What Replicators Will Be Looking For, and Why

Lars Vilhuber and David Wasser

April 26, 2019

Motivation

- We want to be able to verify the work that has been done.

- In order to do that, we need some direction.

- This document describes:
  - What authors should check before providing data and code to journals
  - What verifier teams should check for in the data and code provided to them for the purpose of verification

The README file

- Make sure the contents of the README match the contents of the submission

- Cross-check against the paper's data section

- Make a list of:

The README file

- Make sure the contents of the README match the contents of the submission

- Cross-check against the paper's data section

- Make a list of:
  - Data sets referenced or provided

The README file

- Make sure the contents of the README match the contents of the submission

- Cross-check against the paper's data section

- Make a list of:
  - Data sets referenced or provided
  - Tables
  - Figures

The README file

- Make sure the contents of the README match the contents of the submission

- Cross-check against the paper's data section

- Make a list of:
  - Data sets referenced or provided
  - Tables
  - Figures
  - In-text numbers

The README file

- Make sure the contents of the README match the contents of the submission

- Cross-check against the paper's data section

- Make a list of:
  - Data sets referenced or provided
  - Tables
  - Figures
  - In-text numbers

- You can exclude a list of tables, figures, and in-text numbers that appear in an online appendix

Documenting data

- For each data source, verify:
  - ▸ The data set has a clear name
  - ▸ Licensing and access information: accessible to others? right to redistribute?
  - ▸ The data is cited in **both** the paper and the README

Documenting data (continued)

- For each data set, verify:
  - All variables are labeled or a codebook is provided
  - For confidential data, summary statistics are provided **NOTE FOR LARS: It might be helpful to clarify what we mean by summary stats**
  - No potentially sensitive information in the data set, including but not limited to:
    - Names, SSNs, credit card numbers, addresses etc.

- Remember: Someone else is reading your code and using your data. Variable abbreviations that make sense to you might not make any sense to someone else.

Documenting tables, figures, and in-text numbers

- Can you identify the part of the code that generates that table, figure, or number?

Documenting tables, figures, and in-text numbers

- Can you identify the part of the code that generates that table, figure, or number?

- How about the line?

Documenting tables, figures, and in-text numbers

- Can you identify the part of the code that generates that table, figure, or number?

- How about the line?
  - Add comments to your code and specific references in your README

Documenting tables, figures, and in-text numbers

- Can you identify the part of the code that generates that table, figure, or number?

- How about the line?
  - Add comments to your code and specific references in your README

- Does the code produce an **identifiable** output that contains the table, figure, or number?

- Remember: Someone else is reading your code and using your data. You might know that Column 1 of Table 2 is created by one piece of code while Column 2 is created by another piece, but they do not.

Code verification: where the rubber hits the road

- Create a directory that only contains the provided programs and data

- Create a *config* file that stores file paths and relevant system/software parameters in one place

- Install all identified requirements
  - Easy to forget that `estout` or `outreg2` was once manually installed

- Run the code using the instructions in the README

- Identify all error messages

- Identify all outputs as per the README and list of table, figures, and in-text numbers

- Compare outputs to those in the paper