Creating a curriculum centered on reproducible research

Phil McAleer

School of Psychology, University of Glasgow

#PsyTeachR

philip.mcaleer@glasgow.ac.uk @mcaleerp @uofgpsychology

RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. Replication effects were half the magnitude of original effects, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

http://science.sciencemag.org/content/349/6251/aac4716.full.pdf+html

100 Published Papers Reviewed
36% replicated original finding
64% found no effect, a weaker effect, or the opposite effect to the original finding.

2016

'Changing these incentives ***[for publication of positive results etc]*** *requires a cultural shift in both thinking and practice. Improved doctoral and postdoctoral research methods training is vital (Munafò et al., 2014).* ***However, changing scientific culture can begin at the undergraduate level, instilling the principles of transparency and scientific rigor at the grassroots.'***

Instilling scientific rigour at the grassroots

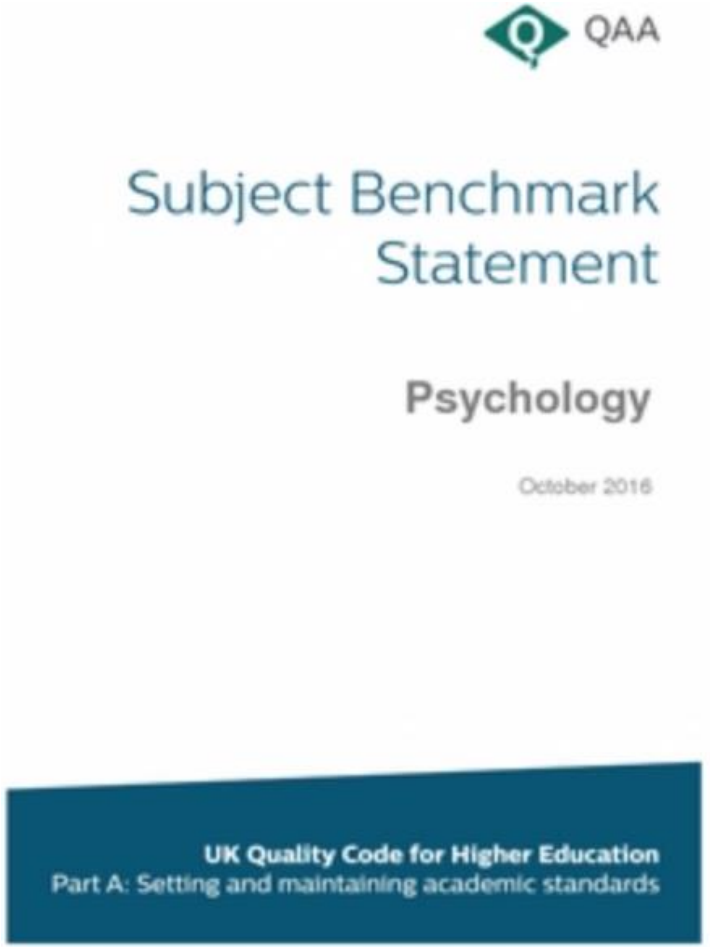
A letter from our March edition advocates consortium-based undergraduate projects.

There is increasing awareness of the problem of unreliable findings across social, psychological and biomedical research. The 'publish or perish' culture, and the bias towards generating novelty and positive results, may incentivise running multiple small studies measuring multiple outcomes. This, combined with flexible analytical procedures, can generate a large number of positive results, but many will be false

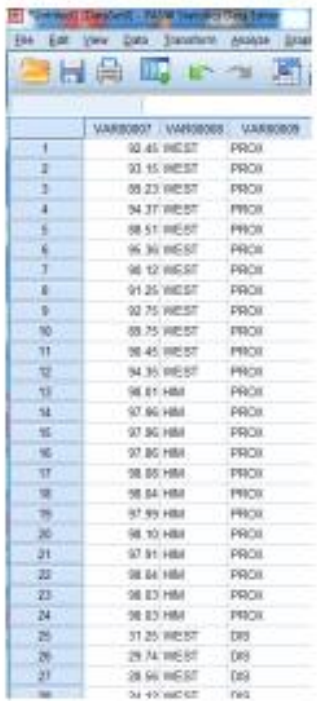
https://thepsychologist.bps.org.uk/volume-29/march-2016/instilling-scientific-rigour-grassroots

2016

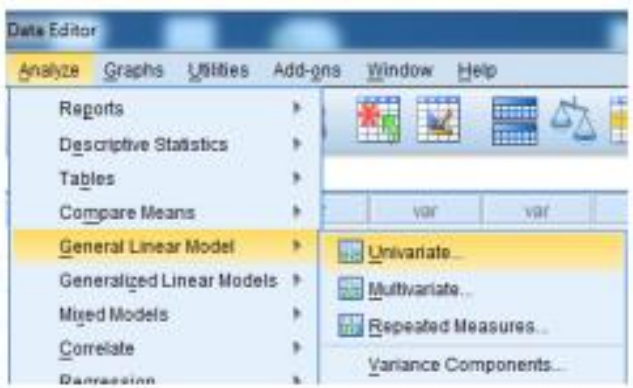
*‘carry out an extensive piece of empirical research that requires them **individually** to demonstrate a range of research skills including planning, considering and resolving ethical issues, analysis and dissemination of findings.’*

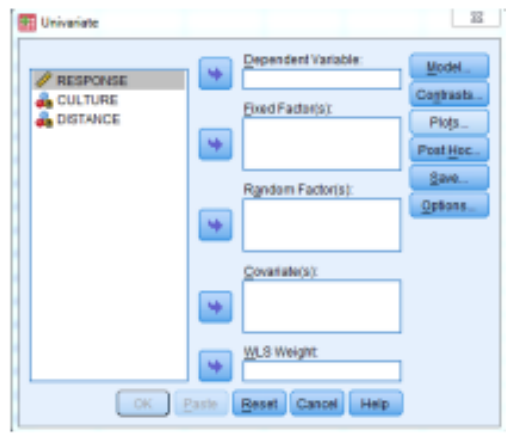


https://www.qaa.ac.uk/

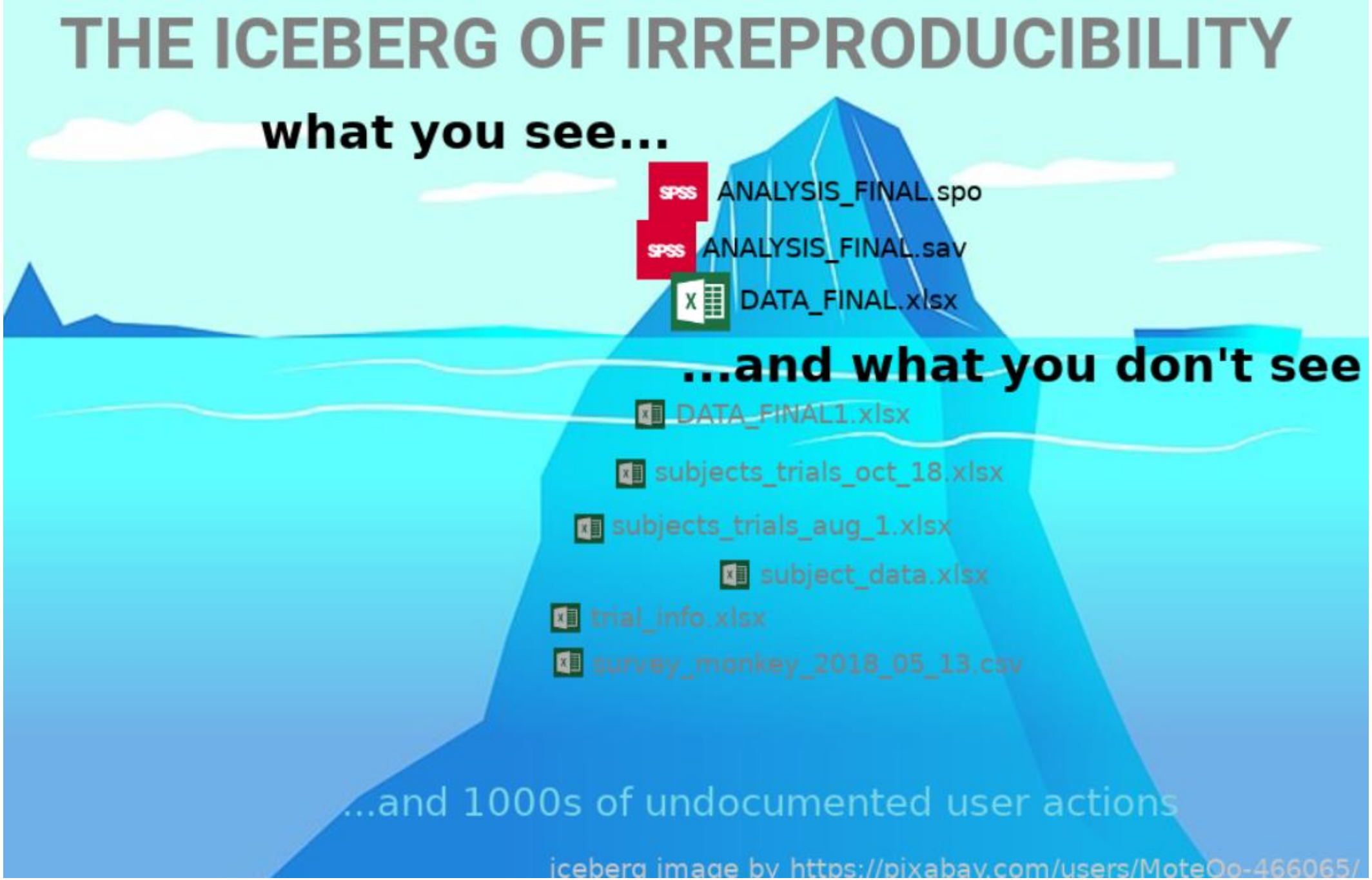


6.2. Then select, '**General Linear Model**', and then '**Univariate**':





https://talklab.psy.gla.ac.uk/barr_sips_2019/

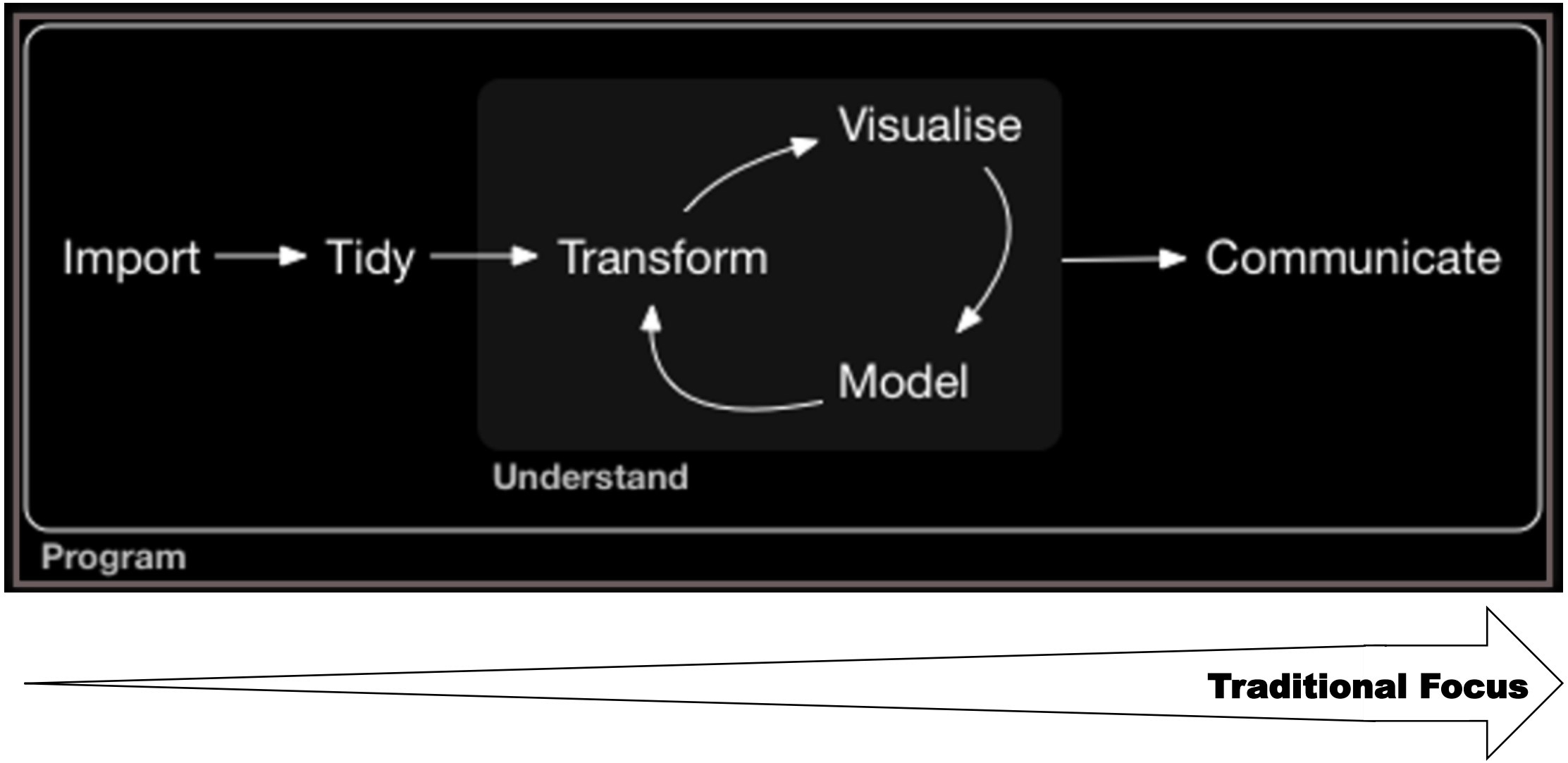


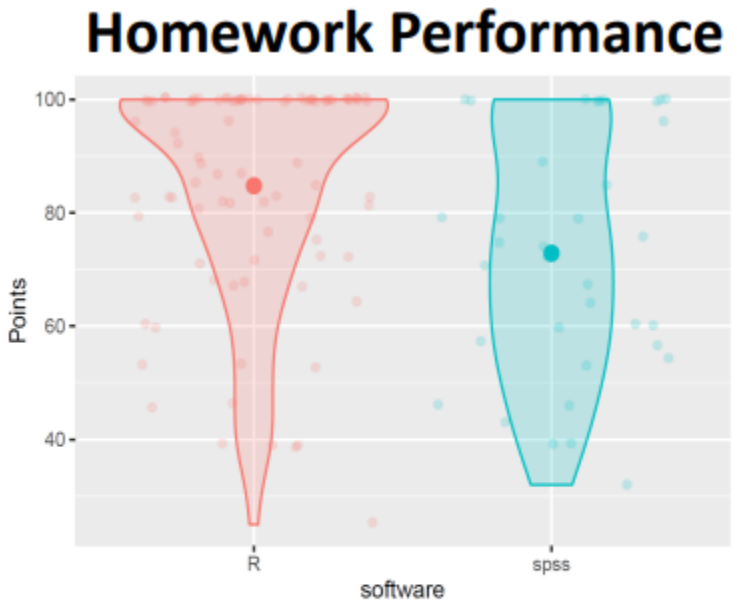
https://talklab.psy.gla.ac.uk/barr_sips_2019/

Knowledge and Skills Gap

- no formal instruction on structuring, transforming, and visualizing data, despite these activities being:
  - the most time consuming
  - the most complex and thus error prone
- absence of formal instruction leads to
  - shallow reproducibility of projects at best
  - over-confidence in some areas as well as insecurity and negative self-image
- lack of knowledge of reproducibility and issues in the field
  - again no strong formal training outside lectures on topics

Grolemund & Wickham (2017), *R for Data Science*

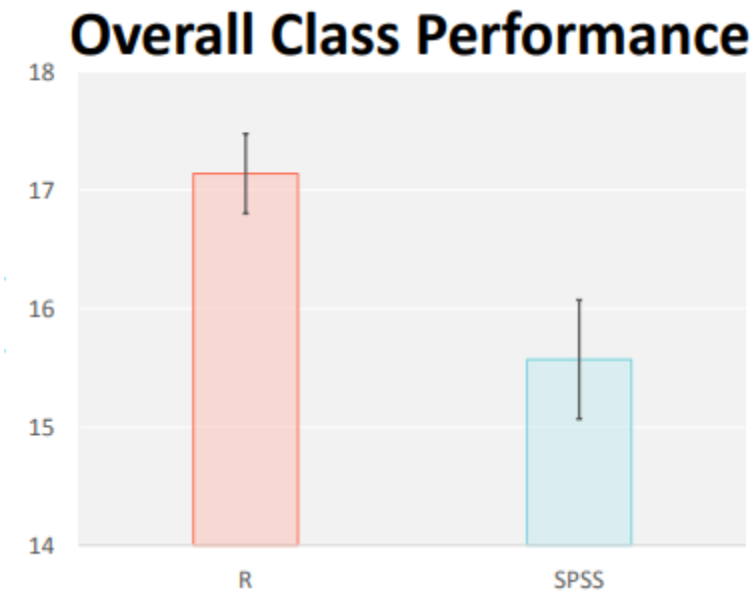


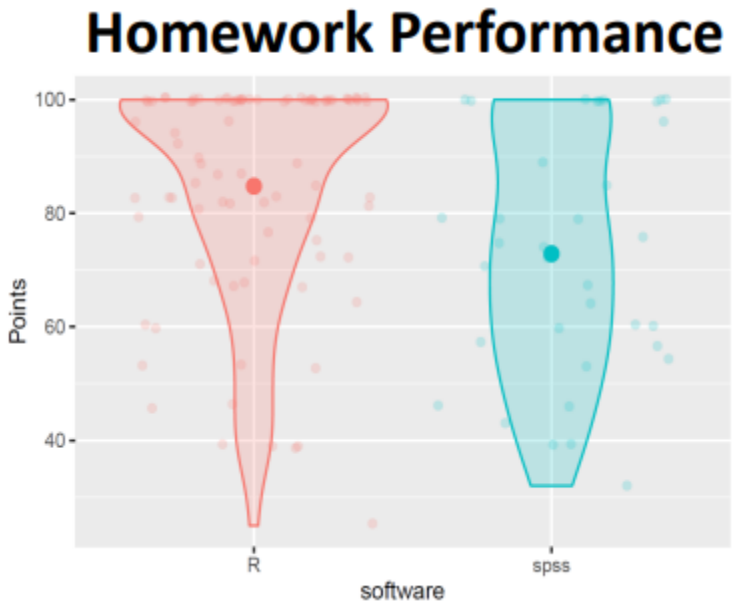


- Overall performance significantly higher in R group than SPSS group, p = .011
- Homework performance significantly higher in R group than SPSS group, p = .006. Difference remained after controlling for student performance on other assignments not related to data-wrangling.
- Overall Anxiety (not shown) significantly higher in SPSS group (*M* = 2.35) than in R group (*M* = 1.97), p = .005

To R or to SPSS: Does autonomous choice of learning technology affect competency & anxiety in Psychology undergraduates?

Dale J. Barr, Phil McAleer, Niamh Stack and Maxine V. Swingler





"We learned the background and theory behind the tests that we run, rather than blindly pressing buttons in a Lab.."

"Additionally, by having to hand in the homework every week we actually have to spend time thinking about the material and really get to properly understand it, instead of trying to cram everything 2 days before the exam."

To R or to SPSS: Does autonomous choice of learning technology affect competency & anxiety in Psychology undergraduates?

Dale J. Barr, Phil McAleer, Niamh Stack and Maxine V. Swingler

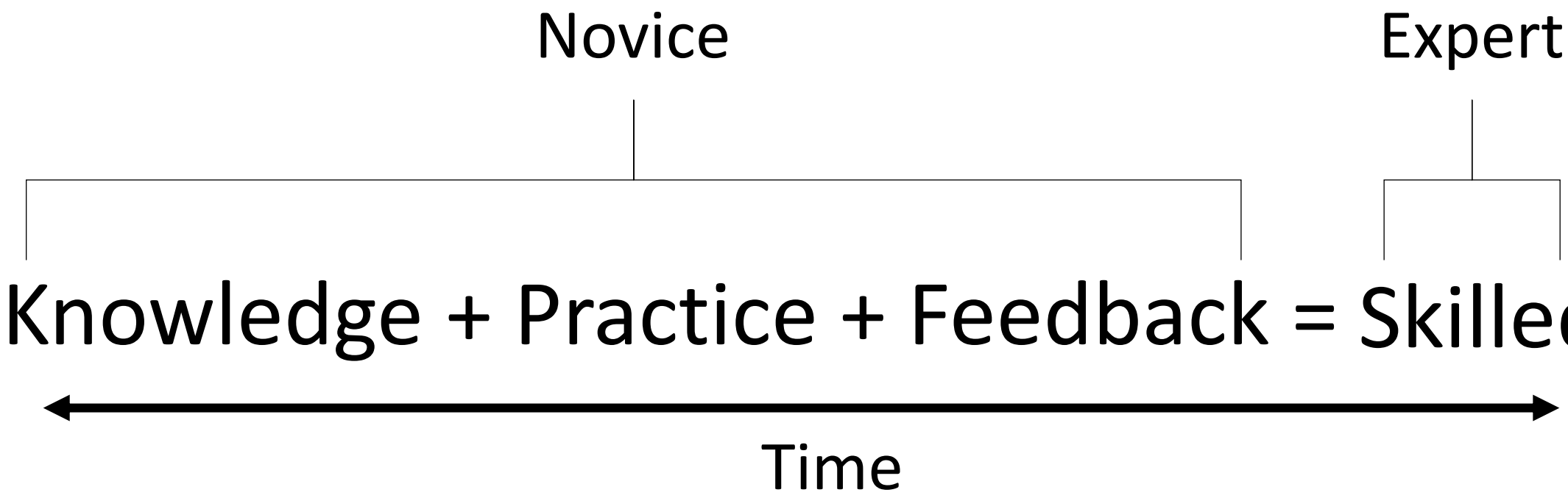
https://psyteachr.github.io/

Our curriculum now emphasizes **essential 'data science' graduate skills** that have been overlooked in traditional approaches to teaching, including programming skills, **data visualisation, data wrangling and reproducible reports**. Students learn about probability and inference through data simulation as well as by working with **real datasets**.

Working with data and research in a reproducible fashion

=

R is our tool as it allows us to teach the skills we want our students to have.
You have to be clear on what it is you want students to learn regardless of software.
Added benefit of being free to use allowing for greater student independence.



- True for any practical skill including data wrangling, analysis, visualisation, placement of citations, report writing and paragraph structure, etc.
- Teaching is about breaking down the skills into knowledge and supplying opportunities for practice and feedback
- Additional element of flexibility – **the skills stay the same; only the data changes**

Didau: https://learningspy.co.uk/learning/novice-expert-model-learning/

2.3.3 Task 3: Load in the Data

Now we have to load in the `.csv` datafiles using the `read_csv()` function and save them as variables in our environment. For example, to load in the `responses` we would type:

```
responses <- read_csv("responses.csv")
```

1. Add the following lines of code to your script and complete them to load in all four `.csv` datafiles. Use the above code as an example and name each variable the same as its original filename (minus the .csv part), again as above, e.g. `responses.csv` gets saved as `responses`. Remember to run the lines so that the data loaded in and is stored in your environment.

```
responses <-  read_csv()    # survey responses
qformats <-                 # question formats
scoring <-                  # scoring info
pinfo <-                    # participant information
```

Portfolio Point - Haven't I read_csv before

As you work with data and functions you will find there are functions with similar names but that give different results. One of these is the `read` function for `csv`. Make sure to always use `read_csv()` as your function to load in `csv` files. Nothing else. It is part of the `readr` package automatically brought in with `tidyverse`.

- **use a novice friendly syntax**
- **real data wherever possible**
- **PreClass, InClass, Assignment Knowledge, Practice, Feedback**

https://psyteachr.github.io/ug2-practical

1. From the options, choose the correct citation for the AQ 10 question questionnaire:
   Allison, Auyeung, and Baron-Cohen, (2012) ▾
2. Complete the sentence, the higher the AQ score... the more autistic-like traits displayed ▾
3. Type in the AQ score (just the number) of Participant ID No. 87:
4. Type how many participants had an AQ score of 3 (again just the number):
5. The cut-off for the AQ10 is usually said to be around 6 meaning that anyone with a score of more than 6 should be referred to for diagnostic assessment. Type in how many participants should we refer from our sample:

Explain This - I dont get these answers

1. From the link above you can see that an appropriate citation for the AQ10 would be (Allison, Auyeung, and Baron-Cohen, (2012))
2. As mentioned, the higher the score on the AQ10 the more autistic-like traits a participant is said to show.
3. You could do this by code with `filter(aq_scores, Id == 87)`, which would give you a tibble of 1x2 showing the ID number and score. If you just wanted the score you could use `pull()` which we havent shown you that yet: `filter(aq_scores, Id == 87) %>% pull(AQ)`. The answer is an AQ score of 2.
4. Same as above but changing the arguement of the filter. `filter(aq_scores, AQ == 3) %>% count()`. The answer is 13. Remember you can do this by counting but the code makes it reproducible and accurate every time. You might make mistakes.
5. `filter(aq_scores, AQ > 6) %>% count()` or `filter(aq_scores, AQ >= 7) %>% count()`. The answer is 6.

- **Integrate theory, research methods, terminology as well as data skills**
- **Allow for formative self-assessment using webex package and RMarkdown**

https://cran.r-project.org/web/packages/webex

https://psyteachr.github.io/ug2-practical

2.3.7 Task 7: Calculating the AQ Scores.

We have now created `rscores` which has information on how each participant responded to each question and how each question should be coded and scored, all within the one tibble. All we need now is to sum the scores for each participant to get their AQ score.

1. Based on your Preclass knowledge, copy the below line into your code and complete it to obtain individual `aq_scores` for each participant.
2. Save your script and run it all again from the start to make sure it works!

```
aq_scores <- rscores %>%
             group_by() %>% # how will you group individual participants?
             summarise(AQ = sum()) # which column will you sum to obtain AQ scores?
```

Helpful Hint

Each participant could be grouped by their Id.

If we summed up the value for each Score we might get a full AQ Score for each particpipant.

2.5.2.3 InClass Task 7

```
aq_scores <- rscores %>%
             group_by(Id) %>% # group by the ID number in column Id
             summarise(AQ = sum(Score)) # sum column Score to obtain AQ scores.
```

Return to Task

- **Increase complexity and reduce hints as chapters progress**
- **Solutions are at the end of each chapter to allow people to check themselves and to move on**

https://psyteachr.github.io/ug2-practical

Year	Topics	N	College/Additional Subjects
1	IDE, Coding, data wrangling, visualisation, probability, descriptives	~500-600	Psychology + College Specific Modules: • **Arts:** English/History/Languages/etc • **Social Science:** Business/Sociology/etc • **Science:** Biology/Chemistry/Computing
2	Recap L1 skills, simulate data, power, inferentials (correlations, regressions, t-tests), the GLM	~300	
3	Simulation based approach to mixed-effects models, psychometric data	~150-200	Various modules related to Psychology
4	Advanced topics such as factor analysis, bootstrapping and permutation tests		

- **Build the skills as you go with focus on becoming confident and competent**
- **Build a solid foundation in key skills (wrangle, visualise) before moving on to analyses**
- **It is ok to slow it down. You do not have to teach everything; teach a few things well**

- One year course; N = ~200
- Diverse graduate skill sets within cohort from degrees in arts to science with varying goals
- Very few have data skills or research methods background
- About 30 hours to learn methods, stats and communication

- 10 week course; N = ~ 60
- Diverse graduate skill sets from undergrad psych and neuroscience degrees & plan to do a PhD.
- Some have data skills, many have research methods and stats knowledge, many don't have any of this
- 20 hours of Data skills that prepare students for advanced stats course

Regardless of duration of course what is fundamental is that you teach the knowledge and skills you want your students to gain in your course

Task 2: Reshaping the STARS

Having had a look at `stars_raw` you will see that it is every answer to every question from all 37 participants, but in wide format. This is often how data comes out of experiments or questionnaires but it is rarely how we want data. We want our data to be in tidy format to begin with, i.e. one column with the participant IDs, one column with the question numbers, and one column with the responses.

- In the `T2702` code chunk below replace the `NULL` with code pipeline (e.g. %>%) to gather all the questions and responses into tidy format.
  - hint: pivot_longer might work
  - store the output as a tibble in `stars_gather` with three columns and 1887 rows (i.e. 37 participants * 51 questions = 1887 rows)
  - The first column should be called `ID` and show the participant ID number
  - The second column should be called `question` and show the question numbers (e.g. Q01, Q51).
  - The third column should be called `response` and show the response to each individual question by each participant.
  - Be exact on column order (ID, question, response) and spelling of column names - do not sort the columns.

```
stars_gather <- NULL
```

Task 3: Combining questions with subscales

Now we are going to have to eventually calculate a mean score for each participant for each subscale but first we need to map each question onto the appropriate subscale.

- In the `T2703` code chunk below replace the `NULL` with code that combines the information in `stars_gather` with the information in `subscales` and store it in `stars_sub`.
  - hint: inner join by the common column in both tibbles. There is only one common column
  - `stars_sub` should have the following column order in this exact order: `ID`, `question`, `response`, `subscale`
  - there will be 4 columns of 1887 rows
  - `stars_sub` should be a tibble

```
stars_sub <- NULL
```

Regular assignments in RMarkdown that tests the knowledge and skills, building complexity week on week, with strong notion of the code must be reproducibile.

Task 6: The descriptives

Looking Good! Next, when writing up experiments we need to have a note of measures of spread (e.g. standard deviation) and central tendency (e.g. mean) for each condition, as well as how many people are in each condition.

- Using **one pipeline** replace the `NULL` in the T6 code chunk below to recreate the descriptives table shown below. Pay particular attention to the names and order of columns, and order of rows.

Store the output in the tibble called `descriptives`. Your table should match what is shown below.

Your answer and feedback

```
descriptives <- combined %>%
  group_by(country, condition) %>%
  summarise(n = n(), mean = mean(correct_pct), sd = sd(correct_pct))
```

- your table `descriptives` matched the solution table

Solution

```
descriptives <- group_by(combined, country, condition) %>%
  summarise(n = n(), mean = mean(correct_pct), sd = sd(correct_pct))
```

country	condition	n	mean	sd
Canada	hooligan	47	46.31206	12.93301
Canada	professor	40	41.75000	13.83640
United Kingdom	hooligan	38	54.91228	13.41723
United Kingdom	professor	30	53.44444	11.49657

- Again goes back to the idea of wrangling. Key skills that you have developed over the semester by asking the same task with different datasets.
- We have also introduced a slightly different mindset here where we show you what we want and ask you to reproduce it. It is all about thinking about what you want to achieve, and then thinking the steps through.

Students get individual as well as generic feedback allowing us to reinforce corrections to common mistakes.
E.g. Not looking at output.

https://github.com/dalejbarr/assessr

Group	GUID	MoodleID	Research	Evaluation	Communication	ForNextTime	Grade
Lab 3 Tues		1E+08	You have ol	Your evaluati	Writing was good; y	You asked about r	C2
Lab 3 Tues		1E+08	You have ol	Evaluation w	Communication wa	You asked about s	B2
Lab 3 Tues		1E+08	You have ol	Evaluation w	Communication wa	You asked about r	B2
Lab 3 Tues		1E+08	You have ol	Evaluation w	Communication wa	You asked about h	B3
Lab 3 Tues		1E+08	While you (	Evaluation w	Communication wa	In general you ne(	C3

Cognition: Question 2 | Perception: Question 4 | Psychobiology: Question 8 | Distribution of Marks

Question: Under what conditions may some people be susceptible to false memories?

Individual Feedback

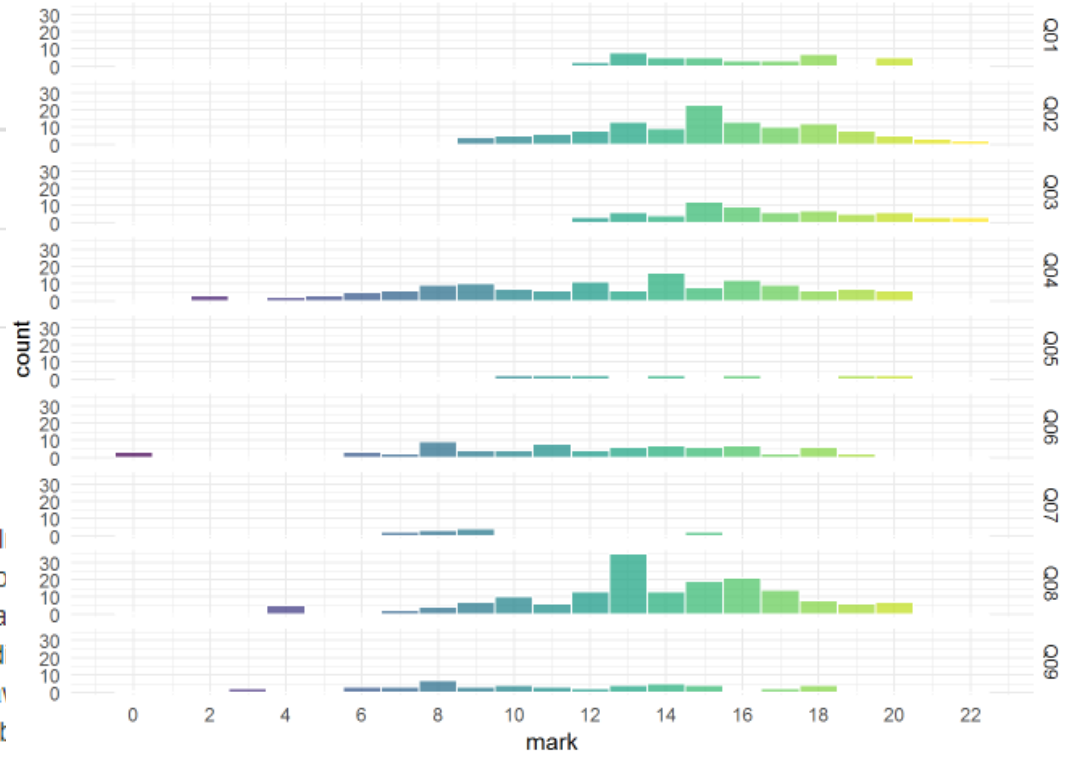
NB: Stars represent marker's assessment of your attainm

Cognition: Question 2 | Perception: Question 4 | Psychobiology: Question 8 | Distribution of Marks

Category	***Excellent*** **keep this up**	***Very Good*** **can be improved by further refining**
Knowledge & Research		*
Comprehension & Evaluation		*
Academic Communication		

Generic Feedback

False memories can occur due to a range of conditions. I
allows for errors and intrusions to take place. Examples o
memorized lists (the DRM procedure); contextual or situa
mistaken identification of Bugs Bunny at Disneyland); ind
believe in reincarnation). Good answers explored and ga
scientific literature, and showed evidence of scholarship t



Demanding on staff to learn this approach and novel ideas.

- **Integrate where possible to improve uptake**
- **Coding skills benefit staff both in terms of own research and admin**
- **markR package**

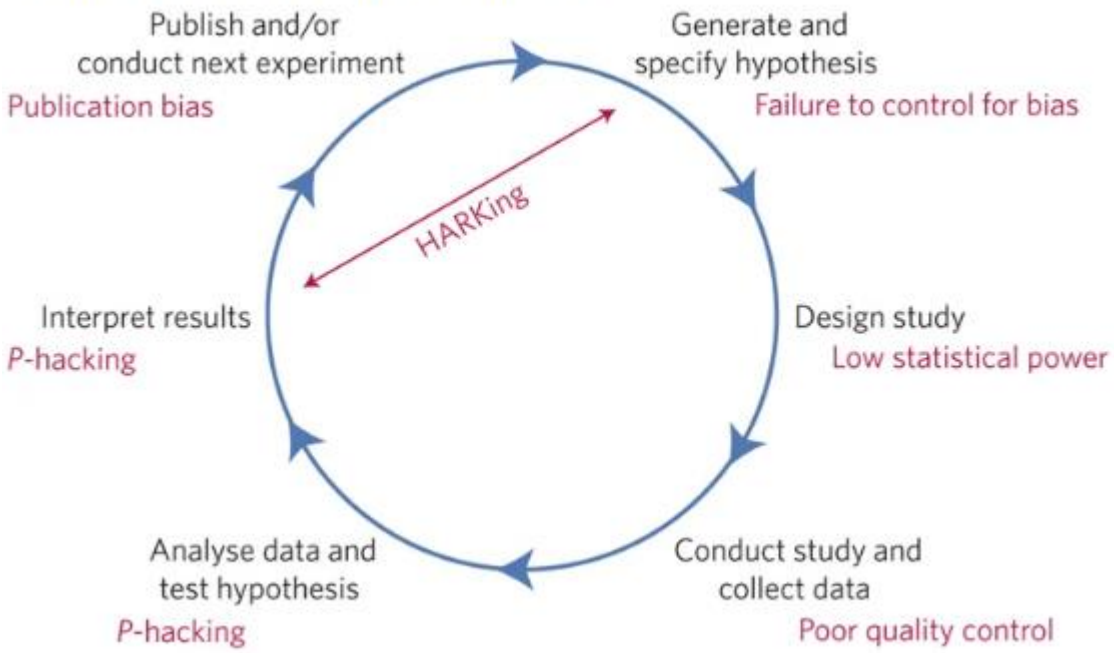
https://debruine.github.io/markr2/

A manifesto for reproducible science

Marcus R. Munafò ✉, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware & John P. A. Ioannidis

Nature Human Behaviour **1**, Article number: 0021 (2017) | Download Citation

Figure 1: Threats to reproducible science.



Not just about coding

Box 2: Distributed student projects.

Student assessment requirements, and limited access to populations of interest, may hinder extensive collaboration within a single institution, but
institutions in the form
model, academics and s
form a consortium, coll
protocol and analysis p
collection. The protoco
each participating cent
analysis. Consortium m
could be used to integr
offering opportunities f
Conclusions based on r
preparation for wider d
conventions such as th
would learn rigorous re
participation in researc
conducted to be genuir
team science would be

Box 3: Registered Reports.

The Registered Reports (RR) initiative seeks to eliminate various forms of bias in hypothesis-driven research[52,53], and in particular, the evaluation of a study based on the results. Unlike conventional journal articles, RRs split the peer review process into two stages, before and after results are known. At the first stage, reviewers and editors assess a detailed protocol that includes the study rationale, procedure and a detailed analysis plan. Following favourable reviews (and probably revision to meet strict methodological standards), the journal offers in-principle acceptance: publication of study outcomes is guaranteed provided the authors adhere to the approved protocol, the study meets pre-specified quality checks, and conclusions are appropriately evidence-bound. Once the study is completed, the authors resubmit a complete manuscript that includes the results and discussion. The article is published at the end of this two-stage process. By accepting articles before results are known, RRs prevent publication bias. By reviewing the hypotheses and analysis plans in advance, RRs should also help neutralize P-hacking and HARKing (hypothesizing after the results are known) by authors, and CARKing (critiquing after the results are known) by reviewers with their own investments in the research outcomes, although empirical evidence will be required to confirm that this is the case.

Not just about coding

Assignments based on reproducible research methods to develop the understanding and knowledge of different practices

- **Essays:**
  - Help develop understanding of terminology and improve critical thinking in the area
  - Discuss whether reproducible research is important to Psychology or not?
  - Evaluate whether a ManyLabs approach is beneficial in developing new theory?
  - Evaluate how the task may influence findings in personality perception research?
- **Peer-Review of Preprints**
  - Shows that the published articles are not error free
  - Develops a healthy skepticism of published research
  - Develops writing skills, critical thinking on methods, and professional courtesy
  - Using open access peer-reviews and pre-printed submissions to improve students' comprehension of academic writing: http://eprints.gla.ac.uk/162868/7/162868.pdf

Not just about coding

- **Registered Reports**
  - Starter topics to develop just the introduction and methods on new research
  - Focus on the research question and approach and not the outcome
  - Develops students understanding of justification, design, and rationale
  - Encourage forward planning and thinking about reproducibility from the start of a project
- **Secondary Data Analyses and Pre-Registrations of a known dataset**
  - Real data that real psychologists have gathered (often in our School)
  - Helps show the worth of the skills we have been developing
  - Shows data is big and messy, and that the findings in a research article are partly based on the decision a researcher makes in the design and analysis
  - Teaches about documenting analysis plan and deviations from it

Not just about coding

Added benefit: the conversations & discussions that these assignments spark

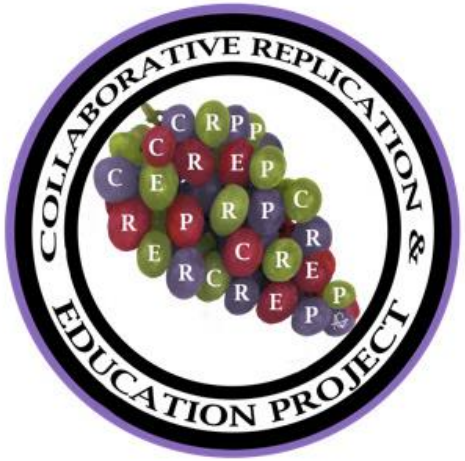
- E.g. Registered Report
  - How many participants would you include?
  - How do you decide that? Power Analysis vs Previous Research
  - The previous research uses less participants than the Power Analysis suggests yet finds a significant effect. What does that mean?
  - Leads to discussion of issues of false positives, false negatives, replications, in a context that students understand better because it is in a project they are interested in and are developing themselves. Real issues and real problems to understand and think about approaches to.
- Students become more aware of the issues in research and how to do better in future assignments and projects.

The Many Lab

Contributors: Charles R. Ebersole, Richard A. Klein, Olivia E. Athertor

Affiliated institutions: Center For Open Science

Date created: 2014-05-15 08:41 PM | Last Updated: 2019-03-27 06:17 PM



Psychological Science Accelerator

A Distributed Laboratory Network

PLOS ONE | TENTH ANNIVERSARY

Collabra: Psychology

OSF PREPRINTS

The open preprint repository network

PeerJ

ROYAL SOCIETY OPEN SCIENCE

PsyArXiv

A free preprint service for the psychological sciences.

Bringing the tools and ideas from our research communities into our teaching to develop the skills we want students to have.

Building a community

You can change as much as you like but a sense of community is vital to making it last.

- Group work dominates a lot of our classes with students working with peers as we would work with colleagues in a lab
- Support creates trust with people becoming more open to talking about mistakes and errors, and asking for help.
- Seminar and Workshop Series based around methods and metascience that is open to both students and staff to present ideas and questions: https://psyteachr.github.io/mms/

Changes don’t have to be complex. Even changing a lab environment from isolated desks to a layout that promotes discussion, as we would have in our own lab meetings, helps create a community.

https://slack.com/intl/en-gb/

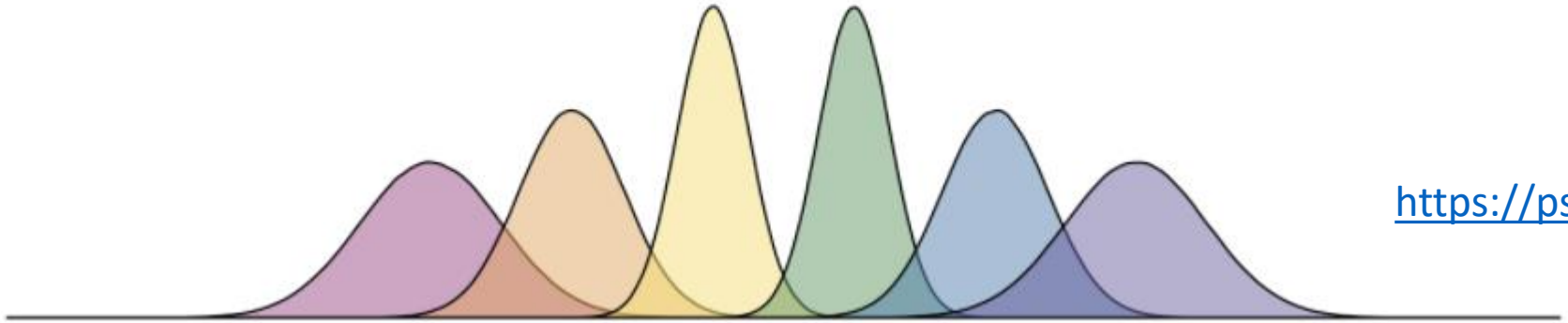
https://www.microsoft.com/en-gb/microsoft-teams/group-chat-software

- Student-led channels where both students, and staff, can help out other students.
- Staff don’t give answers but refer students back to previous assignments and previous feedback, just like we would if we were stuck
- Staff-only channels to help transition
- Acknowledge that not all students, or all staff, are from the same mould and different people work at different speeds and need different support

- We employ our Postgraduates as Graduate Teaching Assistants in our labs to help both students and staff and to serve as a link between the two
  - There is no need to be an expert
  - Read through the code (the tidyverse helps here)
  - Don’t touch the keyboard
  - You are not learning R. You are learning skills using R.

Hack your data beautiful

Coding to creatively and credibly communicate complex ideas



https://psyteachr.github.io/hack-your-data/

fancy.R

Source on Save | Run

```
install.packages("super_fancy_R_club")
install.packages("I_love_pirates")

library("everyone_needs_R")

loveR <- read_csv("tell_me_what_I_want_what_I_really_really_want.csv")
weirdpeople <- read_csv("Ill_tell_you_what_you_want_what_you_really_really_want.csv")
madskillz <- read_csv("madskillz.csv")

why_though <- becauseitscool(madskillz, there, you, go)
and_how <- loveR %>% fancy(contact, GURU) %>%
  superfancy(come, to, GURU %in% BOB517) %>%
  beamazed(learn, the, cool, stuff)

# scared? coding is weird, but R is super accessible and pretty quickly anyone can make the best
# wrangle the hell out of your data

advantages <- madskillz %>%
  CV(knowingRisareallycoolpointonyourCV) %>%
  useful(inaworldwithevergrowingdata,
         itwouldbenaivetothinkthatyouwonthavetodealwithit = sorry) %>%
  friends(youwillgothroughsomeintenseemotions,
          butwewillbetheretosupportyou) %>%
  patience(no.way = willallfrustrationsnotextendyourpatience)
```

@GuruGlasgow

Final Thoughts

- Changing a curriculum is challenging to say the least
  - Know the rationale & explain that clearly to your students
  - Students want to learn and be challenged, and have chosen your course for a reason - most likely interest in the topic. Make the most of that to teach them good practice in your field.
- There is a learning curve for both staff and students so show the benefits to both parties, and never underestimate the importance of community
  - Increased job opportunities and desirable graduate skillset
  - Student independence creates staff time and allows for better projects
- Not everyone is in the position to make wholesale changes
  - Find the one small thing you can change that will benefit your students and change that.

Student Perspective

Student 1: Knowing about reproducibility and issues in reproducibility early on in the undergrad has helped me to be more critical about the papers that I read, but also helped me in designing my own studies. You know the pitfalls, issues that prevent reproducibility, so you know the things to look out for.

Student 2: Learning the skills from level one has given us the tools to evaluate research when writing essays. It has allowed us to take more of a critical look at a paper's methods and therefore question its results. An important skill for doing well in coursework….. Although challenging at first, perseverance with the principles and tools of open science and reproducibility will open so many opportunities during your undergraduate and beyond. It's like a muscle group. The more you work at it, the stronger it gets and you'll see the results.

Conclusion

- Identify the knowledge and skills you want your students to have
- Identify the tools and assignments that allow you to teach those aspects
- Integrate into teaching, research, and admin
- Everyone benefits from teaching based around reproducible methods as the skillset goes beyond the classroom.

@mcaleerp & @uofgpsychology
philip.mcaleer@glasgow.ac.uk

#PSYTEACHREPRODUCIBLE RESEARCH

Community

These changes have truly been a team effort between staff and students and every single one of them should be thanked for helping to improve the teaching of research methods in our School.

Special mention to:

Dale Barr, Lisa DeBruine, Emily Nordmann, Helena Paterson, Niamh Stack, & Heather Cleland Woods. Find them all on twitter as they are all awesome!

@mcaleerp & @uofgpsychology
philip.mcaleer@glasgow.ac.uk

#PSYTEACHREPRODUCIBLE RESEARCH