Introduction to Python for Scientist and Engineers

(a boot camp for data science)

Craig Group
Virginia

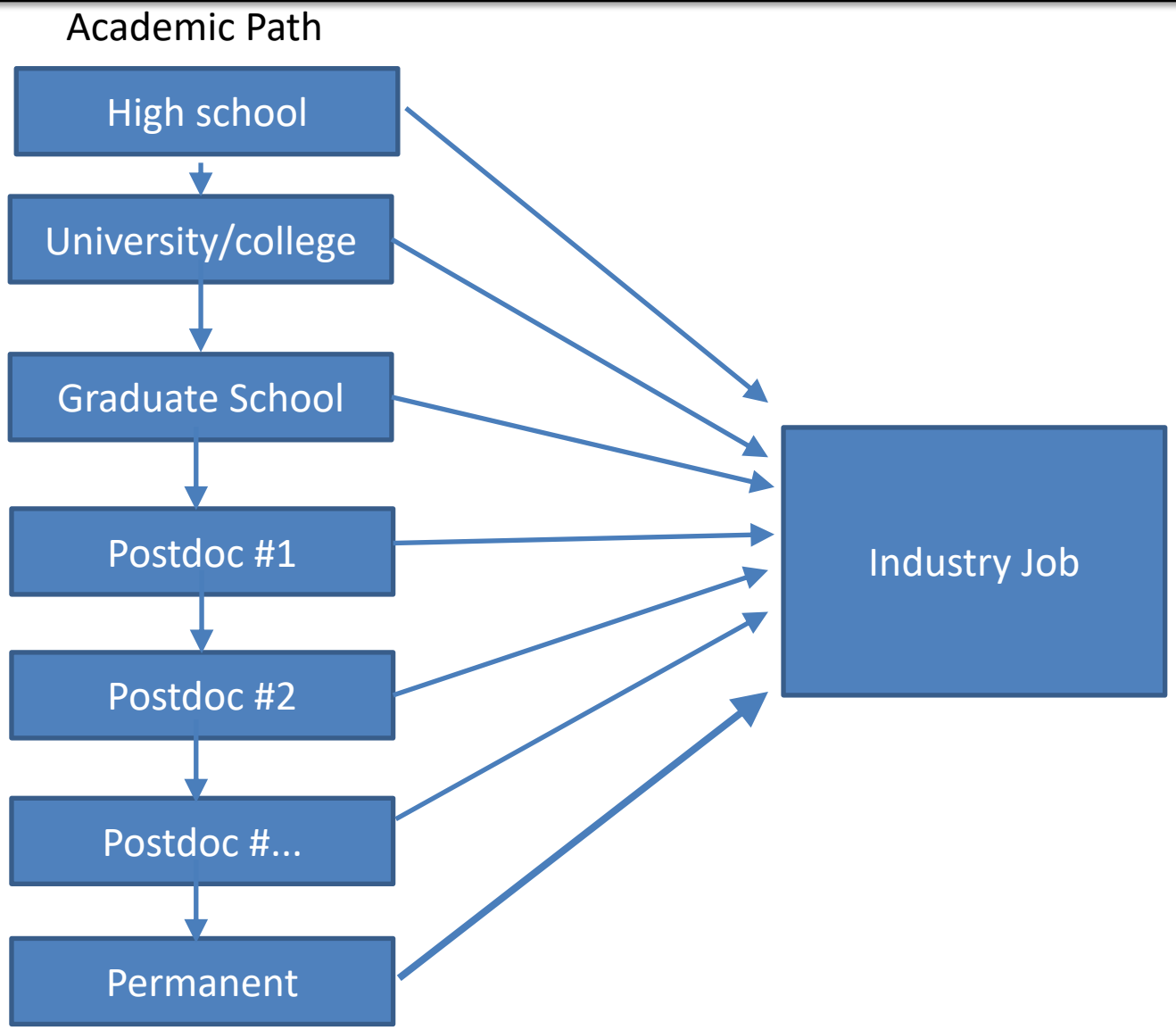
But first…

- Please consider joining us for the **Virginia Academy of Science** Centennial meeting.
- I am the chair of the Astronomy, Math, and Physics section, and we would be happy to have some Physics education talks in our session.
- Also, it is a great place for undergraduates to come and give their first talk.
- Coordinates:
  - May 25th at William and Mary
  - Request a talk here: https://vacadsci.org/ by 4/10/23

Valuable Skills for Students

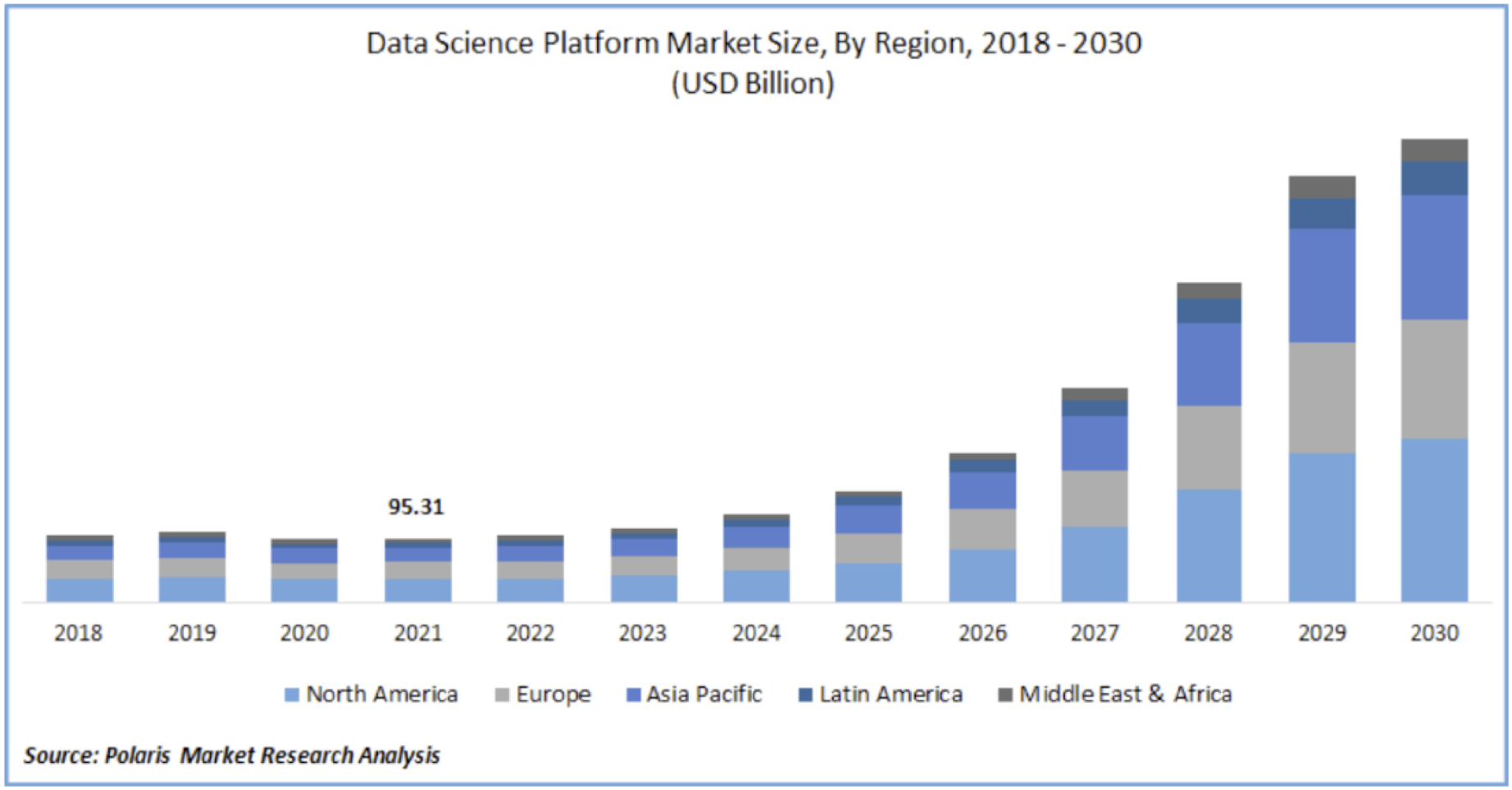
- As educators, planning the curriculum, we have an obligation to prepare our students to contribute to society and and have productive careers.
- Physicists need to program and analyze data, and these skills are extremely marketable.
- We can incorporate these skills directly into physics classes, or have dedicated classes for students to develop these skills.
- We should offer these options early and often!

”Typical” physics career path



Data Science Jobs

Physicists need data science skills, and so does industry!



Topical Group on Data Science

- The American Physical Society has a Topical Group on Data Science
- Within that group is a newish initiative within the Data Science Education Community of Practice (DSECOP)
- 2nd workshop, which will be held from **June 26–28, 2023**, at the University of Maryland, College Park. https://dsecop.org/workshops/
- Workshop purpose:
  - Strategies for integrating data science concepts into the undergraduate physics curriculum
  - Best practices for teaching data science to students with varying levels of experience
  - Exploring the latest tools and technologies for data analysis in the physics classroom
  - Sharing successful case studies of data-driven physics instruction
- I attended last year, and this helped me improve my class!

Interested? fill out the application form below:
https://forms.gle/iFLo7y8HXBmTdoxe8

Past: Scientific Computing at UVA

- Introduction to Scientific Computing
  - For about 15 years we have had a 2000-level class that was required for our BS physics majors.
  - Based on C and Gnuplot
  - Two main thrusts:
    - Introduction to C programming language
    - Computational/statistical analysis techniques largely based on examples in physical sciences
- Computational Physics
  - Two semester 5000-level sequence of more advanced techniques (some C, python, and C++)

New python course

- Our chair approached me about starting a new large-enrollment python course in the physics department.
  - 1000-level
  - No prior knowledge required
  - Not focused on physics
  - Any student could take it to satisfy one of their general-education requirements.
- Taught this class for the first time in Spring 2022 for the first time.

Course Structure

- Couse meets twice per week for 1 hr 15 min.
- Use the UVA High Performance Computing cluster
- Flipped classroom:
  - Reading (usually) required before class – tutorials too
  - Short lecture on “theory”
  - ~1 hr for in-class work:
    - Usually Jupyter notebook with several built-in exercises
    - Students work in pairs
    - Me and two undergraduate students roam the room and help (and ask annoying questions)
  - Weekly HW builds on the in-class examples.

The class

Tables of 4, teams of 2.

Data Science?

- Since I could not focus the course on physics problems, I decided to **teach python with the goal of learning to analyze data.**

  → **Data Science**

- The course has three main focuses:
  - Basics of Python
  - Introduction to statistics
  - Using data science tools to analyze datasets

Basics of Python

Course: Introduction to Data Science for Scientists and Engineers

class day	Day	Date	Topics	In-class	HW	Reading
1	W	8/24	Computers, Linux, Linux Tutorial	Login to Rivanna / Linux		Linux
2	M	8/29	Rivanna - Will from Research Computing will attend??	.basrc, hello.py		Emacs, ways to run python: Sundnes Ch.1
3	W	8/31	Why Python? Python Tutorials - scripting v/s interactive mode v/s notebooks.	Using the Emacs editor, running Python code, python calculator.	HW1: Linux Tutorial	
4	M	9/5	Labor Day - PG Travelling - no class meeting	Labor Day - PG Travelling		Sundnes Ch.2; Ch.3; 7.4; Wood Ch. 3;
5	W - drop	9/7	Variables, memory, "for" loops, strings (Group away -TAs)	command line input and strings	HW2: Python Tutorials, string manipulation, and user input.	
6	M	9/12	Random numbers and Monte Carlo integration	Math module, and random numbers		Sundnes 2.3 and 6.1; Wood Ch.4 and Ch. 5; MC integration
7	W	9/14	f-strings, lists/tuples,	lists, strings, dictionaries, fstrings	HW3: Calculate pi with Monte Carlo methods	
8	M	9/19	numpy arrays v/s lists, Ufuncs	numpy, v/s lists		Sundnes: Ch. 5, For Numpy Ref VandePlas Ch. 2
9	W	9/21	File Input/Output	File I/O with Iris/co2 datasets. Too short? Add to this?	HW4: pi again, but with numpy arrays. Vollume of a sphere	

Statistics

class day	Day	Date	Topics	In-class	HW	Reading
10	M	9/26	Intro to Statistics	Flow control, pairs.py		Sundnes functions Ch.4 and
11	W	9/28	Functions - modular programming I	Functions, pi_functions.py	HW5: Reading and processing Iris datafile.	
	M	10/3	FALL BREAK!!!	FALL BREAK!!!	FALL BREAK!!!	FALL BREAK!!!
12	W	10/5	Classes - modular programming II	Classes - particle class	None!	Sundnes: classes Ch.8;
13	M	10/10	Plotting - matplotlib	plotting examples including Iris		Sundnes: 6.2 --> 6.5, Wood Ch. 10 Stat dists (???)
14	W	10/12	Statistical Distributions	Probability Distributions	HW6: Classes and Functions - gravity problem	
15	M	10/17	Chi^2, probablility distributions	Chi^2 distribution notebook		Wood Ch. 11. Chi^2 and fitting (???)
16	W	10/19	Fitting I	Fitting notebook	HW7: Bite simulation, Gausian	
17	M	10/24	Fitting II (fit quality)	Fitting with errors and pull distributions		None??
18	W	10/26	Fitting III	Fits with parameter errors	HW8: Simulated falling Gaussian fits	

Data Science Tools

class day	Day	Date	Topics	In-class	HW	Reading
19	M	10/31	Classification	Iris - correlation, 2D distributions		None??
20	W	11/2	Recursion/Integration	Recursion/Integration	HW9: Error addition and pair plot	
21	M	11/7	VPython??	Vpython (??)		Wood: Ch. 12
22	W	11/9	VPython??	Vpython (??)	HW10: 2D Integration, Gravity animation	
23	M	11/14	Batch Jobs	Batch Jobs - add a notebook?		For Pandas reference see VanderPlas Ch. 3. For plotting with Pandas see
24	W	11/16	Pandas	Pandas	HW11: Batch jobs	
25	M	11/21	Machine Learning I	Blobs and SVM		For Machine Learning reference see VanderPlas Ch. 5
	W	11/23	THANKSGIVING BREAK	THANKSGIVING BREAK	THANKSGIVING BREAK	
26	M	11/28	Machine Learning II	Gaussian Bayes classifier		For Machine Learning reference see VanderPlas Ch. 5
27	W	11/30	Machine Learning II	Neural Network	HW12: Pandas and ML ???	

Validation?

After teaching the class for two semesters I wanted to make sure I was getting it right…

The question: "What topics should be covered in an introductory class on Python for scientists and engineers?"

Who can answer this question?

Validation?

After teaching the class for two semesters I wanted to make sure I was getting it right…

The question: "What topics should be covered in an introductory class on Python for scientists and engineers?"

Who can answer this question?

→ ChatGPT of course!

According to ChatGPT

1. Basic Python syntax
2. NumPy: A library for numerical computing in Python - arrays and linear algebra
3. Matplotlib: Plotting!
4. Pandas: Data analysis and manipulation, including reading in data and basic statistics
5. SciPy: A library for scientific computing
6. Object-oriented programming: An introduction to classes, objects, and methods
7. Data visualization: Creating effective and informative visualizations
8. Scientific computing: Using Python for numerical simulations, solving differential equations, and other common scientific computing tasks
9. Debugging and error handling: Strategies for identifying and fixing bugs
10. Advanced topics: Depending on the interests of the class, more advanced topics could be covered, such as machine learning with scikit-learn or TensorFlow

Final exam/project

- I gave them a new data file (from the Sloan Digital Sky Survey) in CVS format and had them:
  - Read in the data, and print out a summary table
  - Plot/fit various distributions
  - Study the quality of the fit
  - 1D-classification and confusion matrix
  - Multi-D “pair-plot” with feature comparison
  - Reduce/simplify the dataset
  - Train and assess a Neural Network

  → I was very impressed by what most students could do on their own!

Student Feedback

- ~90% of the class seemed to like it and felt like they learned a great deal! ☺

Student Feedback

- ~90% of the class seemed to like it and felt like they learned a great deal! ☺
- ~10% of the class felt like I didn't teach them anything and hated it! ☹
  - I get the feeling this is often the case for a "flipped" classroom.
  - After all, I did not teach them to program. I just gave them the exercises that allowed them to learn to do it. And, I was there to clear things up.
- Not sure it is possible to fix this?
- Several students have come back to tell me how useful the class was when they started doing research. ☺

School of Data Science

- I met with the leadership of the new School of Data Science at UVA.
- They will:
  - Count this new python course in the physics department toward their currently-offered minor.
  - Count this course toward their data science major (under planning at the moment)
- Hopefully many students of science and engineering will also decide to get a data-science minor after taking this course.
- And hopefully some data-science students will decide to take some physics after this course.

Future: Scientific Computing at UVA

- **Introduction to Python for Scientist and Engineers**
  - Physics department decided to make this the minimal computing requirement for BS majors.
- Computational Physics I
  - The current 2000-level class will be combined with the first of the 5000-level class and re-branded as a 3000/5000-level class.
  - Students can take this after the python class for an elective, focus on computation physics, or **can start here if they already have advanced computing skills**
- Computational Physics (5000-level)
  - One semester sequence of more advanced techniques (some C, python, and C++)

Outlook

- This is my third semester, and I think it is going well!
- We are attempting to scale this class up to <99 students in an active-learning classroom.
- So far:
  - Spring 2022: 20
  - Fall 2022: 40 → ☺
  - Spring 2023: 20 → ☹
    - Poor time slot selected for class
    - But, this allowed me to spend extra time improving
  - Fall 2023: ??