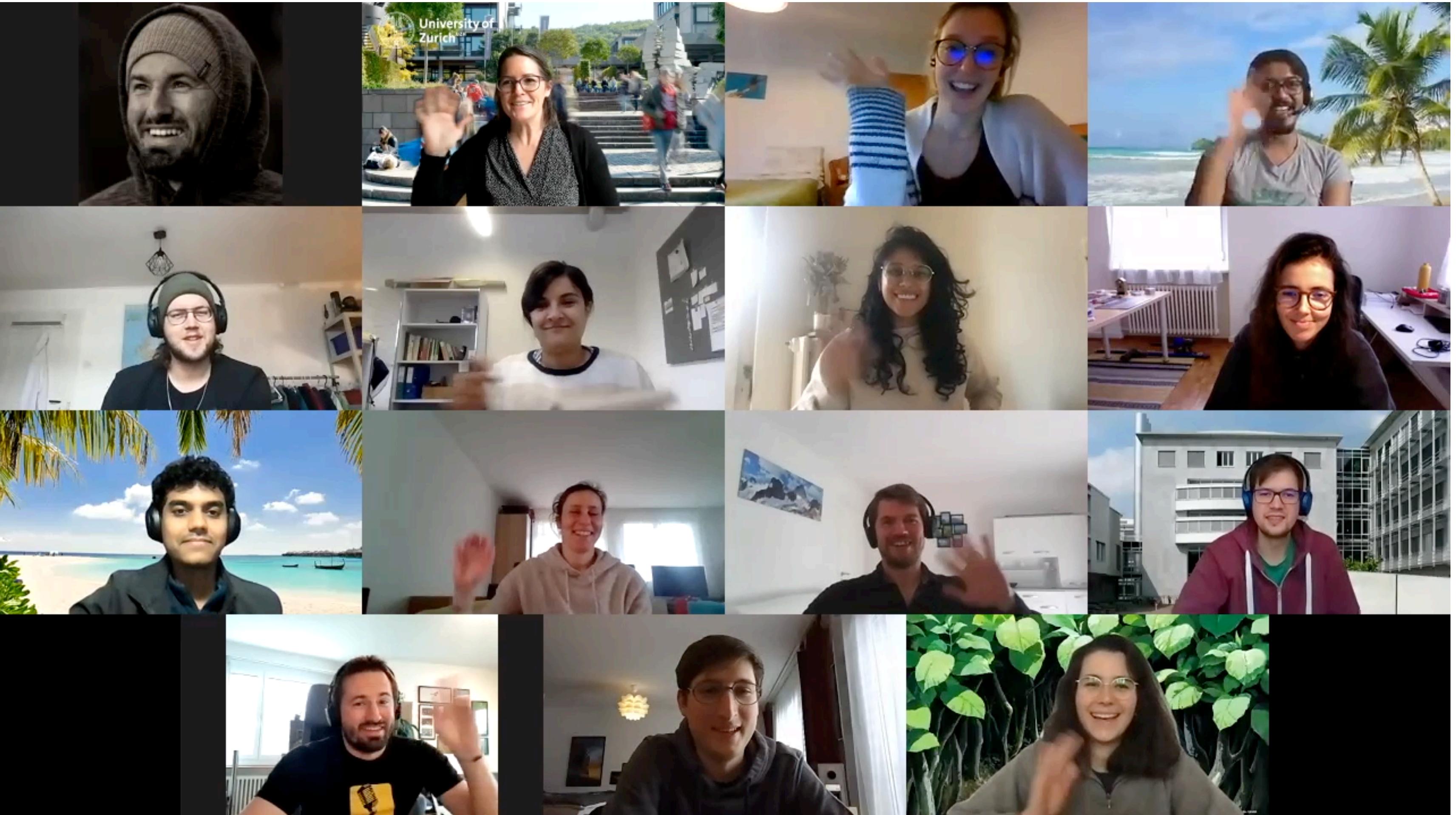
David Hofman



Ewa Merz

Numerous
teaching
assistants

Numerous teaching assistants



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



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 for first time in 2021.

**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 in 2021: Prof. Damien Farine, Dr Hanja

You

Us





The screenshot shows the OLAT course interface for '22FS BIO144 Data Analysis in Biology (Preparation)'. The top navigation bar includes links for Courses, Groups, Authoring, Campus courses, Question bank, and the current course title. On the right, there are icons for RSS feed, progress (0/181), print, search, email, and user profile.

The main content area features a sidebar on the left with sections for 'BIO144' (About the course, Previous knowledge, Wiki: FAQ), 'Forum' (Unit 01 to Unit 05), and a 'Course info' section with links for Course info, Calendar, Participant list, Participant infos, E-Mail, Blog, Wiki, Forum, Documents, Glossary, User role, and My course. Below the sidebar is a 'Course chat' and a 'Course search' bar.

The main content area displays an 'Overview of topics' section with a 'Search' bar. It shows a list of 3 entries in the discussion forum:

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

A large black arrow points upwards from the bottom of the page towards the 'Discussion topics' table.

Ask questions / make requests here.

Short break

Live data analysis demonstration

BIO144
Week 1

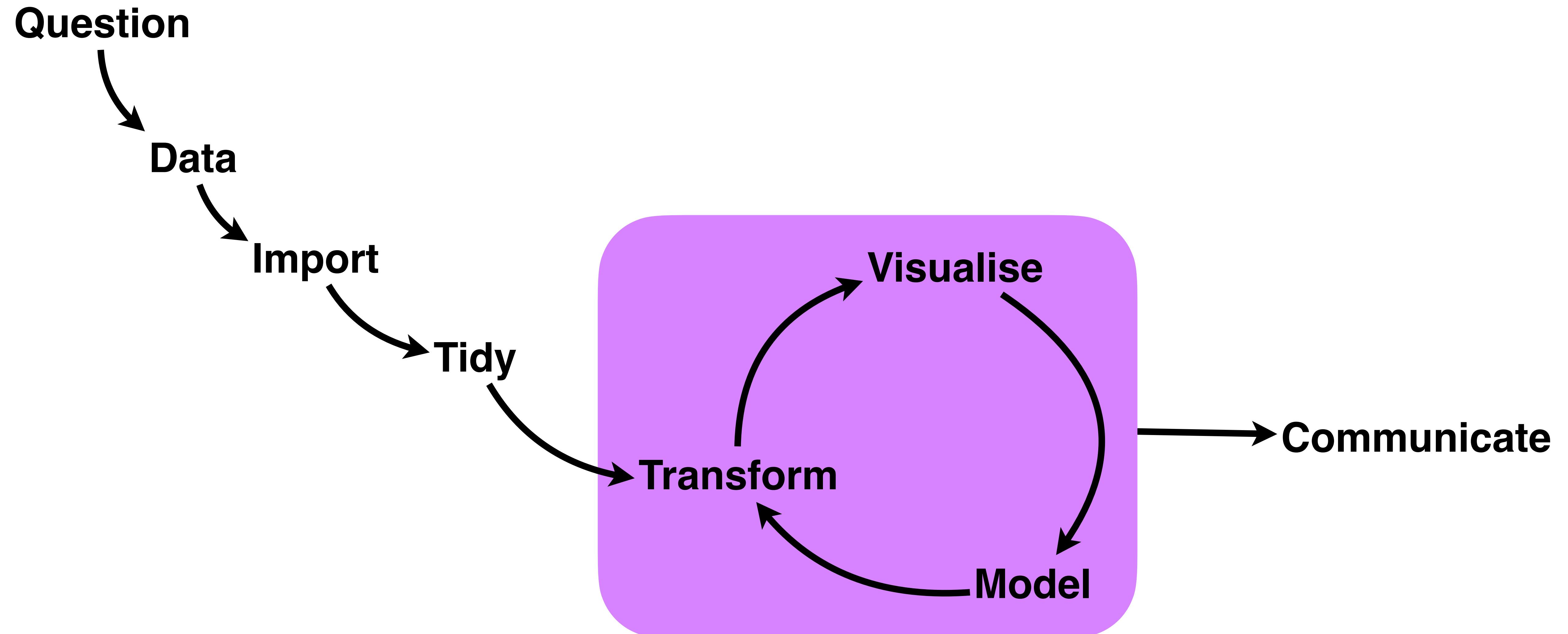
Its a demonstration...

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

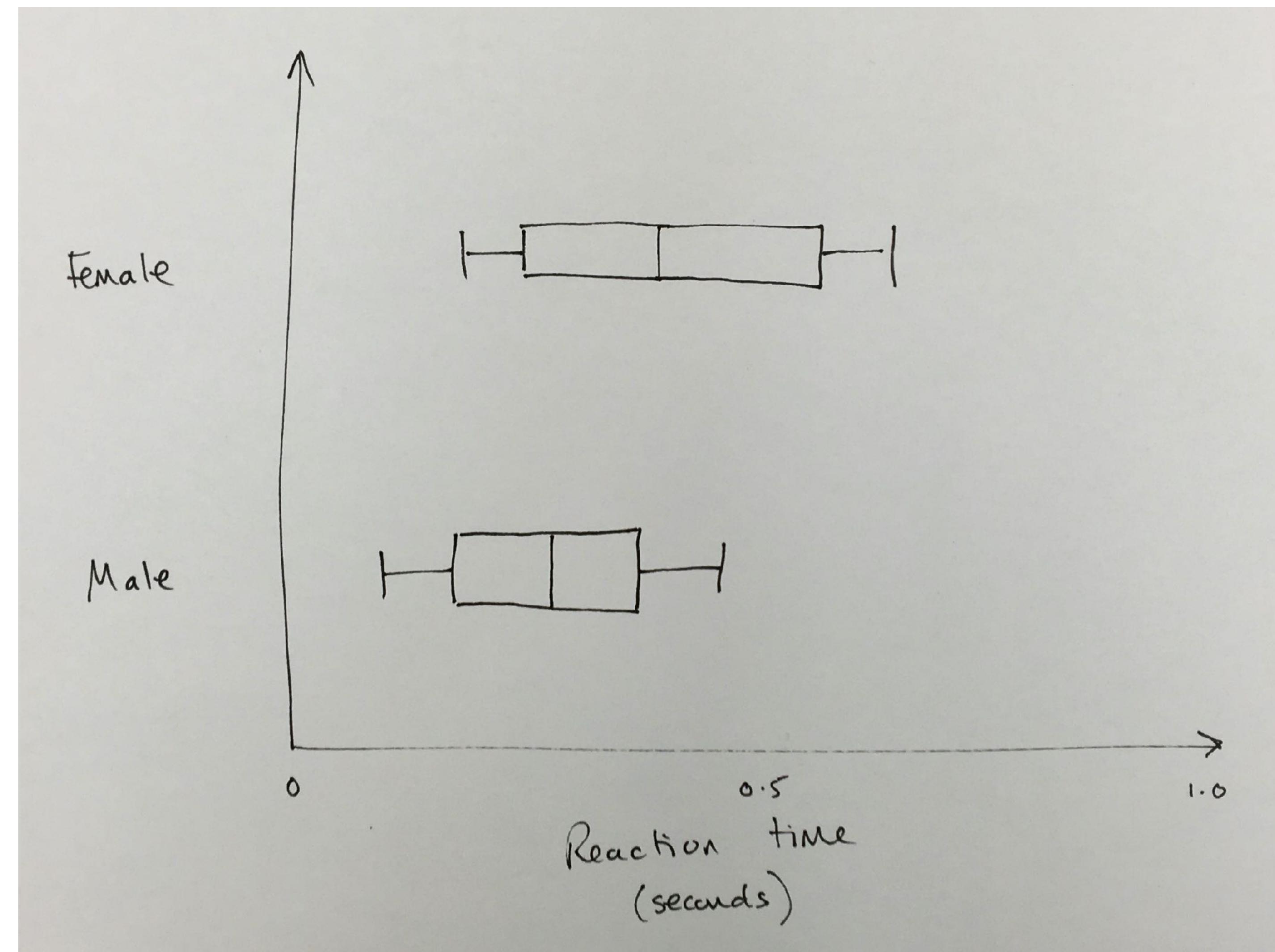
Data wrangling

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

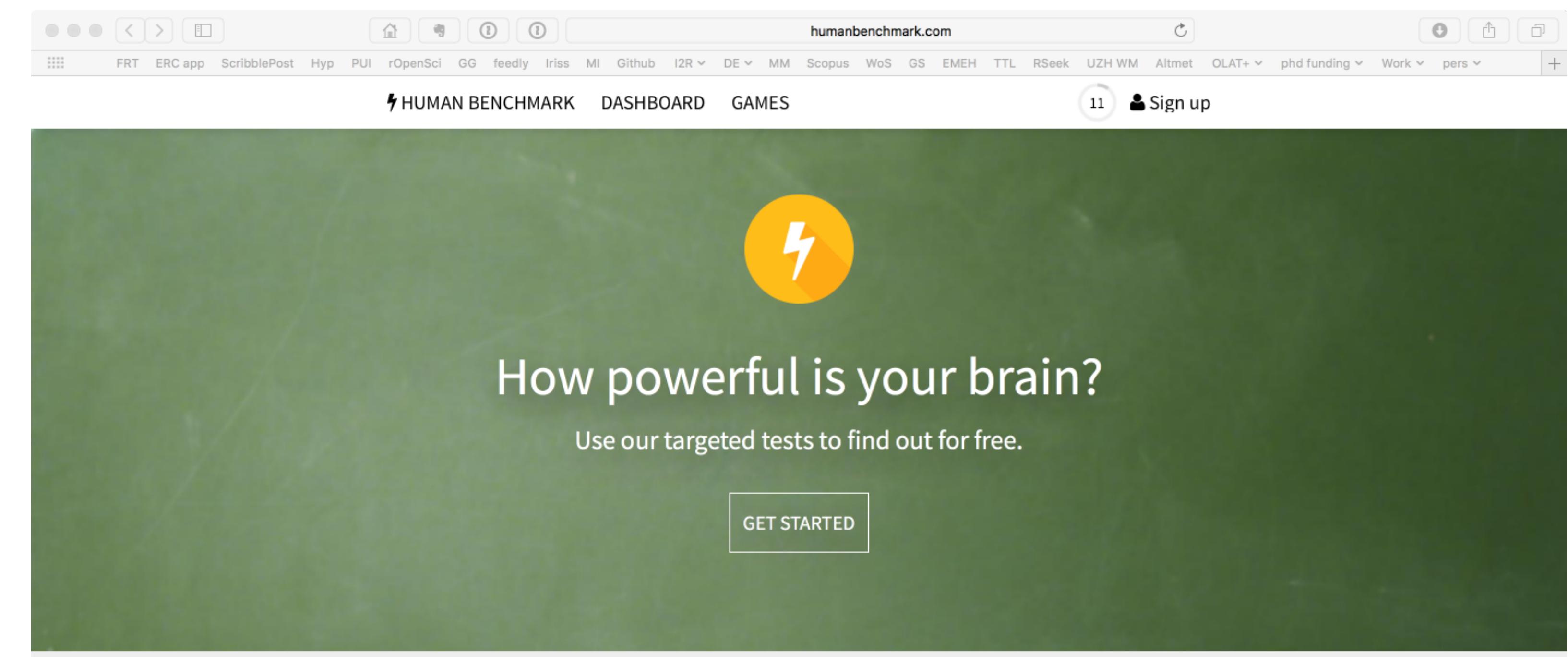
Visualise

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

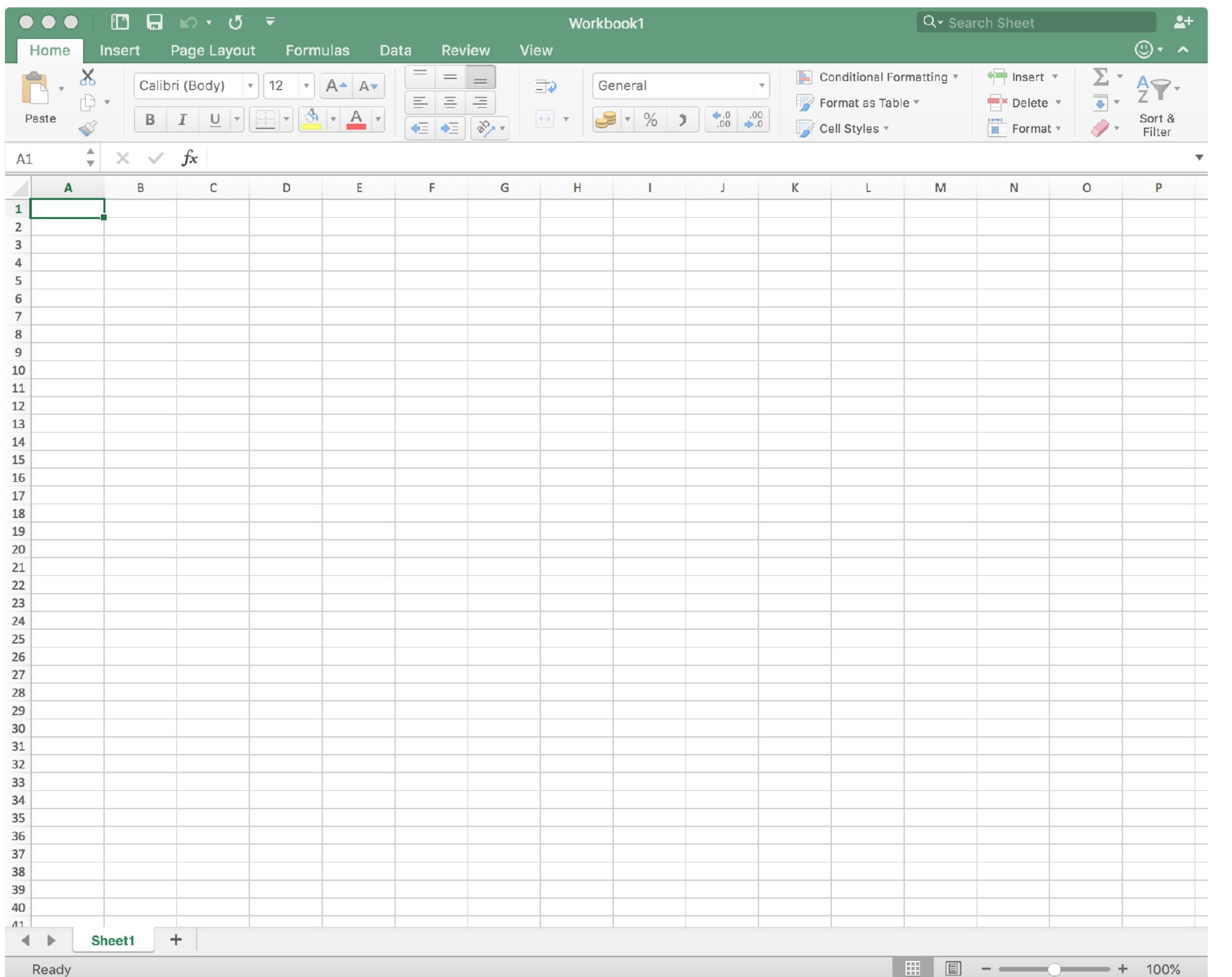
- Data wrangling
- Visualise
- Statistical test
- Critical thinking
- Report / communicate

Check the data in the spreadsheet

Efficient
Consistent
Repeatable
Reliable
Readable
Robust
Persistent
Sharable
Scalable

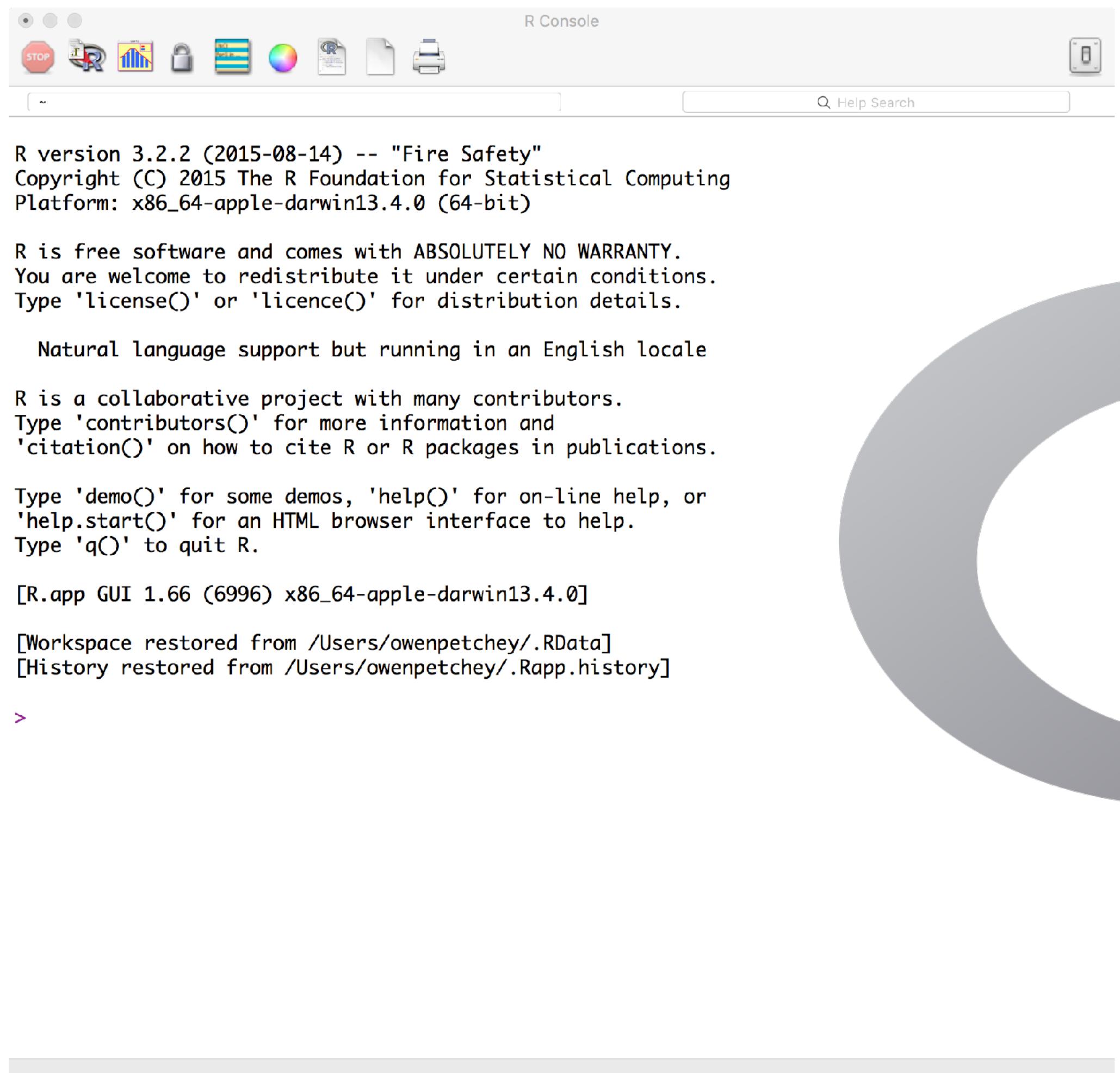
—
—





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

