Skills for Open Science: the impact and rewards of training with The Carpentries

Open Science Days
Max Planck Digital Library
24 October 2021

https://edcarp.github.io

edcarp@ed.ac.uk

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk

Edinburgh Carpentries

Edward Wallace https://ewallace.github.io
Biological Sciences, Edinburgh

Edinburgh Carpentries - https://edcarp.github.io
Thanks to Giacomo Peru, edcarp co-ordinator, and the whole community

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk

Edinburgh Carpentries

- Doing open science requires foundational skills
- The Carpentries approach develops these skills in a community
- It is impactful and rewarding to teach those skills
- A perspective from a biologist in Edinburgh…

https://edcarp.github.io

edcarp@ed.ac.uk

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk

How did I get to run a lab? Mostly by writing code

Edinburgh Carpentries

- Mathematics PhD, University of Chicago 2005-10
  - *stochsimcode,* MATLAB for stochastic simulations of neural networks, *PLoS Computational Biology*
- Systems Biology postdoc, Harvard 2010-13
  - *codonFits*, bad R package for evolution of protein-coding sequences, *Molecular Biology and Evolution*
- Biochemistry postdoc, U. Chicago 2013-15
  - R code for analysing/visualising protein aggregation, *Cell* & *Dryad*
- Informatics / Cell Biology fellow, Edinburgh 2016-17
  - R code for analysing RNA splicing data, *RNA*

https://edcarp.github.io **edcarp@ed.ac.uk**

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk

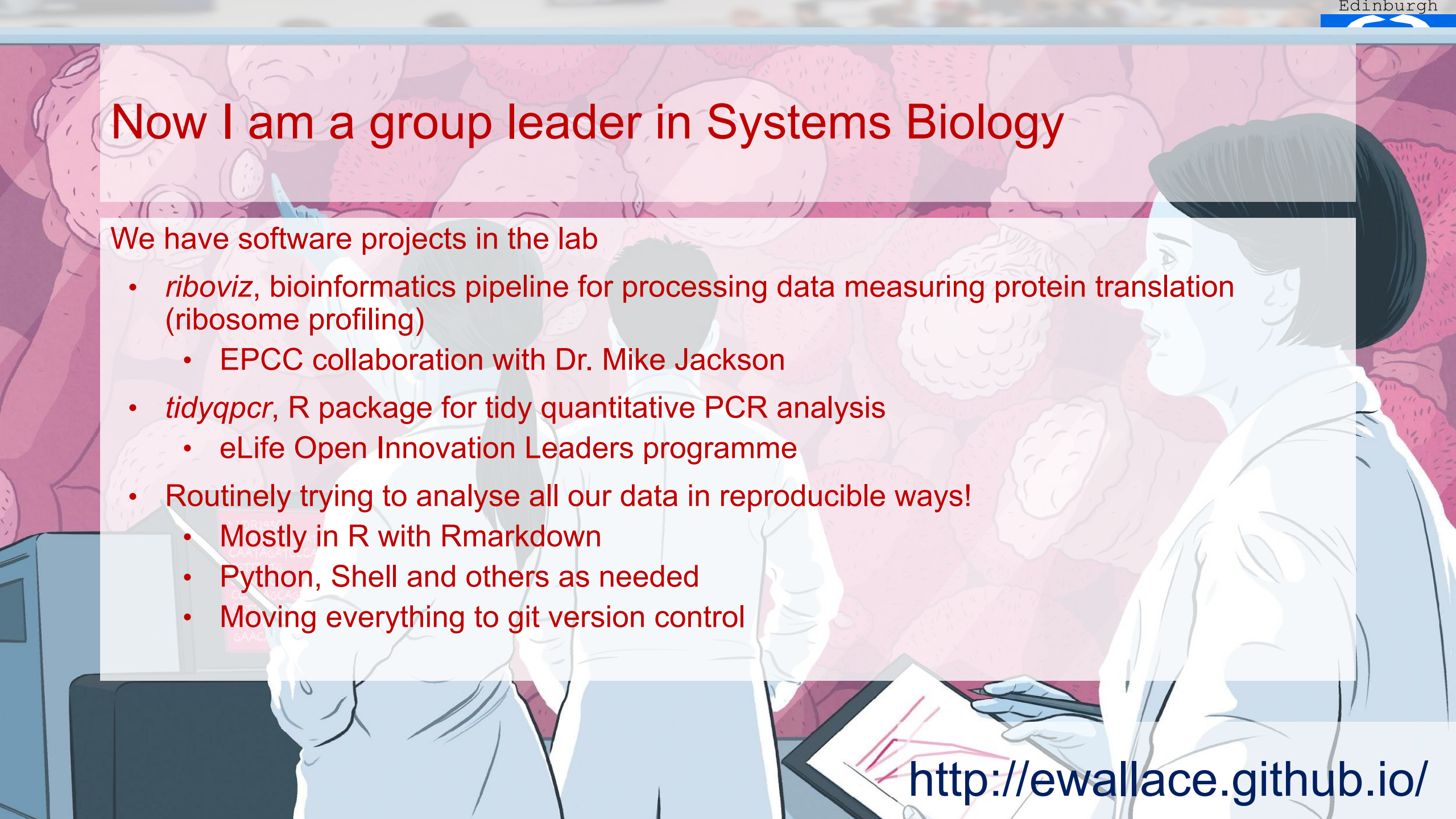
Now I am a group leader in Systems Biology

http://ewallace.github.io/

Now I am a group leader in Systems Biology

We have software projects in the lab

- *riboviz*, bioinformatics pipeline for processing data measuring protein translation (ribosome profiling)
  - EPCC collaboration with Dr. Mike Jackson
- *tidyqpcr*, R package for tidy quantitative PCR analysis
  - eLife Open Innovation Leaders programme
- Routinely trying to analyse all our data in reproducible ways!
  - Mostly in R with Rmarkdown
  - Python, Shell and others as needed
  - Moving everything to git version control



http://ewallace.github.io/

January 3, 2021

Software Open Access

ewallace/pseudonuclease_evolution_2020: Repeated evolution of inactive pseudonucleases in a fungal branch of the Dis3/RNase II family of nucleases

Edward Wallace

This repository forms the supplementary data and analysis for the manuscript:

Repeated evolution of inactive pseudonucleases in a fungal branch of the Dis3/RNase II family of nucleases; E.R. Ballou, A.G. Cook, E.W.J. Wallace. Molecular Biology and Evolution, 2020. DOI: 10.1093/molbev/msaa324

For this release, we clarified some of the documentation. The underlying data remain unchanged. This is the final expected release.

Preview

pseudonuclease_evolution_2020-v1.2.zip

ewallace-pseudonuclease_evolution_2020-0831b11

.gitignore	50 Bytes
LICENSE	11.4 kB
README.md	1.6 kB

My problems as a new lab head

- I have less time to code than I used to
  - I go to meetings, run my lab, write papers & grants, teach
- Everyone in my research group needs to code
  - Even wet-lab biologists need to wrangle and plot their data
  - And to share their data when publishing

How can I promote good practices in research software, when I am writing less code myself?

THE UNIVERSITY of EDINBURGH
School of Biological Sciences
Edward.Wallace@ed.ac.uk

The problem is bigger than me

- Actually, all 21st-century researchers need to code
  - Reproducibly, reliably, efficiently
  - At all career stages.
  - How are they going to learn?

THE UNIVERSITY of EDINBURGH
School of Biological Sciences
Edward.Wallace@ed.ac.uk

The underlying problem: skills for open science

Edinburgh Carpentries

- Working open needs confidence and capability
  - “It’s a nice idea but I don’t do that”
- Reproducible research uses a large software stack
- **Open science rests on foundational skills**
  - **coding**
  - **data science**
  - **project organisation**

https://edcarp.github.io **edcarp@ed.ac.uk**

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk

What do biologists need? We asked them

Edinburgh Carpentries

2019 School of Biological Sciences research computing survey:

- Inform research computing training for students, staff, faculty
- Find out what data & software people use
- Find out what skills & training they think they need
- Input to UKRI/BBSRC data-intensive bioscience review

- We used https://www.onlinesurveys.ac.uk/
- Designed 1-page survey completable in 5 minutes, April 2019
- We can share the survey design for **you** to adapt

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

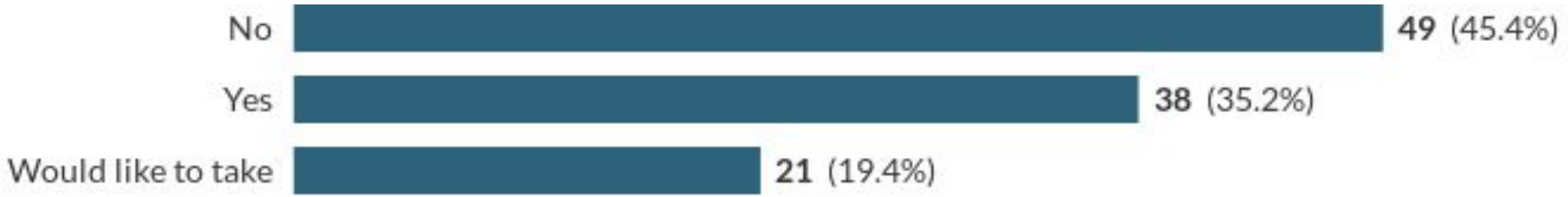
Edward.Wallace@ed.ac.uk

www.software.ac.uk

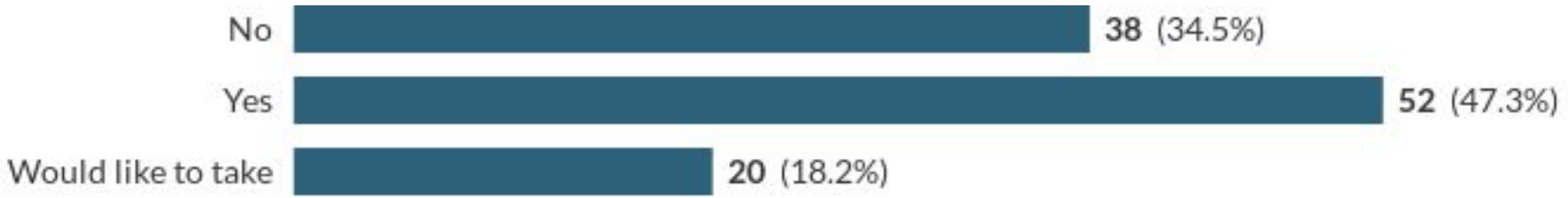
Many of us do not have formal training

Edinburgh Carpentries

In computing:



Or statistics:

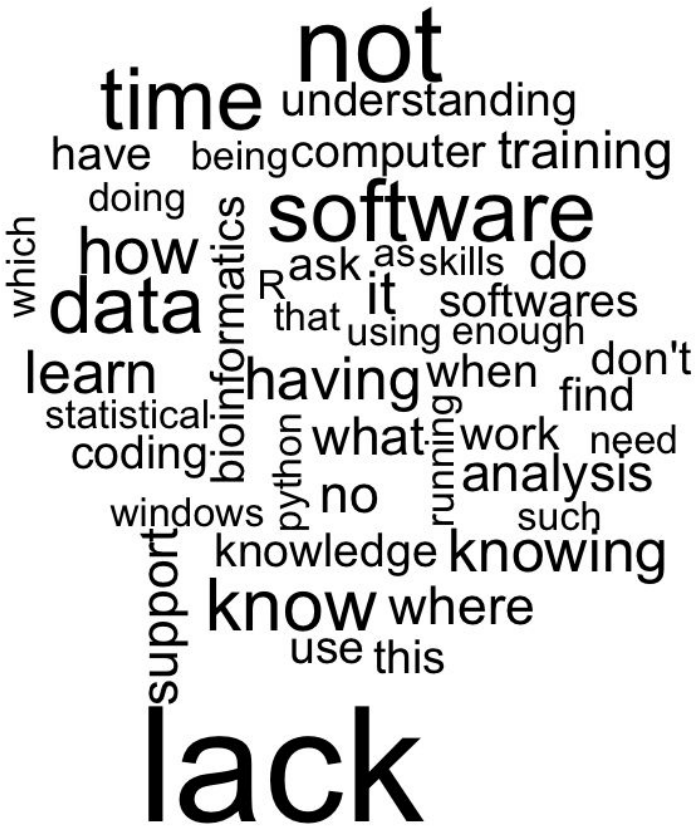


THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk

Your biggest frustration in research computing?



https://edcarp.github.io

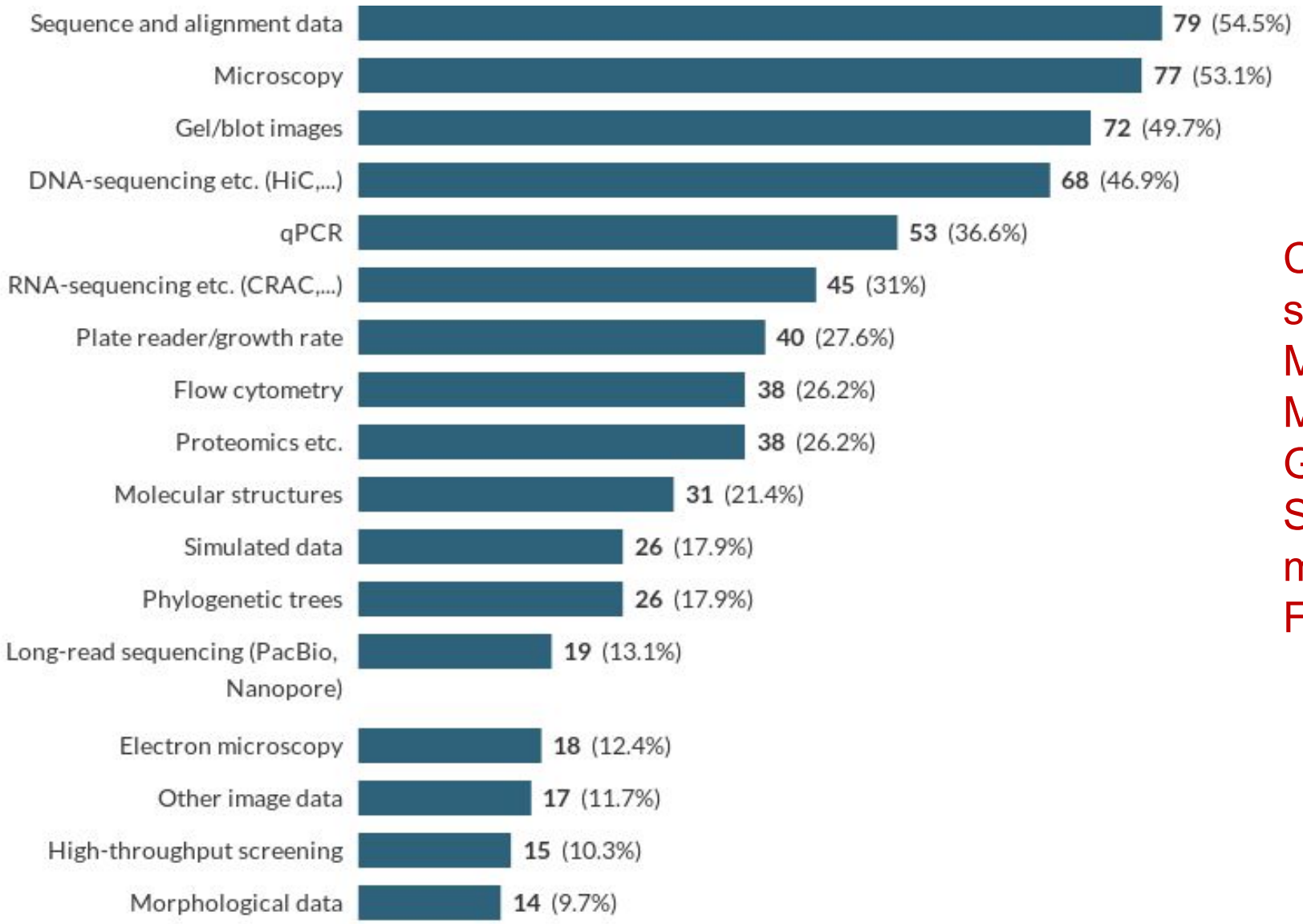
edcarp@ed.ac.uk

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk

We use many kinds of biological data



Correspondingly diverse software:
MS Excel, SPSS, R, python, MATLAB, ImageJ, ImageStudio, Genome Browsers, Benchling, Snapgene, Pymol, BLAST, multiple sequence alignment, FlowJo, ...

edcarp@ed.ac.uk

Your biggest need in computing training?

Researchers need foundational skills

https://edcarp.github.io

edcarp@ed.ac.uk

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk

THE CARPENTRIES

We teach foundational coding and data science skills to researchers worldwide.

DATA CARPENTRY

library carpentry

software carpentry

https://edcarp.github.io

edcarp@ed.ac.uk

What The Carpentries teaches

Edinburgh Carpentries

software carpentry

library carpentry

DATA CARPENTRY

Focus on software development

unix shell, git, ...

Focus on digital librarianship

Focus on data workflows

spreadsheets, R, ...

https://github.com/carpentrieslab

https://github.com/carpentries-incubator

https://edcarp.github.io

edcarp@ed.ac.uk

OPENNESS, ETHOS, PEDAGOGY

Edinburgh Carpentries

OPEN COMMUNITY

resources are developed collaboratively on GitHub by volunteers and available under CC-BY license

STRONG ETHOS

volunteering, inclusivity and respect, Code of Conduct are fundamental aspects of the community

PEDAGOGICAL DRIVE

structured pathway for the development of members, from learner, to helper, to instructor to trainer

https://edcarp.github.io

edcarp@ed.ac.uk

Edinburgh Carpentries - EdCarp

Coordinated approach to teach Carpentries in Edinburgh

Community effort across departments

Provides support with planning and delivery of workshops

Financial support from SoPA, IS and EPCC

Beyond UoE to Heriot-Watt and further

Balance between demand and capacity

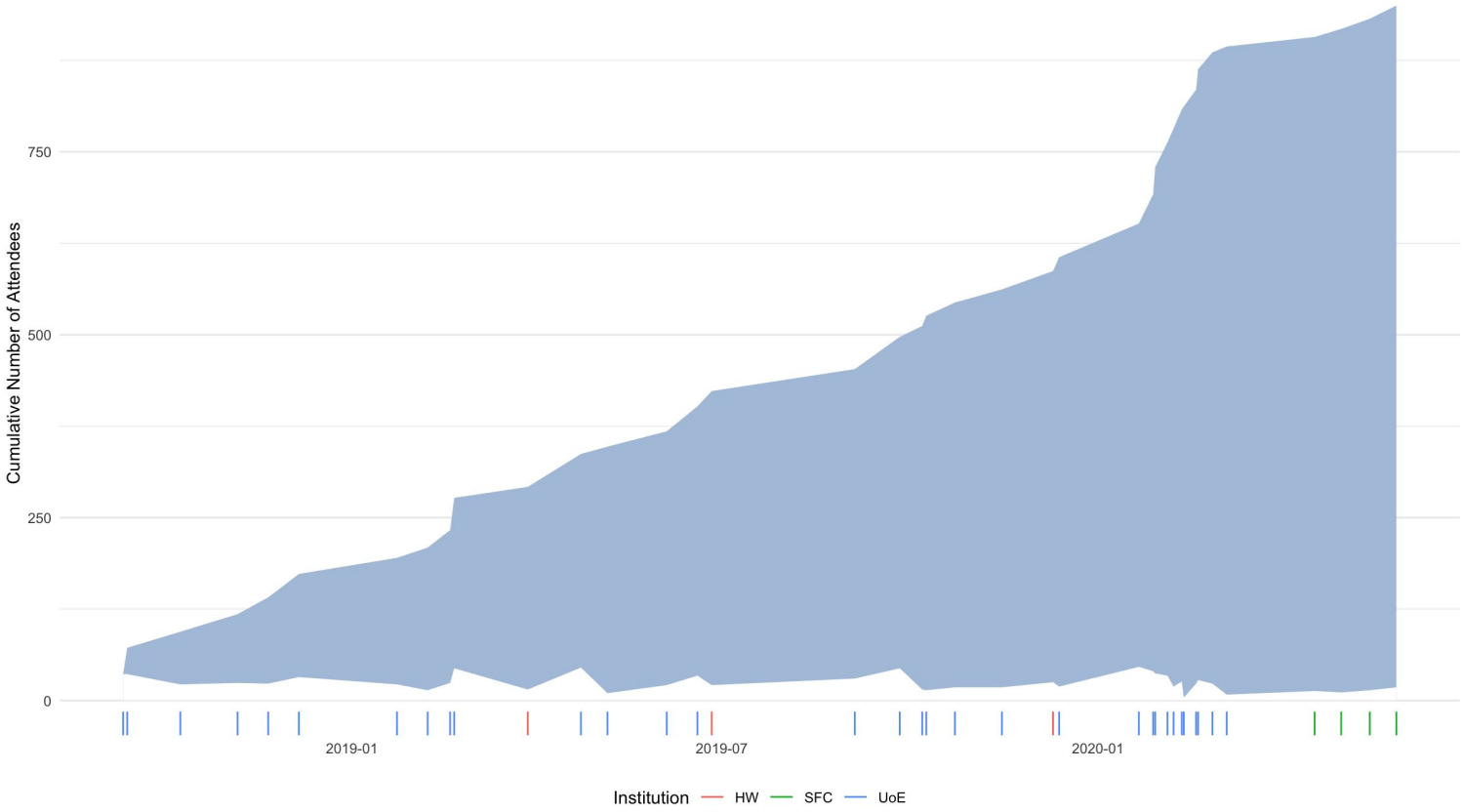
https://edcarp.github.io

edcarp@ed.ac.uk

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

Edinburgh Carpentries is Growing



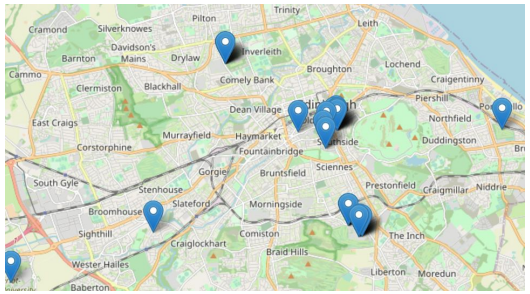
Sep 2018 – May 2020

37 events

45 days of training

900 learners

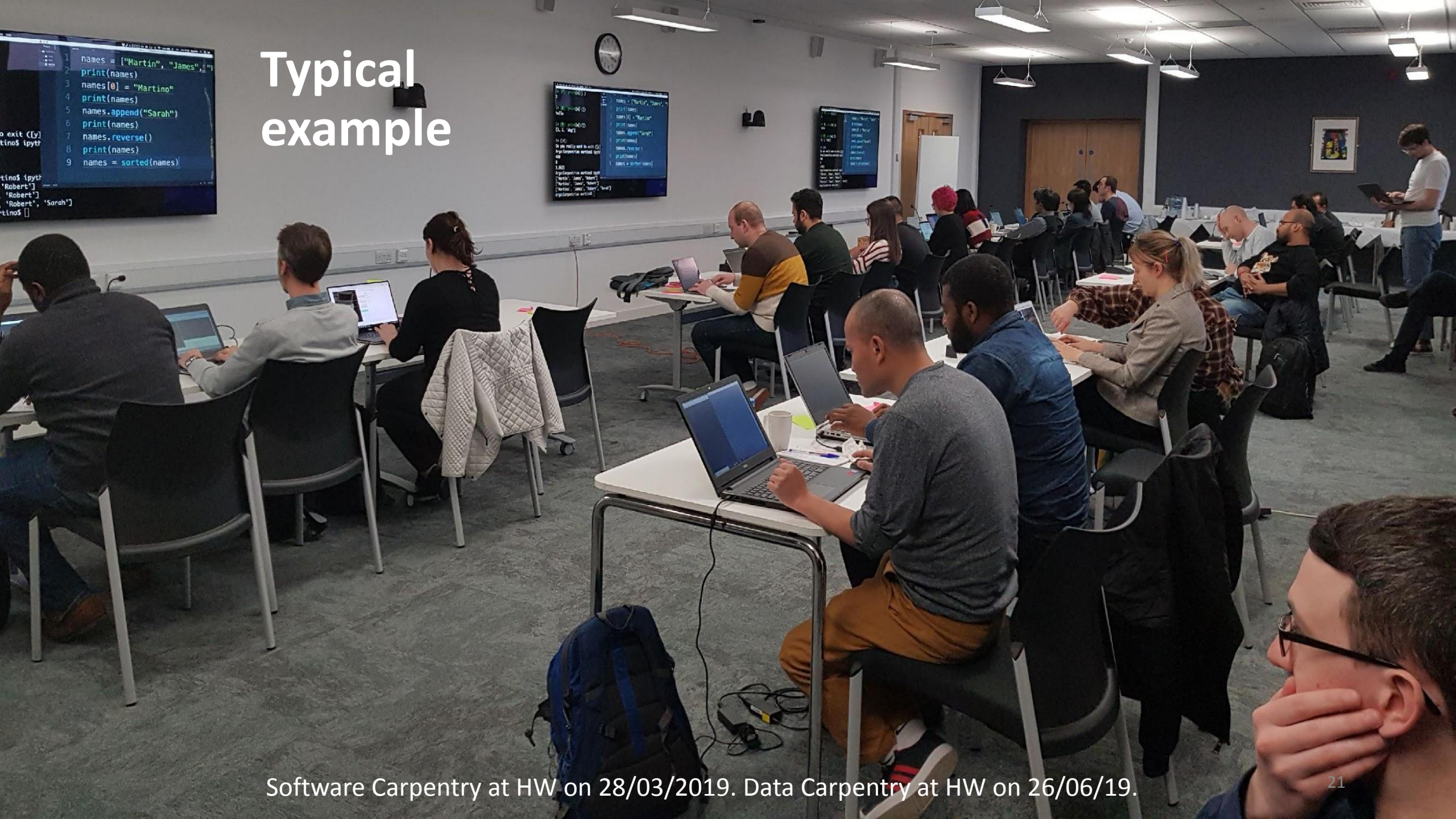
dozens of helpers



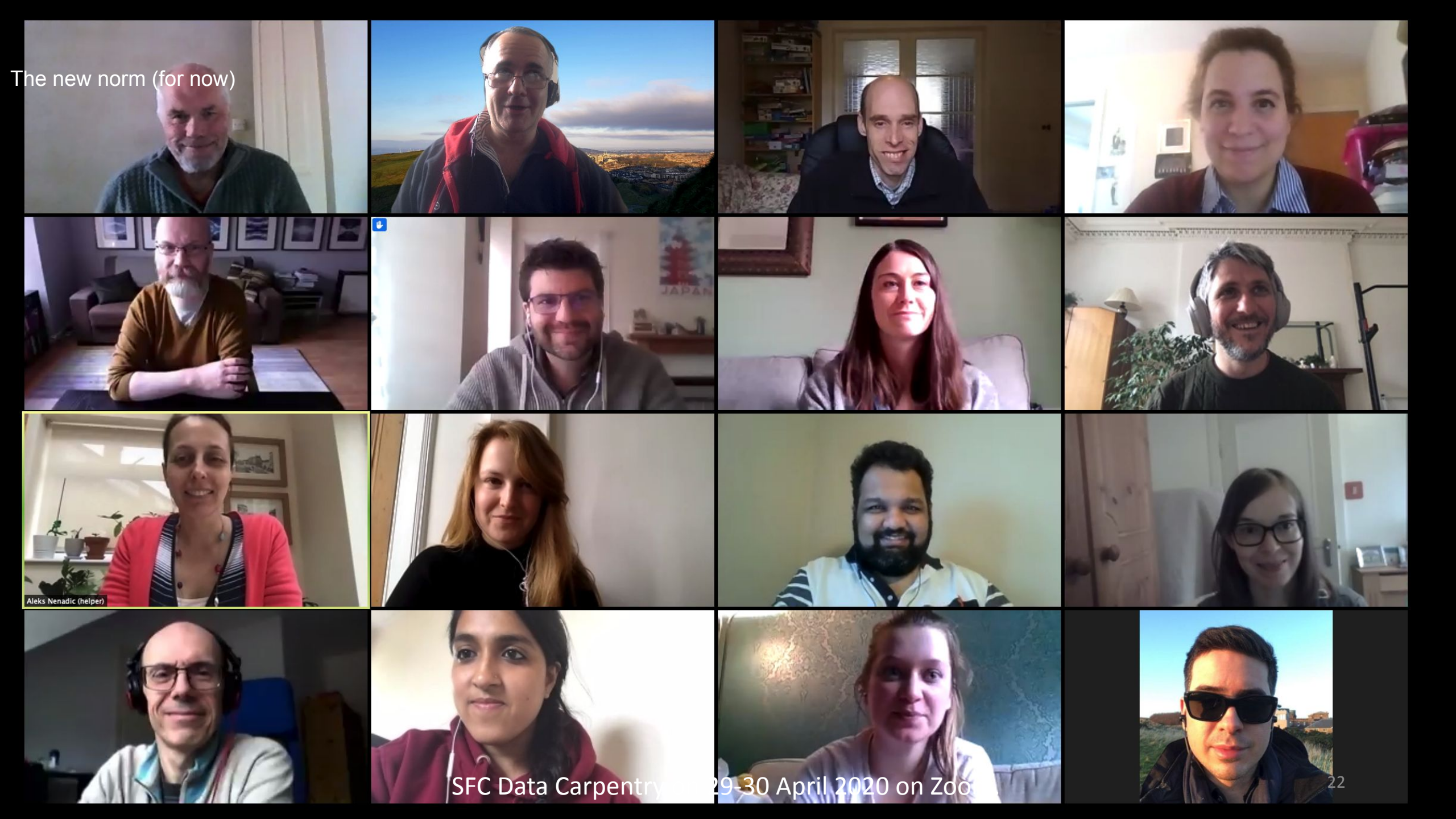
https://edcarp.github.io

edcarp@ed.ac.uk

Typical example



Software Carpentry at HW on 28/03/2019. Data Carpentry at HW on 26/06/19.


The new norm (for now)
Aleks Nenadic (helper)

2020 national review of data-intensive bioscience

Edinburgh Carpentries

Bioscience has emerged as a data-rich discipline, in a transformation that is spreading as widely now as molecular biology in the twentieth century.

In this report, the UKRI-BBSRC research community marks the moment and recognises some of the changes required.

We look forward to supporting new research careers, where data are valued and shared widely, where new software is a natural part of Biology, and where re-analysis and modelling are as creative as experimentation in understanding the rules of life and their applications.

- Andrew Millar, chair

https://edcarp.github.io

edcarp@ed.ac.uk

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk

2020 BBSRC review of data-intensive bioscience

Edinburgh Carpentries

Recommendation 1: UKRI-BBSRC should take specific actions to increase the UK capacity in mathematical and computational skills within the biosciences.

https://www.ukri.org/news/bbsrc-publishes-review-of-data-intensive-bioscience/

UKRI Biotechnology and Biological Sciences Research Council

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk

Ed-DaSH: UKRI funded grant to expand training

Edinburgh Carpentries

- Develop peer-reviewed open training modules for biological data science
- Deliver 98 days of remote training of our new workshops
- Offer a clinic after every workshop for learners to ask advice regarding their own projects.
- Train 30 new instructors to deliver these workshops, building a scalable training community.

School of Biological Sciences
School of Mathematics
College of Medicine and Veterinary Medicine

MRC Human Genetics Unit

EIDF Edinburgh International Data Facility

https://edcarp.github.io **edcarp@ed.ac.uk**

New open teaching materials for open science

Edinburgh Carpentries

Statistics

- Basic and Intermediate Statistical Skills (2 days)
- High Dimensional Statistics (2 days)
- Introduction to Machine Learning (2 days)

FAIR principles and data management

- Hands On Open Science, FAIR Principles and Data Management (2 days)
- Reshaping Research: How Adopting FAIR Increases Productivity (1 day)

Data science computing with workflows

- Good Enough Practice in Scientific Computing (0.5 days)
- Introduction to Conda for (Data) Scientists (0.5 days)
- Workflow Management with (one of) Snakemake or Nextflow (2 days)

https://edcarp.github.io **edcarp@ed.ac.uk**

Edinburgh Carpentries

- Doing open science requires foundational skills
- The Carpentries approach develops these skills in a community
- It is impactful and rewarding to teach those skills
- We have a new grant to expand skills training and grow the community in Edinburgh

https://edcarp.github.io

edcarp@ed.ac.uk

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk

What is in it for me?

- Teaching Carpentries improves my own skills and work
- Instructor Training is good evidence-based pedagogy
- My involvement helps to improve the skills of people in my lab
- Training helps to get grants funded
  - **this is impact, write it in the impact section**
- The Carpentries is a nice, inclusive, activist community
- Teaching colleagues skills that they need is satisfying

THE UNIVERSITY of EDINBURGH
School of Biological Sciences
Edward.Wallace@ed.ac.uk

THANKS

Edinburgh Carpentries

Everyone is welcome to get involved!
EdCarp sign-up: http://eepurl.com/gI4MsX

Thank you from the Edinburgh Carpentries Team

Giacomo Peru, Neil Chue Hong, Graeme Grimes, Mario Antonioletti and all Organising Committee, Steering Committee, Instructors, Helpers, Organisers

UKRI grant: Alison Meynert, Catalina Vallejos, Alex Twyford, Andrew Millar

https://edcarp.github.io **edcarp@ed.ac.uk**

THE UNIVERSITY of EDINBURGH
School of Biological Sciences

Edward.Wallace@ed.ac.uk

www.software.ac.uk