Topic Outline: Data + Python

Revised: January 25, 2016

Materials

- First-day handouts: Syllabus, Project Guide, Due Dates
- Today's handouts: this outline, three ideas, book chapters, red/green stickers
- All posted on *Topic list & links* page of website (except the stickers).

About the course

- Data + Python = Magic!
  - Arthur C. Clarke, Jessica, Tim
- What?
  - ... are you doing here?
  - Skills are nice, coding is literacy for the modern age
  - Something to show potential employers
- Why?
  - Why data?
  - Why code?
  - Why Python?
  - Why bootcamp?
  - Why you?
- Things we believe
  - Anyone can do this. Target audience is programming newbies — with courage.
  - It's ok to be lost. We've all been there, it's not permanent.
  - This is fun. Really.
- Rules to live by
  - Don't panic. It will seem overwhelming at first, but stick with it and you'll be fine.
  - One step at a time. Don't rush this. In six weeks you'll know a lot.
  - Learn by doing. Same directions as Carnegie Hall, no shortcuts.
  - Ask for help. Don't be a hero, let us know if you could use some help.

- Course materials
  - Google "nyu data bootcamp"
  - Website (thanks, Spencer): http://databootcamp.nyuecon.com/ (bookmark me!)
  - Book
  - Topic list & links
  - Discussion group
  - Data page
  - GitHub repository: https://github.com/DaveBackus/Data_Bootcamp

Anaconda

- Install the Anaconda distribution
  - Put red sticker on your laptop
  - Distribution?
  - Google "anaconda download" or borrow a USB drive
  - Download or copy installer to your computer — **Python 3.5!**
  - Run installer
  - Start Launcher (use search box)
  - Replace red sticker with green when Launcher opens
- Environments
  - Environments? (Analogy: Word is an environment for creating Word docs.)
  - Spyder: classic coding environment with editor and output windows
  - Jupyter: environment for creating IPython notebooks, which combine code with text and output

Run test program – twice

- Create `Data_Bootcamp` directory/folder on your computer. **Raise your hand if you're not sure what that means or how to do it.**
- Test program code:

```
"""
Test program for Data Bootcamp course @ NYU Stern
"""
import sys

print('Welcome to Data Bootcamp!')
print('Python version:')
print(sys.version)
```

- Run test program in Spyder
  - Put red sticker on your laptop
  - From Launcher, launch Spyder (labelled "spyder-app")
  - Look around (editor, IPython console, Object inspector)
  - Enter test program in editor (on the left)
  - Save in `Data_Bootcamp` directory as `bootcamp_test.py` (File, Save as, look for folder)
  - Run program (click on large green triangle)
  - Look for correct output (last line should be `3.5.x etc`)
  - Switch to green sticker if it works
- Run test program in Jupyter
  - Put red sticker on your laptop
  - From Launcher, launch Jupyter (labelled "ipython-notebook")
  - Navigate to `Data_Bootcamp` directory
  - Open a new IPython notebook (New, Python 3)
  - Change name from `Untitled` to `bootcamp_test`
  - Look around (toolbar, menubar, code cells)
  - Enter test program in code cell
  - Run program (Cell, Run All)
  - Look for correct output (last line should be `3.5.x etc`)
  - Switch to green sticker if it works
- **More than we need (mtwn)**
  - Ignore if you find confusing, we'll cover later
  - Explain: `import`, `print()`

Practice and review

Put red sticker on your laptop, replace with green when you're done. Discuss with your neighbor. Raise your hand if you could use some help.

1. Fill in the blanks in this table:

Environment	File or Object
MS Word	Word document
MS Excel	Excel file
iTunes	
Spyder	
	IPython notebook

2. Run the `Maddison_data_input.py` Python code example.
   - Go to the `Data_Bootcamp` GitHub repository (link above).
   - Navigate to the `Code` directory and `Lab` subdirectory.
   - Click on `Maddison_data_input.py`.
   - Click on the Raw button and save file in your `Data_Bootcamp` directory.
   - Open file in Spyder (File, Open).
   - Run it by clicking on large green triangle.
   - What do you see?

Thinking about data

- Data + Picture = a compelling way to tell a story
- Where we're headed
  - Think of a **graph** you'd like to produce – a "visualization"
  - And the **story** it tells
  - And the **data** that went into it
- Examples (links on *Topic list & links* page) [Gapminder]
- Questions about graphs
  - What did you learn, what is the **story**?
  - **What else** would you like to know?
  - Where did the **data** come from?
- Examples revisited, answer the questions
- Course projects
  - Course structure: tools, project
  - Opportunity to show off your skills (Projects directory of GitHub repo)
  - First step: develop project ideas (**ideas are developed, not discovered**)
  - What interests you? (finance? movies? soccer?)
- Idea machines
  - Start with an idea or subject (what interests you?)
  - Start with a dataset (you'll know more shortly)
  - Start with an example (see link on data page)
- Three ideas
  - Put red sticker on your computer
  - Goal: write down three ideas, 1-2 sentences each (see handout)
  - Use your imagination, don't overthink it (improv: what's your name?)

- Talk to your neighbors, bounce ideas around
- Or look at the *Data sources* page of the course website
- When you're done, switch to green sticker
- Share an idea with the class, ask for suggestions for developing further
- Save ideas for future reference

After class

- Required
  - Read Syllabus, Project Guide.
  - Mark Due Dates on your calendar.
  - Skim chapters 1-3 of the book.
- Recommended
  - If you haven't already: join the discussion group, take the entry poll.
  - Post a link to an interesting graph on the discussion group.
  - Look through the IPython notebook `bootcamp_examples.ipynb` in the IPython directory of the GitHub repo. What graphs interest you? What data? Do they suggest anything else you might explore?