

```
library(dplyr)  
  
rladies_global %>%  
  filter(city == 'Riverside')
```



| Kickoff meeting



Welcome!





Worldwide organization that
promotes **diversity** in the
#rstats community via meetups
and mentorship in a **friendly**
and **safe** environment

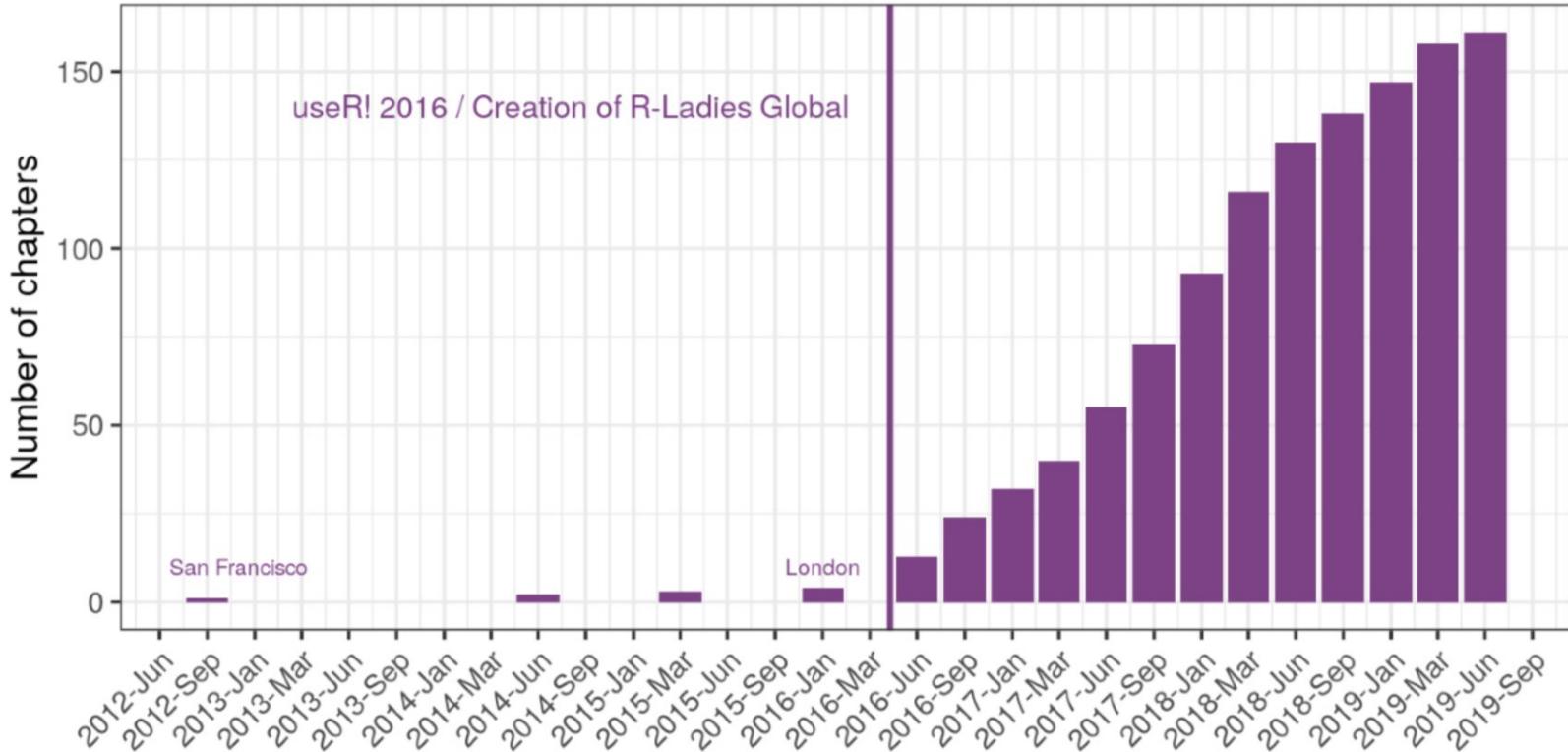
A photograph of the Golden Gate Bridge in San Francisco, California. The bridge's iconic red towers and suspension cables are visible against a backdrop of a cloudy sky and green hills. The water of the San Francisco Bay is in the foreground.

First R-Ladies meetup

San Francisco, CA

October 2012

R-Ladies Chapter Growth



R-Ladies at a glance



47
countries

159 cities

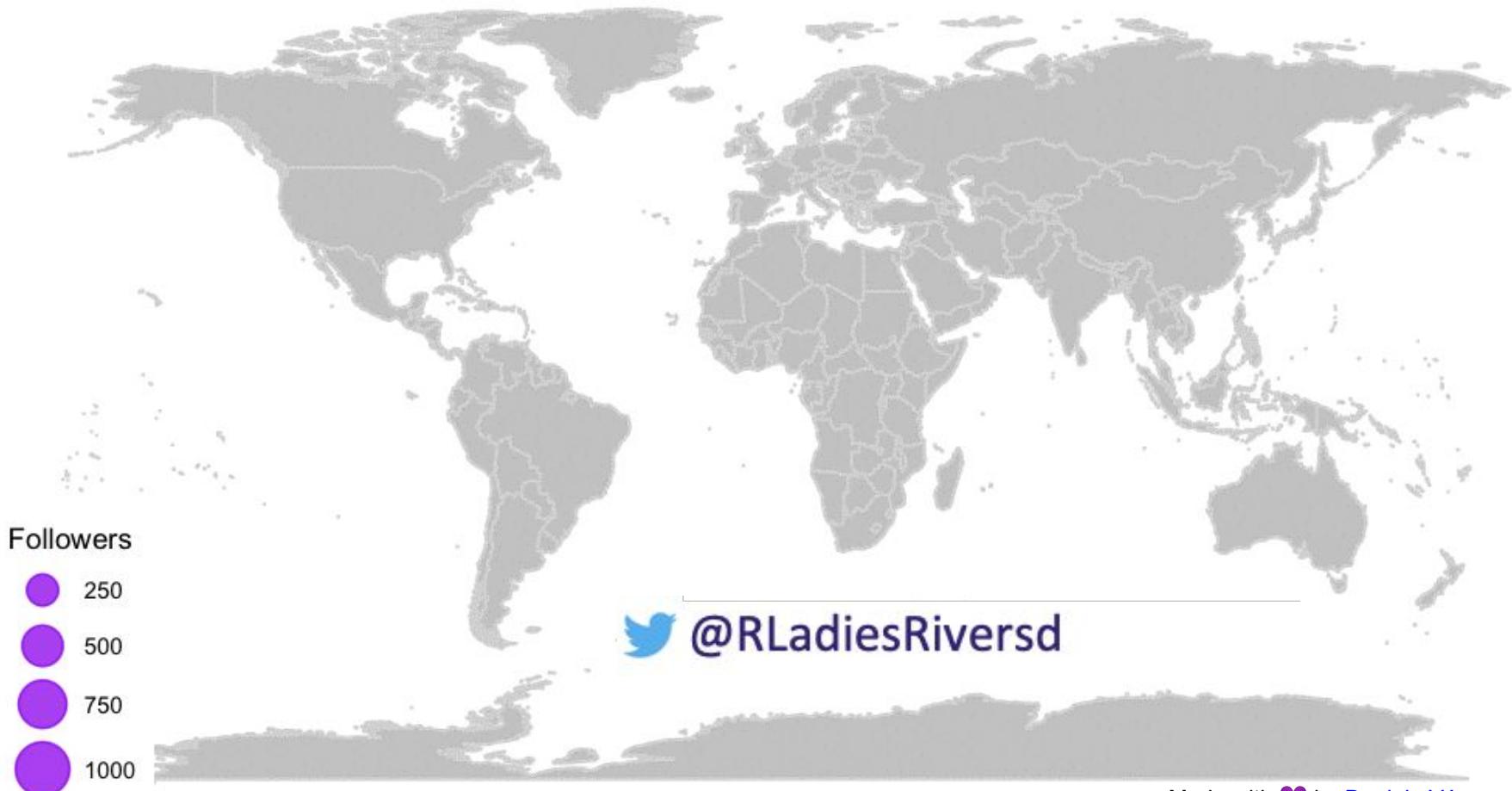
47,000+
members

1500+
events

(May 2019)

2011-09-01

Twitter followers by each chapter





Our mission

Achieve proportionate representation by encouraging, inspiring, and empowering people of genders currently underrepresented in the R community.





Code of conduct

<https://rladies.org/code-of-conduct/>

Riverside chapter co-founders



Marília P. Gaiarsa
UC Chancellor's Postdoc
Entomology

Favorite function: merge
Favorite package: dplyr
 magaiarsa



Daniela Cassol
Bioinformatics Postdoc

Favorite function: sapply
Favorite package:
systemPipeR
 dcassol



Teresa Bohner
Ph.D. Candidate
Botany and Plant Sciences

Favorite function: str
Favorite "package":
tidyverse



| Who else is here?

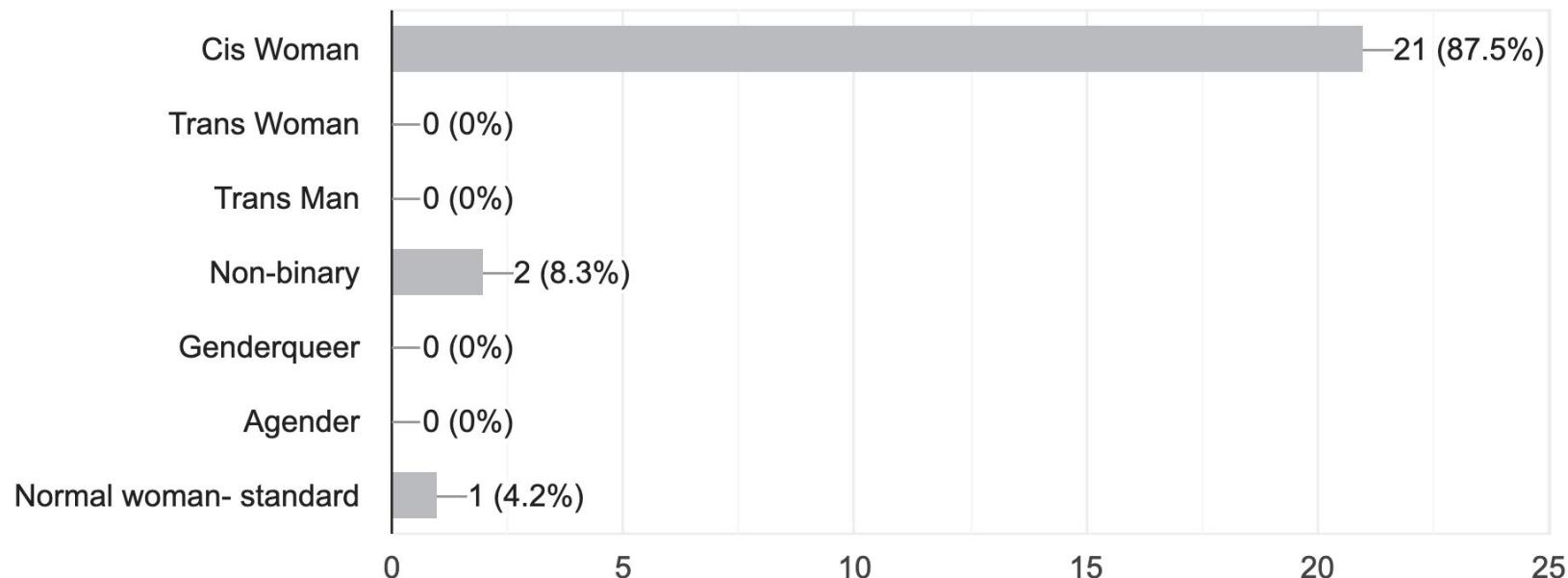


**26 responses to
questionnaire**

(3pm today)

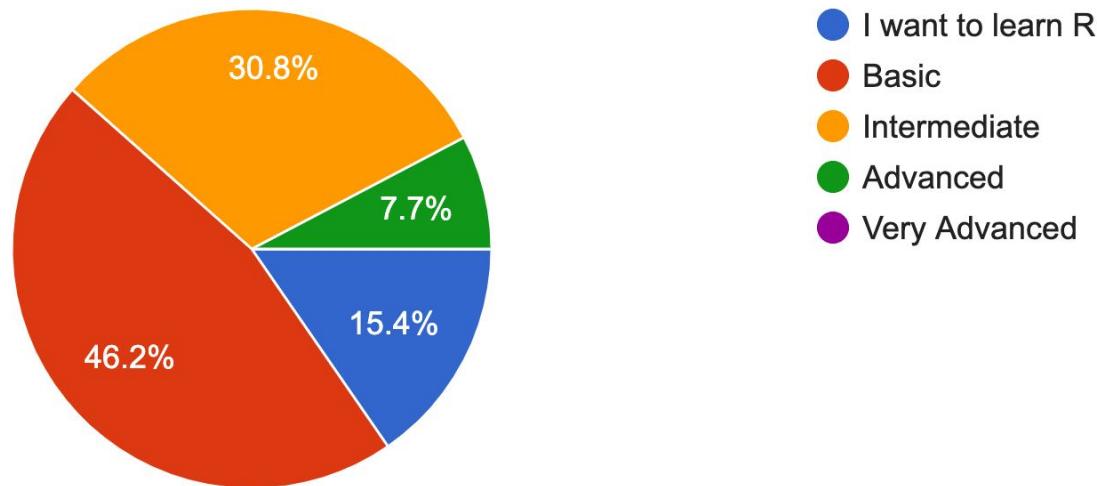
How do you self-identify?

24 responses



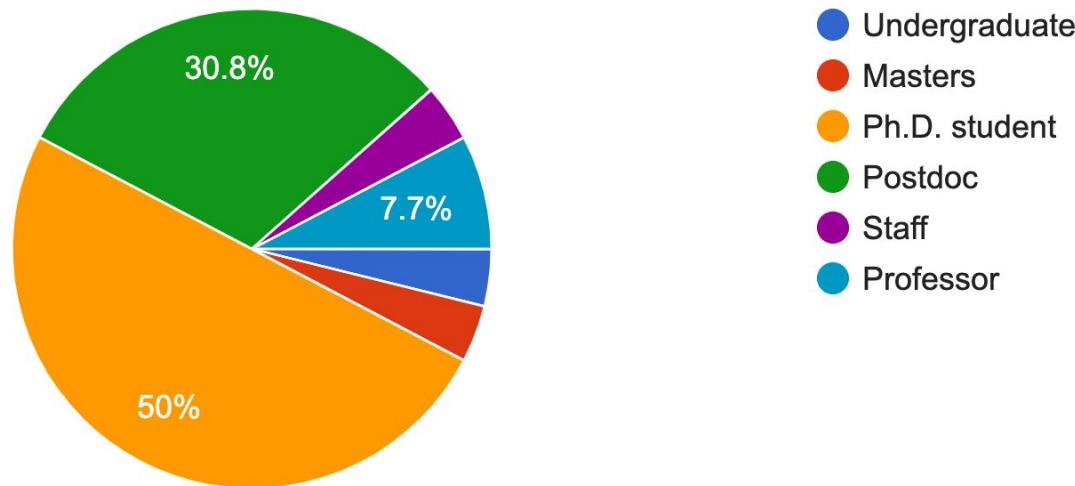
What is your level of experience with R?

26 responses



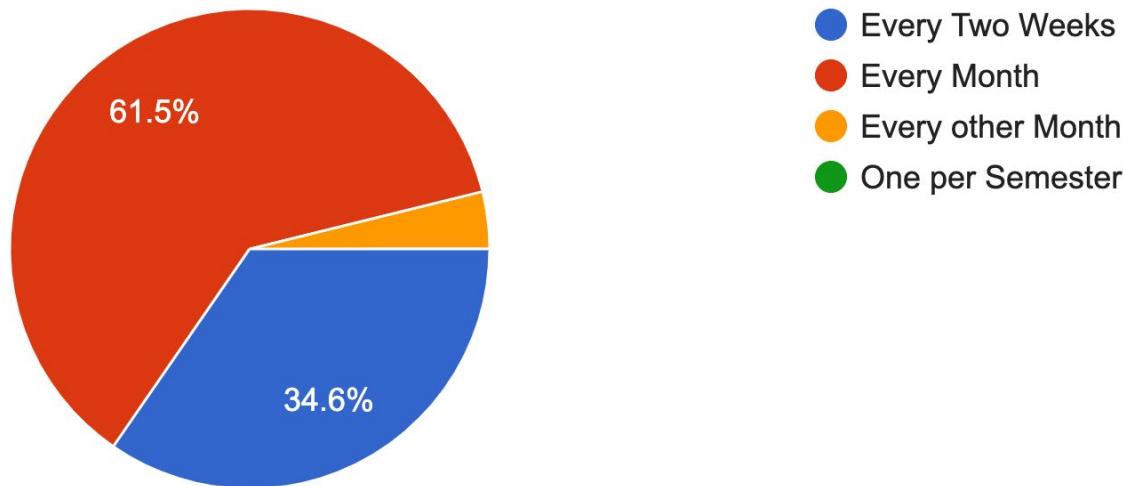
What is your academic stage?

26 responses



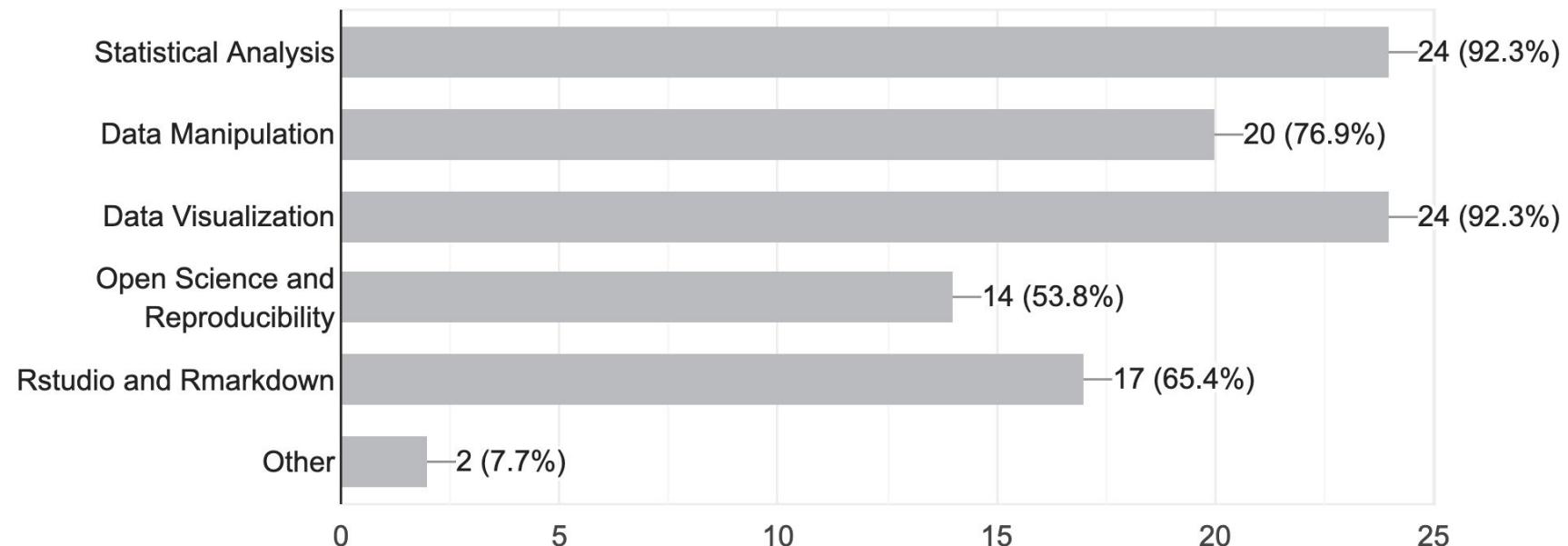
What would be the perfect interval between meetings?

26 responses



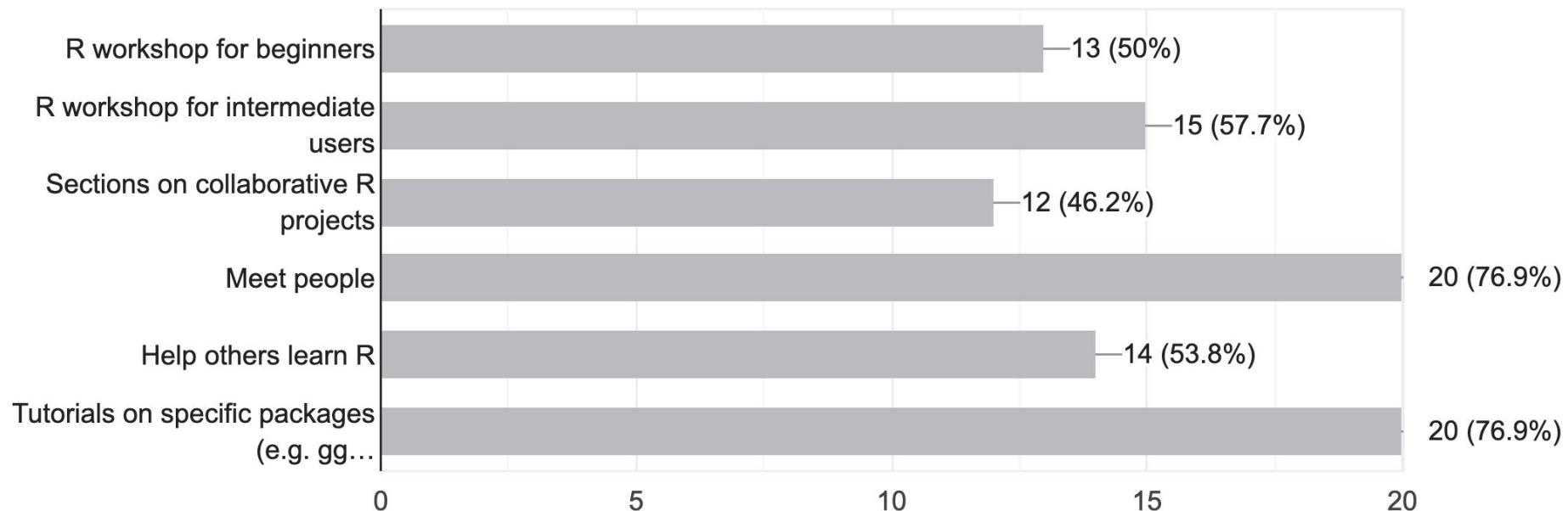
Which themes are you most interested in (check all that apply)?

26 responses



Which activity would you like to do in our meetings (select all that apply)?

26 responses





Our sponsors



Yelda Serin

GradQuant



Kat Koziar

*Data Librarian at
UCR Library*



Our sponsors



Yelda Serin
GradQuant



Kat Koziar
*Data Librarian at
UCR Library*

Hacky Hours: "During Hacky Hours you can talk with us about Python, R, Git, Bash, OpenRefine, data management, cleaning, visualization, & analysis, as well as quantitative methods."

Fridays, 1–3 PM - Hacky Hours



Lightning talks

Marília P. Gaiarsa: (I <3) Using R to simulate global change biology

Teresa Bohner: I <3 ecological data wrangling and making pretty figures

Daniela Cassol: I <3 open science and reproducibility

Magda Argueta: How do I use R and how I want to use it

Danelle Seymour: How R facilitates research in genetics and genomics



Using R to simulate global change biology

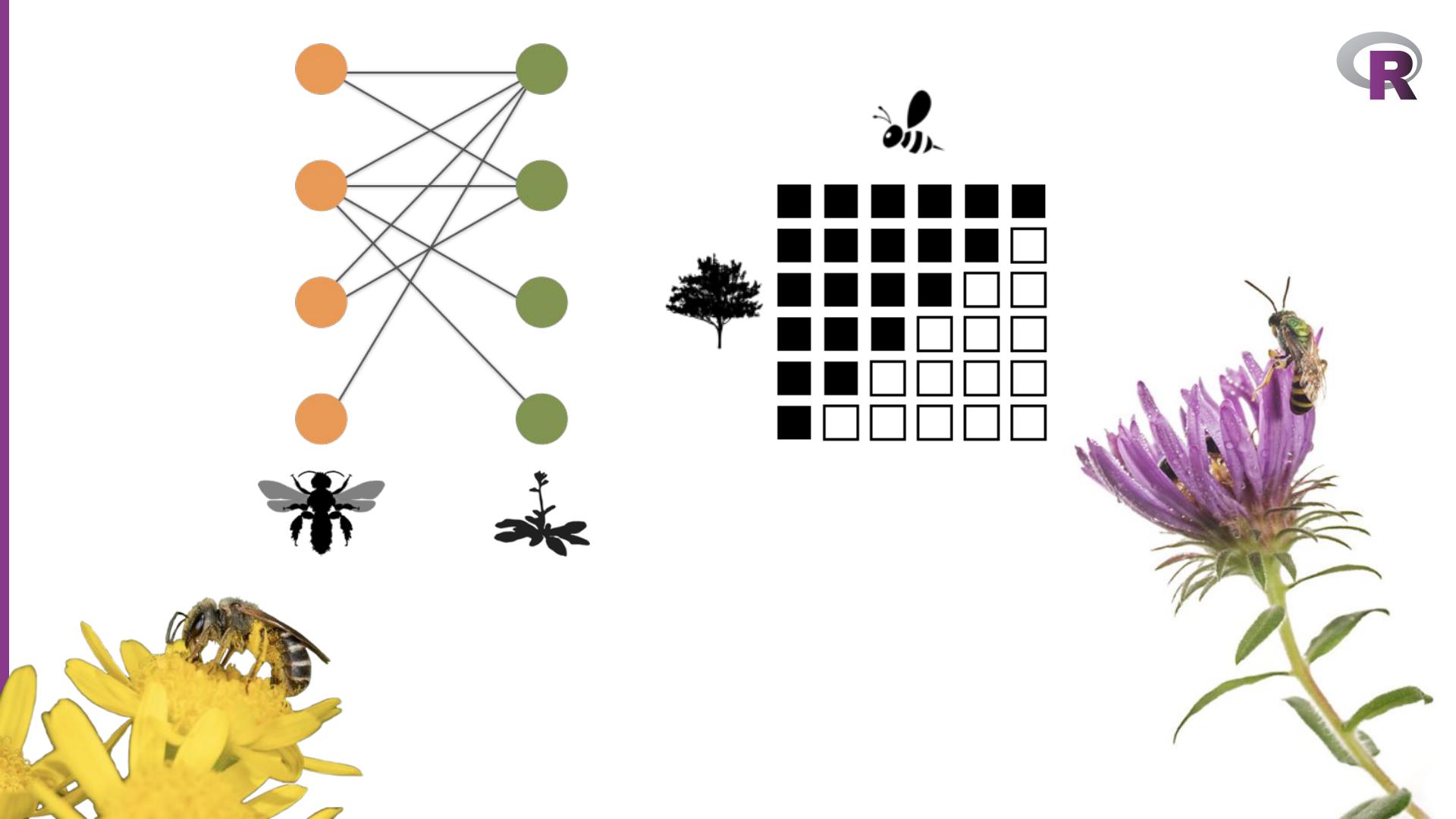
- Ethical and easier
- Tweak parameters



Using R to simulate global change biology & How to keep track of all the stuff you have already done

- Ethical and easier
- Tweak parameters

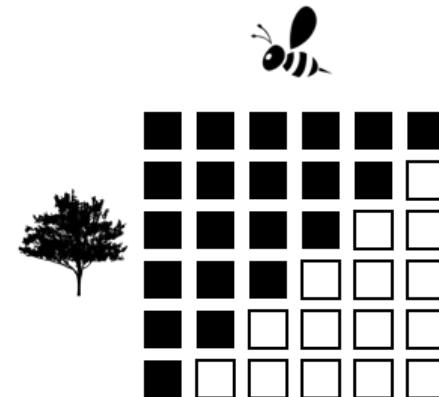






```
#' Function to calculate extinctions (modified from bipartite::extinct)
#' Always considers species from both groups and only binary matrix and inherits the parameters from
#' the secondary extinct function; the extinction probability is defined by
#' @param redeExtinct an interaction matrix
#' @return a matrix with a species extinct
#' number of primary extinctions, total extinctions per time step, nodf, and modularity
#' @author Marilia P. Gaiarsa (gaiarsa.mp@gmail.com)

extinction.marilia <-
  function (redeExtinct)
{
  nr <- NROW(redeExtinct)
  nc <- NCOL(redeExtinct)
  ## calculating the extinction probability for all rows and columns (integers)
  ## specialists have higher probability of extinction
  probab.row <- (max(rowSums(redeExtinct)) - rowSums(redeExtinct))+1 ## added one so there is no zero probability
  probab.col <- (max(colSums(redeExtinct)) - colSums(redeExtinct))+1 ## added one so there is no zero probability
  # sampling whether to exclude rows or cols
  col.or.row <-
    sample(2, 1)
  ## sample a row and a column to be excluded, based on the probability above
  rexcl <- sample(nr, 1, prob = probab.row)
  cexcl <- sample(nc, 1, prob = probab.col)
  if (col.or.row == 1){
    ## if samples a row, remove the row
    redeExtinct[rexcl, ] <- 0
    sec.ext <- sum(colSums(redeExtinct) == 0)}
  if (col.or.row == 2){
    ## otherwise remove column
    redeExtinct[, cexcl] <- 0
    sec.ext <- sum(rowSums(redeExtinct) == 0)}
  return(list(redeExtinct=redeExtinct,
             col.or.row=col.or.row,
             sec.ext=sec.ext))
}
```





```
---  
title: "Multiple drivers - a guide through the analysis"
```

```
author:
```

```
- Gaiarsa, Guimarães & Tylianakis
```

```
date: `r format(Sys.time(), '%d %B %Y')`"
```

```
output: html_document
```

```
---
```

```
```{r setup, include=FALSE}  
knitr::opts_chunk$set(echo = TRUE)
```
```

Creating the networks

The first step is to create the networks with fixed richness and connectance, and varying degrees of nestedness and modularity. All files are in the `"multipleDrivers/code/creatingNetworks/"` directory. We created the function `"make.nested.web.R"` and ran in parallel (`"createNwParallel.R"`). This code create nested networks and from the nested networks we used the code `"creatingModularNetworks.R"` to reorganize interactions and `"multipleDrivers/output/nws/sensitivity"` for the sensitivity analysis and in the paper}}\$

Multiple drivers - a guide through the analysis

Gaiarsa, Guimarães & Tylianakis

12 September 2019

Getting the mod and nodf of the networks

I created a code that calculated nestedness using NODF and the nestednodf (using XXX. The output is a data frame with the names of the networks (random names) and modularity.

Running the simulations scenarios

Perturbation spreading

The probability of being affected by the perturbation is inversely proportional to the number of interaction partners a species has. This probability remains constant through time, given that the network structure does not change in this simulation scenario.

Creating the networks

The first step is to create the networks with fixed richness and connectance, and varying degrees of nestedness and modularity. All files are in the `"multipleDrivers/code/creatingNetworks/"` directory. We created the function `"make.nested.web.R"` and ran in parallel (`"createNwParallel.R"`). This code create nested networks and from the nested networks we used the code `"creatingModularNetworks.R"` to reorganize interactions and change modularity. All created networks are in `"multipleDrivers/output/nws/sensitivity"` for the sensitivity analysis and in the folder **XX for the ones used in the paper**

****Code**:** script `*perturbation.R*` sources the functions `*perturbationSpread*` number of iterations as the number of dimensions, and each element is a matrix with the order of species affected at each time step.

Getting the mod and nodf of the networks

I created a code that calculated nestedness using NODF and the nestednodf (anticipating reviews comments), and modularity using XXX. The output is a data frame with the names of the networks (random names) and the richness, connectance, nestedness, and modularity.

Running the simulations scenarios

Perturbation spreading

The probability of being affected by the perturbation is inversely proportional to the number of interaction partners a species has. This probability remains constant through time, given that the network structure does not change in this simulation scenario.

Code: script `perturbation.R` sources the functions `perturbationSpreading.R` and `spreading.R`. Returns a list with number of iterations as the number of dimensions, and each element is a matrix with the order of species affected. Thus, `rowsums = number of species affected at each time step.`





I ❤️ Open Science
and Reproducibility





Hello!

I am Daniela Cassol

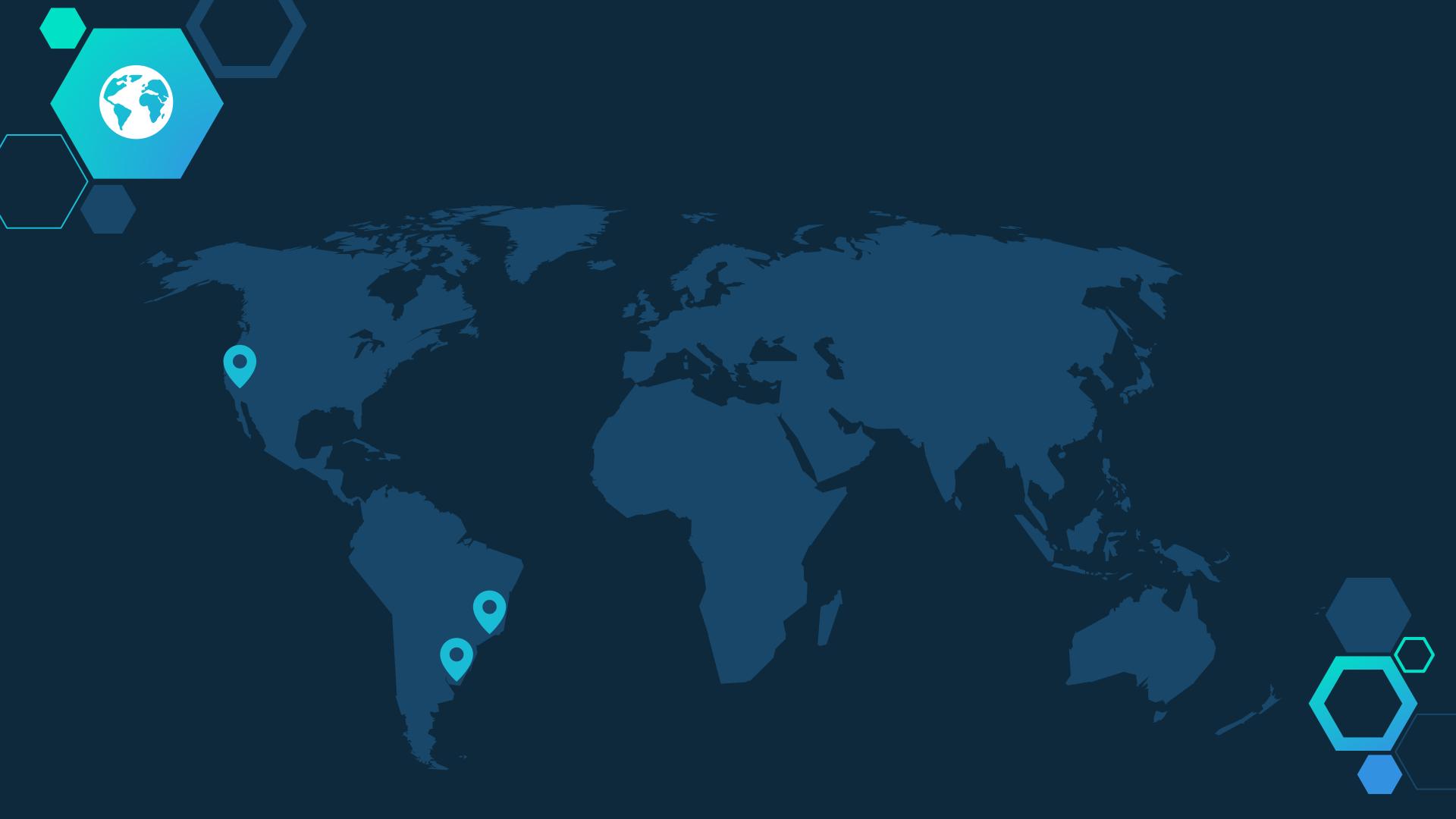
You can find me at

 @dcassol

 danielac@ucr.edu

 <https://www.dcassol.com/>







Open Science and Reproducibility

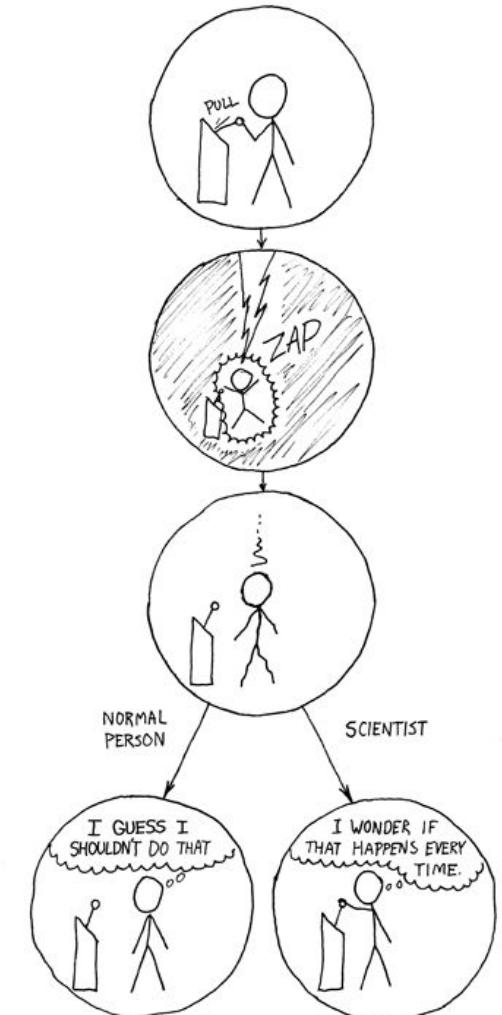


Reproducibility of Science...

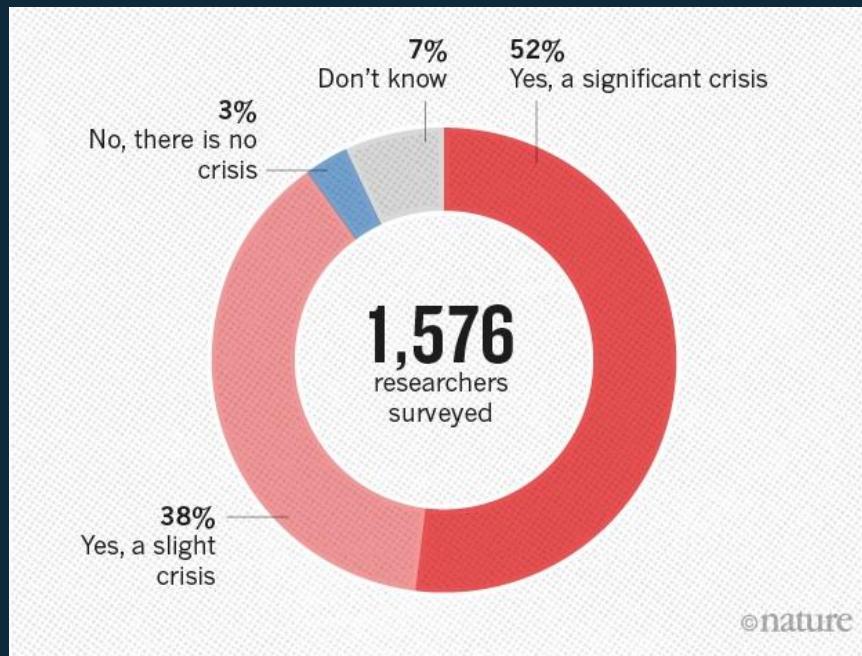
... A fundamental given of the
Scientific Method.

“An experiment is reproducible until
another laboratory tries to repeat it.”

Alexander Kohn

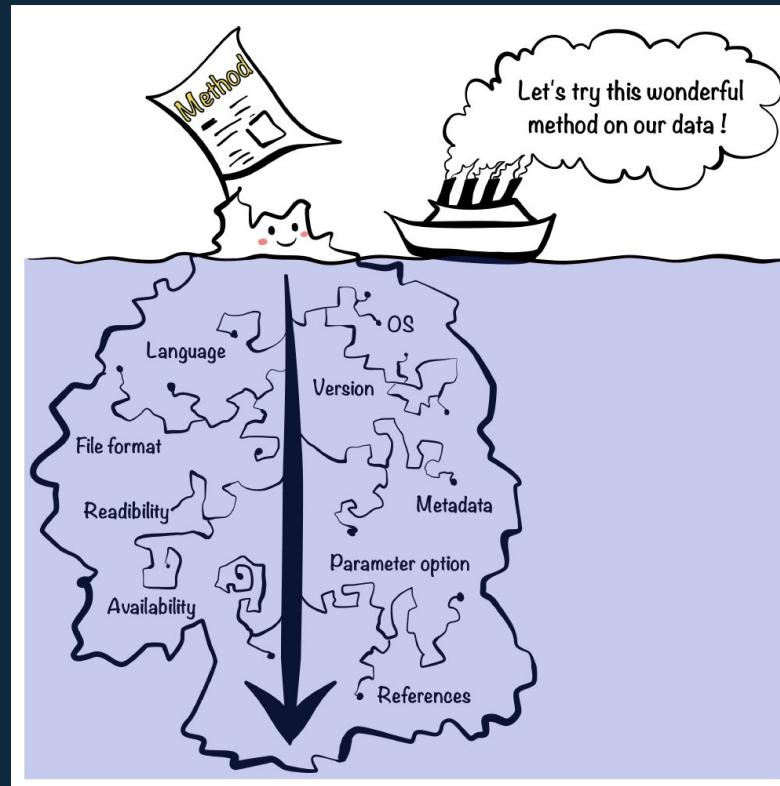


Is there a reproducibility crisis?



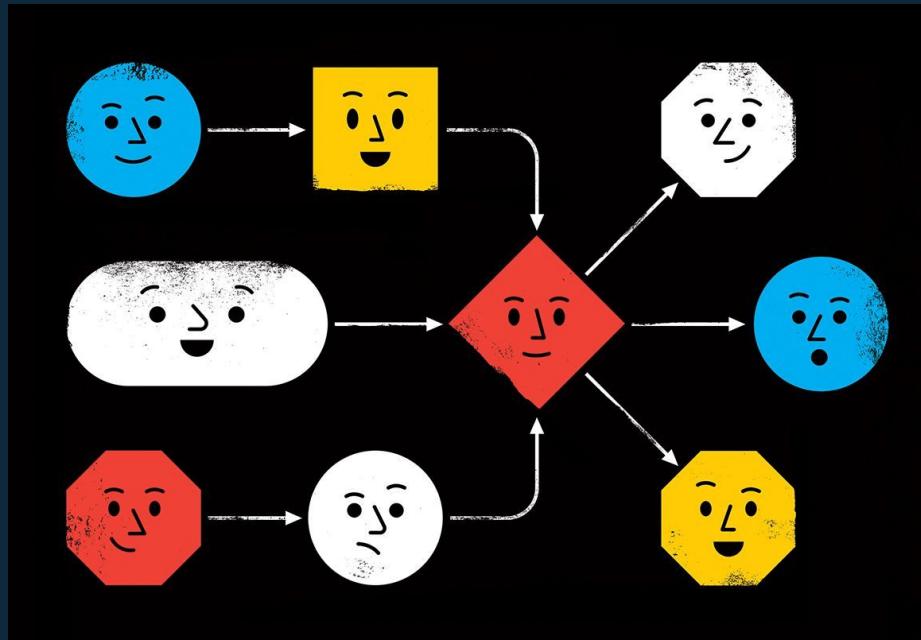


Reproducible Research in Computational Science





Workflows



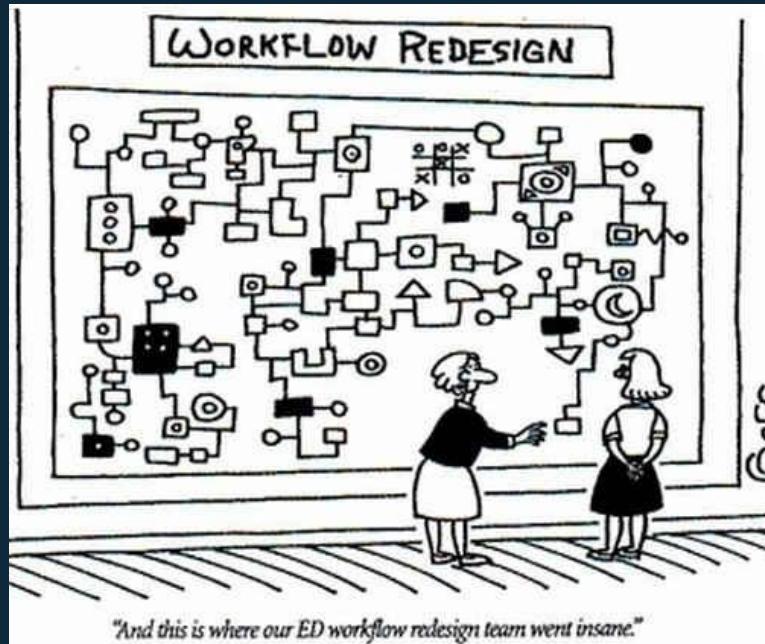


Step 1

Step 2

Step n

Workflows



Solution:



- ❖ End-to-end analysis pipelines
- ❖ Support for R and command-line software
- ❖ Single machines and compute clusters
- ❖ Uniform interface, sample handling and annotation
- ❖ Automated report generation
- ❖ Flexibility to support custom changes and new software/tools



Thanks!

You can find me at:

- ◊ @dcassol
- ◊ danielac@ucr.edu
- ◊ <https://www.dcassol.com/>
- ◊ 1207E Genomics Building



I <3 Ecological Data Wrangling and Making Pretty Figures

Teresa Bohner

September 30, 2019

Often I am happiest working when I am working in R

Skills and packages I use on a regular basis

- .. Data wrangling - **dplyr, tidyverse**
- .. Visualization - **ggplot2**
- .. Statistical modeling - **lme4, brms**
- .. Workflows and collaboration - **rmarkdown, shiny**
- .. Specialized things: GIS - **ggmap, raster, rgdal** Tree-rings
dplR
- .. You get the idea

Why data wrangling?

- .. Although everyone focuses on analysis and visualization, often the time you spend before those steps (data wrangling) is not trivial!
- .. I often work with data from multiple sources so things can get particularly messy

But who doesn't like making figures?!

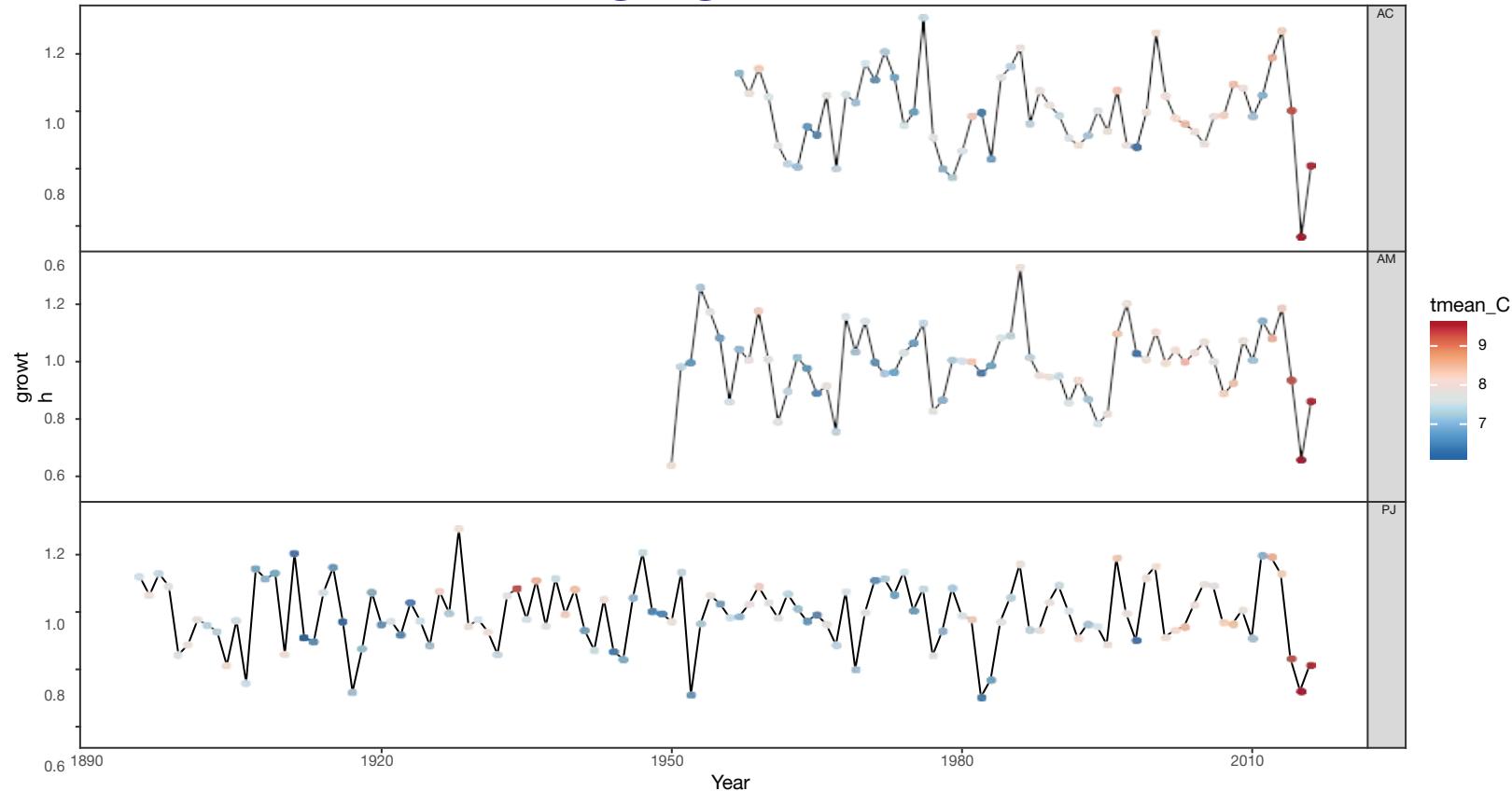


Figure 1: Here's some random tree-ring data

... And interactive figures?!

Drought Event Selection

How should we select drought events?

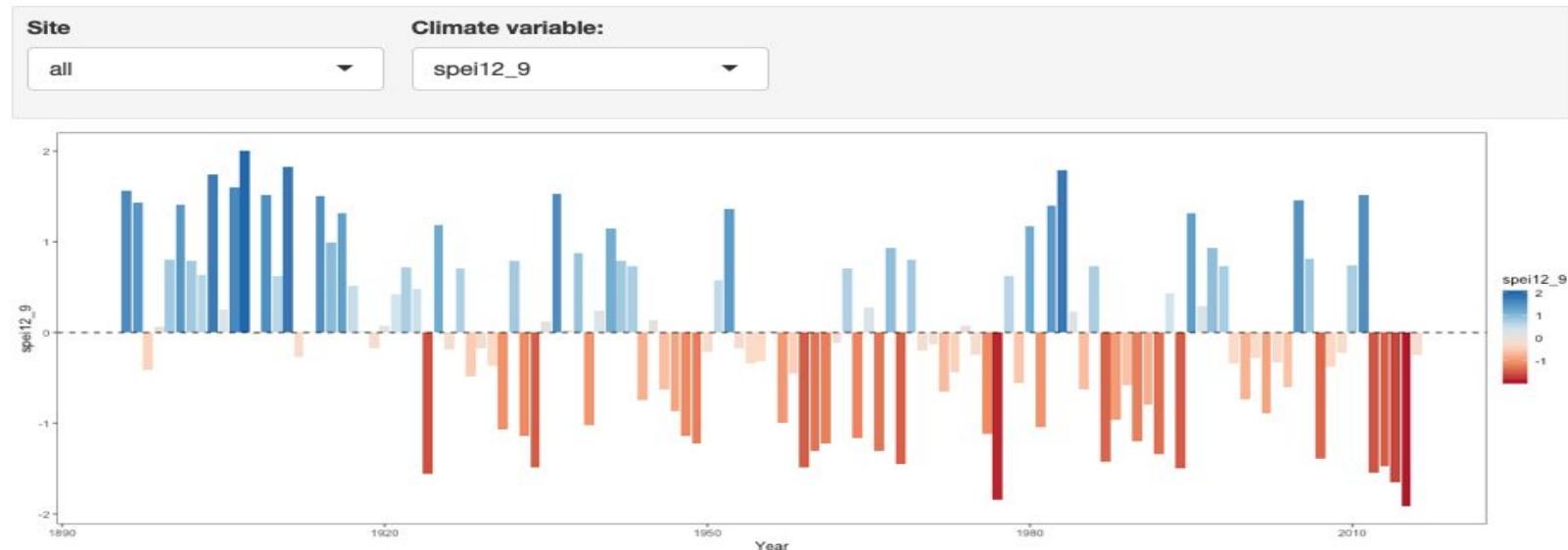


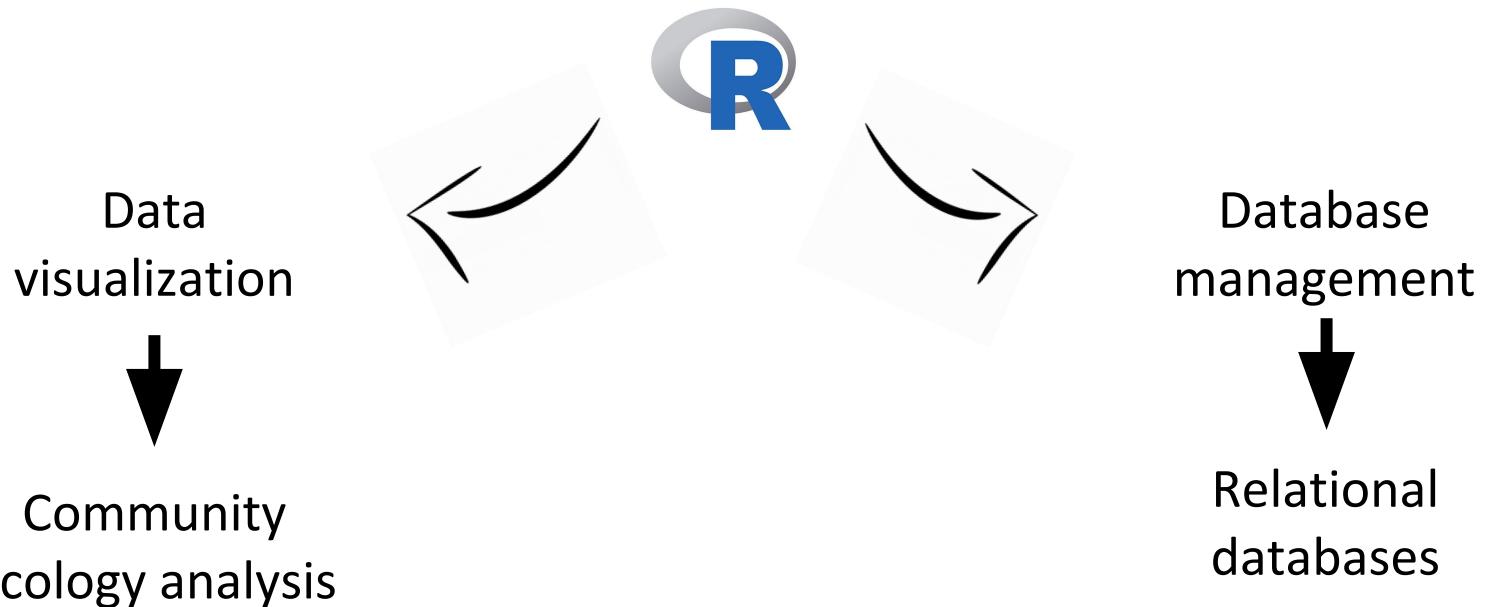
Figure 2: Another thing I was working on last week

In conclusion:

.. I am very excited to nerd out, *ahem *learn* with all of you!

How a field ecologist has used
R and how wants to use it now





- Vegan
- Ecodist
- Bipartite
- Igraph
- Ggplot

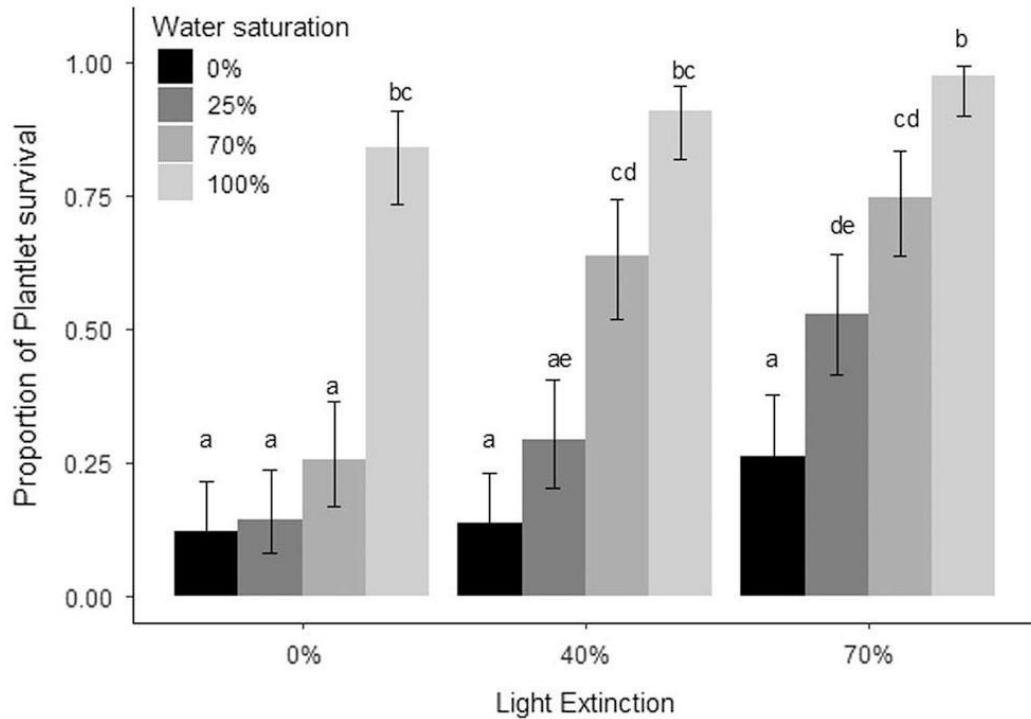
- RSQLite

Basic statistics: `glm()` function

What: Survival of an invasive plant under water and light stress.

How: 25 pots with three plantlets each per treatment. Survival of plantlets was recorded every 3 to 5 d for 60 days.

Analysis in R: Generalized linear model with binomial errors with water (4 levels) and light (3 levels) as factors.



`glm(formula, family=familytype(link=linkfunction), data=)`

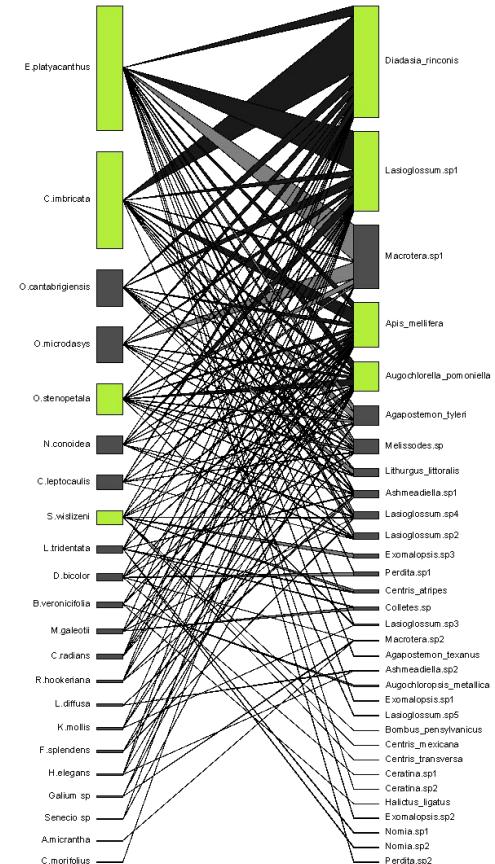
```
glm(survival~water*light,binomial(link="logit"),data=mydata)
```

Community ecology: Network analysis of plant-bee interactions in Central Mexico

What: Study the niche overlap among bee and plant species.

How: Field collection of bees and registration of their interactions with flowers in 2015.

Analysis in R: Creation of an interaction matrix and calculation of a dissimilarity index (Horn 1966) in **bipartite** package. Then, comparison of the niche overlap with random networks (keeping the number of interactions and species) to test for its significance.



I have learnt the little bit I know of stats by using
R...

Now, what I really need to learn is how to
work with metagenomes in R

Genome analysis

ASAR: visual analysis of metagenomes in R

Askarbek N. Orakov^{1,2,*†}, Nazgul K. Sakenova^{1,2,†}, Anatoly Sorokin^{3,4,†}
and Igor I. Goryanin^{1,5,6,*}

Citrus genetics and genomics: How R facilitates my research

Danelle Seymour
University of California, Riverside

Connecting genotype to phenotype is central to breeding

G
E
N
O
T
Y
P
E

ATGCTATCG
ATGCAATCG
ATGCTATCG
ATGCAATCG
ATGCTATCG
ATGCAATCG

P
H
E
N
O
T
Y
P
E



Image from: www.idtools.org

How R facilitates my research

1) Experimental design

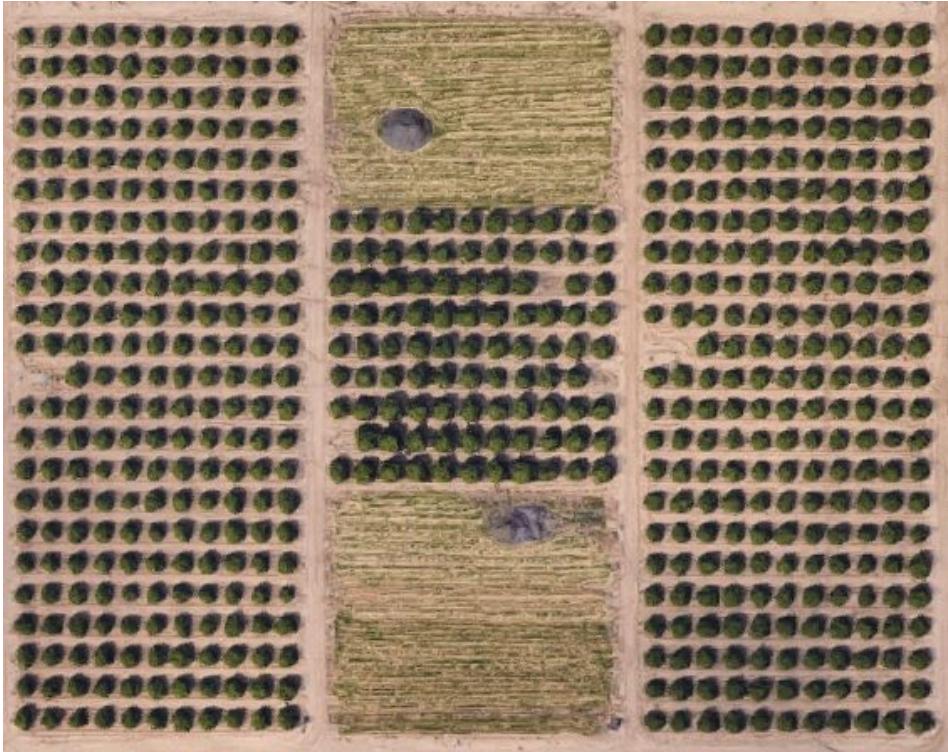
2) Data manipulation

3) Statistical analysis (w/ and w/o application-specific packages)

4) High-quality figure generation

Experimental design

- Randomization of samples in a specific experimental design.
- Sample size determination based on prior knowledge of phenotypic variance.
- Identification of environmental gradients



Data manipulation

- Manipulate lists of IDs (correct errors in directed manner).
- Analyze phenotypic distributions of traits.
- Sliding/neighboring window analyses of genomic data (list all methylated vs unmethylated sites).

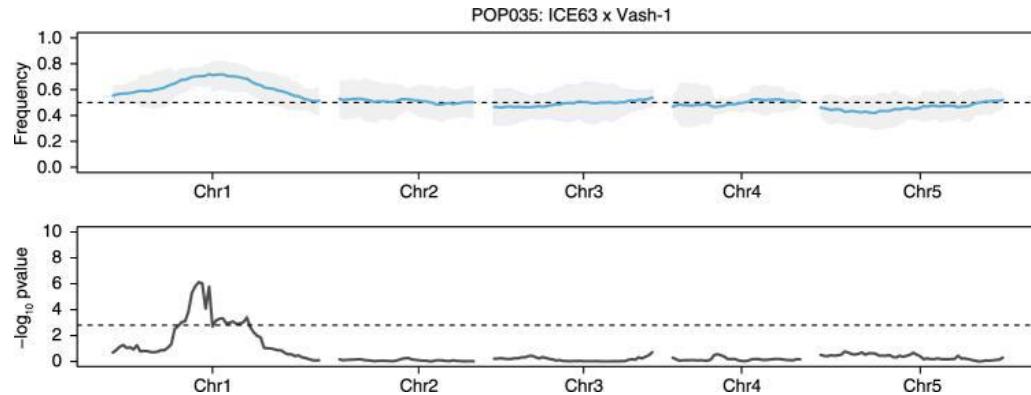


In collaboration with: Dan Chitwood, MSU
Tracy Kahn, UCR

Statistical analyses

Statistical analysis w/o specific package:

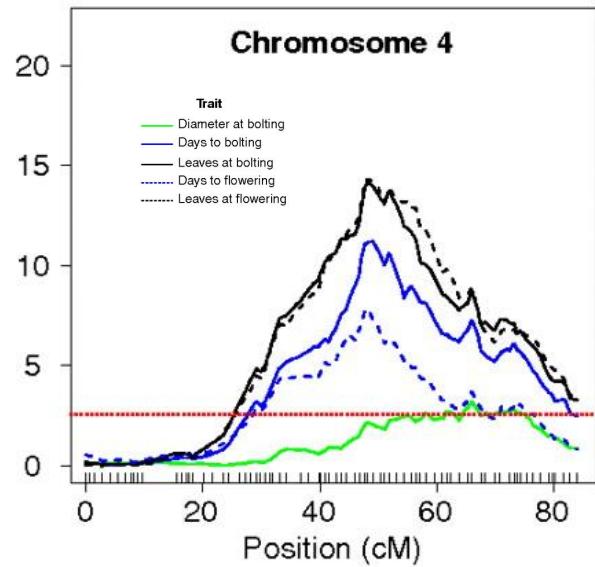
- Model allele frequency in F2 population bulks in windows across the genome.
- Test for significant deviation of allele frequency/



POP035: ICE63 x Vash-1

Statistical analysis w/ specific package:

- R/qtl, edgeR, DESeq2, snp2go, topGO, ...





A couple of questions before we go...



Join us for our next meeting: 11/20*** (Wednesday)

Thank you