Topic Outline: Data + Python

Revised: September 6, 2017

Materials

- Today's handouts: Syllabus, this outline, red/green stickers

- All posted on 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 me?
  - 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

    - Required: practice, exam, project
    - Google "nyu data bootcamp"
    - Website, topic list & links (thanks, Spencer):
      https://nyu.data-bootcamp.com/undergrad_outline/ (bookmark me!)
    - Book:
      https://www.gitbook.com/book/nyudatabootcamp/data-bootcamp/details
    - Discussion group:
      https://groups.google.com/d/forum/databootcamp_fall2017_undergrad
    - Data page:
      https://nyu.data-bootcamp.com/data/
    - GitHub repository: https://github.com/NYUDataBootcamp/Materials

- You

    - Come to class
    - After class: **write** and **read**
    - Practice
    - Have fun

Setting up your computer

- Create `Data_Bootcamp` directory/folder on your computer. This is a place for you to save stuff to, work from, etc.

- So do this...

    - Put red sticker on your laptop
    - Find your main hard drive, on PCs typically the "C" drive
    - In the C drive, create a folder called `Data_Bootcamp`
    - Replace red sticker with green!

- Now if you every have to call this file you know it is `c:\Data_Bootcamp`

Atom

- A versatile, text editor. Must have tool for serious work.

- We will use it to open files, look at them, learn some markdown.

- Lets install it
  - Put red sticker on your laptop
  - Google "atom download" or borrow a USB drive
  - Download or copy installer to your computer
  - Run installer
  - Start Atom
  - Replace red sticker with green when Atom opens
- Go and open new file
- type "my first file", then comma, then birth year, then comma, high school graduation year, then comma, college graduation year (or expected graduation year). For me it looks like this

  `my first file, 1978, 1996, 2001`
- Save as: `my_first_file.csv` in your `Data_Bootcamp` file.
- **Pro tip: naming convention and file type is important in this course**

GitHub

- What I will use it for. . .
  - Source of ALL course materials
  - Place for you to grab materials on the fly. Save files by cut and paste, clever save as, or "Raw" (ask about this)
- You need to create an account and email me your username
- What you will use it for. . .
  - Post your homework and projects; I will "pull" them from there.
  - Long run: Think of this like an artists portfolio. Here you can post your code and projects and show potential employers, family, friends what you have done.
- Now lets use it
  - Put red sticker on your laptop
  - Sign into github (if you do not have an account, create one)
  - Create a new repository and name it `my_first_repository`,
  - place the `my_first_file.csv` file you created in it,
  - Great job! Replace red sticker with green!
    Next class, I'm going to try and "pull" the file and then report some statistics about the class to you.

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 3.6!**
  - Run installer
  - ONLY FOR PC USERS: on Windows Advanced Installation options, click the box "Add Anaconda to my PATH environment"
  - Replace red sticker with green when installation is complete

- 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

- 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 `3.6.x` etc)
- Switch to green sticker if it works

- Spyder startup summary

  - Open by typing Launcher in search box (spotlight on Macs), then choose spyder-app.
  - Or just type Spyder in search box

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
	Excel file
iTunes	
Atom	
Spyder	

2. Open Atom and a new file.

   - In a similar fashion as above, create a `.csv` file with the following information separated by commas
     - Expected major (or concentration)
     - Scale of 1-5 (5 being a lot, 1 being none) experience with programming
     - preferred form of social media
     - career path (e.g. Finance, Consulting, Technology, Fashion, public policy, rock star)
   - Name as `my_second_file.csv` and post on GitHub in `my_first_repository`

3. 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.
   - Get `Maddison_data_input.py`
     - Cut and paste into blank file

– Or: Save file in `Data_Bootcamp` directory (ask how)

- Open file in Spyder (File, Open).

- Run it by clicking on large green triangle.

- What do you see?

4. *Only if you have time.* Try this program: `OECD_health_indicators.py`. What do you see? What questions does it raise? (There are other files in the same directory, but some of them don't work yet.)

After class

- Required

  – Read Syllabus and 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.

  – Explore the website. Make sure you can find the book, due dates, topic outlines, assignments, and data sources.

  – Post a link to an interesting graph on the discussion group.

  – Look through the IPython notebook `bootcamp_examples.ipynb` in the `Code/IPython` directory of the GitHub repo. What graphs interest you? What data? Do they suggest anything else you might explore?