

# Analysis of the Carpentries' Long-Term Feedback Survey

Authors: Kari L. Jordan, Ben Marwick, Belinda Weaver, Naupaka Zimmerman, Jason Williams, Tracy Teal, Erin Becker, Jonah Duckles, Beth Duckles, Elizabeth Wickes

October 2017

## Introduction

Software Carpentry is a worldwide volunteer organization whose mission is to make scientists more productive, and their findings more reliable, by teaching them foundational computing skills. Established in 1998, it runs short, intensive workshops that cover task automation using the Unix shell, structured programming in Python and R, and version control using tools such as Git. Its sibling organization, Data Carpentry, teaches foundational data science skills. To date, the majority of Software and Data Carpentry workshops have been run in the United States, Canada and the UK. However, there is growing interest elsewhere, and there are active 'Carpentries' communities in Australia, New Zealand, South Africa, the Netherlands, Norway and in other countries in Africa and in Central and South America.

While most workshops are favorably assessed by learners at the time of delivery, no systematic, long-term follow-up study has previously been done on the efficacy of the training delivered, nor of the short or longer term impact that such training might have had on learners' work practices, further skills acquisition, or subsequent career paths. There has also been no useful demographic profiling of learners.

Why should that matter?

We are outcome-driven organizations interested in continually improving the workshop experience for both our learners and our instructors. Additionally, we are largely volunteers. In order to continue our valuable work of teaching skills to researchers, we need supportive funding either from grant-making bodies or from member institutions, or - ideally - from both. Our case for funding is strengthened if we can provide impartial evidence that proves our workshops have the outcomes we claim. While it is wonderful to have amassed what must now amount to container-sized loads of positive 'sticky note' workshop feedback over the years, funders generally require more solid evidence of achievement before they are willing to cut a check. Establishing value is not just important for funding though: it is important for our growing community of trainers, instructors, lesson maintainers and helpers, all of whom volunteer their time because they believe in the importance of what we do. We owe it to them to show that their time is not wasted - that we are genuinely furthering the cause of efficiently organized, reproducible science. Lastly, we need to demonstrate to our learners that precious time carved out to master computational and data science skills will pay off many times over in time saved further down the track. Data Carpentry's post-workshop survey results tell us that our respondents are either enthusiastically or very involved in our workshops, they learn a great deal of practical knowledge, and agree that they can immediately apply what they learned at the workshop. Our interest is in establishing, long-term, what impact workshops are having on learners' confidence in the skills they were taught.

Therefore, through the generosity of the Moore Data Driven Discovery Initiative, assessment work has been undertaken across both Carpentries to help build an evidence base to complement the large body of existing anecdotal evidence that Carpentry-style training is both useful and effective in improving researchers' work practices.

To gather the required evidence, Data and Software Carpentry launched a long-term assessment survey in March 2017. We first engaged in a community consultation to determine how best to design and word this survey. The responses from that consultation then guided the development of our long-term assessment strategy. You can read more about the consultation process in this blog post.

The main goal of the resulting survey was to ask our learners to describe concrete changes they had implemented to their research practices as a result of completing a Carpentries workshop. We also asked whether they now had greater confidence in the tools they had been taught, and whether they had progressed

in their careers as a result. The inclusion of multiple choice questions around programming in R or Python helped make the evidence of training efficacy more concrete, comparable and measurable, taking it out of the realm of ‘opinion’ or ‘feeling’, which, while interesting, is not as robust or reliable a marker of success as demonstrable evidence. For instance, ‘I can write a FOR loop to rename and move a batch of files’ is a much more reliable metric of achievement than ‘I can use the shell’.

All of the data collected in this survey was self-reported. It should be noted that there are disadvantages to self-reported surveys. For one, respondents may exaggerate their achievements. Additionally, a respondent’s state of mind while taking the survey may affect her/his answers. Survey results can potentially be biased because those feeling most positive are also those more likely to respond, while learners whose experience was less positive – or even negative - may not bother to answer. To account for this, we compared the results of the long-term survey with that of Data Carpentry’s post-workshop survey results and Software Carpentry’s post-workshop survey results. We found consistent patterns of increased confidence and self-efficacy in our learners.

The survey questions and the data used in this analysis are located in the carpentries/assessment repo on GitHub. We have already received several pull requests from community members interested in this data. Feel free to use the data and tell us about your findings.

This analysis includes 476 observations. Not all respondents answered every one of the 26 questions.

## Highlights

The long-term survey assessed confidence, motivation, and other outcomes **more than six months** after respondents attended a Carpentry workshop. Provided below are a few highlights from the data.

- 77% of respondents reported being more confident **in the tools** that were covered during their workshop compared to before the workshop.
- 54% of respondents have **made their analyses more reproducible** as a result of completing a workshop.
- 65% of respondents have gained confidence in **working with data** as a result of completing a workshop.
- 74% of respondents have **recommended our workshops** to a friend or colleague.

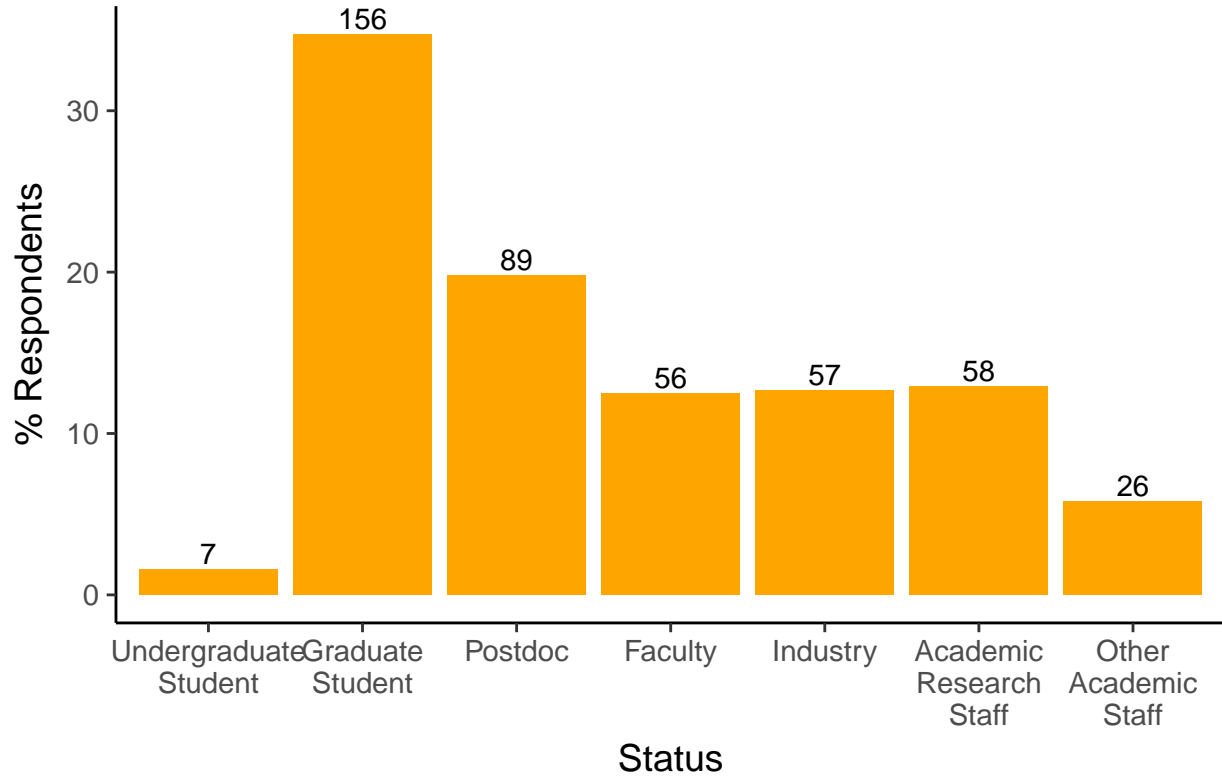
## Respondent Demographics

Carpentry learners represent a wide range of disciplines ranging from the sciences to engineering. Respondents were asked to indicate their field of research, work, or study by checking all that apply from a list of various disciplines. A breakdown of their responses is provided in the table below. Many of the respondents work in Life Sciences.

Field	n	%
Life Sciences	166	36.2
Biomedical/Health Sciences	95	20.7
Agricultural or Environmental Sciences	74	16.1
Physical Sciences	57	12.4
Earth Sciences	48	10.5
Engineering	41	8.9
Mathematics or Statistics	41	8.9
Computer Science	38	8.3
Social Sciences	24	5.2
Library Sciences	20	4.4
Humanities	12	2.6

Field	n	%
Business	6	1.3

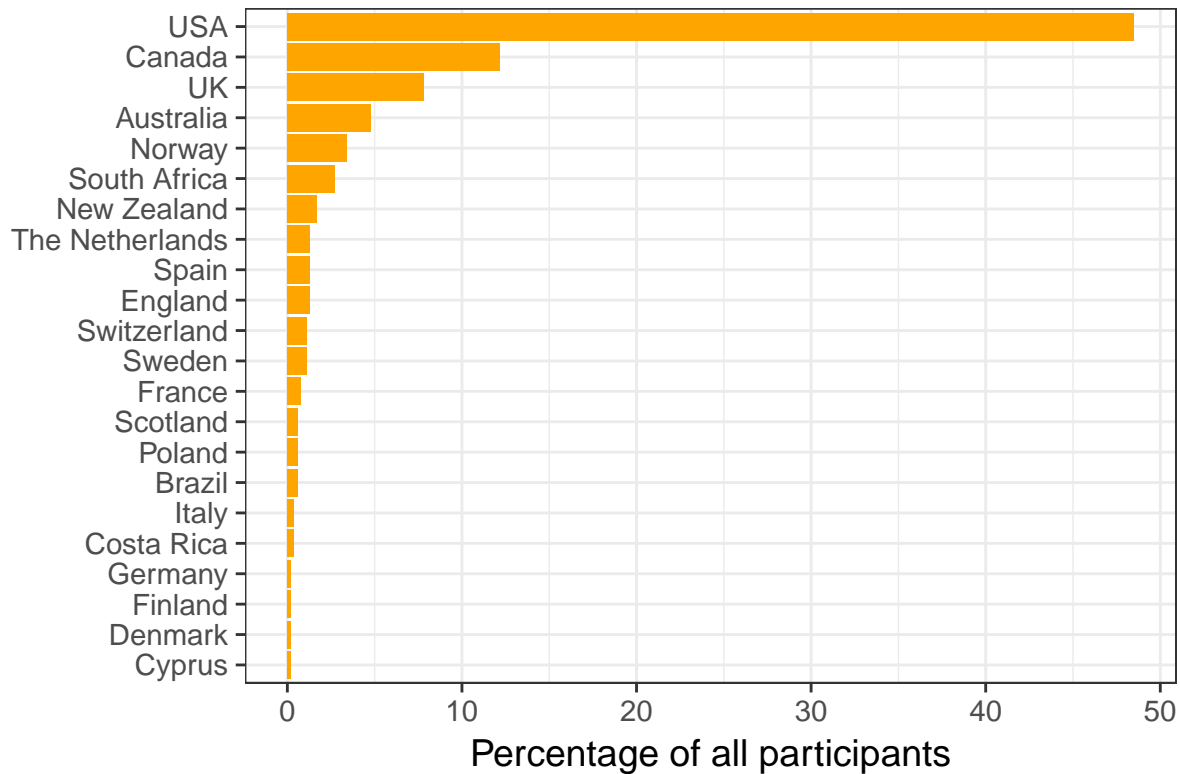
## Majority of Respondents were Graduate Students



Carpentries workshops are open to individuals from all backgrounds and fields. Attendees vary from students (undergraduate and graduate) and faculty to staff and persons working in industry. 35% of our respondents were graduate students.

Below is a breakdown of our respondents by the country in which they attended a Carpentries workshop.

## Carpentry Workshop Respondents by Country



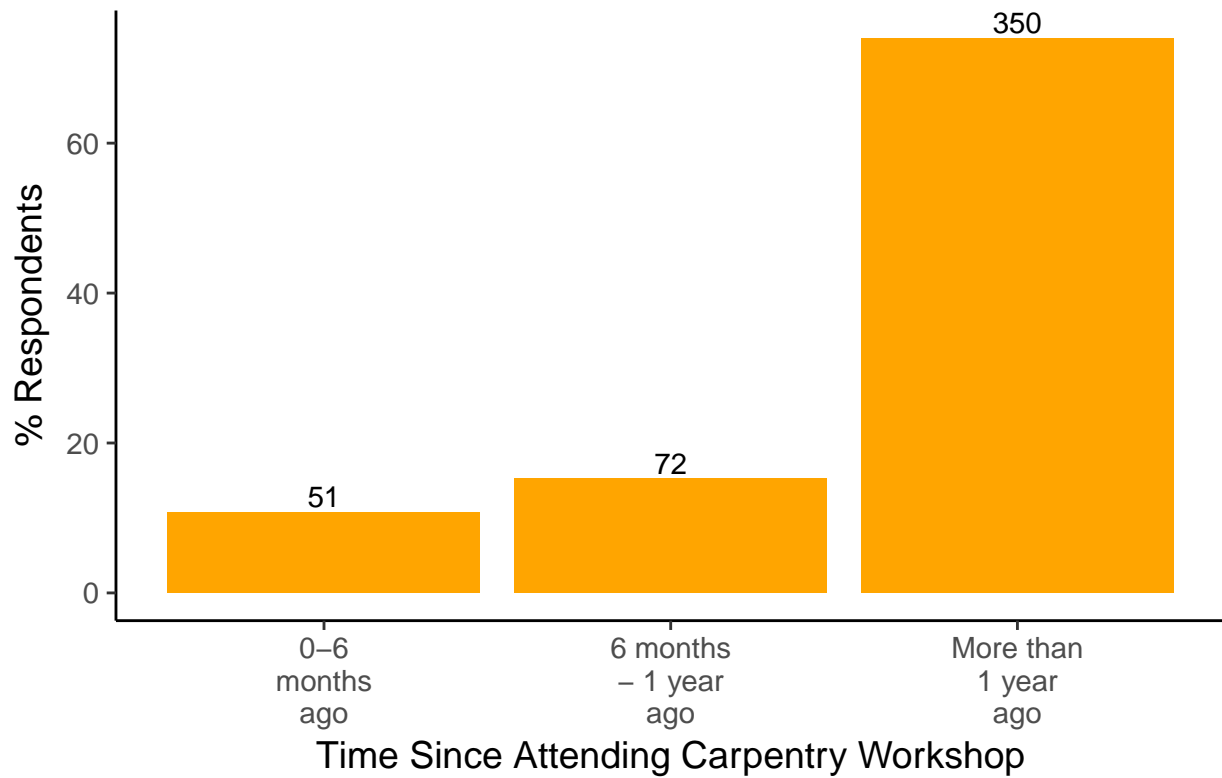
A large portion of Carpentry learners responding to the survey attended a workshop in the United States (48.5%), followed by Canada (12.2%) the UK (7.8%), and Australia (4.8%).

The Carpentries constantly strive to improve our workshop content and operations, which means a workshop run today might be different in some ways from a workshop run six months ago or run last year. If we know how many workshops respondents attended, and how long it has been since they completed that workshop or workshops, we can take those changes into account when we assess learner responses. Locating workshops temporally allows us to account for spikes in data trends that may be a result of changes in our workshop operations.

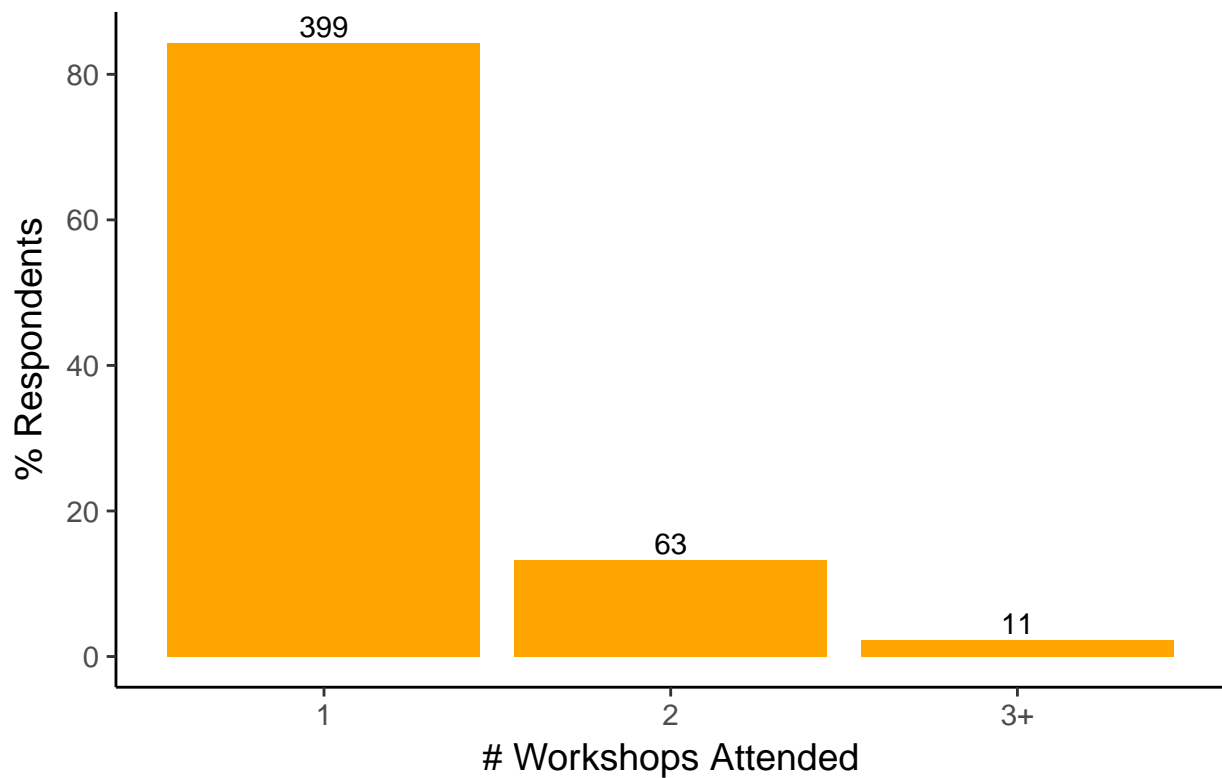
74% of respondents participated in a Carpentry workshop more than one year ago, and 84% of respondents have attended only one Carpentry workshop.

The response rate from learners who attended a workshop more than a year ago speaks to their level of involvement with the Carpentries. 120 of the survey respondents are subscribed to the Carpentries newsletter, indicating their ongoing interest in our work. In survey research, it can be difficult to collect responses from participants a whole year after their involvement in an event. It is great to see that learners feel positively enough about their experience to, firstly, be receptive to our email communication and, secondly, to have taken the time to complete the survey in full. As mentioned in the introduction, this could be a potential disadvantage to self-reported data, as those who may have negative experiences may not have completed the survey.

## Majority Attended Workshop Over a Year Ago



## Majority Attended One Carpentry Workshop



## Workshop Content

Data Carpentry’s lessons include data organization in spreadsheets, data cleaning with OpenRefine, data management with SQL, and data analysis and visualization in R and Python.

Software Carpentry’s lessons include the Unix Shell, version control with Git and Mercurial, programming with Python, R, and MATLAB, databases with SQL, and Automation and Make. Provided below is a breakdown of the tools respondents identified as being taught in the workshop they attended.

A large majority of respondents learned Git (n = 362), Python (n = 289), and the Unix Shell (n = 274). On the low end were spreadsheets (n = 20), cloud computing (n = 11), MATLAB (n = 5), and Mercurial (n = 3). The fact that OpenRefine, spreadsheets, and cloud computing were on the low end is an indicator that the majority of our respondents attended a Software Carpentry workshop. This is most likely because Data Carpentry is a newer organization, starting in 2014, so there have been fewer workshops held.

Table 2: Combination of Tools Covered

Frequency of Tools Covered	n	%
Git Python Unix Shell	92	19.3
Git Python	41	8.6
Git Python SQL Unix Shell	39	8.2
Git R Unix Shell	32	6.7
Git R	25	5.3
Git Unix Shell	20	4.2
Git Python R Unix Shell	18	3.8
Git Python SQL	16	3.4
R	15	3.2
Python	12	2.5

In examining which combinations of tools stood out in the data, we can see from the matrix below that Git was frequently taught alongside Python, R, and/or the Unix Shell. Additionally, SQL was often taught with Git and/or the Unix Shell. OpenRefine, Spreadsheets, and Cloud Computing were on the low end, a clear indicator that the majority of survey respondents attended a Software Carpentry workshop, rather than a Data Carpentry workshop.

Table 3: Matrix of Common Tools Covered

	Cloud Computing	Git	MATLAB	Mercurial	OpenRefine	Python	R	Spreadsheets	SQL
Cloud Computing	11	9	0	0	2	6	6	2	2
Git	9	362	4	2	13	247	132	8	110
MATLAB	0	4	5	0	0	4	2	0	3
Mercurial	0	2	0	3	0	3	0	0	1
OpenRefine	2	13	0	0	28	12	23	13	19
Python	6	247	4	3	12	289	63	5	89
R	6	132	2	0	23	63	186	19	64
Spreadsheets	2	8	0	0	13	5	19	20	16
SQL	2	110	3	1	19	89	64	16	134
Unix Shell	7	244	3	3	9	193	96	10	81

## Programming Usage Pre- and Post Workshop

Understanding respondents' programming usage both before and after attending a Carpentries workshop was one goal of this assessment study. Our hope is that the workshop learners attended favorably influenced their use of the programming tools they learned.

17% of the learners who responded to our survey had not been using the tools covered in their Carpentries workshop before they attended the workshop. This decreased to 5% post-workshop.

The plot below is a comparison of respondents' usage of the tools covered in their workshop both pre- and post-workshop.

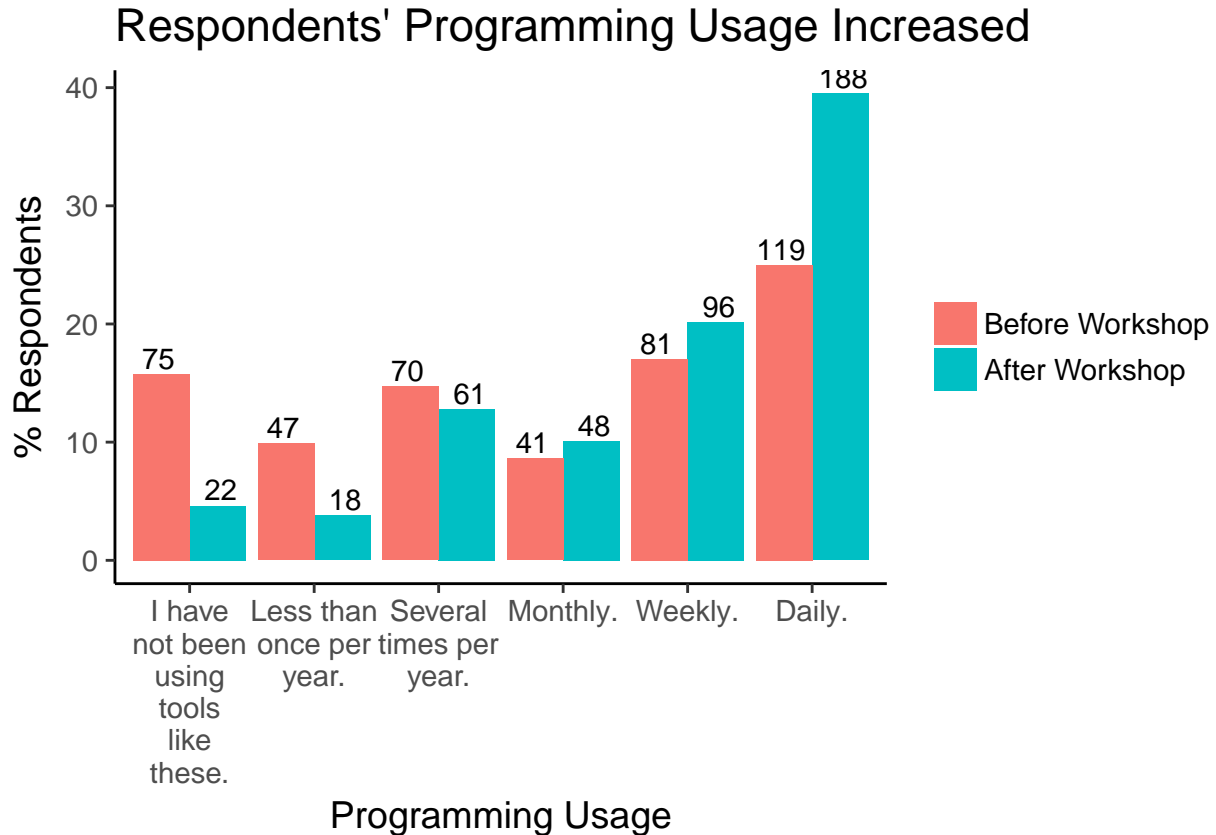
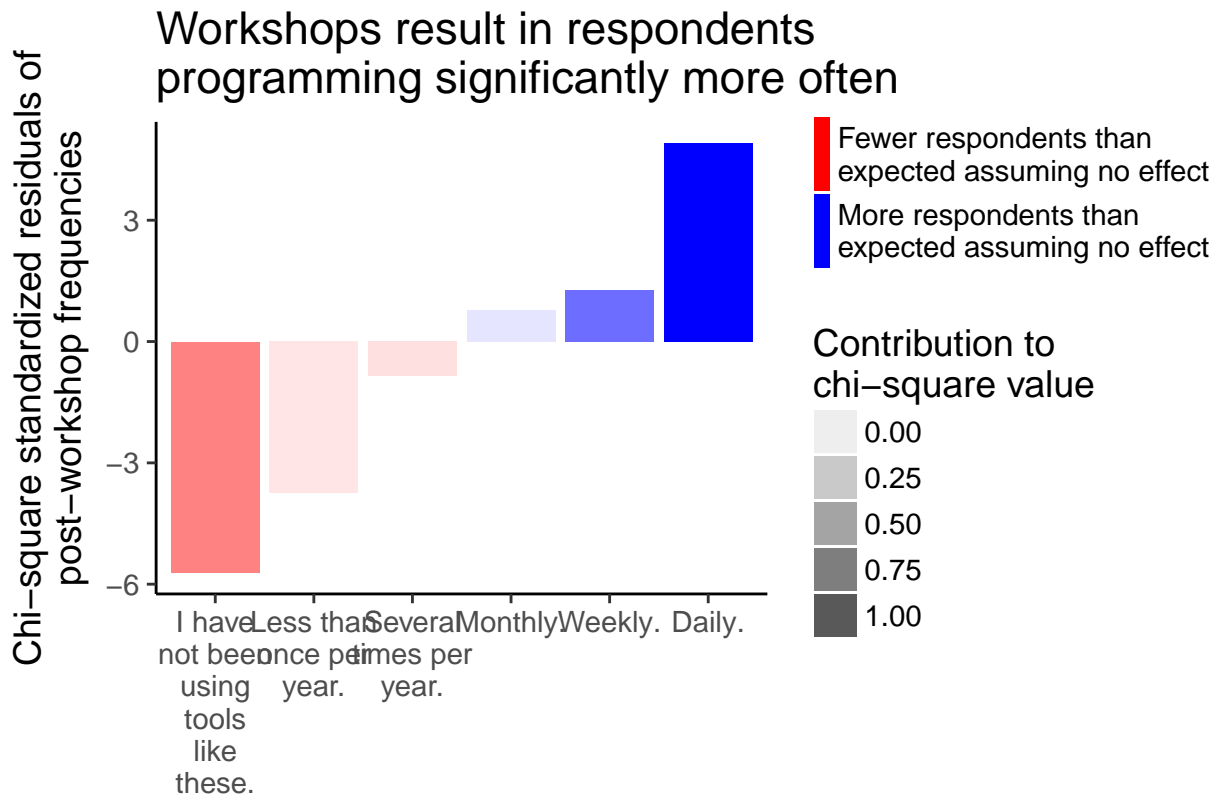


Figure 1: Synopsis: Respondents were asked how often they use programming languages (R, Python, etc.), databases (Access, SQL, etc.), version control software and/or the shell before completing a Carpentries workshop, and since completing a Carpentries workshop.

The most compelling (and pleasing) change in responses was a decline in the percentage of respondents who 'have not been using these tools' (-11.1%), and an increase in the percentage of those who now use the tools on a daily basis (14.5%) at least six months after they attended a Carpentry workshop.



A chi-square test indicates that the use of programming significantly increases post-workshop. The chi-squared standardized residuals for the post-workshop values show that significantly more respondents program daily six months (or more) after the workshop than would have been expected had the workshop had no effect on their practice. Similarly, significantly fewer respondents program less than once per year six months (or more) after the workshop.

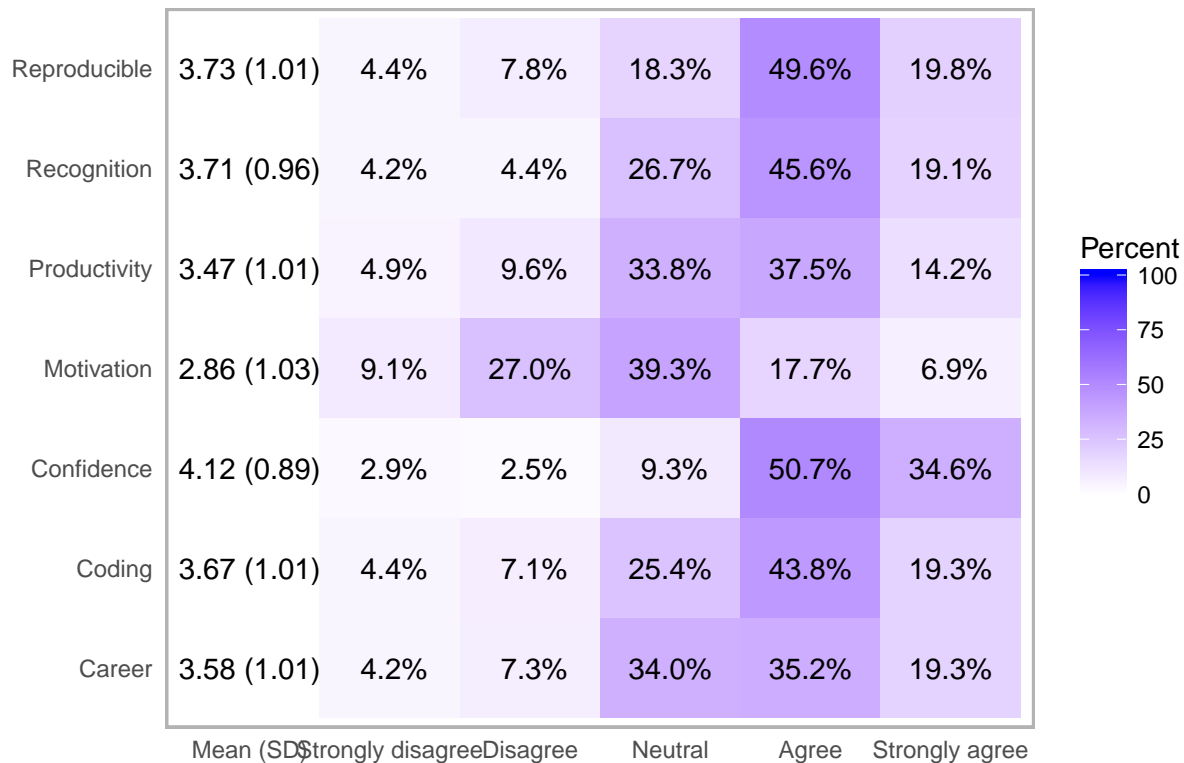
## Workshop Impact

The figure below shows respondents' perception of workshop impact on several factors, including career, confidence, and continuous learning. Respondents were asked to rate their level of agreement (1-Strongly disagree to 5-Strongly agree) with the statements below. The x-axis labels for the figure are in bold, and correspond to the statement following.

- **Reproducible:** I have made my analyses more reproducible as a result of completing the workshop.
- **Recognition:** I have received professional recognition for my work as a result of using the tools I learned at the workshop.
- **Productivity:** My research productivity has improved as a result of completing the workshop.
- **Motivation:** I have been motivated to seek more knowledge about the tools I learned at the workshop.
- **Confidence:** I have gained confidence in working with data as a result of completing the workshop.
- **Coding:** I have improved my coding practices as a result of completing the workshop.
- **Career:** I have used skills I learned at the workshop to advance my career.



## Perception of Workshop Impact



We see an overwhelmingly positive indication that, after taking the workshop, respondents feel motivated to seek more knowledge and have gained confidence in working with data. Additionally, respondents are making their analyses more reproducible and improving their coding practices (i.e. keeping raw data raw, sharing source code openly, and using scripts). As this survey was administered to learners who had taken a workshop at least six months ago, this feeling of confidence has obviously persisted. This is quite impactful, as both the Data Carpentry and Software Carpentry post workshop survey results show learners having self-reported improved understanding of how to import and work with data.

Additionally, 69% of respondents agree or strongly agree that they have improved their coding practices, made their analyses reproducible, or improved their research productivity. They also believe the skills they learned helped them advance their career: 65% of our respondents have received professional recognition as a result of using the tools they learned in a Carpentries workshop.

### Behaviors Respondents Adopted

We asked respondents to identify the behaviors they adopted as a result of completing a Carpentries workshop. We are happy to report that more than half the respondents who answered this question have improved their data management and project organization practices, have used programming languages for automation, and have used version control to manage code. Additionally, respondents are **more confident** now in using the tools than before they completed the Carpentries workshop.

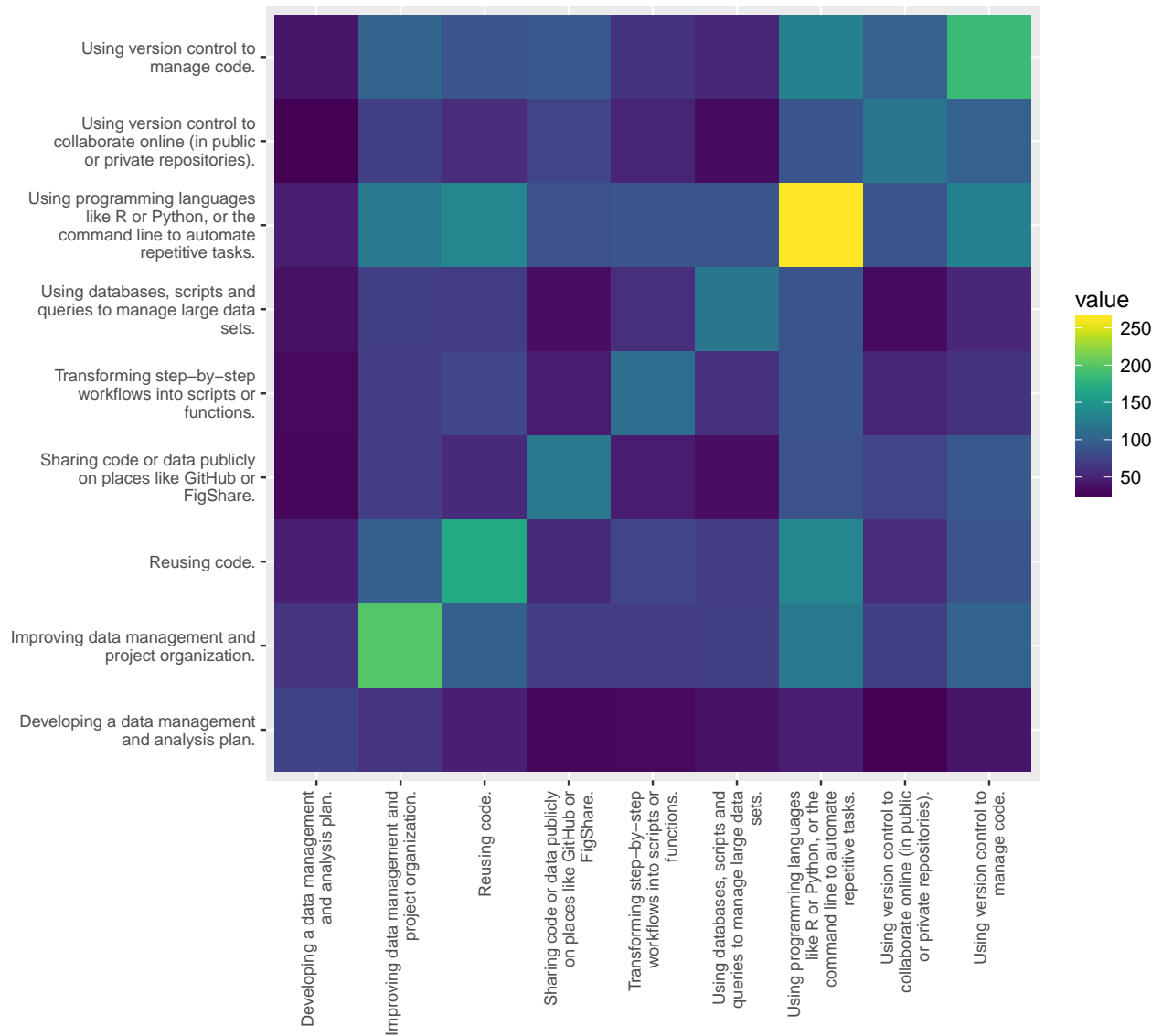
Behaviors	n	%
Using programming languages like R or Python, or the command line to automate repetitive tasks.	260	66.3
Improving data management and project organization.	198	50.5
Using version control to manage code.	185	47.2
Reusing code.	169	43.1

Behaviors	n	%
Sharing code or data publicly on places like GitHub or FigShare.	122	31.1
Using databases, scripts and queries to manage large data sets.	119	30.4
Using version control to collaborate online (in public or private repositories).	119	30.4
Transforming step-by-step workflows into scripts or functions.	111	28.3
Developing a data management and analysis plan.	74	18.9

The matrix below shows a count of the highest combinations of behaviors adopted. Note that the 102 respondents who reported improving their data management and project organization also now use version control to manage code.

	Developing a data ma
Developing a data management and analysis plan.	
Improving data management and project organization.	
Reusing code.	
Sharing code or data publicly on places like GitHub or FigShare.	
Transforming step-by-step workflows into scripts or functions.	
Using databases, scripts and queries to manage large data sets.	
Using programming languages like R or Python, or the command line to automate repetitive tasks.	
Using version control to collaborate online (in public or private repositories).	
Using version control to manage code.	

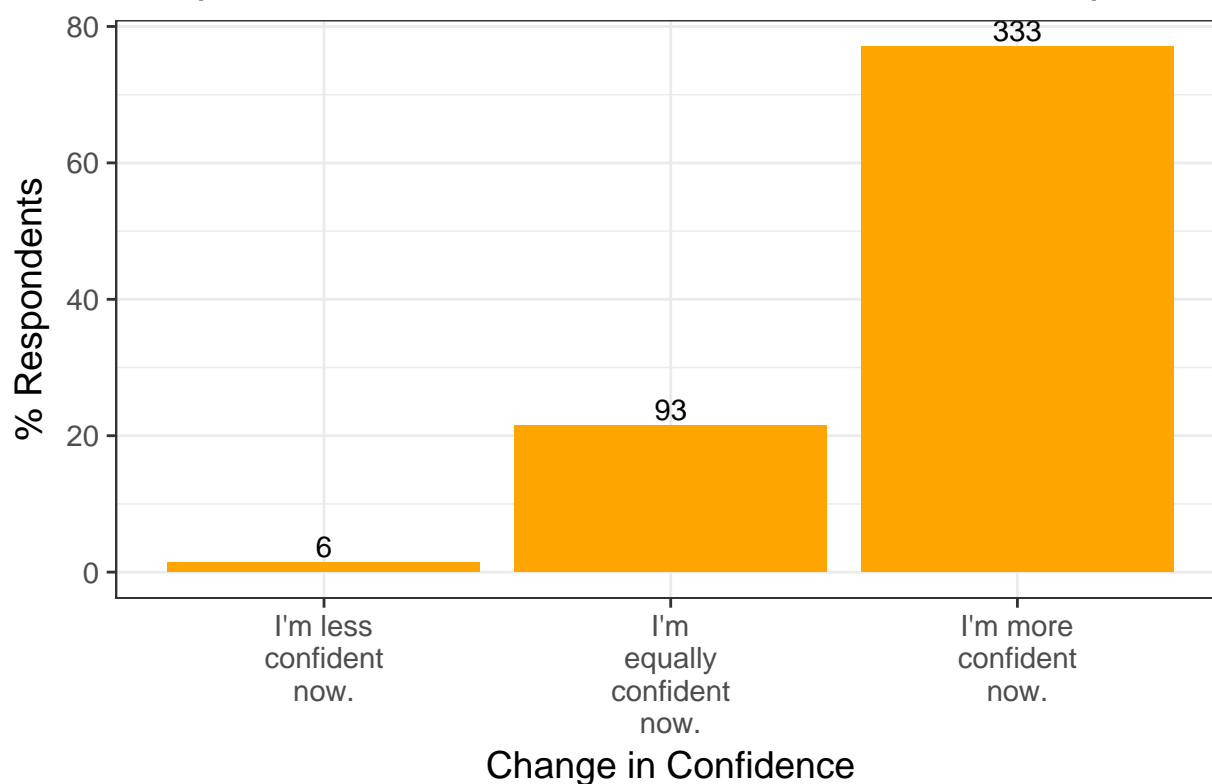
Heatmap of behaviour co-occurrences



### Change in Confidence

Our goal is for learners to leave a workshop with increased confidence about using the tools they were taught. More than 75% of the respondents are now more confident in using the tools they learned than before attending the workshop.

## Respondents are More Confident Post-Workshop



### Usage of Tools for Research and/or Work

We identified specific outcomes directly related to research and/or work, and asked learners if they had achieved these outcomes post-workshop. Respondents reported that the tools they learned improved their overall efficiency, as well as their ability to manage and analyze data.

How Tools Covered Have Helped	n	%
They are improving my overall efficiency.	245	60.5
They are improving my ability to analyze data.	217	53.6
They are improving my ability to manage data.	203	50.1
I am not using the tools I learned.	61	15.1
The tools I learned have not helped me with my work.	28	6.9

Only 28 respondents said the tools they learned have not helped them, and 61 respondents have not been using the tools that were covered in their workshop.

### Contributions to Academic Writing

Another possible outcome of attending a Carpentries workshop is in the use of tools learned to contribute to academic writing (i.e. a grant proposal, journal article).

Contributed-To-Writing	n	%
No.	193	44.5

Contributed-To-Writing	n	%
Not sure.	123	28.3
Yes.	118	27.2

Only 27% of our respondents said that the tools they learned contributed to their academic writing. This is an opportunity for us to explore resources we can offer the community to help use the tools for academic writing purposes.

### Continuous Learning

A key finding is that learners continue their learning after completing a workshop. This can take many forms, including participating in short courses (in-person and online) and using self-guided material. We asked respondents to tell us which learning activities (for data management and analysis) they have participated in since completing a Carpentries workshop. The majority of respondents have used non-Carpentries, self-guided material, though 68 responded having used Carpentries' self-guided material. Additionally, greater participation in meetups and in-person short courses has been reported by respondents.

Continuous Learning	n	%
Used non-Carpentry self-guided material.	127	35
Used self-guided Carpentry lesson material.	68	19
Participated in an in-person short course.	59	16
Participated in an online short course.	45	13
Participated in a Meetup.	35	10
Participated in a semester long course.	24	7

The matrix below provides a breakdown of the combination of continuous learning activities respondents participated in. For example, 35 respondents have used both Carpentry and non-Carpentry self-guided material since attending a workshop.

Table 9:

	Participated in a Meetup.	Participated in a semester long course.	Participated in both
Participated in a Meetup.	35	8	8
Participated in a semester long course.	8	24	8
Participated in an in-person short course.	12	8	8
Participated in an online short course.	16	8	8
Used non-Carpentry self-guided material.	17	12	12
Used self-guided Carpentry lesson material.	9	3	3

### Involvement in the Carpentries

Learners often become actively involved with Software and/or Data Carpentry after completing a workshop. This involvement can take many forms - joining a mentoring group, becoming a workshop helper, or even becoming an instructor. The table provided below shows how respondents have involved themselves with the Carpentries since completing a workshop. Respondents were asked to check all that apply.

Table 10: Respondents Self-Reported Involvement in the Carpentries Post-Workshop

Involvement Post-Workshop	n
Subscribed to the newsletter.	120
Became a workshop helper.	31
Became a Carpentry instructor.	26
Attended at least one community call.	18
Contributed to a Carpentry lesson.	18
Joined a mentoring group.	12
Participated in a Twitter chat.	11
Joined a committee.	6

The matrix below displays frequent combinations of post-workshop involvement. For example, 16 of the respondents who became Carpentry instructors have attended at least one community call.

	Attended at least one community call.	Became a Carpentry instructor.	Became
Attended at least one community call.	18		16
Became a Carpentry instructor.	16		26
Became a workshop helper.	9		15
Contributed to a Carpentry lesson.	12		17
Joined a committee.	2		3
Joined a mentoring group.	6		9
Participated in a Twitter chat.	4		4
Subscribed to the newsletter.	14		17

## Growth Opportunities

We are very excited to know that our workshops are having an impact on learners six months to a year after their attendance. Though the results of this survey are compelling, we do recognize issues for improvement. For example, 48.5% of respondents completed a workshop in the United States. We are continually discussing ways to broaden participation of our workshops in communities outside of the U.S.

We also realize there are some lessons that are not being taught as frequently as others, namely, the Shell lesson and the SQL lesson. We would like to understand why, and what we can do as a community to see an increase in this lesson being taught.

Lastly, 27% of the respondents indicated feeling **neutral** about receiving professional recognition as a result of participating in a Carpentries workshop, and 9% either disagreed or strongly disagreed that they had received any professional recognition. We would like to explore community development opportunities that will benefit learners' personal as well as professional endeavours, so that time spent acquiring important skills for research is adequately recognized and possibly even rewarded.

## Summary

When our learners have successful experiences in our workshops, they are quick to share this positive experience with others. We asked respondents if they had already recommended our workshop, and 74% said yes!

This initial analysis of how Carpentries workshops have impacted learners long-term has been extremely insightful. In general, our workshops are helping learners improve their efficiency with managing and analyzing data. Learners are taking advantage of online resources to improve their skills, and many see value in becoming involved longer term with our community.

We will revisit this data when we compare the responses of learners who took a workshop more than a year ago with those who have taken a workshop less than six months ago and from six months to one year ago. Additionally, we will continue to use this survey every six months to collect data from new learners so we can monitor their progress, and add to our growing evidence base on assessment.