Computing in Context: Fall 2024

Lecture 9 | Review & Contextualizing

COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

Dr Sarah Ashcroft-Jones | Mailman School of Public Health

What have we learned?

- Python
- Data analytics
- Public health context
- Python plus public health data becomes policy suggestions …

Hard skills

- How to script in Python
- How to handle data
- How to build a pipeline to visualization
- Discussion of data ethics in public health

Hard skills contd.

- Iteration
- Conditional and logical statements
- Functions and functional programming
- Interacting functions to complete a larger goal

Hard skills contd.

- Data cleaning and tidy data
- Data visualization
- Clean versus “smelly code”
- IDEs and Jupyter notebooks

Also…..

- Debugging
- Data types
- Error decoding
- Loading data
- Object references
- Deep v shallow copies
- Accessible data viz
- Pandas syntax
- Color palettes
- Randomization
- While loops
- Sentinels
- Unit testing
- Commenting code
- Etc.

Soft skills

- Structured thinking
- Logical reasoning
- Breaking complex tasks into small steps –
  - Minimal Viable Product
- Problem solving

Soft skills

A computational mindset

Soft skills contd.

- Patience and humility
- Safe experiences of failure
- Resilience
- Creativity – thinking about the box

What did you learn? Anything unexpected?

How to talk about this?

Resume – hard skills

- Be clear on specifics of your coding infrastructure
  - Name modules: numpy, pandas, seaborne, etc.
  - Name IDEs – Spyder, Jupyter, Anaconda
  - Virtue signal through these keywords

- Be clear on specifics of coding knowledge
  - Functional programming
  - Iteration and conditionals
  - Data wrangling visualization

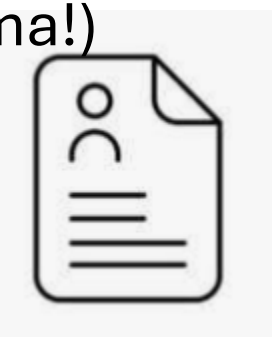
Resume – soft(er) skills

- Broader framing
  - Data management – wrangling, cleaning,
  - Building a pipeline for analysis
  - Understanding the limits of your data and code

- Public health context
  - Data in context – limits, richness, and data gaps
  - Breaking complex dataset down into digestible chunks
  - Asking better question about health

Resume - soft skills

- New ways of thinking – creative and structured/logical
- New ways of organizing – file, folder, directory management matters, it becomes a habit
- New ways of checking work – thoughtfulness / quality assurance
- New standards for work - meticulous attention to detail required
- New ways to ask question – why doesn’t this work, what could I change, ideating new problem solutions ..
  - Flipping challenges into opportunities with every line of code (the drama!)



Interview – Python is the perfect “challenge”

- This may be the first time you have really struggled with a skill
- This may have required a lot of independent learning
- This may have been hard to juggle time wise
- It’s a new way of thinking, a new language, and a courseload all at once – this may have been tough on your cognitive load
- It may have challenged how you think about data in health
- It may have just really sucked and you had to do it anyway!
- Maybe it was just plain hard.

Interview – Python as your success story

- Independence and autonomy thrust on you
- Ownership of work – sink or swim (no group to lean on)

- Success despite all the challenges
  - (feel free to throw me under the bus "*terrible hands-off Professor who never explained anything and barely showed up to class*"– just not in evals please!)

Questions?

Where to next?

Where to next? Python

- Change your Spyder to dark mode – a true programmer.
- Writing unit tests
- Increasing fluency of data wrangling (find a filthy dataset!)
- Implement your stats models in Python (lean into NumPy, built in statistics, pandas, and SciPy.stats)
- Expand your data viz – dynamic visualization, interactive data viz

Where to next? R

- Consider R (in rStudio IDE) for other data analytics
- Pandas maps to dplyr
- Seabourne and plotnine map to ggplot2
- RMarkdown maps to jupyter notebooks
- Recreate your project in R? A personal challenge?

Where to next? SQL

- A language for interacting with databases (big in healthcare)
- Can use SQL queries to pull data from databases
- Then use Python to explore that data – often used in tandem
- Expanding your health data lifecycle

Where to next? GitHub

- Sharing your code
- Collaborating on coding projects
- Building “branches” of code and selecting best option later
- Version control and “reverting” changes

Where to next? AI

- Autocomplete on for suggestions for
- Build test cases for ChatGPT/ NLP model suggestions – did they make a semantic error?
- Seeing how AI would restructure your “smelly code” another way of thinking about the same problem
- Don’t “turn off” the good thinking skills you have built!

Do you have a plan to keep growing this skill?

- Find a project
- Find a buddy
- Find a coding mentor
- Find a rubber duckie

Questions?