The Carpentries: Teaching data science skills to researchers and people working in library- and information-related roles worldwide

Chris Erdmann, Library Carpentry Community and Development Director

Charleston Library Conference
November 7, 2018
Skills and perspectives to work with software and data are increasingly important as we generate more data.

With the emergence of our ability to generate increasing amounts of data, research and work in almost every domain has a data and computational component, including the whole new field of data science.
REALIZING THE POTENTIAL OF DATA SCIENCE

Final Report from the National Science Foundation Computer and Information Science and Engineering Advisory Committee Data Science Working Group

Francine Berman and Rob Rutenbar, co-Chairs
Henrik Christensen, Susan Davidson, Deborah Estrin, Michael Franklin, Brent Hailporn, Margaret Martonosi, Padma Raghavan, Victoria Stodden, Alex Szalay

December 2016

The function of Federal advisory committees is advisory only. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the Advisory Committee, and do not necessarily reflect the views of the National Science Foundation.
69% of business leaders in the United States will prefer job applicants with data skills by 2021.

23% of college and university leaders say their graduates will have those skills.
Rise of data science initiatives in academia

From the Data Science Community Newsletter by Noren & Stenger:

Brigham Young University, Caltech, Carnegie Mellon, College of Charleston, Columbia, Cornell, Dartmouth UMass, George Mason University, Georgetown University, Georgia Tech, Harvard, Illinois Wesleyan University, Johns Hopkins, Mid America Nazarene University, MIT, Northeastern University, Northern Kentucky University, Northwestern, Northwestern College in Iowa, Ohio State University, Penn State University, Princeton, Purdue, Stanford, Tufts University, UC Berkeley, UC Davis, UC Irvine, UC Merced, UC Riverside, UC San Diego, UCLA, UIUC, University of Iowa, University of Michigan, University of Oregon, University of Pennsylvania, University of Rochester, University of San Francisco, University of Warwick, University of Washington, UT Austin, UW Madison, Vanderbilt University, Virginia Tech, Washington University in St. Louis, Middle Tennessee State University, NYU, Amherst College, Brown, CU Boulder, Duke, Illinois Institute of Technology, Lehigh University, Loyola University - Maryland, Rice University, SUNY at Stony Brook, UC Santa Barbara, UC Santa Cruz, UCSF, UMass Amherst, UNC - Wilmington, University of Vermont, University of Arizona, University of British Columbia, University of Chicago, University of Virginia, USC, Worcester Polytechnic, Yale

70 and counting...
Importance of research software & training

- 92% of academics use research software
- 69% say that their research would not be practical without it
- 56% develop their own software (worryingly, 21% of those have no training in software development)

Current unmet needs

Forget Excel: This Was Reinhart and Rogoff's Biggest Mistake

Correlation is not causation

theatlantic.com

2:59 PM - 18 Apr 2013
Our path to better science in less time using open science tools

Reproducibility has long been a tenet of science but has been challenging to achieve—we learned this the hard way when our old approaches proved inadequate to efficiently reproduce our own work. Here we describe how several free software tools have fundamentally upgraded our approach to collaborative research, making our entire workflow more transparent and streamlined. By describing specific tools and how we incrementally began using them for the Ocean Health Index project, we hope to encourage others in the scientific community to do the same—so we can all produce better science in less time.

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
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

Note: “Carpentry” means “the basics” like learning how to nail two boards together or put up a wall straight.
Workshops

- 2-days, active learning
- Feedback to learners throughout the workshop
- Trained instructors
- Friendly learning environment
<table>
<thead>
<tr>
<th>Carpentry</th>
<th>Domain Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software Carpentry</td>
<td>Domain agnostic, research workflow/software-related</td>
<td>Command line, version control, programming</td>
</tr>
<tr>
<td>Data Carpentry</td>
<td>Domain-specific, research data-related</td>
<td>Ecology, Genomics, Geospatial, Social Science, Atmospheric Science</td>
</tr>
<tr>
<td>Library Carpentry</td>
<td>Library and information/workflow-related, Carpentries onboarding, community outreach and advocacy-driven</td>
<td>Data introduction, command line, version control, data wrangling</td>
</tr>
</tbody>
</table>
Workshop goals

- Teach skills
- Get people started
- Introduce them to what’s possible
- Build confidence in using these skills
- Encourage people to continue learning
- Positive learning experience
Our workshops. Our learners.

Since 2012, The Carpentries has seen 58 Trainers badged and 1,480 Instructors certified who have taught 1,332 Carpentries workshops reaching over 37,000 learners in 44 countries.

n = 1,350 on all 7 continents yes, even Antarctica
Who takes workshops?

Data Carpenter Respondents by Career Stage

- Undergraduate Student: 48
- Graduate Student: 592
- Postdoctoral Researcher: 183
- Faculty: 101
- Industry Employee: 49
- Government Employee: 80
- Research Staff: 200
- Management/Administrator: 20
- Retired/Not Employed: 18
- Other: 79

% Respondents
# Figure: Perception of Workshop Impact

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Mean (SD)</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reproducible</td>
<td>3.73 (1.00)</td>
<td>4.2%</td>
<td>8.1%</td>
<td>18.2%</td>
<td>49.7%</td>
<td>19.9%</td>
</tr>
<tr>
<td>Recognition</td>
<td>3.72 (0.96)</td>
<td>3.9%</td>
<td>4.4%</td>
<td>26.6%</td>
<td>45.8%</td>
<td>19.2%</td>
</tr>
<tr>
<td>Productivity</td>
<td>3.45 (1.00)</td>
<td>4.6%</td>
<td>9.7%</td>
<td>35.0%</td>
<td>37.0%</td>
<td>13.7%</td>
</tr>
<tr>
<td>Motivation</td>
<td>2.86 (1.03)</td>
<td>9.0%</td>
<td>26.9%</td>
<td>39.4%</td>
<td>17.9%</td>
<td>6.7%</td>
</tr>
<tr>
<td>Confidence</td>
<td>4.12 (0.88)</td>
<td>2.8%</td>
<td>2.5%</td>
<td>9.5%</td>
<td>50.7%</td>
<td>34.5%</td>
</tr>
<tr>
<td>Coding</td>
<td>3.67 (1.00)</td>
<td>4.4%</td>
<td>6.9%</td>
<td>25.4%</td>
<td>44.3%</td>
<td>18.9%</td>
</tr>
<tr>
<td>Career</td>
<td>3.57 (1.02)</td>
<td>4.2%</td>
<td>7.9%</td>
<td>33.7%</td>
<td>35.3%</td>
<td>18.9%</td>
</tr>
</tbody>
</table>

Instructor training involves

Carpentries pedagogy via:

- 2-day online training (some are in-person)
- Edit a lesson
- 1-hr discussion
- Demo

[www.carpentries.github.io/instructor-training/](http://www.carpentries.github.io/instructor-training/)
Open, collaboratively developed lessons
Community at your fingertips

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
- Social media and chat
23 Research Data Things allowed thousands of librarians to familiarize themselves with research data topics. Library Carpentry aims to do the same for software and data skills.
DATA SCIENCE ROLES

Data Librarian, Data Journalist, Data Analyst, Data Engineer, Data Steward, Data Archivist

38 mentions of the Carpentries as an example and recommendation
The New England Software Carpentry Library Consortium (NESCLiC)
The Strategic Value of Library Carpentry and The Carpentries to Research Libraries

By Elaine L. Westbrooks
Carpentries-based Workshop
“FAIR Data and Software”
July 9 - 13, 2018 in Hannover

Instructors
Katrin Leinweber, Angelina Kraft, Konrad Förstner, Martin Hammitzsch, Luke Johnston, Mateusz Kuzak

Helpers
Chris Erdmann

General Information
This workshop aimed to train junior scientists in implementing the FAIR principles for research data & software management & development. We want to help you identify similarities and differences between these two scientific objects and apply respectively appropriate good practices in preparing, publishing and archiving your work.

It was a new, experimental workshop format that contextualises the highly practical lesson material from the Software and Data Carpentries with the FAIR principles.
How can I get started?

- Host, help, teach.
- Join community initiatives.
- Contribute to a lesson.
- Become a member.

Thank you. Questions?

chris@carpentries.org

@libcce