

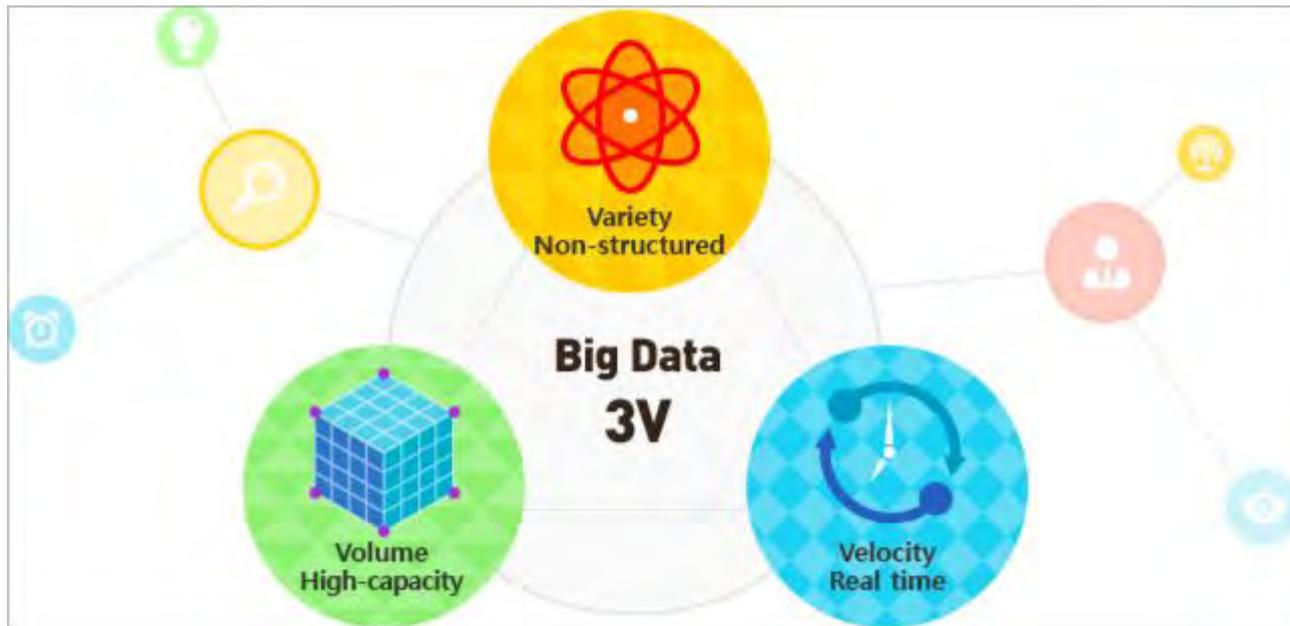
# The Carpentries:

Software, Data and Library Carpentry

Rabea Müller | 2021-09-16



Skills and perspectives to work with software and data are increasingly important as we generate more data.

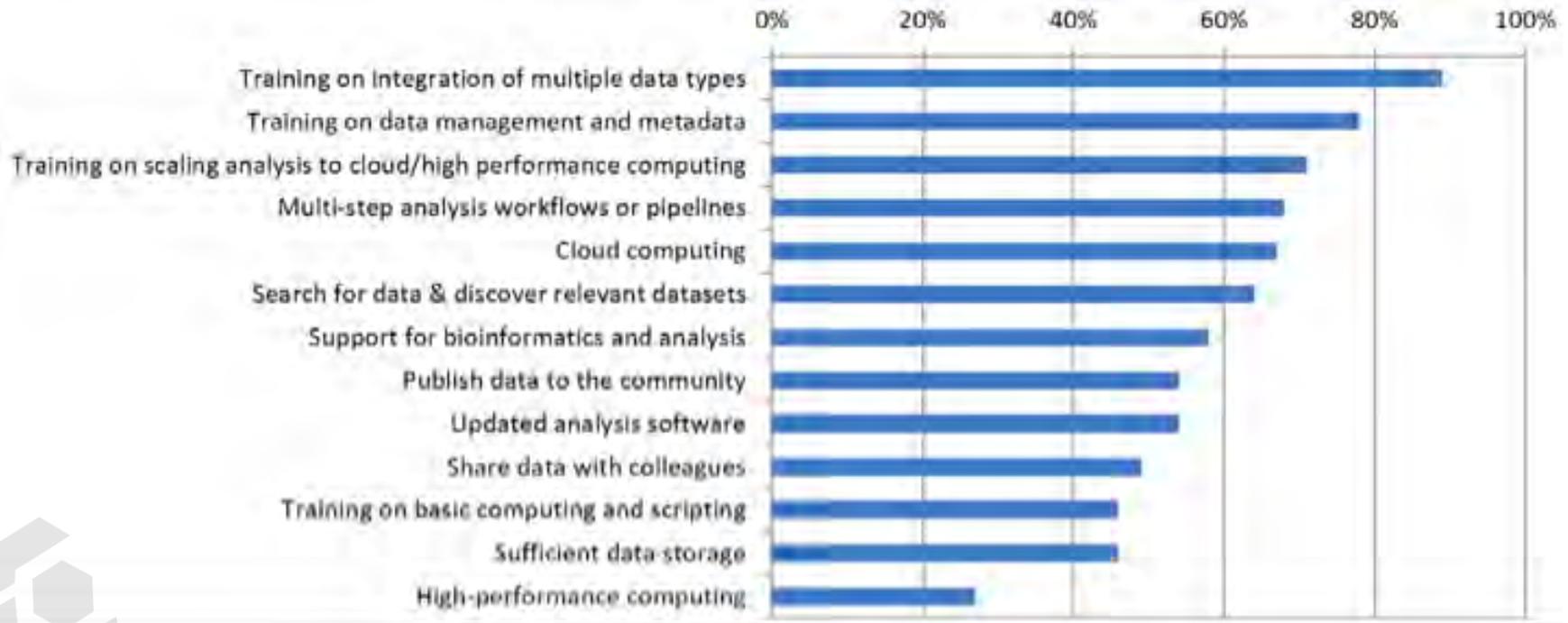


# Researchers are very interested in learning these skills

*Survey by Bioinformatics Resource Australia on what it would be most useful for them to offer*



# Current Unmet Needs



Barone L, Williams J and Micklos D. Unmet Needs for Analyzing Biological Big Data: A Survey of 704 NSF Principal Investigators (2017)



---

**How do we scale data and software skills  
along with data production?**

---



# Building Skills and Community

- Creating training ‘in the gaps’ that is accessible, approachable, aligned and applicable
- Peer-led hands-on intensive workshops
- Volunteer instructors
- Open and collaborative lesson materials
- Creating and supporting community





# THE CARPENTRIES

Non-profit organization that:

- Trains people in software development and data science skills for more effective work and career development
- Builds community and local capacity for teaching and learning these skills and perspectives



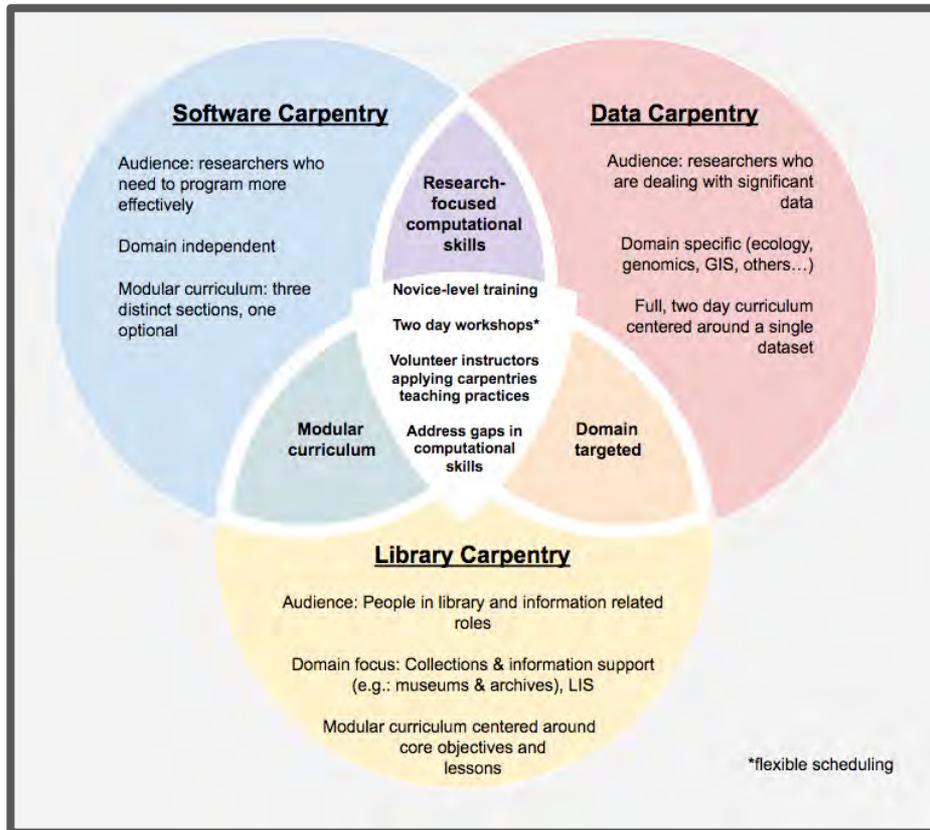
# Carpentries?



# Workshops

- 2-days, active learning
- Feedback to learners throughout the workshop
- Trained instructors
- Friendly learning environment





# Software Carpentry Lessons

- Software development best practices
  - Command line
  - Version control with github
  - Programming in Python or R



# Data Carpentry Lessons



- Working effectively with data and includes domain-specific content
- Domain content
  - **Ecology:** working with tabular data, ecological data
  - **Genomics:** cloud computing, genomic data organization, working with bioinformatics tools at the command line
  - **Geospatial:** organizing and working with geospatial data in R
  - **Social science:** tabular data with social science data
  - More in development



# Library Carpentry Core Objectives



Library Carpentry workshops teach people working in library- and information-related roles how to:

- Cut through the jargon terms and phrases of software development and data science and apply concepts from these fields in library tasks;
- Identify and use best practice in data structures;
- Learn how to programmatically transform and map data from one form to another;
- Work effectively with researchers, IT, and systems colleagues;
- Automate repetitive, error prone tasks.



# Workshop goals

- Teach skills
- Get people started and introduce them to what's possible
- Build confidence in using these skills
- Encourage people to continue learning
- Positive learning experience



# Instructors

Instructor training program that teaches educational pedagogy.  
How to teach generally as well as for Carpentries workshops.

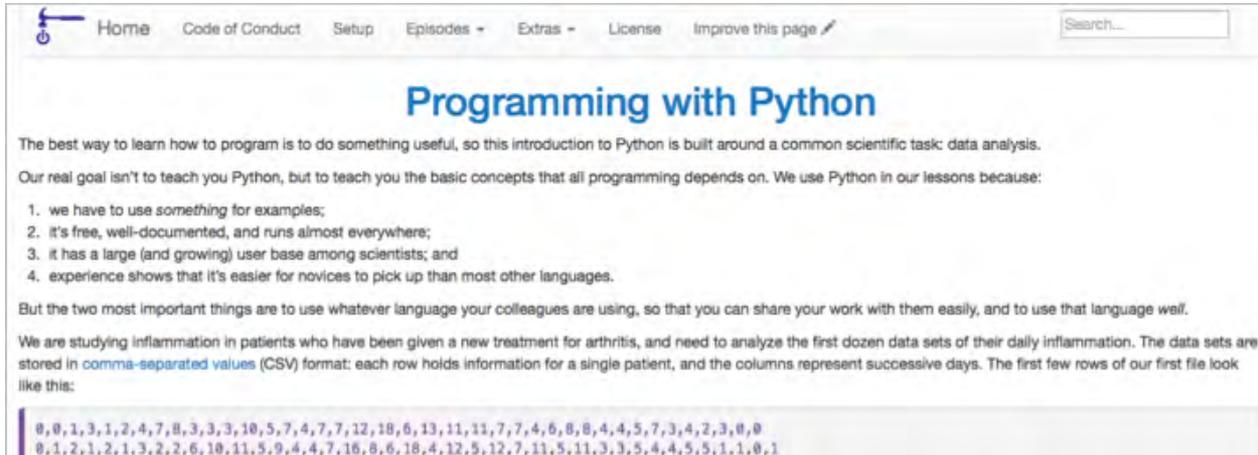
<http://carpentries.github.io/instructor-training/>

Over 1,000 volunteer instructors on 6 continents



# Curriculum

- Open and collaboratively developed
- Continual improvement and up-to-date



Home Code of Conduct Setup Episodes - Extras - License Improve this page

## Programming with Python

The best way to learn how to program is to do something useful, so this introduction to Python is built around a common scientific task: data analysis.

Our real goal isn't to teach you Python, but to teach you the basic concepts that all programming depends on. We use Python in our lessons because:

1. we have to use *something* for examples;
2. it's free, well-documented, and runs almost everywhere;
3. it has a large (and growing) user base among scientists; and
4. experience shows that it's easier for novices to pick up than most other languages.

But the two most important things are to use whatever language your colleagues are using, so that you can share your work with them easily, and to use that language well.

We are studying inflammation in patients who have been given a new treatment for arthritis, and need to analyze the first dozen data sets of their daily inflammation. The data sets are stored in [comma-separated values](#) (CSV) format: each row holds information for a single patient, and the columns represent successive days. The first few rows of our first file look like this:

```
0,0,1,3,1,2,4,7,8,3,3,3,10,5,7,4,7,7,12,18,6,13,11,11,7,7,4,6,8,8,4,4,5,7,3,4,2,3,0,0
0,1,2,1,2,1,3,2,2,6,10,11,5,9,4,4,7,16,8,6,18,4,12,5,12,7,11,5,11,3,3,5,4,4,5,5,1,1,0,1
```



# Community

A group of people excited about software and data skills and about sharing them with others

- Mentoring program and instructor onboarding
- Discussion groups and community calls
- Email lists
- Teaching at other institutions

Connecting with The Carpentries:

<https://carpentries.org/connect/>



# Outcomes

Short and long term surveys show that people are learning the skills, putting them into practice in their work and have more confidence in their ability to do computational work.

The tools I learned in my Carpentry workshop:

*“helped me to reshape my workflow into a far more efficient and robust process.”*

*“are improving my ability to share data and code.”*

*“helped facilitate my understanding of the problems and solutions to accessing and transforming data.”*

*“[are] useful tools for training my own team.”*

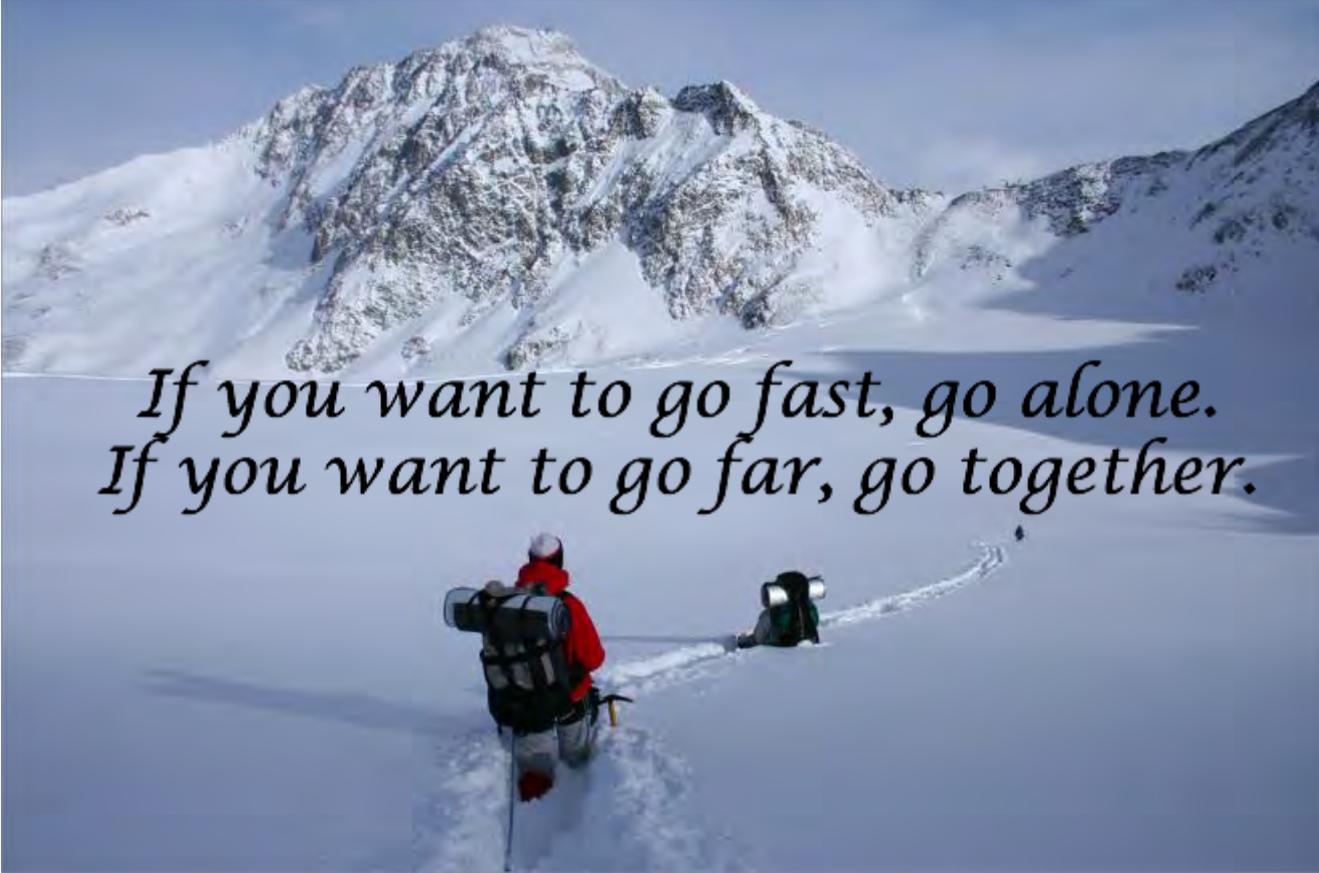


# People like the workshops



Promoter Score	n	%
Detractor	14	5.511811
Passive	49	19.291339
Promoter	191	75.196850





*If you want to go fast, go alone.  
If you want to go far, go together.*



# Thank you!



## Support



## Socials

Twitter

@MueRabea

Email

muellerr@zbmed.de