

## Workshop Overview

Data Carpentry workshops introduce computational skills needed for data management and analysis. Participants are guided through the steps of the data lifecycle through hands-on exercises beginning with data cleaning and organization and through data visualization in R or Python. We assumed no prior computational experience for this workshop, which was open to WSU students and staff in addition to non-WSU participants.

We hosted the [Data Carpentry workshop](#) in October of 2018. Participants were taught how to organize and clean data in spreadsheets using OpenRefine, manage data with SQL, and analyze and visualize data with R or Python. This offering of Data Carpentry was unique in that we ran concurrent breakout groups for R and Python languages, instead of offering only one language during the workshop. The workshop syllabus can be found in Appendix B.

The workshop was team taught between five instructors with the assistance of nine helpers. The helpers have previous experience with the software being taught and were present to troubleshoot participants' problems as they arose. The workshops are generally composed of tutorials that include hands-on exercises with live coding demonstrations. Challenge questions interspersed throughout each tutorial allow participants to practice throughout the workshop and gain confidence in their ability to apply the techniques being taught. We also provided many 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.

Since we hosted a self-organized Carpentry workshop there was not an administration fee to pay. We had a total of 40 registrants in this workshop, with a higher proportion of female attendees than male. Most participants were graduate students, although faculty, postdoctoral researchers, university staff, government employees, and private analysts were also represented. Of those attendees from WSU, members from 11 departments took part in the workshop (Table 1). For participants by college, see Table A1. Thirty-eight participants attended day one, and 26 attended day two.

*Table 1. Summary of Data Carpentry Workshop Demographics*

| <b>Number of:</b>        |    |
|--------------------------|----|
| Registrants              | 40 |
| Male                     | 13 |
| Female                   | 21 |
| Graduate student         | 28 |
| Faculty                  | 2  |
| Staff                    | 3  |
| Postdoc                  | 1  |
| Other positions          | 5  |
| Depts represented        | 11 |
| Colleges represented     | 6  |
| WSU campuses represented | 2  |

*Participant totals may not sum to 40 if not all participants responded to demographic questions.*

## 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 Carpentries to assess the effectiveness of their workshops by measuring participant responses before and after participating in Data Carpentry. Pre-workshop survey response rates were higher than post-workshop (90% vs. 53%). An overview of survey findings is given in Table 2.

Almost 75% of respondents had some previous experience with a programming language. However, nearly 45% had never used a statistical analysis program with a graphical user interface. Post-workshop survey results indicated that most participants (74%) felt they gained immediately applicable knowledge from the workshop. On average, participants said post-workshop that they were very likely (92%) to recommend the workshop to a friend or colleague. Additionally, participants appreciated that the helpers and instructors were attentive and knowledgeable, and the enthusiasm of the instructors.

*Table 2. Overview of survey responses*

|  |   |
|--|---|
|  | Pre-survey response rates (90%) were higher than post-survey response rates (53%)   |
|  | Data carpentry participants came into the workshop with some previous experience programming, and left feeling that they had gained immediately usable knowledge. |
|  | A significant portion of the participants did not have prior experience with a statistical analysis program.  |
|  | Awareness of Data Carpentry came primarily from emails or flyers, followed by friends or colleagues, and then advisors/supervisors.                               |
|  | Participants provided very positive feedback on the instructors' and helpers' performance.  |
|  | Some participants suggested pre-workshop reading materials to reference for functions and intro programming topics.   |
|  | Participants with less programming experience would like greater focus on foundational concepts for R and Python.   |

### Verbal Feedback

We received direct feedback from some participants after the workshop. Several participants were from the Food Science department. They were excited about the workshop and indicated that they would like something like this to be offered to all incoming graduate students. Additionally, several Entomology students attended who had previous exposure to R programming. They mentioned having gained a lot from the dplyr and ggplot sections of the R syllabus.

## Appendices

### Appendix A. Table

Table A1. Participants by College

| College   | Count     |
|---|-----------|
| Center for Institutional Research Computing (CIRC)            | 1         |
| College of Agricultural, Human, and Natural Resource Sciences | 11        |
| College of Arts and Sciences                                  | 3         |
| College of Veterinary Medicine                                | 1         |
| Murrow College of Communication                               | 1         |
| Not listed or non-WSU   | 15        |
| School of the Environment                                     | 2         |
| Voiland College of Engineering and Architecture               | 6         |
| <b>Grand Total</b>  | <b>40</b> |

## **Appendix B. Workshop syllabus**

### **Data Organization**

- Organizing data in Excel/spreadsheets
- Data cleaning with OpenRefine
- Introduction to relational 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

### **Programming in Python**

- Using libraries
- Working with arrays
- Reading and plotting data
- Creating and using functions
- Loops and conditionals
- Defensive programming