PSY 525.001 Spring 2020 Syllabus

Transparent, Open, and Reproducible Research Practices in the Social and Behavioral Sciences

Sidney Harris, New Yorker Magazine

Instructor

Rick O. Gilmore, Ph.D. Professor of Psychology 114 Moore Building

+1 (814) 865-3664 rogilmore AT psu DOT edu Schedule an appointment

http://www.personal.psu.edu/rog1 http://gilmore-lab.github.io http://databrary.org

Meeting Location and Time

Tues 09:05-11:50 pm, 350 Moore January 6 - May 1, 2020 course 17294

Resources

- Awesome, free-to-you, data science tutorials from .
- Psychology Department page and discussion group about data science and reproducibility.

About the course

Is there a crisis of reproducibility in psychological science? What does it mean to ask the question? What are transparent, open, and reproducible research practices? Should one implement them? How? This course will seek answers to these questions. Students will read about recent failures in research transparency and reproducibility and discuss the ethics of open practices in scientific research. Through a series of guided exercises students will learn how to use new research tools (e.g. RStudio, Jupyter/iPython, the Open Science Framework, GitHub, command line scripting) that make it easier to implement open and reproducible research practices. At the end of the course, students will be capable of implementing one or more new research practices into their own workflows. Evaluation will be based on in-class participation, short papers, and project assignments. No prior software development experience is required, but a willingness to learn new skills is essential.

Schedule

Week 1: Tuesday, January 14, 2020

Topics

- Introduction to the course
- Needs/interests assessment, and goal setting
- Discussion: Why trust science?
- Setting up computational environments
    - On your local machine
    - In the cloud

Homework

Due by the start of class on 2020-01-21.

1. In a paragraph or two, discuss whether you think researchers in your field *do* and *should* embrace "Mertonian norms." Why or why not?
2. In a paragraph or two, describe your current knowledge of computer programming languages, and at least three learning goals you have for building upon that base.
3. Create a GitHub account, send me your account name in an email, and see if you can create your own copy (fork or clone) of the course repository. If you succeed, take a screen shot of the repository in your local GitHub account.
4. Write these items up in a Word (.docx) file, and email it to me. Make sure to use a sensible file name, e.g., `psy525-YOUR_LAST_NAME-2020-01-21.docx`.

Week 2: Tuesday, January 21, 2020

Topics

- The values of science
- Cases of scientific misconduct
- What is reproducibility? Are we in a crisis?

Readings/webinars

- The values of science (read 1; skim the others)
    - Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia I: Opening scientific communication. *Psychological Inquiry*, *23*(3), 217–243. Retrieved May 9, 2015, from http://dx.doi.org/10.1080/1047840X.2012.692215
    - Kim, S. Y., & Kim, Y. (2018). The ethos of science and its correlates: An empirical analysis of scientists' endorsement of Mertonian norms. *Science, Technology and Society*, 23(1), 1–24. SAGE Publications India. Retrieved from https://doi.org/10.1177/0971721817744438
    - Brakewood, B., & Poldrack, R. A. (2013). The ethics of secondary data analysis: Considering the application of Belmont principles to the sharing of neuroimaging data. *NeuroImage*, *82*, 671–676. Retrieved from http://dx.doi.org/10.1016/j.neuroimage.2013.02.040
- Cases of scientific misconduct (read 1; skim the other)
    - Hauser

– Stapel

- Reproducibility (skim)

    – Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

- Supplemental (not required)

    – Heesen, R., & Bright, L. K. (2019). Is peer review a good idea? *The British Journal for the Philosophy of Science*, *40*. eprints.lse.ac.uk. Retrieved January 7, 2020, from http://eprints.lse.ac.uk/101242/
    – Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, 8(341), 341ps12–341ps12. https://doi.org/10.1126/scitranslmed.aaf5027
    – http://www.stats.org/what-do-we-mean-by-reproducibility/

Homework

Due by the start of class on 2020-01-28.

1. Choose one of the verification bias examples from *Flawed Science*. In a paragraph or two, propose ways you might avoid this sort of bias in your own research.
2. Choose one of the TOP guideline categories where either your own research practices have room to improve or you are doing rather well. In a paragraph or two, explain your reasoning.

Week 3: Tuesday, January 28, 2020

Topics

- Evaluating Munafò et al.'s (2017) "Manifesto for reproducible science", https://doi.org/10.1038/s41562-016-0021
- Workflows and *methods* reproducibility
- Tidy data

Readings/webinars

- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, N. P. du, ... Ioannidis, J. P. A. (2017). A manifesto for reproducible science. Nature Human Behaviour, 1, 0021. https://doi.org/10.1038/s41562-016-0021. (**Required**)
- Reproducible workflows (**Optional**)

    – http://datasci.kitzes.com/lessons/python/reproducible_workflow.html
    – http://grunwaldlab.github.io/Reproducible-science-in-R/index.html

- Tidy data (**Optional**)

    – Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59(10). http://dx.doi.org/10.18637/jss.v059.i10.

- File naming (**Optional**)

    – http://guides.nyu.edu/data_management/file_org
    – http://kbroman.org/dataorg/

- Spreadsheets (**Optional**)

- https://github.com/jennybc/sanesheets
- http://www.datacarpentry.org/2015-03-09-ISI-CODATA/lessons/excel/ecology-examples/00-intro.html
- https://github.com/jennybc/scary-excel-stories

Homework

Due by the start of class on 2020-01-28.

1. ~~Pick one of the recommended elements from Table 1 in Munafò, et al. (2017). Evaluate the recommendation. Do you agree that it would mitigate one or more threats to reproducibility. Why or why not? Do you agree with the assessment about the degree to which stakeholders have adopted the recommended practice?.~~ We covered most of this material in class.
2. Edit/create a text-based workflow for an active project you are working on.

   - Annotate the workflow to indicate where it could be made more reproducible, transparent.
   - If you are feeling super-ambitious, you may want to try creating a graph-based workflow using the `diagrammeR` package.

Week 4: Tuesday, February 4, 2020

Topics

- Pregistration and registered reports
- Introduction to RStudio and R Markdown

Readings/webinars

- Pre-registration and registered reports

  - Chambers, C., Munafò, M., & signatories, more than 80. (2013, June 5). Trust in science would be improved by study pre-registration. *The Guardian.* Retrieved from https://www.theguardian.com/science/blog/2013/jun/05/trust-in-science-study-pre-registration
  - Registered Reports. (n.d.). Retrieved January 24, 2017, from https://cos.io/rr/?_ga=1.163722943.1251838540.1458403228
  - Mathot, S. (2013, March 26). The pros and cons of pre-registration in fundamental research. Retrieved January 24, 2017 from http://www.cogsci.nl/blog/miscellaneous/215-the-pros-and-cons-of-pre-registration-in-fundamental-research
  - (Optional) Frank, M. (2016, July 22). Preregister everything. http://babieslearninglanguage.blogspot.com/2016/07/preregister-everything.html
  - (Optional) Claesen, A., Gomes, S. L. B. T., Tuerlinckx, F., & Vanpaemel, W. (2019, May). Preregistration: Comparing Dream to Reality. Retrieved from psyarxiv.com/d8wex.

- R Markdown exercise

  - Getting started with R Markdown

- Optional

  - Authoring R Markdown Reports. 3 hrs.
  - Working with the RStudio IDE (Part II). 3 hrs.
  - Working with the RStudio IDE (Part II)

Homework

Due by the start of class on 2020-02-11.

1. Find a preregistration document for a study relevant to your research interests on `aspredicted.org` or `osf.io`. In a few paragraphs, comment on what was and what was not included. Would the preregistration provide researchers sufficient structure to carry out the research without 'HARKing'? What downsides do you see to preregistration?
2. Create a template for a reproducible research report or a pregistration document in R Markdown.
3. Convert some portion of your workflow from last week's assignment into an R Markdown.
4. Make sure to use some R Markdown features you haven't used before. Flag the features you used and explain how they'd be useful in an actual use case you can envision.
5. Either email me the R Markdown file, or if you've pushed your file to a GitHub repo, send me a review request.

Week 5: Tuesday, February 11, 2020

Topics

- Version control
- git
    - GitHub
    - GitHub pages

Readings/webinars

- Hardwicke, T. E., & Ioannidis, J. P. A. (2018). Mapping the universe of registered reports. *Nature Human Behaviour*, *2*(11), 793–796. nature.com. Retrieved from http://dx.doi.org/10.1038/s41562-018-0444-y
- Claesen, A., Gomes, S. L. B. T., Tuerlinckx, F., & Vanpaemel, W. (2019, May). Preregistration: Comparing dream to reality. Retrieved from https://psyarxiv.com/d8wex.
- GitHub and RStudio
- Jenny Bryan's http://happygitwithr.com/

Homework

Due by the start of class on 2020-02-18.

1. Create GitHub repo for the project you completed last week
    - Open an issue flagging `@rogilmore` so I know to look at your repo and document.
2. Create a repo for your final course project
    - Create an R Markdown document where you start to outline the possible directions that your final project might take.
    - Open an issue so I can take a look.
3. Clone a repo; fix/change something; make a pull request.
    - Option 1: http://psu-psychology.github.io/psy-525-reproducible-research-2020/
        - Suggestion: Add something about yourself to `students.html` by editing `students.Rmd` and then rebuilding the site via `rmarkdown::render_site(encoding = "utf8")`
    - Option 2: https://psu-psychology.github.io/data-science-and-reproducibility/
        - Suggestion: Add or edit `resources.html` by editing `resources.Rmd`.

Week 6: Tuesday, February 18, 2020

Topics

- Simulation as a tool for reproducible and transparent science
- Visualization tools in R

Readings/webinars

- Required. Data Visualization with ggplot2 (Part 1)
- Desirable. Data Visualization with ggplot2 (Part 2)

Homework

Due by the start of class on 2020-02-25.

1. Create your own simulated data set for a real or proposed study.

   - You may adapt or build upon the examples used in class.
   - Put the results in an R Markdown (.Rmd) file.
   - Commit the .Rmd file to your private repo on GitHub and either raise an Issue on GitHub or submit a pull request.

2. Plot the results of your simulation using ggplot2 commands.
3. Make sure that your simulation has the following sub-sections:

   - Introduction/Motivation
     - What are you simulating and why? Where do the parameter estimates come from? The literature or your best guess?
   - Plots
   - Statistical Analyses
   - Discussion/Conclusions
     - What did you discover or demonstrate?

- Please label the r chunks in your R Markdown files

Week 7: Tuesday, February 25, 2020

Topics

- Doing other useful things with R and R Markdown

Readings/webinars

Homework

Due by the start of class on 2020-03-03.

1. Create a set of HTML talk slides using R Markdown.
2. Try rendering the slides as a Word document, PowerPoint document, or PDF.
3. Create a new repo and generate a simple website using R Markdown.

Week 8: Tuesday, March 3, 2020

Topics

- Interactive visualizations using Shiny apps

Readings/webinars (recommended)

- How to start with Shiny (Part 1)
- How to start with Shiny (Part 2)
- How to start with Shiny (Part 3)

Homework

Due by the start of class on 2020-03-17.

1. Complete a 1-2 page write-up describing your plans for your final project.

Spring Break March 9 - 12, 2020

NO CLASS

Week 9: Tuesday, March 17, 2020

Topics

- Python, the other language of data science
  - Intro to Jupyter
  - Jupyter for research, teaching, talks

Readings/webinars

- Importing data in Python (Part 1)
- Importing data in Python (Part 2)

Homework

Due by the start of class on 2020-03-24.

TBD

Week 10: Tuesday, March 24, 2020

Topics

- Tools for reproducible data-gathering
  - E-Prime, Matlab (Psychophysics Toolbox), PsychoPy.
  - jsPsych
  - MTurk
    - * psiTurk

Readings/webinars

- de Leeuw, J.R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1-12. doi:10.3758/s13428-014-0458-y

Homework

Due by the start of class on 2020-03-31.

1. Open a PsychoPy demo program and save it with a new name. Add additional documentation to the demo program that explains what's happening. Change one or more parameters to make the program do something slightly different, and explain what parameters you changed.

Week 11: Tuesday, March 31, 2020

Topics

- Using APIs
  - U.S. Census
  - Google Drive
  - Box
  - Wikidata
  - Databrary
  - OSF

Homework

Due by the start of class on 2020-04-07.

1. Create a Jupyter notebook where you document your exploration of the U.S. Census, Box, or Google Drive APIs.

Week 12: Tuesday, April 7, 2020

Topics

- Where to share?
- Publishing data
- Challenges to sharing
- Your open science portfolio
- Funder policies

Readings/webinars

- Meyer, M. N. (2018). Practical Tips for Ethical Data Sharing. Advances in Methods and Practices in Psychological Science, 2515245917747656. SAGE Publications Inc. Retrieved from https://doi.org/10.1177/2515245917747656
- Gilmore, R.O. et al. (2020).

Homework

Due by the start of class on 2020-04-07.

1. Choose two outlets for sharing research data or materials and compare and contrast their strengths and weaknesses.

Week 13: Tuesday, April 14, 2020

Topics

- Catch-up week

Week 14: Tuesday, April 21, 2020

Topics

- Preparation for student projects

Week 15: Tuesday, April 28, 2020

Topics

- Student project presentations

Evaluation

PSY 525 course performance will be evaluated based on the following scheme:

Component	Points	% of Grade
Class participation	2 pts/class * 15 weeks = 30	25
Assignments	5 pts * 12 assignments = 60	50
Final projects	30	25
TOTAL	**150**	**100**

Resources

How-to's

- How to connect to Penn State's RStudio Server Pro instance to run RStudio in the cloud.
- How to connect to Penn State's Jupyter notebook server.
- How to use Google's colaboratory Jupyter notebook service.

Web courses

- Data Carpentry, now The Carpentries

  - OpenRefine for Social Science Data by Data Carpentry.
  - R for Reproducible Scientific Analysis
  - Programming with Python
  - Plotting and Programming in Python
  - Version Control with Git

Books and E-Books

- Chang, W. (2012). *R Graphics Cookbook.* http://ase.tufts.edu/bugs/guide/assets/R%20Graphics%20Cookbook.pdf
- Gandrud, C. (2015). *Reproducible Research with R and R Studio, Second Edition.* CRC Press. https://www.crcpress.com/Reproducible-Research-with-R-and-R-Studio-Second-Edition/Gandrud/p/book/9781498715379
- VanderPlas, J. (2016). *Python Data Science Handbook.* https://jakevdp.github.io/PythonDataScienceHandbook/
- Wickham, H. & Grolemund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* O'Reilly Media. http://shop.oreilly.com/product/0636920034407.do.
- Wickham, H. (2019). *Advanced R, 2nd Ed.* Chapman & Hall/CRC. https://adv-r.hadley.nz/.
- Xie, Y., Allaire, J.J., & Grolemund, G. (2019). *R Markdown: The Definitive Guide.* Chapman & Hall/CRC. https://bookdown.org/yihui/rmarkdown/.

Blogs/Projects

- R bloggers https://www.r-bloggers.com/
- Open science in practice. https://inattentionalcoffee.wordpress.com/2017/01/03/open-science-in-practice/
- Practicing what I preach: The open Ph.D. experiment. http://www.librarianinthecity.com/2016/12/practicing-what-i-preach-the-open-phd-experiment/
- Open Stats Lab. Use data published with psychology papers to teach statistics.
- Jupyter, Zeppelin, Beaker: The Rise of the Notebooks. https://opendatascience.com/blog/jupyter-zeppelin-beaker-the-rise-of-the-notebooks/
- Open science and free culture. http://www.simoncolumbus.com/publications-2/open-science-and-free-culture/
- Collaborative Replications and Education Project (CREP).

Software tools

- OpenRefine, https://openrefine.org
- LaTeX on the web

  - ShareLaTex http://sharelatex.com
  - Overleaf http://www.overleaf.com
  - Authorea http://www.authorea.com

- Experimental Design Assistant for planning animal studies https://eda.nc3rs.org.uk/
- p-hacker Shiny app http://shinyapps.org/apps/p-hacker/

Articles

- Schweinsberg, M., Madan, N., Vianello, M., Sommer, S. A., Jordan, J., Tierney, W., . . . Uhlmann, E. L. (2016). The pipeline project: Pre-publication independent replications of a single laboratory's research pipeline. Journal of Experimental Social Psychology, 66, 55–67. https://doi.org/10.1016/j.jesp.2015.10.001
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, N. P. du, . . . Ioannidis, J. P. A. (2017). A manifesto for reproducible science. Nature Human Behaviour, 1, 0021. https://doi.org/10.1038/s41562-016-0021.
- Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., . . . Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. Nature Reviews Neuroscience, advance online publication. https://doi.org/10.1038/nrn.2016.167.

Citation Management

Comparisons

- http://guides.lib.uchicago.edu/c.php?g=297307&p=1984557

- https://en.wikipedia.org/wiki/Comparison_of_reference_management_software

Software

- Mendeley https://www.mendeley.com/careers/?dgcid=google_paid-search_Mendeley.com-Search-Branded-USA&gclid=Cj0KEQiAifvEBRCVx5up6Ojgr5oBEiQALHw1TviV-SG0qOSpmmSkoIbrYrn1ZvPeZTinianXeiC
- Zotero https://www.zotero.org/
- Paperpile https://paperpile.com/