Unit 3: Core tools

Reasons for Coding

Lesson 25

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

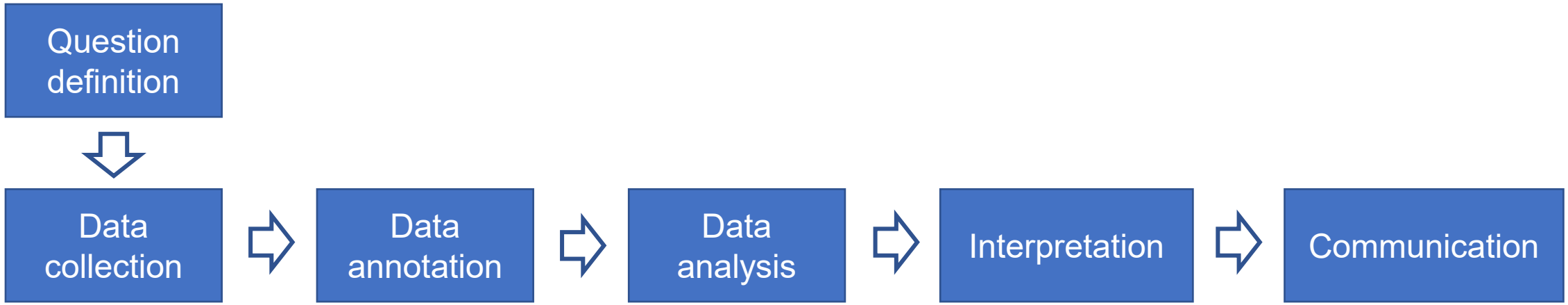
Objectives

- Understand the different reasons we code in a data science project
- Understand why most coding should be “tooling”

Outline

- Review data science project phases
- Ways of coding
- Why exploratory coding should be limited

Data science project phases



Remember: reproducibility!!!

Ways of coding (in data science)

Exploratory

trying lots of things

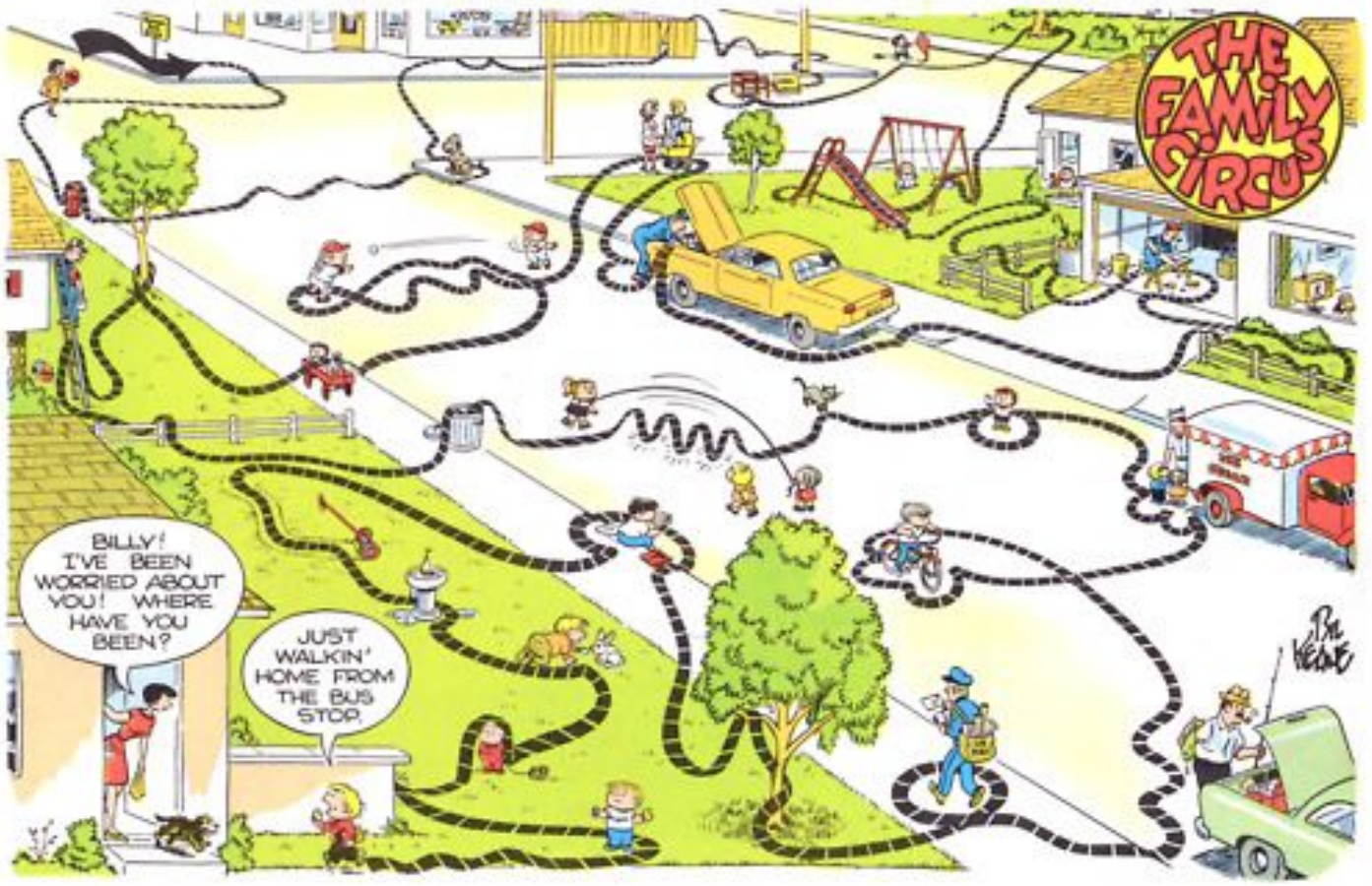
Tool building

Pipelining

figure out the process and stitch it together

Remember: reproducibility!!!

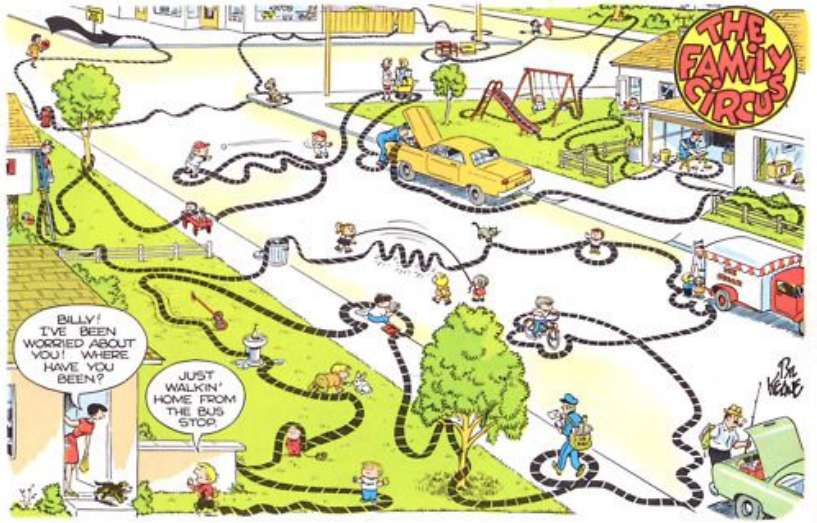
Exploratory coding is like...



Remember: reproducibility!!!

Scripts vs. Jupyter Notebook

units of work
run independently



Lesson wrap-up

Takeaways

- There's a constant tension between coding to tool and explore
- Jupyter is a powerful exploration tool
- Scripts are powerful for tooling … but used the right way, are good for exploration as well.

Up next

- Writing scripts