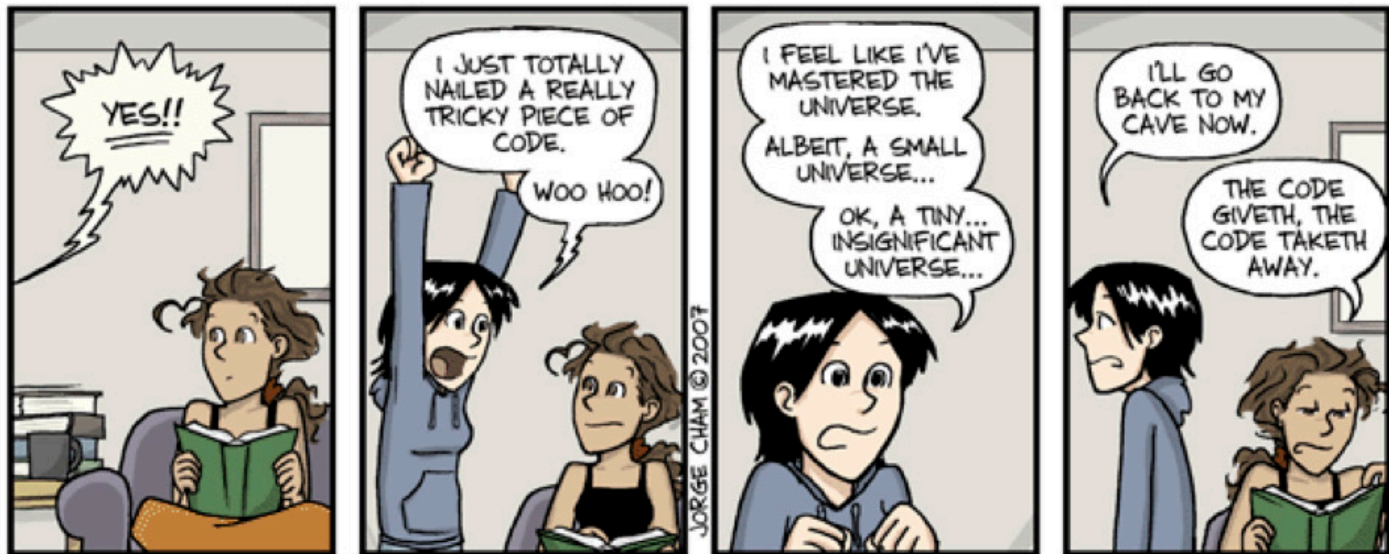


Welcome to BIOL 501: Quantitative Methods in Ecology and Evolution

Dr. Beth Volpov, b.volpov@oceans.ubc.ca; UBC 2022 Term 2

Piled Higher and Deeper by Jorge Cham

www.phdcomics.com



As you come in, please **make a name tent**,
and **write in BIG font**



Outline for today

- Meet your Instructor and TAs
- Course organization and schedule
 - Learning Objectives
- Health and well-being
- Why we use R
- Organizing data for analysis in R
- Review of foundational concepts in statistics
- Wrap-up
 - To Do List before 1st Workshop
 - Pre-course survey
 - Sign up for Discussion/Moderating (1 each)
 - First Discussion paper

Discussion

How do you feel about statistics and coding?

Go to Jamboard:

https://jamboard.google.com/d/1nZNk2e_nyinPRZ1TRfw632mcPQJBhYGyL4RTDjOhw1k/edit?usp=sharing

Instructions to add a sticky note (anonymous)

1. Click the 4th icon down on the left "sticky note".
2. Type in the sticky note, sentences, phrases
3. Click "SAVE" to submit

You can submit multiple sticky notes



BIOL 501 Teaching Team



Instructor:
Dr. Beth Volpov
b.volpov@oceans.ubc.ca



Instructor:
Avery Kruger
avery.kruger@botany.ubc.ca



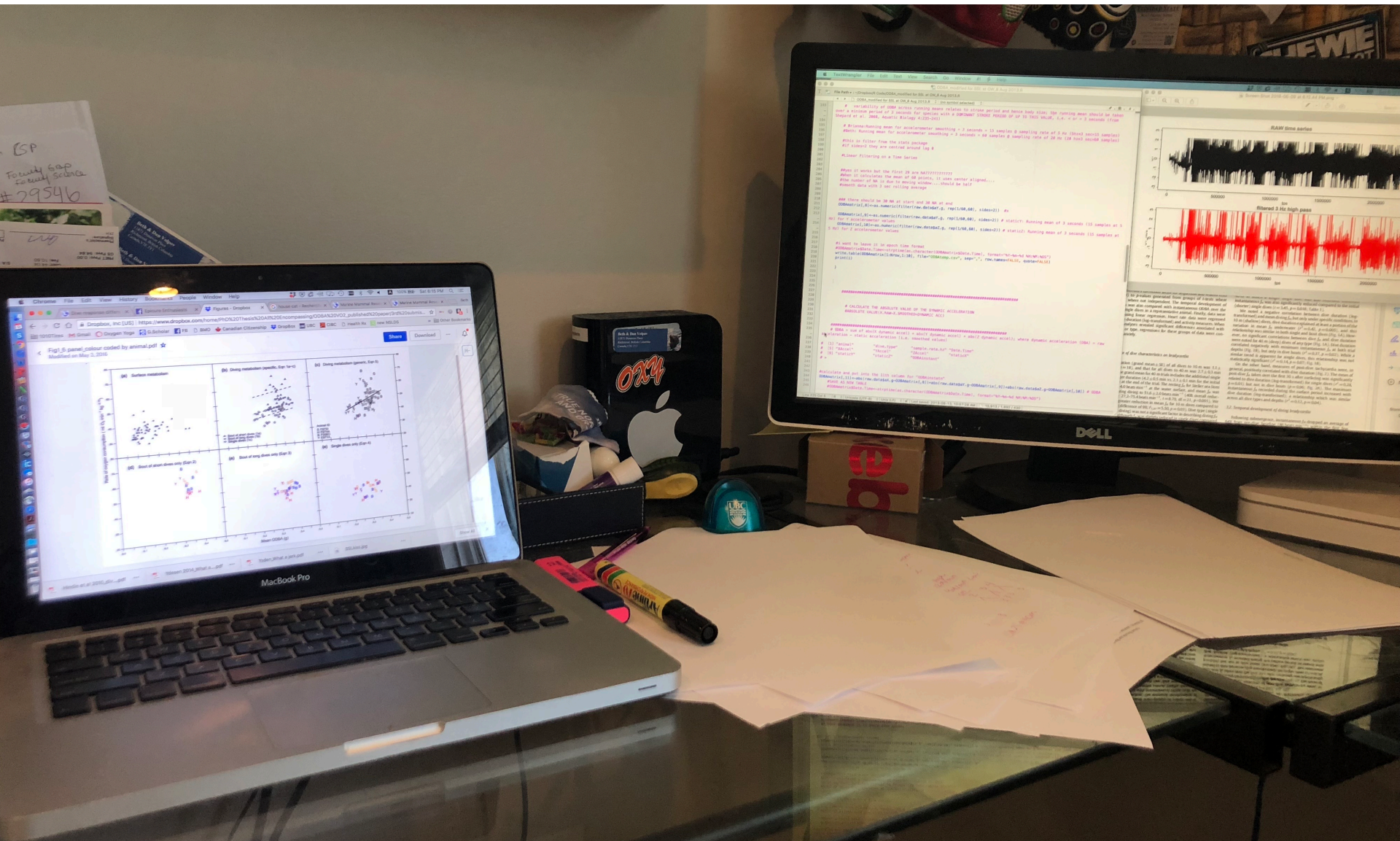
- **Research:** marine mammal physiology & ecology
- **Stats:** linear mixed-effects models, big data
- **Family:** 2 young kids



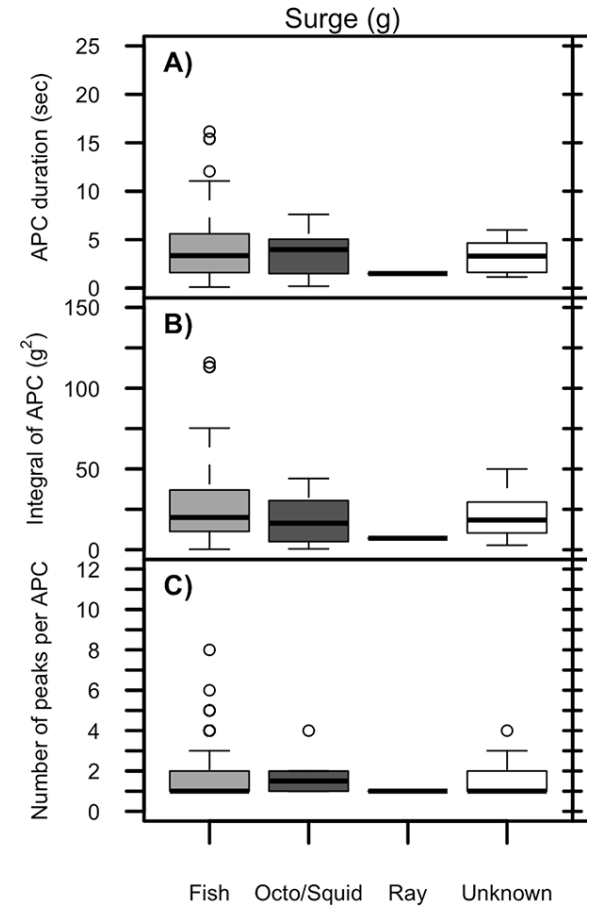
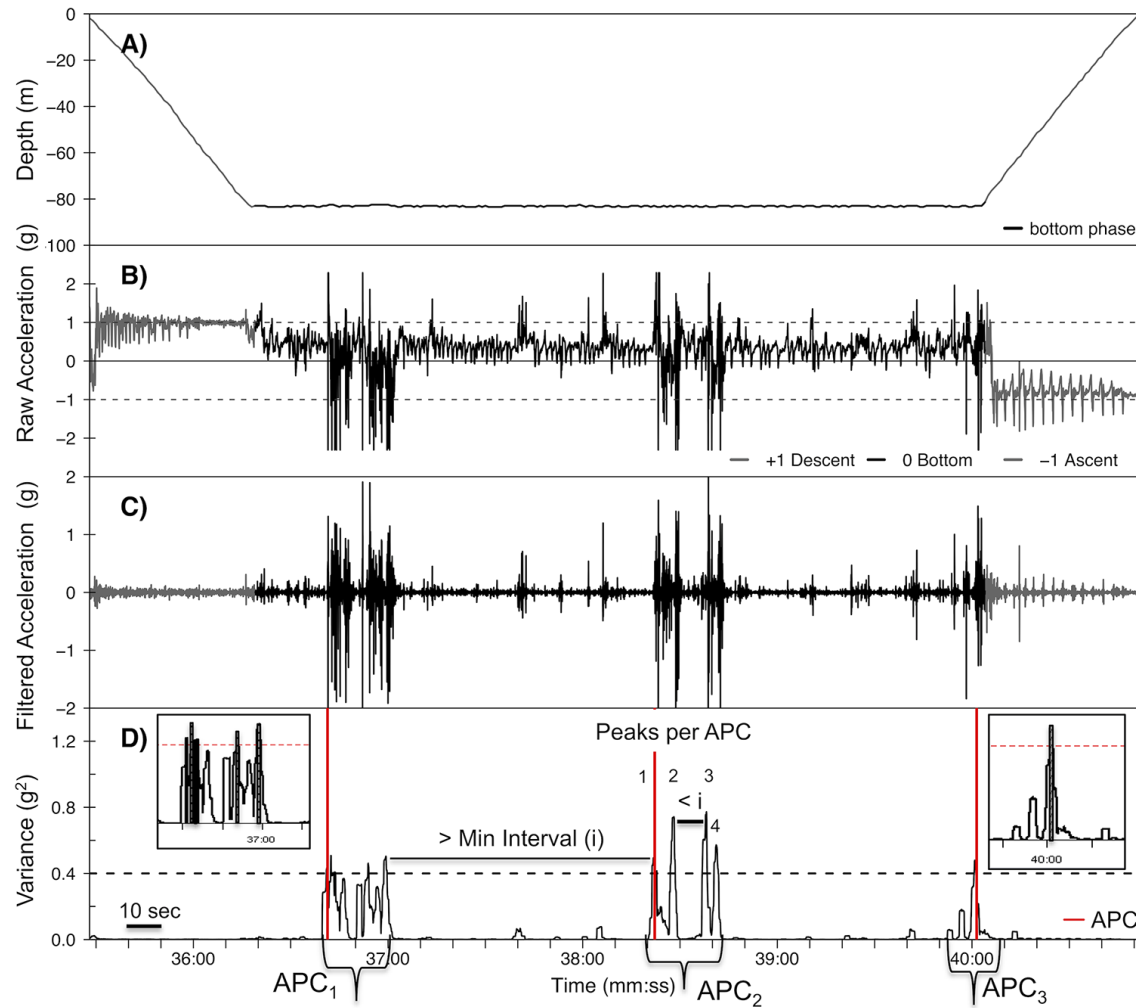
Critter Camera Video



... this is the majority of what a biologist does

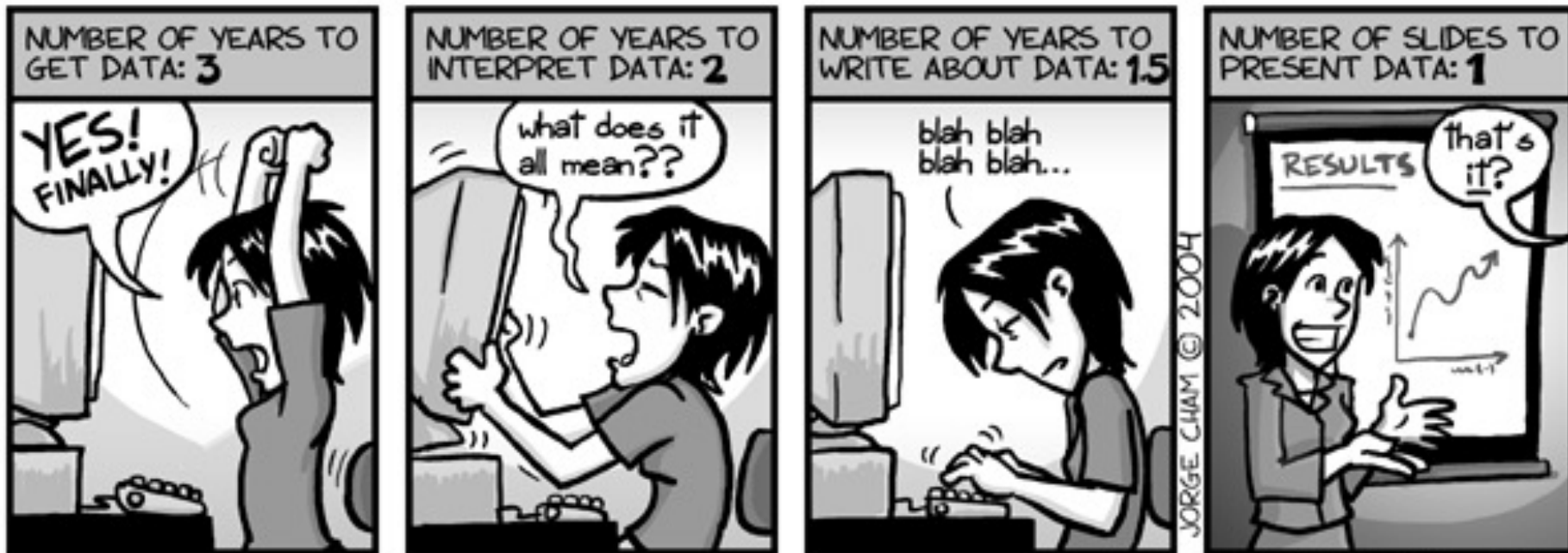


High-volume 3D acceleration data

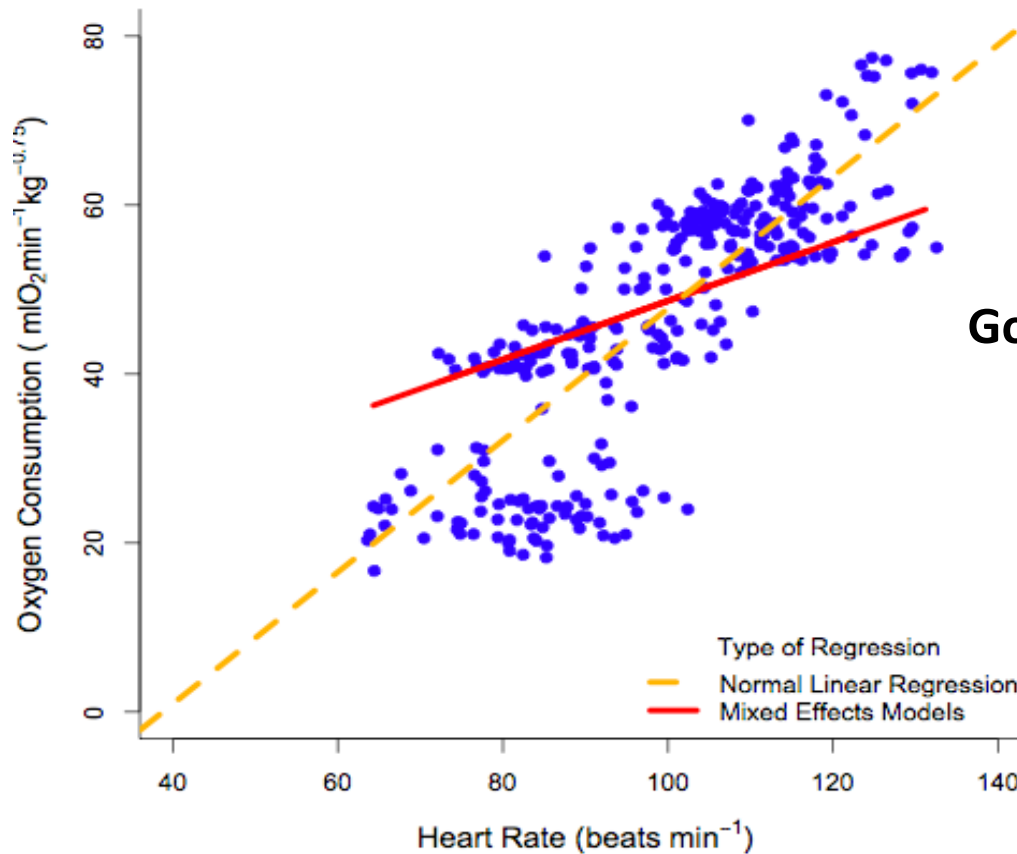


Millions of lines of data distilled into a single data point

DATA: BY THE NUMBERS



My first love in R: Linear Mixed-Effect Models



**Goal is to infer about
population NOT
individuals**

- Treat animal as random sample of population
- Different than pooling of normal linear regressions
- Model animal differences in variance, errors etc.
- > power of with small n

My R experience

- I'm not a statistician, but genuinely enjoy using stats to find patterns that our eyes can't see
- I learned on Base R before Rstudio existed
- I am not a complete R expert. I have a head start on most of you. Maybe I will hold this edge until the end of term.
- You will discover things that I don't know about. Please share these on Piazza/Slack and with me.

TA: Avery Kruger



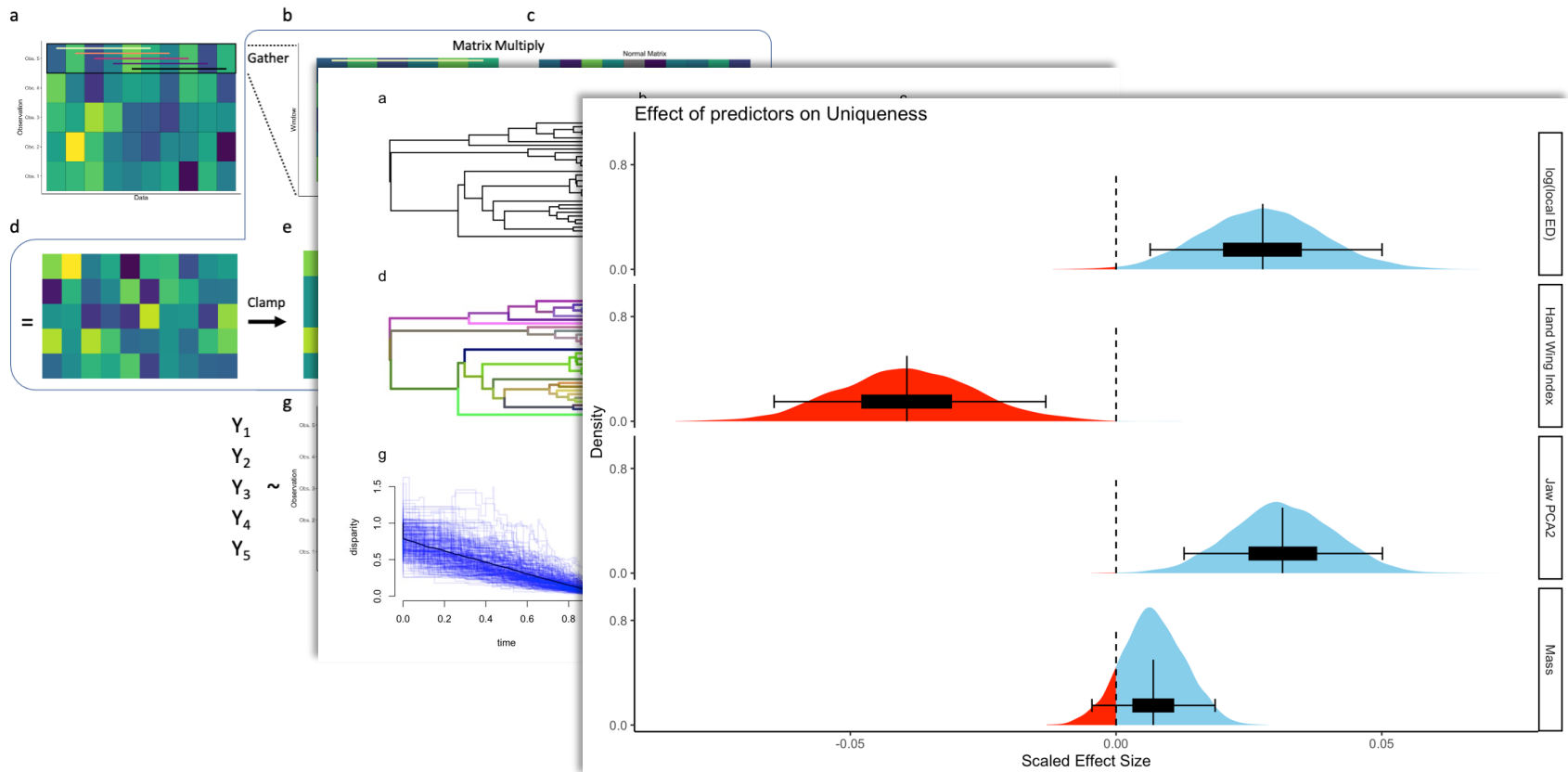
Research

- The relationship between phylogeny and ecology
- I do a lot of coding—and I like it!
- I'm very good at writing functions (**I've written a package for R!**)

This is what I used to do →



Some plots I've made



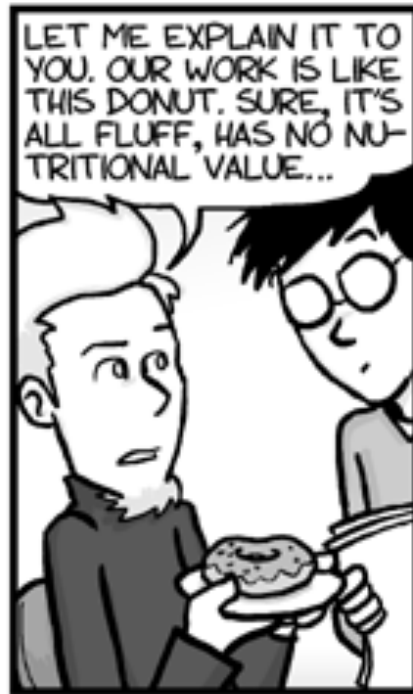


Tell us about you next.....

Pre-course survey posted on Canvas

- Who are you academically?
- Who are you outside of your thesis?
- Familiarity with R?
- Familiarity with stats theory?
- What are some potential types of data you want to collect for your thesis?
- Types of analysis you are interested in learning in R?
- **Student's choice last lecture**

Your thesis is like a donut



Outline for today

- About the course
 - Housekeeping
 - Course organization
 - Grades
 - Assignments
 - Master overview calendar

Waitlisted Students

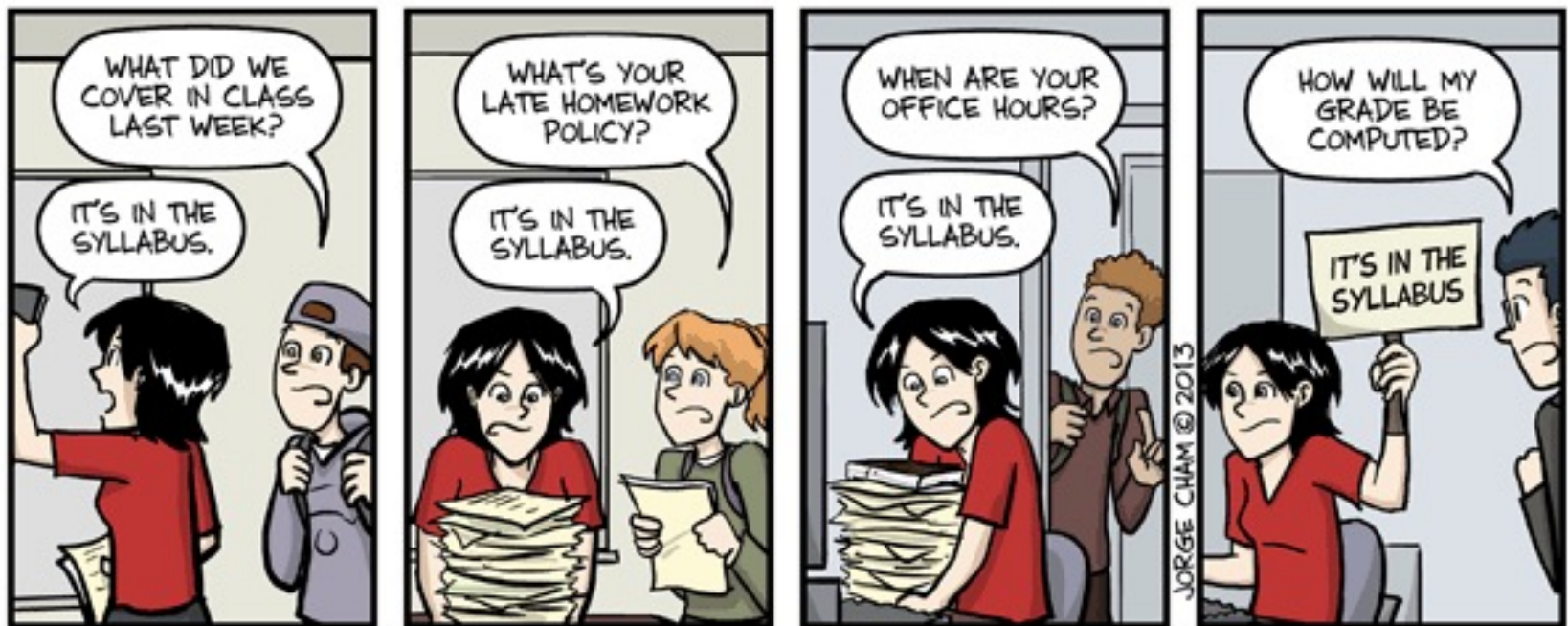
- It's not possible to access Canvas until you are registered
- I will email you when/if spot is available
- Get material from the external BIOL 501 webpage
- **Keep attending class to avoid getting behind**
- Yes, sign up for a Discussion/Moderation spot (> Jan 20)

Discussion (show of hands)

Do you prefer Slack or Piazza

- Optional student collaboration and discussion
- Monitored by TA if get stuck, but mostly student-led
- Not part of your grade
- Choose one but not both

Please see Canvas syllabus for details Today is just highlights



IT'S IN THE SYLLABUS

This message brought to you by every instructor that ever lived.

WWW.PHDCOMICS.COM

Course Background

- Course was developed in ~ 2009 by Dolph Schluter in response to needs identified by grad students to the BRC
- Help us improve it, I will be giving you surveys for feedback
- This is a “second” course in data analysis, to take you beyond the most basic, introductory level, which I’m assuming you have already done

Learning Objectives

- Prepare you for research by reviewing the basic principles for designing good studies, gathering and organizing data, and properly analyzing those data.
- Increase your understanding of concepts in data analysis
- Introduce you to innovative approaches used in biology to analyze data.
- Broad coverage of current methods, rather than deep foundation on few topics.

Course components

Discussion & Lecture (Tues)

- **1-2 pm Paper Discussion**
- (~ 5 min break)
- **2-3pm Lecture on same topic as Discussion**

Workshops (Thurs)

1-3pm Thursdays

Self-paced

No Lecture Content

Don't have to turn in



Ideally, start the workshop on Wednesday or Thursday morning

Lecture and Discussion Days (Tues)

- All students read the paper and participate in Discussion
- **Presentation of paper (20-25 min)**
 - 2 students present summary of paper
 - Analyze the topic, use reading handout as starting point
 - Explain key points, figures, additional opinions
- **Discussion of paper (~ 30 min)**
 - 2 students moderate the discussion
 - Whole class participation
- (~5 min break)
- **Lecture on same topic (50 min)**

Each person please sign up for 1 Discussion and 1 Moderation spot on the Google Doc posted on Canvas Home Page before Jan 16

*Waitlisted students sign up for dates > Jan 20th

Workshop Days (Thurs)

- First workshop is this Thursday on Introduction into R.
- Use own computers. Have latest R version installed (4.1.2).
- Work through problems on the web site during the workshop.
- **HINT:**The “R tips” web pages contain most clues needed to carry out workshops.
- 2 hours is allocated, but may take longer. There is no specified portion of workshops that must be completed. The further you go, the more you will learn. You control that.
- Go back and get more practice as needed.

You don't have to turn in anything from the workshops.

To do before workshop this Thurs

- Have latest R version installed (4.1.2) (Rstudio preferred)
- Set your R aesthetics and window panes as you prefer
 - Rstudio/Preferences/Appearance
 - Rstudio/Preferences/Pane Layout
- Mac Users also need XQuartz for some R commands
- Set up "your system" of how you will organize R scripts, figs, tables etc
- See R Tips page: Calculate and Data sections will be useful for this week

Have Script started for Workshop 1 and the datasets downloaded and imported prior to workshop

Creating an inclusive and respectful classroom is important

- **All** voices are important in lecture and discussions
- We are all at different points in learning R and have different strengths
- We work together to create a safe inclusive learning space by brainstorming guidelines for communication, participation, and discussions
- Examples: raise hand to talk, don't interrupt

Discussion

What are some specific group guidelines for us to create a respectful, inclusive space in lecture and discussions where *all* voices have input?

Go to SAME Jamboard as prior (click next slide >):

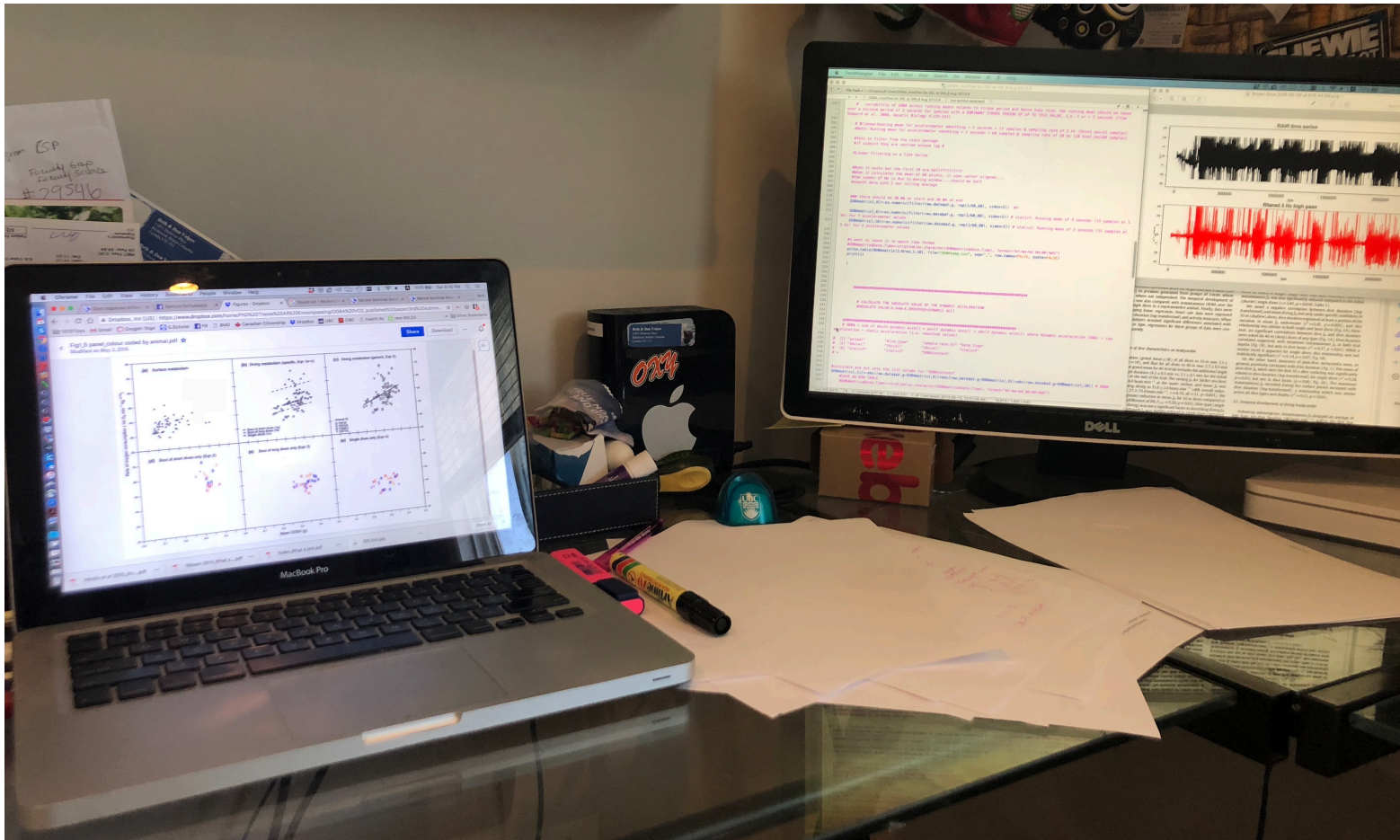
https://jamboard.google.com/d/1nZnk2e_nyinPRZ1TRfw632mcPQJBhYGyL4RTDjOhw1k/edit?usp=sharing

- Please add at least 2 specific contributions
- You can add more than 2
- We will review as a group prior to the first discussion



Useful Tips for Coding

- Get 2 monitors in your office *and* at home
- Vertical monitors are amazing for coding



Where to find data to learn R and do assignments?

- Ask your lab members for their datasets that are published
- Many online journals require database submission (QA/QC is variable)
- Manually enter data from published papers figs and tables
- No copyright on published data (it's ok to use for R practice or meta-analysis)
- Graphics tool: <https://www.datathief.org/>
- Online data archives, e.g., Ecological Archives, <https://esapubs.org/archive/>, Genbank, Dryad (<https://datadryad.org/search>)
- Permissions/conditions may be required to publish results from archives.
- **I am happy to share my published MSc and PhD datasets**

1. Master Course Calendar
2. Tour BIOL 501 External webpage before Thursday
3. Subscribe to Google Calendar

BIOL 501. January 9th-April 13th 2022

Week #	Day	Date	Description	Discussion Paper and Handout (Tuesdays)	Presentors	Moderators	Homework Assignment
1	Tues	Jan 10th	Lecture: Introduction	NA	NA	NA	
	Thurs	Jan 12th	Workshop: Introduction into R				
2	Tues	Jan 17th	Lecture: Graphics	How to display data badly (Wainer 1984)	1 _____	1 _____	#1. Improve a Graph
	Thurs	Jan 19th	Workshop: Graphics		2 _____	2 _____	
3	Tues	Jan 24th	Lecture: Experimental Design	Pseudoreplication and design of experiments (Hurlbert 1984)	1 _____	1 _____	
	Thurs	Jan 26th	Workshop: Plan Experiments		2 _____	2 _____	
4	Tues	Jan 31st	Lecture: Linear models	Confidence Intervals (Colegrave & Ruxton 2003)	1 _____	1 _____	
	Thurs	Feb 2nd	Workshop: Linear Models		2 _____	2 _____	
5	Tues	Feb 7th	Lecture: Mixed-effects models	Regression towards the mean (Kelly & Price 2005)	1 _____	1 _____	#2. Analyze a linear model
	Thurs	Feb 9th	Workshop: Mixed-effects models		2 _____	2 _____	
6	Tues	Feb 14th	Lecture: Likelihood	Simplicity in data analysis (Murtaugh 2007)	1 _____	1 _____	
	Thurs	Feb 16th	Workshop: Likelihood		2 _____	2 _____	
Midterm Break Feb 20-24th (no classes)							
7	Tues	Feb 28th	Lecture: GLMs	Controlling false discovery rate (Verhoeven 2005)	1 _____	1 _____	
	Thurs	March 2nd	Workshop: GLMS		2 _____	2 _____	
8	Tues	March 7th	Lecture: Model Selection	Why do we still use stepwise regression? (Whittingham et al 2006)	1 _____	1 _____	
	Thurs	March 9th	Workshop: Model Selection		2 _____	2 _____	
9	Tues	March 14th	Lecture: Bayesian Methods (guest lecture Evan Sidrow)	The earth is round (Cohen 1994)	1 _____	1 _____	
	Thurs	March 16th	Workshop: Bayesian inference		2 _____	2 _____	
10	Tues	March 21st	Lecture: Computer Intensive (Bootstrapping & resampling)	Working with multiple hypothesis (Dochtermann & Jenkins 2011)	1 _____	1 _____	
	Thurs	March 23rd	Workshop: Bootstrap and randomization		2 _____	2 _____	
11	Tues	March 28th	Lecture: Meta Analysis	Publication bias in fluctuating asymmetry (Palmer 1999)	1 _____	1 _____	

External webpage vs Canvas

- Primary webpage is external to Canvas
- www.zoology.ubc.ca/~bio501/R/
- **Updated regularly** – hit your browser refresh button before starting assignment/workshop.
- Assignments are on both, but **only submitted via Canvas**
- The “R tips” help pages.
- The *Calculate* and *Data* R tips pages will be useful for 1st week’s workshop
- Sync Google calendar to your phone to get alerts

Course components

- Zoom is emergency backup only if instructor is sick or UBC closed
- **Books:** No textbook, but various sources, many available online.
See www.zoology.ubc.ca/~bio501/R/

How to get help in R

- Try solving the workshops by yourself first.
- If it is not working, ask your neighbor.
- Ask on Piazza/Slack
- Then ask instructor or TA.
- Admit if you are stuck. You are not alone.

Grading Scheme

	Percent (%)	Description
All Assignments	50	
Assignment #1	15	Improve a graph
Assignment #2	25	Analyze a linear model
Assignment #3	10	Data Manipulation
Presenting a talk on paper	15	
Moderating a discussion on paper	15	
Overall participation in discussions	15	
Feedback, surveys, & engagement	5	
TOTAL	100	

No final exam, but last assignment is due April 16th.

- Try Rmarkdown but not required
- Workshops are not turned in or for participation credit
- Rubrics for each assignment provided (clear expectations)
- Presenting and Moderating: combination of peer-evaluation plus Instructor, TA evaluation

Email Policy for Instructor, TA, Students

- Yes, we all have smartphones, but an instantaneous response is not realistic
- We will try our best to respond to course emails **within 48 hours.**
- Weekend/holidays response time 48-72 hours.
- Include **“BIOL 501 Student:” at the start of your subject line for all emails.**

Office Hours



JORGE CHAM © THE STANFORD DAILY

- We are here to help you learn!
- Come prepared with questions
- Attempt the Workshop/Assignment before asking
- **We won't just paste in the code for you.**
- Try posting on Piazza/Slack also for peer feedback

Well-Being and Health Are Important

- Don't come to class if you are sick
- BIOL 501 doesn't exist in isolation
- Good-posture/desk setup while coding is important
- Come talk to us if you have any concerns, 100% confidential

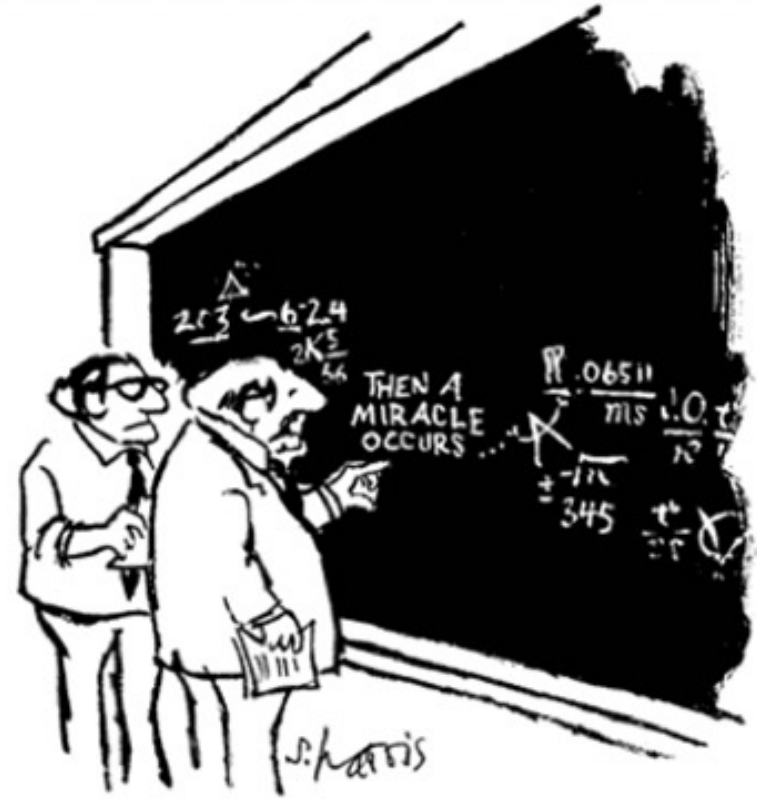
Work-life balance is important. Enjoy the sunsets & what makes you happy.



Teaching Philosophy

Active learning requires your effort.

- Instructors or TA's won't do it for you.
- Using code you don't understand is ultimately useless



"I think you should be more explicit here in step two."

Less guidance is given in a grad course than an undergrad course.

What is R?



- R is a language and environment for statistical computing and graphics.
- Free, collaborative
- The current R is the result of a collaborative effort involving contributors from all over the world.

Discussion

Brainstorm “Good” and “Bad” things about using R

Go to *SAME* Jamboard as before (click next slide >):

https://jamboard.google.com/d/1nZNk2e_nyinPRZ1TRfw632mcPQJBhYGyL4RTDjOhw1k/edit?usp=sharing

Use a **green sticky note** for "Good"
and **pink sticky note** for "Bad"



How will you organize and standardize R Scripts and Data?

Think about **long-term organization and storage** of R scripts, data, outputs, figures

- Your computer may change during grad school and afterwards
- I have had 3 different laptops since my MSc and can still find and run R code from 2008
- **How?** organization of files and **lots of annotations in scripts**
- It can be years between data analysis and journal revisions (redo analysis, edit figures)

How will you organize and standardize R Scripts and Data?

Think about **long-term organization and storage** of R scripts, data, outputs, figures

- For your thesis, I'd recommend organizing by data chapters which is equivalent to published papers.
 - Sometimes 1 pre-processing script (ie 1 script per animal)
 - 1 Main analysis script with full figures in it (stats, models on full dataset)
 - 1 separate script for only final figures (at the end)

How will you organize and standardize R Scripts and Data?

- I use **dropbox** to store data and run R scripts
 - Import/export data from dropbox directly into R with file pathways and workind directories
 - Can access it anywhere (i.e. field work, home, University)
 - Large volume data doesn't slow down my laptop
 - Easy for collaborations (share/remove users as collaborations change)
 - Easily searchable for specific code
- Github options also

Outline for today

- Review of foundational concepts in statistics
- Wrap-up
 - To Do List before 1st Workshop
 - Pre-course survey
 - Sign up for Discussion/Moderating (1 each)
 - First Discussion paper

On scrap paper in groups of 2-3, answer these questions (1-2 word answers)

1. What is the definition of probability?
2. The chance difference between an estimate from a random sample and the population parameter being estimated is called _____.
3. A systematic discrepancy between estimates of a parameter from a sample and the true value of the population parameter is called _____.
4. The probability distribution of values of a sample estimate that we *might* obtain when we sample a population, and their probability of occurrence, is called the _____.
5. The standard deviation of the probability distribution for a sample estimate is called the _____.
6. A range of values surrounding the sample estimate that is likely to contain the true value of the population parameter is called a _____.
7. The probability of obtaining a discrepancy from the null hypothesis as extreme as that observed, if the null hypothesis were true, is called the _____.

Self-reflection

- **How did you do so far?**
- If you need, review these core concepts in your stats book (e.g, Chapters 4 and 6 of Whitlock and Schluter (2015)).
 - Estimation vs hypothesis testing
 - Sample estimate (“effect size”) vs population parameter
 - Probability
 - Sampling distribution
 - Standard error
 - Confidence interval
 - Effect size
 - P -value

First discussion paper Tues Jan 17th

- Wainer (1984) How to display data badly.
- Need two presenters for next week: 20-25 minute presentation
- Need two discussion moderators
- Evaluated for participation by TA, Instructor, and peers
- See Canvas and talk to me after class for guidelines and tips on presenting and moderating

To do before workshop this Thurs

- Have latest R version installed (4.1.2) (Rstudio preferred)
- Set your R aesthetics and window panes as you prefer
 - Rstudio/Preferences/Appearance
 - Rstudio/Preferences/Pane Layout
- Mac Users also need XQuartz for some R commands
- Set up "your system" of how you will organize R scripts, figs, tables etc
- I use dropbox to store data and run R scripts
 - You can import/export data from dropbox directly into R with file pathways
 - Can access it anywhere
 - Large volume data doesn't slow down my laptop
 - Easy for collaborations (share/remove users as collaborations change)
 - Easily searchable

Have Script started for Workshop 1 and the datasets downloaded and imported prior to workshop