

Data Science Training for All

Sharing the Cost (and Benefits) of a Carpentries Membership

Kristin Lee | Tufts University
Joshua Dull | Yale University

Who are we?

Kristin Lee

Research Data Librarian

Tisch Library,
Tufts University



Joshua Dull

Digital Scholarship
Specialist

Yale University Library



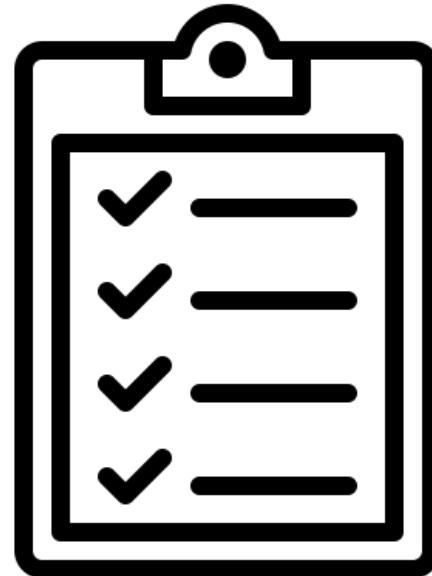
Etherpad

For collaborative notes and questions:

<https://pad.carpentries.org/2019-03-05-erl>

Agenda

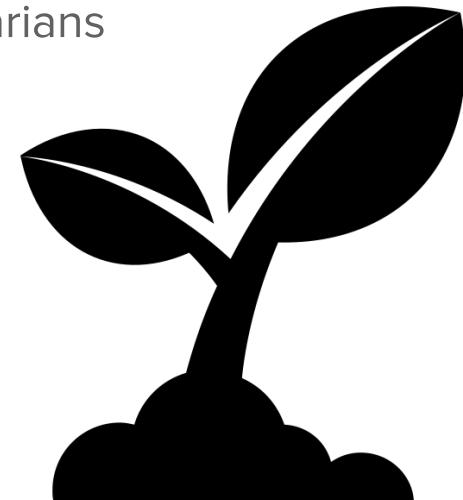
1. Origin Story
2. Year One
3. Year Two
4. The Future?
5. Hands-On Activity



Created by Setyo Ari Wibowo
from Noun Project

Another Origin Story...

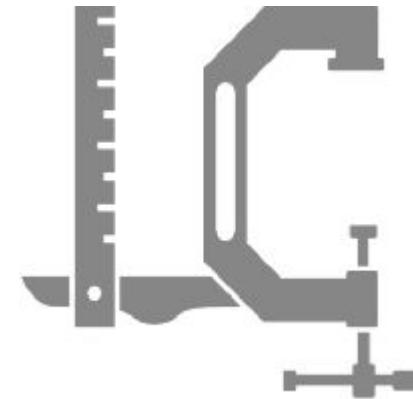
- No radioactive spiders, just an idea
- Desire for more advanced/technical training for librarians
 - For teaching students
 - For librarians to have the skills too
- Wanted training to teach coding and data skills
 - Carpentries offered this opportunity
 - Long waiting list for training
 - Shorter with a membership
- Expensive to go alone
 - Less risk as a group
- Conversation at RDAP
 - Could we all join together?



Created by Vectors Market
from Noun Project

What are The Carpentries?

- “The Carpentries teach foundational coding, and data science skills to researchers worldwide.”
 - Software Carpentry, Data Carpentry & Library Carpentry
 - Merged into one organization over 2018
- <https://carpentries.org/>



Year One: Timeline



MIT Workshop

NEASIST & MIT host a Library Carpentry workshop

NESCLiC Formed

First NESCLiC meeting held; first-year schools chosen

Joined Carpentries

NESCLiC officially joins Software Carpentry as a member organization

1st Workshop

NESCLiC members attend SWC workshop at Tufts

Instructor Training

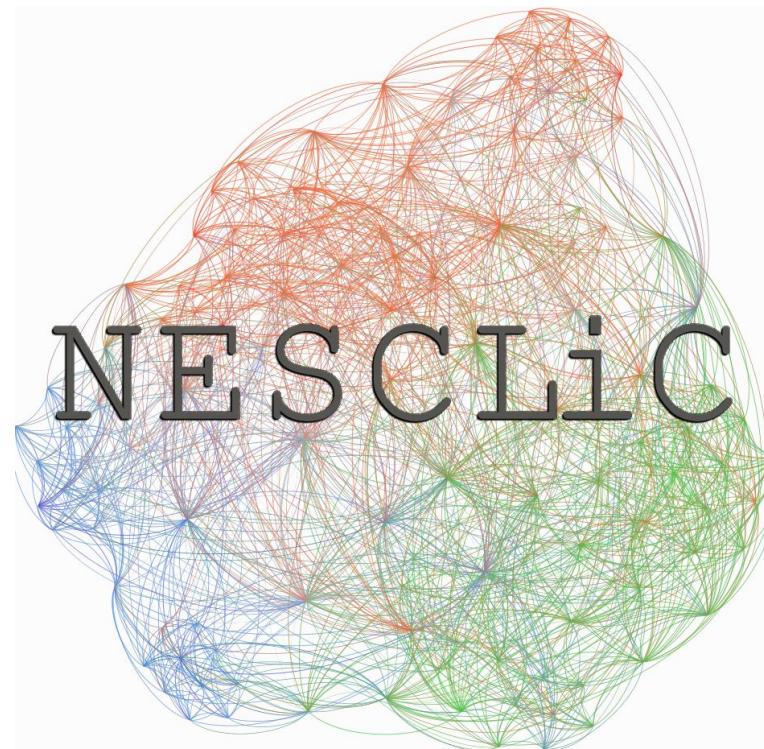
NESCLiC members attend Instructor Training at Brown & Amherst

Year Two

NESCLiC began planning year-two; reached out to new schools for interest

Year One: 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 One: NESCLiC Members



BROWN

Andrew Creamer

Bruce Boucek

Patrick Rashleigh



DARTMOUTH

Lora Leligdon

Christian Darabos

James Adams

Julie Goldman

Ceilyn Boyd

HARVARD
LIBRARY



Kristin Lee

Ari Gofman

Kyle Monahan



Thea Atwood



Joshua Dull

Catherine DeRose

Kate Nyhan

Sarah Oelker

James Burke



Year One: Our First Workshop

- Got together to take a workshop
 - Led by James Adams from Dartmouth
 - Hosted by Tufts
 - Costs distributed between Tufts, Harvard, and Dartmouth
 - Participants brought snacks and coffee to share
- Standard Software Carpentry curriculum
 - Shell, Git, and Python
 - Learners helped each other



Year One: Instructor Training

- Carpentries' Instructor training
 - A different way to look at pedagogy!
 - 2-Days
- Two locations
 - Brown & UMass Amherst
 - Virtual instruction



Year One: Instructor Training

- Introduce you to evidence-based best-practices of teaching
- Teach you how to 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



Year One: Workshops Around New England

- Between October 1, 2017 and October 1, 2018:
 - Participated in/hosted 6 workshops
 - 2 Library Carpentry
 - 4 Software Carpentry
 - ~ 110 total attendees



Home



Yale University Library

August 15-16, 2018

9:00 am - 5:00 pm

Instructors: Catherine DeRose, Joshua Dull,
Kate Nyhan

Year One: Using the Lessons

- Carpentries lesson materials are CC-BY
- Not official “workshops”, but great building-blocks
- Examples
 - go.tufts.edu/messydata
 - nesclic.github.io/2018-04-18-YUL

Introduction to OpenRefine (based on OpenRefine for Ecology)

Working with OpenRefine

OpenRefine PTB1881_wyear [Permalink](#)

Facet / Filter Undo / Redo 1 / 1

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

1270 rows Extensions: Wikidata ▾

Show as: rows records Show: 5 10 25 50 rows first < previous 1 - 50 next > last ▾

YEAR	Month	Route Date	Location
1881	May	7	Jersey City, N.J.
1881	May	16	Bridgeport, Conn.
1881	May	17	Waterbury, Conn.
1881	May	18	New Haven, Conn.
1881	May	19	Meriden, Conn.
1881	May	20	Hartford, Conn.
1881	May	21	Holvoete, Mass.

➤

Overview

Teaching: 5 min
Exercises: 10 min
Questions

- How can we bring our data into OpenRefine?
- How can we sort and summarize our data?
- How can we find and correct errors in our raw data?

Objectives

- Create a new OpenRefine project from a CSV file on the web.
- Recall what facets are and how they are used to sort and summarize data.
- Recall what clustering is and how it is applied to group and edit typos.
- Manipulate data using previous steps with undo/redo.
- Employ drop-downs to split values from one column into multiple columns.
- Employ drop-downs to remove white spaces from cells.

Year One: Pilot Baby!

Calliope Jane Atwood-Curtis

Born February 14, 2018 at 3:24 am



Year Two: NESCLiC Members



BROWN

Kelsey Sawyer

Ashley Champagne



DARTMOUTH

Pamela Bagley

Stephen Gaughan

Meghan Kerr



Susan Remondi

TBD



Ann Kardos

Erin Jerome



Sally Gore

Tess Grynoch



David Cirella

Barbara Esty

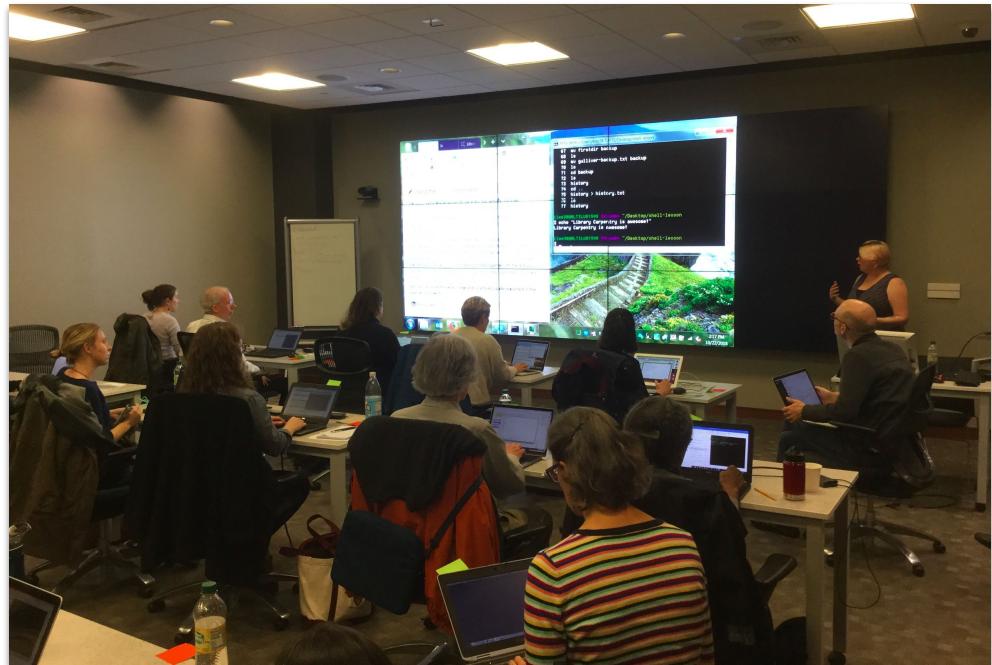
Sawyer Newman

Jennifer Chaput



Year Two: Even more workshops!

- Between October 1, 2018 and May 1, 2019:
 - 9 workshops
 - 2 Library Carpentry
 - 5 Software Carpentry
 - 2 Data Carpentry
- Repurposing materials for local instruction



The Future: Unexpected Benefits

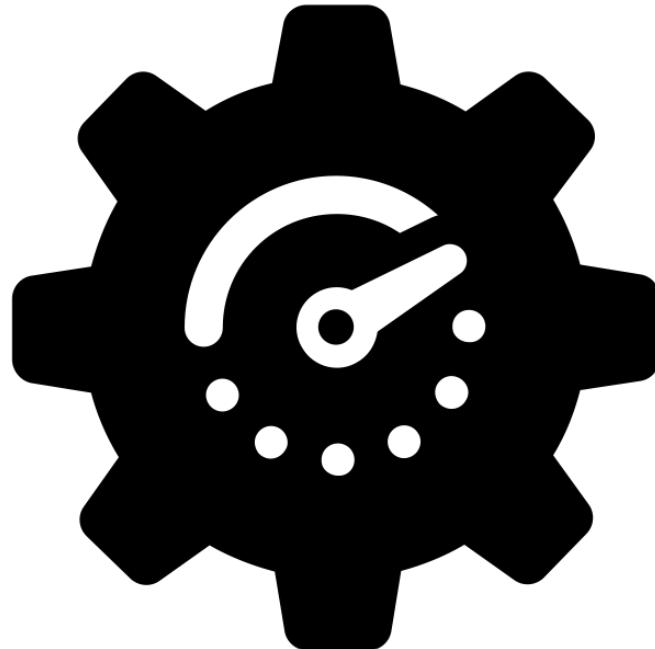
- Not just instructor training
 - Keep data/coding skills sharp
 - A group of people who help each other
- Expanded Local Network
 - Communicate and collaborate on other projects
- Collaboration with other regional organizations
 - NEASIST
 - NNLM New England Region



Created by Laili Hidayati
from Noun Project

The Future: Moving forward

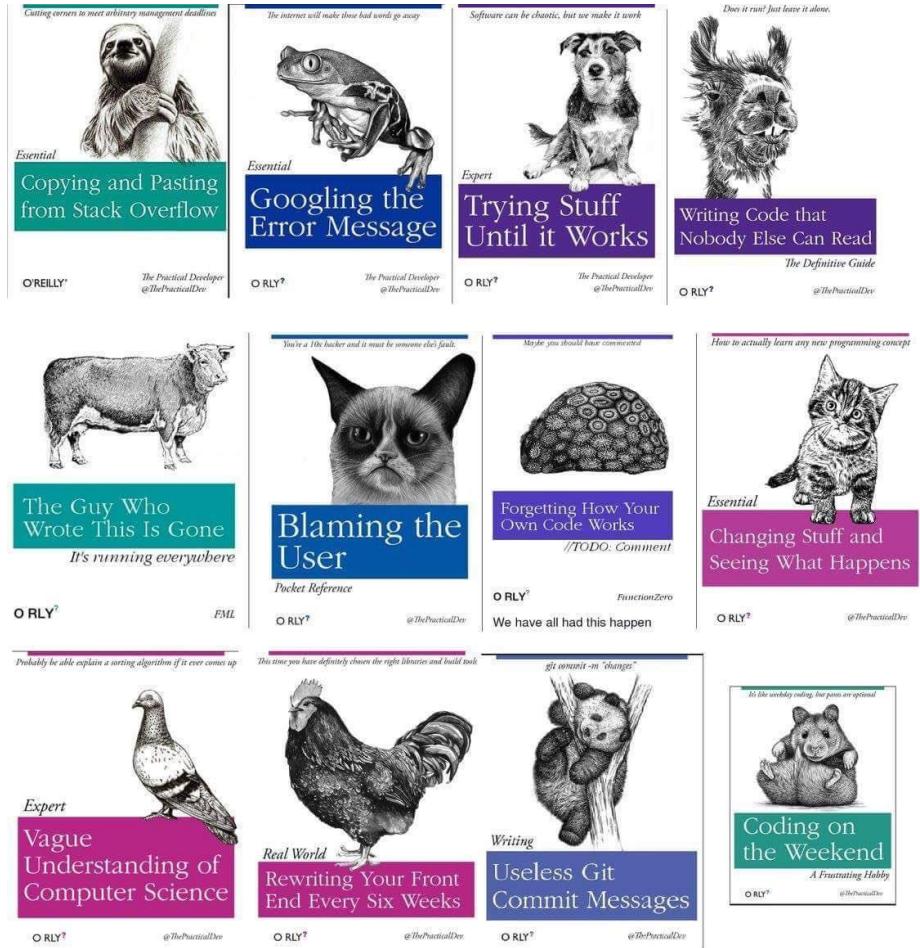
- The Carpentries
 - Evolving as an organization
- Changing roles within libraries
 - Evolution of Library Carpentry
 - More ‘non-traditional’ library training
- Train-the-Trainer
 - Train others in pedagogy
- Further participation in lesson development
 - Becoming lesson maintainers



Created by Gregor Cresnar
from Noun Project

Jargon Busting

1. Pair up!
2. Write down any terms, phrases, or ideas around code or software development in libraries (3 mins).
3. Get into groups of 4-6!
4. Make a list of all the problematic terms, phrases, and ideas each pair came up with.
5. Spend 10 minutes trying to explain what the terms, phrases, or ideas on your list mean.
6. Identify the terms your groups resolves as well as those you are still struggling with.



Questions?

nesclic.github.io/home

joshua.dull@yale.edu

kristin.lee@tufts.edu