

Workshop Overview

CEREO typically offers Data or Software Carpentry workshops once per semester, covering basic computational skills for data management, analysis, or research computing. The goal of this offering was slightly different: We wanted to provide hands-on instruction for higher level R programming techniques targeted at graduate students in the sciences. In contrast to Data and Software Carpentry workshop offerings, this workshop required basic knowledge of R programming prior to the workshop, took place fully over Zoom, and spanned five mornings instead of two full days. The curriculum was developed by the instructors, with some topics adapting material published by the Carpentries, and other topics being developed from scratch. We limited the attendance to 20 participants due to the online format. All attendees were students at WSU, and we were able to include students from multiple campuses.

The workshop took place from November 16th – 20th, 2020. We invited attendees to join us for all five days of instruction, but did not require this. Broad topics by day included: 1) Git/GitHub intro + data manipulation and visualization with tidyverse R packages, 2) R Markdown, 3) basic modeling with R, 4) workflow management with the drake package, and 5) an overview of making a free personal website with GitHub. While attendance to every lesson wasn't required, we did build upon (and reuse) skills learned throughout the week so that participants could see how techniques could be put into practice. For example, Git and GitHub were taught at the start of the workshop and integrated into each day's lesson plan to a varying degree. The workshop syllabus can be found in Appendix B.

The workshop was taught by a team of five instructors with the additional assistance of four helpers. Helpers generally had previous experience with the techniques or packages being taught and were present to troubleshoot technical problems, monitor the Zoom call, and manage the shared notetaking document (a Google Doc). To manage instructor and helper communications throughout the week of the workshop we created a Slack workspace used by the full team. Like similar CEREO workshops, the lessons were generally composed of tutorials that included hands-on exercises with live coding demonstrations. Challenge questions were interspersed throughout most tutorials, allowing participants to practice throughout the workshop and gain confidence in their ability to apply the techniques being taught. We also provided many

Table 1. Summary of Workshop Demographics based on survey responses

Number of:	
Registrants	20
Male	1
Female	11
Prefer not to say	4
Graduate student	19
Faculty	0
Postdoc	1
Undergraduate student	0
Other positions	0
Depts represented	3
Colleges represented	2
WSU campuses represented	4

Subsection totals may not sum to 20 if not all participants responded to demographic questions. Data were gathered from multiple sources with different response rates.

of the scripts in a format that would be accessible online to the participants after the workshop, allowing them to revisit the same material afterwards.

We sold out all 20 spots in the workshop within 48 hours of opening registration. There was a much higher proportion of female attendees than male, and all but one attendee (a post-doc) were graduate students (Table 1). Members of three separate departments, two colleges, and four WSU campuses were represented. For participants by college, see Table A1. Approximate attendance counts in order of workshop day were: 19, 15, 16, 16, and 11.

Unlike previous workshops, we also provided a two-hour Zoom call the day before the workshop to assist participants with installing the software if they needed help. In addition to this we sent out a PDF describing software needed in advance, some installation tips, and ways to check if difficult software had been installed successfully. None of the participants attended the pre-workshop installation call, and very few installation issues arose.

Summary of Participant Feedback

Survey Results

Links to confidential surveys were sent to the workshop participants immediately before and after the workshop. The surveys were created by the instruction team to replace surveys typically used for Carpentries workshops, but which were not available for this non-Carpentry offering. The goal was to assess participant satisfaction and solicit feedback for improving future iterations of this workshop. Pre-workshop survey response rates were higher than post-workshop (80% vs. 40%). An overview of survey findings is given in Table 2.

All of the pre-workshop respondents had previous experience in a programming language, and 94% of those had previous experience with R. However, only 25% of the respondents had previous experience with a statistical analysis program that used a graphic user interface (GUI), such as JMP or SPSS. Post-workshop survey results indicated that participants appreciated the live-coding style of the workshop. On average, attendees found the workshop very helpful (avg. rating 8.6/10) and would be very likely to recommend it to someone else (9.75/10). Additionally, respondents appreciated the helpers' and instructors' clarity and guidance during the workshop. When asked for areas for potential improvement, the most commonly cited issue was feeling rushed at the end of the workshop day or needing a slower pace.

In-workshop Feedback

Participants were also asked to provide simple short answer feedback about the workshop at the end of each day. Response rates varied day-to-day, and sometimes highlighted different strengths of the workshop than those in the post-workshop survey. For example, only one survey respondent mentioned R Markdown as the most useful section of the workshop, but three participants gave enthusiastic feedback in support of the R Markdown content at the end of the day it was covered. Suggestions for improvement here commonly mentioned pacing, relevance to respondents' research, and requests for access to lesson materials.

Table 2. Overview of survey responses

Overview
Response rates (80%) were higher than post-survey response rates (40%)
Pre-survey responses (n = 16)
100% of pre-workshop respondents came in with previous experience programming and 94% had prior experience with R.
25% of pre-workshop respondents had prior experience with a GUI-based statistical analysis program.
Most pre-workshop respondents considered themselves beginners (56%) or intermediate-level users (38%) with R.
Pre-workshop respondents mentioned working with data (25%), learning about Git/GitHub (31%), and increasing their comfort/understanding/familiarity with R (31%) as common reasons for attending.
Post-survey responses (n = 8)
Respondents rated the workshop as very useful (8.6 / 10 for usefulness on average) and would be very likely to recommend the workshop to someone else (9.75 / 10 on average).
Respondents commonly reported that live coding was a major strength of the workshop along with guidance/clarity from those leading it.
Modeling was mentioned most often as the most useful topic covered, followed by sections of the lesson on data manipulation and visualization with the tidyverse.
Pace was the most often cited area for improvement. Respondents felt that instructors tended to run out of time at the end of their days, could generally take a slower pace, and could allow for more catch-up time.

Appendices

Appendix A. Table

Table A1. Participants by College

College	Count
College of Agricultural, Human, and Natural Resource Sciences	9
Murrow College of Communication	0
College of Arts and Sciences	11
College of Veterinary Medicine	0
Not listed or non-WSU	0
College of Education	0
Voiland College of Engineering and Architecture	0
Grand Total	20

Appendix B. Workshop syllabus

Day 1: Brief intro to Git/GitHub. Intro to Data Manipulation and Visualization with the tidyverse

- Git/GitHub
 - Set up Git & create a repository
 - Set up GitHub & connect your local Git repository
- tidyverse
 - Select(), filter(), piping, mutate()
 - Pivot_wider(), pivot_longer()
 - Basic ggplot syntax
- Pushing to GitHub

Day 2: Intro to R Markdown

- Introduction to the markdown language
 - Formatting: Headers, super/subscripts, bold/italics
 - Inserting images, figures, equations
 - Practice reading, analyzing, and plotting data
 - Methods for making tables
 - Exporting the results as HTML

Day 3: Model Building with R

- Building models
 - Building a model and assessing fit (predictions vs data)
 - Assessing model residuals
 - Incorporating interactions with continuous and categorical variables
 - Including polynomial terms
 - GAMs
- Purrr + broom for model building
 - Nested dataframes/list-columns
 - Run models iteratively using map functions
 - Combining map with glance to extract model parameters
 - Visualize model parameters

Day 4: Workflow management in R with drake

- Introduction to workflow management
- Introduction to functions
- Introduction to drake
 - File structure
 - Splitting a workflow into functions
 - Writing core drake scripts
 - Organizing a full analysis plan
 - Editing or recovering a workflow

Day 5: Intro to making a website with Git/GitHub

- GitHub Pages background
- Finding and cloning templates
- Editing template documents locally and online
- Customizing website layout and formatting
- Uploading images and HTML files