REPRODUCIBLE PSYCHOLOGY: AN OVERVIEW

Tal Yarkoni

with Richard Morey, Jon Grahe, Eshin Jolly, & JB Poline

PLAN

- Workshop format
  - Best practices (~1.5 hours)
  - Demos (~1.5 hours)
  - Open discussion and/or hands-on session (~1 hour)
- Please ask lots of questions!

REQUIREMENTS

- There aren't really any
- Ideally, this would all be hands-on
  - But, logistically, that comes at a big cost
- Some demos (the OSF) will be web-based
- For others, you may be able to follow along interactively if:
  - You installed the Anaconda distribution (Jupyter demo)
  - You have git installed on your machine (git/GitHub demo)

OVERVIEW

- Definitions
- Motivation
- Good hygiene
- Interactive notebooks
- Version control
- Sharing
- Further resources

I. DEFINITIONS

- What do we mean by reproducibility?
- Well, what do *you* mean?

REPRODUCIBILITY, *N*.

- For our purposes: the ability to duplicate/repeat the same *procedures*
- Distinct from replicability!
  - Replicability: the ability to obtain the same result
  - Note: in some usage, these terms are reversed
  - "**Reproducibility** is the ability of an entire analysis of an experiment or study to be duplicated, either by the same researcher or by someone else working independently, whereas reproducing an experiment is called **replicating** it.[1] Reproducibility and replicability together are among the main principles of the scientific method." —Wikipedia
- But it gets even more complicated…
  - http://languagelog.ldc.upenn.edu/nll/?p=21956

REPRODUCIBILITY VS. REPLICABILITY

- Reproducibility is necessary but not sufficient for replicability
- Much of what makes research reproducible is domain-general
- Replicability is harder, and heavily field-dependent
- We'll focus almost exclusively on reproducibility

CAVEATS

- A lot of this is "do as I say, not as I do"
- Not always feasible to do *everything* covered here
- Don't let the perfect be the enemy of the good

II. MOTIVATION

- Why does reproducibility matter?
- One answer: how could it not?
  - Wikipedia defines science as "a systematic enterprise that builds and organizes knowledge in the form of testable explanations and predictions about the universe."
- How can we *systematically* learn anything about the world if our experiments are irreproducible?

BENEFITS OF REPRODUCIBILITY

- In the modern context, there are other reasons to do science in a reproducible way:
  - Makes your own life easier in many ways
  - Sends a signal that others can trust your work
  - Prevents catastrophic (or non-catastrophic) failures

III. DATA HYGIENE

@jj_bauer

GOOD HYGIENE IS ESSENTIAL

- The best data in the world is no use to anyone if it's incomprehensible
  - E.g., http://okfnlabs.org/bad-data/
- A set of practices that will make your data much more maintainable and reusable

PRESERVE RAW DATA

- There is almost never a reason to edit/modify an original dataset in-place
- Need to clean/sanitize/improve on a dataset?
  - Save a copy!
- Better: create derivative datasets on-the-fly
  - More on this later…

USE CLEAR NAMES

- Ideally, variable and file names should be self-explanatory
- In practice, clarity and brevity are in tension
- Pick a naming convention and stick with it
  - E.g., lowerCamelCase, separated_by_underscores, etc.
- Avoid special characters and whitespace

DOCUMENT YOUR DATA

- Variables
  - Despite the previous slide, your variables are probably *not* self-explanatory
  - Even if the names are, the values are not
  - A codebook is your friend!
    - More detail is better!
    - E.g., https://osf.io/yw3pq/
    - E.g., https://osf.io/4e2ex/

Cookbook for a Codebook

Kai T. Horstmann

20 2 2017

Introduction

For any research project, codebooks have to exist. A codebook is a table that contains all the information someone would need to start analyzing the corresponding data frame. That being said: For each data frame, there has to be a codebook.

Codebooks come in many simple and annoying forms, such as "use variable v1 to v15 to create the scale of extraversion" or "every fifth variable consitutes an individual scale." This documentation is about how (and why) making codebooks easier can save your time, nerves, and even friends.

A codebook is of course independent from the platform on which the data are collected. However, formr.org provides a nice integration a codebook and data collection platform. This means that while creating a study (survey) in formr, you will already have created 95% of your codebook. Furthermore, using formr ensures that the codebook can be used to analyse the data later on.

We also provide some functions throughout this document that will help you turning the codebook + data into whatever you need.

> The general idea is that `function(codebook, data) = results`.

https://osf.io/72hrh/

DOCUMENT YOUR DATA

- Preprocessing steps
  - Often, you'll share only derived data
  - It's still important to document how the derivatives were generated from the original/raw data
  - How did you compute 'age_adjusted_stroop_score'?
  - How did you aggregate trial-level data into subject-wise RT values?

USE SENSIBLE FILE FORMATS

- It's tempting to just hit "Save"
- But some file formats are proprietary/costly to use
  - .mex, .sav, .psd, .xls, etc.
  - Only use these if there's no open/free equivalent
  - Don't assume everyone else uses the same software

BACK UP YOUR STUFF!

- "I lost the data in a hard drive crash"
- Come on—it's 2017!

- Your data should be *automatically* backed up
- Ideally, in the cloud: Dropbox, Google Drive, etc.
- It costs ~$3/month! How can you afford not to?

PROJECT HYGIENE

- What goes for datasets extends to entire projects

BE CONSISTENT

- Use consistent template(s) for your projects (e.g., data/, code/, figures/, analyses/, etc.)
- You can set up a template and copy it over
- In many languages, there are prebaked solutions
  - R: Project template (http://projecttemplate.net/)
  - Python: https://github.com/uwescience/shablona

Organization of the project

The project has the following structure:

```
shablona/
  |- README.md
  |- shablona/
     |- __init__.py
     |- shablona.py
     |- due.py
     |- data/
        |- ...
     |- tests/
        |- ...
|- doc/
   |- Makefile
   |- conf.py
   |- sphinxext/
      |- ...
   |- _static/
      |- ...
|- setup.py
|- .travis.yml
|- appveyor.yml
|- LICENSE
|- ipynb/
   |- ...
```

https://github.com/uwescience/shablona

README

- Ever been pointed to a project and thought, "I don't know where to begin?"
- Put a README in the project root
  - Plain text, but can also use markup (e.g., Markdown, RST)
  - Can extend the same principle to every folder
- Many online repositories will display this automatically

DOCUMENT AS YOU GO

- Keep track of what you do as you do it
  - Log analyses, edits, file modifications, etc.
- The gold standard: make your entire workflow a document (more on this later)

Psychological Assessment

1040-3590/14/$12.00 http://dx.doi.org/10.1037/pas0000032

Using a Genetic Algorithm to Abbreviate the Psychopathic Personality Inventory–Revised (PPI-R)

Hedwig Eisenbarth
University of Colorado Boulder
and University of Regensburg

Scott O. Lilienfeld
Emory University

Tal Yarkoni
University of Texas at Austin

Some self-report measures of personality and personality disorders, including the widely used Psychopathic Personality Inventory–Revised (PPI-R), are lengthy and time-intensive. In recent work, we introduced an automated genetic algorithm (GA)-based method for abbreviating psychometric measures. In Study 1, we used this approach to generate a short (40-item) version of the PPI-R using 3 large-*N* German student samples (total $N = 1{,}590$). The abbreviated measure displayed high convergent correlations with the original PPI-R, and outperformed an alternative measure constructed using a conventional approach. Study 2 tested the convergent and discriminant validity of this short version in a fourth student sample ($N = 206$) using sensation-seeking and sensitivity to reward and punishment scales, again demonstrating similar convergent and discriminant validity for the PPI-R-40 compared with the full version. In a fifth community sample of North American participants acquired using Amazon Mechanical Turk, the PPI-R-40 showed similarly high convergent correlations, demonstrating stability across language, culture, and data-collection method. Taken together, these studies suggest that the GA approach is a viable method for abbreviating measures of psychopathy, and perhaps personality measures in general.

Keywords: psychopathy, genetic algorithm, abbreviation, personality

WORKFLOW AS DOCUMENT

- All* data, code, and materials needed to generate all analyses/figures in the paper are publicly available
  - https://github.com/tyarkoni/precis/
- Many other examples here: https://github.com/jupyter/jupyter/wiki/A-gallery-of-interesting-Jupyter-Notebooks

DOCUMENT AS YOU GO

- The gold standard is not always feasible
  - That's fine! Do what you can!
- E.g., keep a running analysis log in a plaintext file
  - As much detail as you can practically manage
- At bare minimum, keep a "change log" briefly describing all major changes to the project

COMMENTING

- Good code is well-commented code
- Comment your code as if you were writing it for someone else—because you are
- What will be unclear to you six months from now?
- Useful no matter what language or environment you work in
  - Even if you construct your analyses in a GUI!

SPSS SYNTAX EXAMPLE

- SPPS syntax courtesy of Karen Adolphs (via Rick Gilmore)

LICENSE YOUR WORK

- You own the copyright on IP as soon as it's created
- Any use by others (except "fair use") requires a license
  - Note: data doesn't play by the same rules!
- Be explicit about what you allow or don't allow
  - Include a license file in your project
- There are many online resources to help you choose
  - E.g., https://creativecommons.org/licenses/

Choose a license

This chooser helps you determine which Creative Commons License is right for you in a few easy steps. If you are new to Creative Commons, you may also want to read Licensing Considerations before you get started.

Get Started

https://creativecommons.org/share-your-work/

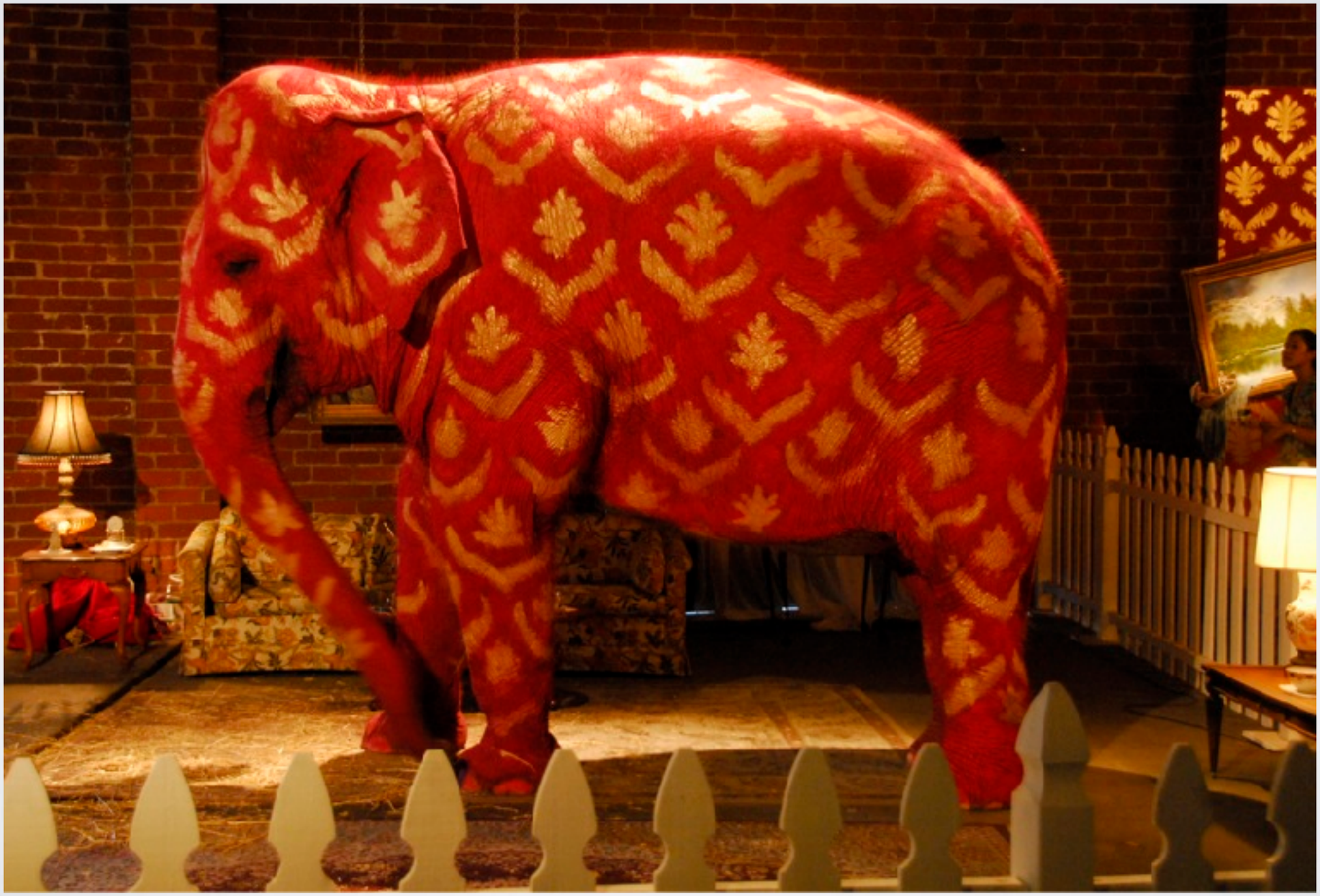
BE PERMISSIVE

- It's tempting to place restrictions on usage
  - E.g., many people want non-commercial clauses
- In general, science operates best with permissive licenses

IV. INTERACTIVE NOTEBOOKS

- A modern version of the lab notebook
- Integrates text, analysis code, figures, results, etc.
- The most widely used platform is Jupyter (jupyter.org)
- The goal: a fully automated workflow

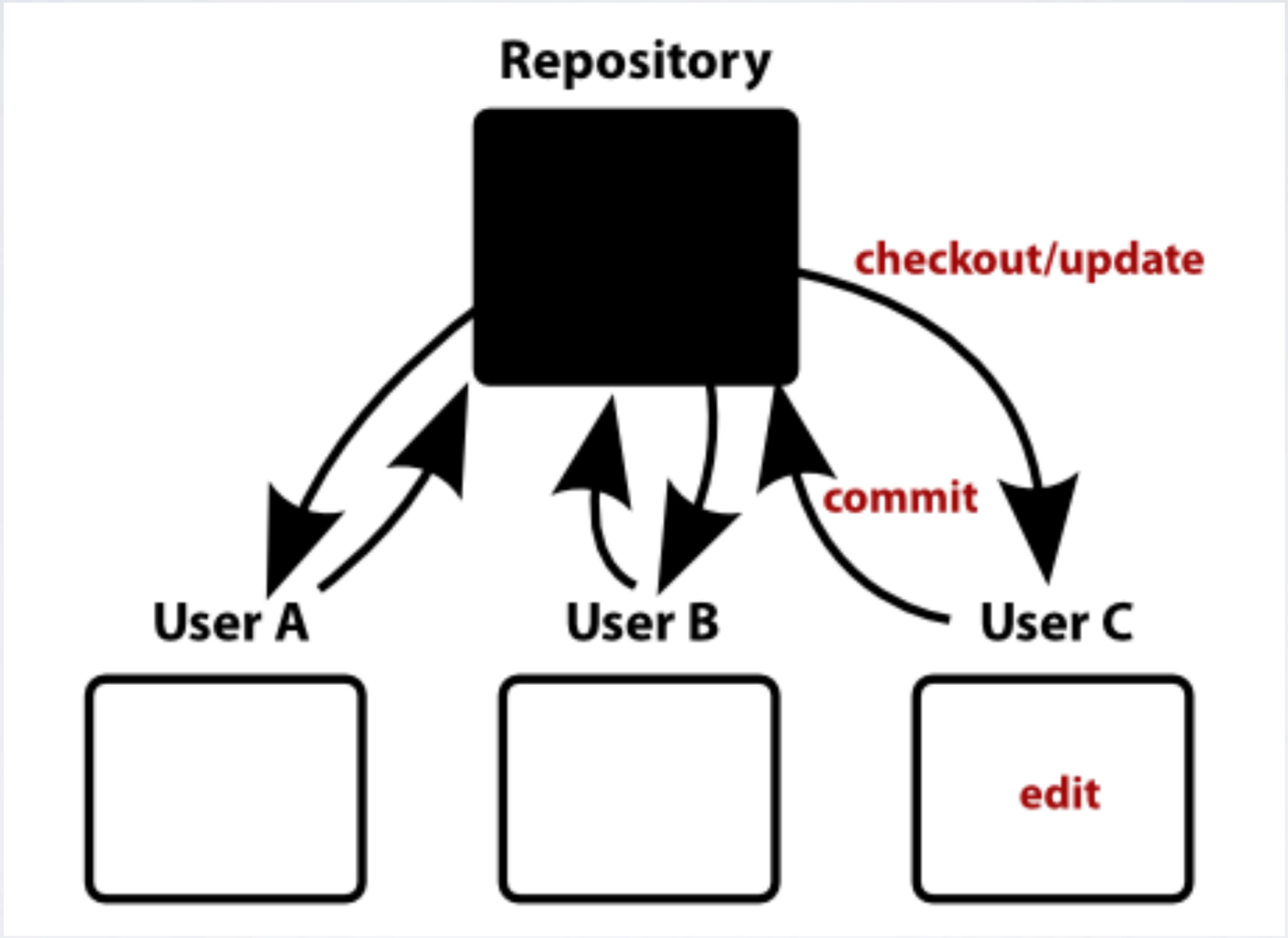
THE ELEPHANT IN THE ROOM



THE ELEPHANT IN THE ROOM

- To do *fully* reproducible science, it's almost essential to be able to program to some degree
  - You do *not* need to be a software developer
- "Programming is the new statistics"
  - Maybe even *more* important than statistics for modern psychology

V.VERSION CONTROL



VERSION CONTROL

- Keeping track of changes is essential in science
- Different versions of datasets, materials, analysis code, figures, etc.
- Bad:
  - manuscript_draft_final_final_final_V4.doc
- Better:
  - Restore specific file version via Dropbox, Google Drive, etc.
- Best:
  - Any platform that supports systematic and comprehensive tracking and updating of files in a project

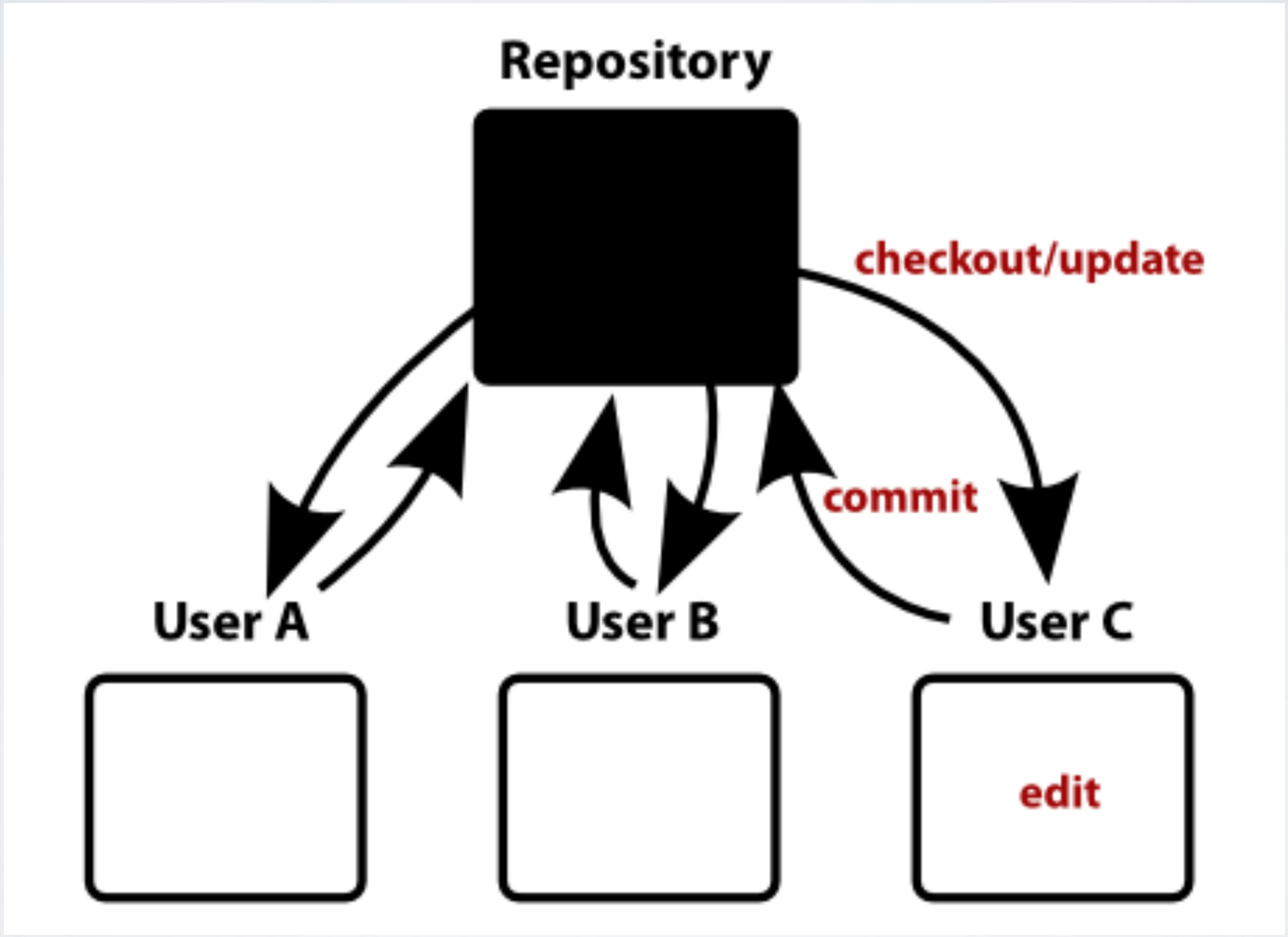
WHY VERSION CONTROL?

- Allows multiple people to collaborate without friction
- Can revert project to any moment in history
- Allows multiple versions to coexist
- Full trace of all changes to every file

GIT

- Git: the most widely used version control platform
- Available on all platforms (https://git-scm.com/)
- Distributed: you can "clone" any repository
- Has a relatively steep learning curve, but lots of resources!
  - https://try.github.io
  - http://rogerdudler.github.io/git-guide/
  - Cheat sheet: https://services.github.com/on-demand/downloads/github-git-cheat-sheet.pdf

DISTRIBUTED VERSION CONTROL



- GitHub.com: a public repository built on git
- Benefits:
  - Provides a more intuitive web interface
  - Adds useful project management tools
  - Makes collaborative work, credit assignment & tracking much easier
- Dominates the open-source software space
  - But it isn't just for code! Supports all kinds of files
  - Scientists can reap almost all of the same benefits


KEEP
CALM
AND
SHARE
STUFF

SHARE YOUR STUFF

- Give away the products of your labor
  - Data, results, code, etc.
- Why?
  - Methods sections are rarely sufficient for full reproducibility
  - More citations (Piwowar, Day, & Fridsma, 2007; Piwowar & Vision, 2013)
  - Greater visibility and influence
  - Sends positive signal to others
  - Less work required to respond to requests

OPEN ACCESS Freely available online

PLoS one

Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results

Jelte M. Wicherts*, Marjan Bakker, Dylan Molenaar

Psychology Department, Faculty of Social and Behavioral Sciences, University of Amsterdam, Amsterdam, The Netherlands

Abstract

Background: The widespread reluctance to share published research data is often hypothesized to be due to the authors' fear that reanalysis may expose errors in their work or may produce conclusions that contradict their own. However, these hypotheses have not previously been studied systematically.

Methods and Findings: We related the reluctance to share research data for reanalysis to 1148 statistically significant results reported in 49 papers published in two major psychology journals. We found the reluctance to share data to be associated with weaker evidence (against the null hypothesis of no effect) and a higher prevalence of apparent errors in the reporting of statistical results. The unwillingness to share data was particularly clear when reporting errors had a bearing on statistical significance.

Conclusions: Our findings on the basis of psychological papers suggest that statistical results are particularly hard to verify when reanalysis is more likely to lead to contrasting conclusions. This highlights the importance of establishing mandatory data archiving policies.

Citation: Wicherts JM, Bakker M, Molenaar D (2011) Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. PLoS ONE 6(11): e26828. doi:10.1371/journal.pone.0026828

Editor: Rochelle E. Tractenberg, Georgetown University Medical Center, United States of America

Received May 20, 2011; **Accepted** October 4, 2011; **Published** November 2, 2011

Funding: The preparation of this article was supported by three grants (021-001-124, 451-07-016, and 400-08-214) from the Netherlands Organization for Scientific Research (NWO). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: j.m.wicherts@uva.nl

- ☐ I have read the Instructions to Authors contained on the Journal of Applied Psychology's website (www.apa.org/pubs/journals/apl).
- ☐ I have included a cover letter in the cover letter text box that contains the names and contact information of all authors. Inclusion of such information indicates that these individuals have agreed to be an author.
- ☐ I attest that this manuscript is not being considered by another journal nor has it been published elsewhere *and* that the data on which the manuscript is based are original.
- ☐ Or: If the manuscript has been previously published in conference proceedings I have explained this in my cover letter.
- ☐ I attest that the data on which this manuscript is based are original data and have not been previously published. I further attest that the data were not simultaneously collected from the same sample along with other data (e.g., data that are parsed, collected at different time points etc.) that have been published or that I intend to subsequently submit for publication.
- ☐ Or: Multiple uses of data collected from the same sample (e.g., data that overlap, data that are parsed or collected at different time points, etc.) have the potential to constitute "duplicate publication." If applicable, I have provided information in my cover letter about related manuscripts that have been, are, or will be submitted for consideration to the same or another journal. I have also edited the manuscript to be transparent with respect to prior, current, and / or any additional uses of the data set. In order to maintain masked review, I have incorporated a table to provide data transparency, following the sample format provided.
- ☐ I verify that if this manuscript describes a study wherein humans were participants, the treatment of those participants was in accordance with established ethical guidelines and appropriate institutional approval has been obtained.
- ☐ I have followed instructions to adequately conceal the identity of all authors in the manuscript itself to ensure masked review.
- ☐ This manuscript has been prepared to be consistent with the 6th edition of the APA Publication Manual guidelines.
- ☐ I have obtained permission to reproduce or adapt any copyrighted material from other sources and am able to provide documentation for this permission.
- ☐ I have disclosed any information in the cover letter about any interests or activities for myself or my coauthors that might be seen as influencing the research (e.g., financial interests in a test or procedure, funding by a 3rd party for research).
- ☐ If my manuscript submission includes supplemental materials, I have explained the relevance of these materials in the cover letter.
- ☐ I agree to comply with APA Ethics Code Standard 8.14a, Sharing Research Data for Verification, allowing other qualified professionals to confirm the analyses and results should my manuscript be accepted for publication. As suggested by APA guidelines, I will retain the raw data for a minimum of 5 years after the publication of this research.
- ☐ If my manuscript is accepted for publication, I agree to transfer copyright to APA according to the guidelines on the APA Publication Rights Form: http://www.apa.org/pubs/authors/publication-rights-form.pdf.

"I CAN'T SHARE THIS…"

- People give many reasons for not proactively sharing their data/materials/code/etc.
- Good reasons not to share:
  - It would violate subjects' privacy or confidentiality
  - It would break the law

"I CAN'T SHARE THIS…"

- Bad reasons not to share:
  - "I don't know how."
    - After this workshop, you do!
  - "It's hard."
    - True, but so are many other things in science that you still do
  - "I could get scooped."
    - You should be so lucky!
  - "People will laugh at me."
    - Possible, but it's much more likely that they'll thank you

GOOD NEWS

- If you do all the other stuff we've talked about, sharing is usually a no-brainer
  - "git push"

OTHER TOPICS

- If time allows, we can talk about:
  - Automated testing
  - Environments
  - Containers
  - Editors and IDEs
  - Field-specific specifications and standards
  - Googling for technical help
  - Programming basics
  - Comparison of programming languages

VI. WHERE TO NEXT?

- Feeling overwhelmed?
  - That's totally normal!
  - This is a continuous, long-term skill-building process
- This is a great time to learn reproducible science
  - The tools are great, and getting better rapidly

RESOURCES

- Google and Stack Overflow
- OSF.io
- Code Academy (https://www.codecademy.com/)
- Software Carpentry and Data Carpentry
  - Try to organize a workshop at your institution!