6-9-2022

Library/Software/Data Carpentries

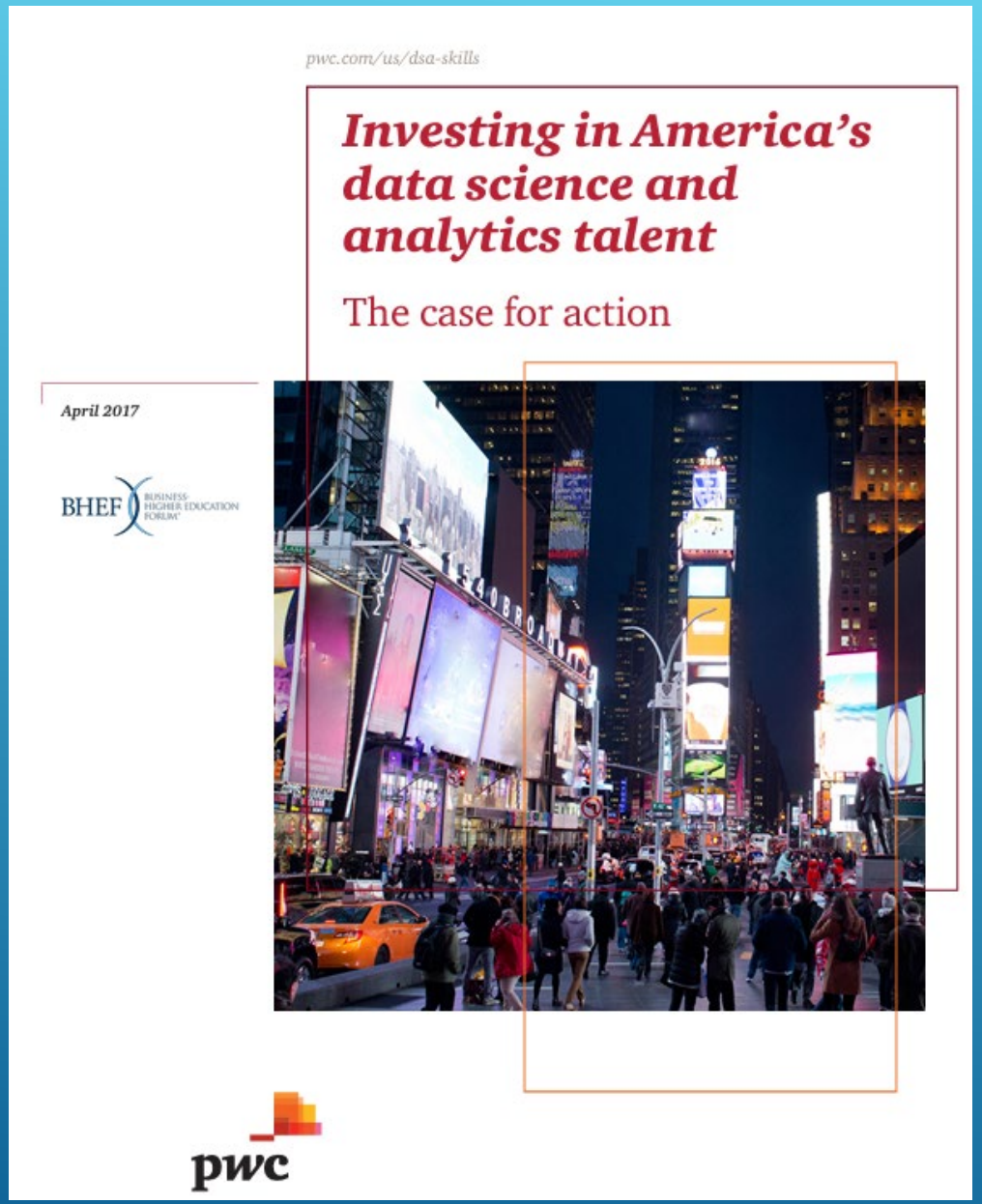
Lori D. Sloane
University of New Mexico Health Sciences Library and Informatics Center, lsloane@salud.unm.edu

LIBRARY/SOFTWARE/DATA CARPENTRIES

By Lori Sloane

The Carpentries is a global non-profit organization that teaches practical data science skills to researchers through active learning workshops.

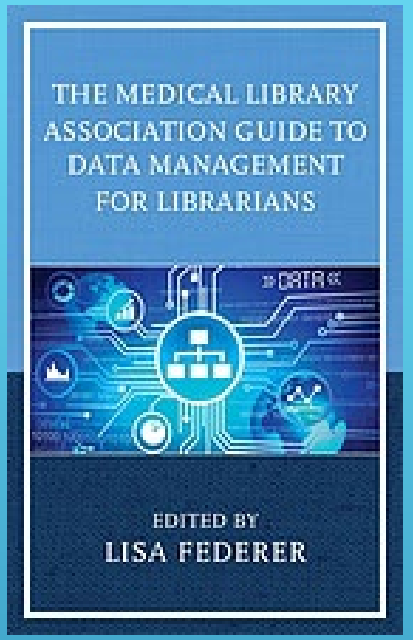
June 9th HSLIC Knowledge Café

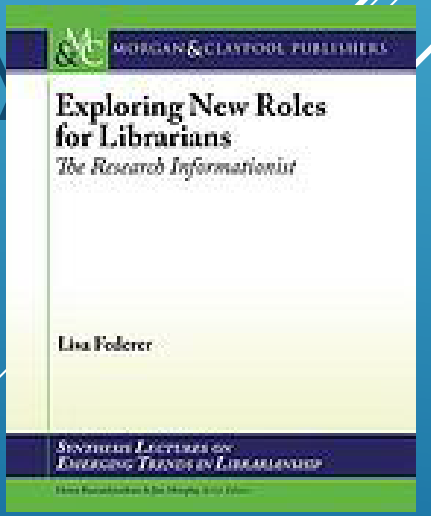


69% of business leaders in the United States will prefer job applicants with data skills by 2021.

23% of college and university leaders say their graduates will have those skills.

- Big-Data Skills: Bridging the Data Science Theory-Practice Gap in Healthcare
- Realizing The Potential of Data Science
- Defining data librarianship: a survey competencies, skills, and training





A VERY BRIEF HISTORY OF THE CARPENTRIES

1998

Software Carpentry is founded in 1998 by Greg Wilson and Brent Gorda to teach researchers better software development skills.

2005

Lesson materials are made open source with support from the Python Software Foundation.

2012

Software Carpentry workshop efforts scale with support from the Alfred P. Sloan Foundation and the Mozilla Science Lab.

2013

The first Software Carpentry for Librarians workshops are organized in the US and Canada.

2014

Data Carpentry is founded by Karen Cranston, Hilmar Lapp, Tracy Teal, and Ethan White with support from the National Science Foundation.

James Baker receives support from the Software Sustainability Institute to develop and implement Library Carpentry.

Software Carpentry Foundation is founded under the auspices of NumFOCUS.

2015

Data Carpentry workshop efforts scaled with support from the Gordon and Betty Moore Foundation.

2018

In January, Software Carpentry and Data Carpentry merge to form The Carpentries, a fiscally sponsored project of Community Initiatives.

In November, Library Carpentry joins as a Lesson Program.

Software
Sustainability
Institute

GORDON AND BETTY
MOORE
FOUNDATION

NUMFOCUS

The Carpentries Mission

The Carpentries builds global capacity in essential data and computational skills for conducting efficient, open, and reproducible research. We train and foster an active, inclusive, diverse community of learners and instructors that promotes and models the importance of software and data in research. We collaboratively develop openly-available lessons and deliver these lessons using evidence-based teaching practices. We focus on people conducting and supporting research.

MISSION & VISION

Act Openly

Created by Dairy Free Design
from Noun Project

Value All Contributions

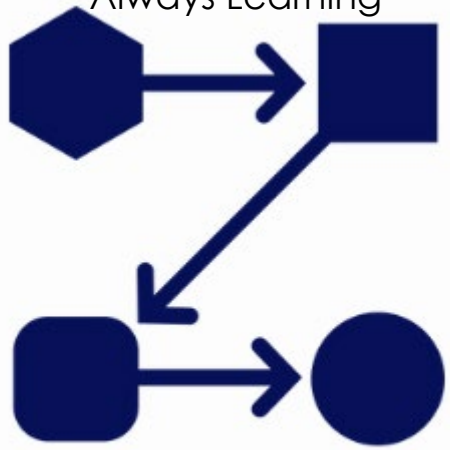
Created by hafiudin
from Noun Project

Inclusive of All

Created by Yu luck
from Noun Project

Empower One Another

Always Learning



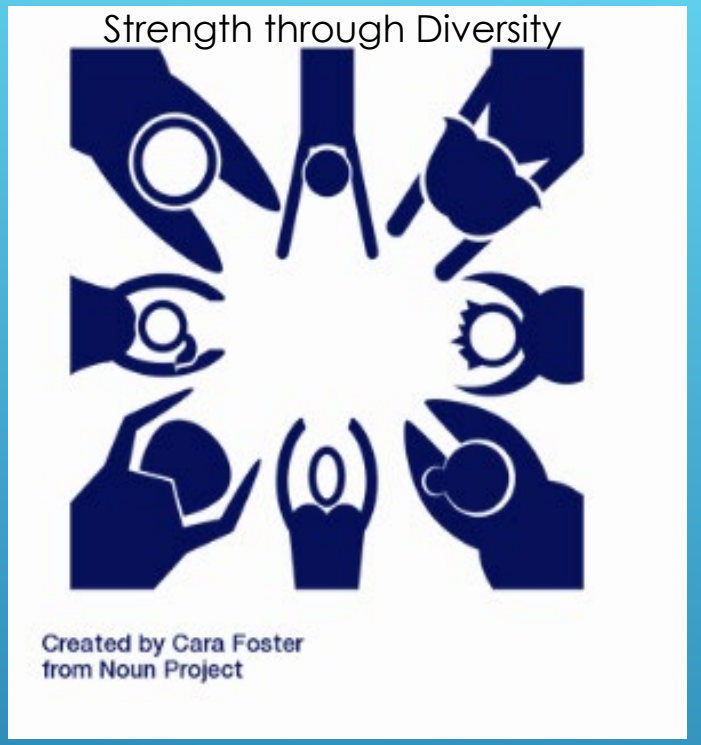
Created by Bold Yellow
from Noun Project

People First

Created by Adrien Coquet
from Noun Project


Access for All
Created by Laura Amaya
from Noun Project


Community Collaboration


Strength through Diversity
Created by Cara Foster
from Noun Project

- Use welcoming and inclusive language
- Be respectful of different viewpoints and experiences
- Gracefully accept constructive criticism
- Focus on what is best for the community
- Show courtesy and respect towards other community members

CODE OF CONDUCT

Library Carpentry

Data Introduction, command line, version control, data wrangling

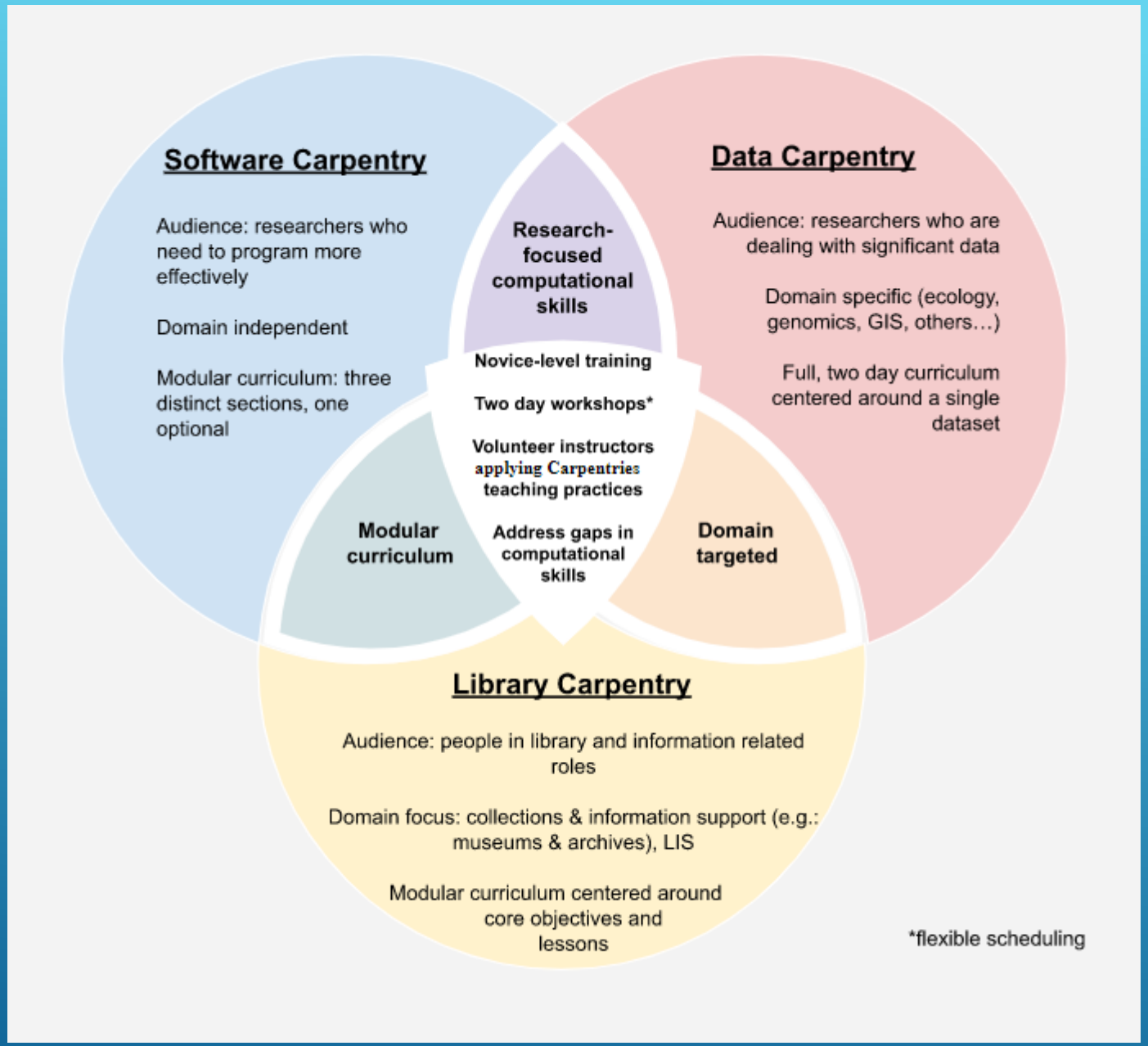
Software Carpentry

Command line, version control, programming

Data Carpentry

Ecology, Genomics, Geospatial, Social Science, Atmospheric Science

GENERAL OVERVIEW


Software Carpentry
Audience: researchers who need to program more effectively
Domain independent
Modular curriculum: three distinct sections, one optional
Data Carpentry
Audience: researchers who are dealing with significant data
Domain specific (ecology, genomics, GIS, others…)
Full, two day curriculum centered around a single dataset
Research-focused computational skills
Novice-level training
Two day workshops*
Volunteer instructors applying Carpentries teaching practices
Address gaps in computational skills
Modular curriculum
Domain targeted
Library Carpentry
Audience: people in library and information related roles
Domain focus: collections & information support (e.g.: museums & archives), LIS
Modular curriculum centered around core objectives and lessons
*flexible scheduling

- 2-days, active/hands-on learning
- Trained/certified instructors

WORKSHOPS

WORKSHOP GOALS

- Teach skills
- Get people started and introduce them to what's possible
- Build confidence in using these skills
- Encourage people to continue learning
- Positive learning experience

Lessons

Lesson	Site	Repository	Reference	Instructor Notes
Workshop Overview				
Introduction to Working with Data (Regular Expressions)				
The UNIX Shell				
OpenRefine				
Introduction to Git				

LIBRARY CARPENTRY WORKSHOP OVERVIEW

Our Core Lessons in English

Lesson	Site	Repository	Reference	Instructor Notes
The Unix Shell				
Version Control with Git				
Programming with Python				
Plotting and Programming in Python				
Programming with R				
R for Reproducible Scientific Analysis				

SOFTWARE CARPENTRY WORKSHOP OVERVIEW

DATA CARPENTRY WORKSHOP OVERVIEW

HOSTING A WORKSHOP

THE UNIVERSITY OF NEW MEXICO

WORKSHOP for DATA ANALYSIS

REGISTRATION LINK: nmepscor.org/form/data-carpentry-registration-hsc
INFORMATION LINK: nmcarpentries.github.io/2021-11-15-unm-online

When: November 15 - 19, 4:00 - 7:00 PM

Where: Zoom/online

Who: Free and open to all HSC researchers and students, no previous knowledge of R required

PARTICIPANTS WILL LEARN:

- Data organization
- Data management
- R for reproducible scientific analysis
- Logistic regression using R

INSTRUCTORS:

- Harry Snow, CTSC Informatics
- Jessica Gross, CTSC Biostatistics
- Lori Sloane, HSLIC
- Todd Quinn, UNM Libraries
- Jon Wheeler, UNM Libraries

Sponsored by UNM Libraries, Health Sciences Clinical & Translational Science Center, Health Sciences Library & Informatics Center, and New Mexico EPSCoR

THE UNIVERSITY OF NEW MEXICO

R for Reproducible Scientific Analysis

REGISTRATION LINK: https://bit.ly/3iUcN5Q
INFORMATION LINK: nmcarpentries.github.io/2022-05-16-unm

When: May 16 - 17, 8:00 AM - 5:00 PM

Where: HSLIC room 428
(in-person - COVID-19 permitting)

Who: Open to all HSC researchers and students

Cost: Free

No previous knowledge of R required

Class limited to 20 - Sign Up Today!

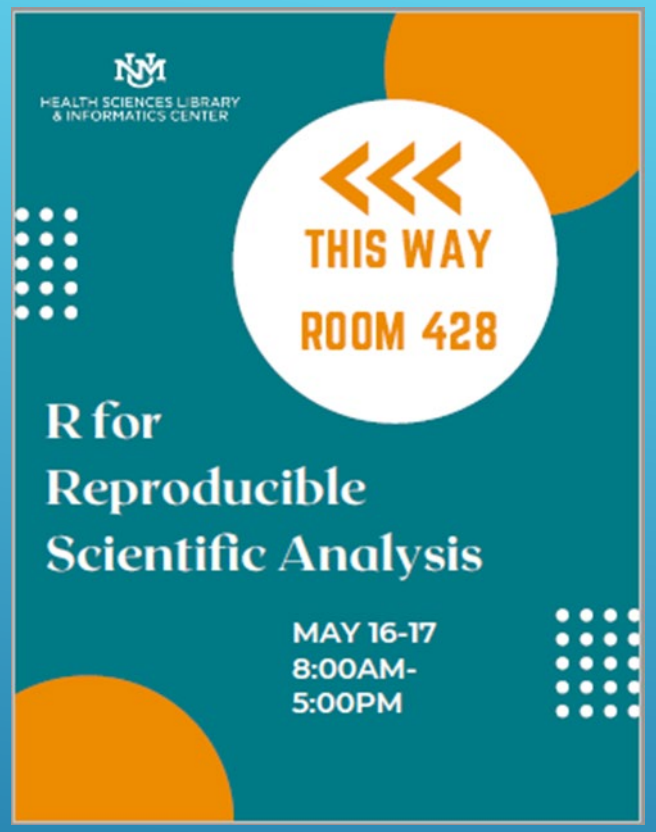
IN THIS SOFTWARE CARPENTRY WORKSHOP, PARTICIPANTS WILL LEARN:

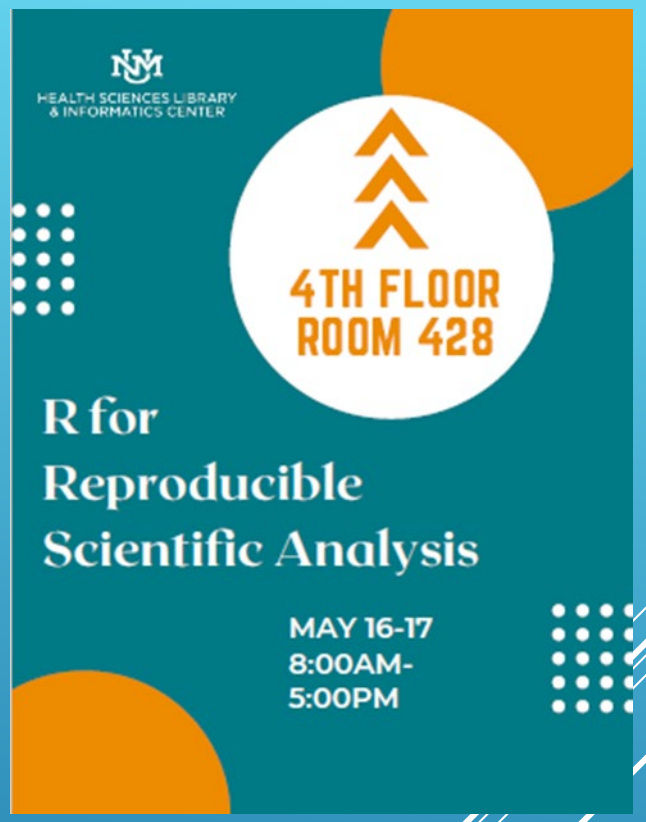
- R programming fundamentals
- Vectorization
- Creating Publication-Quality Graphics
- Data Frame Manipulation

BREAKFAST, LUNCH & SNACKS PROVIDED

Sponsored by UNM Libraries, UNM Clinical & Translational Science Center, Health Sciences Library & Informatics Center, and New Mexico EPSCoR


HEALTH SCIENCES LIBRARY
& INFORMATICS CENTER
WELCOME!
R for
Reproducible
Scientific Analysis
MAY 16-17
8:00AM-
5:00PM


HEALTH SCIENCES LIBRARY
& INFORMATICS CENTER
THIS WAY
ROOM 428
R for
Reproducible
Scientific Analysis
MAY 16-17
8:00AM-
5:00PM


HEALTH SCIENCES LIBRARY
& INFORMATICS CENTER
4TH FLOOR
ROOM 428
R for
Reproducible
Scientific Analysis
MAY 16-17
8:00AM-
5:00PM

University of New Mexico Health Sciences Center

May 16-17, 2022

8:00 am - 5:00 pm MDT

Instructors: Lori Sloane, Harry Snow, Todd Quinn, Qingqing Li

Helpers: Jon Wheeler

General Information

Software Carpentry aims to help researchers get their work done in less time and with less pain by teaching them basic research computing skills. This hands-on workshop will cover basic concepts and tools, including program design, version control, data management, and task automation. Participants will be encouraged to help one another and to apply what they have learned to their own research problems.

For more information on what we teach and why, please see our paper "Best Practices for Scientific Computing".

Who: The course is aimed at graduate students and other researchers. **You don't need to have any previous knowledge of the tools that will be presented at the workshop.**

Where: HSLIC Room 428, University of New Mexico, Albuquerque, NM, 87131. Get directions with OpenStreetMap or Google Maps.

When: May 16-17, 2022. Add to your Google Calendar.

Requirements: Participants must bring a laptop with a Mac, Linux, or Windows operating system (not a tablet, Chromebook, etc.) that they have administrative privileges on. They should have a few specific software packages installed (listed below).

Accessibility: We are committed to making this workshop accessible to everybody. For workshops at a physical location, the workshop organizers have checked that:

- The room is wheelchair / scooter accessible.
- Accessible restrooms are available.

Materials will be provided in advance of the workshop and large-print handouts are available if needed by notifying the organizers in advance. If we can help making learning easier for you (e.g. sign-language interpreters, lactation facilities) please get in touch (using contact details below) and we will attempt to provide them.

Contact: Please email LSloane@salud.unm.edu or jwheel01@unm.edu for more information.

Schedule

Day 1

Before	Pre-workshop survey
08:00	Setup & Introduction to R and RStudio
09:00	Project Management with RStudio
09:30	Seeking Help
10:00	Morning break
10:15	Data Structures
12:00	Lunch break
13:00	Exploring Data Frames
13:30	Subsetting Data
14:30	Afternoon break
14:45	Control Flow
15:45	Creating Publication-Quality Graphics with ggplot2
16:30	END

Day 2

08:00	Vectorization
09:00	Functions Explained
10:30	Morning break
10:45	Writing Data
11:00	Splitting and Combining Data Frames with plyr
12:00	Lunch break
13:00	Data Frame Manipulation with dplyr
14:30	Afternoon break
14:45	Data Frame Manipulation with tidyr
15:30	Producing Reports with knitr
16:00	Wrap-up
16:30	Post-workshop Survey
16:40	END

Feedback to learners throughout the workshop – STICKY NOTES

Welcome to The Carpentries Etherpad!

This pad is synchronized as you type, so that everyone viewing this page sees the same text. This allows you to collaborate seamlessly on documents.

Use of this service is restricted to members of The Carpentries community; this is not for general purpose use (for that, try https://etherpad.wikimedia.org).

Users are expected to follow our code of conduct: https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html

All content is publicly available under the Creative Commons Attribution License: CHAT https://creativecommons.org/licenses/by/4.0/

Welcome to the UNM HSC Online Data Carpentry workshop!
Some important links:
The workshop website: https://nmcarpentries.github.io/2022-05-16-unm/
The Code of Conduct: https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html

Software Carpentry R for Reproducible Scientific Analysis: https://swcarpentry.github.io/r-novice-gapminder/

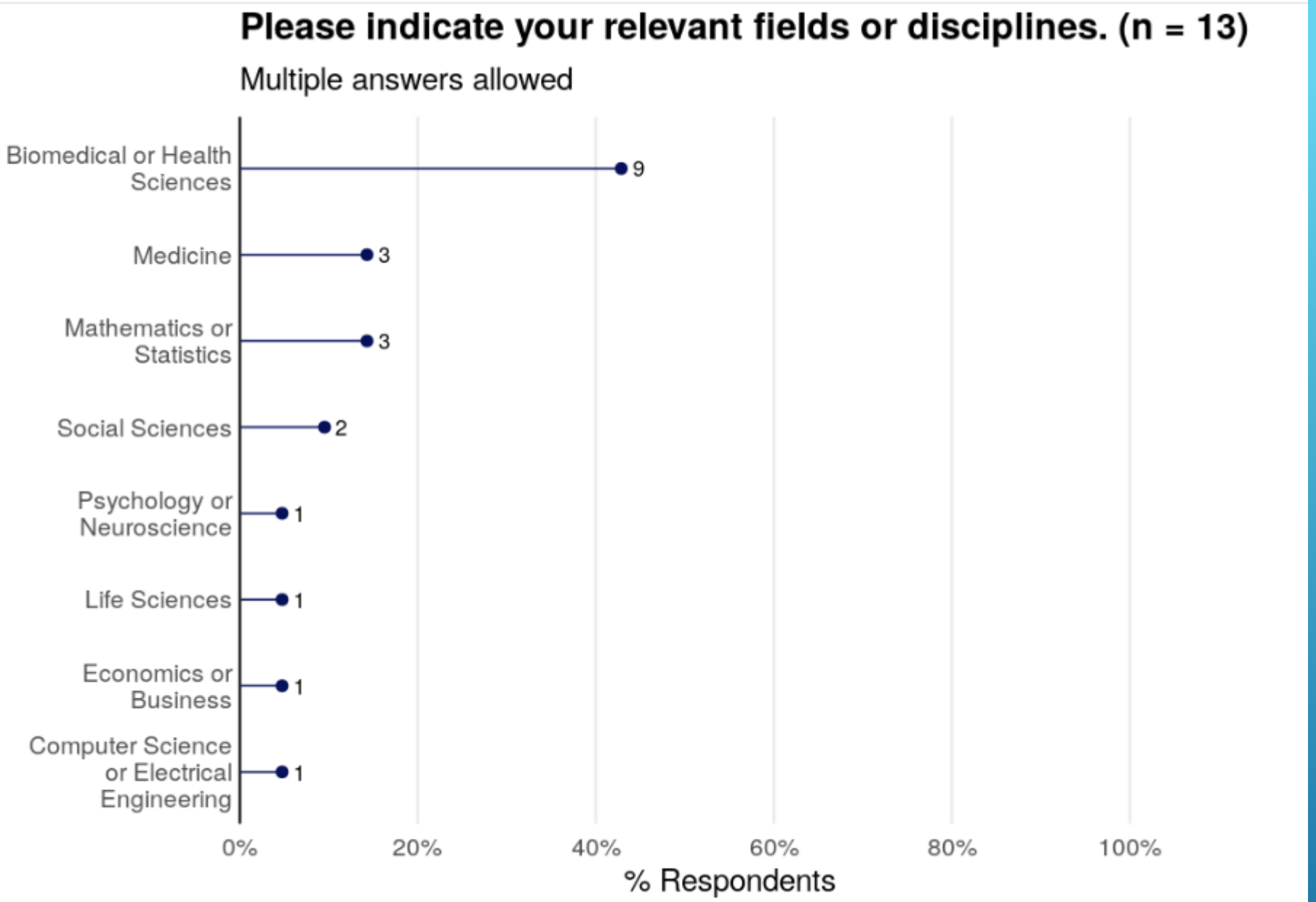
Everyone (instructors and helpers, too!):
Please complete the NM EPSCoR demographic survey: **https://www.surveymonkey.com/r/TVGTCXY**

Keyboard shortcuts
https://support.rstudio.com/hc/en-us/articles/200711853-Keyboard-Shortcuts-in-the-RStudio-IDE

Day 1
Introductions:
Please add your name, pronouns (if you are comfortable doing so), job title, and your favorite emoji (copy and paste from https://getemoji.com/)

EitherPad Community site for documenting and sharing information

https://pad.carpentries.org/2022-05-16-unm



Occupation

Research staff: 6

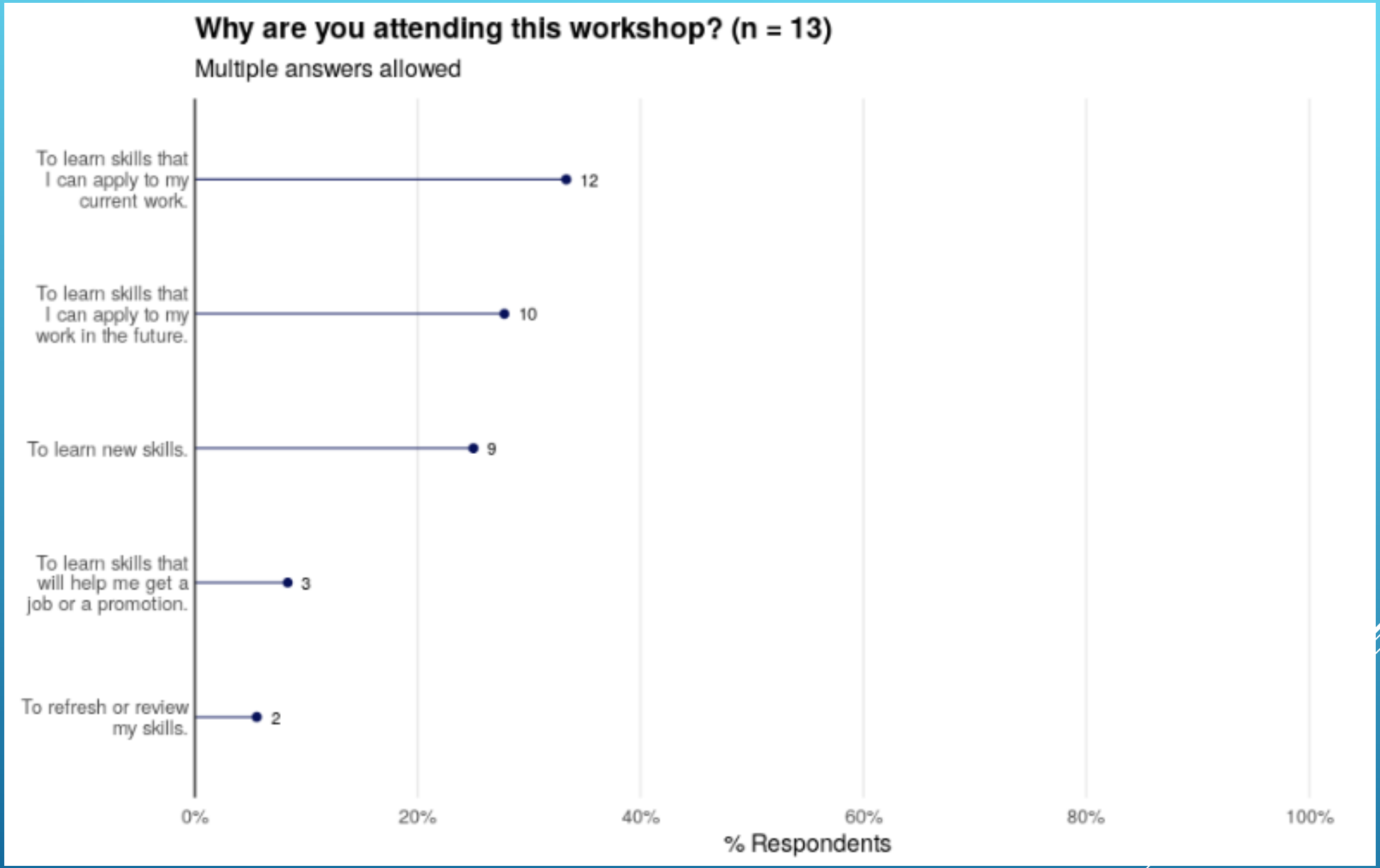
Analyst: 3

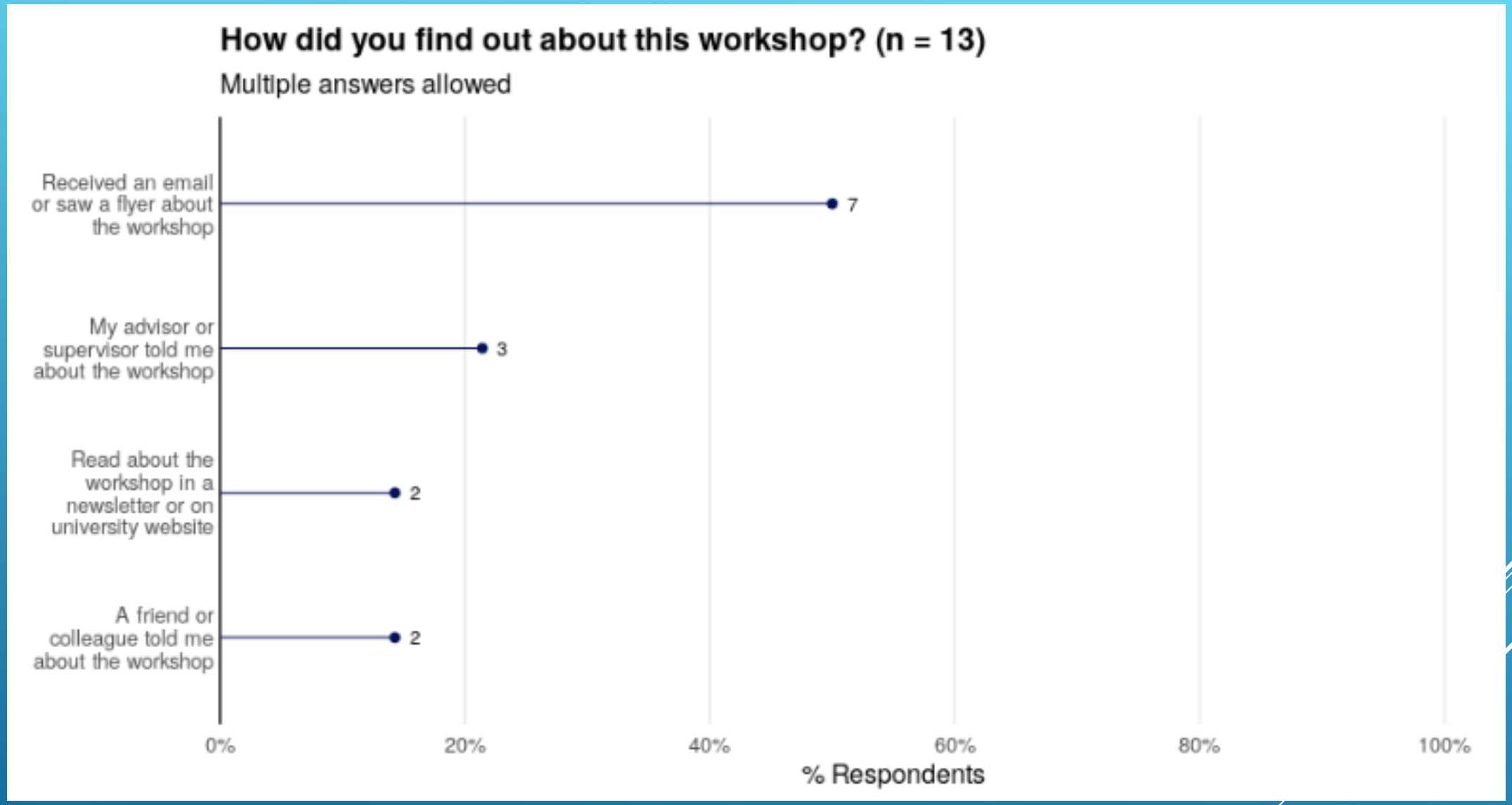
Faculty: 2

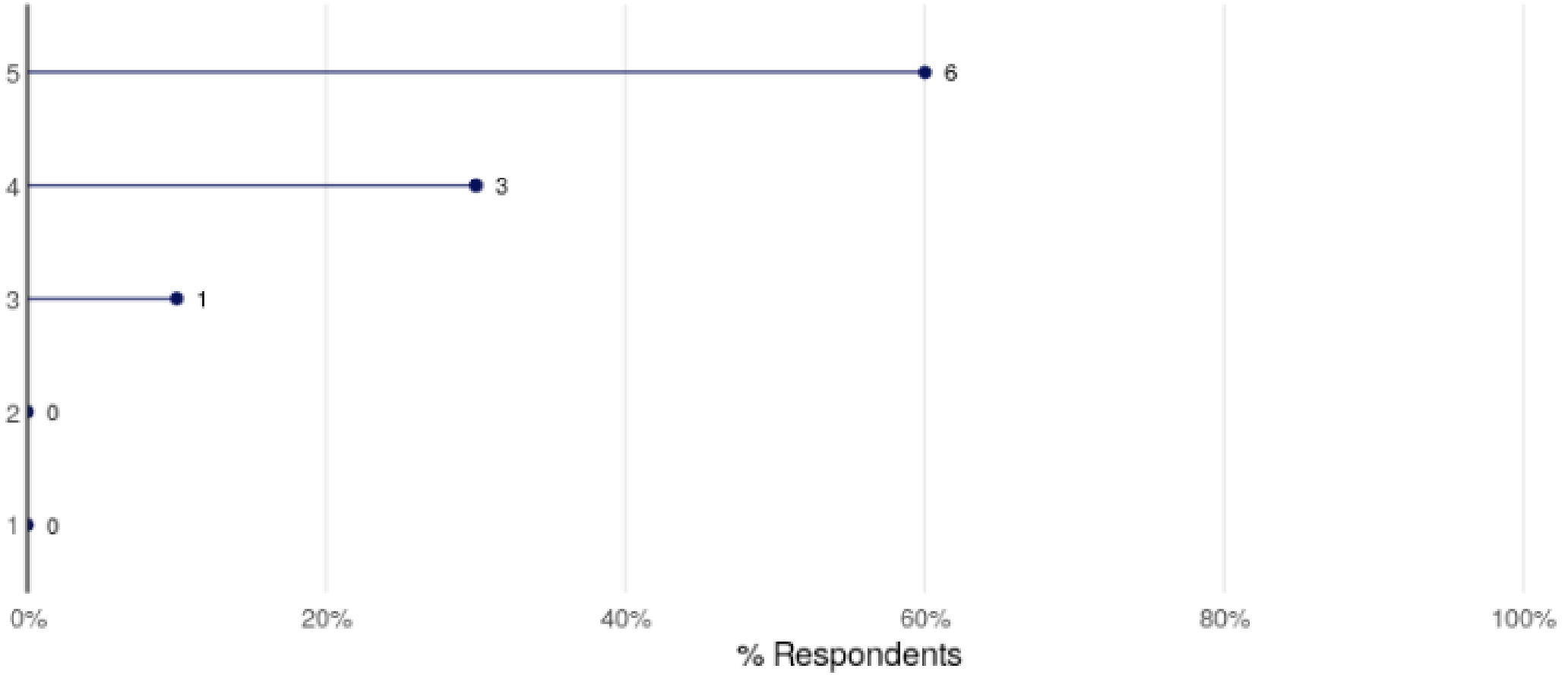
Administrative staff: 2

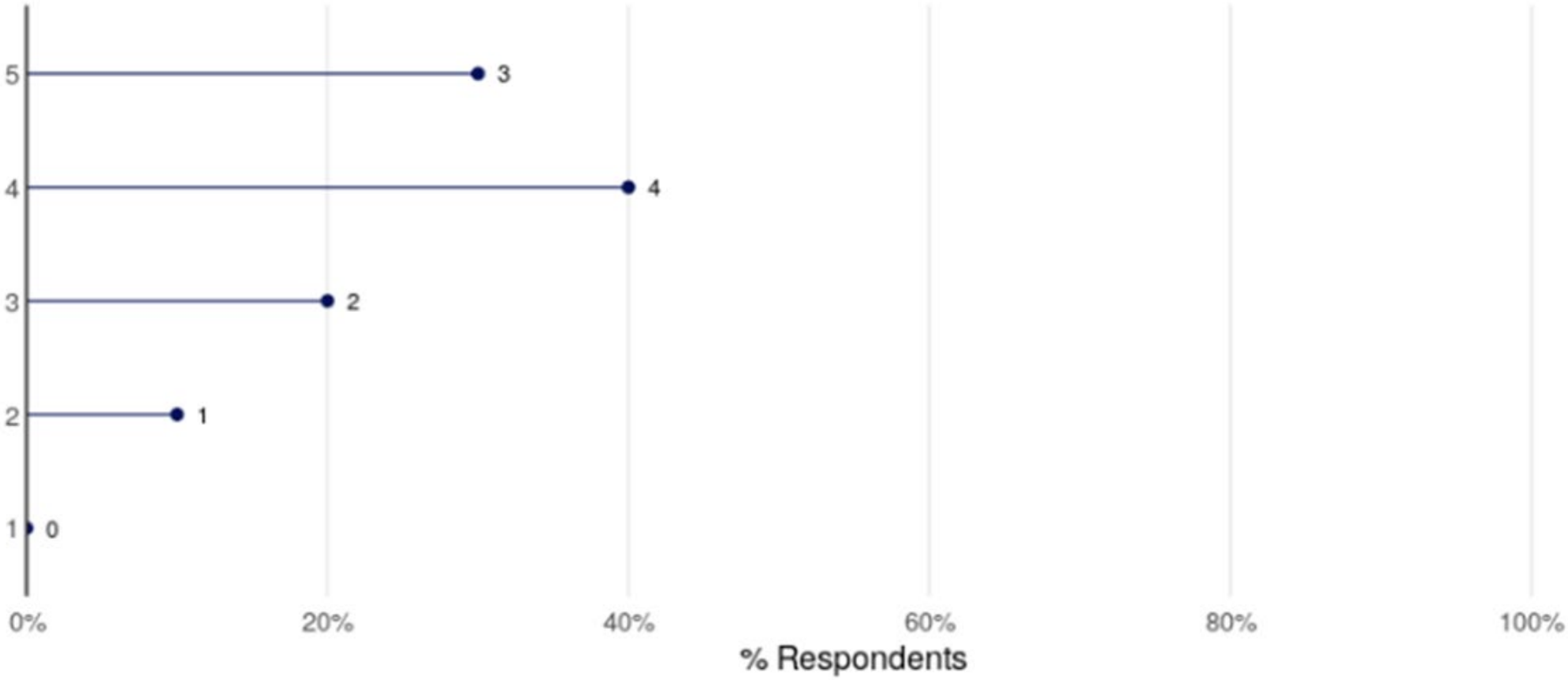
Grad Student: 1

(multiple answers allowed)


Why are you attending this workshop? (n = 13)
Multiple answers allowed
To learn skills that I can apply to my current work.
12
To learn skills that I can apply to my work in the future.
10
To learn new skills.
9
To learn skills that will help me get a job or a promotion.
3
To refresh or review my skills.
2
0%
20%
40%
60%
80%
100%
% Respondents


How did you find out about this workshop? (n = 13)
Multiple answers allowed
Received an email or saw a flyer about the workshop
7
My advisor or supervisor told me about the workshop
3
Read about the workshop in a newsletter or on university website
2
A friend or colleague told me about the workshop
2
0%
20%
40%
60%
80%
100%
% Respondents

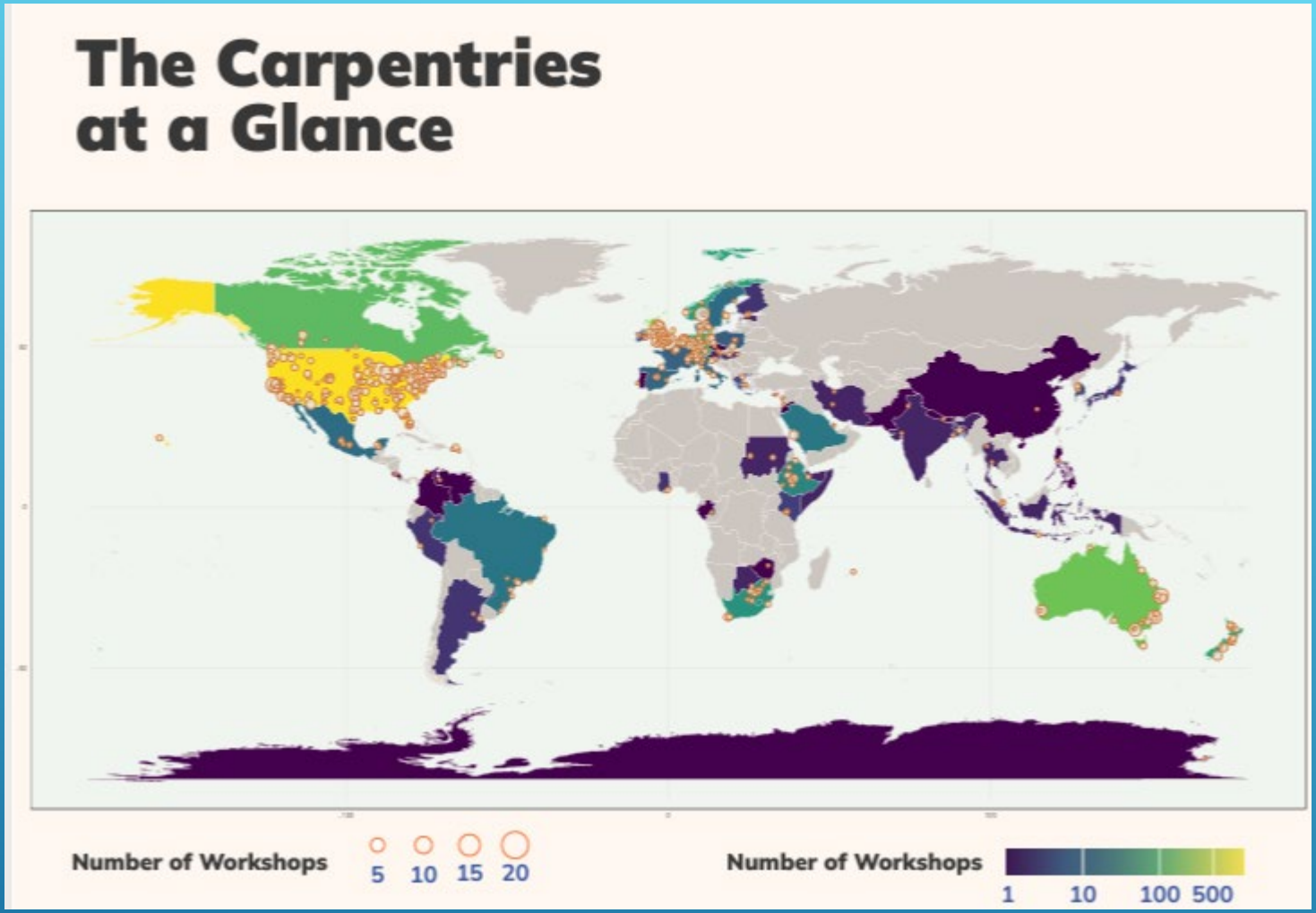

I felt comfortable learning in this workshop environment.
(1: Strongly disagree, 3: Neutral, 5: Strongly agree)
5
6
4
3
3
1
2
0
1
0
0%
20%
40%
60%
80%
100%
% Respondents


I can immediately apply what I learned at this workshop.
(1: Strongly disagree, 3: Neutral, 5: Strongly agree)
5
4
3
2
1
3
4
2
1
0
0%
20%
40%
60%
80%
100%
% Respondents

Please list the major strengths of this workshop
The learn together by doing together is a great way to learn how to code. The presenters were great.
Great instructors, pace modified as students needed more help
Being able to ask questions and confer with the instructors and fellow classmates.
The web-based guideline and the etherpad. The instructors are approachable
The instructors were incredibly knowledgeable, and the course was well-organized and accessible.
the instructors were fantastic. The workshop and resources were well thought out and planned
great instructors
The instructors are so organized and funny! Harry's voice projected extra well via Zoom.
Great facilitators and Instructors

LESSONS LEARNED

- Need more helpers for hybrid workshops
  - One to take care of the chat and someone separate to take care of the EitherPad postings
- Have diagrams to provide for any concepts drawn in the whiteboard
- Having the Public Service for laptop checkout
- Shorter days would be better



Since 2012 up-to-date, we have run **3,832** workshops in **64** countries and trained **3,512** volunteer instructors to deliver our **45** collaboratively developed, open lessons to **87,800** novice learners at our **92** member sites.

Get Involved

You Belong in The Carpentries!

The Carpentries works to help institutions and individuals spread skills for data analysis, computational thinking, and research software development through building local and global communities of practice.

Our work is made possible by the efforts of our amazing volunteer Instructors, Trainers, Mentors, Maintainers, curriculum advisors, helpers, workshop hosts and organisers, community champions, and member organisations.

https://carpentries.org/volunteer/

REFERENCES

- Investing in America's Data Science and Analytical Talents
- Big-Data Skills: Bridging the Data Science Theory-Practice Gap in Healthcare
- Realizing the potential of Data Science
- Defining data librarianship: a survey of competencies, skills, and training
- Unmet Needs For Analyzing biological big data: A Survey of 704 NSF principal investigators
- Labor and skills gap analysis of the biomedical research workforce
- Data Communities: A Solution To The Data Literacy Gap
- FILLING THE GAP: BOOSTING DATA SKILLS
- The History of Software Carpentry
- THE CARPENTRIES
- Analysis of Software and Data Carpentry's Pre- and Post-Workshop Surveys
- Workshop website for my session on May 16-17, 2022.
- Chris Erdman presentation