Software and Data Carpentry: Building Skills and Perspectives to Enable Data-Driven Research

Jonah Duckles and Tracy Teal

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

Available at http://jduck.net/presentations

Are we using the right tools?

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

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

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

Research organization leaders seek to build sustained capacity

Researchers need to build and expand on data skills

Technical consulting

doesn't scale

Online training and MOOCS

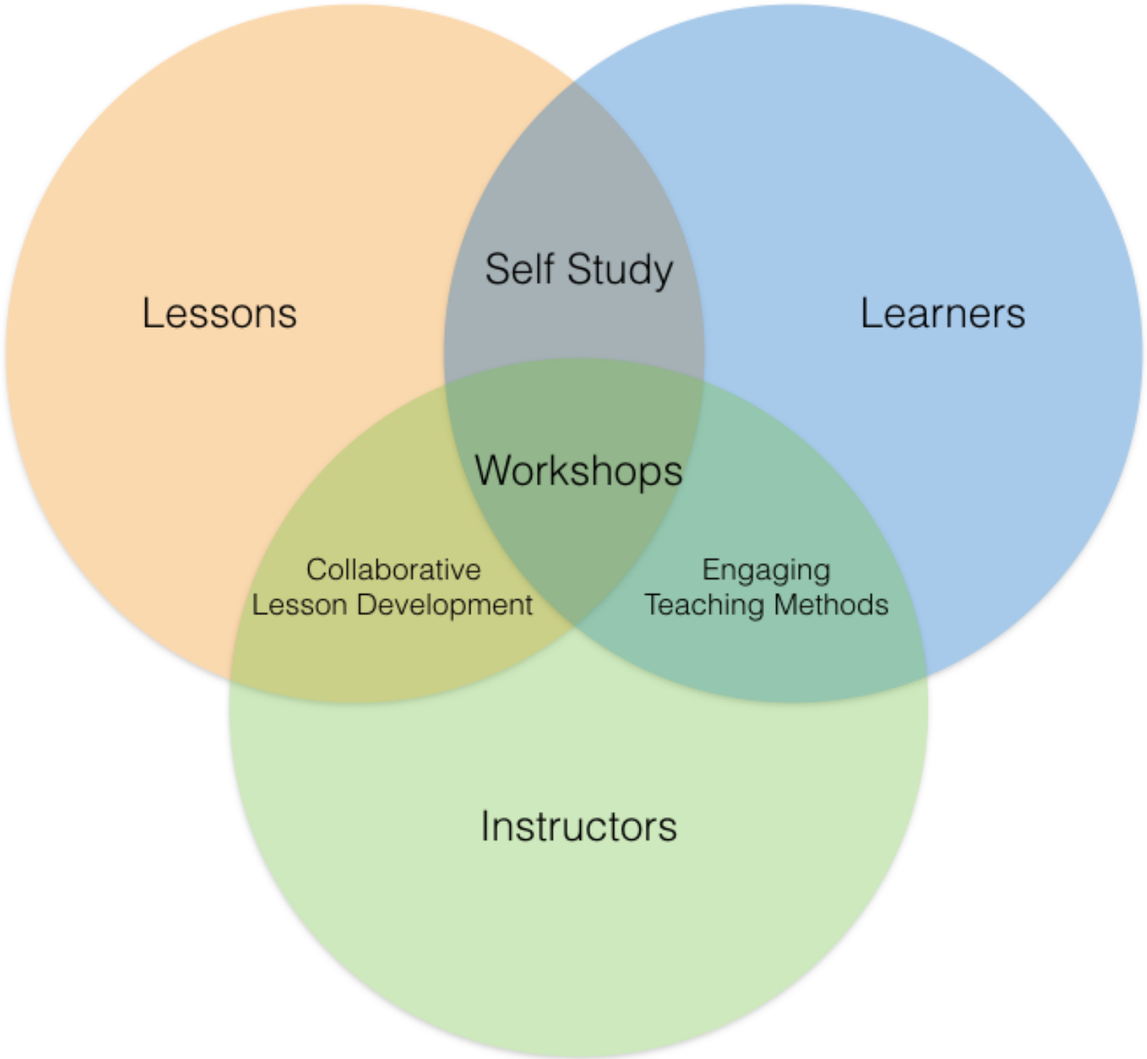
can leave people frustrated, without a community to help them

Our approach is to build community

and capacity through peer instruction

What goes into preparing a workshop with

IMPACT


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

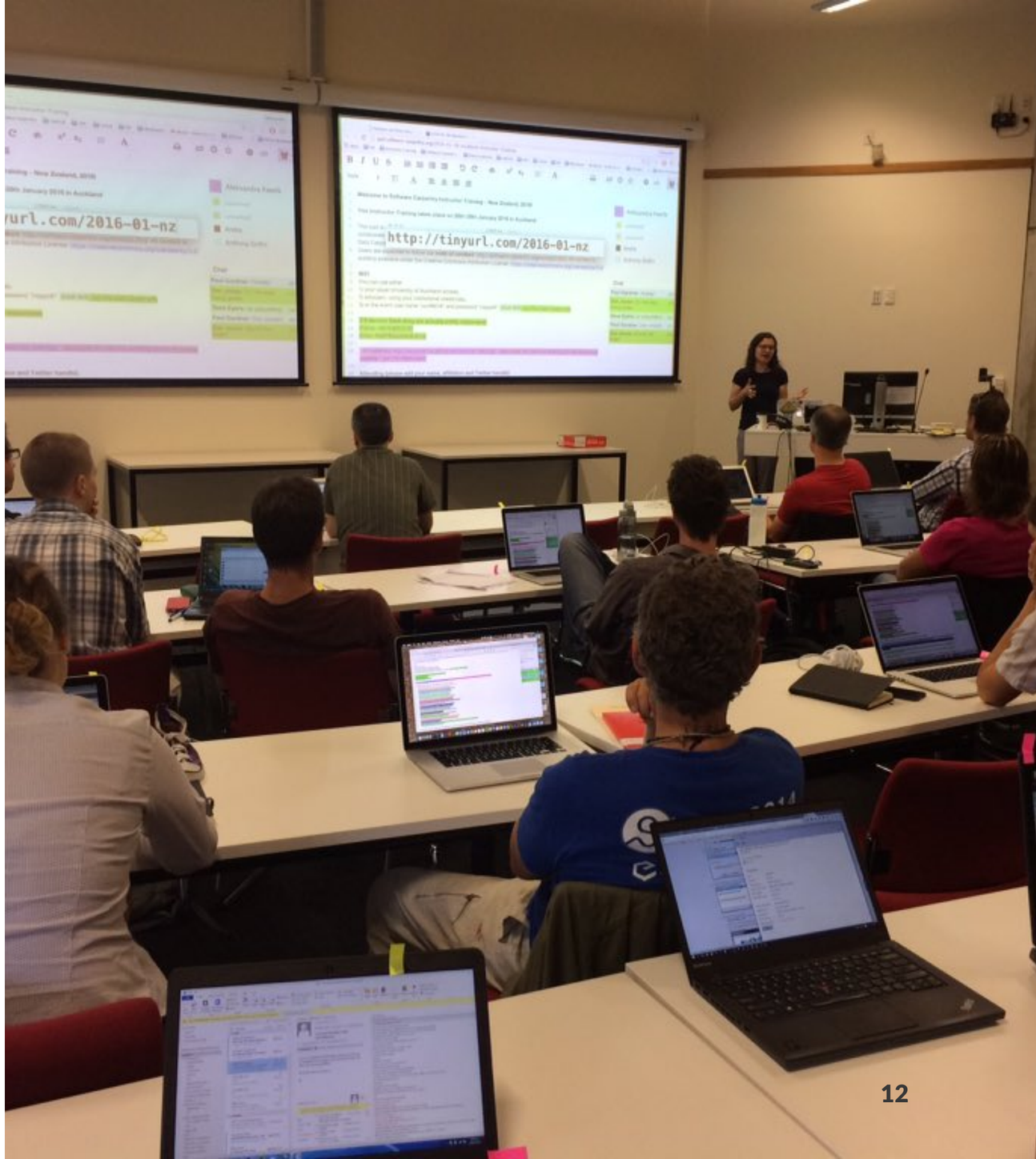
Impactful Workshops

- Trained volunteers
- Drawn from the research communities we serve
- Good 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

- We train researchers how to teach technical skills:
  - pedagogical approaches
  - take into account learner's background
  - reducing cognitive load
  - give and receive constructive feedback



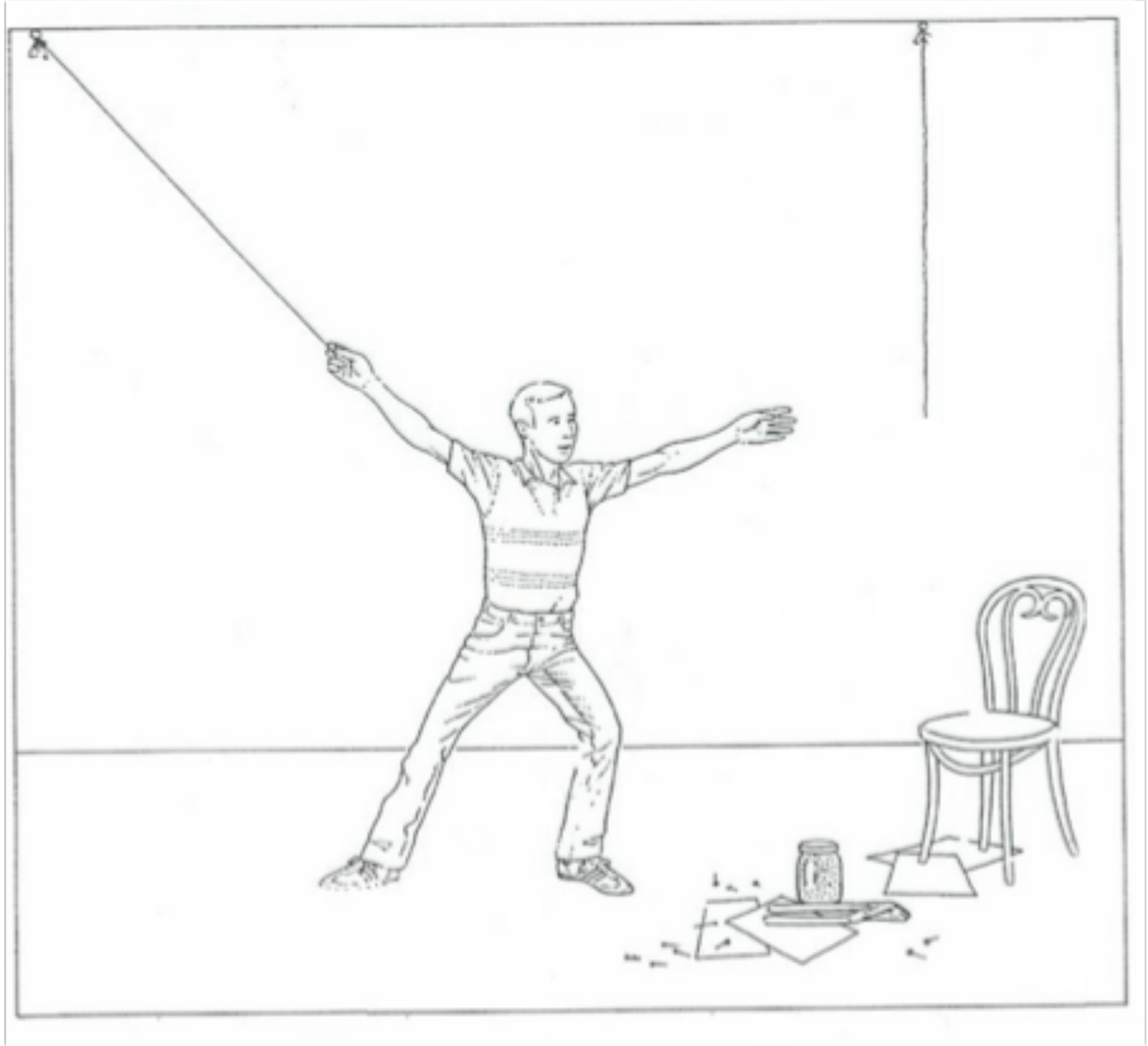
Teaching as performance art

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

Lessons

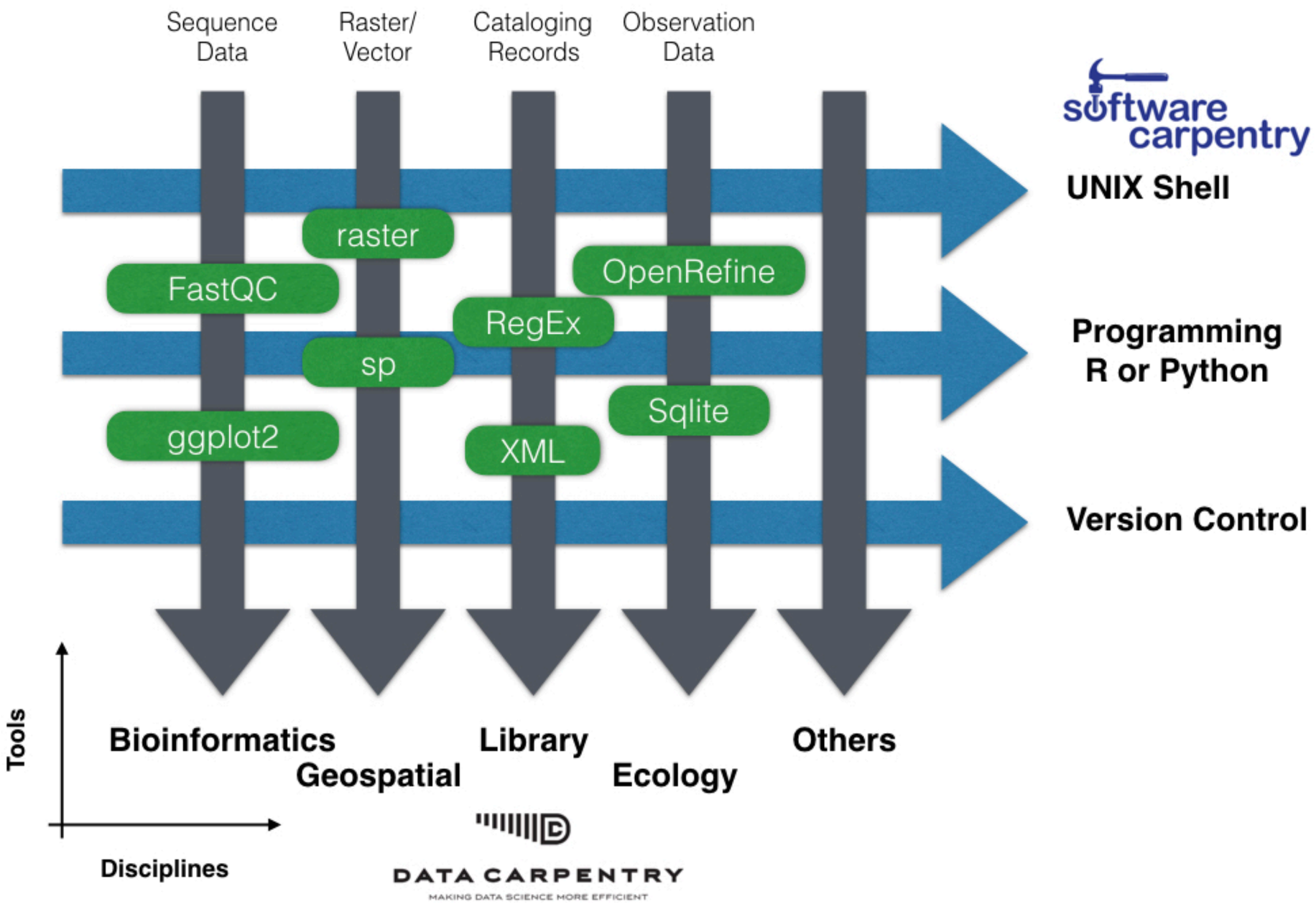
Preparing the unconscious mind

- Two day constraint
- N.R.F. Maier's two cords experiment
- Problem solving skills can be inaccessible to the conscious mind

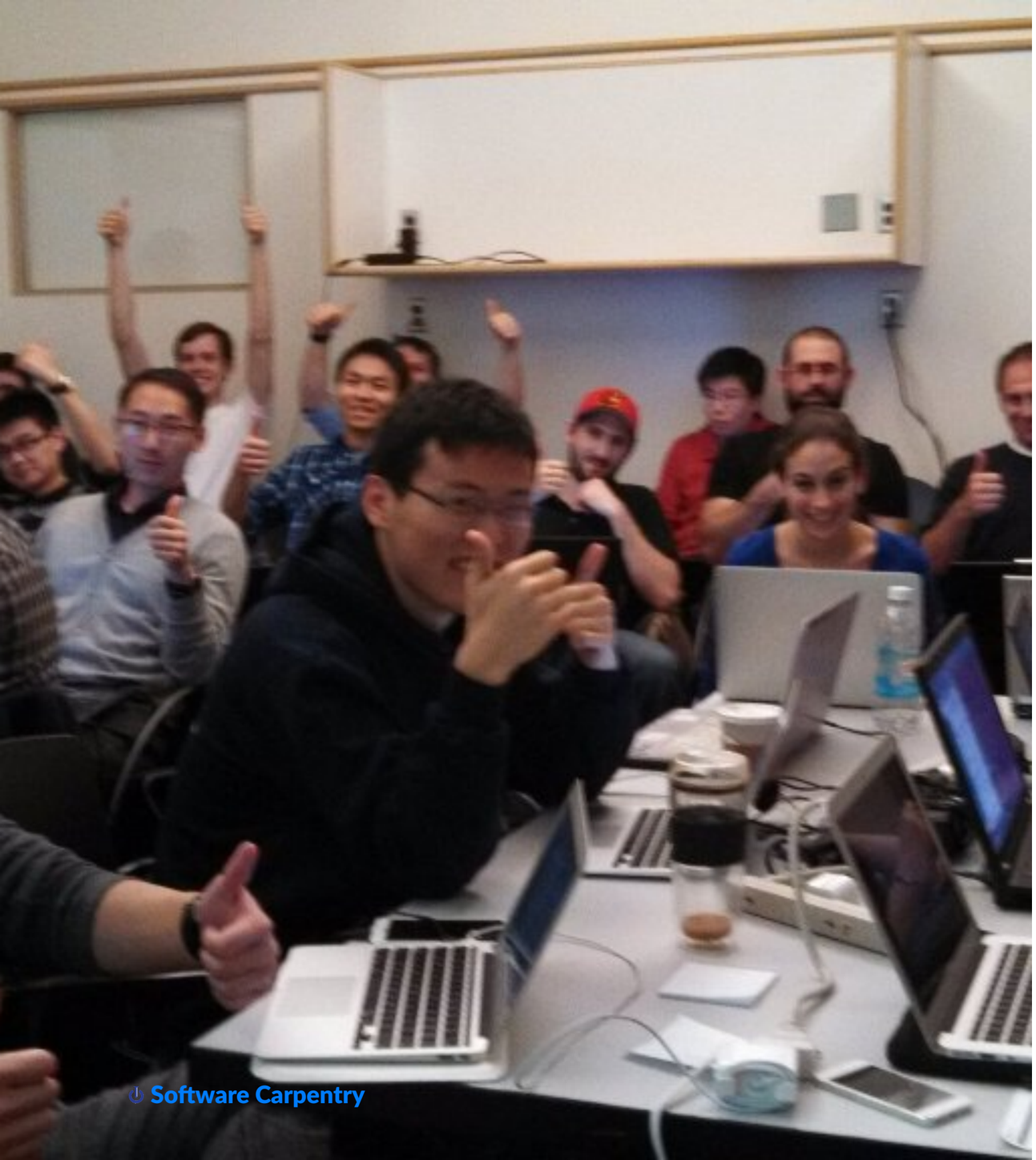


Jugyokenkyu - lesson study

- 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


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

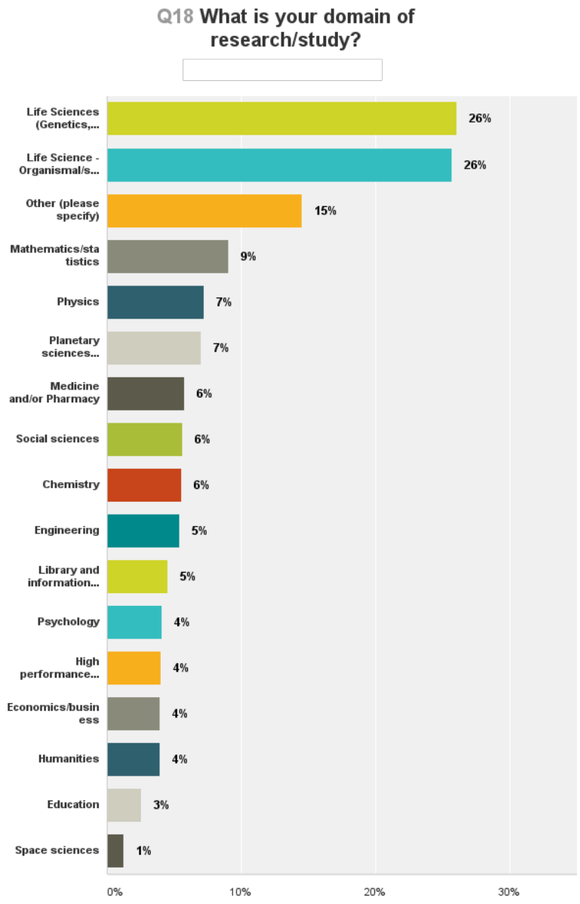


Community

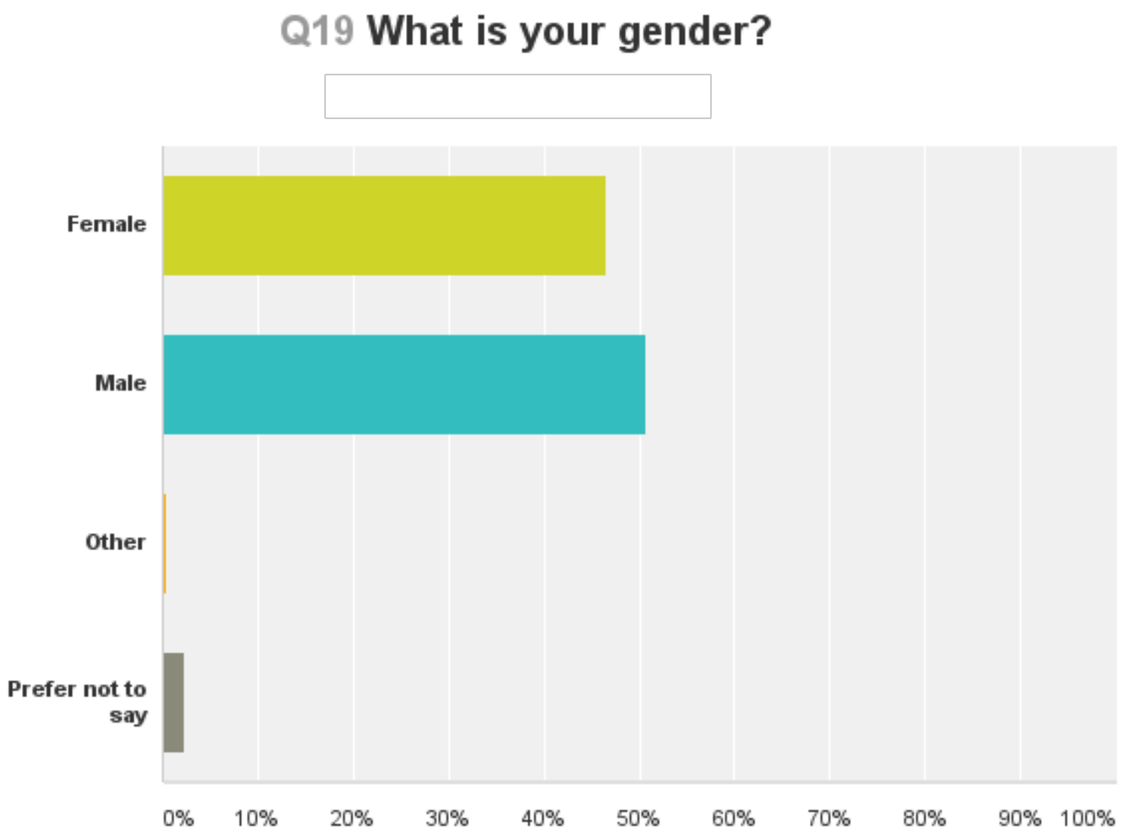
- Cultivating a community and culture requires ongoing and intentional work
  - Code of Conduct and enforcement
  - Active mentoring of instructors
  - Collaborative lesson development

Learners

Our workshops are welcoming and relevant to all disciplines



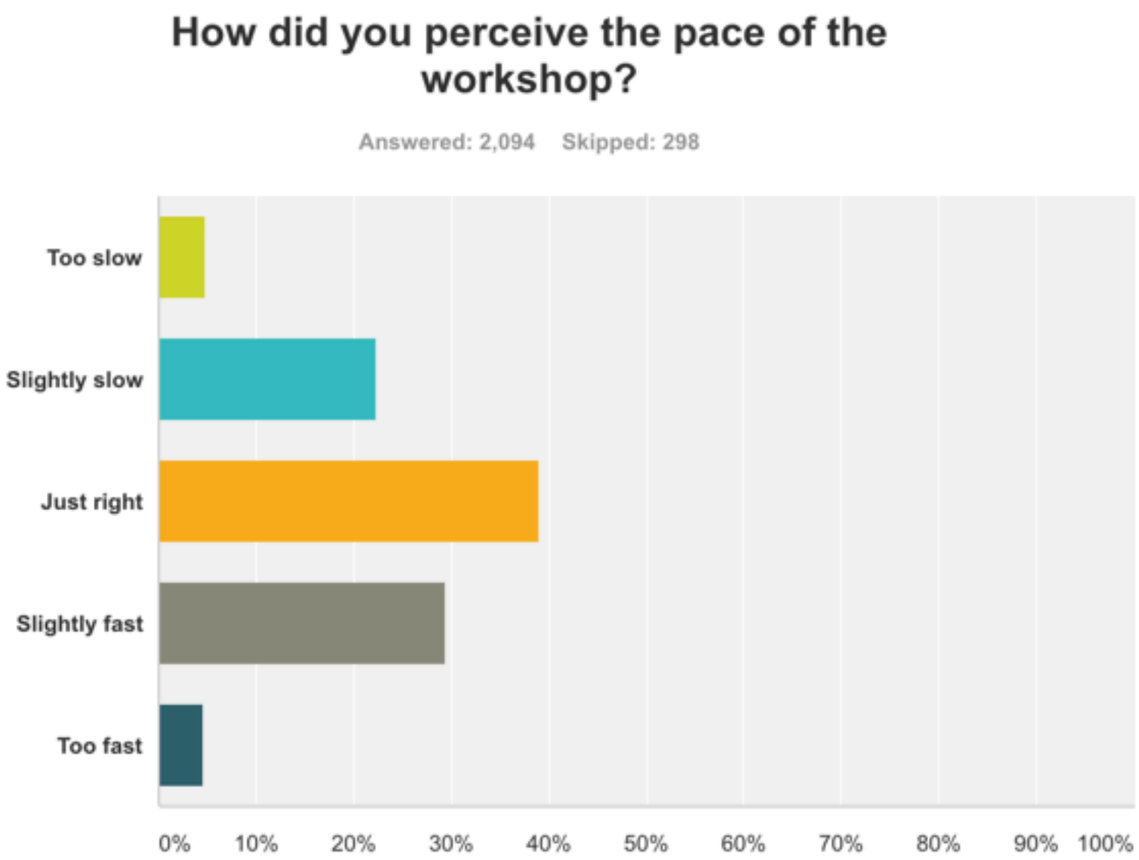
We achieve a very good gender balance



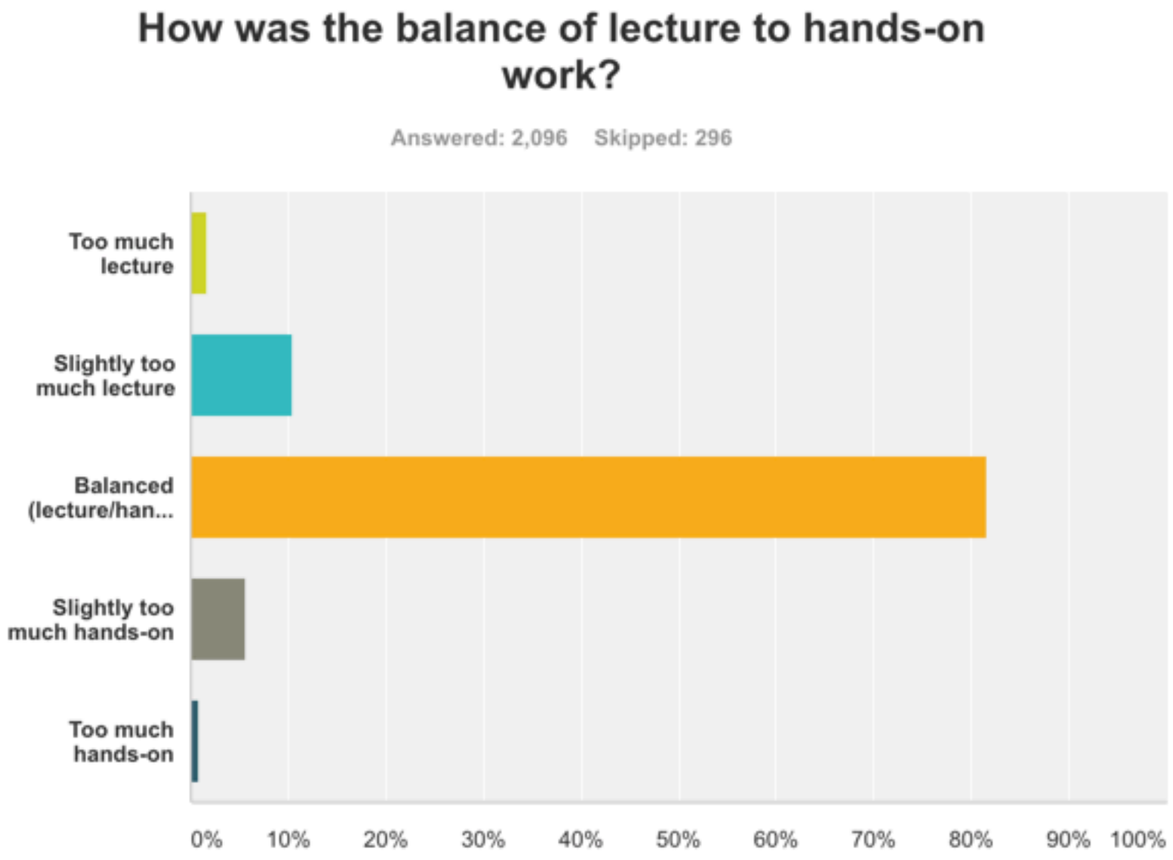
Results

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	Total	Weighted Average
The amount of information covered at the workshop was reasonable for allotted time	1.72% 36	7.74% 162	8.50% 178	58.07% 1,216	23.97% 502	2,094	3.95
The overall atmosphere of the workshop was welcoming	0.96% 20	0.72% 15	1.19% 25	29.91% 626	67.22% 1,407	2,093	4.62
I learned skills that I will be able to use in my research/work	1.34% 28	1.63% 34	9.18% 192	47.94% 1,003	39.91% 835	2,092	4.23
The material presented matched the workshop description	1.10% 23	2.67% 56	6.97% 146	45.03% 943	44.22% 926	2,094	4.29
I would recommend this workshop to a friend/colleague	1.44% 30	2.20% 46	6.94% 145	40.02% 836	49.40% 1,032	2,089	4.34
The workshop was worth my time	1.63% 34	2.83% 59	9.05% 189	43.39% 906	43.10% 900	2,088	4.24

Adapting to learners



Balancing demonstration and application



Learners perception of Instructors

	Never	Rarely	Sometimes	Often	All of the time	Total	Weighted Average
Instructors gave clear answers to your questions	0.24% 5	0.34% 7	4.55% 94	43.28% 895	51.60% 1,067	2,068	4.46
Instructors were considerate	0.19% 4	0.15% 3	2.42% 50	21.77% 450	75.47% 1,560	2,067	4.72
Instructors were good communicators	0.19% 4	0.68% 14	9.59% 198	36.56% 755	52.98% 1,094	2,065	4.41
Instructors were enthusiastic	0.19% 4	0.29% 6	3.87% 80	25.91% 535	69.73% 1,440	2,065	4.65

Motivating Future Learning

	Less motivated	Slightly less motivated	No change in motivation	More motivated	Much more motivated	N/A - Topic not covered	Total	Weighted Average
Unix shell	1.04% 20	1.35% 26	20.89% 402	45.11% 868	28.69% 552	2.91% 56	1,924	4.02
Git (version control)	1.81% 35	2.38% 46	15.63% 302	43.74% 845	34.01% 657	2.43% 47	1,932	4.08
Python	0.77% 14	1.37% 25	13.29% 243	28.72% 525	31.89% 583	23.96% 438	1,828	4.18
R	0.35% 6	0.58% 10	12.86% 223	17.19% 298	24.34% 422	44.69% 775	1,734	4.17
SQL	0.61% 10	0.79% 13	16.25% 266	10.45% 171	7.64% 125	64.26% 1,052	1,637	3.66

How to get yourself involved?

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

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

How to get your organization involved?

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

Partnership Tiers

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
Joint Fee (annual)	$5,000	$7,500	$15,000	Contact us
SC or DC Only (annual)	$4,000	$6,500	$12,500	Contact us

Thank You to our supporters!

Thank You!

Jonah Duckles, Executive Director

jduckles@software-carpentry.org

Become a partner: http://software-carpentry.org/scf/join/

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

Twitter: @swcarpentry @jduckles