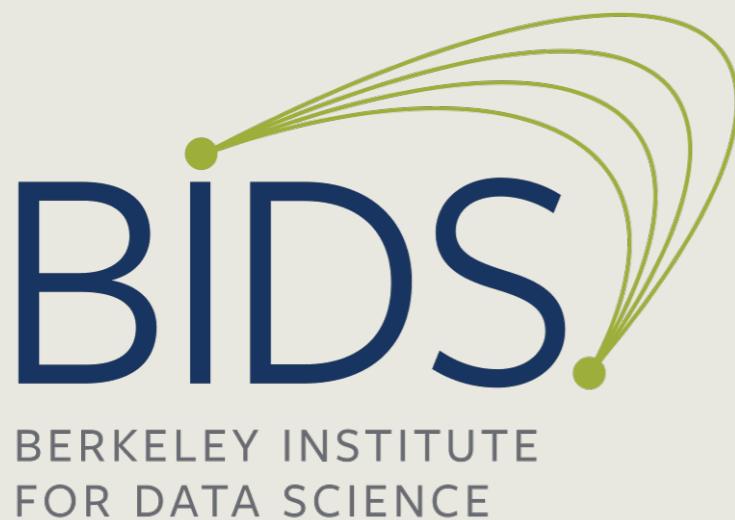


Unleashing the power of biodiversity data with data science

April 17, 2019

Ciera Martinez & Sara Stoudt



Mission: Safer Internet

Foundation Corporation



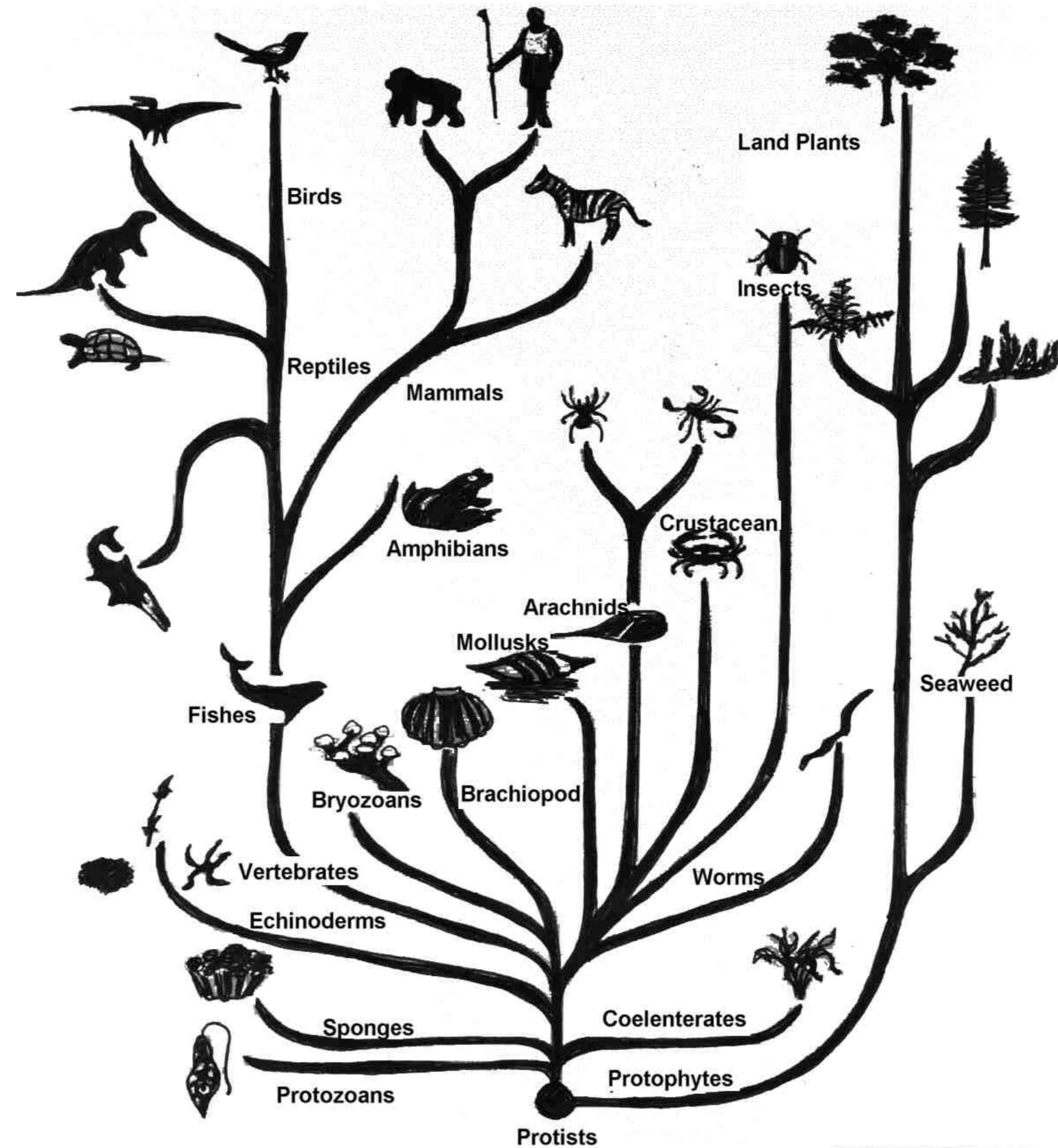
moz://a

Mission: Safer Internet



moz://a

1. Decentralization
2. Privacy and security
3. Digital Inclusion
4. Web Literacy
5. Openness
 - Science
 - Publishing
 - Hardware
 - Software
 - Ethics
 - Data

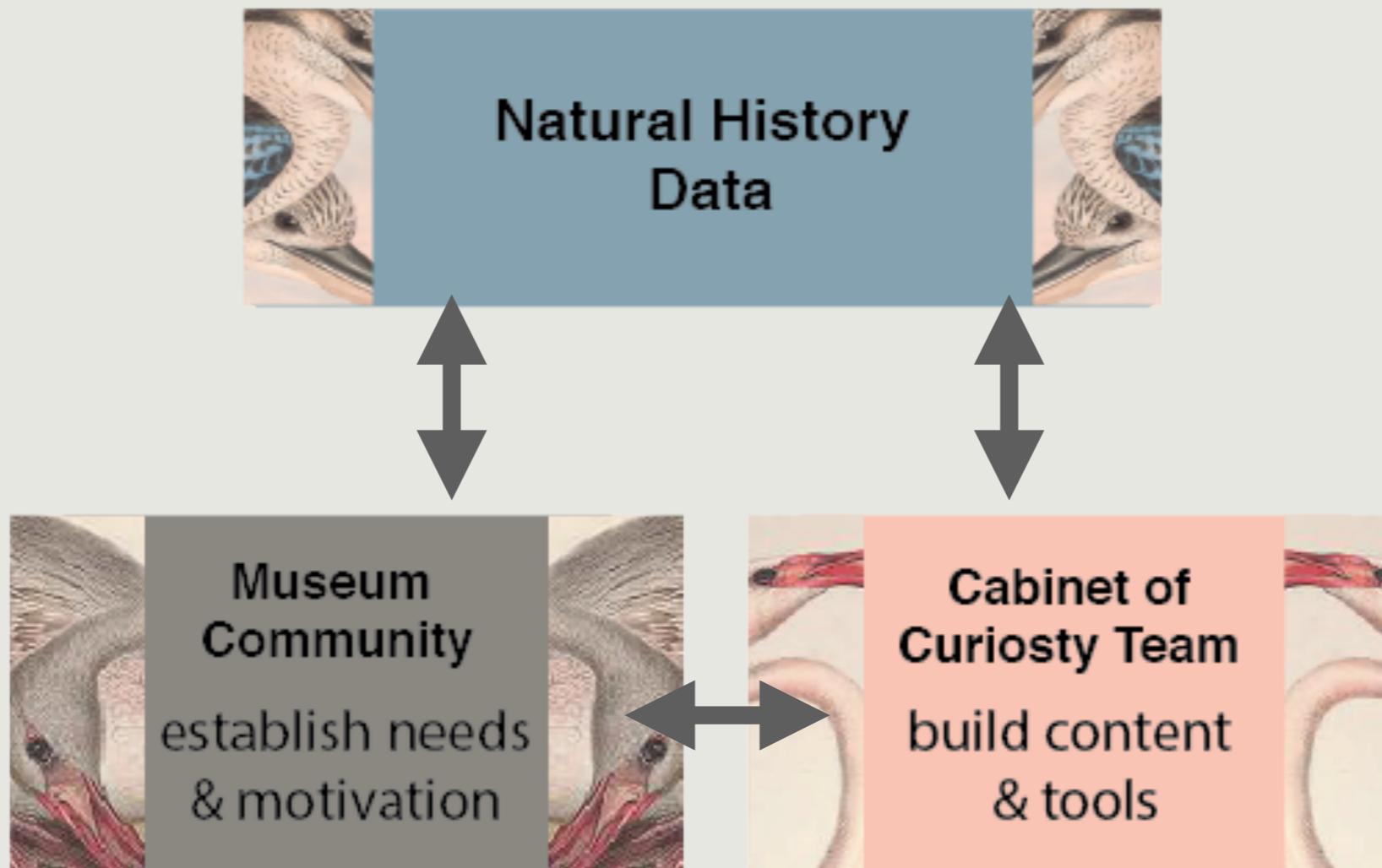






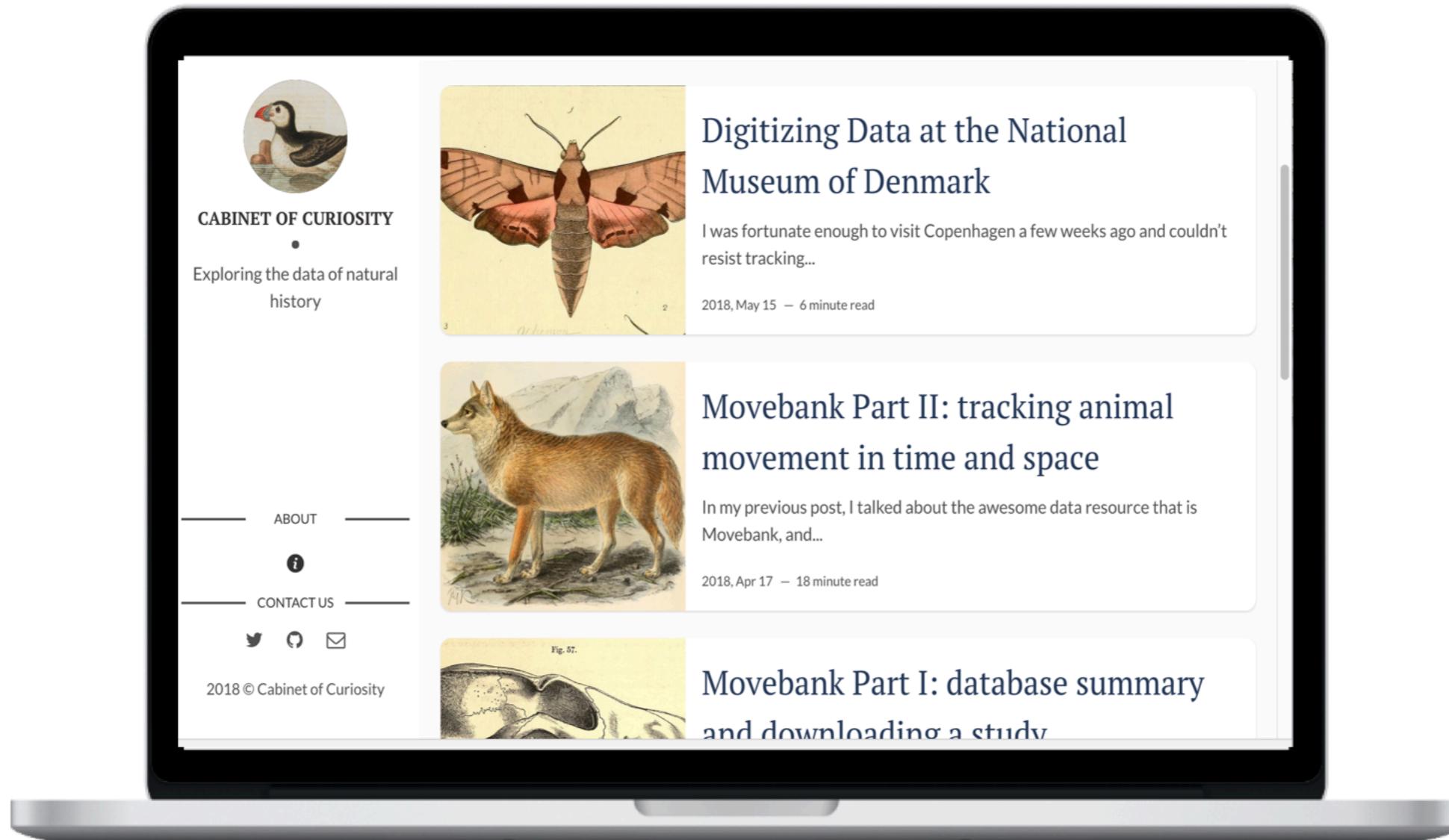
Mozilla Project Goals

1. Empower programmers to build tools to help researchers use data.
2. Create tutorials, curriculum, and data standards for Natural History Data community and classrooms.
3. Reinforce and support continuation and sustainability of digitization efforts.



Cabinet of Curiosity

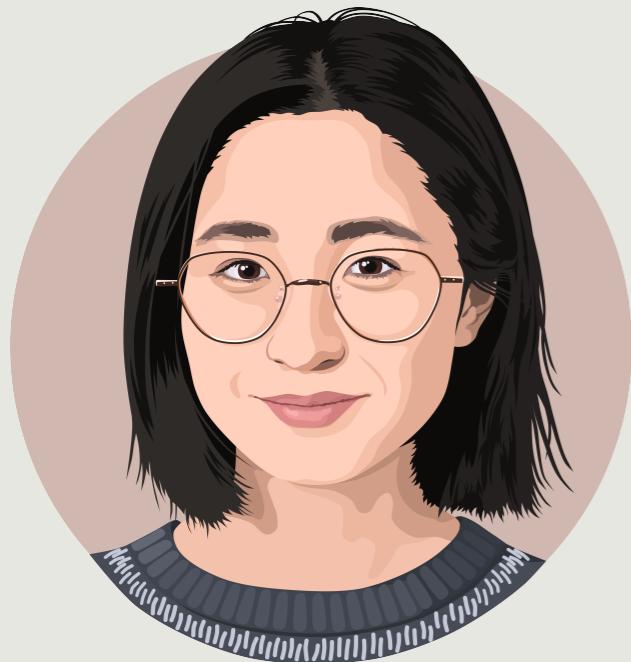
Exploring the data of natural history



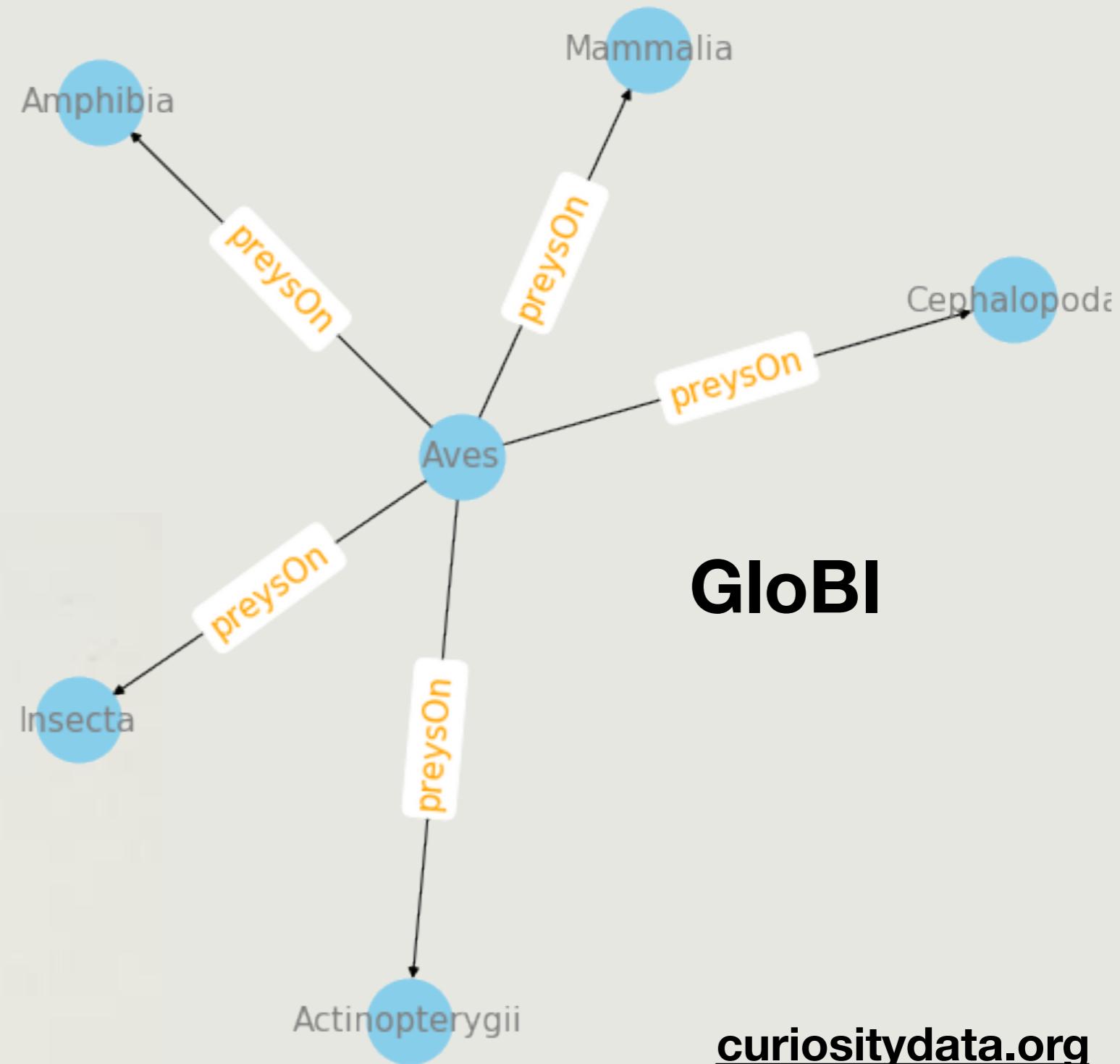
curiositydata.org/



Ask how animals interact with network graph visualization



Yikang Li



GloBI

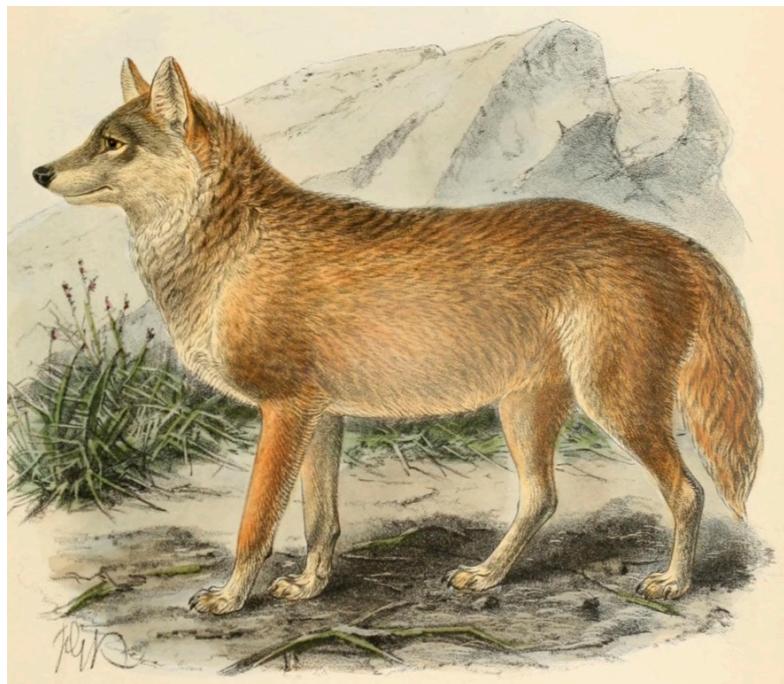
curiositydata.org

Tracking animal movement in time and space

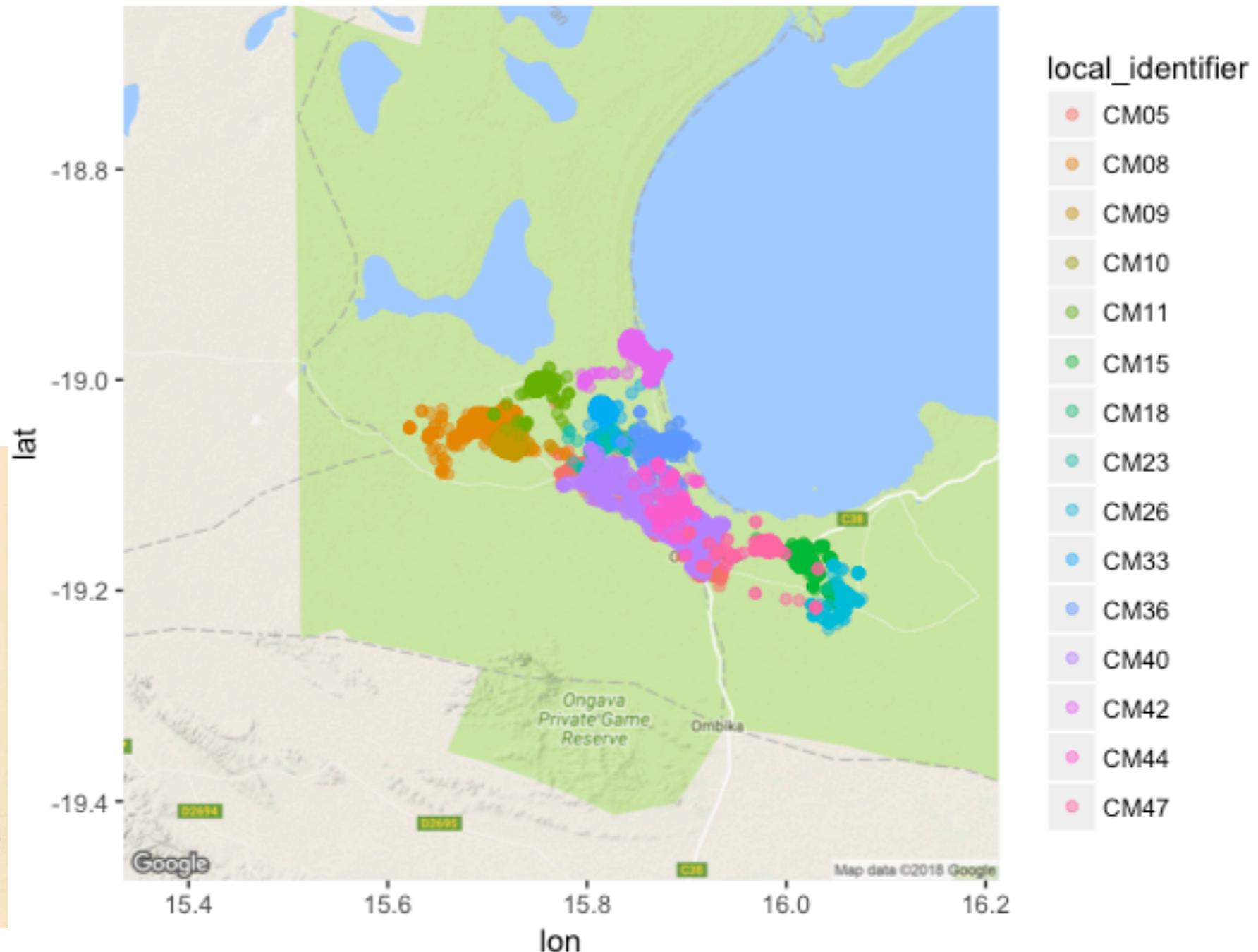
MoveBank



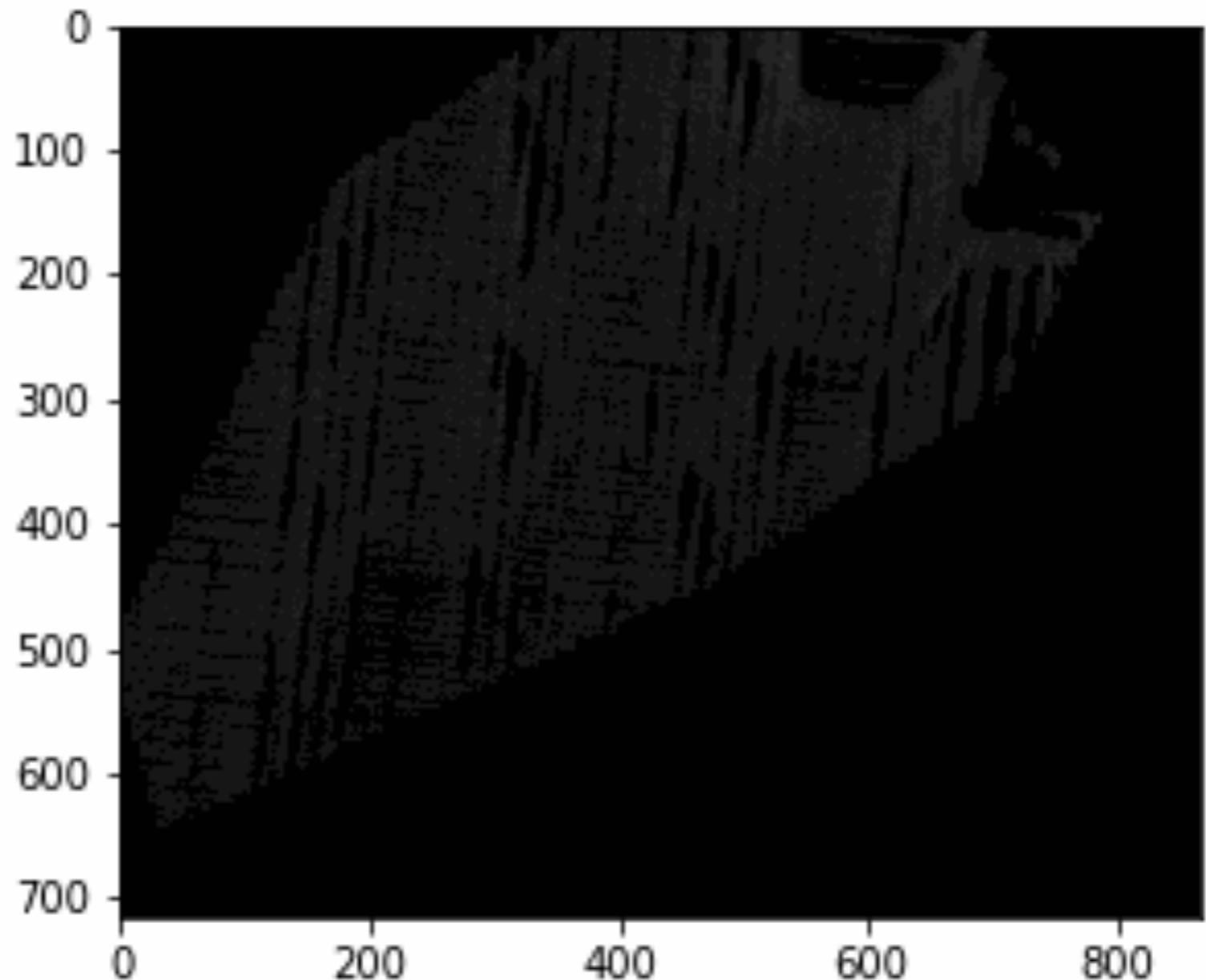
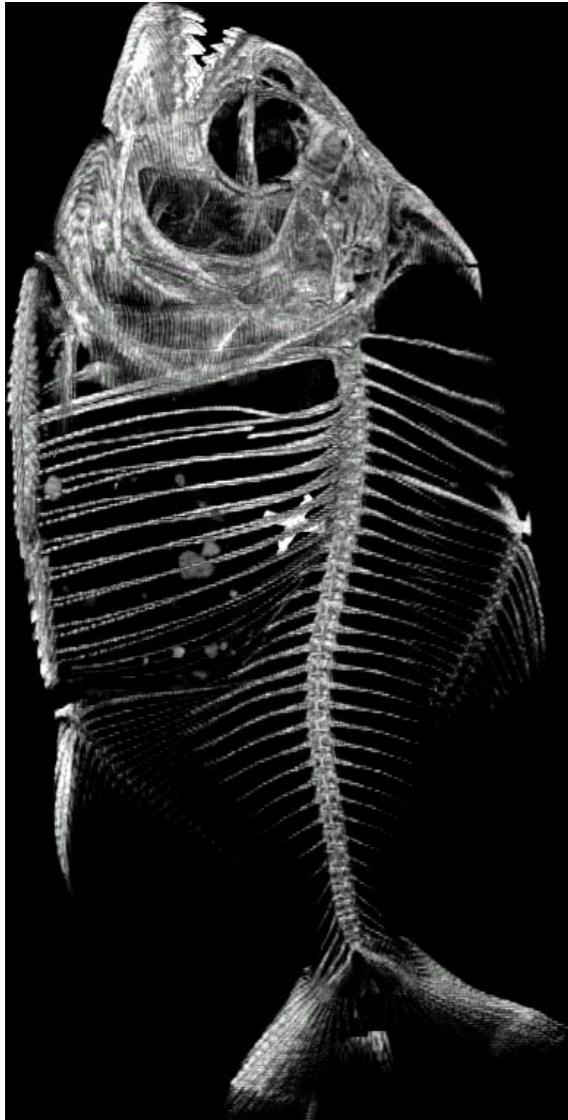
Caryn Johansen



Jackals in April 2009



Soon: Explore 3D animal skeletons using Python Morphosource



Other data

- Occurrence
- Catalog
- Ecological
- Specimen phenotype
- Environmental
- Specimen
- Genotype

List of Databases with free data: bit.ly/data_list

Other data

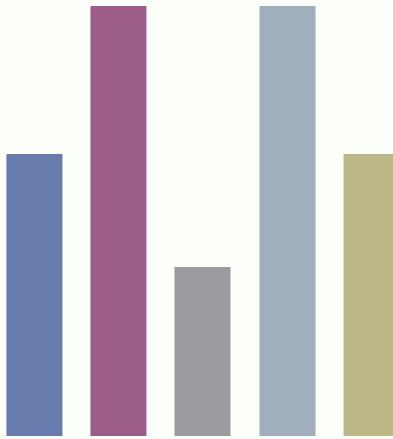
- Occurrence
- Catalog
- Ecological
- Specimen phenotype
- Environmental
- Specimen
- Genotype

List of Databases with free data: bit.ly/data_list

Become a Data Scientist

Data Scientist Job Postings





Find me at

Mozilla project



@cierareports



@iamciera

cieramartinez.org

curiositydata.org

mozillasciencefellows.org

How to contribute: bit.ly/coc_homebase

Who am I?

- Sara Stoudt
- PhD Candidate in Statistics at UC Berkeley
- Berkeley Institute for Data Science Fellow
- @sastoudt
- <https://sastoudt.github.io/>
- **Coming Soon**: LA Times Data Desk Intern
- **Coming Soon**: Communicating with Data: The Art of Writing for Data Science (with Deb Nolan)

Why work in the open?

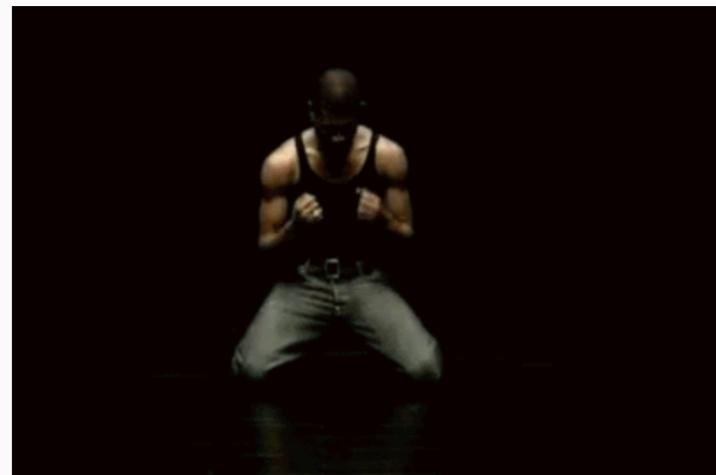
My experience blogging about #rstats

Get Motivated

Confessions of a Tidyverse Straggler

PUBLISHED ON APR 18, 2018

These are my confessions...



If I'm gonna tell it, then I gotta tell it all.

Obligatory Usher Reference

I learned R before the tidyverse (shout out to @BaumerBen's Multiple Regression class Fall 2016) and have failed to update my skills. There, I said it. I'm a statistics PhD student whose code is riddled with dollar signs (*pull* Sara!) and the occasional loop that I've never wrangled into a proper apply statement (which should now be switched to *purrr?*). And although I have picked up some *ggplot2* and *dplyr* (I'm not THAT much of an R hermit), I really don't know what new functionality is available. This prevents my code from keeping up with the times. It has always been on my to-do list to work through the tidyverse documentation so that I can start using the packages in my own work, but I've always prioritized "done" over "pretty".

But no longer! My plan of attack is to start blogging about my tidyverse immersion, one package at a time. Each week(ish) I will read up on a tidyverse package and ideally, update some gross code from one of my current or previous projects,

Retweeted 16 times



Sara Stoudt @sastoudt · 18 Apr 2018

Confessions of a Tidyverse Straggler: I can't back down from my goal now; it's in writing. [sastoudt.github.io/blog/2018/04/1...](https://sastoudt.github.io/blog/2018/04/18-confessions-of-a-tidyverse-straggler/)

He/Him



Emily Robinson @robinson_es

Data Scientist @DataCamp, formerly @Etsy. #rstats. She/her

Following



Hadley Wickham ✅ @hadleywickham

R, data, visualisation.

Following



David Robinson @drob

Chief Data Scientist at @DataCamp, #rstats fan/evangelist

Following



Get Help



Sara Stoudt
@sastoudt

Tardy **#TidyTuesday** still in progress... **#rstats** users give me strength. Can someone please show me how to make the geom_bar fill order match the legend?



sastoudt/tidytuesday

Repo for initial setup of the #tidytuesday visualization project - sastoudt/tidytuesday
github.com

9:24 PM - 17 May 2018



Hadley Wickham ✅ @hadleywickham · 18 May 2018

Replying to [@sastoudt](#)

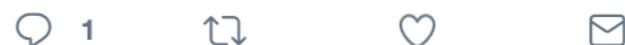
I think dodging happens in the order of rows in the data frame, so try `arrange()`ing?



Thomas Mock 🏠 @thomas_mock · 18 May 2018

Replying to [@sastoudt](#)

Hey [@sastoudt](#)! did you have any luck with the other code?



Sara Stoudt @sastoudt · 18 May 2018

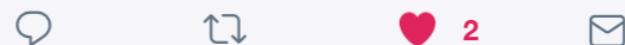
Yup, just figured it out & updated the page. Unfortunately the order parameter that [@ibddoctor](#) suggested didn't work, but a mix of `arrange` (as [@hadleywickham](#) suggested) and `as.data.frame` did the trick (although why the latter is necessary remains a mystery to me).



Peter Higgins @ibddoctor · 18 May 2018

Replying to [@sastoudt](#)

Try learnr.wordpress.com/2010/03/23/ggp...



<https://sastoudt.github.io/tidytuesday/week7/>

Get Comfortable



Sara Stoudt
@sastoudt

you can teach a base R dog tidy tricks: some reflections on my adventures in tidyverse
sastoudt.github.io/blog/2018/09/01-adventures-in-tidyverse-reflections/ ... Stay tidy my friends!



Home About Blog

STAT 198 Blogging for Data Science

Spring 2018

<https://stat198-spring18.github.io/>

UC Berkeley

Professor Deborah Nolan

Graduate Student Instructor: Sara Stoudt



cabinetofcuriosity

<https://github.com/cabinetofcuriosity>

Repositories 14 People 0 Projects 0

Dismiss

Grow your team on GitHub

GitHub is home to over 31 million developers working together. Join them to grow your own development teams, manage permissions, and collaborate on projects.

Sign up

Find a repository... Type: All Language: All

coc_lewis_clark_workshop

Dockerfile GPL-2.0 Updated 7 hours ago

Top languages

Jupyter Notebook HTML CSS Dockerfile

<https://sastoudt.github.io/blog/2018/09/01-adventures-in-tidyverse-reflections/>

Get Noticed

Retweeted 5 times X

 **Sara Stoudt** @sastoudt · 18 Aug 2018
Tardy #TidyTuesday Take-Away: The followers and following patterns of Russian trolls on Twitter break Benford's Law (inspired by @jengolbeck). Thoughts @ollie? github.com/sastoudt/tidy...

 **Tejendra P Singh | ते** @tejendrapsin...
जेन्द्र
Chess Enthusiast, Vishy Anand fan,
Aspiring Economist, Data Science Lover,
R, #Stata #rstats. Retweets ≠ Endorsement

 **Michael Lopez** @StatsbyLopez
Data & Analytics @NFL, by way of Skidmore, Brown, UMass, and Bates.

 **Oliver Roeder** ✅ @ollie
senior writer, @fivethirtyeight. writing a book about games for @wwnorton.

- I got my interview and internship offer for the LA Times all in DMs on Twitter.
- My blog and Tidy Tuesday entries served as my data journalism portfolio.

<https://sastoudt.github.io/tidytuesday/week20/>

Resources

- <http://varianceexplained.org/r/start-blog/>
- <https://alison.rbind.io/post/up-and-running-with-blogdown/>
- <https://github.com/rfordatascience/tidytuesday>