

Workshop Overview

The Carpentry workshops are designed to provide basic training for scientists who want to learn to build, use, validate and share data using versatile open-source software. By teaching participants fundamental software skills and best practice techniques for working with and analyzing data, workshops help researchers spend more time doing useful research.

CEREO hosted two workshops, a Software Carpentry and a Data Carpentry workshop during the Spring 2015 semester. While somewhat similar, the Software Carpentry workshop led by Kara Woo (WSU) and Karl Broman (University of Wisconsin) focused on data management and version control. Participants were taught how to use R for data analysis, task automation and visualization, and Git for version control to record and archive changes in a file or set of files over time. The Data Carpentry workshop led by Kara Woo (WSU) and Naupaka Zimmerman (ASU) covered data processing and analysis. Participants were taught how to organize and clean data in spreadsheets using OpenRefine, import data into other analysis software, and how to use SQL and R for data analysis and visualization. Syllabi for each workshop can be found in Appendix B.

The structure of each workshop played a large role in its success. Workshops were team-taught by two instructors with the assistance of four or more “helpers” who were familiar with the software being used and could trouble-shoot problems while the instructors were teaching. Each workshop was organized into short tutorials with hands-on practical exercises and a capstone project to complete on the end of the second day. This hands-on teaching approach and cumulative test of what was learned allowed participants to walk away feeling more confident about their ability to pursue their research.

The Carpentry workshops are non-profit organizations and therefore requested a small administration fee (\$1,500) that supports creation of new lesson plans, instructor training and workshop coordination. Host institutions (CEREO) are expected to cover instructor travel, lodging and per diem expenses. Since these hands-on workshops lose much via remote participation, CEREO and CAS provided travel and lodging funds nine WSU participants from non-Pullman campuses who wanted to attend the workshops.

In total, these CEREO-sponsored workshops were attended by 63 participants, with an even distribution of male and female participants. Participants were mainly graduate students, although some staff, faculty and postdocs attended each. In total, 17 different departments spanning 6 colleges and 5 campuses took advantage of these workshops (Table 1). For participants by college, see Table A1.

Table 1. Summary of Carpentry Workshops

Number of:	Software	Data	Software & Data	Total
Attendees	29	34	8	63
Males	19	13	2	32
Females	10	21	6	31
Graduate student	22	22	5	44
Faculty	3	3	1	6
Staff	2	4	1	6
Postdoc	2	5	1	7
Depts Represented	12	13	5	17
Colleges Represented	5	6	5	6
Campuses Represented	5	2	2	5

Summary of Participant Feedback

Survey Results

Surveys were created by the Carpentry Foundation, and the questions used varied between workshops. CEREO was only offered the opportunity to add additional questions to their pre- and post-surveys for the Data Carpentry workshop. The Software Carpentry workshop surveys were mainly used to determine the skill level of participants for planning purposes. The Data Carpentry workshop survey focused more heavily on how participants viewed their experience in the workshop. In both cases, pre-survey responses (~75% response rate) were significantly higher than post-survey responses (~30% response rate). An overview of survey findings is given in Table 2.

Software Carpentry survey results provide limited useful information for assessment purposes (as they were designed mainly to assess familiarity with software for course planning purposes). Most of the participants surveyed before the workshop were at least familiar with programming, although the platforms represented were variable. Post-survey results indicated that overall, participants were pleased with the pace and scope of the material presented.

The Data Carpentry survey provided more insight into how participants felt about the skills they acquired during the workshop. In contrast to the Software Carpentry workshop, fewer participants indicated familiarity with programming at the start of the workshop. However, post-survey results showed that nearly all of the participants felt they gained a great deal of practical and immediately applicable knowledge from the workshop and that it helped them improve their data management and analysis skills. Participants who responded to the post-survey questionnaire all agreed or strongly agreed that this workshop was a worthwhile use of their time, and appreciated the interactive nature of the workshop and the logical flow of course material (working from data collection to visualization).

Table 2. Overview of survey responses

- Pre-survey response rates (~75%) for both workshops were higher than post-survey response rates (~30%)
- Software Carpentry surveys were primarily for course planning and not as useful for workshop assessment as Data Carpentry surveys.
- Data Carpentry participants came into the workshop knowing less about programming, but left feeling that they gained much practical, applicable knowledge for data management and analysis.
- Awareness of the Data Carpentry workshop came primarily from graduate student advisors.
- Data Carpentry participants all agreed or strongly agreed that the workshop was a worthwhile use of their time
- Data Carpentry participants provided very positive feedback on the structure and flow of the workshop.

Verbal Feedback

In addition to survey results, CEREO also directly received feedback from participants. Several graduate students and faculty enthusiastically expressed their desire for CEREO or WSU to institutionalize workshops like these into short-courses for incoming graduate students. This feedback came from participants in the physical and social sciences, with the latter interested in seeing workshops or short-courses developed and/or expanded

to include a comprehensive overview of how to work with qualitative data (the Software and Data Carpentry curriculum focus heavily on quantitative data analysis).

Participants also highly praised the structure and hands-on approach of these workshops. They particularly appreciated working through a logical progression of data analysis (as with the Data Carpentry event) and having the opportunity to exercise all the skills they had learned via a capstone project during the workshop event, so that if they got stuck there were still helpers around to guide them. Several participants even suggested a third day for full immersion within a large capstone project with immediate access to helpers would have been desirable. A summary of this feedback is given in Table 3.

Table 3. Overview of Verbal Feedback

- Participants expressed great interest in seeing the product of these two workshops institutionalized at WSU as a short-course for income graduate students
- Workshops developed or expanded to include qualitative data analysis were also suggested
- Capstone projects were commonly mentioned as one of the most positive pieces of the workshop- particularly because participants worked on them with help immediately available.
- Several participants suggested a full day for practicing their newly developed skills around helpers and instructors would have been desirable.
- The use of helpers, in addition to instructors, was viewed as critical by most participants to keep everyone moving at the same pace.

Appendices

Appendix A. Tables

Table A1. Participants by College

College	Software	Data
Voiland College of Engineering & Architecture	10	8
College of Agricultural, Human and Natural Resources	4	14
College of Arts & Sciences	12	6
Office of Research and Economic Development	2	1
Office of the Provost	1	4
Murrow College of Communication	0	1
<i>Total</i>	<i>29</i>	<i>34</i>

Appendix B. Workshop Syllabi

B1. [Software Carpentry Syllabus](#)

The Unix Shell

- Files and directories
- History and tab completion
- Pipes and redirection
- Looping over files
- Creating and running shell scripts
- Finding things

Programming in R

- Working with vectors and data frames
- Reading and plotting data
- Creating and using functions
- Loops and conditionals
- Using R from the command line

Version Control (Git)

- Creating a repository
- Recording changes to files
- Viewing changes
- Ignoring files
- Working on the web
- Resolving conflicts
- Open licenses
- Where to host work, and why

B1. [Data Carpentry Syllabus](#)

Data Organization

- Organizing data in Excel
- Data cleaning with OpenRefine
- Introduction to databases
- Combining and querying data using SQL

Programming in R

- Working with vectors and data frames
- Reading and plotting data
- Creating and using functions
- Intro to *dplyr*
- Visualizing data with *ggplot2*