Software and Data Carpentry

Jonah Duckles

Executive Director
Software Carpentry
jduckles@software-carpentry.org
@jduckles

The Carpentries

Jonah Duckles

Director of Membership
Software Carpentry
jduckles@carpentries.org
@jduckles

Are we using the right tools for data driven research?

Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics

Barry R Zeeberg†, Joseph Riss†, David W Kane, Kimberly J Bussey, Edward Uchio, W Marston Linehan, J Carl Barrett and John N Weinstein

† Contributed equally

BMC Bioinformatics 2004 5:80 | DOI: 10.1186/1471-2105-5-80 | © Zeeberg et al; licensee BioMed Central Ltd. 2004

Received: 05 March 2004 | Accepted: 23 June 2004 | Published: 23 June 2004

Automatic conversion of gene symbols to dates and floating-point numbers is a problematic feature of Excel software.

— Zieman et al.

What are the right tools?

With growing rate of data accumulation, there is an acute need for all researchers to learn about:

— Repeating common tasks

— cleaning/reading/processing data

— sharing code for common tools and methods

— collaborating with code

Departure point

Every discipline is accumulating data at unprecendented rates

Skills lag far behind needs in ~~many~~ most disciplines

These skills are incredibly marketable in both science & business careers

Technical consulting

doesn't scale

Online training and MOOCS

can leave people frustrated, without a community to help them

Our approach is to build community

and capacity through peer instruction

What does it take to have a workshop with

IMPACT

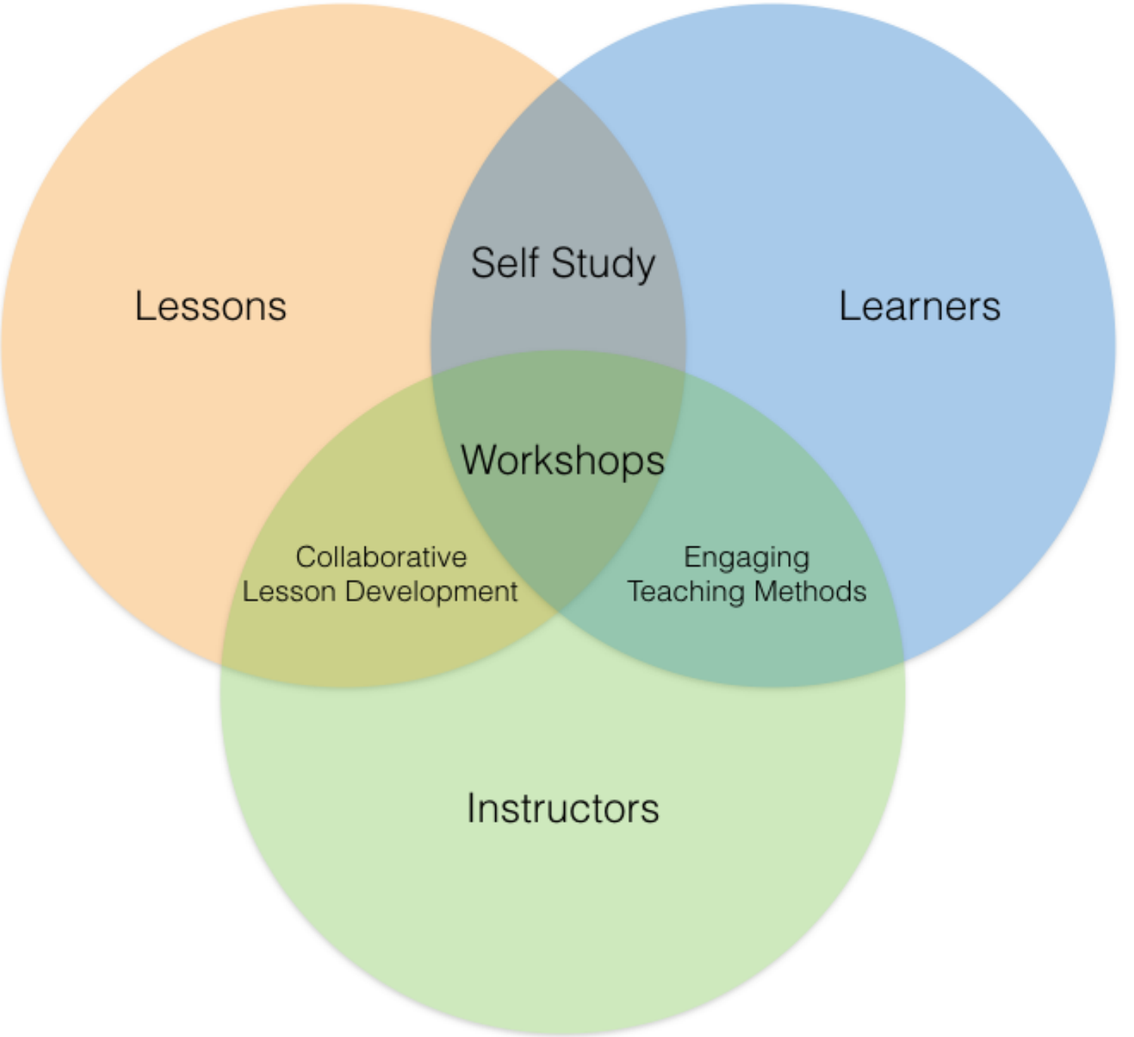
The Carpentries

Quality LESSONS

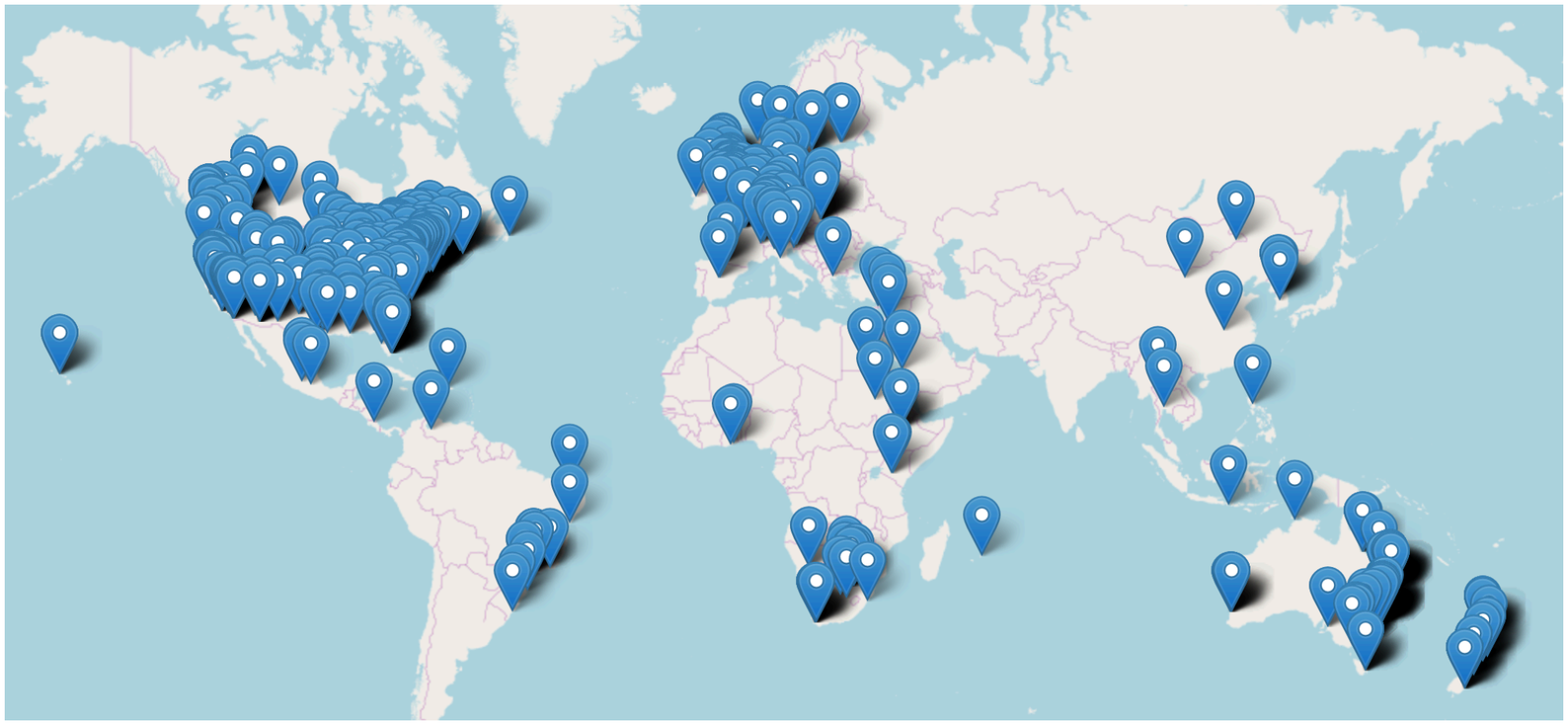
Motivated LEARNERS

Trained INSTRUCTORS

Lead to workshops with IMPACT


Lessons
Self Study
Learners
Workshops
Collaborative
Lesson Development
Engaging
Teaching Methods
Instructors

It's working



n = 1,350 on all 7 continents yes, even Antarctica

The Carpentries

Impactful Workshops

— Trained instructors who are themselves researchers

— Community developed lessons

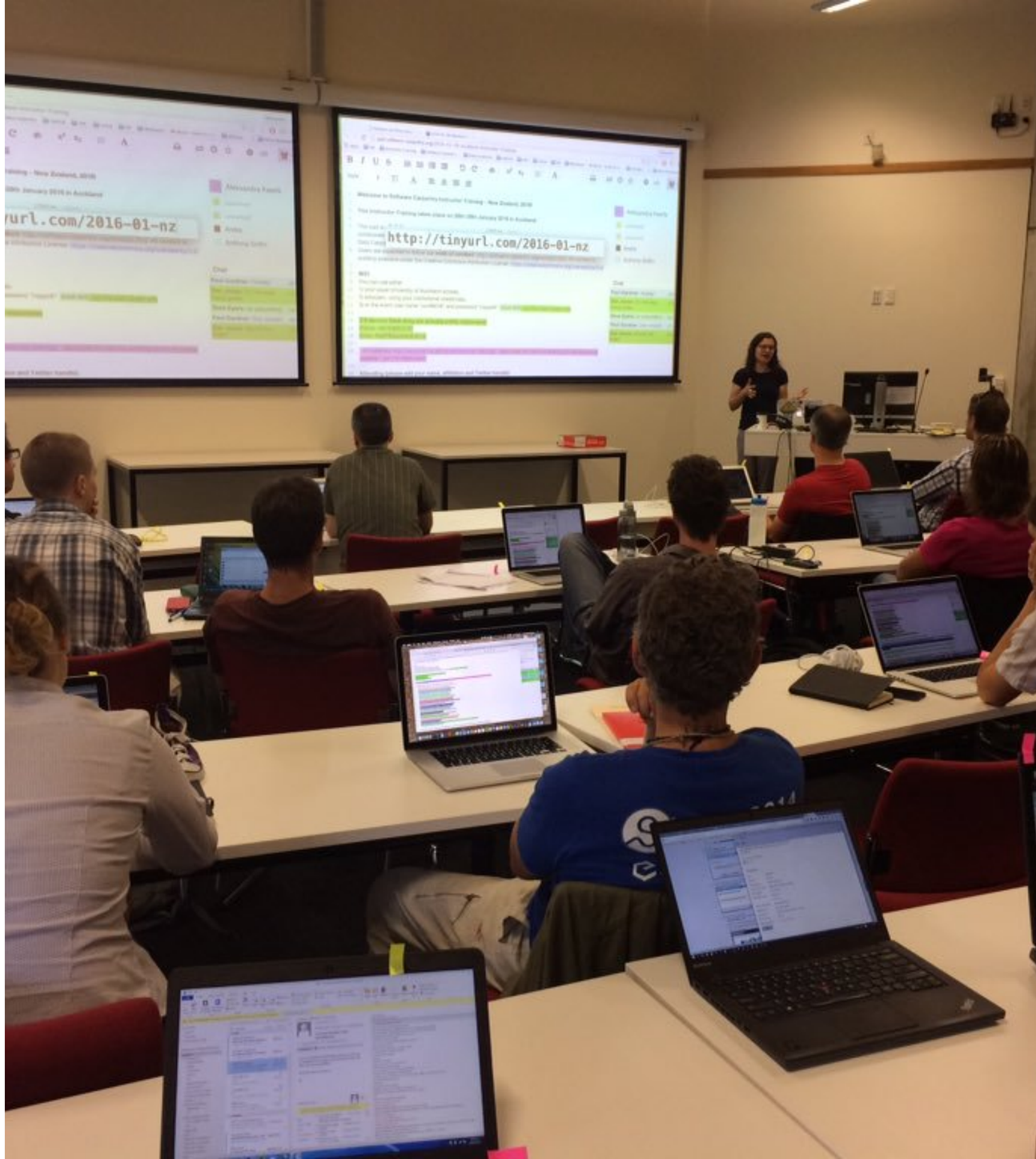
— Passion to apply the skills and spread the community

— A mission to change the culture of how research is done

Instructors

Instructor training

- Instructor training prepares researchers to impactfully teach technical skills:
  - live coding
  - take into account learner's background
  - reducing cognitive load
  - improve self as an instructor over time
  - give and receive constructive feedback



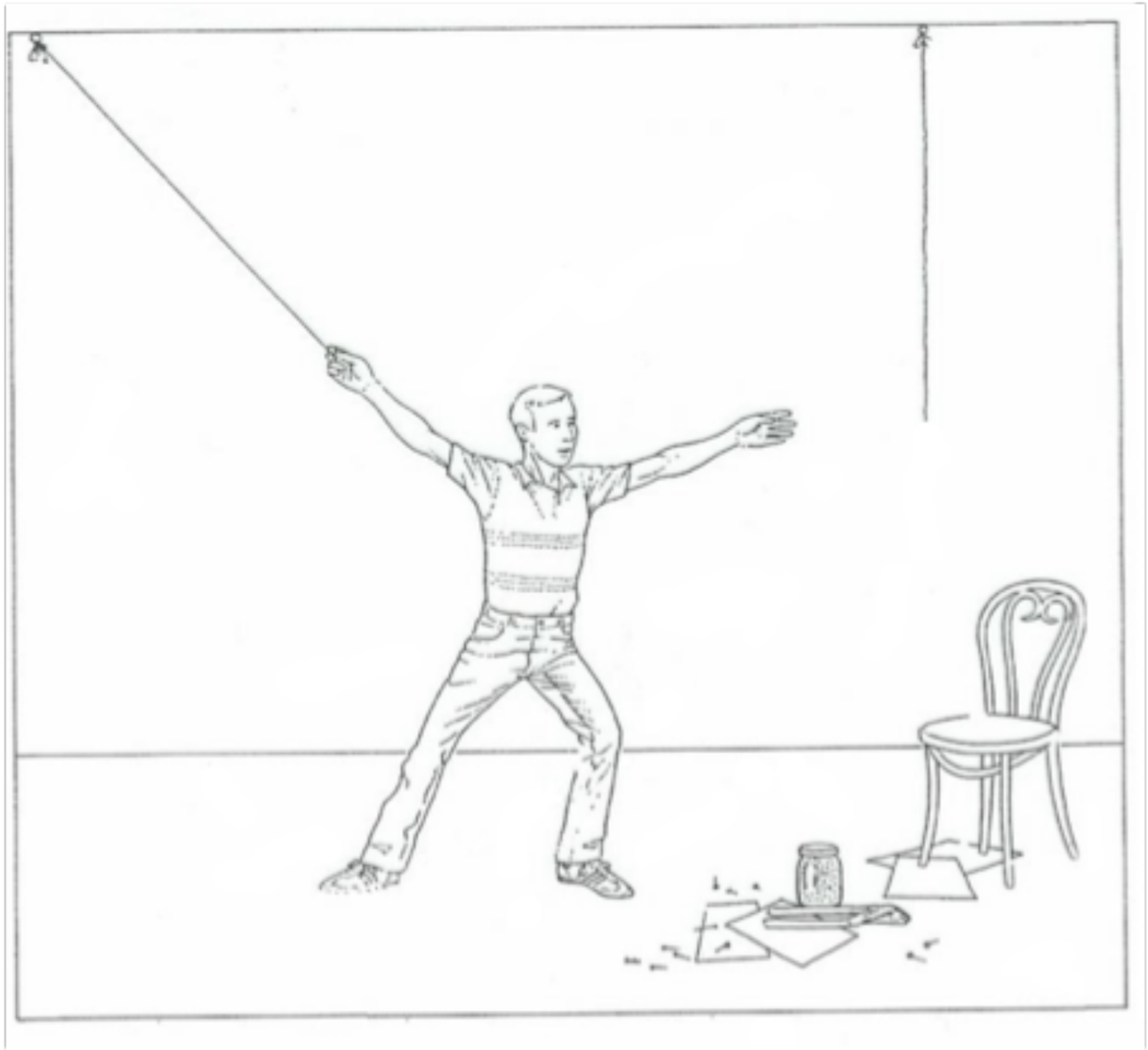
Teaching as performance art

— Excitement
— Engagement
— Passion for the topic
— Improv
— Lessons are a loosely sketched script

Lessons

Preparing the unconscious mind

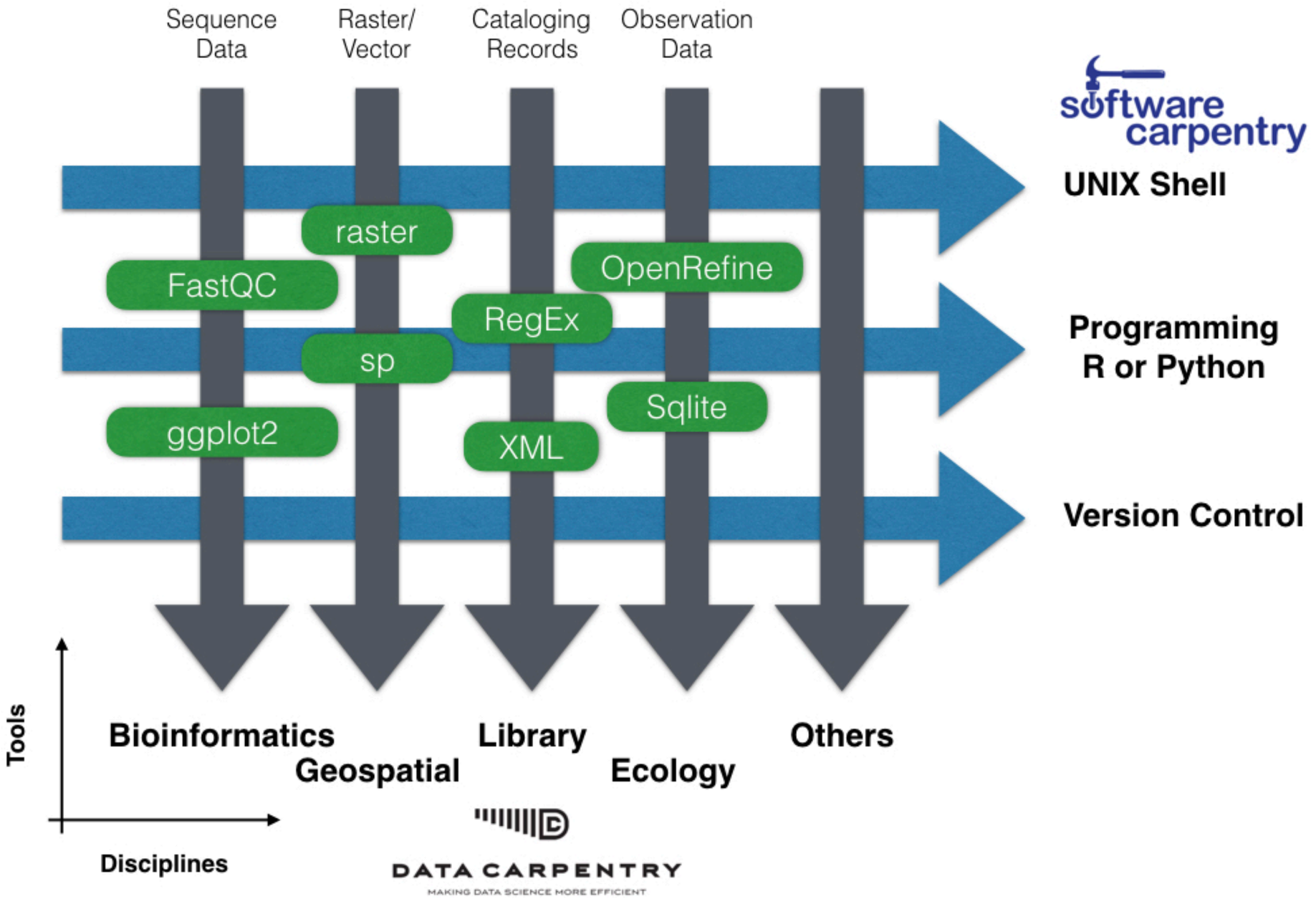
- Two day constraint
- N.R.F. Maier's two cords experiment
- Problem solving skills can be inaccessible to the conscious mind
- You can't teach all the things



Jugyokenkyu - lesson study

- Coordinated collaboration, testing and continuous improvement of lessons.
- Collaboration on lessons, conversation about teaching of lessons
- Instructor community that discusses the ongoing improvement of a lessons
- Kaizen of teaching

Workshops

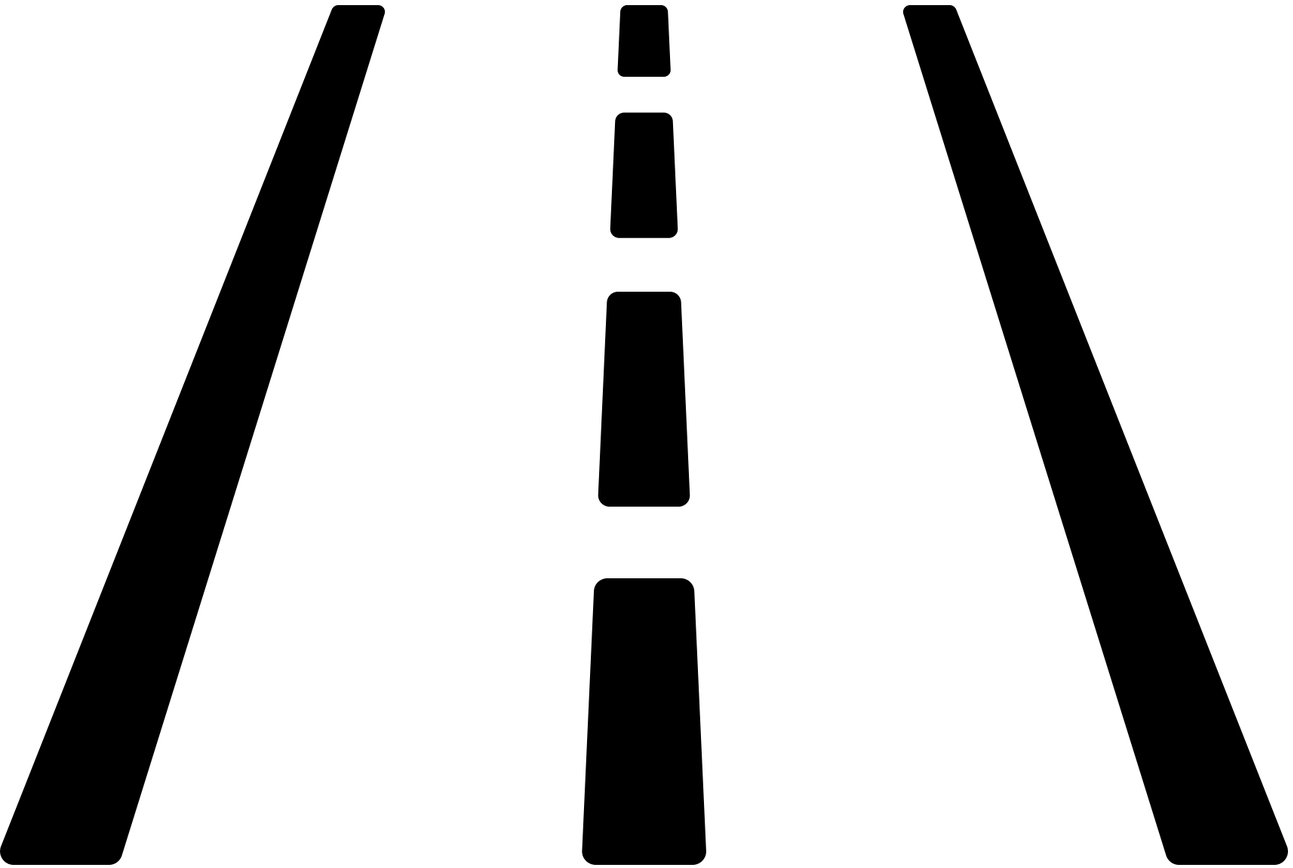

Sequence Data
Raster/ Vector
Cataloging Records
Observation Data
software carpentry
UNIX Shell
raster
FastQC
OpenRefine
RegEx
sp
Programming R or Python
Sqlite
ggplot2
XML
Version Control
Tools
Bioinformatics
Library
Others
Geospatial
Ecology
Disciplines
DATA CARPENTRY
MAKING DATA SCIENCE MORE EFFICIENT

Lesson	Site	Repository	Reference	Maintainer(s)
The Unix Shell				Gabriel Devenyi, Ashwin Srinath
Version Control with Git				Ivan Gonzalez, Daisie Huang
Version Control with Mercurial				Doug Latornell
Using Databases and SQL				Abigail Cabunoc Mayes, Sheldon McKay
Programming with Python				Trevor Bekolay, Valentina Staneva
Programming with R				Daniel Chen, Harriet Dashnow
R for Reproducible Scientific Analysis				Thomas Wright, Naupaka Zimmerman
Programming with MATLAB				Isabell Kiral-Kornek, Ashwin Srinath
Automation and Make				Gerard Capes
Instructor Training				Greg Wilson

Community

- Cultivating a community and culture requires ongoing and intentional work
  - Code of Conduct and enforcement
  - Active mentoring of instructors
  - Collaborative lesson development

Getting Started



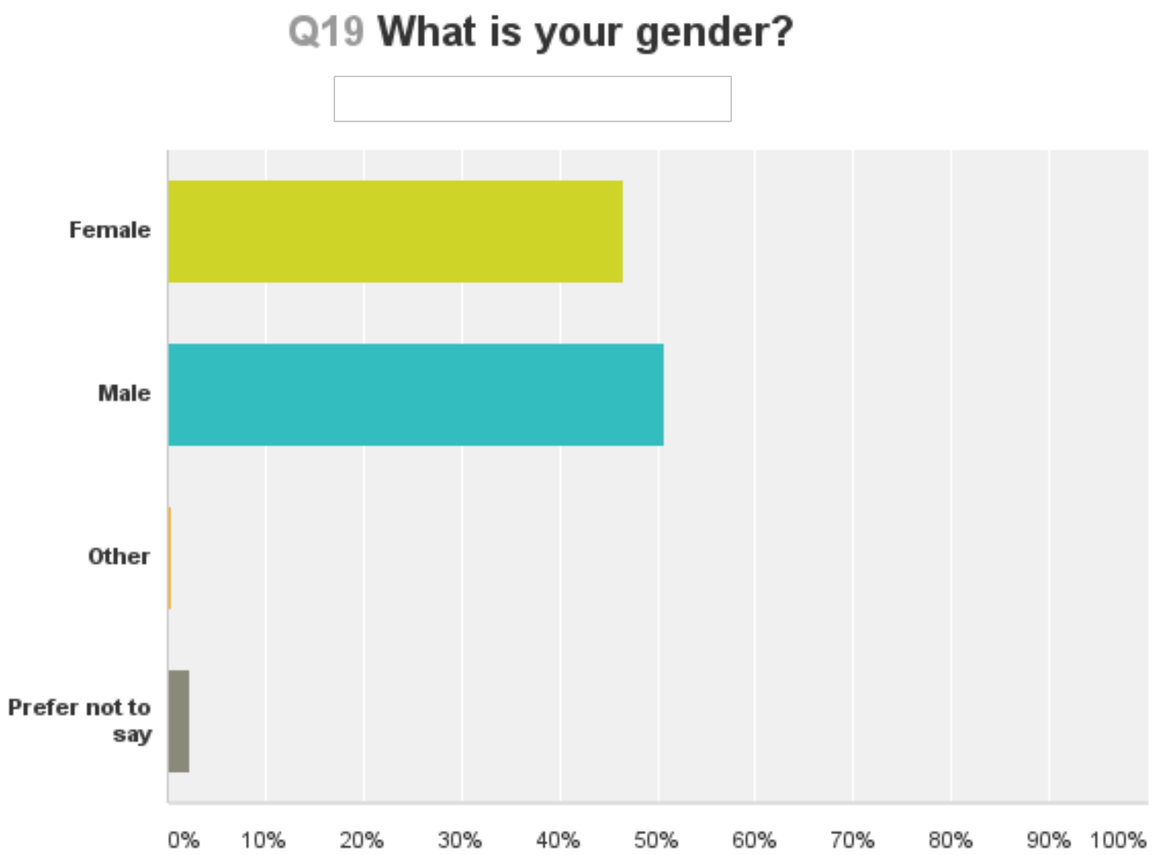
- Run a local workshop
- Build a support coalition
  - Libraries, Deans, Chancellors, Department chairs
- Train local instructors
- Build it into your organization
- Connect to the global community

Learners

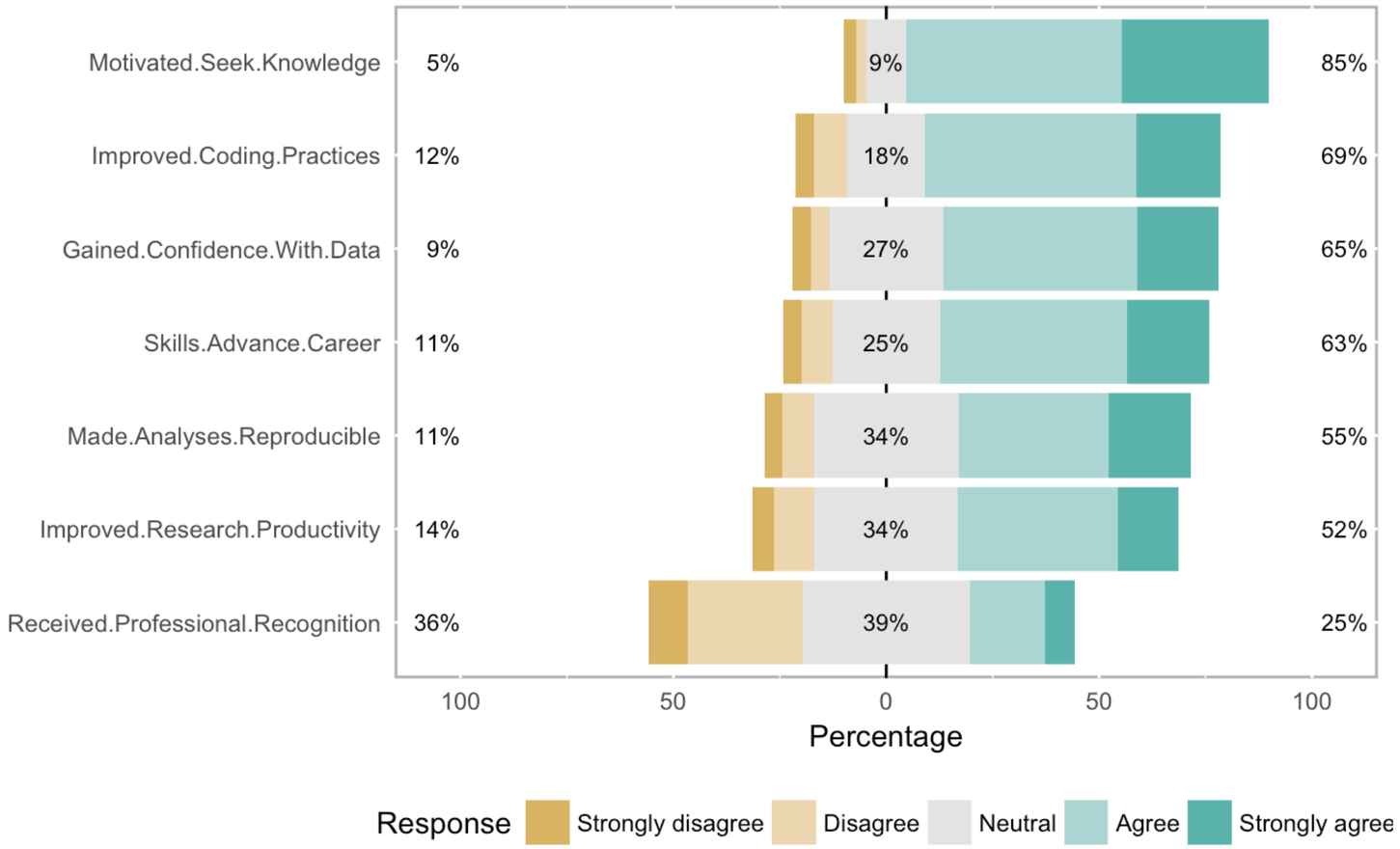
Our workshops are welcoming and relevant to all disciplines

Research Domain	%
Life Sciences (Genetics, genomics, bioinformatics)	24.9
Life Science - Organismal/systems (ecology, botany, zoology, microbiology, neuroscience)	24.0
Planetary sciences (geology, climatology, oceanography, etc.)	6.6
Mathematics/statistics	6.0
Physics	5.8
Civil, mechanical, chemical, or nuclear engineering	4.5
Medicine and/or Pharmacy	4.3
Chemistry	4.0
Social sciences	4.0
Library and information science	3.2
Economics/business	2.6
Humanities	2.6
Psychology	2.4
Education	2.1
High performance computing	2.1
Space sciences	0.9

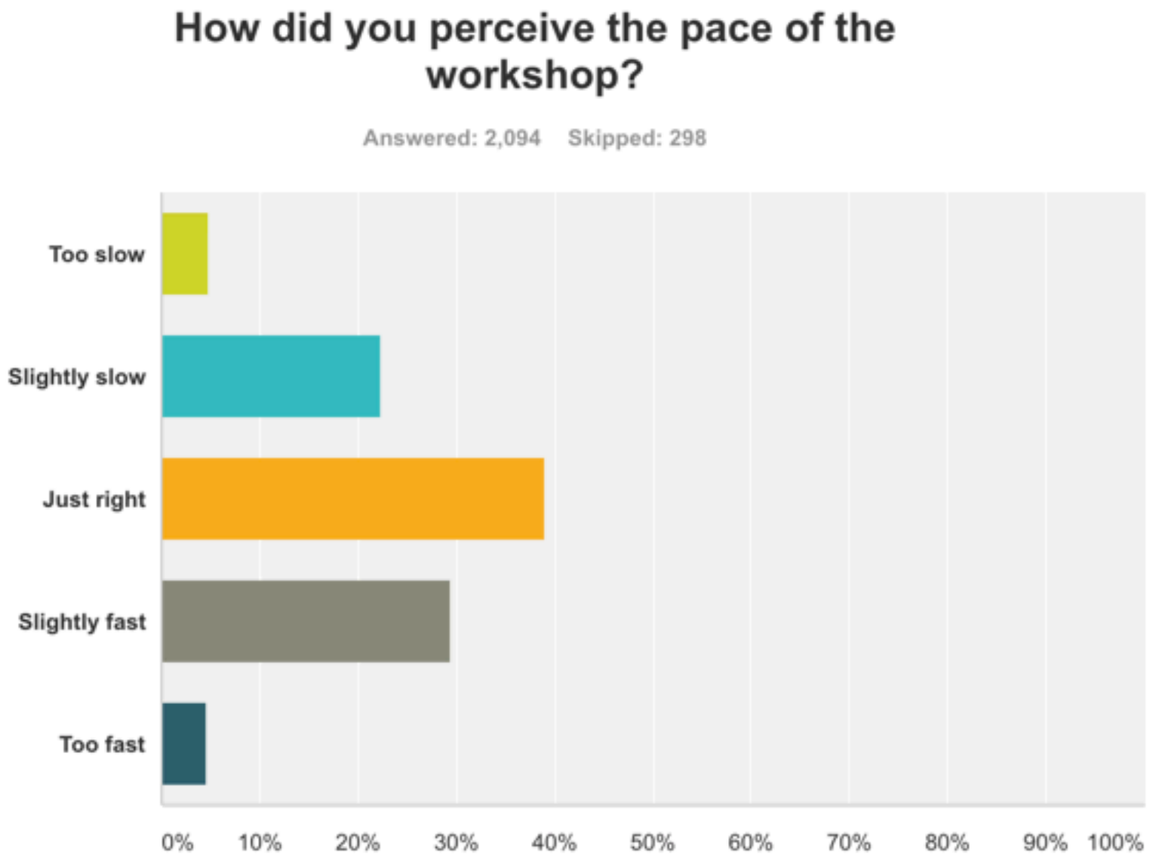
We achieve a very good gender balance



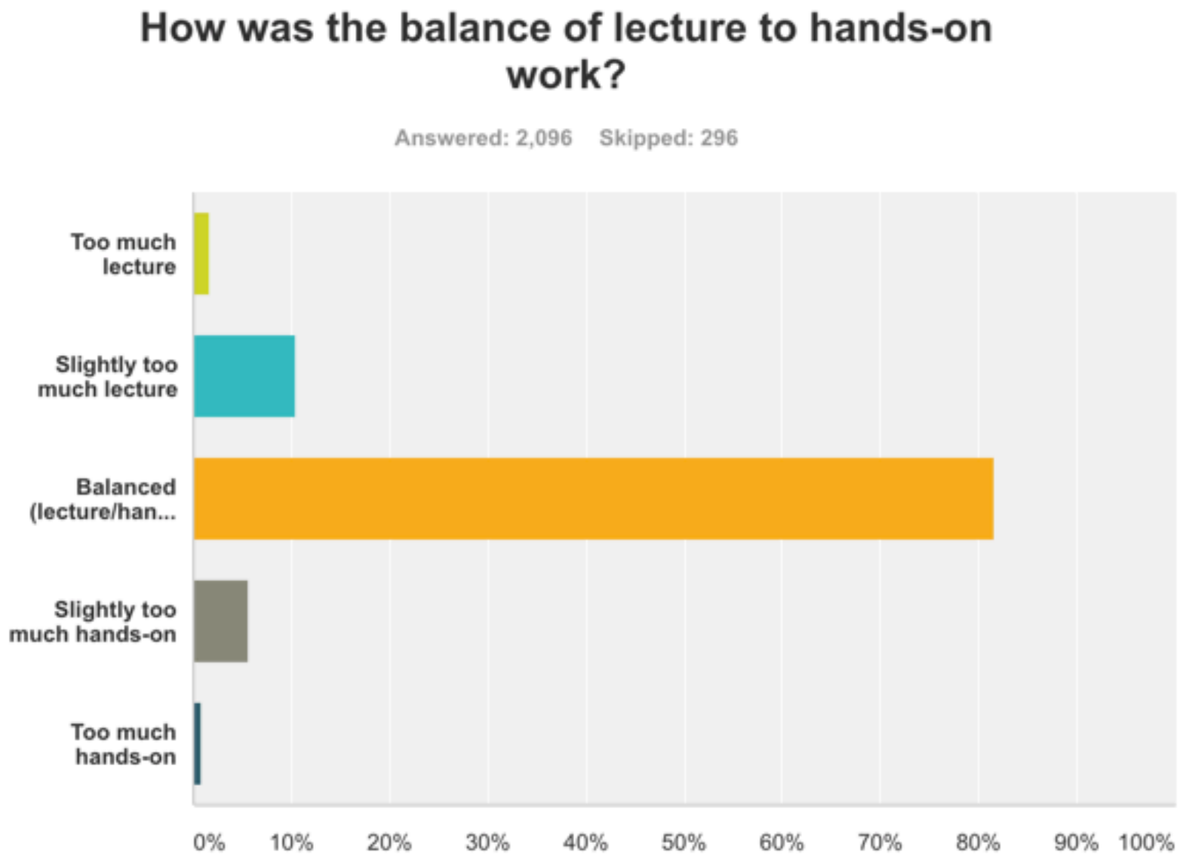
Long-term outcomes for learners



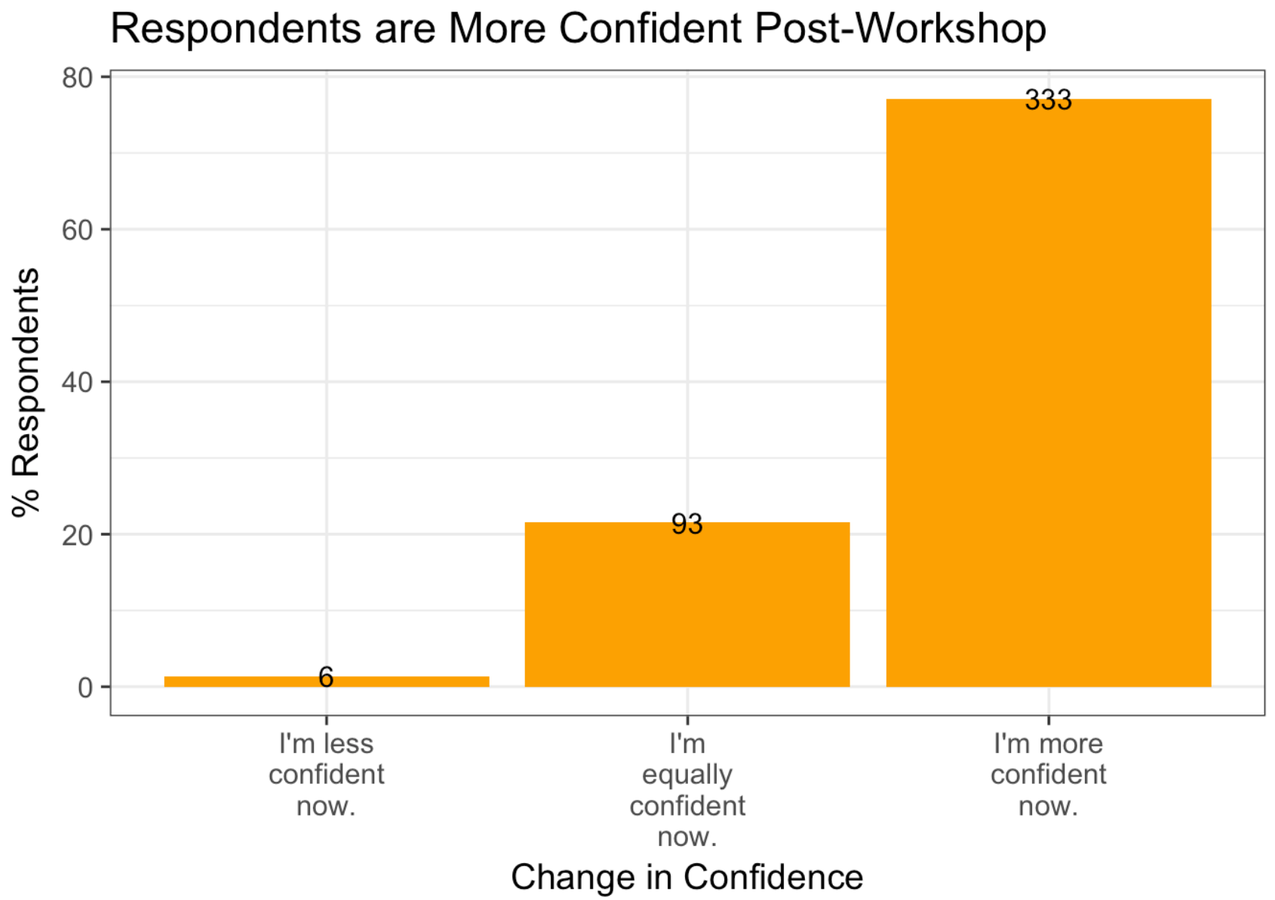
Adapting to learners

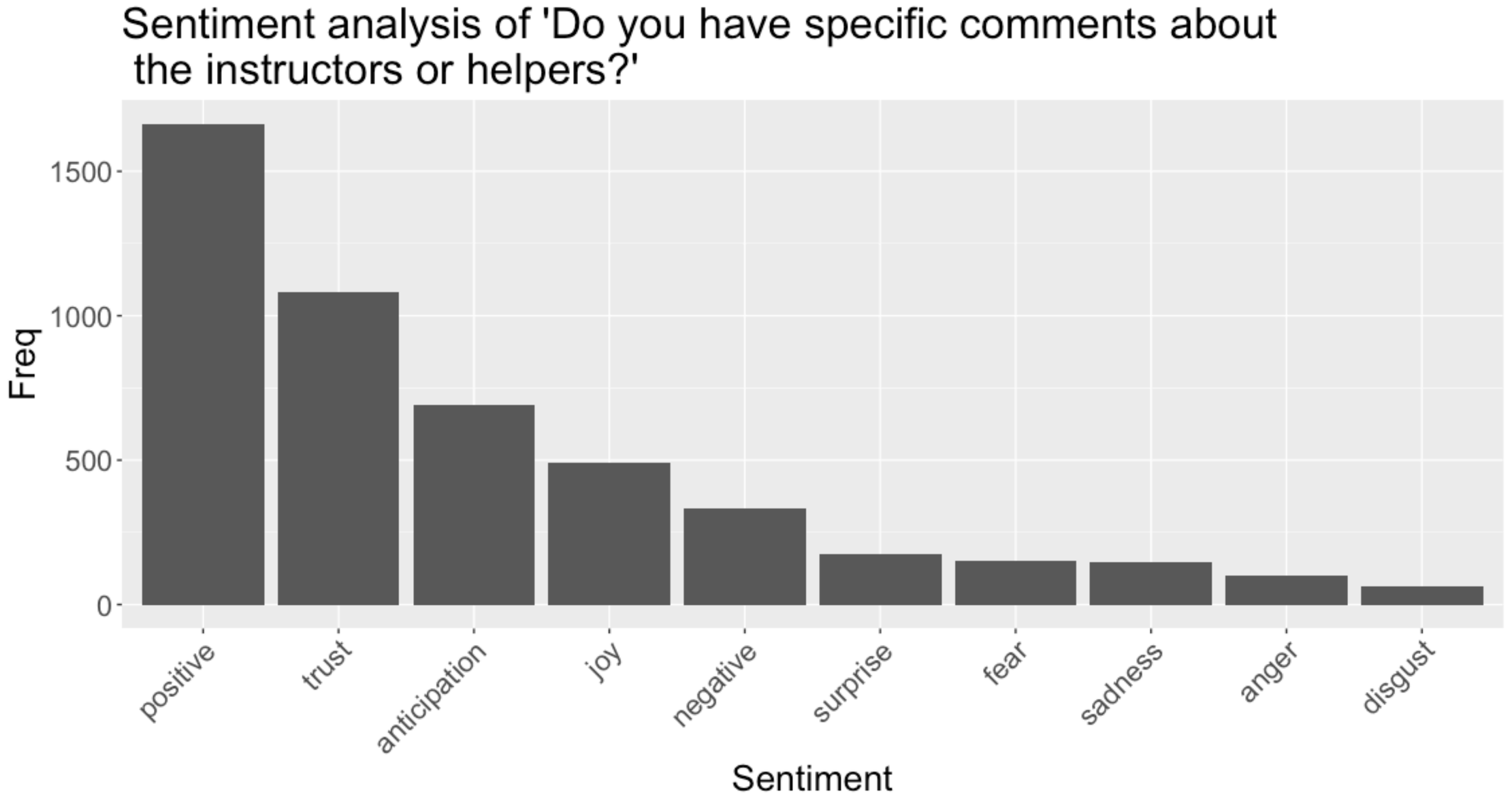


Balancing demonstration and application

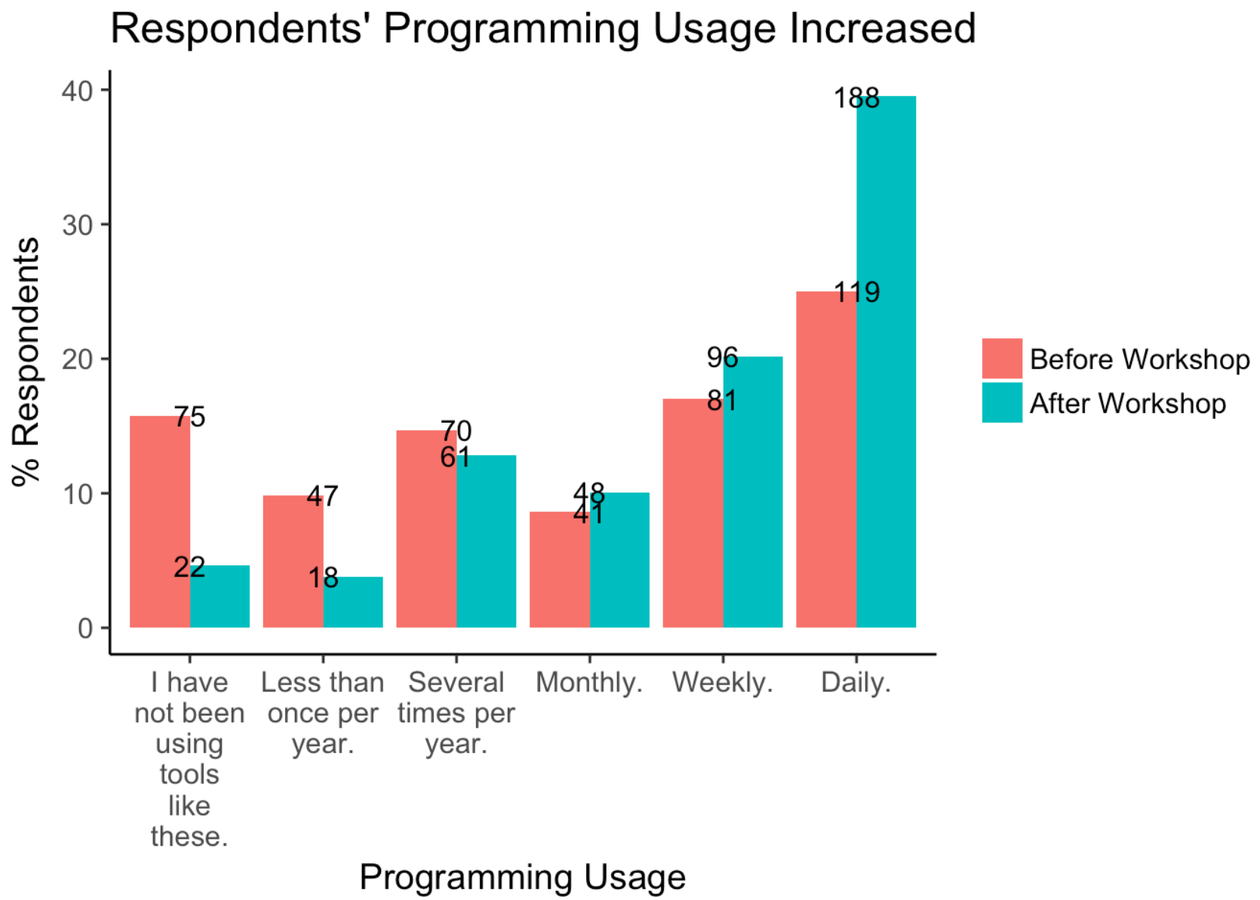


Building confidence

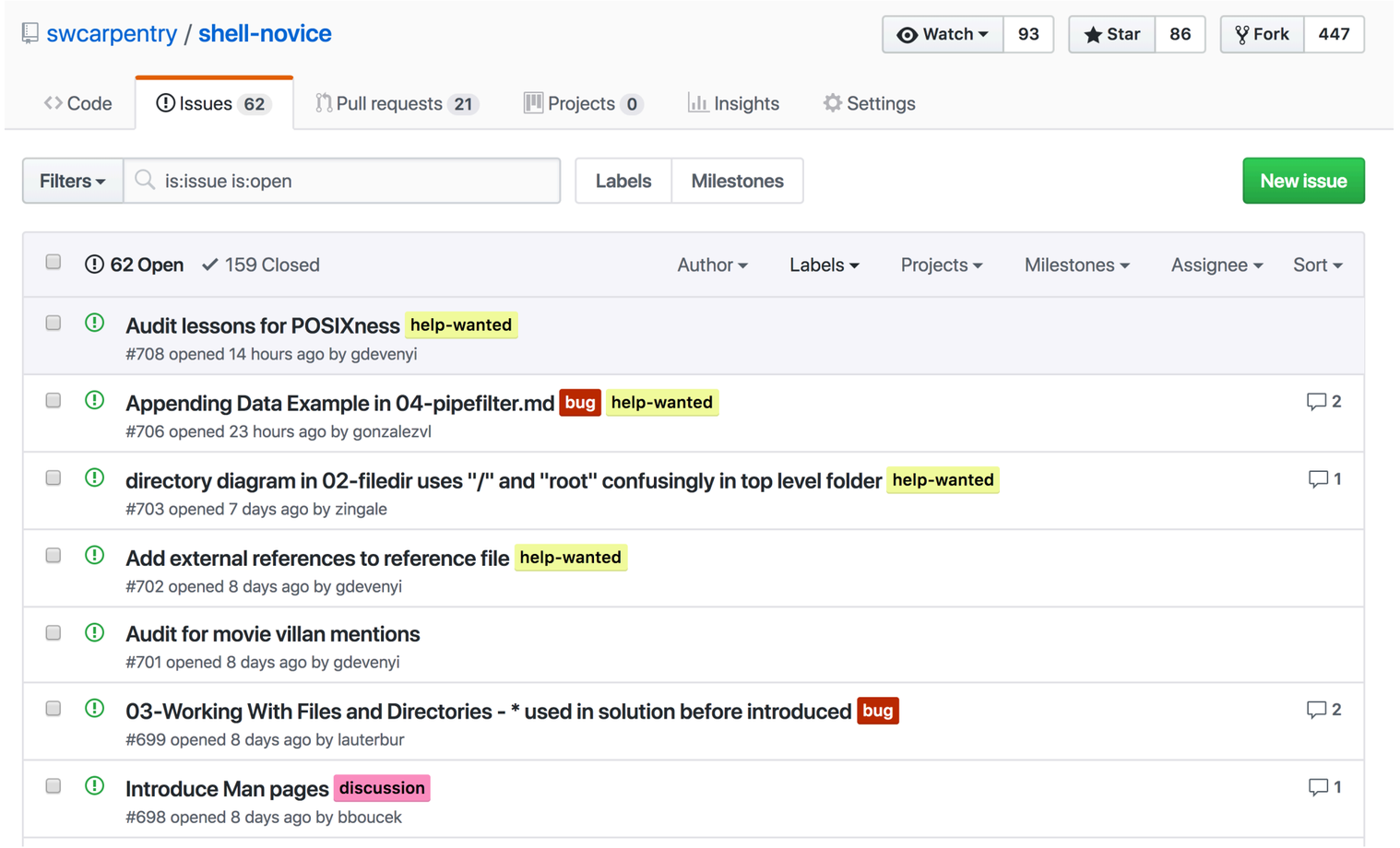



Sentiment analysis of 'Do you have specific comments about the instructors or helpers?'
Freq
1500
1000
500
0
positive
trust
anticipation
joy
negative
surprise
fear
sadness
anger
disgust
Sentiment

Self-assesment of workshop impact on practices



Collaborative lesson development



01-basics - Update language for clarity #473

Edit

Open **caseyyoungflesh** wants to merge 1 commit into `swcarpentry:gh-pages` from `caseyyoungflesh:youngflesh-branch`

Conversation 0 | Commits 1 | Files changed 1 | +16 −17

caseyyoungflesh commented 10 days ago

Changed some of the language to be more clear - changes in sentence structure, defining jargon. Also changed the use of the 'tape' analogy to 'recording' (everyone knows what a recording is!) in reference to issue #461 .

 `01-basics - Update language for clarity` Verified c765238

Add more commits by pushing to the **`youngflesh-branch`** branch on **caseyyoungflesh/git-novice**.

This branch has no conflicts with the base branch

Merging can be performed automatically.

Merge pull request ▾ You can also open this in GitHub Desktop or view command line instructions.

Reviewers

Suggestions

gvwilson

 iglpdc

Assignees

No one—assign yourself

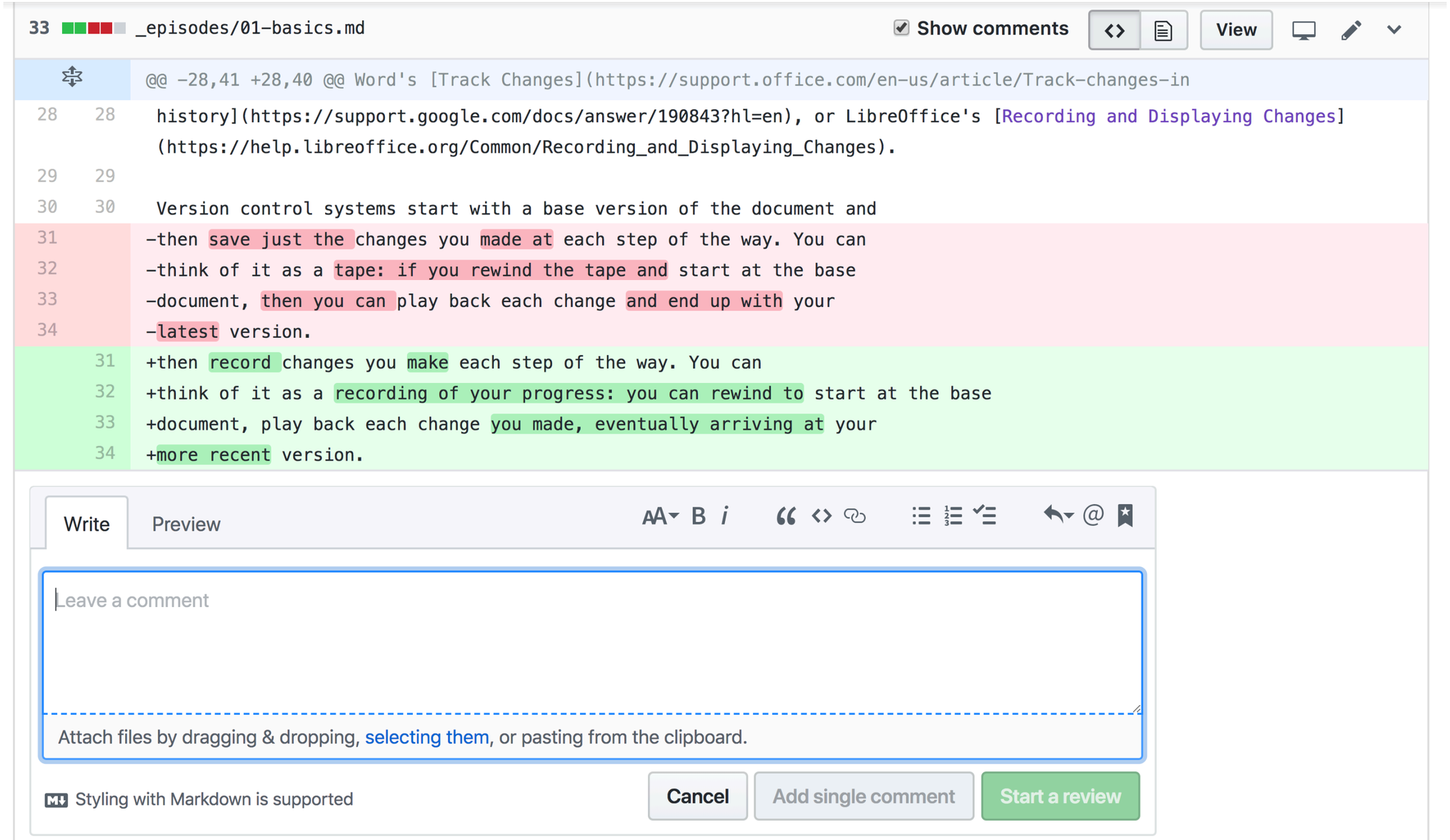
Labels

None yet

Projects

None yet

Milestone


33 _episodes/01-basics.md
Show comments
View
@@ -28,41 +28,40 @@ Word's [Track Changes](https://support.office.com/en-us/article/Track-changes-in
history](https://support.google.com/docs/answer/190843?hl=en), or LibreOffice's [Recording and Displaying Changes]
(https://help.libreoffice.org/Common/Recording_and_Displaying_Changes).
Version control systems start with a base version of the document and
-then save just the changes you made at each step of the way. You can
-think of it as a tape: if you rewind the tape and start at the base
-document, then you can play back each change and end up with your
-latest version.
+then record changes you make each step of the way. You can
+think of it as a recording of your progress: you can rewind to start at the base
+document, play back each change you made, eventually arriving at your
+more recent version.
Write
Preview
Leave a comment
Attach files by dragging & dropping, selecting them, or pasting from the clipboard.
Styling with Markdown is supported
Cancel
Add single comment
Start a review

How to get your organization involved?

http://software-carpentry.org/membership/

Partnership Tiers

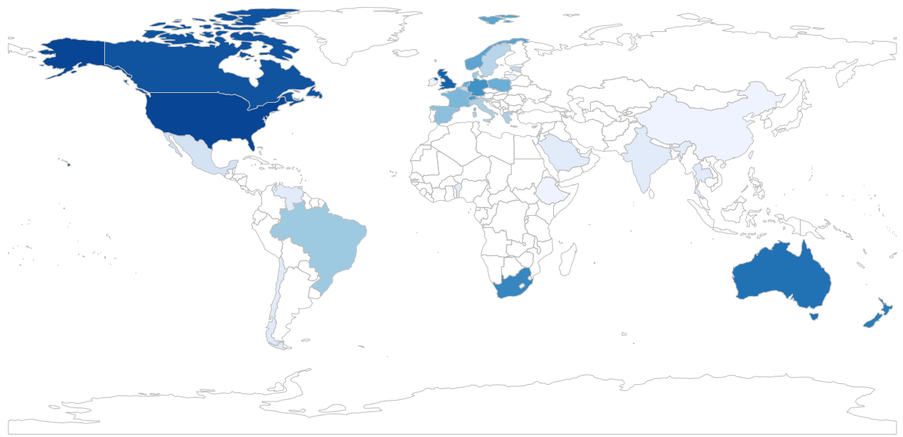
	Bronze	Silver	Gold	Platinum
# of Coordinated Workshops	2	4	6	negotiable
Discount for additional coordinated workshops	20%	33%	50%	negotiable
Self-organized workshops at partner organization **	no-charge	no-charge	no-charge	no-charge
Number of instructors trained ***	0	6 online	15 with possibility for in-person^ training event	negotiable
Seat on the SCF Advisory Board	No	Yes	Yes	Yes
Train an in-house instructor trainer at partner org	No	No	No	Available
Lesson development services	No	No	No	Available
Membership Dues (annual)	$5,000	$7,500	$15,000	Contact us

THE CARPENTRIES

What you Support with Membership

— Instructor training
— Instructor mentorship
— Workshop management database
— Curriculum Development and Maintenance
— Metrics / Assessment infrastructure
— Global communities of practice

CarpentryCon



— When: May 30-June 1, 2018
— Where: Dublin, Ireland
— Moreinfo: carpentrycon.org

The Carpentries Community Cookbook

Introduction

- Introduction
- Building or improving your own local community
- What is a community of practice?
- Codes of Conduct

Initiation

- Getting started

Activities

- Workshops
- Help Sessions
- Hacky Hour
- Carpentries Study Groups
- Events
- Research Bazaar (ResBaz)
- CarpentryCon
- SatRdays
- THATCamp
- International Events
- Vendor Training

Tools

- Communications

Support Structures

- Support Structures
- Sourcing funding for workshops

Appendix

- Appendix

Community Cookbook

https://cookbook.carpentries.org

Thank You to our supporters!

55 Organizations in 10 countries are currently members.

1300 workshops have been taught since 2012 for 35,000 learners by our 1200 instructors from 39 countries.

Other Related Initiatives

— coderefinery.org
— Research Software Engineer Association
— Library Carpentry
— Mozilla Science Lab

Many of our community members and member organizations are active in the above initiatives as well.

Thank You!

Jonah Duckles

Director of Membership
Software Carpentry
jduckles@carpentries.org
@jduckles

Become a member organization:

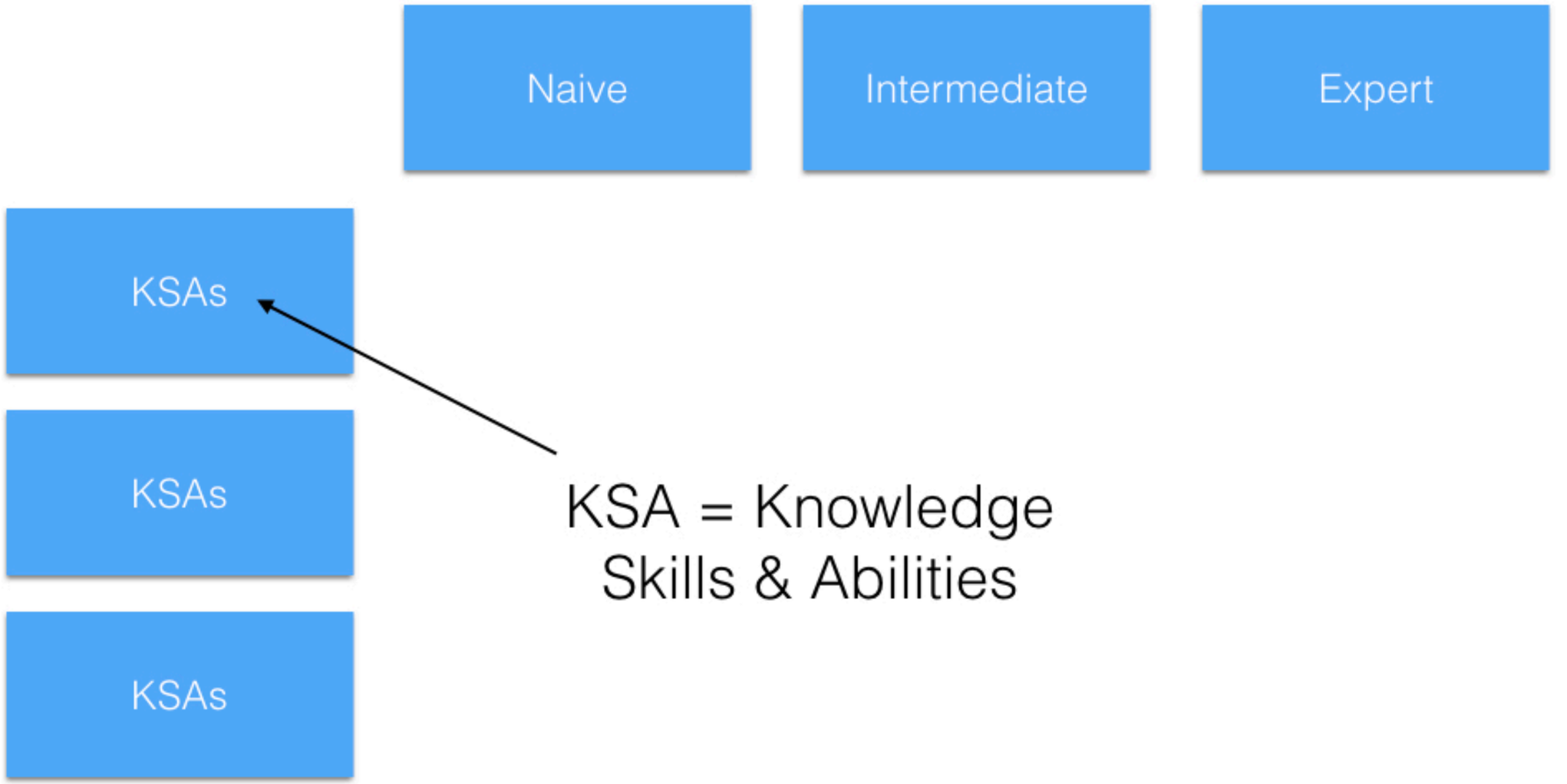
http://software-carpentry.org/membership/

Get involved: http://software-carpentry.org/join/

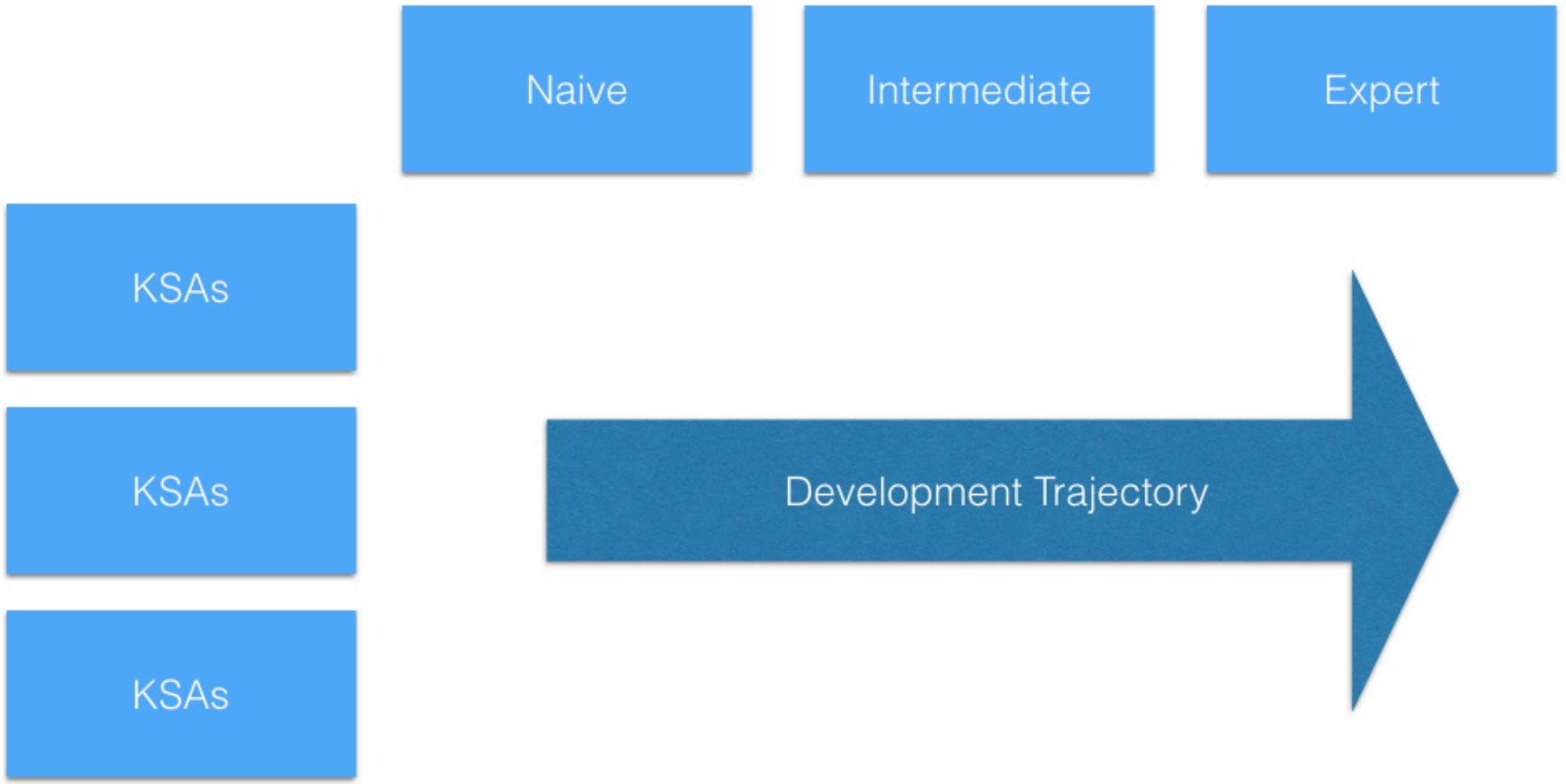
Twitter: @swcarpentry @jduckles

RESERVE SLIDES

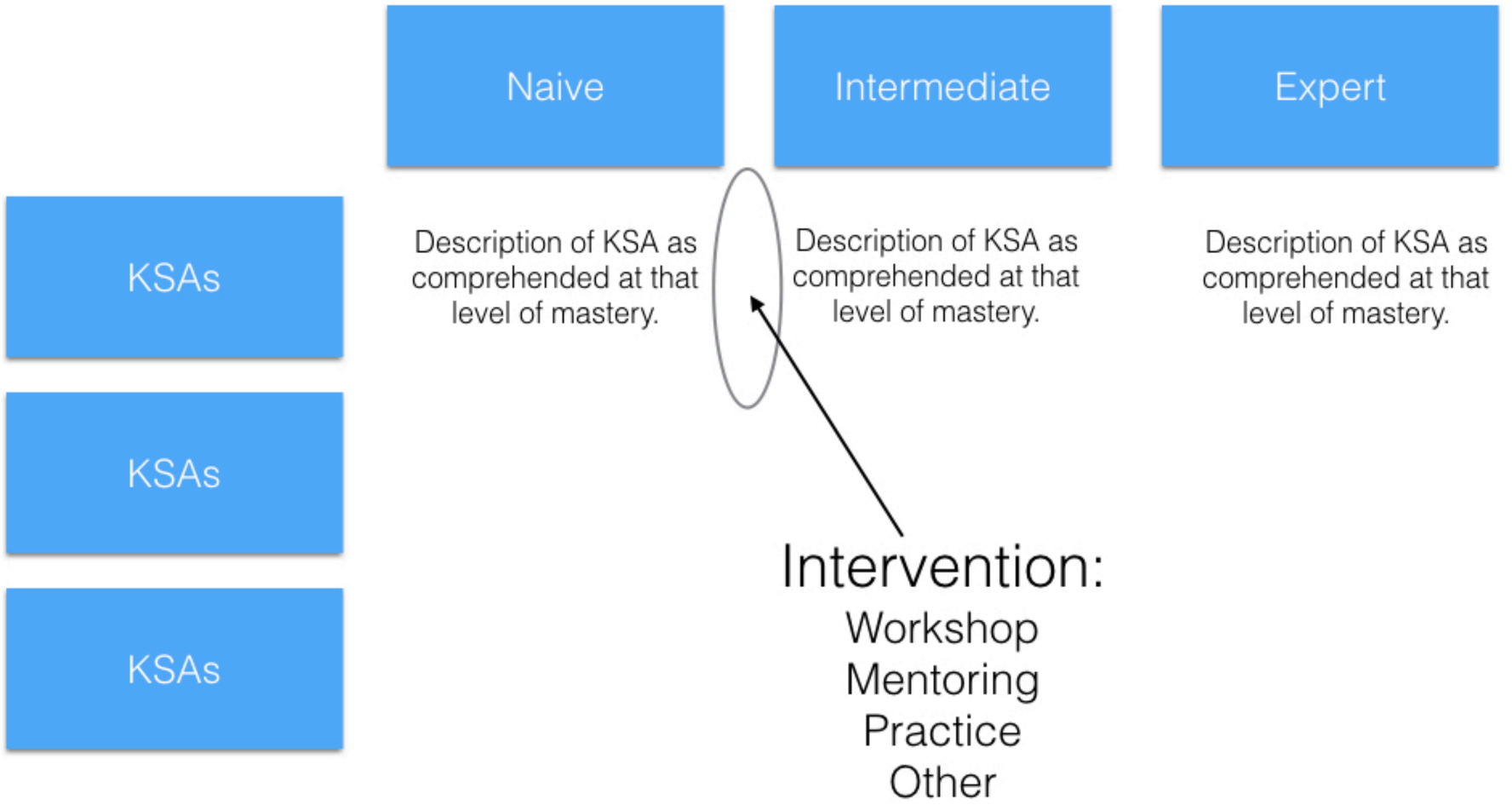
Mastery Rubric



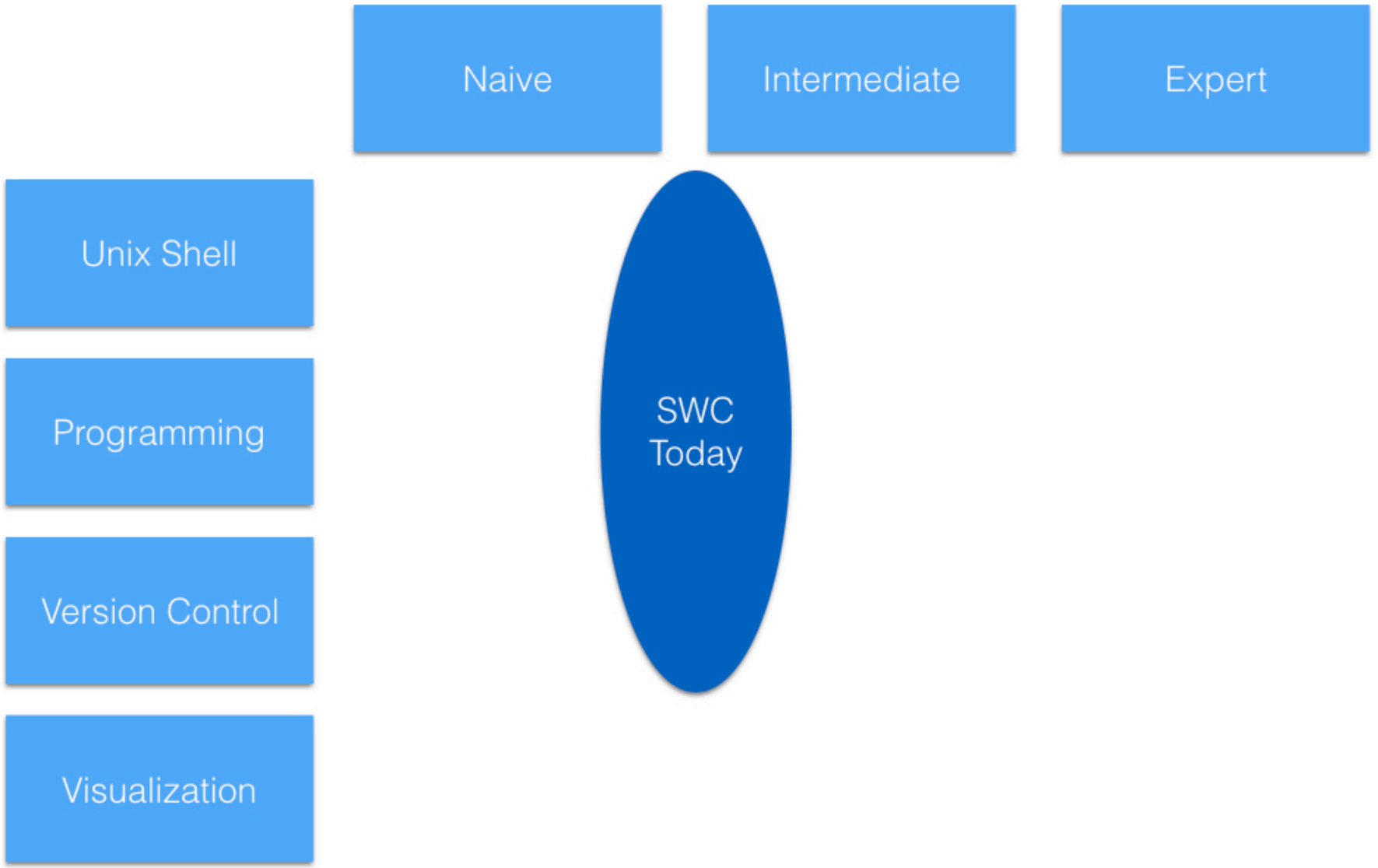
Mastery Rubric



Mastery Rubric



Mastery Rubric



education sciences

MDPI

Article

How the Mastery Rubric for Statistical Literacy Can Generate Actionable Evidence about Statistical and Quantitative Learning Outcomes

Rochelle E. Tractenberg

Collaborative for Research on Outcomes and Metrics; Departments of Neurology; Biostatistics, Bioinformatics & Biomathematics and Rehabilitation Medicine, Georgetown University Medical Center, Suite 207 Building D, 4000 Reservoir Road NW, Washington, DC 20057, USA; rochelle.tractenberg@gmail.com

Academic Editor: James Albright
Received: 7 July 2016; Accepted: 12 December 2016; Published: 24 December 2016

Abstract: Statistical literacy is essential to an informed citizenry; and two emerging trends highlight a growing need for training that achieves this literacy. The first trend is towards "big" data: while automated analyses can exploit massive amounts of data, the interpretation—and possibly more importantly, the replication—of results are challenging without adequate statistical literacy. The second trend is that science and scientific publishing are struggling with insufficient/inappropriate statistical reasoning in writing, reviewing, and editing. This paper describes a model for statistical literacy (SL) and its development that can support modern scientific practice. An established curriculum development and evaluation tool—the Mastery Rubric—is integrated with a new, developmental, model of statistical literacy that reflects the complexity of reasoning and habits of mind that scientists need to cultivate in order to recognize, choose, and interpret statistical methods. This developmental model provides actionable evidence, and explicit opportunities for consequential assessment that serves students, instructors, developers/reviewers/accreditors of a curriculum, and institutions. By supporting the enrichment, rather than increasing the amount, of statistical training in the basic and life sciences, this approach supports curriculum development, evaluation, and delivery to promote statistical literacy for students and a collective quantitative proficiency more broadly.

Keywords: statistical literacy; mastery rubric; collective quantitative proficiency; basic sciences; life sciences; scientific practice; curriculum development; curriculum evaluation; actionable evidence

1. Introduction

Statistical literacy (SL) is widely described as important for full social participation (see [1]; elementary curricula, e.g., [2,3]; higher education and beyond, e.g., [4–6]). Although this is true for all students, there is a special relationship between statistics and scientific research that amplifies the importance of developing appropriate statistical literacy in undergraduate or graduate/post-graduate students in the sciences.

Empirical research relies on statistical methods, and statistics is a wide, dynamic field perpetually propelled by new and improved methods. This far outstrips the capacities of other fields to fully adapt to these innovations, much less to incorporate all "relevant" methods in their own PhD curricula. Recently, Weissgerber et al. (2016) [7] correctly articulate that—and the myriad empirical arguments why—basic scientists *need* training in statistics (see also [8–16]; see also [17]). In fact, science PhD programs face a nearly Sisyphusian task: to adapt to some or any new methods, or even to prepare their students to adapt, so that their non-statistical discipline may exploit the power of new, or justify selecting established, statistical methods. Learning all statistical methods is clearly not feasible; even

Rochelle Tractenberg's paper about a Mastery Rubric for Statistical Literacy:

http://www.mdpi.com/2227-7102/7/1/3/pdf

How to get yourself involved?

http://software-carpentry.org/join/

— newsletter
— mailing list
— github
— open calls for instructor training