Software and Data Carpentries at the UCSB Library

Our First Year

UC SANTA BARBARA LIBRARY

Jon Jablonski

PRRLA

9/2/2019

Who am I?

- UCSB Librarian
- Collaboratory director
- Early literacy advocate
- Geographer
- China Fulbright alum

Strong demand for 'command-line' tools

Carpentries.org

THE CARPENTRIES

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

- Membership benefits:
  - Access to certified instructors
  - 6 - 15 seats in instructor training
  - A strong community with excellent instructional materials
- Fills an identified need on campus
  - Without needing to develop our own materials.

Workshops are…

- Strong pedagogy and fills gaps in curriculum
- 2 full days: a big commitment
- BYO Device
- Created on github.io
- For novices
- Designed to teach reproducible data science

- Year 1 instructor cohort
  all volunteers -1
  - Web programmer
  - Reasearch data curator
  - Copy cataloger
  - Central IT
  - Bio post-doc

New Membership

Software

Data

Library

Spatial

5 workshops Jan – May
extended workshop thru summer

Open to all affiliates:

- Faculty, grads, computing staff*, library staff

*CIO has a large re-training effort for departmental computer support staff

THE CARPENTRIES

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

Data Carpentry

Lessons

Lesson	Site	Repository	Reference	Instructor Notes	Status	Maintainer(s)
Introduction to Data	🗔	🐙	👁	⊕	Stable	Carmi Cronje, Paul Pival, Shari Laster*, Anton Angelo (Past Maintainer: James Baker)
The UNIX Shell	🗔	🐙	👁	⊕	Stable	Danielle Kane*, Nilani Ganeshwaran, John Wright, Anna Oates, Belinda Weaver, Tim Dennis
OpenRefine	🗔	🐙	👁	⊕	Stable	Owen Stephens, Juliane Schneider, Paul Pival, Kristin Lee, Erin Carrillo* (Past Maintainer: Carmi Cronje)
Introduction to Git	🗔	🐙	👁	⊕	Beta	Chris Erdmann, Thea Atwood, Drew Heles*, Katrin Leinweber, Eva Seidlmayer (Past Maintainers: Belinda Weaver, Jez Cope)

Canonical Software Carpentry:

- Bash / Unix commands
- Gis
- R or Python

Extended Curriculum

The following Library Carpentry lessons can also be taught in addition to our core curriculum. These lessons have been taught infrequently and still need further work. We would value any feedback on these lessons.

Lessons

Lesson	Site	Repository	Reference	Instructor Notes	Status	Maintainer(s)
Tidy Data	🗔	🐙	👁	⊕	Beta	Tim Dennis, Thea Atwood, Sherry Lake*, Erika Mias (Past Maintainer: Jez Cope)
SQL	🗔	🐙	👁	⊕	Beta	Kristin Lee, Chris Erdmann, Jordan Pedersen* (Past Maintainers: Elaine Wong, Janice Chan)
Webscraping	🗔	🐙	👁	⊕	Alpha	Joshua Dull*, Thomas Guignard (Past Maintainer: Belinda Weaver)
Introduction to Python	🗔	🐙	👁	⊕	Alpha	Konrad Foerstner, Drew Heles, Elizabeth Wickes, Laura Wrubel* (Past Maintainers: Carlos Martinez, Richard Vankoningsveld)
Introduction to Data for Archivists	🗔	🐙	👁	⊕	Alpha	Jeanine Finn, Katherine Koziar*, and Scott Peterson (Past Maintainers: Jenny Bunn, Noah Geraci, and James Baker)

Library Carpentry

Code of Conduct

We are dedicated to providing a welcoming and supportive environment for all people, regardless of background or identity. By participating in this community, participants accept to abide by The Carpentries' Code of Conduct and accept the procedures by which any Code of Conduct incidents are resolved. Any form or behaviour to exclude, intimidate, or cause discomfort is a violation of the Code of Conduct. In order to foster a positive and professional learning environment we encourage the following kinds of behaviours in all platforms and events:

- Use welcoming and inclusive language
- Be respectful of different viewpoints and experiences
- Gracefully accept constructive criticism
- Focus on what is best for the community
- Show courtesy and respect towards other

Q6

I felt comfortable learning in this workshop environment.

Answered: 9   Skipped: 2
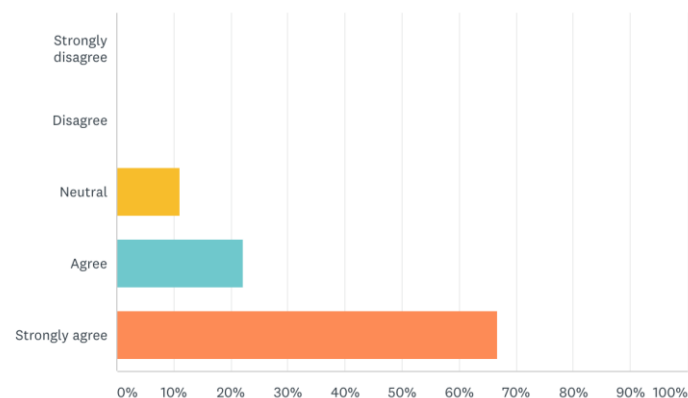
Workshop Logistics

Helpers
Sticky Notes
Hospitality

Surveys

Please be sure to complete these surveys before and after the workshop.

Pre-workshop Survey

Post-workshop Survey

Schedule

Before the Workshop

Complete the Setup steps below to:
- install the software we'll be using in the Workshop
- download the example files we'll be using in the workshop

Day 1: Thursday, April 11, 2019

Before	Pre-workshop survey
08:15	Installation help. Please arrive early if you had problems installing all of the required software.
08:30	Coffee and light breakfast

Day 2: Friday, April 12, 2019

08:30	Coffee and light breakfast
09:00	Introduction to R and Rstudio
10:00	Project Management with Rstudio

Type along

Copy-paste is cheating

Adding columns and rows in data frames

We already learned that the columns of a data frame are vectors, so that our data are consistent in type throughout the columns. As such, if we want to add a new column, we can start by making a new vector:

```
age <- c(2, 3, 5)
cats
```

```
   coat weight likes_string
1 calico    2.1            1
2  black    5.0            0
3  tabby    3.2            1
```

We can then add this as a column via:

```
cbind(cats, age)
```

```
   coat weight likes_string age
1 calico    2.1            1   2
2  black    5.0            0   3
3  tabby    3.2            1   5
```

Challenges

Pick-and-choose during the days

Interactivity every 5 minutes

Cognitive load = regular formative assessment

✏️ Appending Data

We have already met the `head` command, which prints lines from the start of a file. `tail` is similar, but prints lines from the end of a file instead.

Consider the file `data-shell/data/animals.txt`. After these commands, select the answer that corresponds to the file `animalsUpd.txt`:
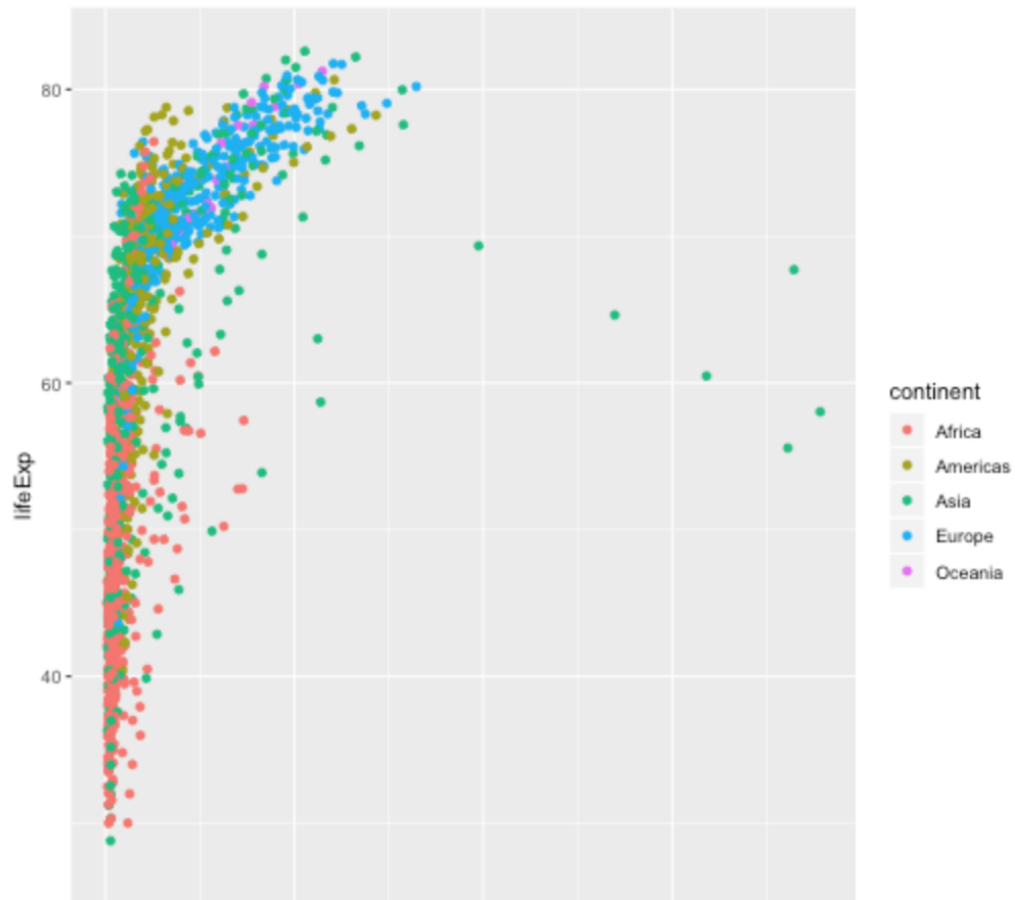
```
$ head -n 3 animals.txt > animalsUpd.txt
$ tail -n 2 animals.txt >> animalsUpd.txt
```

1. The first three lines of `animals.txt`
2. The last two lines of `animals.txt`
3. The first three lines and the last two lines of `animals.txt`
4. The second and third lines of `animals.txt`

👁 Solution ▾

Useful Goals

```
ggplot(data = gapminder, aes(x = gdpPercap, y = lifeExp, color=continent)) +
    geom_point()
```

Similar techniques everywhere!

Δ confidence

Pre- and Post- Assessment

- Overwhelmingly positive
    - With definite room for improvement
- Largest gains:
    - "I can write a small program/script/macro to solve a problem in my own work.
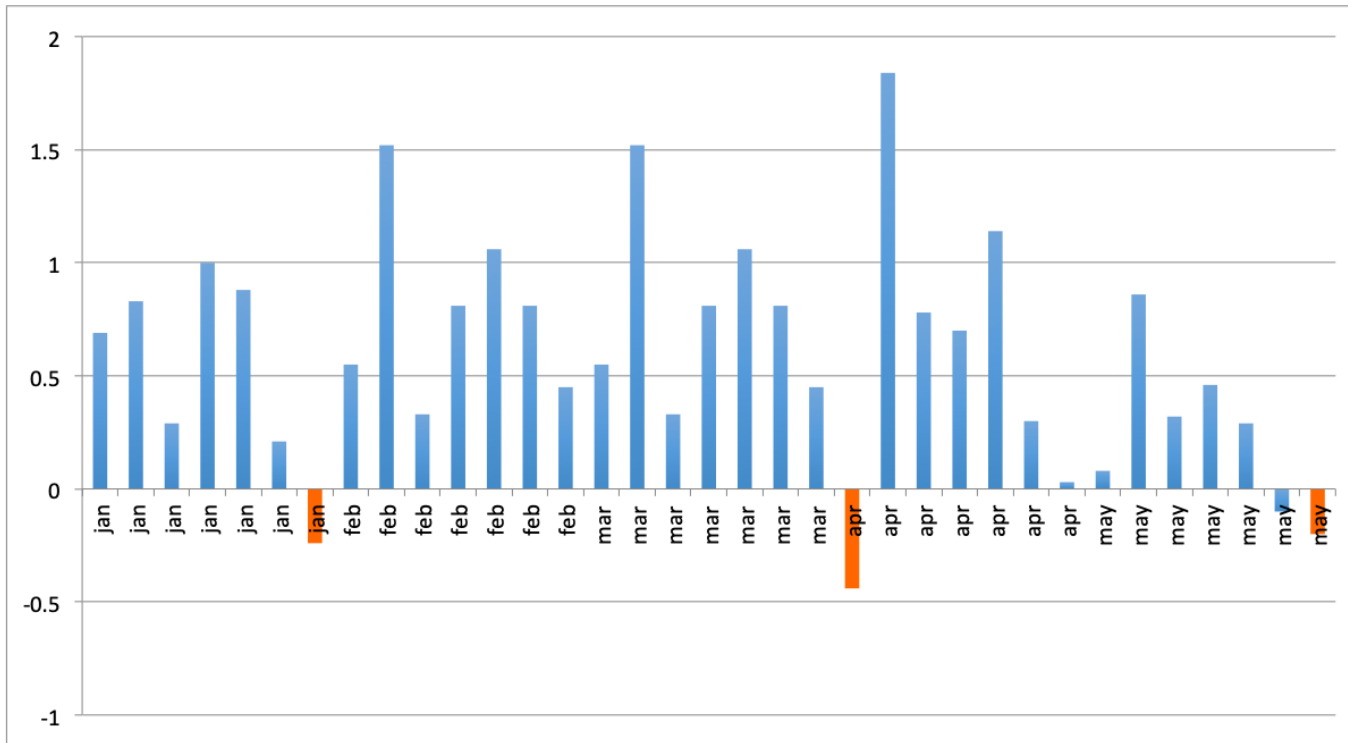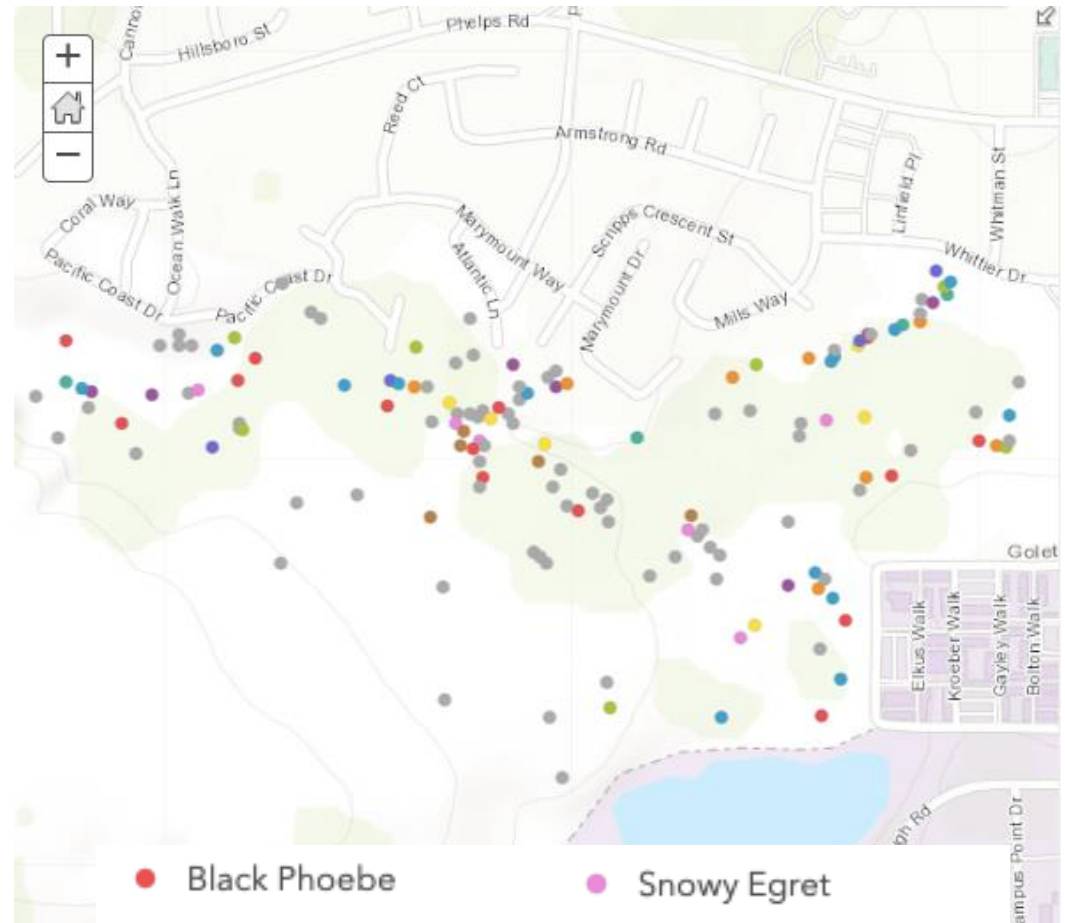
Pre- and Post- Assessment

- Losses:
    - "Using a programming language (like R or Python) can make me more efficient at working with data."
    - "Having access to the original, raw data is important to be able to repeat an analysis."

Moving forward

- Increased membership level to increase instructor pool
- 1 workshop / month is the goal
  - Fall 2019
    - R
    - Python
    - Spatial R
    - New instructors
- Create local examples

Legend:
- Black Phoebe
- Song Sparrow
- Anna's Hummingbird
- Common Yellowthroat
- House Finch
- Mallard
- Snowy Egret
- Killdeer
- Scaly-breasted Munia
- Cliff Swallow
- Other

Thank you!