Python for Data Analysis and Visualization

CS102
Winter 2019

Python

- Very popular general-purpose programming language
- Used from introductory programming courses to production systems

Python Features

- Dynamically typed
  (rather than statically typed like Java or C/C++)
- Interpreted
  (rather than compiled like Java or C/C++)

Python programs are comparatively…

+ Quicker to write
+ Shorter
– More error-prone
– Slower to run

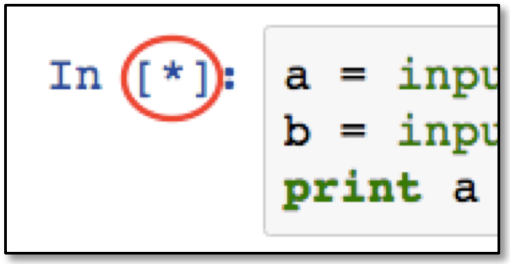
Python for Data

- Fairly easy to read/write/process data using standard features
- Plus special packages for…
  - Numerical and statistical manipulations - numpy
  - Visualization (“plotting”) - matplotlib
  - Relational database like capabilities – pandas
  - Machine learning - scikit-learn
  - Network analysis - networkx
  - Unstructured data – re, nltk, PIL

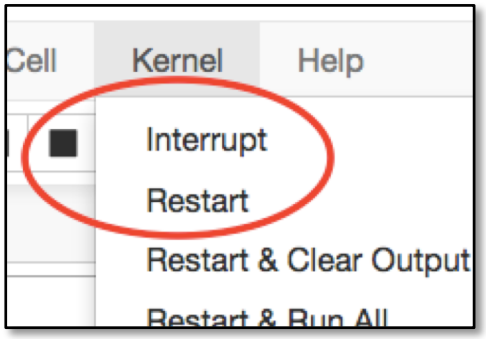
Jupyter Notebook Hints

To execute a cell, click inside the box then type shift-enter (or shift-return)

If nothing happens, some cell is probably still executing



Try Kernel > Interrupt or Kernel > Restart



What We'll Cover

1. Python basics
2. Data manipulation
3. Plotting
4. Pandas

(more in later topics)

For help while working with Python:

Tutorials and help pages (website)

- Web search