Statistics 159/259: Reproducible and Collaborative Statistical Data Science

Philip B. Stark, Department of Statistics, UC Berkeley

www.stat.berkeley.edu/~stark pbstark@berkeley.edu @philipbstark

Office: 403 Evans Hall. Office hours: TBA

This version: 30 September 2018.

Reproducible and Collaborative Statistical Data Science: Overview

This course teaches reproducible and collaborative research techniques through applied statistics, including reproducing published work and re-analyzing the data in that work using other methods–reproducibly.

Examples will be drawn from a variety of fields, including agriculture, health, and climate.

There will be roughly five small assignments and several larger projects. Much of the work will be collaborative in groups of 4-5. You will be asked to review your own contributions and each others contributions to group projects. There will not be a final exam, but there will be final presentations of group work.

Administrativia

Prerequisites

- Statistics 133, 134, 135
- Willingness to pick up programming languages and software tools independently (tools used will include Python; Jupyter Notebooks; the Python "scientific stack" of numpy, scipy, matplotlib, and perhaps pandas and scikit; git; GitHub; Travis CI; Docker; LaTeX, Markdown, pandoc)

Format and assessment

- 3 hours of lecture and 2 hours of lab per week (bCourses will have screencasts of lectures)

- approximately 5 “small” individual assignments (40% of grade)
- 2 larger projects (30% of grade)
- a group term project (30% of grade)

Office hours

- Mondays 12:15-1:15, 403 Evans Hall

Graduate Student Instructor

- Mitch Negus, mitchell_negus@b.e
- Office hours 10-12 Tuesdays and Thursdays, 444 Evans Hall

Submitting assignments: Submit written assignments by making a pull request to your private repository within the git organization for the class, https://github.berkeley.edu/stat-159-259-f18. Use your CalNet credentials to access your private repository. Create a directory for each assignment labeled with the assignment number, e.g., “Assignment1” for the first assignment.

- Text documents should be written in LaTeX or Markdown. A pdf and the source file should be submitted. Microsoft Word is not acceptable.
- Code and analyses should be in python. All code should have accompanying unit tests. In some cases, Jupyter notebooks will be the appropriate thing to submit; in others (more extensive analyses), a collection of .py files will be more appropriate. For term projects, the “deliverable” will include a repository that includes code, data, analyses, unit tests, and coverage tests.

Code of conduct; attribution of work

The high academic standard at the University of California, Berkeley, is reflected in each degree awarded. Every student is expected to maintain this high standard by ensuring that all academic work reflects unique ideas or properly attributes the ideas to the original sources.

These are some basic expectations of students with regards to academic integrity: Any work submitted should be your own individual thoughts, and should not have been submitted for credit in another course unless you have prior written permission to re-use it in this course from this instructor.

All assignments must use “proper attribution,” meaning that you have identified the original source and extent or words or ideas that you reproduce or use in your assignment. This includes drafts and homework assignments! If you are unclear about expectations, ask your instructor.

Do not collaborate or work with other students on assignments or projects unless the instructor gives you permission or instruction to do so.

Disability accommodations

If you need an accommodation for a disability, if you have information your wish to share with the instructor about a medical emergency, or if you need special arrangements if the building needs to be evacuated, please inform the instructor as soon as possible.

If you are not currently listed with DSP (the Disabled Students' Program) and believe you might benefit from their support, please apply online at dsp.berkeley.edu

Resources

- Computing resources
  - We will be using Jupyter notebooks. You can use a hosted notebook at https://datahub.berkeley.edu/ or install Jupyter on your own device. The datahub.berkeley.edu server will have all the packages you need pre-installed. In contrast, if you use the Anaconda distribution, you will need to install some extra things, such as the permute and cryptorandom packages.
  - We will use the campus github server, github.berkeley.edu
  - The class notes and other materials are available at https://github.berkeley.edu/pbstark/S159-f18
  - Assignments should be submitted by pull request to your private repository within the class organization https://github.berkeley.edu/stat-159-259-f18
- Git and git workflows
  - Introduction to Git. This is based on the notes we used in this class, but has a fair amount of additional explanation and detail you may find useful through the semester.
  - Immersion course
  - git-scm guide
  - Statlab development git workflow
- Continuous integration
  - Travis CI for beginners
  - Continuous integration with Travis by Simon Scholz
- Scientific Python, Jupyter
  - Lecture notes on scientific python
  - Python for scientific computing by Fernando Perez
  - https://hplgit.github.io/primer.html/doc/pub/half/book.pdf
  - Elegant SciPy, Stefan van der Walt. The full book and all the notebooks are available.
  - Getting started with Python for research, a gentle introduction to Python in data-intensive research.
  - An introduction to "Data Science", a collection of Notebooks by BIDS' Stéfan Van der Walt.

    - Effective Computation in Physics, by Kathryn D. Huff; Anthony Scopatz. Notebooks to accompany the book.
    - A Whirlwind Tour of Python, by Jake VanderPlas.
    - Python for Data Analysis, 2nd Edition, by Wes McKinney, creator of Pandas. Companion Notebooks
    - Effective Pandas, a book by Tom Augspurger, core Pandas developer.
- Docker
    - https://docs.docker.com/get-started/
    - https://docker-curriculum.com/
- LaTeX
    - https://www.tug.org/twg/mactex/tutorials/ltxprimer-1.0.pdf
- Markdown
    - https://daringfireball.net/projects/markdown/syntax
    - https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet
    - https://www.markdownguide.org/getting-started/
- Pandoc
    - https://pandoc.org/getting-started.html
    - https://pandoc.org/MANUAL.pdf
- Miscellaneous computing tutorials
    - Berkeley Statistical Computing Facility tutorials

Topics

- What is reproducibility?
    - Why is reproducibility an issue?
    - Terms in different fields. Computational, experimental, ...
    - "Preproducibility"
    - What is contributing to lack of reproducibility in science?
    - Importance of replication
    - "Virtual witnessing"
- Attempting computational reproducibility in data analysis
    - data
        - get the data
            - can be hard even if journal/funder requires making data available
        - figure out what format it's really in
            - data dictionaries sometimes help
            - proprietary formats common
        - figure out whether it's the right data
            - sniff tests
            - consistency tests
            - sleuthing

    * pre-processing, cleaning, etc.
  - analysis
    * figure out what they claim to have done
    * figure out what they did
      · usually impossible from just the methods section
      · much harder if the analysis was not scripted
      · might be impossible even with their code
    * figure out what they should have done
    * compare

- Sciencing is hard
  - confirmation bias
  - easiest person to fool is yourself
  - misperceptions of probability
  - confidence and accuracy unrelated
  - shiny models and methods
  - broken reward structure
  - ritualization
  - Cargo-cult science and statistics

- Papers/Datasets for the term
  - papers where there seemed to be a chance to get the data
  - topics with social impact: food, health,
  - some paper I think are bogus
  - some papers whose conclusions I like–scrutiny to avoid confirmation bias

- Software engineering
  - revision/version control
  - documentation
  - modularity and abstraction
    * consistency: APIs, calling signatures, object-oriented coding
    * separating data, computation, presentation
    * how general is the problem your approach can solve?
      · what's the right level of abstraction?
      · does it solving it require other broadly useful tools?
      · consider other approaches to subproblems?
      · don't re-invent the wheel. . . but understand how wheels work
  - unit tests, integration tests, regression tests, coverage tests
  - code review
  - pair programming
  - scripted analyses
  - automation
  - accountability

- Case study: Karp et al., 2015

- access to data
- reproducing the main results from the data
- regression models
  * assumptions required to perform OLS
  * assumptions required for OLS to be unbiased
  * assumptions required to compute SE
  * assumptions required for $\hat{\beta}/SE$ to have a t-distribution
- interpreting P-values
  * what's the null hypothesis?
  * appropriateness of t-tests in regression
  * p-values from observational data: hypothetical randomness
- permutation tests
  * group invariances and exchangeability
  * the Neyman "ticket" model
    · the strong null hypothesis and weak null hypotheses
  * interference
    · when is non-interference a reasonable assumption?
  * null hypotheses, tests, and test statistics
    · key restrictions
    · significance versus power
    · specific alternatives and omnibus alternatives
    · p-values versus fixed-level tests
  * generating random permutations. See Stark 2017
    · generating pseudo-random numbers and pseudo-random integers
    · LCGs, Mersenne Twister, cryptographic PRNGs
    · shuffling algorithms
    · the cryptorandom library
    · problems with R's algorithms for generating random integers and random samples Ottoboni & Stark
  * confidence bounds for p-values by inverting binomial tests See Permute/utils binom_conf_interval
- from reproducibility to replicability, stability, and generalizability
  * transforming data before regression: "Garden of forking paths." See Gelman & Loken
  * sensitivity of conclusions to transformations
  * sensitivity of conclusions to individual data: "influential observations"
  * testing before modeling and post-selection inference (POSI)
  * why reporting everything you tried matters; pre-registration
    · AllTrials.net
    · changing clinical endpoints. Example: PACE trial for CFS/ME

- Sensitivity analysis and sensitivity auditing
- Statistical models and response schedules

  - Response schedules and "physics." See Freedman SMTP Ch6
  - Linear probability models
  - Logit and probit models
  - Poisson regression

- Bayesian estimation and inference

Assignment 1. Due 9/3, 11:59pm:

Getting started reproducing research

1. Look at the data Morabia transcribed from P.C.A. Louis on bloodletting for pneumonia and read Morabia (2006). What do you think of the fact that data from 1828 are available? Reproduce the results below (which Morabia cites); if you cannot reproduce them, say why:

- 77 patients
- 2 comparison groups of 41 and 36 patients
- comparable average age (41 and 38 years, respectively)
- number of patients bled on the first day, who had passed the age of fifty, was nearly twice as great as that of the patients of the same age, who were bled at a later period
- duration of disease was an average of 3 days shorter in those who had been bled early compared with those who had been bled late
- 'three sevenths' (i.e., 44 %) of the patients who had been bled early died
- 'only one fourth' (i.e., 25 %) of those bled late died

Is Louis's work an observational study or an experiment? Do you think it amounts to a "natural experiment"? Why or why not? Give two scientific questions (*statistical hypotheses*) those data might address. What do you think the most important confounding factors would be, for those two hypotheses? What would be the most natural "as-if" randomization to use in analyzing the data to address the hypotheses you formulated, if you were to consider the study to be a natural experiment? What are the controls? Is the experiment blind? Double-blind? (If you are unfamiliar with the notions of confounding, natural experiments, controls, blinding, etc., see the relevant chapters of SticiGui.)

1. Read Karp et al. (2015) and look at the data they provided in Data Dryad. Which figures and tables in the paper could, in principle, be reproduced from the data they provide? Which cannot? Does the methods section describe how they processed the data in adequate detail to reproduce the analyses? If not, what else would you need to know?

Assignment 2. Due 9/9, 11:59pm:

Terminology: Reproducibility, Replicability, Preproducibility, etc.

Read Barba (2018), Buckheit and Donoho (1995), Rokem et al. (2018), Stark (2018). Explain in your own words different senses of the terms "reproducible," "replicable", and "repeatable." In your own words, explain why these concepts are important for science and society. Do you think there's a reasonable case for introducing a new term, such as "preproducible" (not necessarily that word, but a new term)? Why or why not? What would you propose as a solution to the problem that different disciplines use "reproducible," "replicable," and "repeatable" to mean different things? There's no length restrictions for this assignment, but I would expect it to take about 2 pages to do a good but concise job.

Reading assignment, finish before class on 9/24

Saltelli et al. (2015), van der Sluijs et al. (2005), van der Sluijs et al. (2008), van der Sluijs (2016)

van der Sluijs and Saltelli will give guest lectures the week of 9/24.

Assignment 3. Due 9/30, 11:59pm:

Sensitivity Analysis, Sensitivity Auditing, and Public Policy

Read and Berk and Freedman (2001), van der Sluijs (2016), Urban (2015)

- Urban reports, "Overall, 7.9% of species are predicted to become extinct from climate change; (95% CIs, 6.2 and 9.8) (Fig 1)."
  - Urban derives his estimate using "Bayesian meta-analysis"
    - Explain what meta-analysis is, including the assumptions (see Berk and Freedman)
    - If you can, state the additional assumptions of Bayesian meta-analysis (this might require research)
    - If you can figure it out from the paper and supplemental materials, state the prior Urban uses
  - Urban points out that there are several general approaches the underlying 131 studies use to estimate the number of species that will go extinct. Sketch how these work:
    - Species-area relationships
    - Expert opinion
    - Species distribution models
  - Recall that the taxonomy of life is Kingdom, Phylum, Class, Order, Family, Genus, Species. There are about 1.9 million known species of eukaryotes (everything but bacteria), and it is estimated that there are 8.7 million in all. There are estimated to be from millions to trillions of species of bacteria (prokaryotes).
    - Estimate the number of species included in the 131 studies Urban relied on (this might require research: explain how you get your estimate)

    - Are those species a random sample of all known species? Of all species?
    - What animal genus has the most species?
      - Do any of the studies Urban relies on examine that genus?
    - What plant genus has the most species?
      - Do any of the studies Urban relies on examine that genus?
    - Do any of the studies Urban relies on consider bacterial species?
    - What families does Urban's estimate consider? (do your best to figure this out–I don't expect you to read all 131 studies)
  - What data go into Urban's estimate?
  - What, if anything, is random in Urban's estimate?
  - Is the estimate of 7.9% of species unbiased? Why or why not?
  - Is the range (6.2%, 9.8%) a confidence interval? Why or why not?
  - Urban's estimate does not have a timeline: it's "climate change," not "climate change over the next 50y," for instance. How does that make sense?
  - List 5 potentially large sources of uncertainty that Urban did not consider or did not address adequately
  - On balance, do you think the 7.9% ((6.2%, 9.8%) figures are reliable? Useful? Interpretable?

Assignment 4. Due 9/30, 11:59pm:

Algorithms, unit tests, continuous integration

This assignment concerns the chi-square statistic for the "two-sample problem" for categorical data.

The input is two lists, $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_m)$. Imagine concatenating the lists to get a new list $z = (z_1, \ldots, z_N)$ of length $N = n + m$. Let $\{u_k\}_{k=1}^K$ denote the unique values in z and let π_k denote the relative frequency of the value u_k among the elements of z, that is,

$$\pi_k \equiv \frac{\#\{z_j, j = 1, \ldots, N : z_j = u_k\}}{N}.$$

Let $E_k \equiv n\pi_k$ and let $O_k \equiv \#\{x_j, j = 1, \ldots, n : x_j = u_k\}$ (that is, O_k is the number of elements of x that are equal to u_k). The chi-square statistic for these data for the two-sample problem is

$$\chi^2 = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k}.$$

Write three different python functions that (each) take as input x and y and return χ^2. The functions should use different strategies and/or data structures to calculate χ^2.

Write unit tests for the functions to ensure that they work correctly for arbitrary input lists x and y.

Assignment 5. Due 10/7, 11:59pm:

Cargo-Cult Statistics and "researcher degrees of freedom"

Read Gelman and Loken (2013) and Silberzahn et al. (2018).

- How many of the co-authors in Silberzahn et al. are in Statistics departments?
- The basic question the teams are supposed to answer is "are soccer referees are more likely to give red cards to dark-skin-toned players than to light-skin-toned players?"
  - Explain in your own words what "more likely" means here.
  - Is there anything random going on? If so, what? What makes it random?
  - What would "equally likely" look like in the real world?
  - Is the question about individual referees, or referees in general?
- For teams 1–6 and 12 (seven teams) in Silberzahn et al., explain the following:
  - Does the analysis implicitly or explicitly involve a model? If it does:
    - Describe the model
    - List and explain the assumptions of the model
    - Assess the evidence that the model is a response schedule
    - Describe any goodness-of-fit tests the team used to check the adequacy of the model
    - Explain in words what "OR" means for each of the models.
      - What is OR or what parameter does it estimate?
      - Are the confidence intervals really confidence intervals? Why or why not?

Collected Reading List:

Foundations; Statistical Models

1. Feynman, R., 1974. CalTech Commencement Address, http://calteches.library.caltech.edu/51/2/CargoCult.htm

2. Freedman, D.A., 1995. Some issues in the foundations of statistics, *Foundations of Science*, *1*, 19–39. https://doi.org/10.1007/BF00208723

3. Freedman, D.A., 1999. From association to causation: some remarks on the history of statistics, *Statistical Science*, *14*(3), 243–258.

4. Freedman, D.A., and R. Berk, 2001. Statistical Assumptions as Empirical Commitments, http://escholarship.org/uc/item/0zj8s368#page-1 (also in Freedman, D.A.,

2010. *Statistical Models and Causal Inference: A dialog with the Social Sciences*, Cambridge University Press. D. Collier, J. Sekhon, P.B. Stark, eds.)

5. Freedman, D.A., 2008. On types of scientific inquiry: the role of qualitative reasoning, *The Oxford Handbook of Political Methodology*, Box-Steffensmeier, J.M., H.E. Brady, and D. Collier (eds), Oxford University Press, Oxford. DOI: 10.1093/oxfordhb/9780199286546.003.0012. Preprint

6. Freedman, D.A., 2009. *Statistical Models: Theory and Practice*, 2nd edition, Cambridge University Press.

7. Freedman, D.A., R. Pisani, and R. Purves, 2007. *Statistics, 4th edition*, W.W. Norton, New York.

8. Klemes, V., 1989. The Improbable Probabilities of Extreme Floods and Droughts, in O. Starosolsky and O.M. Meldev (eds), *Hydrology and Disasters*, James and James, London, 43–51. https://www.itia.ntua.gr/en/getfile/1107/1/documents/1997_ImprobProbabilities_OCR.pdf

9. LeCam, L., 1977. Note on metastatistics or 'An essay toward stating a problem in the doctrine of chances,' *Synthese*, *36*, 133-160.

10. Stark, P.B., 1997. SticiGui

11. Stark, P.B., 2016a. Pay no attention to the model behind the curtain

12. Stark, P.B., 2016b. The value of P-values, *The American Statistician*, *70*, DOI:10.1080/00031305.2016.1154108

13. Stark, P.B., 2017. Mathematical Foundations, Inequalities, Introduction to permutation tests, Rabbits and Cargo-Cult Statistics, Generating pseudo-random samples and permutations

14. Stark, P.B., and A. Saltelli, 2018. Cargo-cult Statistics and Scientific Crisis, *Significance*, *15*(4), 40–43. https://www.significancemagazine.com/593

Statistical methodology

1. Freedman, D.A., 2008. Randomization does not justify logistic regression, *Statistical Science*, *23* 237–249. DOI: 10.1214/08-STS262 https://arxiv.org/pdf/0808.3914.pdf

2. Hastie, T., R. Tibshirani, and J. Friedman, 2009. *Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition*, Springer-Verlag, NY. https://web.stanford.edu/~hastie/Papers/ESLII.pdf

3. McCullagh, P. and J.A. Nelder, 1983. *Generalized Linear Models*, 2nd edition, Chapman & Hall, NY.

Evidence, Models, and Public Policy

1. Saltelli, A., P.B. Stark, W. Becker, and P. Stano, 2015. Climate Models as Economic Guides: Scientific Challenge or Quixotic

Quest?, *Issues in Science and Technology*, Spring 2015. Reprint: http://www.stat.berkeley.edu/~stark/Preprints/saltelliEtal15.pdf

2. van der Sluijs, J.P., J.S. Risbey, and J.R. Ravetz, 2005. Uncertainty Assessment of Voc Emissions From Paint in the Netherlands Using the NUSAP System, *Environmental Monitoring and Assessment*, *105*, 229–259. doi:10.1007/s10661-005-3697-7

3. van der Sluijs, J.P., A.C. Petersen, P.H.M. Janssen, J.S. Risbey, and J.R. Ravetz, 2008. Exploring the quality of evidence for complex and contested policy decisions, *Environmental Research Letters*, *3*, doi:10.1088/1748-9326/3/2/024008

4. van der Sluijs, J.P., 2016. Numbers Running Wild, Chapter 5 *in The Rightful Place of Science: Science on the Verge*, A. Benessia, S. Funtowicz, M. Giampietro, Â.G. Pereira, J. Ravetz, A. Saltelli, R. Strand, J.P. van der Sluijs, eds., Consortium for Science, Policy & Outcomes, AZ & DC. http://www.andreasaltelli.eu/file/repository/Science_on_the_Verge_FINAL_.pdf

Foundations: Computation, Optimization

1. Goldberg, D., 1991. What every computer scientist should know about floating-point arithmetic, *ACM Computing Surveys*, *23*, 5–48.

2. Lawson, C.L. and R.J. Hanson, 1974. *Solving Least Squares Problems*, Prentice-Hall, NJ.

3. Ottoboni, K. and P.B. Stark, 2018. Random problems with R, ArXiV, https://arxiv.org/abs/1809.06520. also see https://stat.ethz.ch/pipermail/r-devel/2018-September/076817.html

Agriculture, Ecology, and Health

1. Fagan, J., T. Traavik, and T. Bøhn, 2015. The Seralini affair: degeneration of Science to Re-Science?, *Environmental Sciences Europe*, *27*:19, DOI 10.1186/s12302-015-0049-2

2. Karp, D.S., S. Gennet, C. Kilonzo, M. Partyka, N. Chaumont, E.R. Atwill, and C. Kremen, 2015. Comanaging fresh produce for nature conservation and food safety, *PNAS*, *112* (35) 11126-11131. https://doi.org/10.1073/pnas.1508435112

3. LeCanne, C.E., and J.G. Lundgren, 2018. Regenerative agriculture: merging farming and natural resource conservation profitably, *PeerJ* 6:e4428 https://doi.org/10.7717/peerj.4428

4. Morabia, A., 2006. Pierre-Charles-Alexandre Louis and the evaluation of bloodletting, *J. Roy. Soc. Medicine*, *99*, 158–160. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1383766/pdf/0158.pdf

5. Seralini, G.-E., E. Clair, R. Mesnage, S. Gress, N. Defarge, M. Malatesta, D. Hennequin, and J. Spiroux de Vendômois, 2014. Republished study:

long-term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize, *Environmental Sciences Europe*, *26*:14, http://www.enveurope.com/content/26/1/14

Pedestrians and Race

1. Coughenour, C., S. Clark, A. Singh, E. Claw, J. Abelar, and J. Huebnera, 2017. Examining racial bias as a potential factor in pedestrian crashes, *Accident Analysis and Prevention*, *98*, 96-100. http://dx.doi.org/10.1016/j.aap.2016.09.031

2. Goddard, T., K.B. Kahn, and A. Adkins, 2015. Racial Bias in Driver Yielding Behavior at Crosswalks, *Transportation Research Part F: Traffic Psychology and Behaviour*, *33*, 1-6. http://dx.doi.org/10.1016/j.trf.2015.06.002

Earthquake probabilities

1. USGS 2008 Bay Area Earthquake Probabilities. http://earthquake.usgs.gov/regional/nca/ucerf/

2. Cornell, C.A., 1968. Engineering seismic risk analysis, *Bull. Seism. Soc. Am*, *58*, 1583–1606.

3. Mulargia, F., P.B. Stark, and R.J. Geller, 2017. Why is probabilistic seismic hazard analysis (PSHA) still used? *Physics of the Earth and Planetary Interiors*, *264*, 63-75. https://doi.org/10.1016/j.pepi.2016.12.002

4. Stark, P.B. and D.A. Freedman, 2003. What is the Chance of an Earthquake? in *Earthquake Science and Seismic Risk Reduction*, F. Mulargia and R.J. Geller, eds., NATO Science Series IV: Earth and Environmental Sciences, v. 32, Kluwer, Dordrecht, The Netherlands, 201–213. Preprint: http://www.stat.berkeley.edu/~stark/Preprints/611.pdf

Impact of Climate Change

1. Houser, T., R. Kopp, S. Hsiang, M. Delgado, A. Jina, K. Larsen, M. Mastrandrea, S. Mohan, R. Muir-Wood, D.J. Rasmussen, J. Rising, and P. Wilson, 2014. The American Climate Prospectus: Economic Risks in the United States, 2014. http://rhg.com/reports/climate-prospectus

2. Hsiang, S., R. Kopp, A. Jina, J. Rising, M. Delgado, S. Mohan, D.J. Rasmussen, R. Muir-Wood, P. Wilson, M. Oppenheimer, K. Larsen, and T. Houser, 2017. Estimating economic damage from climate change in the United States, *Science*, *356*, 1362-1369 DOI: 10.1126/science.aal4369

3. Ranson, M., 2014. Crime, weather, and climate change, *Journal of Environmental Economics and Management*, *67*(3), 274-302. https://doi.org/10.1016/j.jeem.2013.11.008

4. Urban, M.C., 2015. Accelerating extinction risk from climate change, *Science*, *348*, Issue 6234, 571–573, DOI: 10.1126/science.aaa4984, http://science.sciencemag.org/content/348/6234/571.full

Reproducibility and Scientific Method

1. Ball, P., 2018. High-profile journals put to reproducibility test, *Nature*, 27 August, https://www.nature.com/articles/d41586-018-06075-z, doi: 10.1038/d41586-018-06075-z

2. Barba, L., 2016. The hard road to reproducibility, *Science*, *354*, 142. doi 10.1126/science.354.6308.142

3. Barba, L., 2016. Reproducibility Syllabus, http://lorenabarba.com/blog/barbagroup-reproducibility-syllabus/

4. Barba, L., 2018. Terminologies for Reproducible Research, https://arxiv.org/abs/1802.03311

5. J.B. Buckheit and D.L. Donoho, 1995. Wavelab and Reproducible Research, https://statweb.stanford.edu/~wavelab/Wavelab_850/wavelab.pdf

6. Implementing Reproducible Research, edited by V. Stodden, F. Leisch and R. Peng

7. The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences Kitzes, Turek and Deniz, eds.

8. Reproducibility: a Primer on Semantics and Implications for Research

9. Rokem, A., B. Marwick, and V. Staneva, 2018. Assessing Reproducibility, in *The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences*, University of California Press. https://www.practicereproducibleresearch.org/core-chapters/2-assessment.html

10. Schapin, and Schaffer, 1985. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*, Princeton University Press, NJ ISBN 0-691-08393-2

11. Stark, P.B., 2018. No Reproducibility Without Preproducibility, *Nature*, *557*, 613. https://www.nature.com/magazine-assets/d41586-018-05256-0/d41586-018-05256-0.pdf doi: 10.1038/d41586-018-05256-0

12. Stark, P.B., 2017. Preface to *The Practice of Reproducible Research*, J. Kitzes, D. Turek, and F. Deniz, eds., University of California Press, Berkeley

13. Teytelman, L., 2018. No more excuses for non-reproducible methods, *Nature*, *560*, 411. https://www.nature.com/articles/d41586-018-06008-w, doi: 10.1038/d41586-018-06008-w

Weaponizing Reproducibility

1. Hakim, D. and E. Lipton, 2018. Pesticide Studies Won E.P.A.'s Trust, Until Trump's Team Scorned 'Secret Science', *New York Times*, 26 August. https://www.nytimes.com/2018/08/24/business/epa-pesticides-studies-epidemiology.html