

Unit 3: Core tools

Why these core tools?

Lesson 22

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. Git
8. Jupyter notebooks
9. Jupyter & statefulness
10. Bokeh
11. Advanced bokeh

Lesson overview

Objectives

- Understand the data science activities that necessitate specific kinds of tools
- Understand current best tools for these activities

Outline

- Exploration
- Expressiveness
- Reproducibility
- Collaboration
- Communication

Exploration

- need to explore the data throughout

Need tools that allow quick and easy experiments

- Efficient language
- Good library support

data \Rightarrow questions

data is inherently unwieldy and large

Expressiveness

Need tools that allow you to implement complex algorithms

*data transformations
ML algorithm / training
regime*

Reproducibility

Need to be able to prove how we got a result.

for DS projects
we are accountable for

Collaboration

Need to coordinate and share work with colleagues

Communication

- Need to deliver findings in accessible, engaging ways
- Need to deliver prototypes

Lesson wrap-up

Takeaways

- Data science activities have certain universal needs
- These needs drive the requirement for certain kinds of tools

Up next

- Data science project structure