

Data Analysis in Biology

BIO144
FS 2021



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

Statistical Analysis

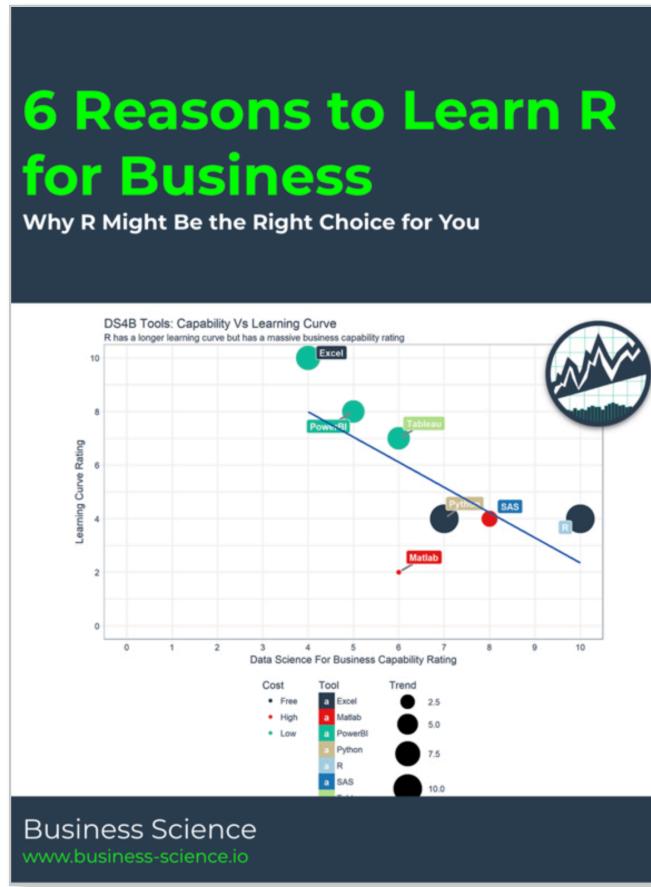
A video player interface with a dark grey background. At the top, there is a light grey header containing the title text. Below the header is a dark grey control bar. On the left side of the control bar is a red progress bar with a white slider. To the right of the slider are three white icons: a play button, a forward button, and a volume button. Next to the volume button is the text "0:06 / 3:00". In the center of the control bar is the text "Source: LinkedIn". On the far right of the control bar are five small white icons: a document, a gear, a square, a screen with a play button, and a double arrow.



WHY BUSINESS SCIENCE? COURSES BLOG CONTACT LOG IN ENROLL

6 Reasons To Learn R For Business [2021]

Written by Matt Dancho on December 17, 2020



Get Articles in Your Inbox

email

Get Articles

Search for Articles

Search

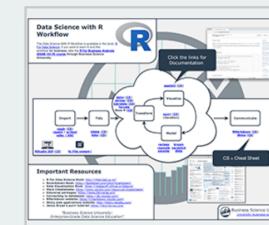


Find Articles By Category



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

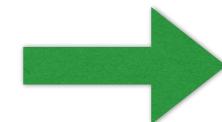
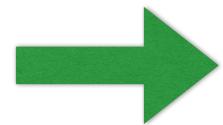
Our hundreds of partners from governments, the private sector, and civil society organizations are joining forces to take action, galvanize political commitment, build trust, and spur innovation in the booming data ecosystems of the 21st century.

[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.

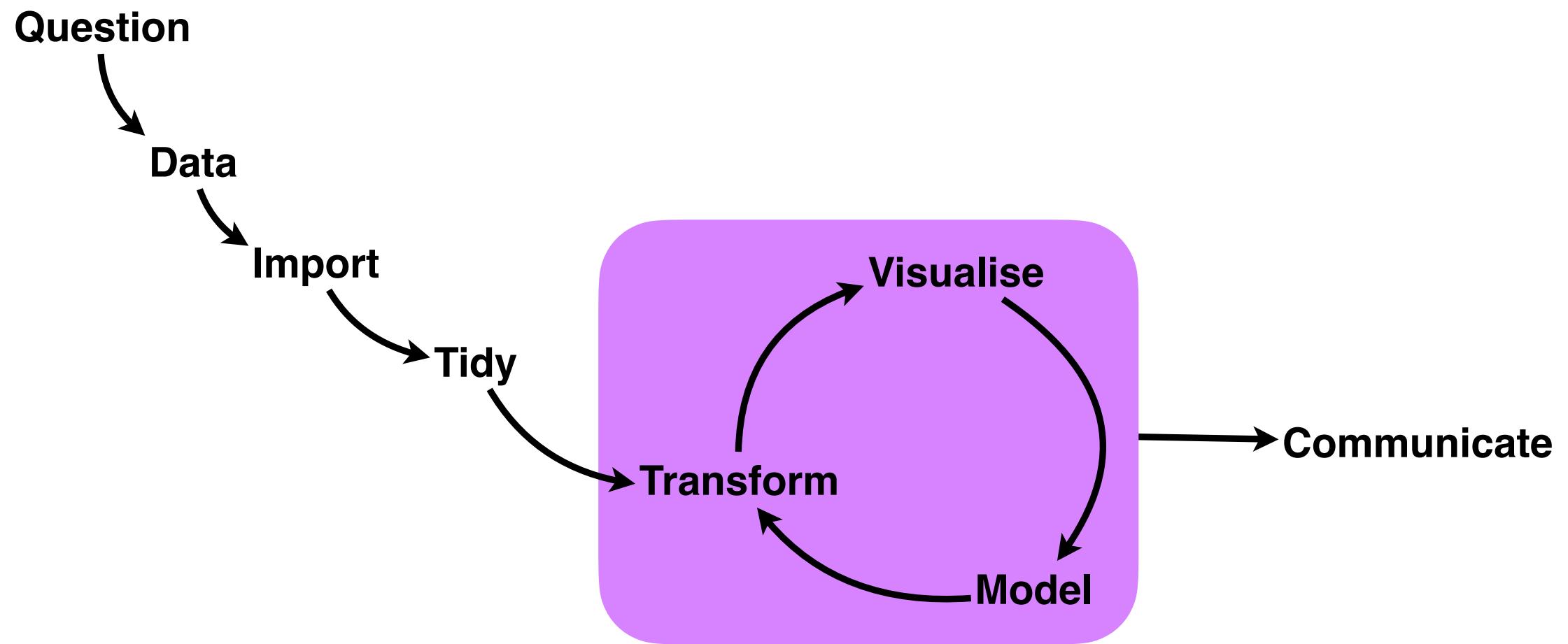
[LEARN MORE](#)



Question
Puzzle
Problem

Data
+
Analysis

Answer
Solution



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"

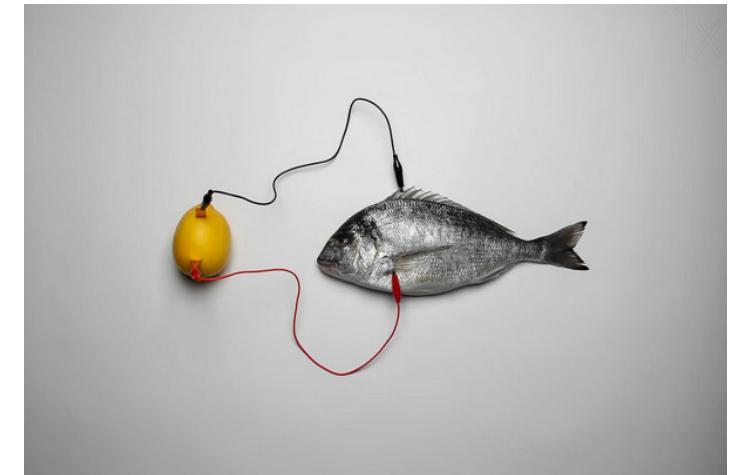
"Feedback is constructive"

"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.

The conceptual side



1 - 2:45pm Mondays

The practical side



1 - 3pm Thurs. or Fri.

Details on openedX

The screenshot shows the openedX platform interface. At the top, there are tabs for Geographie, Chemie, Mathematik, and Biologie. Below these tabs, there is a banner for "WELCOME TO OPENEDX @ UZH!" featuring the University of Zurich logo. The main content area displays several course cards:

- UZH BIO144 DATA ANALYSIS IN BIOLOGY** (Starts: Jul 27, 2016)
- UZH BIO134 PROGRAMMING IN BIOLOGY** (Starts: Sep 01, 2016)
- UZH MULTI_MODULE INTRODUCTION TO R** (Starts: Sep 11, 2016)
- UZH AST241 INTRODUCTION TO ASTROPHYSICS** (Starts: Sep 12, 2016)

At the bottom, there are two additional course cards:

- UZH** (with a grid diagram question)
- UZH** (with a cell diagram question)

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

<https://edu-exchange.uzh.ch/>

Owen sent a welcome email

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.

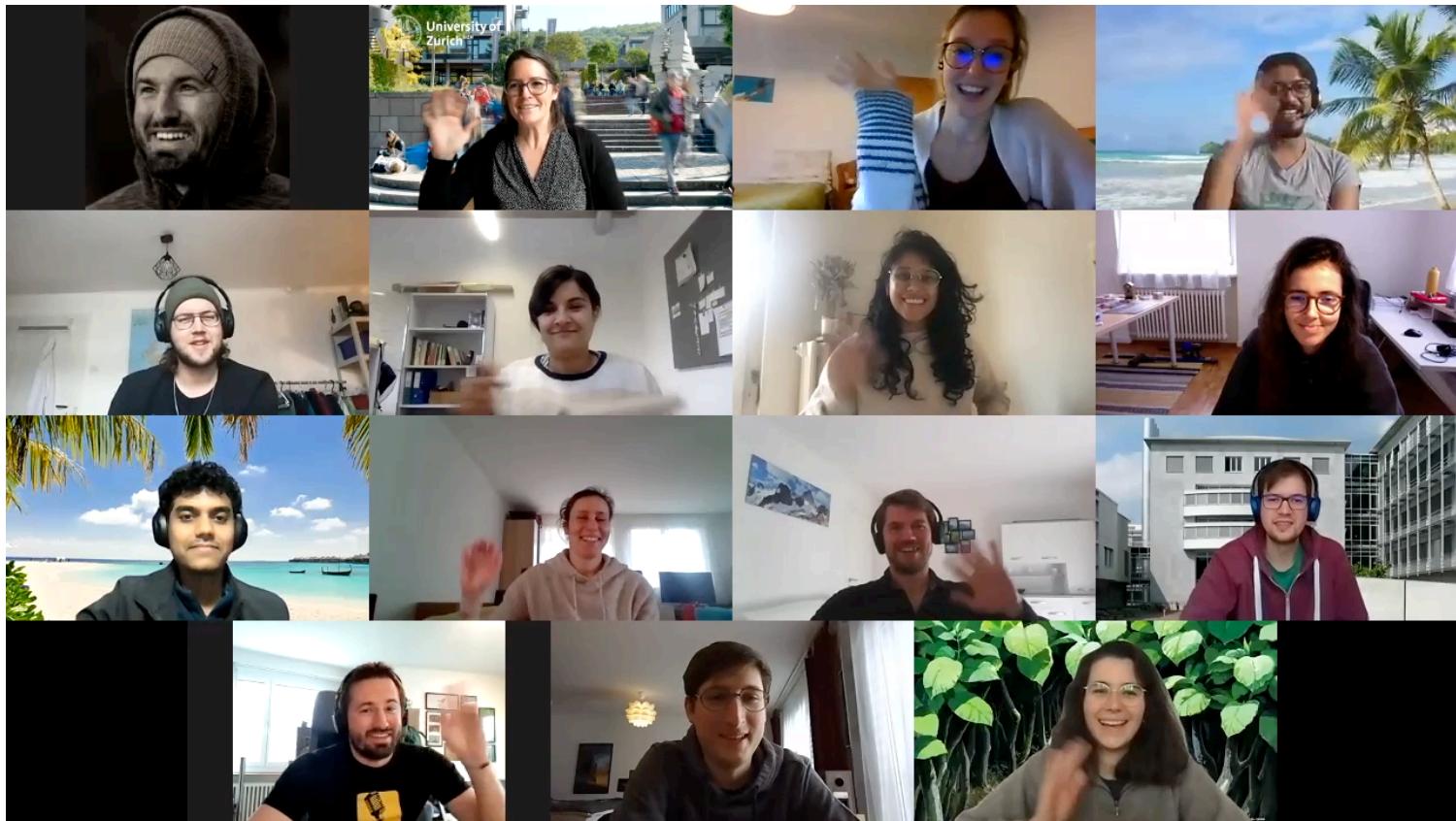
We will return to on-site, in-person, when we can.

*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.)*

The team...



Numerous teaching assistants



Graded assessment questions



Hyperlink

Some “play” questions



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 image shows a background graphic of multiple hands of different skin tones raised, set against a light blue background with several large, semi-transparent question marks in various colors (red, orange, yellow, green, blue) floating around.

University of Zurich Zurich, Switzerland UZH: BIO144 Data Analysis in Biology Discover New

Owen Petchey ▾

View this course as: Staff

Course Discussion Progress Instructor

Course > About the course > The basics > Important links

◀ Previous ▶ Next ▶

Important links

Bookmarked

The google doc for making comments and asking questions during lectures.

The google doc for finding a TA to talk to during a practical.

The lecture podcasts on OLAT (there is nothing else about BIO144 on OLAT).

Example solution scripts.

◀ Previous ▶ Next ▶

STAFF DEBUG INFO

A black arrow points from the text "Ask questions / make requests here." at the bottom of the slide up towards the "Important links" section on the right side of the slide.

Ask questions / make requests here.

Short break and then switch to other presentation

Lecture 1: Introduction and Outlook

BIO144 Data Analysis in Biology

Stephanie Muff & Owen Petchey

University of Zurich

31 January, 2021

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

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?

Data and statistics are essential

A good foundation in statistics **makes you more independent** from consultants or the goodwill of colleagues. Without such a knowledge, you will have to heavily rely on and trust others.

Data analysis/statistics is itself an exciting part of research!

Data analysis is at the **interface between mathematics and biology/medicine** (and many other applied research fields).

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



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, Spanjaardskanaal 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

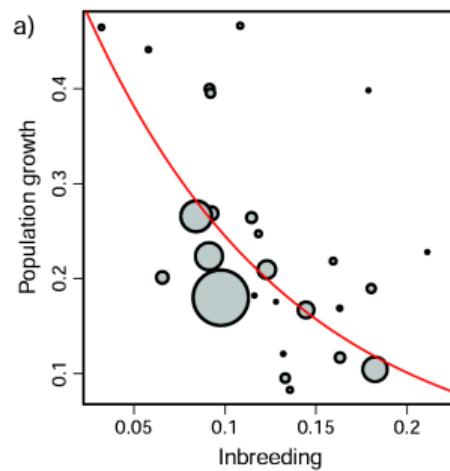
^f Swiss Federal Research Institute WSL, Biodiversity and Conservation Biology, 8903 Birmensdorf, Switzerland

<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)



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. Böss verstehen viele Forscher die Bedeutung dieser Berechnungen gar nicht. Von Patrick Imhasly

Kritische Stimmen können nicht nur Menschen, sondern auch statistische Größen. Sie gilt besonders für den sogenannten *p-Wert*, der von Medizinerinnen und -männern und jede Studentin in Kostüm kommt, wenn sie über ihre Arbeit sprechen. Er ist statistisch an sich gut, insbesondere wenn er mit einer Wissenschaft verbunden ist. Inzwischen ist der *p-Wert* indes auf die schiefen Bahnen geraten. Dies ist nicht ohne Grund, wie der britische Geisteskundler Ronald Fisher, (1925) als eine Art informelles Szenenbild für Statistik bekannt ist. Er war es, der den *p-Wert* in der Praxis erfasst zu einem simplen Lackmuster verkennete.

Bei einer statistischen Analyse von Daten entwirkt ein *p-Wert* < 0,05 (5 Prozent) oder noch besser < 0,01 (1 Prozent), gelten diese als signifikant. Das ist eine so genannte *statistisch Beweislast* zugewandt. Das entscheidet etwa darüber, ob ein neues Medikament wirklich besser ist als ein bestehendes. Ein *p-Wert* sollte seine Stärke in etwas angemessene Fachbücher prahlend anbauen. Aber es kann auch nur ein Zufallsblitz sein, der durch den Zufall ausser Kraft zu setzen, hat sich Rüdiger Wiesner, der Direktor des Allgemeinen Instituts für Statistik (AIS) in Stuttgart, jüngst ehrlich beklagt.

Wissenschaftler verwenden den *p-Wert* unterschiedlich. Einmal, um zu beschreiben, das kostet schlechte Forschung und umgekehrt die Glanzleistung der Wissenschaft. Der Mediziner und Statistiker John Ioannidis, der an der Universität Stanford sprach in einem Konzertsaal von Design abhängig, „der falsche Beleg des *p-Wertes* ist derzeit ein riesiges Problem, so ausdrückt, dass manche wichtig

werden darum – von allen wegen der oft Falschgedanken von Publikationen belohnt werden.“ Angeregt der Massenstaat sich sich die ASA jetzt verabschiedet, zum ersten Mal in ihrer Geschichte, eine Richtlinie für die Beurteilungen zu veröffentlichten, wie man mit einer statistischen Grösse vorsichtig vorgeht.

Wissen ist Macht ist ein aus, was man gewissenvon den erwartet, erklärt der Berner Epidemiologe Peter Flati, der seit kurzem die Leitung des Instituts für Statistik der Universität Toronto tingt ist. Das bedeutet: Der *p-Wert* nimmt die Wahrscheinlichkeit, dass die Hypothese falsch ist, in Betracht, und auch nicht, ob ein bestmögliches Resultat zufällig zustande gekommen ist. Der kontinuierliche Wert kann also sehr unterschiedlich sein. „Um einen signifikanten Unterschied oder Zusammenhang „Ob A-Wirkung“ oder „Zusammenhang „Ob A-Wirkung“ zu beweisen, erklärt Peter Flati, „Durch genau das wünschen viele Forscher und Journalisten.“

Flati konzentriert, dass die Signifikanzgrenze 5 Prozent kein 1 Prozent historisch gewesen ist. „Vor 100 Jahren war es 0,05.“ Der *p-Wert*, überlassen so jedoch als standardisch angenommen, ist nicht so einfach zu definieren, wie es scheint. Ronald Fisher, der Erfinder des *p-Wertes*, überlässt es jedoch als standardisch angenommen, dass die Signifikanzgrenze 0,05, als welcher Grenze ein *p-Wert* in einer Untersuchung Aussagekraft haben soll. „Für diesen kleinen Unterschied zwischen 0,05 und 0,01 gibt es keinen allgemeinen Konsens“, sagt Flati. „Klar ist, dass die Gefahr von Generalisierungen von Forschungsergebnissen, und Leuchtturm-Held vom Institut für Biostatistik und Bioinformatik der Universität Zürich (IBZ).“

Die britischen Statistiker Jonathen Sterne und George Davey Smith haben schon vor 15 Jahren im «British Medical Journal» den

„Daten werden reinigt, von Lauten analysiert, die nicht dafür ausgehoben sind.“

Immer statistisch festgestellt, oder noch gruseliger: Um eine kritisch postulierten beiden untersuchten Medikamenten ist – unter den Annahmen der Hypothese – die Wahrscheinlichkeit, der *p-Wert*, und je geringer er ist, desto weniger spricht für die Hypothese. Wenn der *p-Wert* sehr niedrig ist, dann ist die Hypothese höchstwahrscheinlich falsch. Wenn der *p-Wert* sehr hoch ist, dann gelten die Daten nicht für die Hypothese. Eine Studie ist als «signifikant» zu betrachten, wenn der *p-Wert* sehr niedrig ist. Doch die Grenze ist statistisch gesehen, tatsächlich geprägt. (siehe J.)

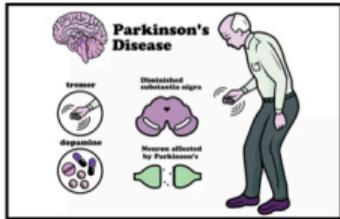
zugeben, die Relevanz von medizinischen Studien nicht mehr als eingeführtes oder reichweitenfakturierendes Werkzeug, sondern als anhand anderer Ergebnisse zu interpretierendes. Generell ist das Verhältnis zwischen den Studien, die in der Praxis in meist noch einschneidendem Strauß festgestellt sind. Denn nach dem britischen Statistiker und dem britischen Ökonomen David Hume ist nichts sicherer als ein Wissenschaftliches Forschung kennen neue Studien zu erschaffen, die *p*-Werte angeben, die in der Praxis nicht mehr zuverlässigen. Gleichzeitig werden Sensitivitätsuntersuchungen zu den festgestellten Werten immer wichtiger. (siehe J., S., T., H.)

Es gibt Alternativen

Das Problem ist dabei nicht nur, dass der *p-Wert* ein trügerisch einfaches statistisches Instrument ist, «einfaches Fehlern kann zu falschen Ergebnissen führen, was man allen Umfrageteilen kann und von ins Wunderland von *p*-Werten kommt», sondern auch, dass es in anderen Fällen durchaus bestreitbare signifikante auch wenn kein effektiv vorhandener Unterschied ist. Und das kann nur verhindern, dass sich bei der Planung und Auswertung von Studien schon lange nicht richtig vorgehen.

Für empirische, die Beobachtung von Studien zuestrebende mit Vertrauensintervallen zu verzieren, ist eine andere Methode, die Unschärfe einer Schätzung zu sehen. Und diese erfordert alternative Maße für statistische Sicherheit, zum Beispiel *Bayesianismus*. Mit ihm trifft man die Wahrscheinlichkeit einer Hypothese anhand der Daten abgessen, statt dass wie beim *p-Wert* nach einer Hypothese gefragt wird.

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:

How do we get insights from data... .

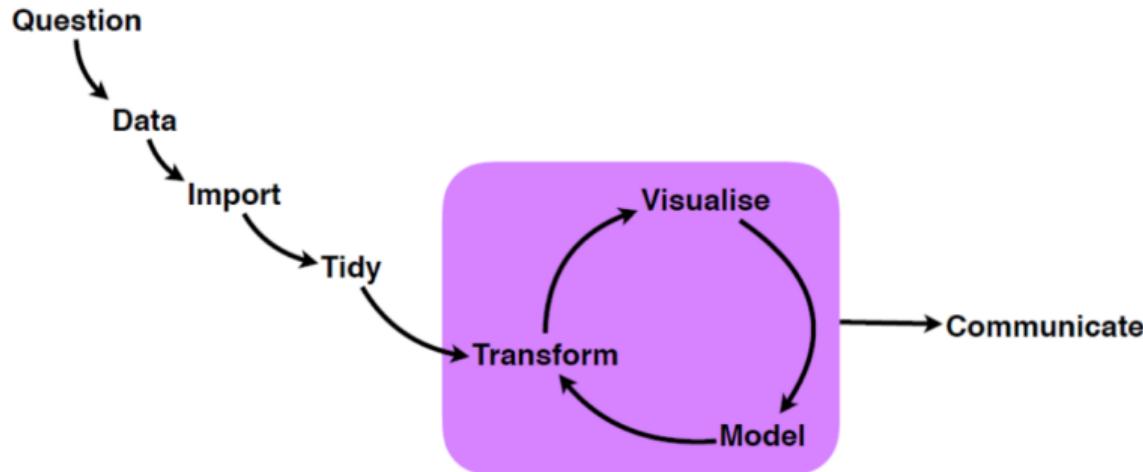
... rather than nonsense.

- ▶ 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).

Steps in a getting insights from data (“work flow”)



Visualising data

You should remember the following options for graphical data descriptions. Several of them appeared already in previous examples.

Representation	Useful for
Scatterplots	Pairwise dependency of continuous variables.
Histograms	Distribution of continuous variables.
Box and whisker plots	Distribution of continuous variables for different categories.

All can be augmented, for example by “conditioning” (e.g. colouring points according to the values of a variable).

A career in visualising data??!!

There are many “fancy” ways to graphically display data (**nice-to-know**):

- ▶ 3D-plots
- ▶ Spatial representations (using geodata)
- ▶ Interactive graphs and animations

Many R packages are available for various purposes. Interactive apps can, for example, be generated with Shiny. Check out the shiny gallery:

<http://shiny.rstudio.com/gallery/>

What is a (statistical) model?

A model is an approximation of the reality. **Understanding how the real world works** is usually only possible thanks to simplifying assumptions.

→ This is exactly **the purpose of statistical data analysis**.

In 2014, David Hand wrote:

In general, when building statistical models, we must not forget that the aim is to understand something about the real world. Or predict, choose an action, make a decision, summarize evidence, and so on, but always about the real world, not an abstract mathematical world: our models are not the reality – a point well made by George Box in his often-cited remark that "all models are wrong, but some are useful."

Goals of the course (part 2)

By the end of the course you will be able

- ▶ to **explore** and **analyze** data with appropriate methods, including statistical models,
- ▶ to **report** and **interpret** the results,
- ▶ to **draw conclusions** from them,
- ▶ to give **graphical descriptions** of the data and the results,
- ▶ to **be critical** about what you see.

Literature

Recommended literature (books available as ebooks from uzh):

1. *Lineare Regression* by W. Stahel (pdf on course webpage)
2. *Getting Started with R, An introduction for biologists* (**Second Edition**)
Beckerman, Childs & Petchey, Oxford University Press (DO NOT USE THE FIRST EDITION!).
3. *The New Statistics With R* by A. Hector, Oxford University Press;

→ See “Course texts/material” on course website.

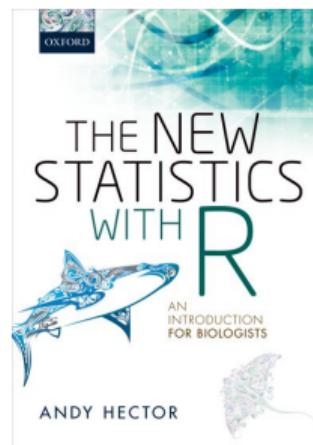


GETTING STARTED WITH **R**

An Introduction for Biologists

SECOND EDITION

ANDREW P. BECKERMAN,
DYLAN Z. CHILDS, AND
OWEN L. PETCHEY



THE NEW
STATISTICS
WITH **R**

AN
INTRODUCTION
FOR BIOLOGISTS

ANDY HECTOR

Live data analysis demonstration

BIO144
Week 1

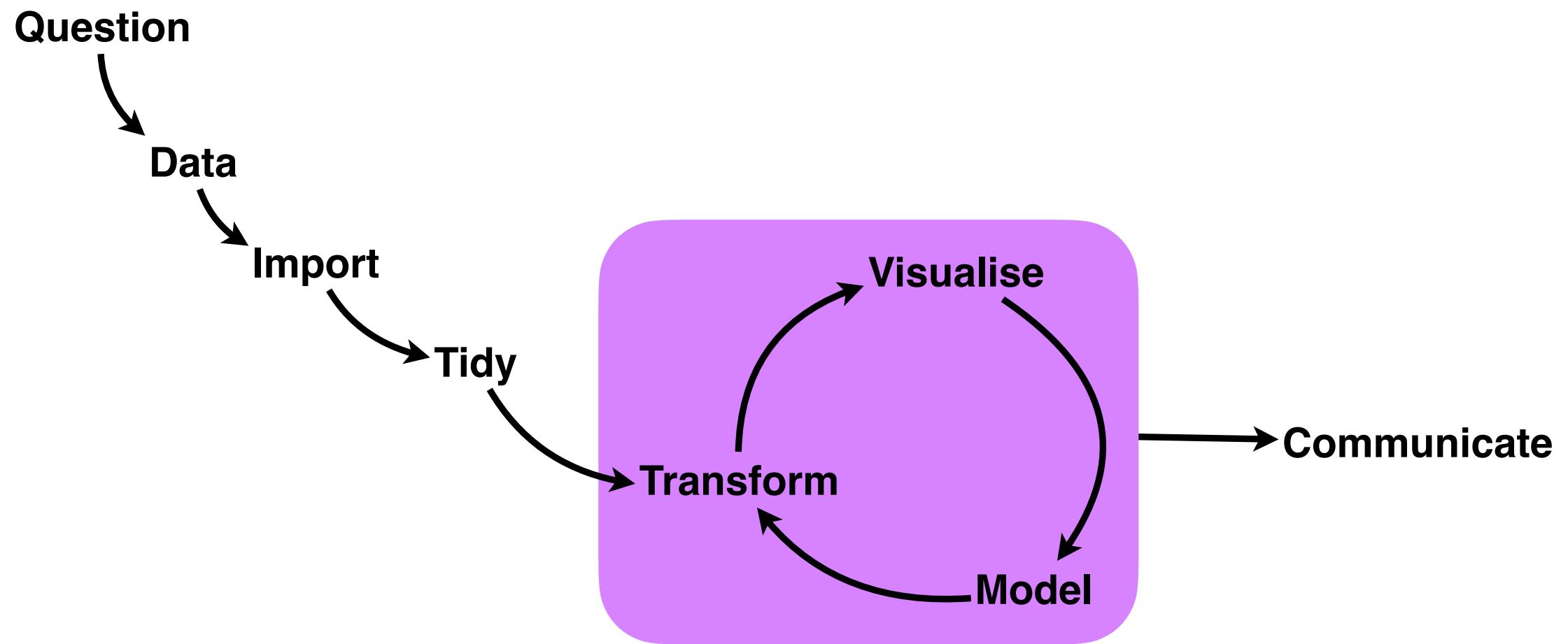
Its a demonstration...

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

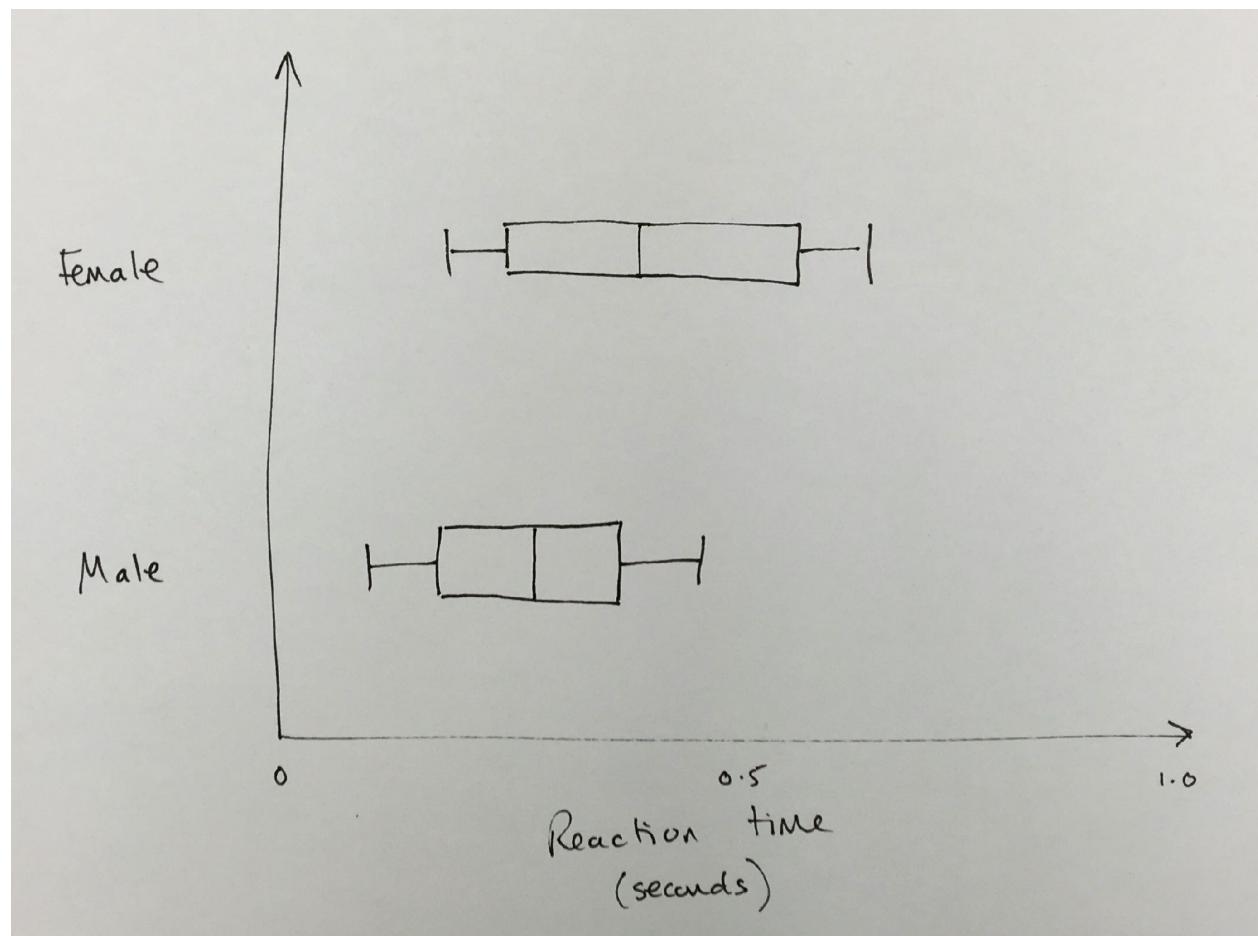
Data wrangling

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?

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 homepage of humanbenchmark.com. At the top, there is a navigation bar with various links such as 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 header with the text "HUMAN BENCHMARK DASHBOARD GAMES" and a "Sign up" button. The main content area features a large green background image of a brain. In the center, there is a yellow circle with a lightning bolt icon. Below it, the text "How powerful is your brain?" is displayed, followed by "Use our targeted tests to find out for free." and a "GET STARTED" button. A modal window is overlaid on the page, containing four circular icons with labels: "Number Memory" (red circle with "123"), "Reaction Time" (yellow circle with a lightning bolt), "Verbal Memory" (green circle with a book), and "Visual Memory" (blue circle with a grid). Each icon has a brief description below it. At the bottom of the page, there is a copyright notice: "Copyright 2007-2016 Human Benchmark contact@humanbenchmark.com 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://bit.ly/2Educm>

The screenshot shows a Google Forms survey titled "My Human Benchmark results". The survey is for a live data analysis demonstration at 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

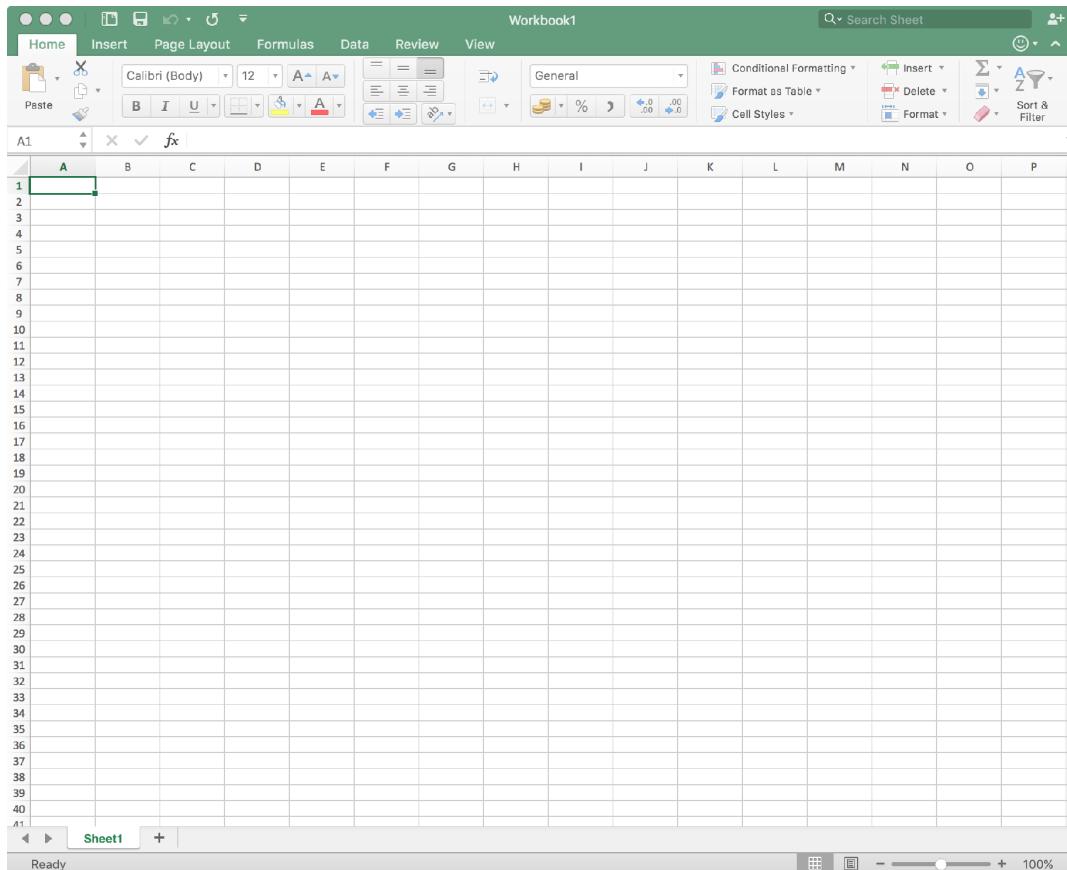
Check the data in the spreadsheet

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**



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]

>

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

