

This workshop is designed to give you a brief introduction to several software tools to help you make your work more reproducible. Specifically, we'll work through a simple example to help you learn a little bit about The Open Science Framework, study registries, R Markdown for dynamic documents, sending Stata regression output automatically to your LaTeX paper, versioning your code with git and GitHub, and sharing your data with Harvard's Dataverse.

We'll model our example off a multi-arm water, sanitation, and hygiene cluster randomized trial in Kenya. There are eight arms to the trial (W, S, H, N, WSH, WSHN, Active control, Passive Control). You enroll pregnant women, give them tools and instruction/encouragement for good environmental health behaviors, and then evaluate the health of their children two years later. Hopefully the kids will be taller and have less diarrhea. (This study is called WASH Benefits, and is ongoing in Kenya and Bangladesh. This is mostly a health study, but the tools are very similar for the social sciences. You can read the published protocol in *BMJ Open* or results of a pilot in the *American Journal of Tropical Medicine & Hygiene*.)

1. You've come up with the idea for the project, and you want to register it. Create an account on the Open Science Framework (if you don't already have one). Create a new project, either from scratch, by using a template, or forking an existing project.
2. In real life, do a power calculation to make sure your study is large enough (80-90% power).
3. Write a simple version of a pre-analysis plan using your text editor of choice.
 - (a) Main hypothesis
 - (b) Primary outcomes
 - (c) Secondary outcomes
 - (d) The precise composition of any families that will be used for mean effects analysis
 - (e) The subgroups that will be analyzed
 - (f) The direction of expected impact if we want to use a one-sided test
 - (g) The primary specification to be used for the analysis
4. Register the trial and the PAP on the OSF.
5. Go to Kenya and run the experiment.
6. The data fairy collected your data and conveniently put it on the Internet. Download the data from my GitHub repository: <http://www.github.com/garretchristensen/riverside>. You don't need an account to download the .zip archive. (If you're ahead of the game, you can clone the repo. If GitHub is new to you, we'll come back to that.)
7. Write code to label variables.
8. Write code in Stata or R to run the regressions you said in the PAP you were going to run, and version control the code with the OSF.
 - (a) Run your first regression and save the code.
 - (b) Upload the code to the OSF project.
 - (c) Change the code a little, and upload it to the OSF again.
9. Write code in Stata or R to run the regressions you said in the PAP you were going to run and version control the code with git.
 - (a) Create a github repo.
 - (b) Run your first regression and save the code, add, commit, and push it to the repo.
 - (c) Change the code a little, add, commit, and push.
 - (d) Have somebody else push to your repo, and pull the changes to sync.

10. Run several regressions in Stata, and put the output directly into your regression table in LaTeX.
 - (a) Write a Stata .do file with regression commands, followed by `outreg2` commands and `tex` option.
 - (b) Write a LaTeX file with the ‘standalone’ package and include the output files.
 - (c) Compile the LaTeX document.
 - (d) Change the regressions, and then recompile the pdf.
11. Run a simple analysis in R Studio using R Markdown, which combines the code and output into one dynamic document.
12. Share your data with the world using Harvard’s Dataverse.
13. Link the OSF project to the data on Dataverse, and the code on GitHub.

End of Workshop!