

University of Massachusetts Medical School

eScholarship@UMMS

National Network of Libraries of Medicine New
England Region (NNLM NER) Repository

National Network of Libraries of Medicine New
England Region

2019-03-29

Data Science Training for the Future: Building a Carpentries Consortium

Julie Goldman
Harvard Medical School

Et al.

Let us know how access to this document benefits you.

Follow this and additional works at: <https://escholarship.umassmed.edu/ner>



Part of the [Library and Information Science Commons](#), and the [Public Health Commons](#)

Repository Citation

Goldman J, New England Software Carpentry Library Consortium. (2019). Data Science Training for the Future: Building a Carpentries Consortium. National Network of Libraries of Medicine New England Region (NNLM NER) Repository. <https://doi.org/10.13028/q2bx-q484>. Retrieved from <https://escholarship.umassmed.edu/ner/68>

Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in National Network of Libraries of Medicine New England Region (NNLM NER) Repository by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

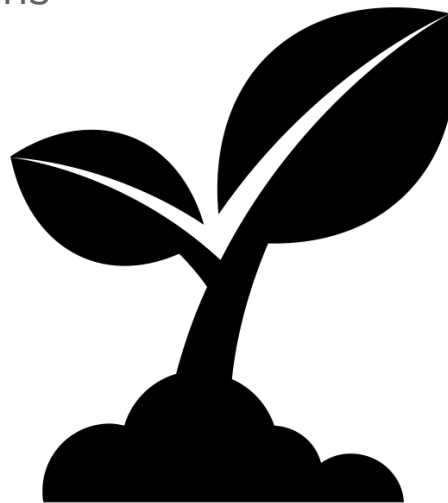
Data Science Training for the Future

Building a Carpentries Consortium

Julie Goldman on behalf of the
New England Software Carpentry Library Consortium
NNLM NER eScience Forum | March 29, 2019

Community of Practice

- Desire for more advanced/technical training for librarians
 - For teaching students
 - For librarians to gain skills too
- Wanted training to teach coding and data skills
 - Carpentries offered this opportunity
 - Long waiting list for training
 - Shorter with a membership
- Expensive to go alone
 - Less risk as a group
- Conversation at RDAP 2017
 - Could we bring together institutional contributions?
 - What would be the benefits to our institutions?



Created by Vectors Market
from Noun Project

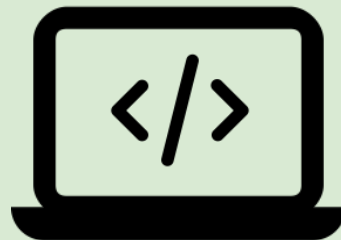
The Carpentries



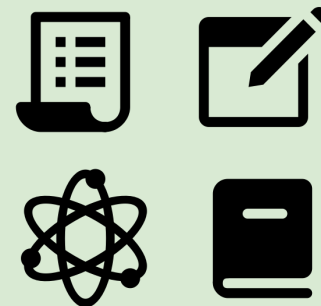
Non-profit
organization and
international
community of
volunteers



that develop
lessons, train
instructors,
and organize
workshops



in **data**
management
and **software**
development
best practices



for and by
researchers and
librarians across
disciplines

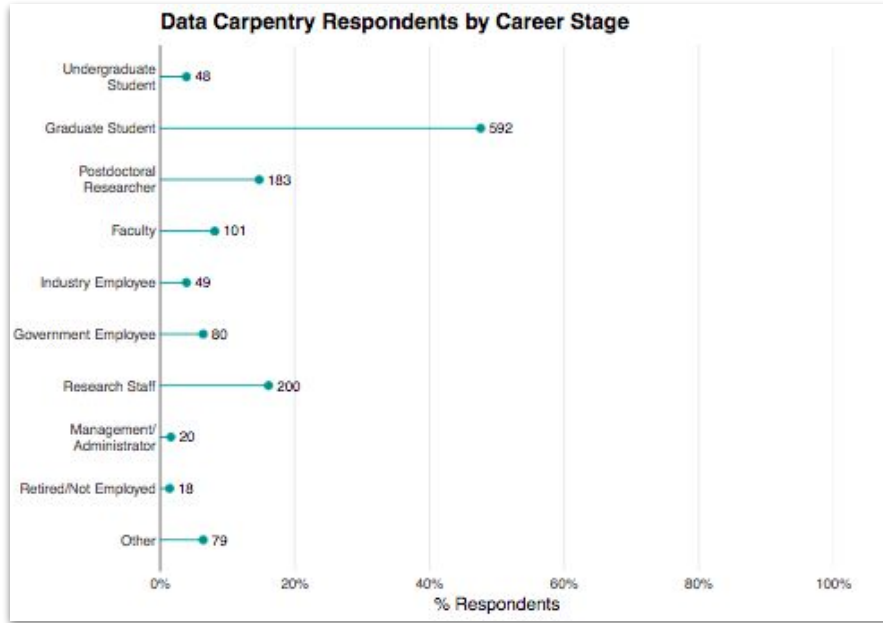
Workshops

- 2-days, active/hands-on learning
- Trained/certified instructors
- Feedback to learners throughout the workshop (Surveys, post-its, +/-)
- Friendly learning environment (Code of Conduct)



Who takes workshops?

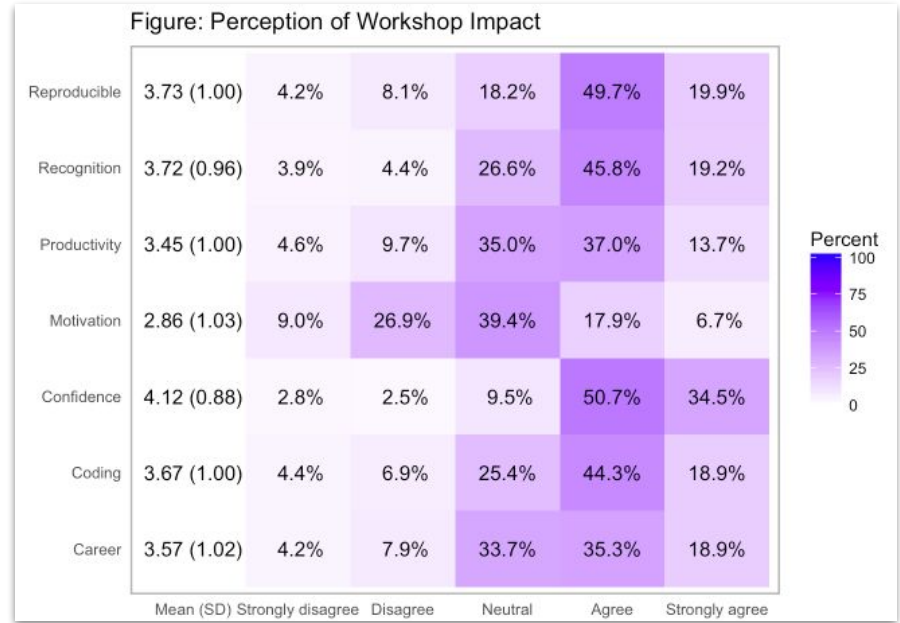
66% of Data Carpentry workshop attendees are early career



Analysis of Software and Data Carpentry's Pre- and Post-Workshop Surveys
<https://doi.org/10.5281/zenodo.1325463>

What's the impact?

85% report they gained more confidence after a workshop



Analysis of the Carpentries Long-Term Feedback Survey: Perception of Workshop Impact
https://carpentries.github.io/assessment/learner-assessment/archives/2018/code/2018_January_long_term_report.html

Forming a Consortium

New England Software Carpentry Library Consortium (NESCLiC)

- Gathered a group of like minded individuals
- Library and IT staff from academic institutions

Administrative Steps

- Written proposals for each school
- Signing/Billing Institution & Letters of Intent
- Gold Tier Membership
 - \$15,000 per year
 - \$1000 per instructor trained



■ Year 1

■ Year 2

NESCLiC Members



Andrew Creamer
Bruce Boucek
Patrick Rashleigh

Kelsey Sawyer
Ashley Champagne



Kristin Lee
Ari Gofman
Kyle Monahan

Susan Remondi



Jennifer Chaput



Lora Leligdon
Christian Darabos
James Adams

Pamela Bagley
Stephen Gaughan



Joshua Dull
Catherine DeRose
Kate Nyhan

David Cirella
Barbara Esty
Sawyer Newman



Thea Atwood

Ann Kardos
Erin Jerome



Julie Goldman
Ceilyn Boyd

Meghan Kerr



Sarah Oelker
James Burke



Sally Gore
Tess Grynoch

Instructor Training

- Virtual instruction in two locations
 - Brown & UMass Amherst
- Introduce you to evidence-based best-practices of teaching
 - A different way to look at pedagogy!
- Create a positive environment for learners at your workshops
- Provide opportunities for you to practice and build your teaching skills
- Help you become integrated into the Carpentries community



Example Lesson: *OpenRefine*

Library Carpentry **OpenRefine**

Introduction to OpenRefine

Overview

Teaching: 15 min
Exercises: 0 min

Questions

- What is OpenRefine? What can it do?

Objectives

- Explain what the OpenRefine software does
- Explain how the OpenRefine software can help work with data files

What is OpenRefine?

OpenRefine is described as "a power tool for working with messy data" [David Huynh](#) - but what does this mean? It is probably easiest to describe the kinds of data OpenRefine is good at working with and the sorts of problems it can help you solve.

OpenRefine is most useful where you have data in a simple tabular format such as a spreadsheet, a comma separated values file (csv) or a tab delimited file (tsv) but with internal inconsistencies either in data formats, or where data appears, or in terminology used. OpenRefine can be used to standardize and clean data across your file. It can help you:

- Get an overview of a data set
- Resolve inconsistencies in a data set, for example standardizing date formatting
- Help you split data up into more granular parts, for example splitting up cells with multiple authors into separate cells
- Match local data up to other data sets, for example in matching local subjects against the Library of Congress Subject Headings
- Enhance a data set with data from other sources

Some common scenarios might be:

- Where you want to know how many times a particular value (name, publisher, subject) appears in a column in your data
- Where you want to know how values are distributed across your whole data set
- Where you have a list of dates which are formatted in different ways, and want to change all the dates in the list to a single common date format. For example:

Data you have	Desired data
1st January 2014	2014-01-01
01/01/2014	2014-01-01
Jan 1 2014	2014-01-01
2014-01-01	2014-01-01

- Where you have a list of names or terms that differ from each other but refer to the same people, places or concepts. For example:

Data you have	Desired data
London	London
London!	London

The screenshot shows the OpenRefine interface with a dropdown menu set to 'chemical elements'. A list of 23 elements is displayed, each with a star icon and a speech bubble icon. The elements listed are: 1. hydrogen, 2. helium, 3. lithium, 4. beryllium, 5. boron, 6. carbon, 7. nitrogen, 8. oxygen, 9. fluorine, 10. neon, 11. sodium, 12. magnesium, 13. aluminum, 14. silicon, 15. francium, 16. phosphorus, 17. iron, 18. sulfur, 19. chlorine, 20. argon, 21. potassium, 22. calcium, and 23. lead. Each element name has a 'Choose new match' link below it.

Unexpected Benefits

- Not just instructor training
 - Keep data/coding skills sharp
 - A group of people who help each other
- Expanded Local Network
 - Communicate and collaborate on other projects
- Collaboration with other regional organizations
 - NEASIST
 - NNLM New England Region



Created by Laili Hidayati
from Noun Project

Moving Forward



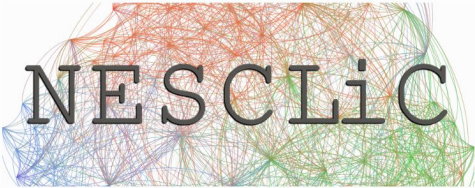
Created by Gregor Greenar
from Noun Project

- The Carpentries
 - Evolving as an organization
- Changing roles within libraries
 - Evolution of Library Carpentry
 - More 'non-traditional' library training
- Train-the-Trainer
 - Train others in Carpentries pedagogy
- Further participation in lesson development & curriculum areas
 - Becoming lesson maintainers
 - Digital Humanities at Yale
 - Biodiversity at Harvard

nesclio.github.io/home

NESCLIC

About Members Publications Workshops



The **New England Software Carpentry Library Consortium** or **NESCLIC** is an association of academic libraries, joined as a community of practice focused on building data science skills in research computing, and extracting, wrangling, storing, analyzing, and visualizing data. NESCLIC allows member institutions to share in a Gold-level membership with the Carpentries.

Our Model

NESCLIC member institutions share a Gold-level Carpentries membership, which affords NESCLIC 6 free coordinated workshops, a 50% discount on the cost for additional coordinated workshops, no charge for self-organized workshops, and fifteen instructors trained, per year.

NESCLIC provides the opportunity for members to:

- Develop instructional materials
- Build a network of data fluent researchers and instructors
- Refine skills by participating as instructors at other member institutions

Financial commitment: Cost of shared Carpentry membership (\$1000/person) plus training expenses (travel or hosting)

Time commitment: (2) – two-day training workshops plus committee work

NESCLIC has established bylaws and members contribute to the consortium by participating in one of several roles:

- Carpentries liaison
- Membership/Recruiting
- Assessment
- Outreach
- Workshop coordination

Become a Member

Interested in joining NESCLIC?
Email: nesclio@googlegroups.com

Workshops

Check back for future workshops offered by NESCLIC members.

This list includes official Carpentries' workshops and workshops adapted from the Carpentries' materials where NESCLIC members served as hosts, instructors, or helpers.

Venue	Dates	Lessons
Tufts University	April 12-13, 2019	Unix, Git, Programming with Python
Tufts University	February 15-16, 2019	Data Organization in Spreadsheets, OpenRefine for Data Cleaning, R for Social Scientists
Brown University	January 16-17, 2019	Data Organization in Spreadsheets, OpenRefine for Data Cleaning, Introduction to R, Data Analysis and Visualization in R, Data Management with SQL
Harvard University	January 15-16, 2019	Data Intro for Librarians, Shell Lessons for Librarians, Tidy Data for Librarians, OpenRefine for Librarians
UMass Amherst	January 8-9, 2019	Unix, Python, Git
Dartmouth College	November 29-30, 2018	Unix, Git, Python
Tufts University	November 9-10, 2018	Unix, Git, Programming with Python
Brown University Library (Hosts & Helpers), Yale and Tufts (Instructors)	October 22-23, 2018	Data Intro for Librarians, Shell Lesson for Librarians, Git Intro for Librarians, OpenRefine for Librarians
Yale University Library	August 15-16, 2018	Tidy Data for Librarians, Shell Lesson for Librarians, OpenRefine for Librarians, Python Intro for Librarians