THE CARPENTRIES

The Carpentries - Building data skills and community

Open Science Fair 2019, Porto

Birgit Schmidt, Daniel Bangert
University of Göttingen

SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

THE CARPENTRIES

- Train people in **foundational** computational and data science skills for more effective work and career development
- Build community and local capacity for teaching and learning these skills and perspectives

https://carpentries.org/

For a very brief history of The Carpentries see https://twitter.com/raynamharris/status/1062516187874189312 (image)

THE CARPENTRIES

76 member organisations
https://carpentries.org/members/

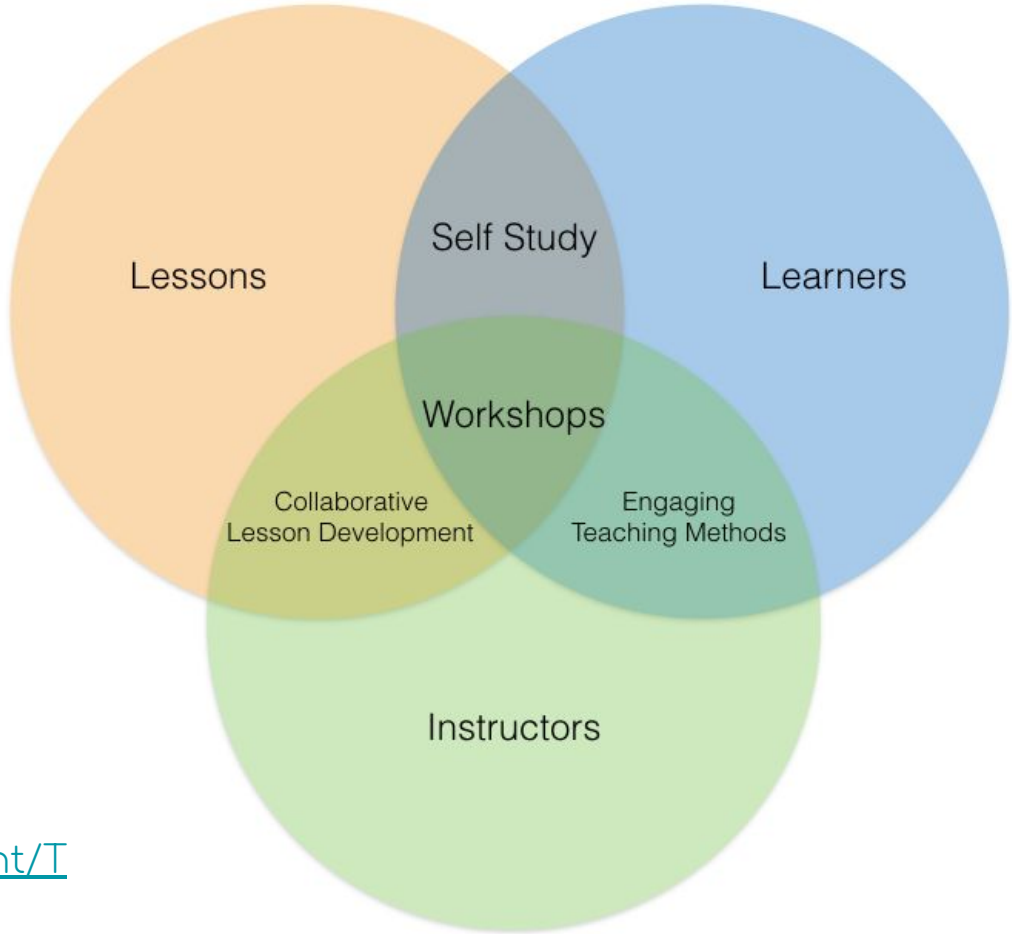
38K learners reached

1.7K workshops

1.6K instructors trained

Workshops on **7 continents**

Source:
https://carpentries.org/files/assessment/TheCarpentries2018AnnualReport.pdf



Building skills and community

- Creating training ‘in the gaps’
- Peer-led, hands-on intensive **workshops**
- Open, collaboratively-developed **lesson materials**
- Creating and supporting **community**

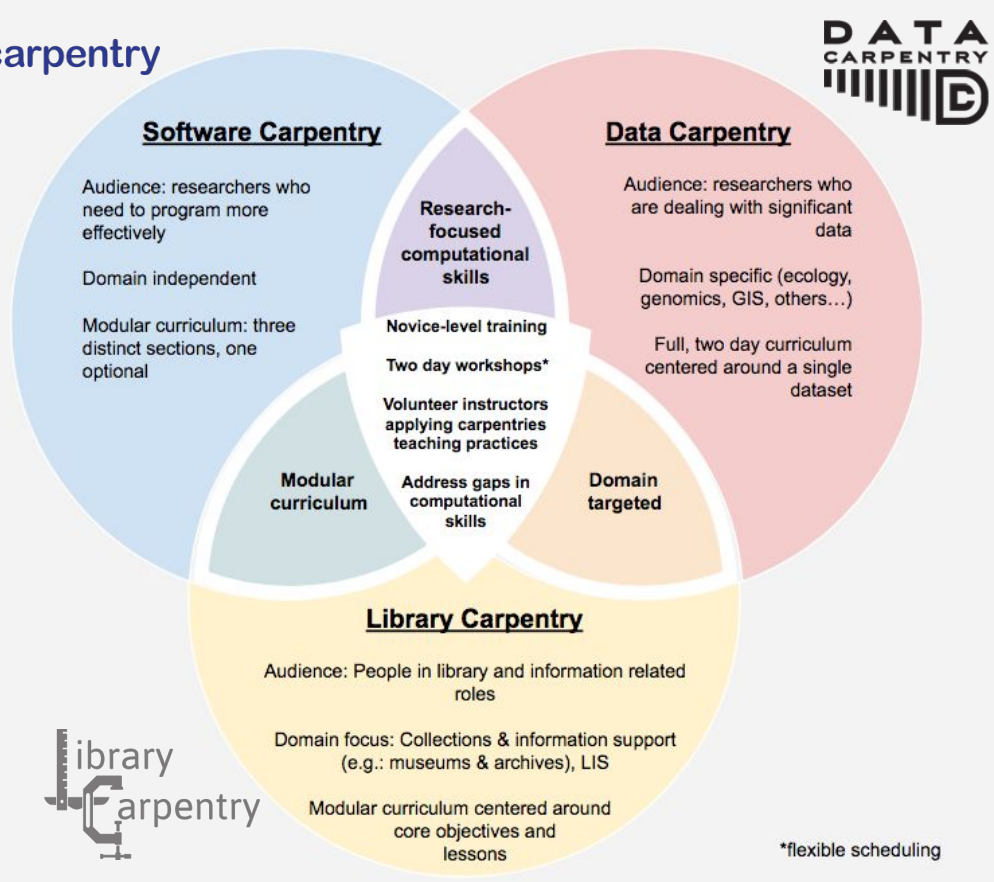
https://carpentries.org/community/

Workshops

- Two-days, active learning
- Welcoming and friendly learning environment
- Trained, volunteer instructors
- Participatory live coding
- Continuous feedback
- Easy to get help
- Collaborative notes
- Pre- and post-workshop surveys

https://carpentries.org/workshops/

Lesson programs



Library Carpentry

Core Curriculum

Data intro for librarians
An introduction to data structures, regular expressions, and computing terms

Unix Shell
An introduction to command line interfaces and task automation using the Unix shell

OpenRefine
An introduction to cleaning up and enhancing a dataset using OpenRefine

Git Intro for Librarians
An introduction to version control using Git and GitHub for collaboration

Extended Curriculum (Beta/Alpha)

SQL for Librarians
An introduction to relational database management using the SQLite tool

Webscraping
An introduction to extracting structured data from websites suing a range of tools

Tidy data for librarians
An introduction to good data organisation, which is the foundation of much of our day-to-day work in libraries.

Introduction to Python
An introduction to Python, a general purpose programming language

Data Intro for Archivists
An introduction to data structures, regular expressions, and computing terms for archivists

Instructors

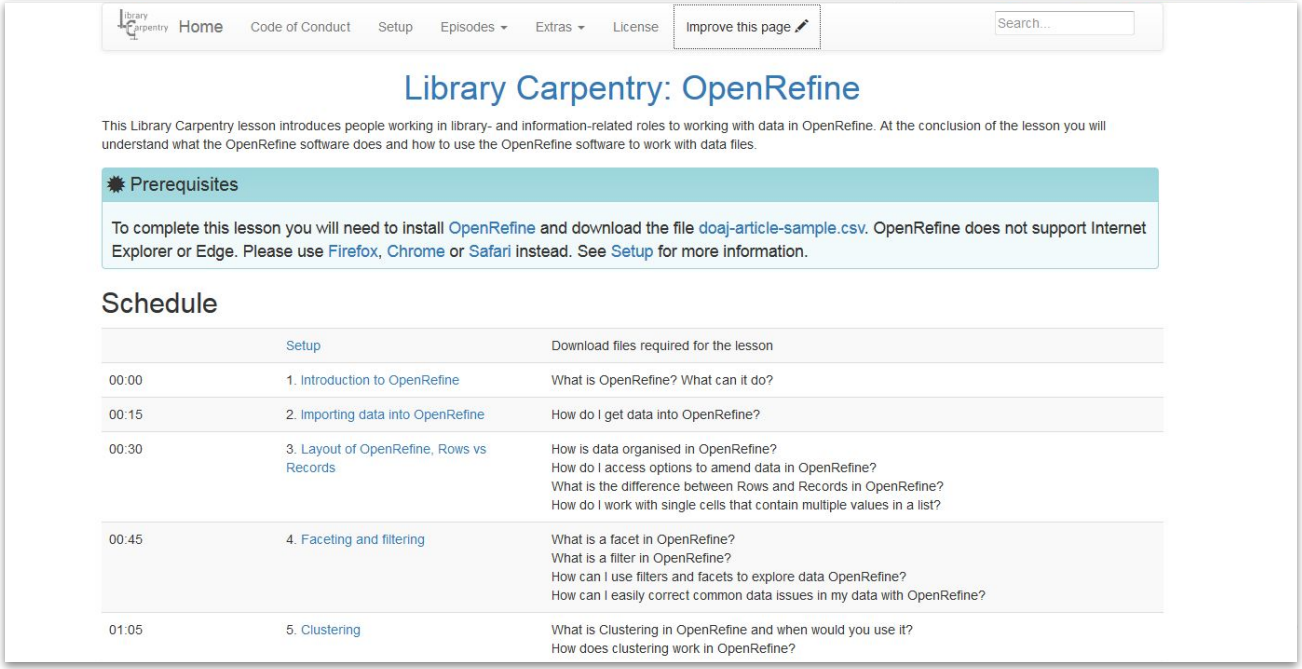
Instructor training program that teaches educational pedagogy. How to teach generally as well as for Carpentries workshops.

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 | Search...

Library Carpentry: OpenRefine

This Library Carpentry lesson introduces people working in library- and information-related roles to working with data in OpenRefine. At the conclusion of the lesson you will understand what the OpenRefine software does and how to use the OpenRefine software to work with data files.

Prerequisites

To complete this lesson you will need to install OpenRefine and download the file doaj-article-sample.csv. OpenRefine does not support Internet Explorer or Edge. Please use Firefox, Chrome or Safari instead. See Setup for more information.

Schedule

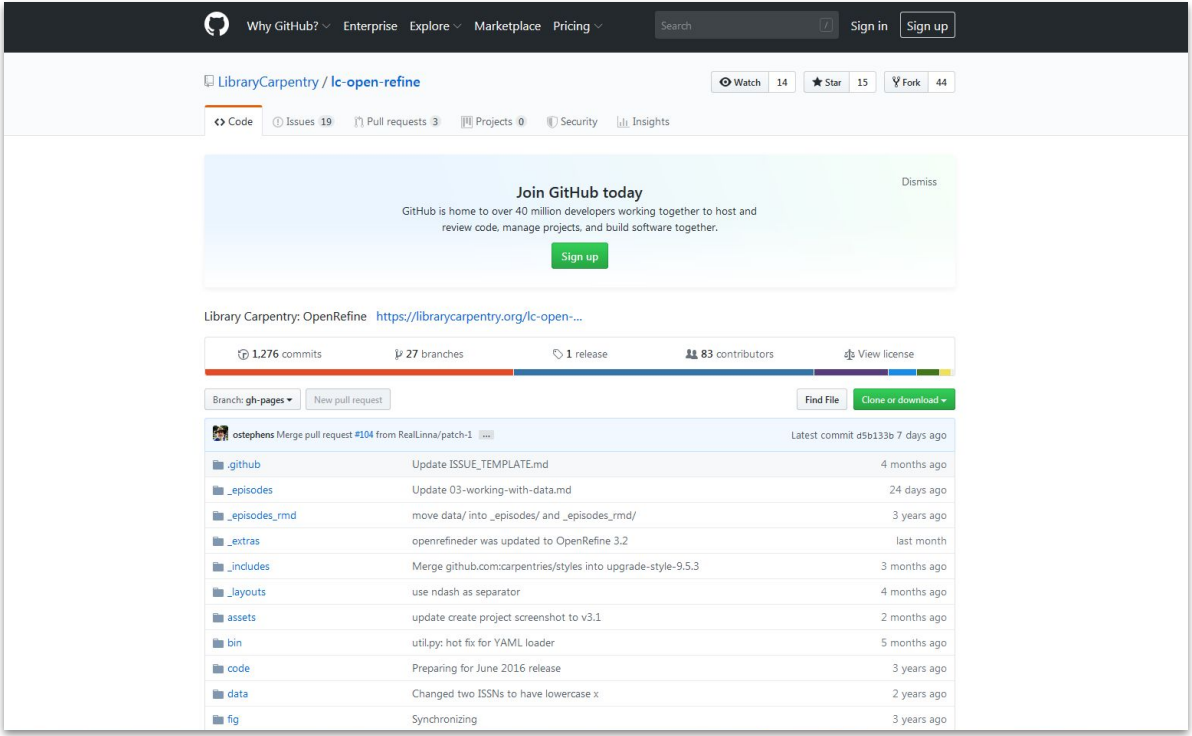
	Setup	Download files required for the lesson
00:00	1. Introduction to OpenRefine	What is OpenRefine? What can it do?
00:15	2. Importing data into OpenRefine	How do I get data into OpenRefine?
00:30	3. Layout of OpenRefine, Rows vs Records	How is data organised in OpenRefine? How do I access options to amend data in OpenRefine? What is the difference between Rows and Records in OpenRefine? How do I work with single cells that contain multiple values in a list?
00:45	4. Faceting and filtering	What is a facet in OpenRefine? What is a filter in OpenRefine? How can I use filters and facets to explore data OpenRefine? How can I easily correct common data issues in my data with OpenRefine?
01:05	5. Clustering	What is Clustering in OpenRefine and when would you use it? How does clustering work in OpenRefine?

Curriculum Development Process

In the process of developing infrastructure and guidelines to support more lesson development.

- Identifying needs for content
- Identifying learning goals and objectives
- Content development and assessment

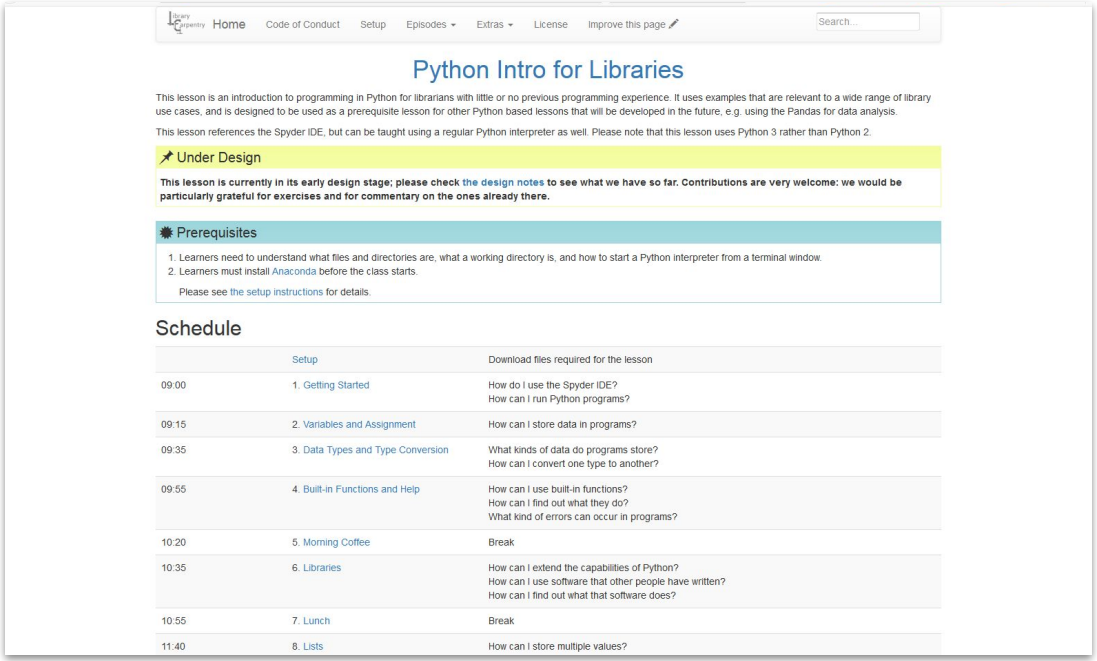
One repository per lesson



Stages of lesson development

Lessons are in various stages of development - stable, beta, alpha, and conceptual.

E.g. Python for Libraries (alpha)



Home Code of Conduct Setup Episodes Extras License Improve this page Search...

Python Intro for Libraries

This lesson is an introduction to programming in Python for librarians with little or no previous programming experience. It uses examples that are relevant to a wide range of library use cases, and is designed to be used as a prerequisite lesson for other Python based lessons that will be developed in the future, e.g. using the Pandas for data analysis.

This lesson references the Spyder IDE, but can be taught using a regular Python interpreter as well. Please note that this lesson uses Python 3 rather than Python 2.

Under Design

This lesson is currently in its early design stage; please check the design notes to see what we have so far. Contributions are very welcome: we would be particularly grateful for exercises and for commentary on the ones already there.

Prerequisites

1. Learners need to understand what files and directories are, what a working directory is, and how to start a Python interpreter from a terminal window.
2. Learners must install Anaconda before the class starts.

   Please see the setup instructions for details.

Schedule

	Setup	Download files required for the lesson
09:00	1. Getting Started	How do I use the Spyder IDE? How can I run Python programs?
09:15	2. Variables and Assignment	How can I store data in programs?
09:35	3. Data Types and Type Conversion	What kinds of data do programs store? How can I convert one type to another?
09:55	4. Built-in Functions and Help	How can I use built-in functions? How can I find out what they do? What kind of errors can occur in programs?
10:20	5. Morning Coffee	Break
10:35	6. Libraries	How can I extend the capabilities of Python? How can I use software that other people have written? How can I find out what that software does?
10:55	7. Lunch	Break
11:40	8. Lists	How can I store multiple values?

Top 10 FAIR Data & Software Things

Global sprints in Nov 2018 & June 2019 - disciplines covered, cf. https://librarycarpentry.org/Top-10-FAIR/

The Top 10 FAIR Data & Software Things are brief guides (stand alone, self paced training materials), called "Things", that can be used by the research community to understand FAIR in different contexts but also as starting points for conversations around FAIR.

- Oceanography
- Research Software
- Research Libraries
- Research Data Management Support
- International Relations
- Humanities: Historical Research
- Geoscience
- Biomedical Data Producers, Stewards, and Funders
- Biodiversity
- Australian Government Data/Collections
- Archaeology
- Music

Zenodo version: http://doi.org/10.5281/zenodo.3409968

Impact

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:

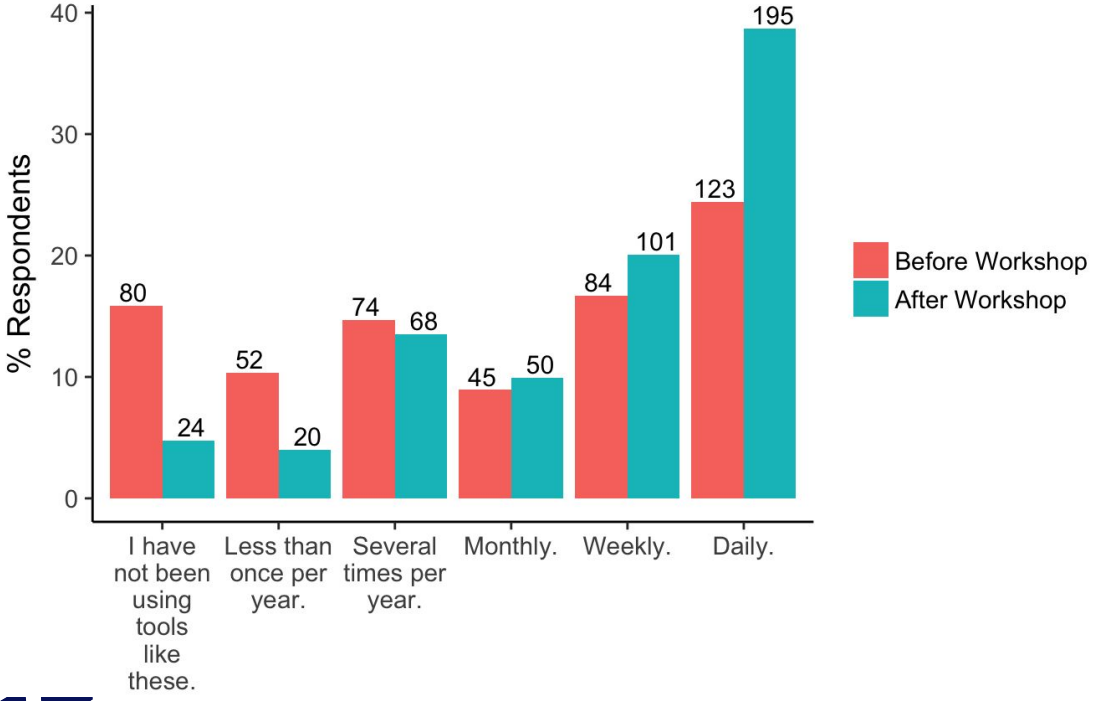
"helped me to reshape my workflow into a far more efficient and robust process."

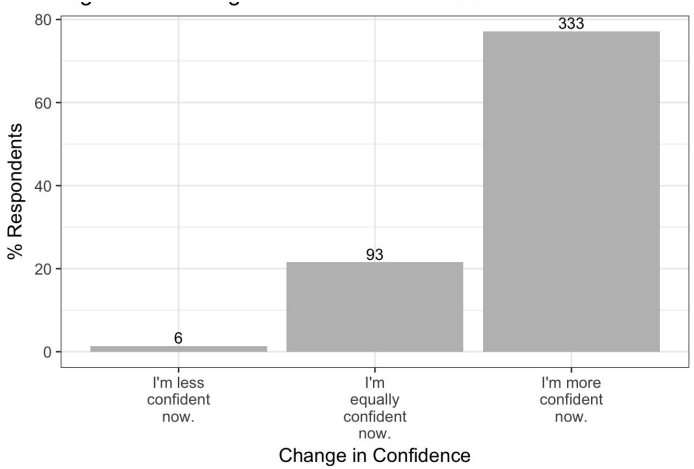
"are improving my ability to share data and code."

"helped facilitate my understanding of the problems and solutions to accessing and transforming data."

https://carpentries.org/assessment/

Skill usage and confidence persists





Questions?

Sources and further reading:

The Carpentries
Library Carpentry
The Carpentries Handbook presentations