

Unit 3: Core tools

Project structure

Lesson 23

Derek Ruths

Overview of unit

Objectives:

- Understanding how data science activities necessitate certain kinds of tools
- Fundamentals with core data science tools

1. Why core tools?
2. Project organization
3. Python
4. Best practice: write CLI tools
5. Best practice: write unit tests
6. Best practice: resource referencing
7. Github
8. Jupyter notebooks
9. Jupyter & statefulness
10. Bokeh
11. Advanced bokeh
12. HW 3

Lesson overview

Objectives

- Understand the basic structure for a data science project

Outline

- The starting structure

Typical structure

<project-name>

README.md

src

data

scripts

← code that's not cli
← not necessarily checked in because of so big
organized: vaudata, outputs, analysis
← data science end user tools
 ↳ eg python tool in HW4
 ↳ built to run
 ↳ cli stuff
 ↳ could import stuff from src to use

Anatomy of a README

Your README is a way to communicate with OTHER people who want to pick this project up/use it. Write with them in mind. If you make unusual knowledge assumptions, say so up front.

- Welcome/overview – what is this project? What is it for?
 - TODOs – notes from the author, to the author/knowledgeable user
 - Gotchas – any nasty surprises someone should know about
- Setup/Installation
- Running

Lesson wrap-up

Takeaways

- Data science project structure isn't complicated
- Keep data separate
- Keep your README up-to-date

Up next

- Python