The Smithsonian Carpentries Group: Data Science Education at the Smithsonian

Data is central to the Smithsonian: from research and collections to public programs and operations, every aspect of the Smithsonian generates and uses data. The Smithsonian's 2017 Strategic Plan includes multiple goals that in order to be successful, require the widespread adoption of digital and data science skills (e.g. Goal 5: "Drive large, visionary, interdisciplinary research and scholarly projects"). The Smithsonian Carpentries Group's purpose is to teach Smithsonian employees and fellows best-practice data science skills and to build a community of learners that continue to connect, share skills, experiences and challenges. The adoption of the Carpentries model at SI has led to improved data literacy and the beginning of an SI community built around data skills that crosses between units and supports Strategic Plan Goal 1: "Be One Smithsonian".

Description of Carpentries Program

The Carpentries is a non-profit organization that teaches foundational computer program coding and data science skills to people worldwide. Carpentries instructors use pedagogical approaches to develop and teach workshops that give learners the skills and perspectives to work effectively and reproducibly with software and data. The workshops are designed for people with little to no prior computational experience; they teach not only introductory programming skills but also thought processes and techniques to apply these skills in real-world contexts. In addition, the Carpentries Instructor Training Program educates participants on how to effectively teach data science skills, enabling them to run workshops of their own and promoting coding and data knowledge among diverse communities.

Worldwide, the Carpentries has delivered more than 2,100 workshops to more than 50,000 learners on seven continents. The lessons designed for Carpentries workshops utilize real scientific datasets and demonstrate skills that researchers would use to answer real scientific questions. Learners report that they are able to model their own work on examples shown during workshops, enabling them to accomplish tasks that had previously been beyond their abilities.

Origin of Smithsonian Carpentries Program in late 2017

Beginning in 2015, a small group of OCIO and NMNH employees and fellows convened an ad hoc "SI Bioinformatics Working Group," which coordinates training, weekly "office-hours" style brown bags, and user support on the SI High-Performance Computing Cluster. Initially, the training was focused on the bioinformatics tools needed for biodiversity genomics research. As adoption of these tools increased across the SI science units, users began to request more general training in data science skills, including programming languages R and Python.

Instead of continuing to create custom workshops for these broad topics (which is both time-consuming and can reflect the subject-matter bias of the person preparing the materials), we decided to become members of the Carpentries organization. As a member organization, we have access to teaching materials developed by academics and data science professionals, we can use our own expertise and experience to augment existing materials and create new

lessons, and we can offer Instructor training to SI staff to grow Smithsonian capacity for effective data science education.

The Smithsonian joined the Carpentries in autumn of 2017 and trained an initial cohort of seven SI staff as Instructors. This group went through a two-day Instructor Training workshop, which focuses on pedagogical techniques for effective teaching. This training allowed us to learn how to make our workshops more inclusive, accessible, and more valuable for audiences across SI -- not just the biologists we had the most experience working with up until that point.

The first Smithsonian Carpentries workshop was held in February 2018 at NMNH. With the guidance of an experienced Instructor from UC Davis, the Smithsonian Carpentries team instructed 37 SI-affiliated participants in commonly used data skills, including working with spreadsheets, cleaning data, programming and graphing in Python, and querying databases with SQL.

Growth of SI Carpentries Program from NMNH-focused to pan-Smithsonian

After the success of our first year, and in order to fulfill the high demand for more workshops going forward, we decided to expand the SI Carpentries group by training 15 additional Instructors in Fall 2018.The increased number of instructors allowed us to start offering a standing quarterly workshop in the SI Castle Library, which we began in February 2019. Every quarterly workshop (February, May, and August 2019) has filled up in hours and had a waiting list. In addition to this standing quarterly workshop, we continue to travel to units to offer workshops (e.g. SERC, SCBI Front Royal) when there is need; three additional workshops have or will be offered in 2019, one at SERC, one at SCBI Front Royal and one specifically for interns and fellows.

Over eight total Data Carpentry workshops, we have had a total of 218 attendees from 25 different Smithsonian units. A full list of workshops delivered by the Smithsonian Carpentries team can be found at datascience.si.edu/carpentries. In Fall 2019, we will increase our SI Carpentries group with a new set of Instructors, drawing from 22 applications from 14 different units.

Impact of the workshops

The Carpentries organization delivers a pre- and post-workshop survey to all workshop attendees, so that impact can be measured. Our Carpentries workshops have shown that after the workshop 80% of the attendees feel they can write a short computer program to aid them in their work, compared to 30% before attending a workshop. Our own survey to Carpentries workshop alumni showed that 65% of respondents have incorporated techniques from the workshops in their regular work.

On a scale of 0 - 100, how likely are you to recommend this workshop to a friend or colleague?
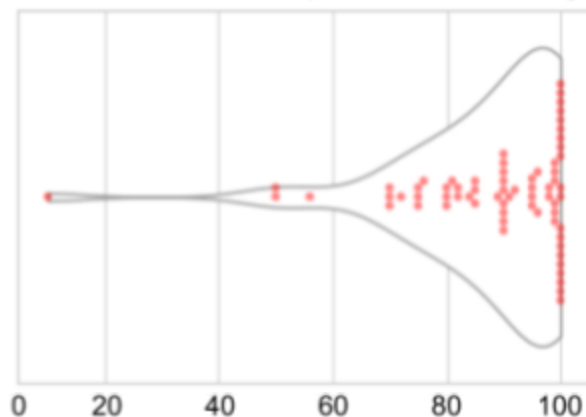
Figure 1: Distribution of workshop attendees' likelihood to recommend to a friend or colleague. Each red dot represents a single respondent (n= 68).
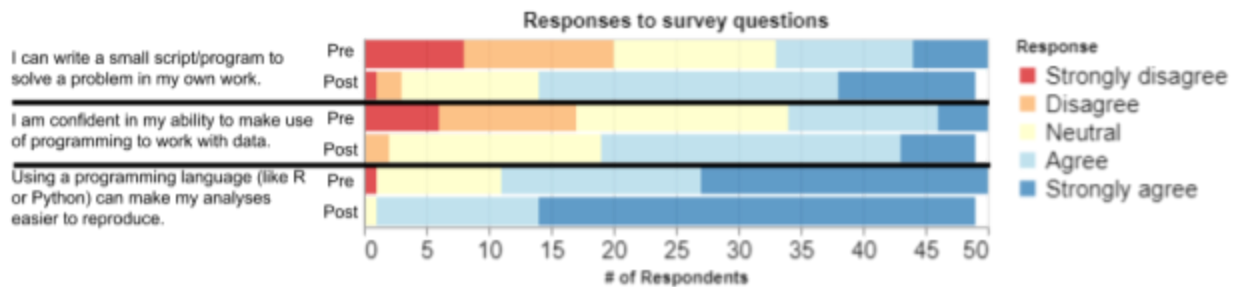
Figure 2: Comparison of responses to pre- and post-workshop survey questions relating to attendees' ability to accomplish data-related tasks. (n=50, which was the subset of survey respondents that could be matched between pre- and post-workshop surveys)

Innovations of SI Carpentries

Because of the diffuse nature of the Smithsonian, we recognized that maintaining a community of practice after each workshop ended was going to be a problem for attendees. To address this issue, the Carpentries team came up with two solutions. The first was to leverage the Smithsonian's subscription to Slack (a popular collaboration and messaging platform) to create a "carpentries-alumni" channel. In this channel, Instructors and participants of Carpentries workshops ask data-related questions, provide links to helpful articles, and share upcoming events. The second solution was to create a monthly Carpentries Brown Bag series. Each month, on the fourth Thursday, Carpentries Instructors and previous attendees of workshops gather together during lunch hour, and someone gives a short presentation related to a Carpentries topic. The rest of the lunch hour is then open for participants to discuss any other topics, or ask for advice.

Another innovation the Smithsonian Carpentries group is working on is to build a Washington DC area network of other local Carpentries groups, such as the Library of Congress, NIST, NIH, and George Washington University.

Summary

The Smithsonian Carpentries team has refined the materials and practices of the Carpentries organization to specifically suit the needs of the Smithsonian. Our workshops have impacted 25 SI units, empowering workshop participants to process and gain new insights into their data. When polled, over 90% of our participants analyze data that directly or indirectly contributes to exhibitions, websites, public programs or other interactions with the public.

Workshops impact both staff within the SI as well as shorter-term interns and fellows. The interns and fellows trained at the SI over the past two years have spread what they've learned beyond the SI to their colleges, universities, and future workplaces. By developing the SI-carpentries-alumni slack channel and brown bag lunches we are building a robust infrastructure for continued interaction, not limiting the Carpentries mission to only the workshops, but making it a long-term part of SI participants' careers.