

Good Enough Practices for Scientific Computing

Pankaj Pandey

SciPy.in 2016
IIT Bombay

Enthought

Overview

- Tools and techniques for accessibility and efficiency of research
- Tried and tested by various researchers and continued adoption
- Software Carpentry, Data Carpentry

Why?

- Good enough practices (not Best?)
- For future self and others
- Quality and reproducibility

Aspects

- Handling data
- Software practices
- Collaboration
- Organizing the work
- Version control
- Report and paper manuscripts

Data Handling

- Save Raw Data
- Understandable: variable names, file format, filename
- Analysis friendly data
- Record data processing: script/document
- Share data

Software Practices

- Document files and programs
- Write code in functions (no global variables)
- Remove duplication/reuse existing software
- Meaningful names
- List dependencies (requirements.txt)
- Use control flow not comments
- Example running code
- Share code

Collaboration

- Overview of project
- TODO, project plan, roadmap
- License it for reuse
- Make it citable
- Collaboration tools
 - software repositories
 - document collaboration

Project organization

- One project one directory
- Separate text, raw data, results, code, external scripts and programs
- Name files appropriately

```
|-- CITATION
|-- README
|-- LICENSE
|-- requirements.txt
|-- data
|   -- birds_count_table.csv
|-- doc
|   -- notebook.md
|   -- manuscript.md
|   -- changelog.txt
|-- results
|   -- summarized_results.csv
|-- src
|   -- sightings_analysis.py
|   -- runall.py
```


Tracking Changes

- Version control: git, github
- Documents: Google drive, Dropbox
- Frequent small changes
- Share frequently: Avoid emails with attachments
- Frequent backups of everything

Manuscripts and Reports

- Single master copy
- Rich text editing: Google docs, MS office
- Git for plain text formats: Latex, markdown
- Publish supplementary materials like tables and figures other than just embedded pdf formats

Summary

- Share everything needed for reproducing
- Track changes
- Tools to make collaboration easier
- Make your work reproducible

Further reading

- <http://software-carpentry.org/lessons/>
- <http://datacarpentry.org/>
- <https://swcarpentry.github.io/good-enough-practices-in-scientific-computing/>
- Best Practices:
<http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001745>



Thank you!

Questions?

@pankaj_tweet