

- Changing the data. Questions like “How stable are my conclusions? What happens to all my figures and tables if I omit the 5 smallest states from my panel of 50 states?” can be rapidly answered. The easier it is to investigate new analyses, the more things you try.
- Effective collaboration. All coauthors can read, run and modify all the code that produced the figures in the current circulated draft.
- Debugging. If your adviser asks “How exactly did this number get produced?” you can give a rapid, precise and accurate answer.
- Updating. If you are presenting code (e.g., lecture notes) and you want to make changes where necessary for a new software version, you simply re-run the document.
- Revisions. 4 months after you submitted the paper, when the referee reports come back, you will be glad if you have your work organized in this way!

What do you think are the most compelling reasons for you, personally, to consider following the compendium approach of Gentleman (2005) for your thesis research?

What do you think are the main reasons why you, personally, might choose not to follow the compendium approach of Gentleman (2005) for your thesis research?

“If my analysis involves the big data set so that the time needed to run the code is very long, their approach would be not that useful. Since you have to wait, say, about a month to see the figures after running the code.”

“If the thesis requires a huge amount of computations and a long code, I might choose not to do so.”

Ironically, computationally intensive research that may benefit most from reproducibility appears like a reason not to do it!

“Ironically, computationally intensive research that may benefit most from reproducibility appears like a reason not to do it!”

Is there a satisfactory resolution to this?

“If the majority of people in academia are not using it, I dont see why I should use it. After all it requires extra work.”

Currently, only a small fraction of Statistics papers are published as a compendium. For what reasons do you think this fraction is not larger?

“A large amount of datasets is not public and the datasets are getting larger and larger now.”

“Readers might be more used to reading the traditional kind of static papers. They might not want to spend much time dealing with compendium. Thus publishing institutes are not really using this at the moment.”

The Journal of Statistical Software
currently has the 2nd highest impact factor
of all Statistics journals (behind JRSSB,
but ahead of JASA, Annals of Statistics,
Annals of Applied Statistics, ...). Why?

The JSS model: papers are Sweave/Knitr
documents describing the theory and practice of
an R package published on CRAN.

“I think one major reason is that people can easily use your result. And this may influence your publication.”

What is the extra effort involved in turning an Sweave or Knitr version of a manuscript into a compendium?

“I’m sure once you know how to do it theres not much extra effort involved, but theres probably a learning curve.”

“I did read the paper carefully, and I am still confusing about the clear definition of compendium. In my opinion, Knitr version may be good enough.”

In the context of R, a compendium is an R package for which the main purpose is the vignette (i.e., the paper!)

“Sweave succeeds in making a navigable document by directly linking your R-code to the production of results and figures needed for the paper. However, as an author, you must make extra effort to ensure the software you have created is usable by others and available.”

“A big emphasis was allowing a reader to recreate any of analysis found in the paper. How does the compendium approach work when the analysis is based on day-long simulations? Especially if those simulations were run remotely? In some ways it seems like this sort of research would benefit the most from a compendium, but it would be more difficult to create an interactive compendium.”

“Are researchers allowed to publish a compendium if theyre publishing in a journal. It seems like that would make the paper freely available and the journal has an interest in limiting its circulation. In general, can researchers disseminate their own work even after it’s published?”

Reproducibility is power.

Demand it!