

Tools for a Reproducible Workflow

Garret Christensen¹

¹UC Berkeley:

Berkeley Initiative for Transparency in the Social Sciences

Berkeley Institute for Data Science

BITSS 2016 Summer Institute

Slides available online at `https:`

`//github.com/BITSS/SummerInstitute2016`



BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES



Workflow

Christensen

Introduction

Workflow

Conclusion

“Reproducibility is just collaboration with people you don’t know, including yourself next week”
—Philip Stark, UC Berkeley Statistics

Christensen

Introduction

Workflow

Conclusion

- Version Control
- Dynamic Documents

Christensen

Introduction

Workflow

Conclusion

- Version Control
- Dynamic Documents



Write your code and your paper in the same file so you won't lose information or make copy and paste mistakes.

- Include tables by linking to a file, instead of a static image.
- Include number by linking to a value calculated by an analysis file, instead of a static number typed manually.
- Automatically update tables and numbers.
- Produce entire paper with one or two clicks.

Possible in Python, R, and to a lesser extent, Stata

- Jupyter—several (many?) languages
- R—use R Studio to manage projects with built-in version control, and R Markdown/knitr for publication-quality dynamic documents.
- Stata—combine with LaTeX for two click workflow
- Stata—use ‘markdoc’ ado for some dynamic ability.



Studio[®]

The Jupyter logo features a stylized orange circle with a gradient, composed of two curved segments. Four dark gray dots are positioned at the top-left, top-right, bottom-left, and bottom-right corners of the circle. The word "jupyter" is centered within the circle in a dark gray, lowercase, sans-serif font.

jupyter

Conclusion

Christensen

Introduction

Workflow

Conclusion

OK, I'm convinced. How do I learn more?

- Work through my demos. [▶ Link](#)
- Software Carpentry's tutorials [▶ Link](#)