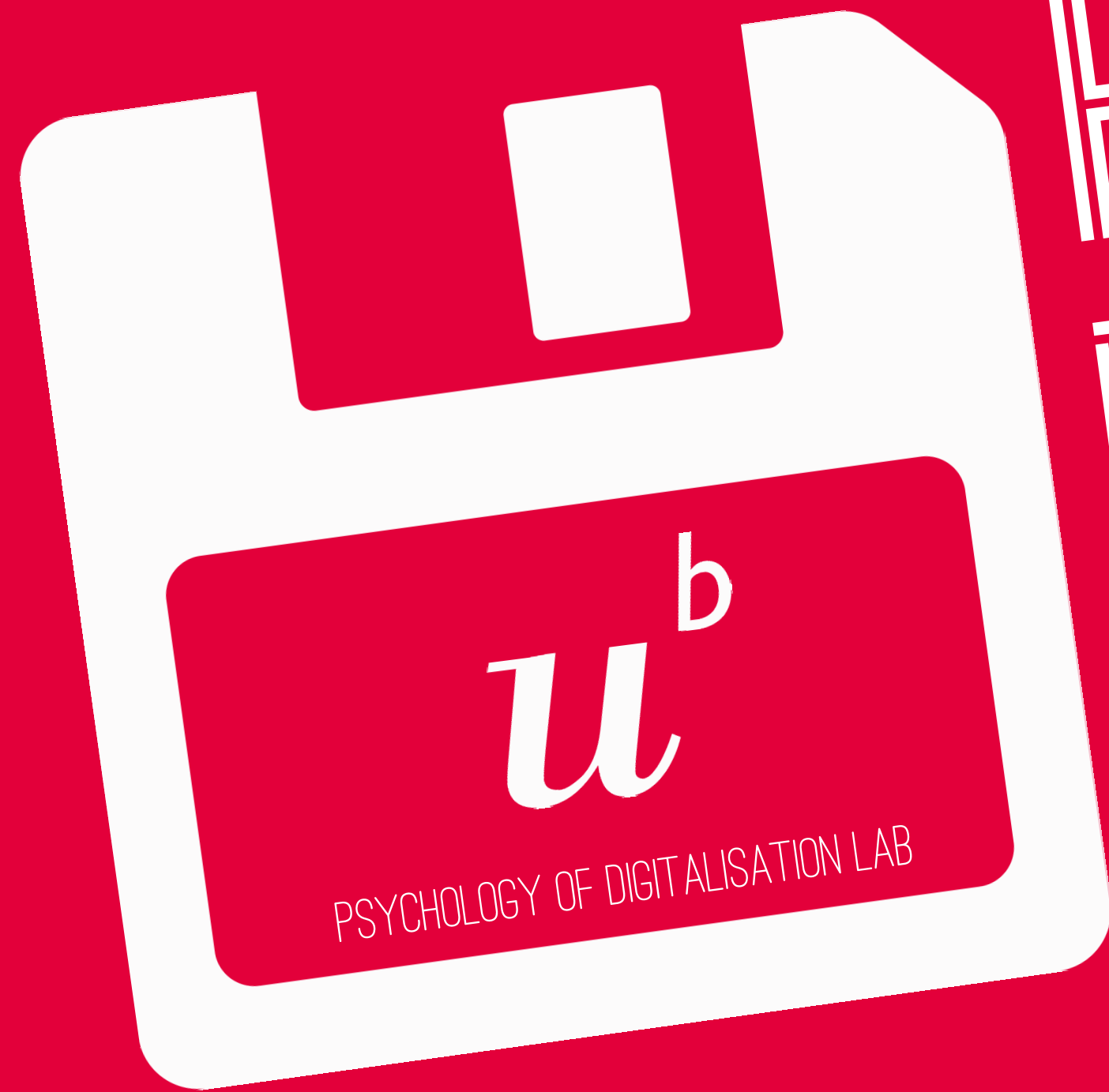




REPRODUCIBLE DATA PROCESSING AND VISUALISATION IN R

WEEK 1: CLASS OVERVIEW AND SETUP

PD DR. IAN HUSSEY

 **EVERYTHING
IS COMPUTER**



 **EVERYTHING
IS COMPUTER**

Technical Skills		
Beginner	Intermediate	Advanced
R you Ready? Reproducible Data Preparation and Analyses in R [Beginner] (S/MS)	Reproducible Data Processing and Visualisation in R [Intermediate] (S/MS)	Improving your Statistical Inferences through Monte Carlo Simulation Studies (S/MS)
	Estimating the Credibility of Past Research (S/MS)	
	Transdisciplinary Data Science (L)	
Substantive Knowledge		
Meta-Science	Privacy and Security	Digitalisation
Psychology of Digitalisation: Everything is Computer (L)		
Digitalisation of Psychology: Computer is Everything (L)		
Psychology in Crisis? (S/MS)	Escape Rooms: Cybersecurity Education at School (S)	Influencer Marketing (S)
	Usable Security (S)	Intervention & Consulting at Lufthansa (S)

GETTING TO KNOW ONE ANOTHER

☐ EVERYTHING
IS COMPUTER

GETTING TO KNOW ONE ANOTHER

- ◆ Aka “help Ian integrate 🇨🇭”
- ◆ How late in the year do you swim in the Aare?
- ◆ What’s the lowest water temperature you swim at?

DER, DIE ODER DAS? WARUM?



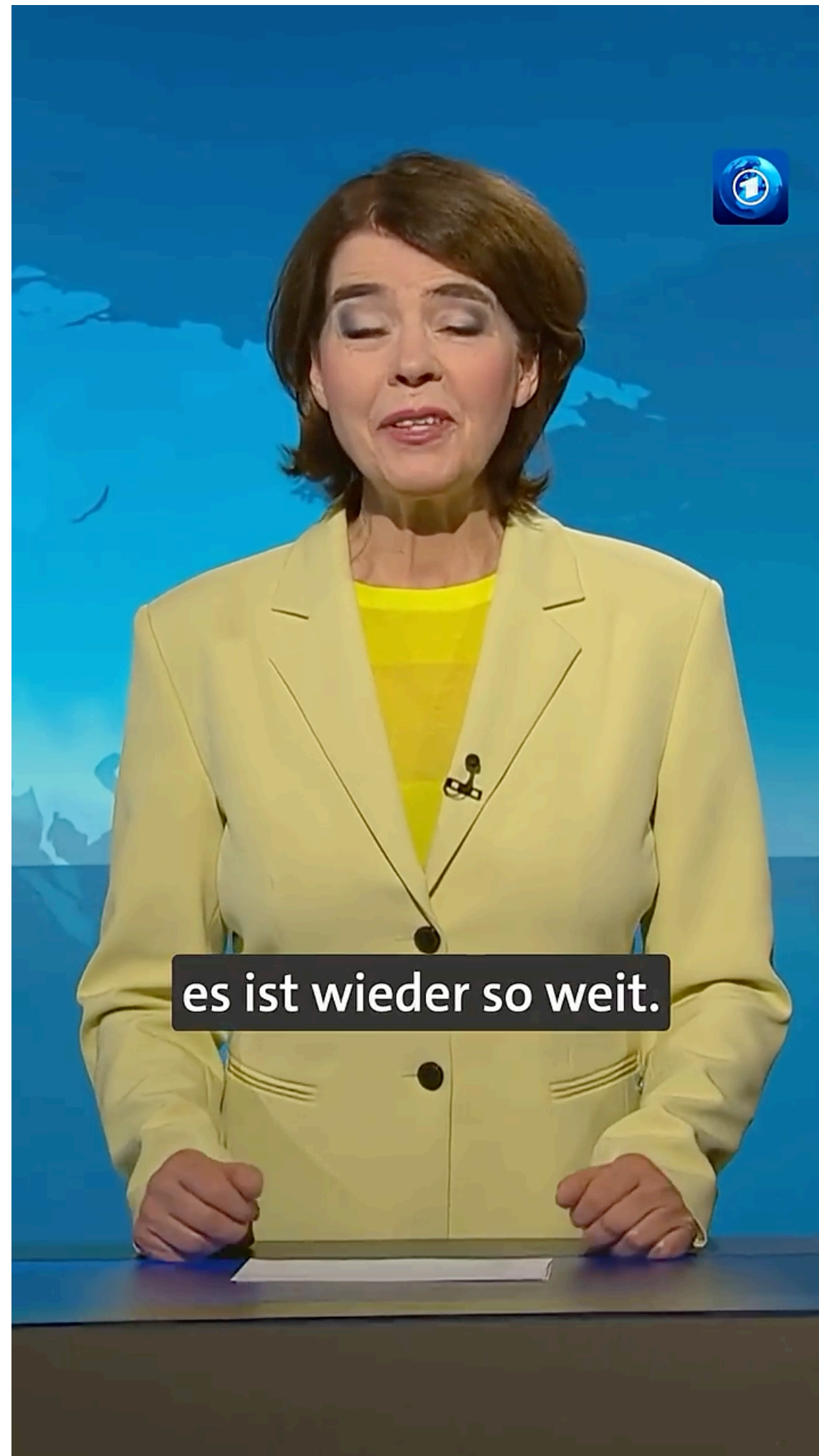
WHO DOES NOT USE LLMS, ALMOST EVER?

WHAT GENDER DO YOU REFER TO LLMS USING?

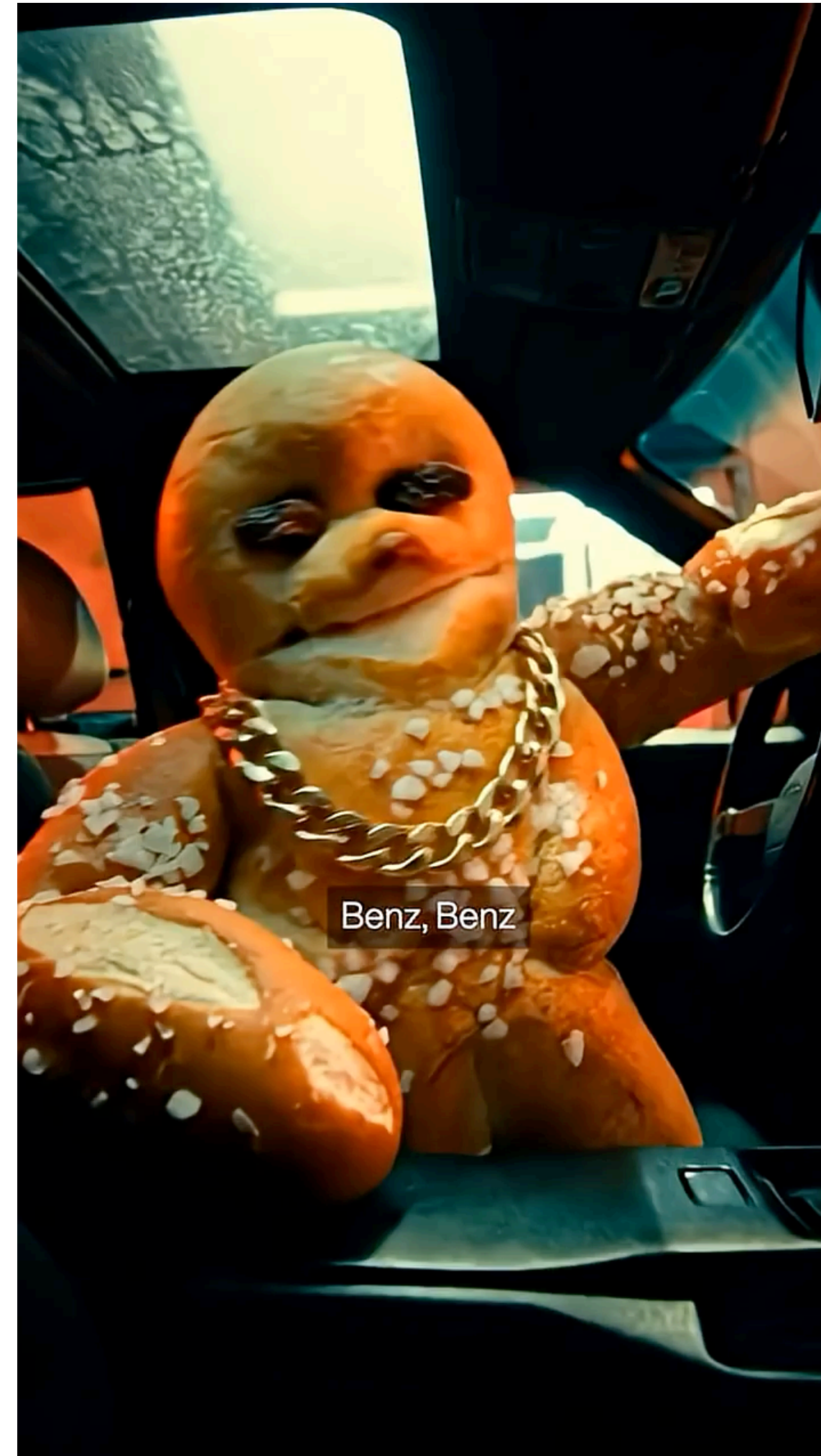
- ◆ “I asked ChatGPT, and [it / he / she / they] sad...” ?

WHICH ONE IS MORE CRINGE?

WHICH ONE IS MORE CRINGE?



WHICH ONE IS MORE CRINGE?



WHICH ONE IS MORE CRINGE?



TikTok
@houseofdrenial

ES ISH GSI

GETTING TO KNOW ONE ANOTHER

- ◆ Communication style
 - Wollen wir uns duzen?
- ◆ Preferred names, pronunciation, pronouns
- ◆ Accessibility & accommodations
- ◆ If you want a break in the middle of class, ask! We often forget
- ◆ Errors we care about in this class
 - English ✗
 - R ✓

PURPOSE OF COURSE

EVERYTHING
IS COMPUTER



Thomas Herndon in 2013

Economics PhD student

Took a class where the professor asked students to examine the results of influential papers



He chose **Reinhart & Rogoff (2010)** 'Growth in a Time of Debt'

Claim:

Economic growth slows dramatically when the size of a country's debt rises above 90% of Gross Domestic Product

Reinhart & Rogoff's article **changed the world**

Influenced economic 'austerity' policies (EU, UK)

Ireland & Greece's economies were sacrificed to stop the spread of the debt crisis to the EU

4 million Irish people took on 190 Billion CHF in debt

- 48k CHF per person

Cut our healthcare, education, and infrastructure

- These still have not recovered



◇	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

Home > Lifestyle

George Osborne plunged UK into austerity due to an ‘error on a spreadsheet’

An academic paper produced by two Harvard economists and relied upon by Cameron and Osborne contained serious miscalculations, it has been claimed.



by Jack Peat — 2022-09-22 15:30 in Lifestyle



George Osborne plunged the UK into austerity “all for nothing” due to an error on an Excel spreadsheet, according to a podcaster.

Hosts James Harkin, Andrew Hunter Murray, Anna Ptaszynski, and Dan Schreiber – from the No Such Thing As A Fish podcast – claim the UK reacted to incorrect data in an academic paper on economic growth and inflation.

Subscribe to our Newsletter

Email address:

Your e-mail address

SIGN UP

View our [Privacy Policy](#) and [Terms & Conditions](#)

More from TLE

- ▶ Watchdog calls for ticket office closures to be delayed over machine upgrades.
- ▶ Richard Curtis uses Al Pacino’s Any Given Sunday speech to inspire global action on climate change
- ▶ James Cleverly holidayed with well-connected Tory lobbyist in Gibraltar
- ▶ Eleveses: Has the Left Abandoned the Working Class?
- ▶ UK recession fears grow as economy contracts
- ▶ Labour calls on Sunak to block Liz Truss’s soon-to-be-published honours list
- ▶ Mick Lynch suggests RMT union will not comply with new anti-strike legislation
- ▶ Russell Brand receives standing ovation at Wembley show following sexual assault

COMPUTATIONAL REPRODUCIBILITY IS LOW

Hardwicke et al. (2021) [10.1098/rsos.201494](https://doi.org/10.1098/rsos.201494)
Crüwell et al. (2023) [10.1177/09567976221140828](https://doi.org/10.1177/09567976221140828)

Stodden et al. (2018) [10.1073/pnas.1708290115](https://doi.org/10.1073/pnas.1708290115)
Obels et al. (2020) [10.1177/2515245920918872](https://doi.org/10.1177/2515245920918872)



COMPUTATIONAL REPRODUCIBILITY IS LOW

36% 9 of 25 psych articles

Hardwicke et al. (2021) [10.1098/rsos.201494](https://doi.org/10.1098/rsos.201494)
Crüwell et al. (2023) [10.1177/09567976221140828](https://doi.org/10.1177/09567976221140828)

Stodden et al. (2018) [10.1073/pnas.1708290115](https://doi.org/10.1073/pnas.1708290115)
Obels et al. (2020) [10.1177/2515245920918872](https://doi.org/10.1177/2515245920918872)



COMPUTATIONAL REPRODUCIBILITY IS LOW

36%

9 of 25 psych articles

7%

1 of 14 psych articles

Replicable

Hardwicke et al. (2021) [10.1098/rsos.201494](https://doi.org/10.1098/rsos.201494)

Crüwell et al. (2023) [10.1177/09567976221140828](https://doi.org/10.1177/09567976221140828)

Stodden et al. (2018) [10.1073/pnas.1708290115](https://doi.org/10.1073/pnas.1708290115)

Obels et al. (2020) [10.1177/2515245920918872](https://doi.org/10.1177/2515245920918872)

COMPUTATIONAL REPRODUCIBILITY IS LOW

36%

9 of 25 psych articles

7%

1 of 14 psych articles

Replicable

26%

53 of 205 Science articles

~~Prestige~~

Hardwicke et al. (2021) [10.1098/rsos.201494](https://doi.org/10.1098/rsos.201494)

Crüwell et al. (2023) [10.1177/09567976221140828](https://doi.org/10.1177/09567976221140828)

Stodden et al. (2018) [10.1073/pnas.1708290115](https://doi.org/10.1073/pnas.1708290115)

Obels et al. (2020) [10.1177/2515245920918872](https://doi.org/10.1177/2515245920918872)



COMPUTATIONAL REPRODUCIBILITY IS LOW

36%

9 of 25 psych articles

7%

1 of 14 psych articles

Replicable

26%

53 of 205 Science articles

~~Prestige~~

34%

21 of 62 Registered Reports

~~Open Science~~

Hardwicke et al. (2021) [10.1098/rsos.201494](https://doi.org/10.1098/rsos.201494)

Crüwell et al. (2023) [10.1177/09567976221140828](https://doi.org/10.1177/09567976221140828)

Stodden et al. (2018) [10.1073/pnas.1708290115](https://doi.org/10.1073/pnas.1708290115)

Obels et al. (2020) [10.1177/2515245920918872](https://doi.org/10.1177/2515245920918872)

STRUCTURE

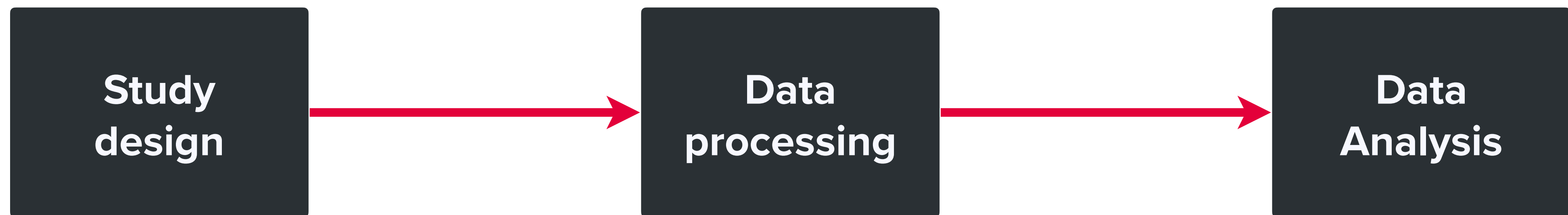
EVERYTHING
IS COMPUTER

COMPUTATIONAL REPRODUCIBILITY IS LEARNABLE

By the end of this class,
you should be able to write more reproducible code
than most psychology researchers

COMPUTATIONAL REPRODUCIBILITY IS LEARNABLE

This course is intended to fill the gap
between other courses





© Sarah Andersen



SCHEDULE

Date	Lecturer	Topic
14.09	Ian	Course structure; Setup, fundamentals, reproducible reports
21.09	Ian	Loading data, pipe, renaming
28.09	Ian	Data transformation I
5.10	Ian	Data transformation II
12.10	Lukas	Data transformation III
19.10	Ian	Structuring projects, strings and factors; Overview assignment 1
26.10	Ian	Shaping and pivots
2.11	Ian	Binding and joining
9.11	Ian	Sharing and privacy; Intro to github
16.11	Lukas local, Saloni remote	Principles of visualization
23.11	Lukas	Plotting using ggplot2; Reporting; Overview of assignment 2
30.11	Lukas	The linear model
7.12	Ian	Oral examination
14.12	Ian	Oral examination

COMMUNICATION

- ◆ Please use Slack
 - You will be sent an invite
 - Class channel where possible, DM if necessary
 - If necessary, ian.hussey@unibe.ch

REQUIREMENTS

- ◆ Weekly attendance (80% minimum)
- ◆ Laptop + recent version of R, RStudio, & {tidyverse}
- ◆ Semi-weekly exercises
- ◆ Self-guided practice

ASSESSMENT

- ◆ Two at-home assignments
- ◆ One in-class 1:1 oral assessment
 - ◆ Pilot to understand LLM-resistant ‘proof of thought’ assessments
- ◆ Each is 1/3 of your grade

ASSESSMENT

- ◆ All assignments must be licensed CC BY 4.0
 - i.e., can be used or modified with attribution
- ◆ In the service of:
 - Our materials are open source, so will yours be
 - Building a cumulative knowledge base of simulation studies
 - Practicing Open Science skills



AI

- ◆ You may use AI (Claude, Gemini, ChatGPT, etc.) within reason and for the purposes of learning but not completing assignments
- ◆ Make it help you learn, don't outsource thinking and understanding to it.
 - “Fix this map() code”
 - “Teach me how to use map()”
- ◆ You take ultimate responsibility for your code and learning
- ◆ We reserve the right to ask you to explain your assignment's code in an oral exam

CONTENT

EVERYTHING
IS COMPUTER

Introduction

- 1 Installation and setup
- 2 Fundamentals
- 3 Reproducible reports
- 4 Loading, viewing, and saving data
- 5 The pipe and renaming
- 6 Data transformation I
- 7 Data transformation II
- 8 Working with strings and factors
- 9 Reshaping and pivots
- 10 Binding and joining
- 11 Structuring projects
- 12 Visualization
- 13 Reporting
- 14 Data privacy and anonymity
- 15 License and citation

Reproducible Data Processing and Visualization

in R and tidyverse

AUTHOR
Ian Hussey

PUBLISHED
August 27, 2025

Introduction

This [Open Source](#) eBook provides materials for the semester-long Master’s seminar course “Reproducible Data Processing and Visualization in R” that I deliver at the University of Bern’s Institute of Psychology.

How to use this book

This book is made up of individual Quarto (.qmd) workbooks. Many of the exercises are easiest to complete in your own local copy of these .qmd files.

I suggest that you [download the .zip file](#) of all the .qmd files and their supporting data and cheatsheets and use the eBook as a reference book. *If the the .zip file does not contain the most recent versions, please contact me and I’ll updated it.* If I’m slow or you’re feeling impatient, you can also [download the most recent versions from GitHub](#), along with the other files that create this eBook.

You can also copy and paste the code for any chapter directly from the website. Click the “</> Code” button on the top right of each page to see the full .qmd file’s code. You can copy and paste this into a .qmd file. However, it’s probably easier to download all the .qmd files and data as mentioned above.

Learning to code is a practice skill. Almost anyone can become competent in writing reproducible code for data processing and visualization with practice. More than anything else, completing this course requires that you practice in your own time, using not only the examples provided but also ones you create yourself. Take real data sets from your own studies, or the thousands available on the Open Science Framework ([osf.io](#)), or create simulated messy datasets with my R package [{truffle}](#) and practice.

</> Code

Table of contents

- Introduction
- How to use this book
- Other learning resources
- Find me elsewhere
- Contributing



Reproducible
Data Processing
and Visualization

in R and tidyverse

Ian Hussey

Introduction

- 1 Installation and setup
- 2 Fundamentals
- 3 Reproducible reports
- 4 Loading, viewing, and saving data
- 5 The pipe and renaming
- 6 Data transformation I
- 7 Data transformation II
- 8 Working with strings and factors
- 9 Reshaping and pivots
- 10 Binding and joining
- 11 Structuring projects
- 12 Visualization
- 13 Reporting
- 14 Data privacy and anonymity
- 15 License and citation

Reproducible Data Processing and Visualization

in R and tidyverse

AUTHOR
Ian Hussey

PUBLISHED
August 27, 2025

Introduction

This [Open Source](#) eBook provides materials for the semester-long Master’s seminar course “Reproducible Data Processing and Visualization in R” that I deliver at the University of Bern’s Institute of Psychology.

How to use this book

This book is made up of individual Quarto (.qmd) workbooks. Many of the exercises are easiest to complete in your own local copy of these .qmd files.

I suggest that you [download the .zip file](#) of all the .qmd files and their supporting data and cheatsheets and use the eBook as a reference book. *If the the .zip file does not contain the most recent versions, please contact me and I’ll updated it.* If I’m slow or you’re feeling impatient, you can also [download the most recent versions from GitHub](#), along with the other files that create this eBook.

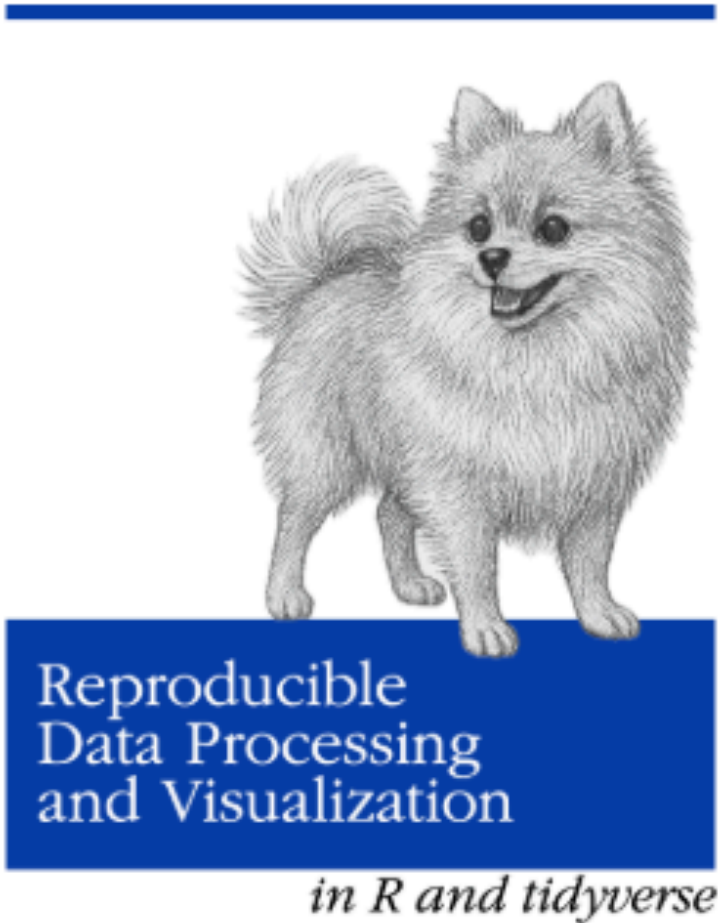
You can also copy and paste the code for any chapter directly from the website. Click the “</> Code” button on the top right of each page to see the full .qmd file’s code. You can copy and paste this into a .qmd file. However, it’s probably easier to download all the .qmd files and data as mentioned above.

Learning to code is a practice skill. Almost anyone can become competent in writing reproducible code for data processing and visualization with practice. More than anything else, completing this course requires that you practice in your own time, using not only the examples provided but also ones you create yourself. Take real data sets from your own studies, or the thousands available on the Open Science Framework ([osf.io](#)), or create simulated messy datasets with my R package [{truffle}](#) and practice.

</> Code

Table of contents

- Introduction
- How to use this book
- Other learning resources
- Find me elsewhere
- Contributing

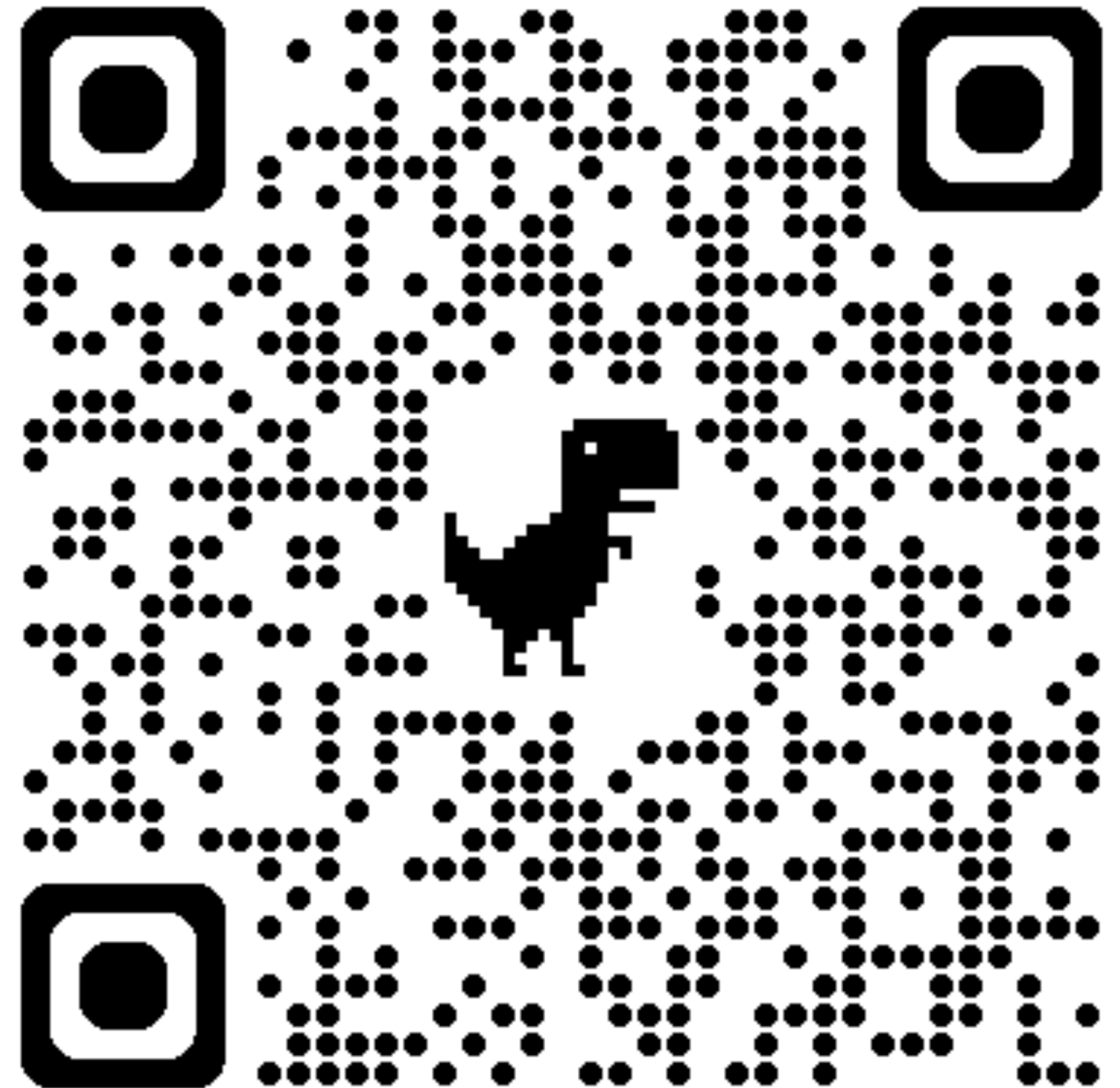


Ian Hussey



BOOK

wrangling.tidyverse.se



BOOK

- ◆ During class, week 1
 - Chapter 1: Installation and setup [reference]
 - Chapter 2: Fundamentals [reference]
 - Chapter 3: Reproducible reports
- ◆ At home, before week 2's class
 - Chapter 4: Loading data
 - Chapter 5: The pipe and renaming