Best Practices: Reproducible Data (Based on Millman and Perez 2015)

Goals:

* To be able to retrace steps

Best Practices:

* When possible, use open source software; proprietary software hinders reproducibility.
* Re-examine datasets with independent and newly developed packages to ensure consistency of results
* Document collaborative research through version control repositories such as Github to track changes and contributions
* Automate executable processes
* Validate code and test often
* Make code machine and human readable
  + Practice literate programming
* Try to implement a continuous integration system
* Share the knowledge of scientific commuting! Make sure others understand.

“Reinhart-Rogoff affair”- how does it relate to open science and reproducible data analysis

The Reinhart-Rogoff affair is an example of why data should be open for public scrutiny, or at least available for peer review prior to publication. In this case, Reinhart and Rogoff have influenced potentially important political decisions on the US economy, including how the budget deficit should be managed, based on erroneous data analysis. Even if their conclusions were correct, the authority from which they were made was affected, perhaps harming other work complementary to their own. The Reinhart-Rogoff affair exhibits obvious parallels with research conducted on climate change and the skepticism that encompasses it. Ultimately, making methods and data available for scrutiny by others aids in reproducibility, and reproducibility is important for scientific validity and policy change.