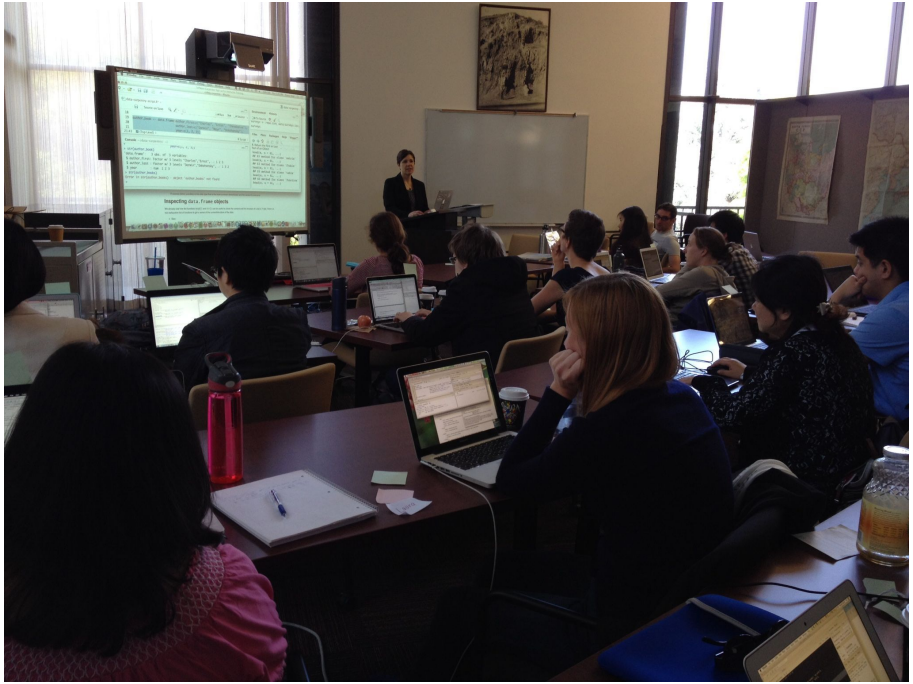
# Creating a Data Carpentry Biodiversity Curriculum

François Michonneau -- The Carpentries -- francois@carpentries.org

Deb Paul -- iDigBio -- dpaul@fsu.edu

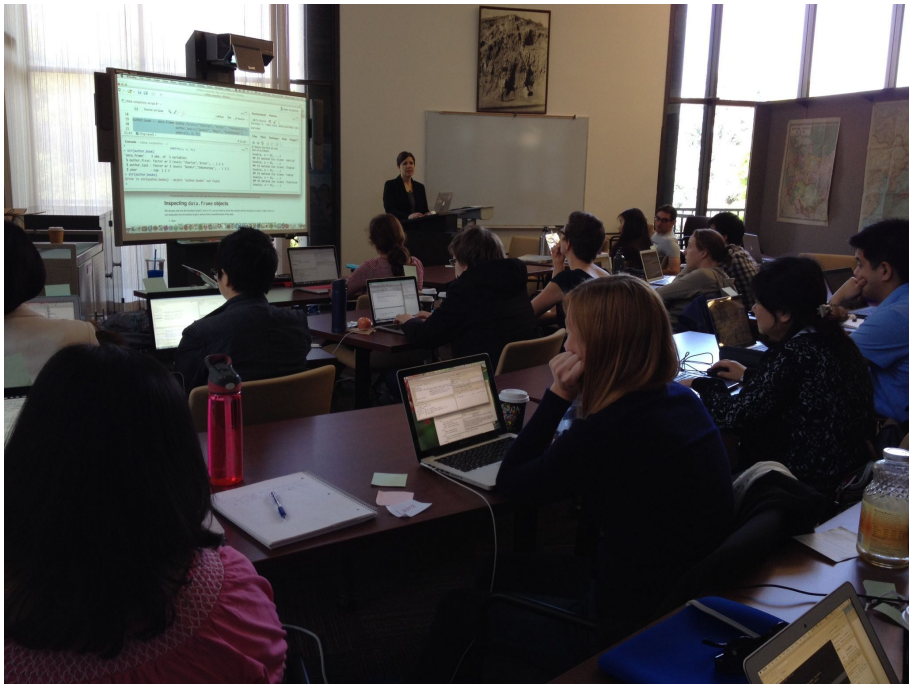
# The Carpentries

- 2-day workshops
- Focus on foundational computing and data science skills for researchers with no or little coding experience
- Volunteer organization
  - 1,300+ instructors worldwide
- Local communities



# The Carpentries Curriculum

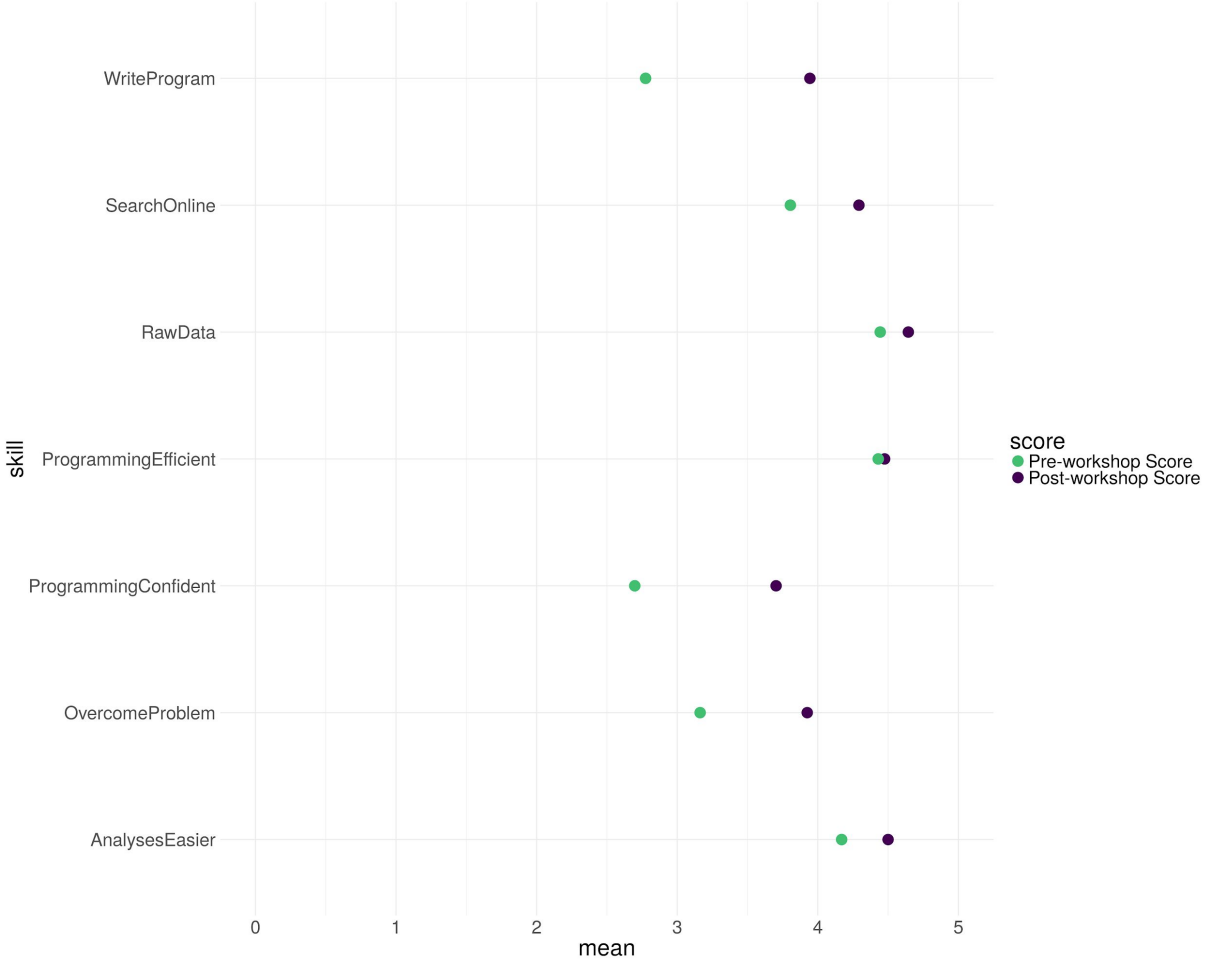
- Software Carpentry
  - Building robust research software
- Data Carpentry
  - Skills for working with data
  - Domain specific
    - Ecology
    - Genomics
    - Social Sciences
    - Geospatial
    - Astronomy
    - Economics
    - Image analysis
    - ....



# Our Workshops. Our Learners.

# Our Reach.

# Our Impact


WriteProgram
SearchOnline
RawData
ProgrammingEfficient
ProgrammingConfident
OvercomeProblem
AnalysesEasier
skill
0
1
2
3
4
5
mean
score
Pre-workshop Score
Post-workshop Score

# Long term

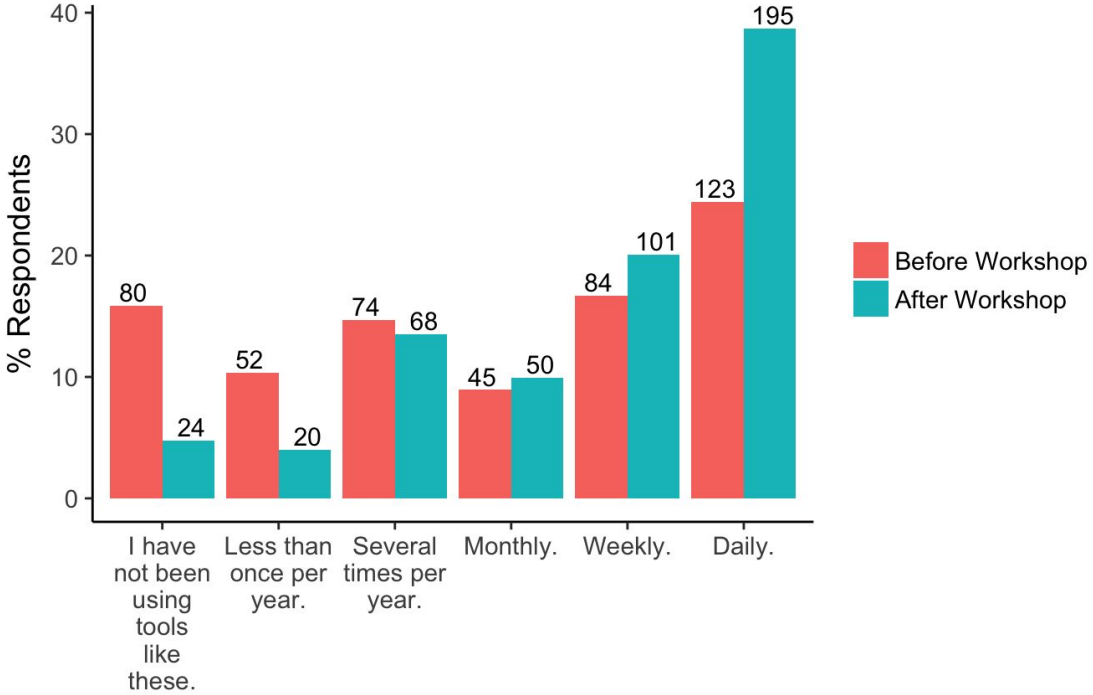
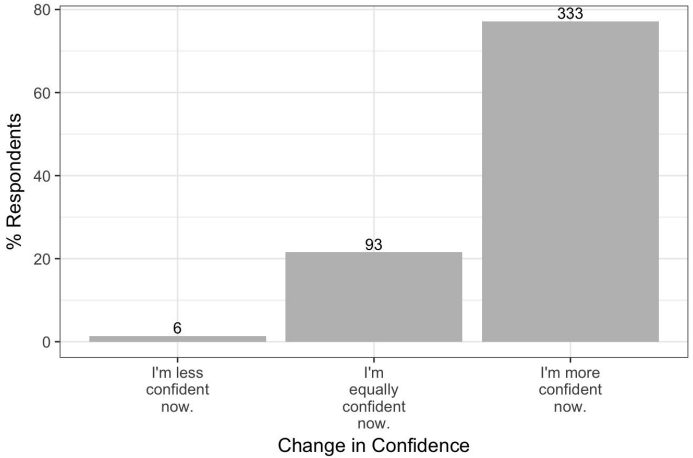



Figure 9: Respondents' Programming Usage Increased

Figure 12: Change in Confidence 6 Months Later

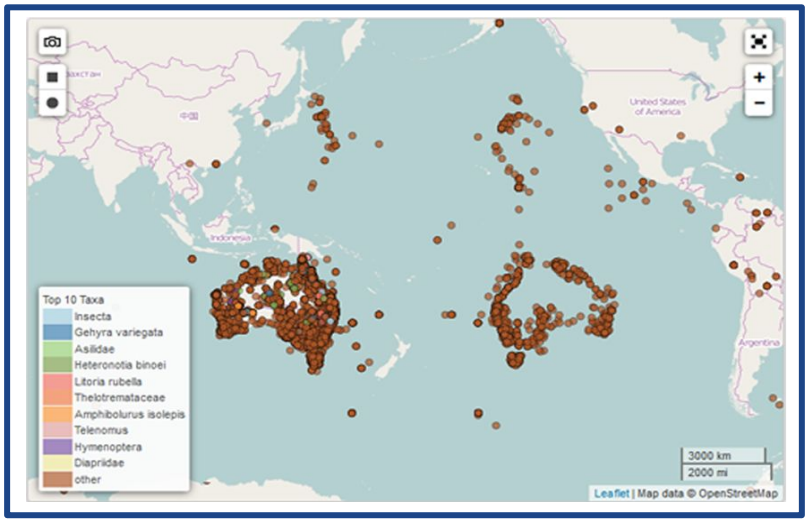


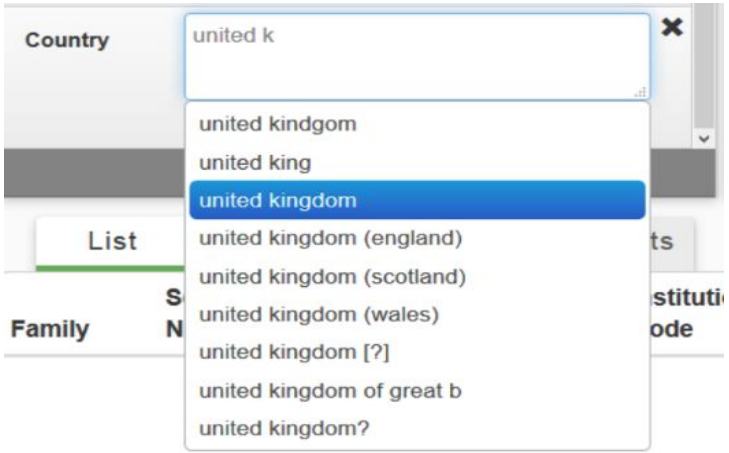
# What does a workshop look like?

- Data organization in spreadsheets
- OpenRefine
- Introduction to R or Python with focus on data manipulation and visualization
- SQL

# What would a biodiversity data curriculum look like?

- Identify skills and learning objectives
  - Challenges
    - Diverse data
    - Data quality and data cleaning




# Data Carpentry Biodiversity data curriculum

- Curriculum developed and sustained by the community
  - Who is interested in getting involved?

- Support from institutional and organizational collaborators
  - Partnership with existing training programs?

- Instructor training
- Maintainers onboarding

# What possible skills to include?

## some known common skill needs

- data evaluation (fitness for use)
- better spreadsheet skills
- (meta) data standards
- data publishing and reproducible research (FAIR)
- using authority files (taxonomy, geography, people)
- georeferencing, geospatial data skills
- data visualization
- data formatting and transformation
- scripting and apis
- specific tools/software: GeoLocate, Open Refine, GBIF tools, Kurator, GNR, collection management software,...
- data and media: management
- data quality (messy data, cleaning, validation, automation), visualization, data life cycle

## repetitive tasks

- data cleaning
  - transcription
  - editing
  - updating
  - mapping
- batch editing
- data import, export