Effective python programming

Monday, May 27, 2019 1:55 PM

In week 1, we dealt mostly with the complexities of Jupyter notebooks as a programming environment.

In this week, we will cope instead with the subtleties of programming in Python for data science.

Elements of effective Python programming

- (Stepwise) refinement.
- Style and consistency.
- Readable variable names.
- Digestible chunks.
- Encapsulation
- Data structures

(Stepwise) Refinement

- Refers to a specific style of programming.
- Sketch out the program in English.
- Write comments describing each step.
- Create Jupyter cells for each step. Label each one with one of the comments.
- Fill in the details, cell by cell.

Style and consistency

- Python Enhancement Proposal (PEP) 8: a set of rules for Python style.
- Indentation.

- White space.
- Line length.
- Defining variables before first use.

Readable variable names

- Describe what is actually contained within the variable.
- Especially important when a variable is the result of a cell.
- Increase reusability of your notebook.
- Substitute for a lot of comments.

Digestible chunks

- Humans are limited in the programs they can write.
- McCabe's observation: if a program has more than 10 if statements -- implied or explicit -- then it is not testable manually.
- Break programs into smaller "cells".
- Debug each one separately.
- Document cell inputs and outputs.

Encapsulation

- Don't repeat code when you need to do something twice.
- Instead, write functions that do the same thing.
- Call the function rather than repeating code.
- Improves reusability and maintainability.

Data structures

- Often, there are multiple ways to achieve a result.
- These can execute in drastically different amounts of time.
- A central issue: which data structures are excellent for which tasks?