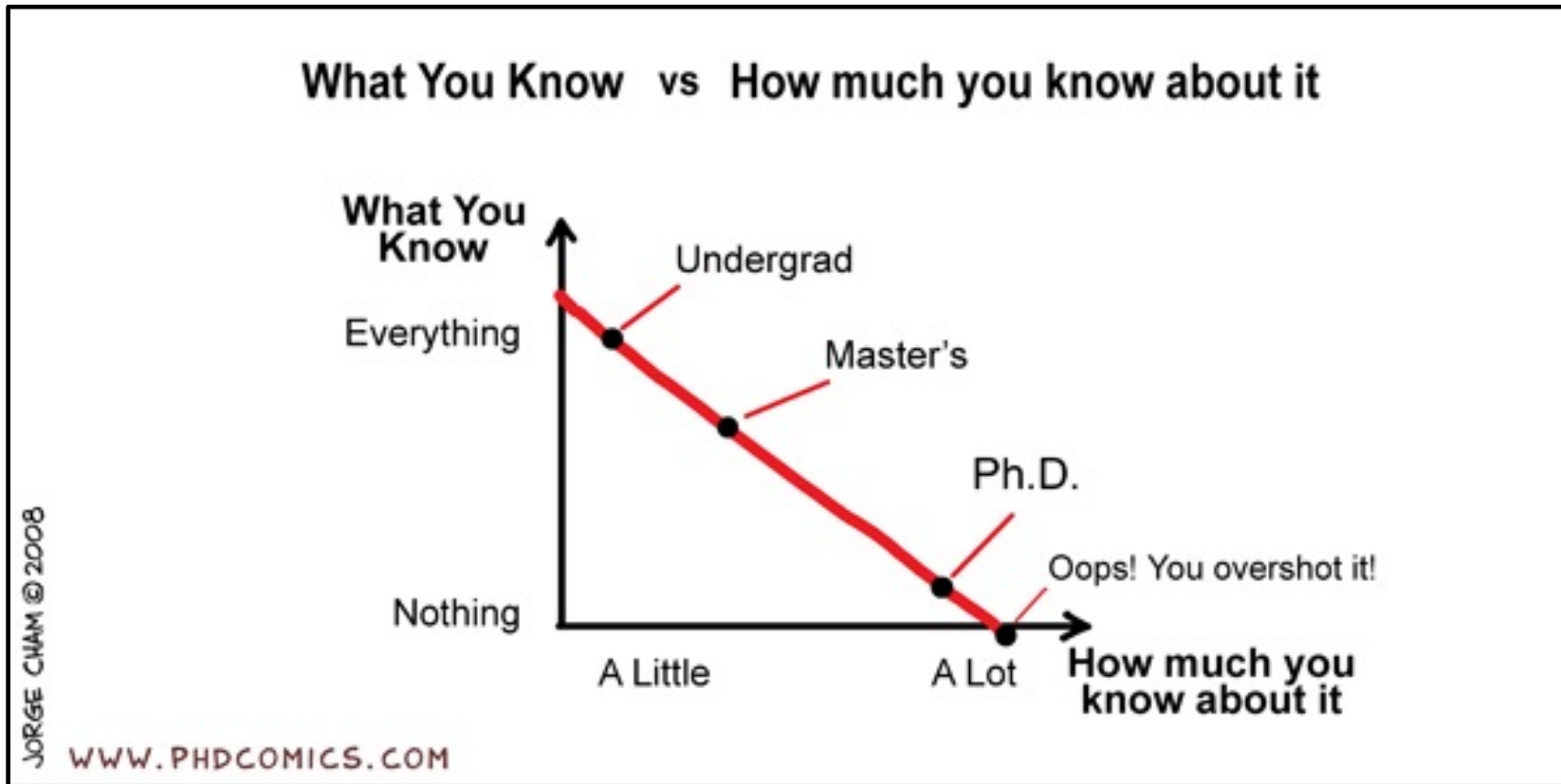# Peer-Feedback Survey
1 survey every Tuesday (constant link)

https://ubc.ca1.qualtrics.com/jfe/form/SV_bvLa6xMXdRk1j6u

BIOL 501: Graphics

Get me out, or make a new one →

FIRST NAME
Preferred pronouns

Outline

- Peer-feedback survey on Discussion Moderation and Presenting
- Why make graphs?
- 4 Principles of effective display
- Types of graphs to achieve these principles
- Why some graphs fail, and what can be done
- Tables
- Interactive Graphs
- Ggplot
- Motion (video)
- Wrap-up

Discussion

What kind of graphs do you have experience making or do you think you will make in your thesis?

Go to Jamboard:

https://jamboard.google.com/d/103Uolikn1ZhLSpJ5ururRoUIRr1JCs sqoGb3ibRPfko/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

Recommendations for data entry

Gastel, B., & Day, R. A. (2016). *How to write and publish a scientific paper* (Eighth ed.). Greenwood, an imprint of ABC-CLIO, LLC.

Best practices for data entry and management
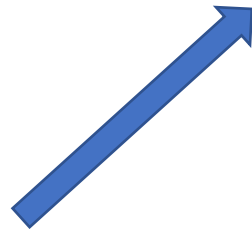
Bring the.csv of your thesis dataset to a workshop, I'm happy to look at the headers, and layout of it *before* you import it into R

A common mistake in organizing databases for categorical variables

- **Question:** Can heart rate predict metabolic rate and does this relationship differ between single dives vs dive bouts?
- X=heart rate (numerical)
- Y=metabolic rate (numerical)
- Treatments in this LME model are fixed effects (categorical variable)
  - Single dive
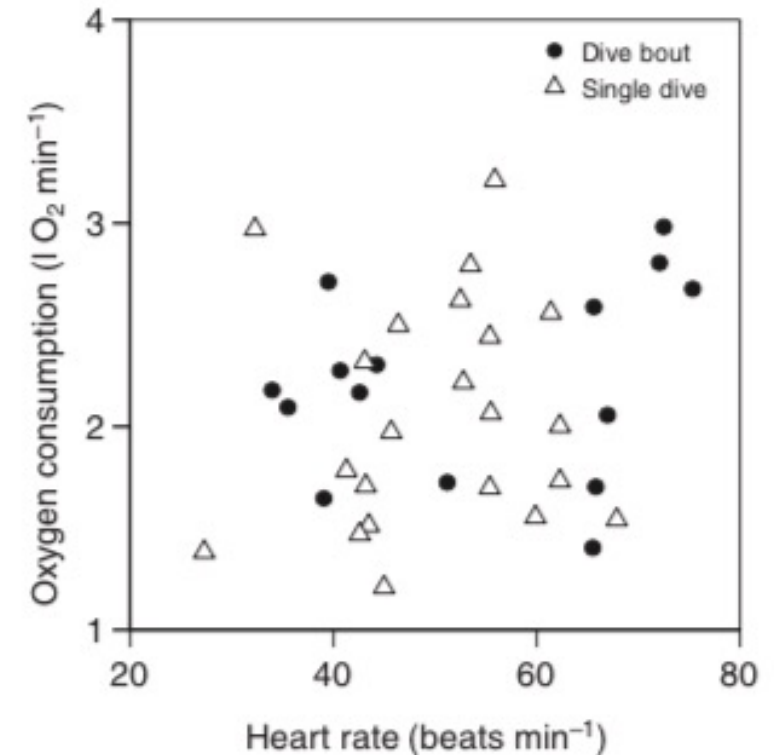 - Dive bout
- Animal is a random effect

Fig. 2. The relationships between diving f_H and DMR for Steller sea lions over a dive bout or over a single dive were not significantly linear.

How would you assess this database?
Is it "R friendly"?

|	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Animal	Day	Month	Year	Depth	Body Mass	Rep	Single Dive	Dive Bout	Dive.Duration	tmrs	tmrs.bin	total.time	Heart Rate Beats/Min	Average Metabolic Rate L O2/Min
2	Boni	7	aug	2008	10	142	1	Single		3.95	4.39	4 plus	8.85	53.24	1.801132156
3	Boni	14	aug	2008	10	144.5	2	Single		4.85	3.41	2 to 4	8.91	52.15	1.664475516
4	Boni	16	sep	2008	40	142	3	Single		7.44	2.26	2 to 4	10.72	85.47	2.037451206
5	Boni	19	aug	2008	40	155.5	1	Single		6.87	2.01	2 to 4	6.87	47.34	1.587817682
6	Hazy	5	aug	2008	10	171.5	1	Single		2.64	4.48	4 plus	7.84	97.23	2.091133827
7	Hazy	7	aug	2008	10	172	2	Single		3.69	2.25	2 to 4	6.35	89.81	2.328101564
8	Hazy	14	aug	2008	10	173	3	Single	Dive Bout	4.39	2.85	2 to 4	7.68	78.64	2.055794036
9	Hazy	4	sep	2008	40	171.5	3		Dive Bout	5.91	3.35	2 to 4	10.34	68.78	1.960670476
10	Sitka	1	aug	2008	10	223.5	1		Dive Bout	4.73	5.29	4 plus	10.44	58.07	1.792121579
11	Sitka	7	aug	2008	10	223	2		Dive Bout	4.2	0.45	less than 1	10.44	53.06	1.727073004
12	Sitka	4	sep	2008	40	215	1		Dive Bout	6.48	5.33	4 plus	12.97	45.92	2.268856079
13															

Outline

- Peer-feedback survey on Discussion Moderation and Presenting
- Why make graphs?
- Principles of effective display
- Types of graphs to achieve these principles
- Why some graphs fail, and what can be done
- Tables
- Interactive Graphs
- Ggplot
- Motion
- Wrap-up

Why make graphs?

- Because they can change the world!

- The human eye is a pattern detector.

- Enable visual comparisons of measurements between groups and expose relationships betwe variables.

- Best method for communicating results to a wid audience

Goal: show a pattern clearly & quickly
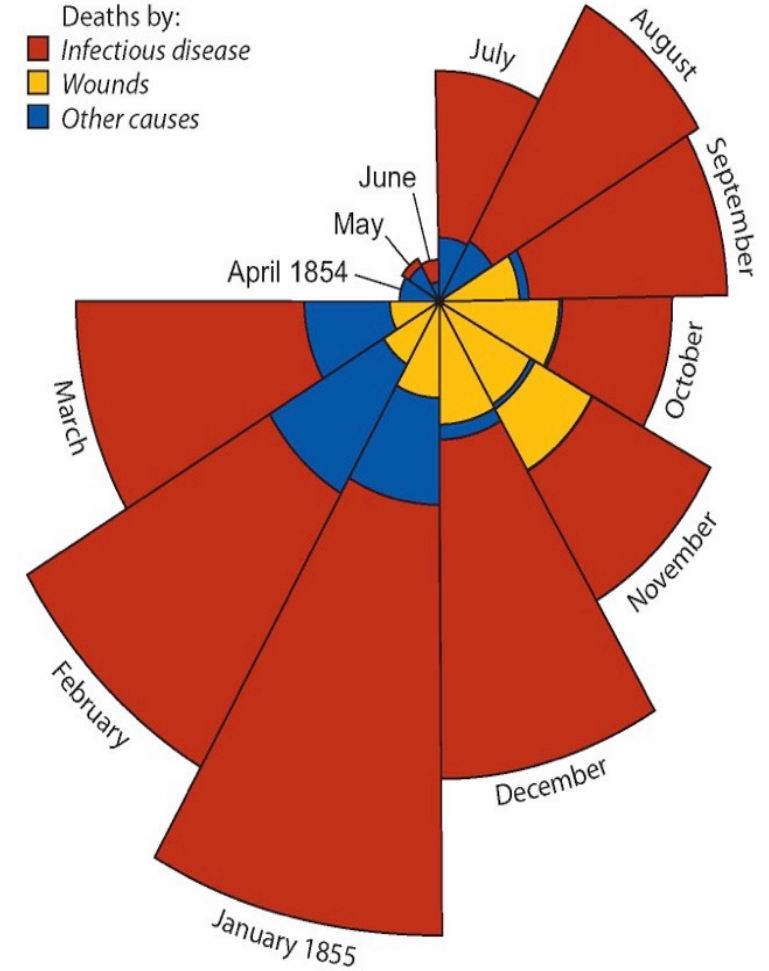
Causes of deaths in the British Army during the
Crimean War (F. Nightingale 1858)
(area of pie = number of deaths)

Resource on writing and publishing scientific papers

- Chapter 16: Tables
- **Chapter 17: Graphs**
- Presentations
- CVs
- Writing manuscripts
- How to submit a manuscript and review process

Gastel, B., & Day, R. A. (2016). *How to write and publish a scientific paper* (Eighth ed.). Greenwood, an imprint of ABC-CLIO, LLC.
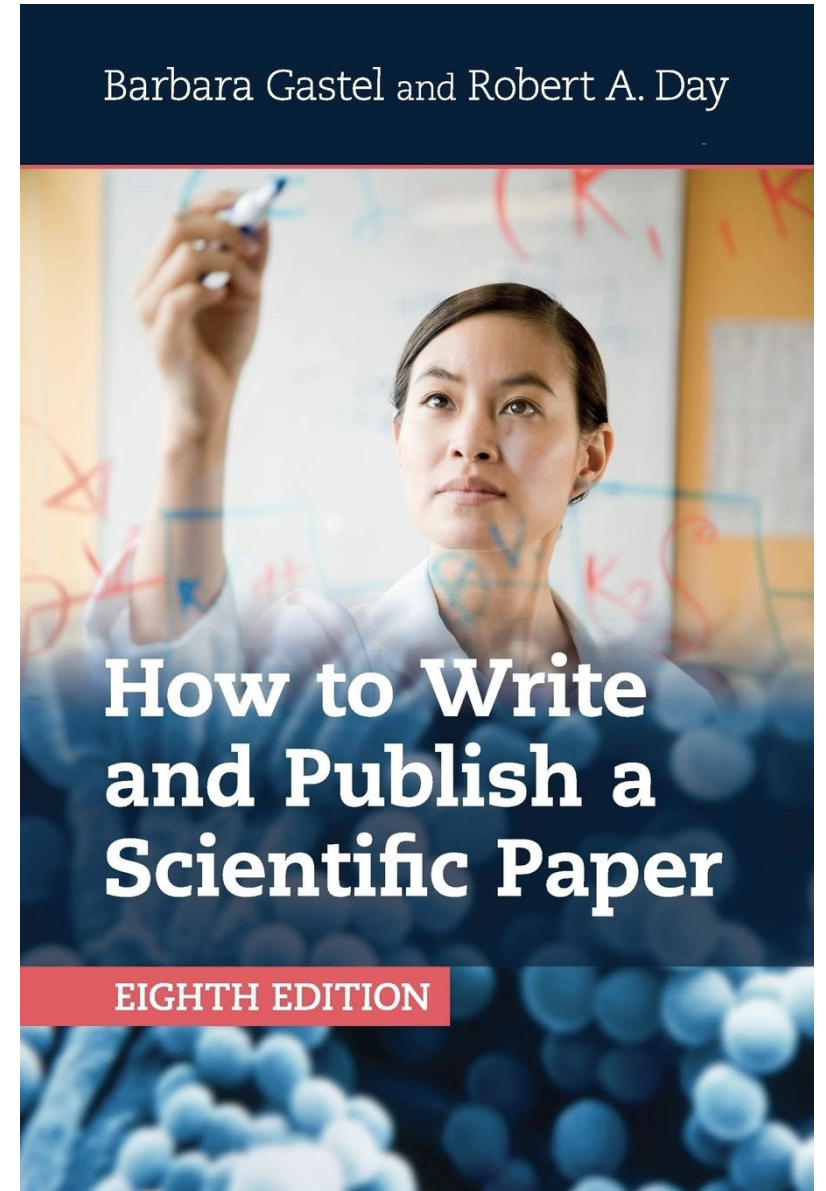
Currently in 9th edition since ~ 1979
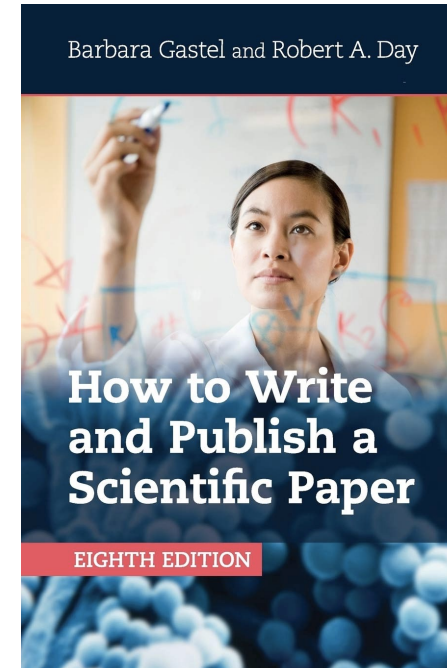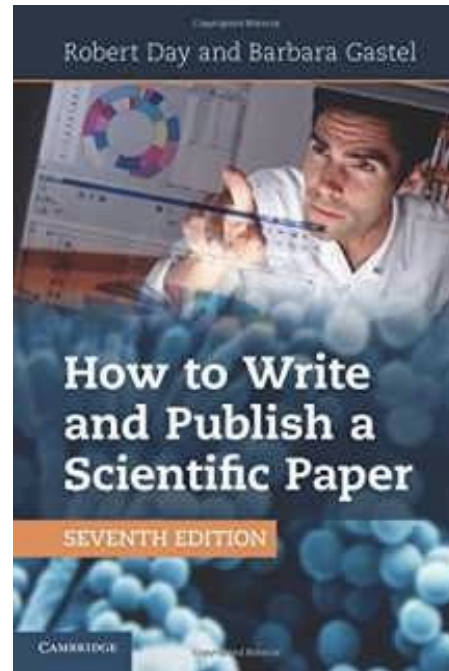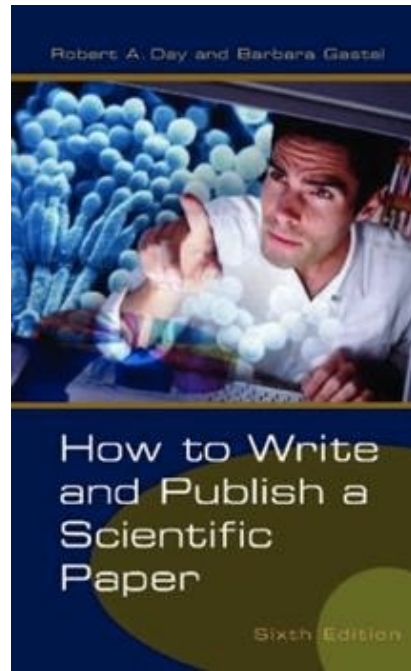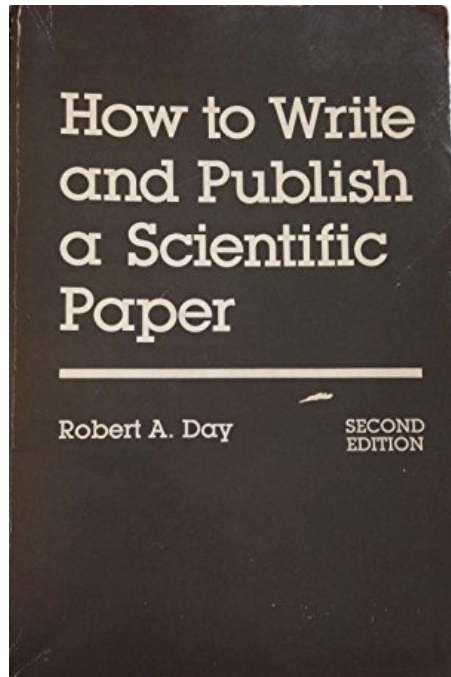
If you want an electronic copy of 8th edition, see me during the workshop

Why do we use graphs?

- To show patterns in data "greatest number of ideas in shortest time"

- To show spread, shape of distribution—frequency distributions

- To show associations between variables
  - The relationship between two or more variables (linear, nonlinear, other)
  - Differences between groups

Graphs should make the viewer goes "Oh!" and not "Huh?"
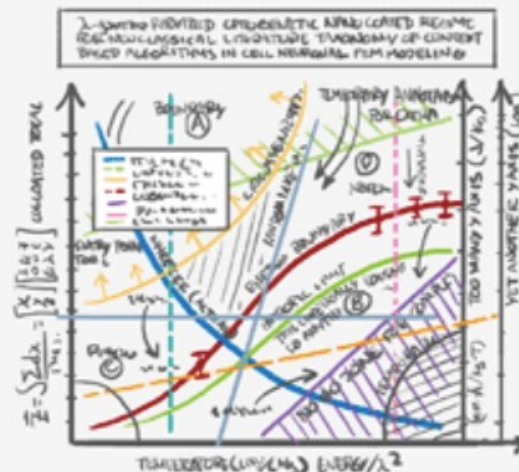
How to achieve graphical excellence?

*"Graphical excellence is that which gives to the viewer the **greatest number of <u>ideas</u> in the shortest time** with the least ink in the smallest space"*
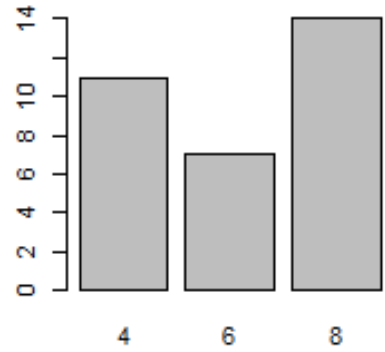   – Tufte (1983)

Edward Tufte the "Leonardo da Vinci of data" wrote 4 books on data visualization

Take home point

- An entire manuscript key points should be understandable with only the figures +tables+captions
  - No undefined abbreviations or unfamiliar terms
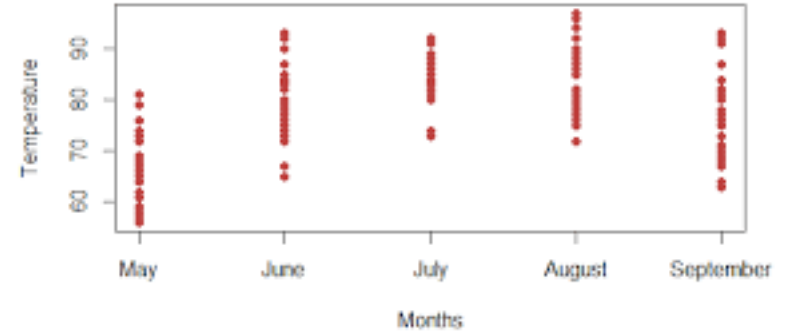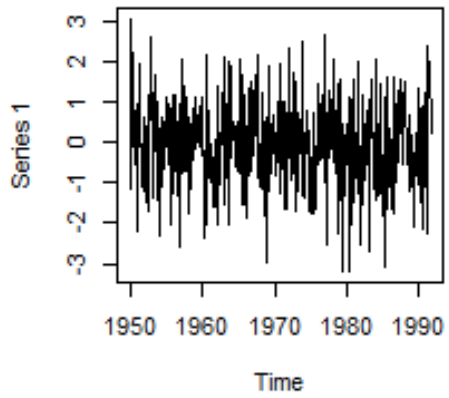 - Include units
  - Include sample sizes
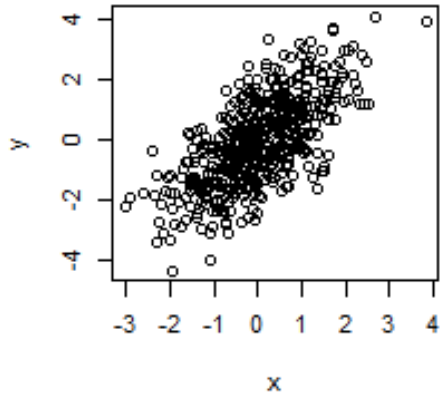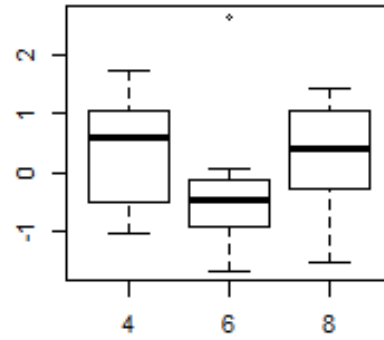
Gastel, B., & Day, R. A. (2016). *How to write and publish a scientific paper* (Eighth ed.). Greenwood, an imprint of ABC-CLIO, LLC.
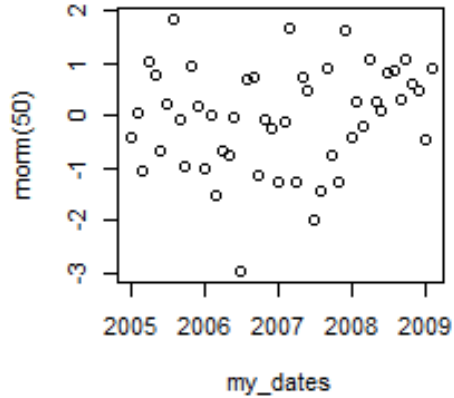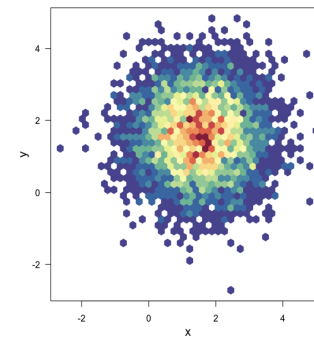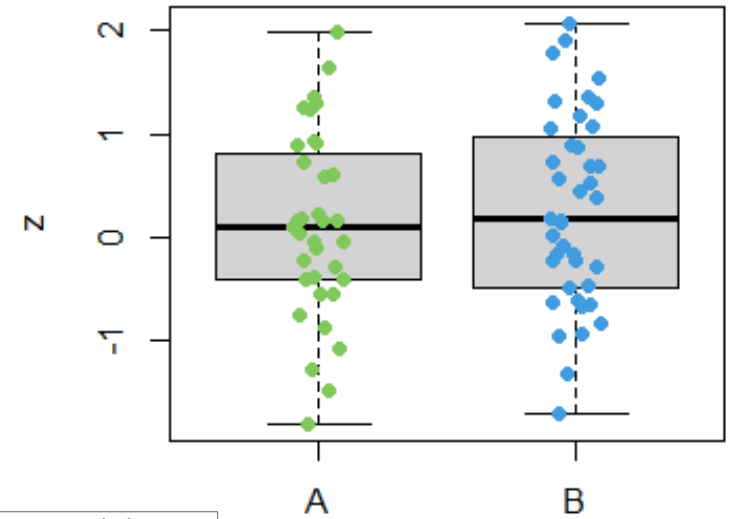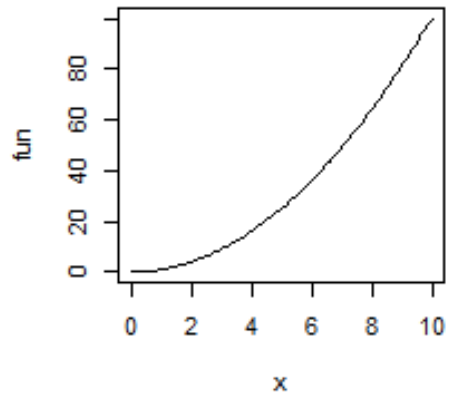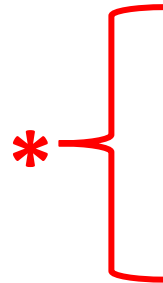
Endless types of graphs available in R

4 Principles to Increase Effectiveness of Graphs

1. Show the data

2. Make patterns in the data easy to see

3. Represent magnitudes honestly

4. Draw graphical elements clearly, minimizing clutter

Reference: Tufte 1983

*Hint: The rubric for Assignment #1 specifically refers to these 4 principles

Assignment number 1: Improve a Graph					
Full name of student:_____					
Student #:_____					
Assignment score out of 15:_____					
Writen intrepretaion of the Bad Graph (2 marks)					
	Excellent (full mark)	Average	Poor	Missing	
Written explanation of the study (enough detail to understand graph, but no too verbose)	1		0.5		**Total per section**
Analyze the "bad graph". Explain what patterns it intended to show and why it was not effective at accomplishing this?	1		0.5		**2**
Writen intrepretation & Plotting the New graph (8 marks)					
	Excellent (full mark)	Average	Poor	Missing	
Written explanation explaining how your improvements display the patterns more effectively	2		1		
1. Show the data (is all the data shown accurately?)	1		0.5		
2. Make patterns in the data easy to see. Are the patterns easier to see in this NEW graph compared to the original?	2		1		
3.Represent magnitudes honestly. Is the baseline (0) reasonable, are the axis scales reasonable?	1		0.5		
4. Draw graphical elements clearly, minimizing clutter. Is the new graph more clear or less cluttered?	1		0.5		**Total per section**
Does the graph have a well-written caption that stands alone from text?	1		0.5		**8**
Quality of the R code (5 marks)					
	Excellent (full mark)	Average	Poor	Missing	
Is the code well annotated (with # comments) so that a new person could understand it?	2		1		
Are the variable names and assignments appropriate (not too long, not too short) and logical if a new person was using this code?	1		0.5		**Total per section**
Could the code be extended to another situation without starting again (is the code reproducible)? Is it "general" to be re-used?	2		1		**5**
					Total Score out of 15
	15				**15**

1. Show the data

"While scatterplots prompt the reader to critically evaluate the statistical tests and the authors' interpretation of the data, bar graphs discourage the reader from thinking about these issues"
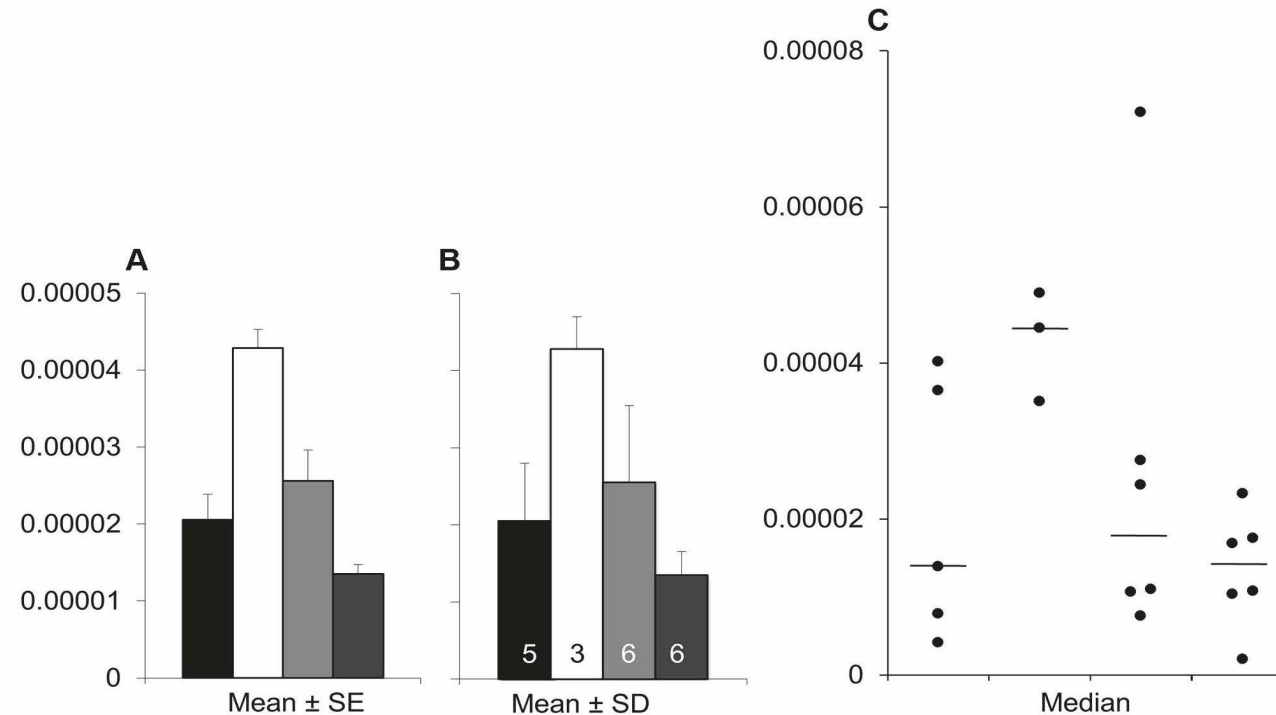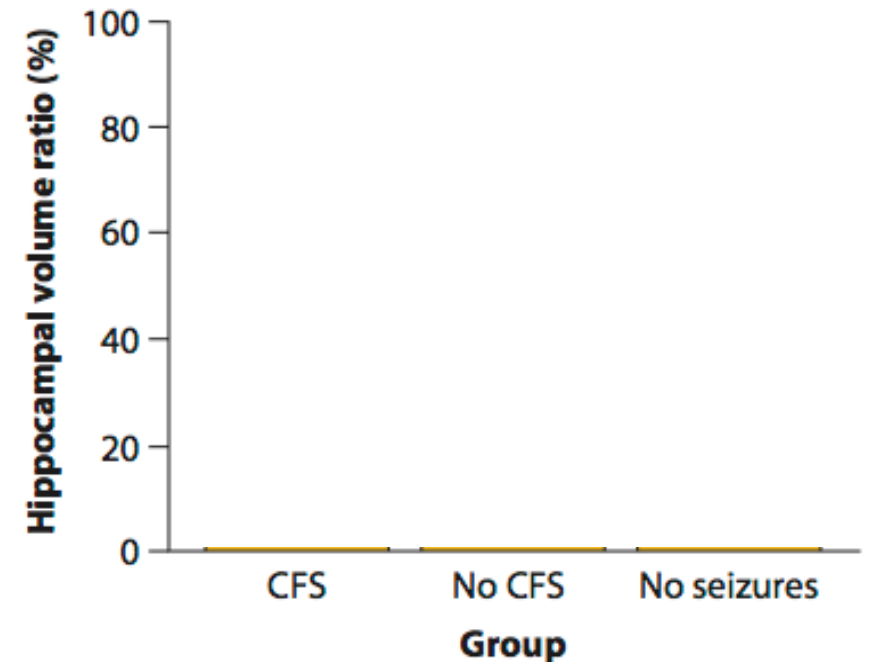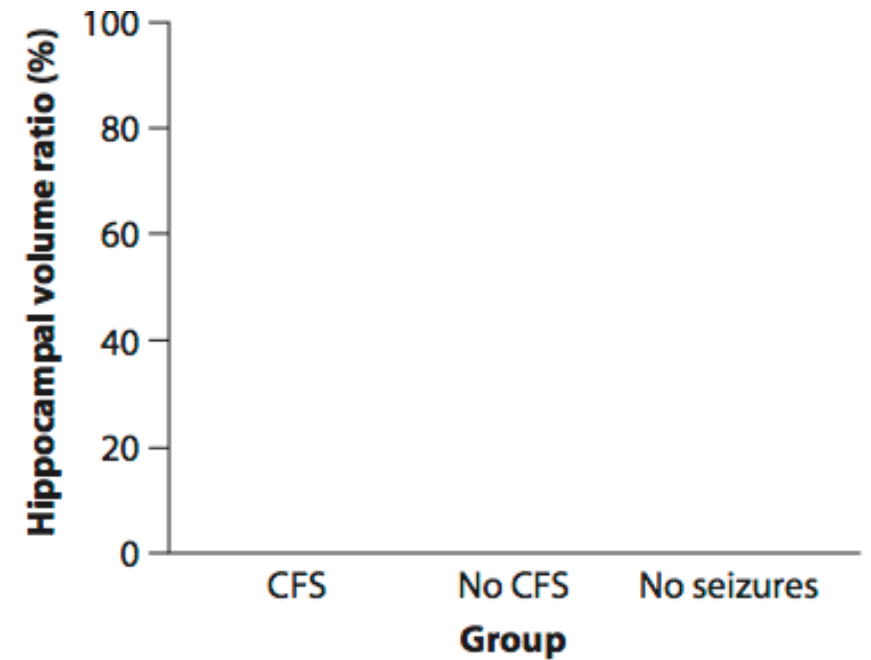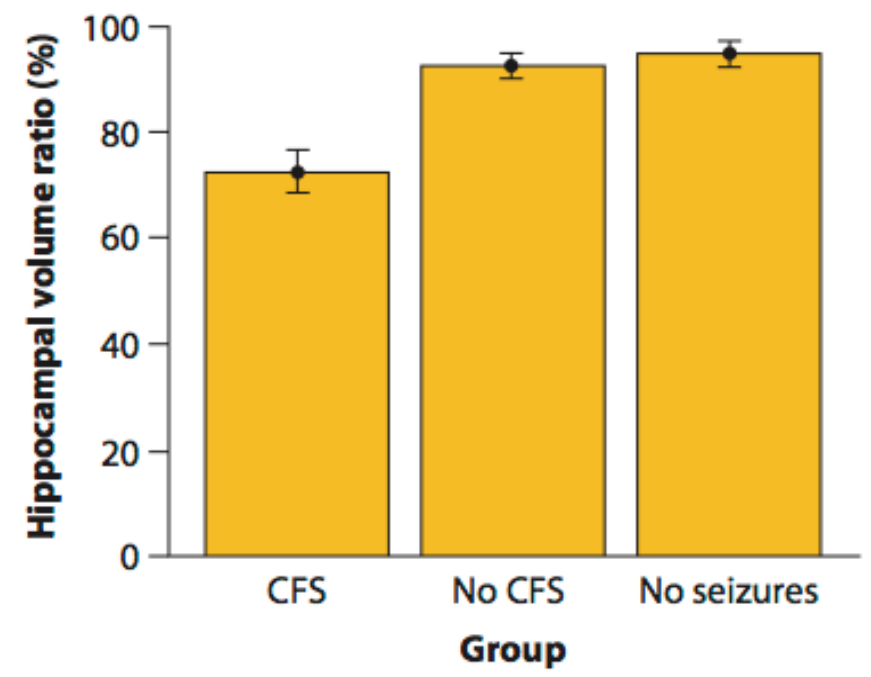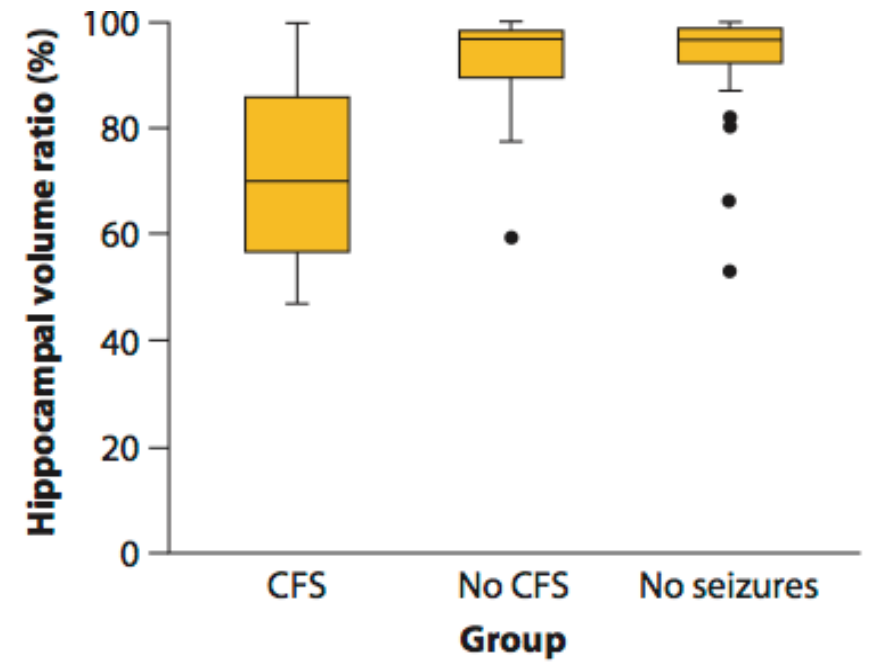
1. Show the data

- Hippocampal volume loss in patients with drug resistant epilepsy
  - Y-variable=numerical
- Treatments (X-variable categorical)
  - Childhood febrile seizures (CFS)
  - Childhood non-febrile seizures (no CFS)
  - No childhood seizures

Cooke et al 1993, n=107 patients

1. Show the data

Discussion:
These show the same data.
Which graph is more effective ? Why?

2. Make patterns easy to see

"Graphical excellence consists of complex ideas communicated with clarity, precision and efficiency" – Tufte (1983)

2. Make patterns easy to see

- Number of bird species and number of distinct human languages present in each square of a grid of continental Africa.

- What is the pattern in these data? How long did it take you to "see"?

- Is it easy to appreciate how strong the relationship is between the variables?

species richness
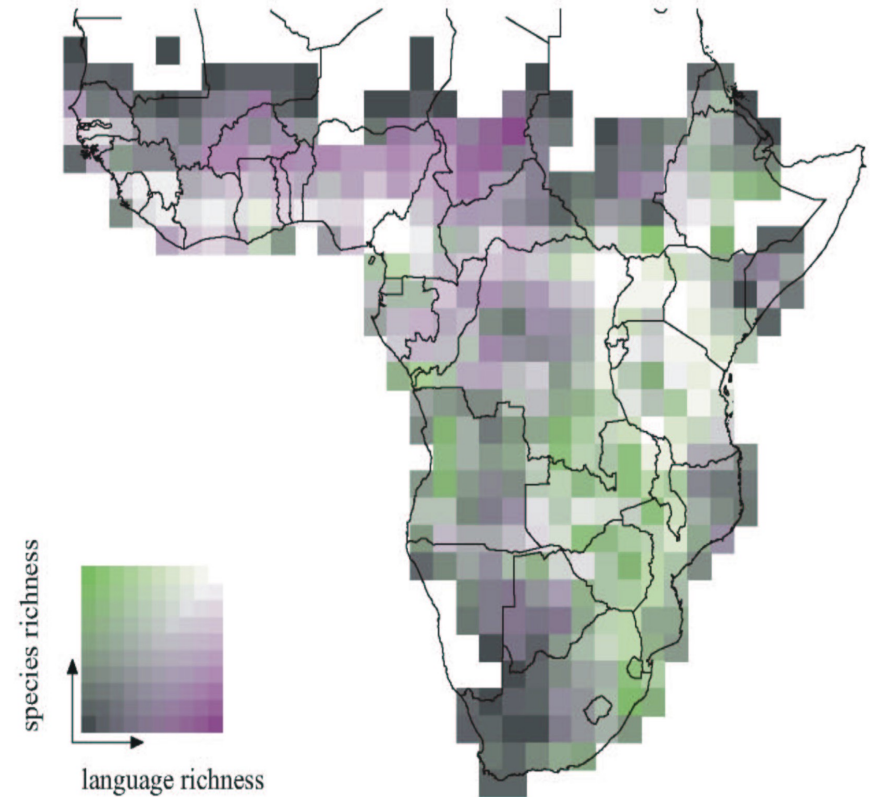
language richness

2. Make patterns easy to see

- Fig. 2. A comparison of language and species richness in sub-Saharan Africa. Each variable has been partitioned into 10 equal-frequency ranks. Cells of equal rank for both variables are coloured in grey scale from black (both low rank) to white (both high rank). The degree of colour saturation indicates that the ranks are different. A cell is green if the species-richness rank is high relative to language richness and blue if language richness is high relative to species richness.
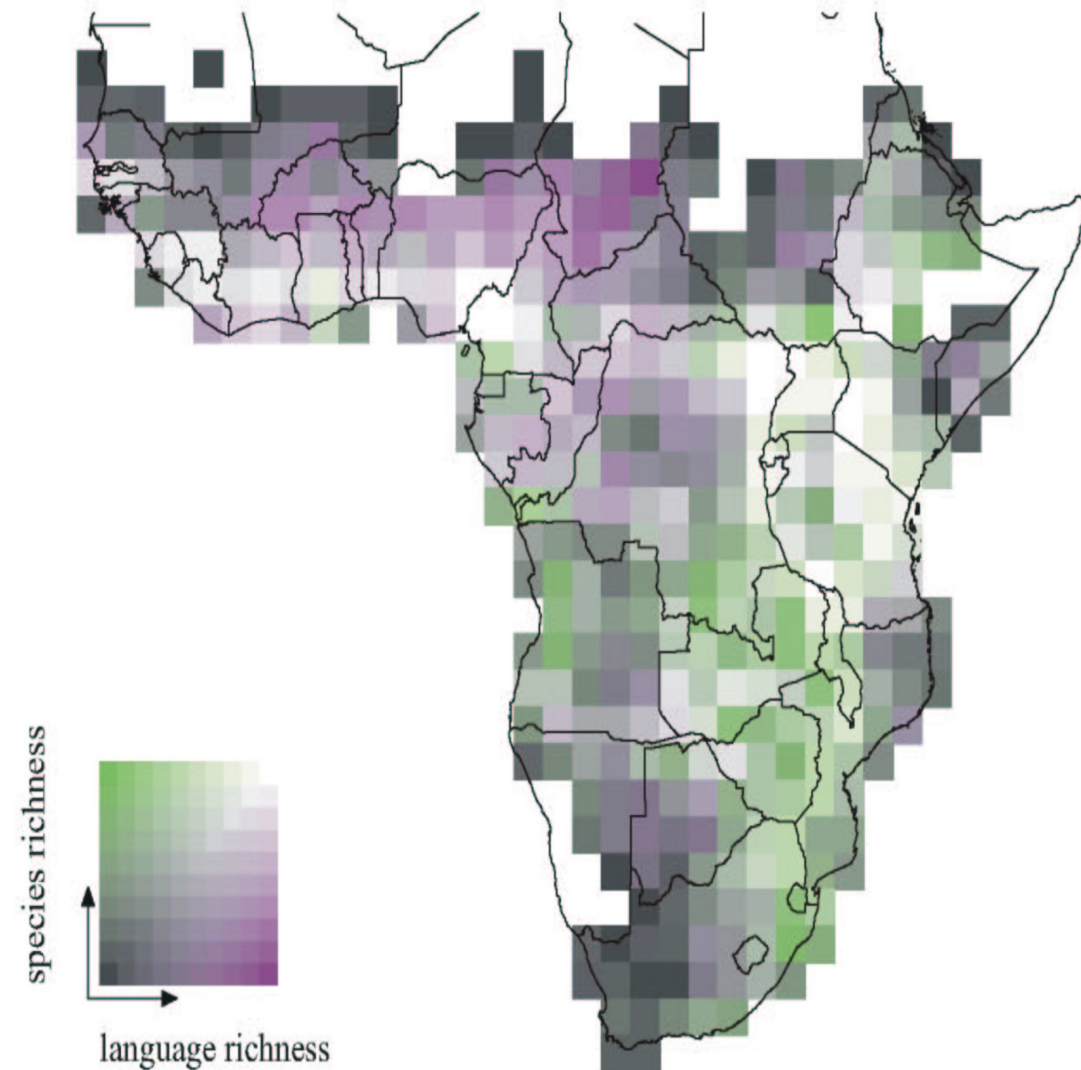
The distribution of cultural and biological diversity in Africa

Joslin L. Moore[1,2*], **Lisa Manne**[3], **Thomas Brooks**[1,4], **Neil D. Burgess**[1,5], **Robert Davies**[6], **Carsten Rahbek**[2], **Paul Williams**[3] and **Andrew Balmford**[1]

[1]*Conservation Biology Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK*
[2]*Zoological Museum, University of Copenhagen, Universitetsparken 15, DK-2100, Copenhagen Ø, Denmark*
[3]*Biogeography and Conservation Laboratory, The Natural History Museum, Cromwell Road, London SW7 5BD, UK*
[4]*Center for Applied Biodiversity Science, Conservation International, 1919 M Street NW, Suite 600, Washington, DC 20036, USA*
[5]*Conservation Science Programme, World Wildlife Fund, 1250 24th Street NW, Washington, DC 20037-1193, USA*
[6]*Llanunwas, Upper Solva, near Haverfordwest, Pembrokeshire SA62 6UJ, UK*

Anthropologists, biologists and linguists have all noted an apparent coincidence in species diversity and human cultural or linguistic diversity. We present, to our knowledge, one of the first quantitative descriptions of this coincidence and show that, for 2° × 2° grid cells across sub-Saharan Africa, cultural diversity and vertebrate species diversity exhibit marked similarities in their overall distribution. In addition, we show that 71% of the observed variation in species richness and 36% in language richness can be explained on the basis of environmental factors, suggesting that similar factors, especially those associated with rainfall and productivity, affect the distributions of both species and languages. Nevertheless, the form of the relationships between species richness and language richness and environmental factors differs, and it is unlikely that comparable mechanisms underpin the similar patterns of species and language richness. Moreover, the fact that the environmental factors considered here explain less than half of the variation in language richness indicates that other factors, many of which are likely to be historical or social, also influence the distribution of languages.

3. Represent magnitudes honestly

- *"A graphic does not distort if the visual representation of the data is consistent with the numerical representation"* – Tufte (1983)

- Bad examples:
  - Compare 2 plots with different Y-axis scales
  - Hide outliers

- Does the data illustrate what the model/analysis was actually doing?

# Example: Does the data show what the model was actually doing?

- **Question:** A) Can heart rate predict metabolic rate on single dives only?
- Linear-mixed effects model (LME) treats animal as a random effect, so **I don't plot symbols by animal, but rather treatment type**
- This was an appendix figure in Thesis, but not in publication
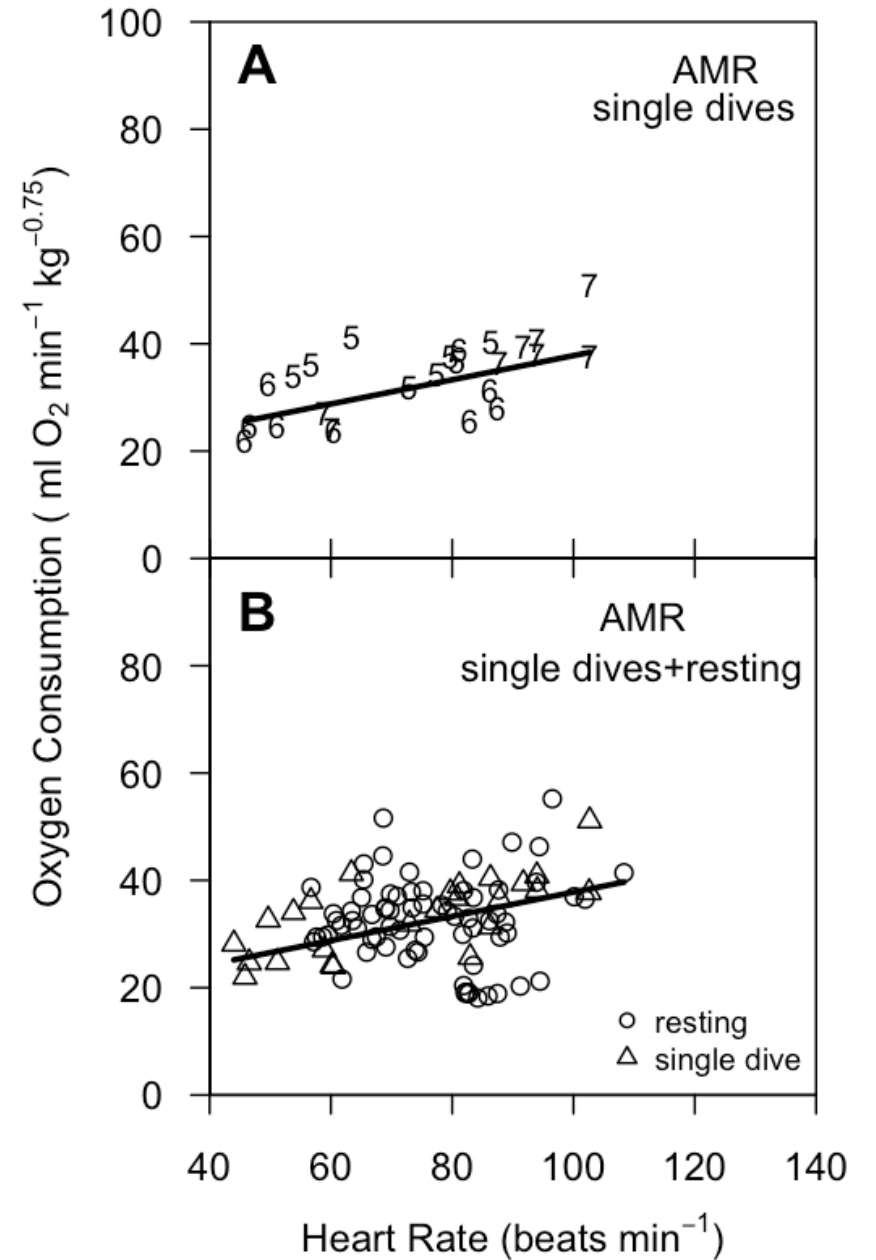
Figure A3. 5 The relationship between heart rate (*fh*) averaged over a dive cycle and average metabolic rate (*V&O2* =AMR) for Steller sea lions over a single dive (A) was not significantly different than when animals were resting in water (B). See Table 3.1 for Animal ID to corresponding animal number.

Example

- Birds sleep in risky situations by sleeping with 1 eye open and one hemisphere of the brain awake
  - Unihemispheric slow-wave sleep

- Some mammals do this too→whales, dolphins, seals, sea lions
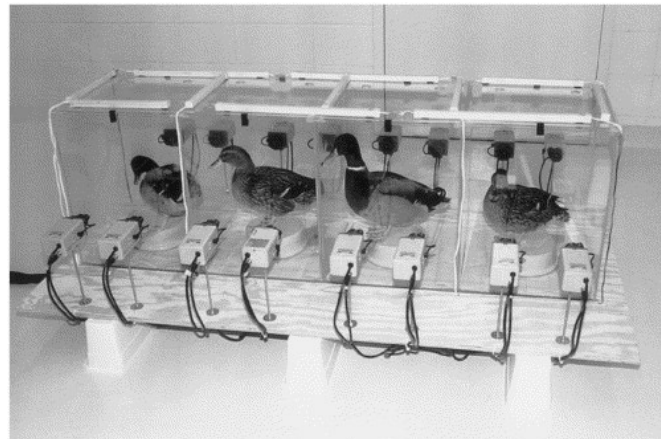
Slow wave sleep in the brain hemispheres of mallard ducks sleeping with one eye open.
From Rattenborg et al. (1999) *Nature*.
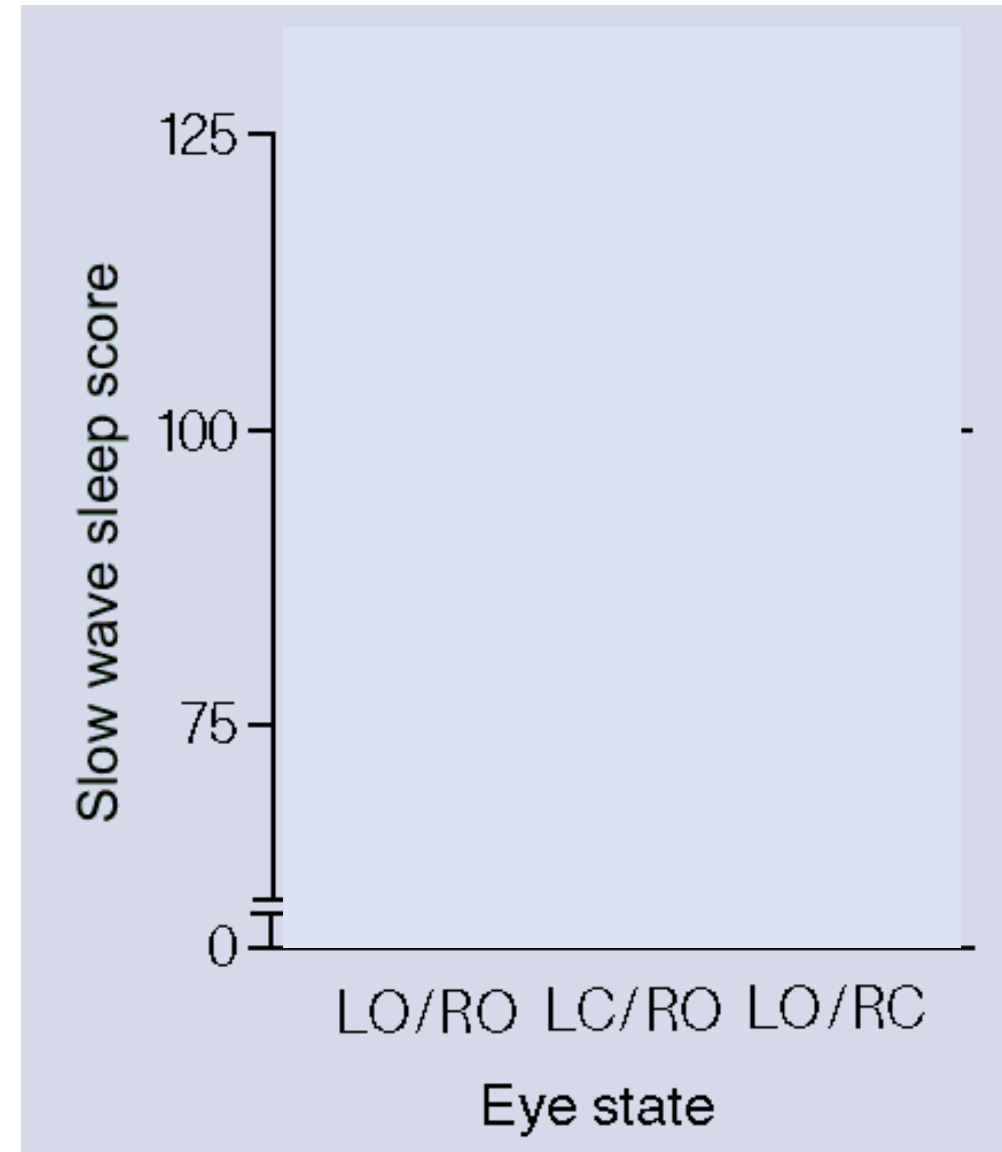
LO/RO=both eyes open
LC/RO=Left closed/Right open
LO/RC=Left open/Right closed

4. Represent magnitudes honestly

1. Are the bars "consistent with the numerical representation"?

2. Is 0 a reasonable baseline for evaluating sleep score?

3. Are there other issues with the graph?

Slow wave sleep in the brain hemispheres of mallard ducks sleeping with one eye open. From Rattenborg et al. (1999) *Nature*.
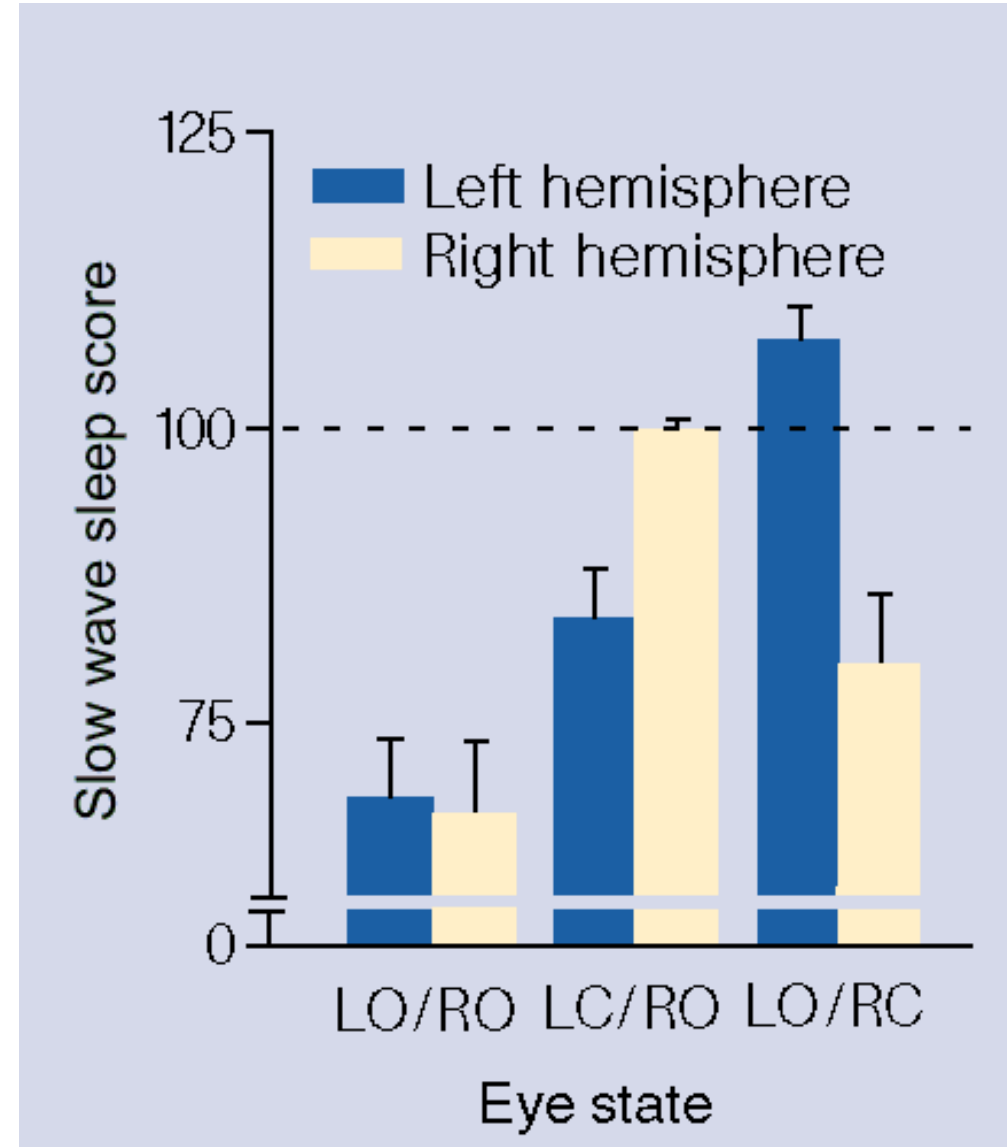
4. Draw graphical elements clearly, minimize clutter

- *Maximize the data-ink ratio, within reason"* – Tufte (1983)

- Originally journals charged for colour

- Modern take with online journals: How will this figure look printed black and white?

- Not a great use of negative space and solid grey background

Strip chart made with ggplot() using default graphical options.
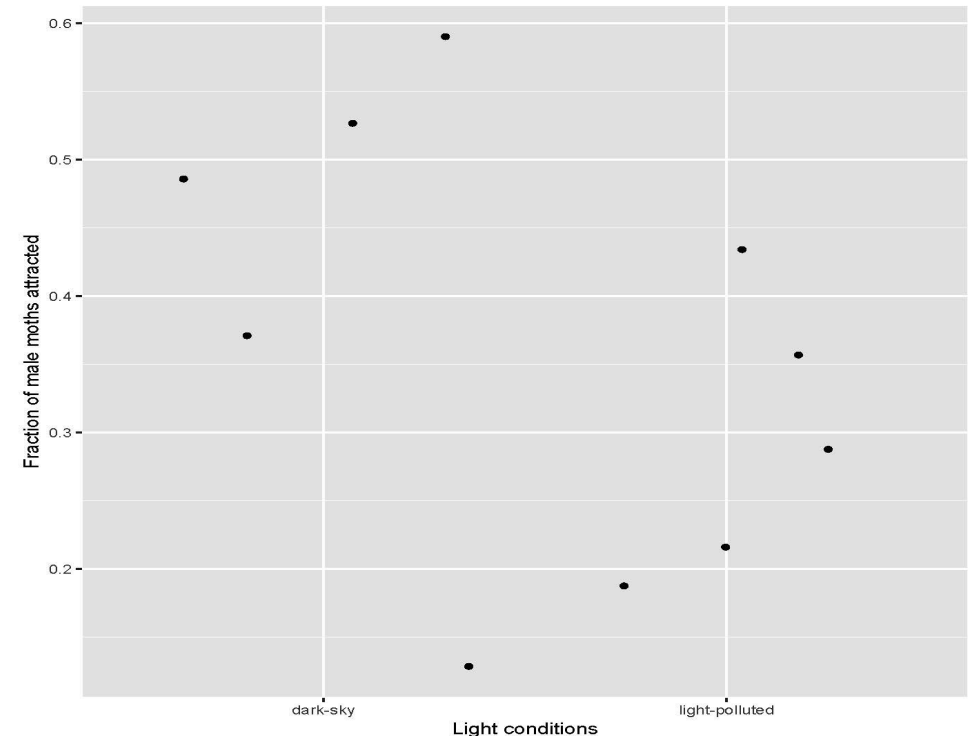
Altermatt and Ebert (2016) measured light attraction of ermine moths (*Yponomeuta cagnagella*) from 10 different populations. Five of the populations were located in urban areas with plenty of human lights. The other 5 populations were located in pristine areas with no light pollution

Default R plots are OK for exploration, but usually not polished enough to publish

4 Principles to Increase Effectiveness of Graphs

1. Show the data
2. Make patterns in the data easy to see
3. Represent magnitudes honestly
4. Draw graphical elements clearly, minimizing clutter

Reference: Tufte 1983

Outline

How to display category frequencies: **Bar graph**

- <u>Height</u> of bars displays frequency distribution of a categorical (grouping) variable

- Zero baseline

- Space between bars emphasize height

- Order of categories – most to least frequent is usually best
  - **R defaults order alphabetically**
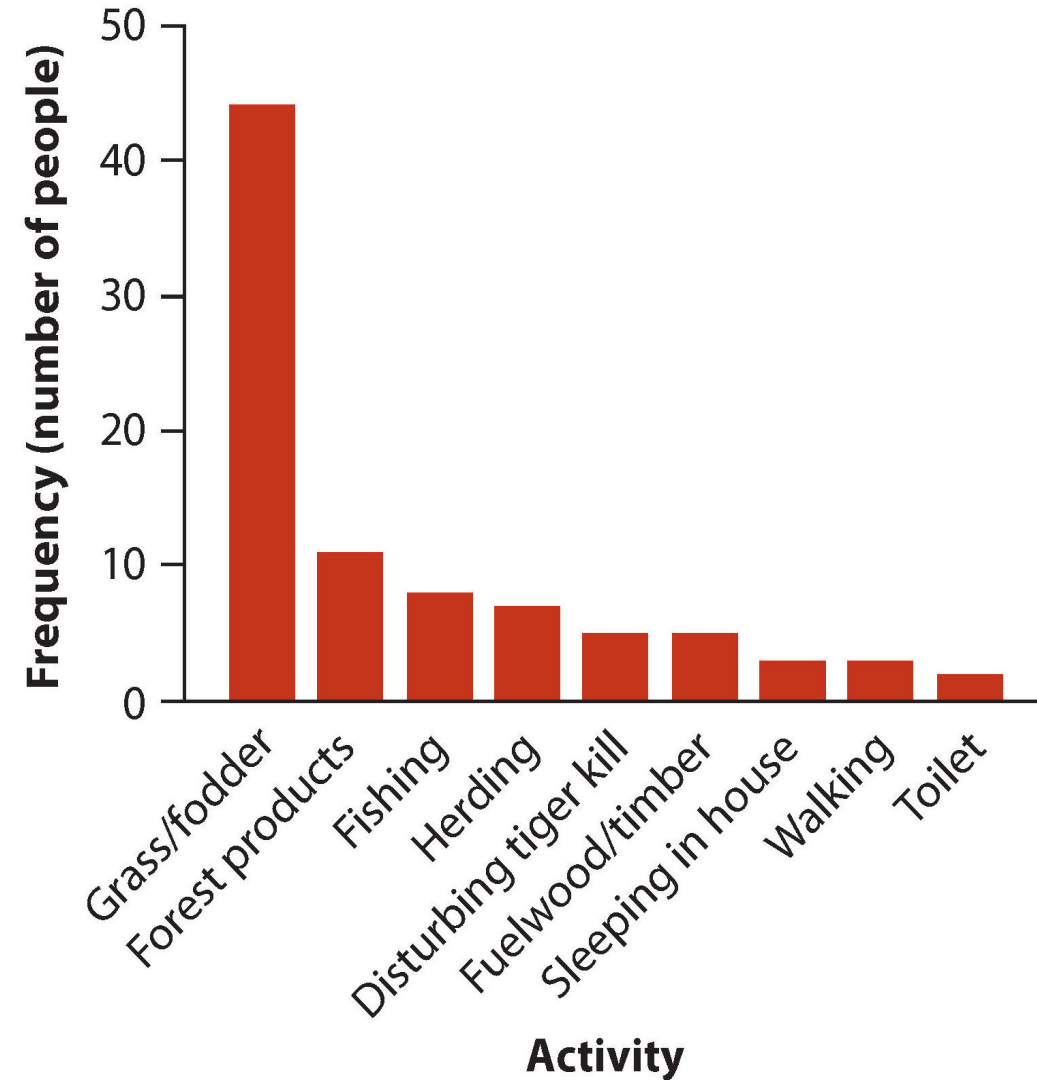
Activities of people at the time they were attacked and killed by tigers near Chitwan National Park, Nepal, between 1979 and 2006

How to display category frequencies
Bar graph vs. pie chart: which is more successful?

- Humans are better at comparing relative areas in bar graphs than pie charts.
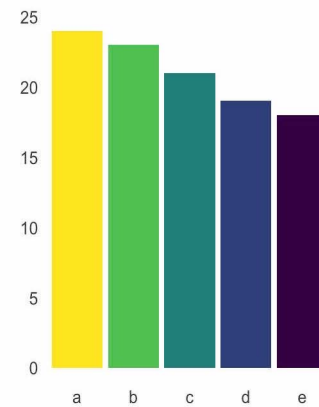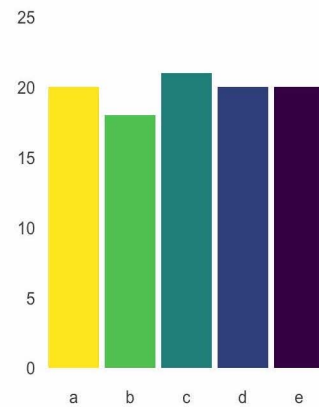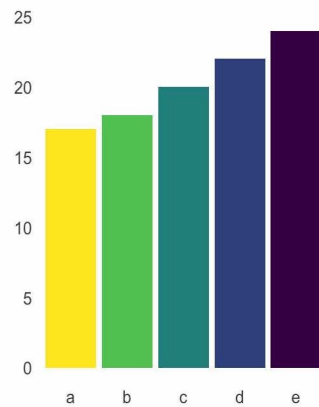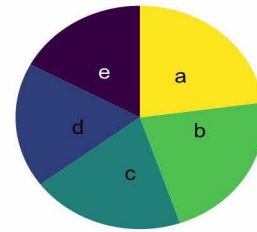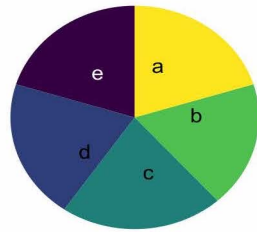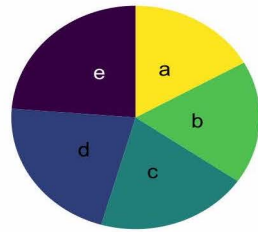
How to display frequency distribution for numeric variable:Histogram

- Uses <u>area</u> of bars to display frequency distribution of a numerical variable

- Zero baseline

- No spaces between bars

- Choice of number of bins and bin width

The frequency distribution of bird species abundance at Organ Pipe Cactus National Monument. *n* = 43 species
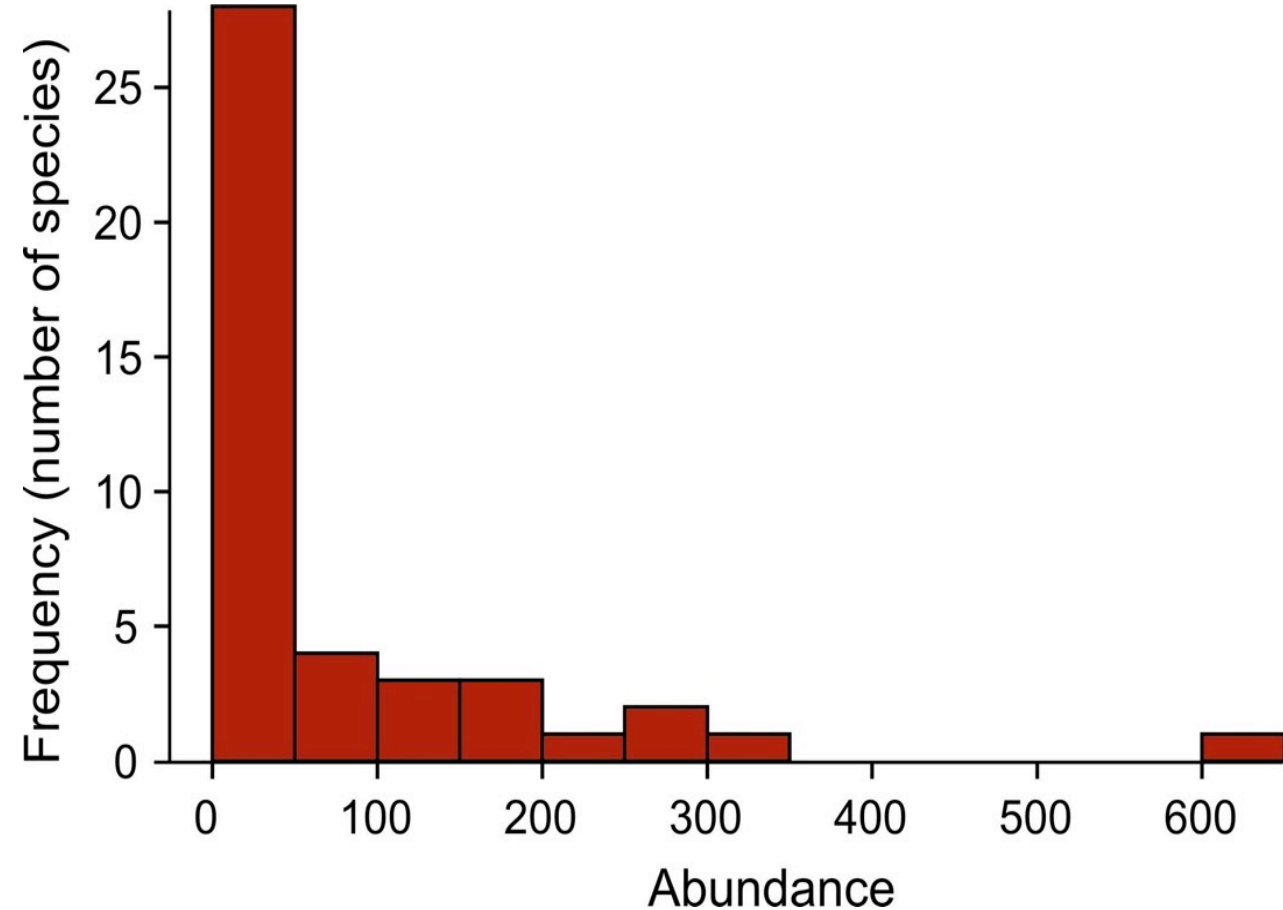
How to display association between 2 numerical variables: Scatterplot

- Non-zero baseline often ok

- Goal is to show association, not height above 0

- Points fill the space available

If you use base R plot(), R will just plot the data by it's best guess of plot type (based on the class of your variables)
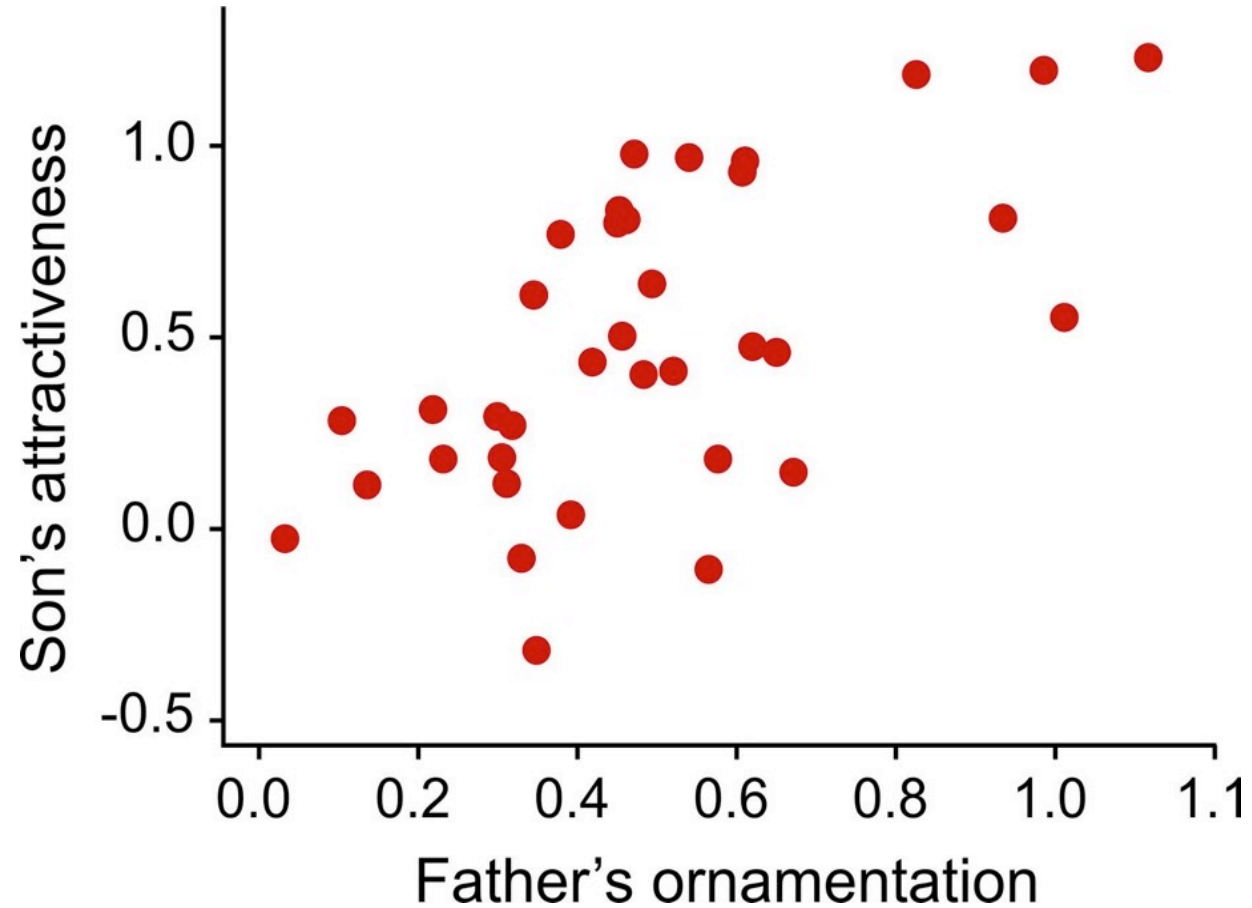
The relationship between the ornamentation of male guppies and the average attractiveness of their sons. *n* = 36 families.

A catterplot
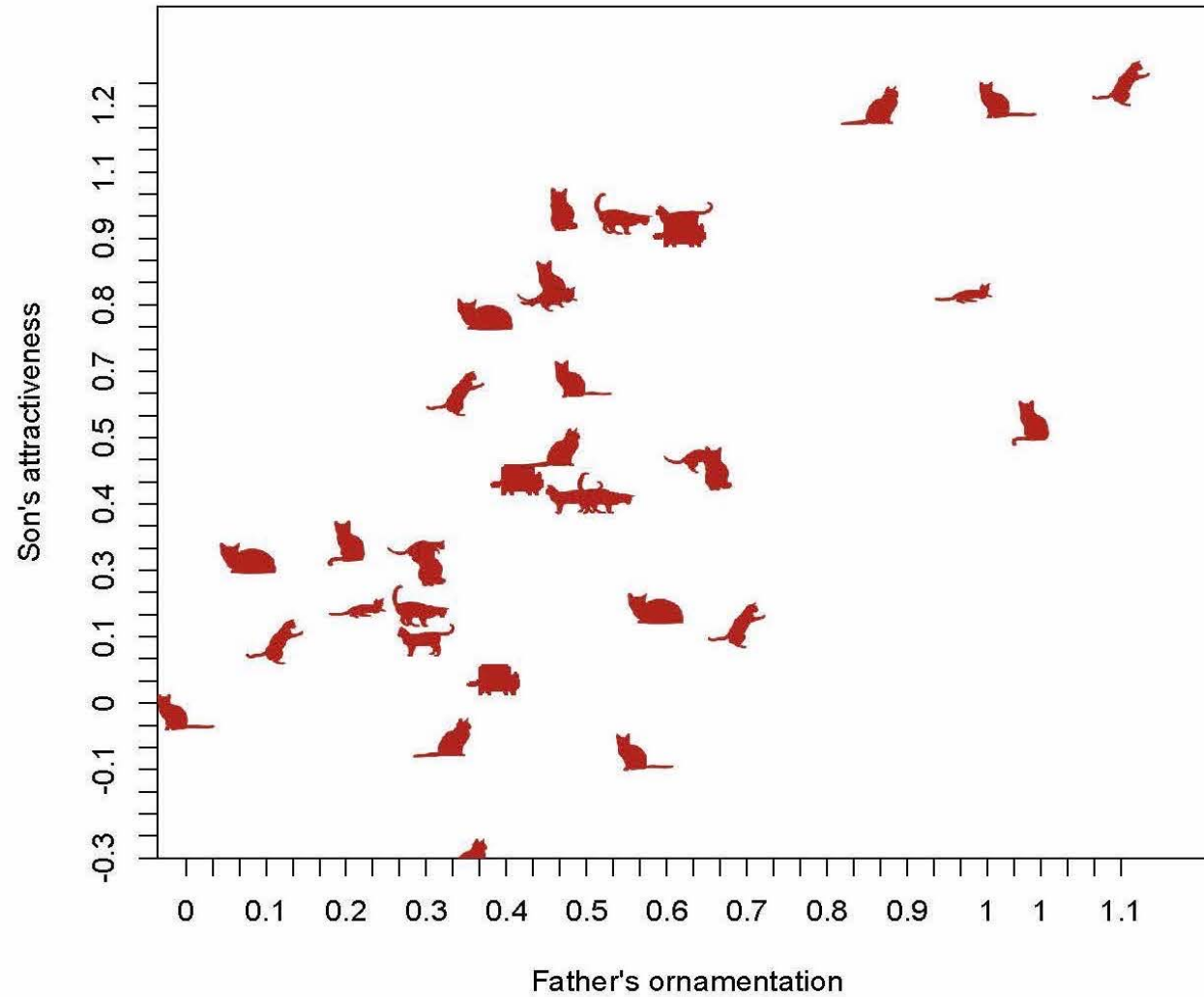
How to display association between categorical variables: Grouped bar graph

- Uses <u>height</u> of bars to display <u>association</u> between two (or more) categorical variables.

- Explanatory variable = outer groups; response variable = inner groups

- Zero baseline (so that height is proportional to frequency)

- Spacing between bars wider between outer groups
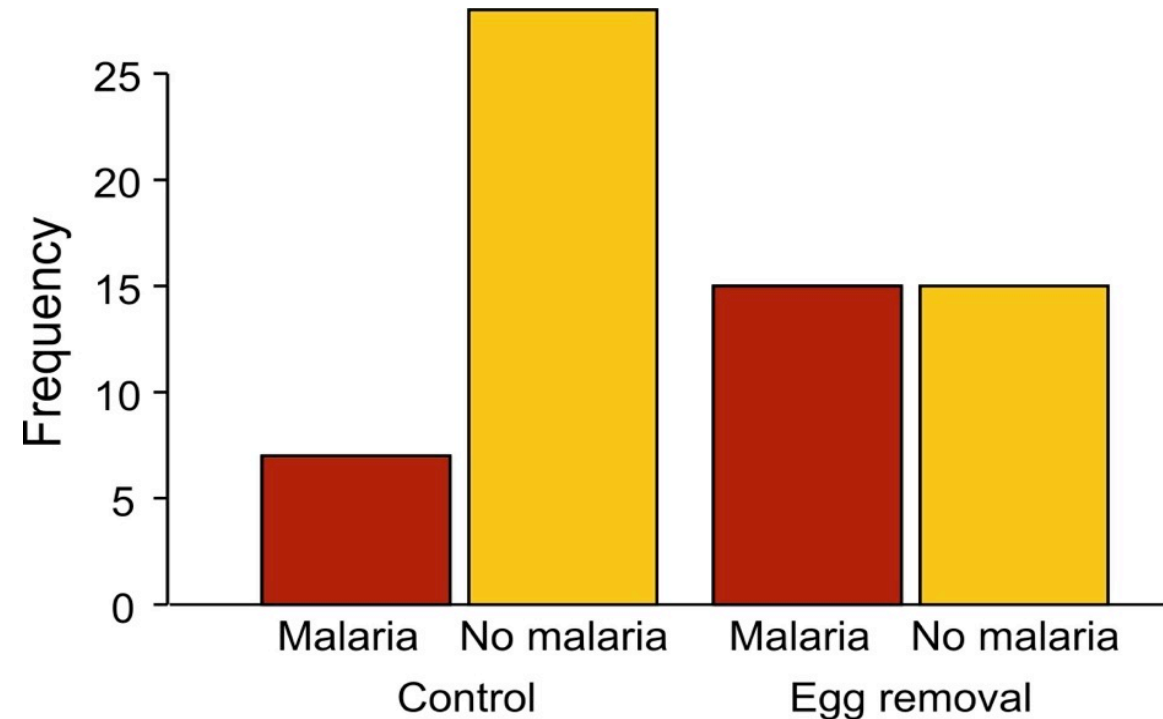
Incidence of malaria in female great tits in relation to experimental treatment. *n* = 65 birds.

How to display association between categorical variables: **Mosiac plot**

- Uses <u>area</u> of rectangles to display <u>association</u> between two (or more) categorical variables

- Explanatory variable along horizontal axis; response variable stacked

- Area proportional to frequency

- Like a graphical representation of a contingency table

Incidence of malaria in female great tits in relation to experimental treatment.
n = 65 birds.

Grouped bar graph vs. mosaic plot
Which is more successful?
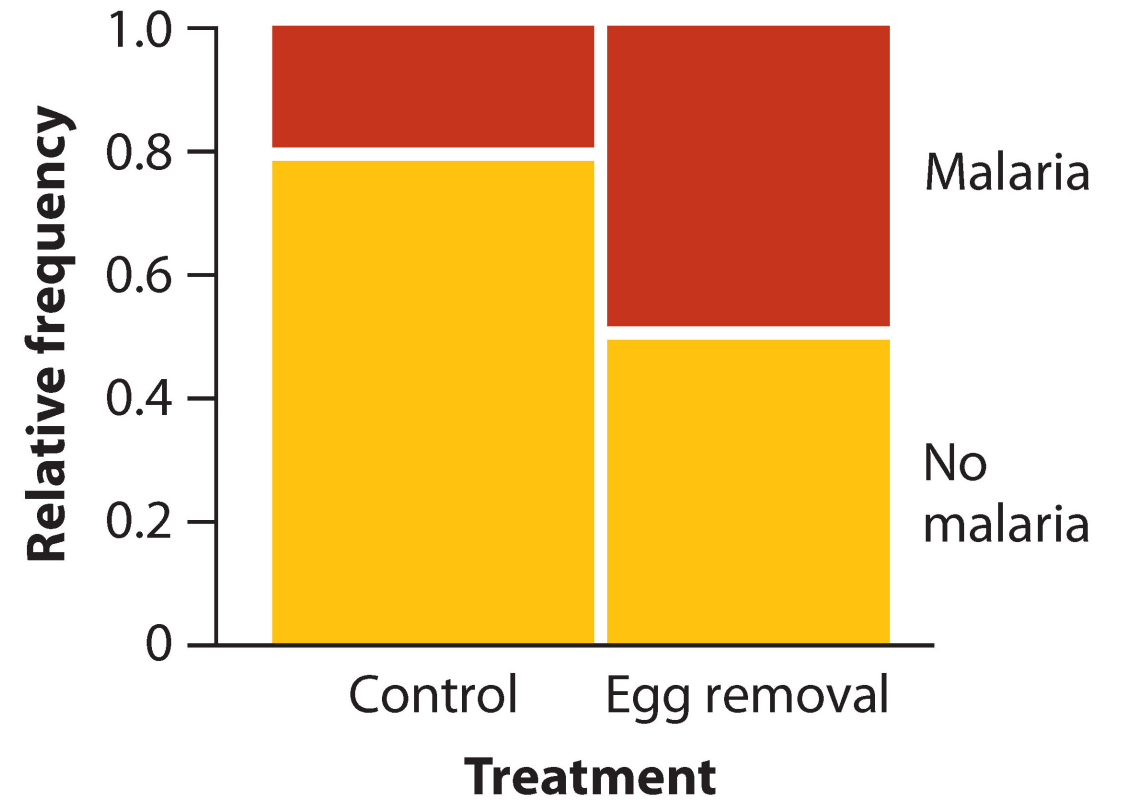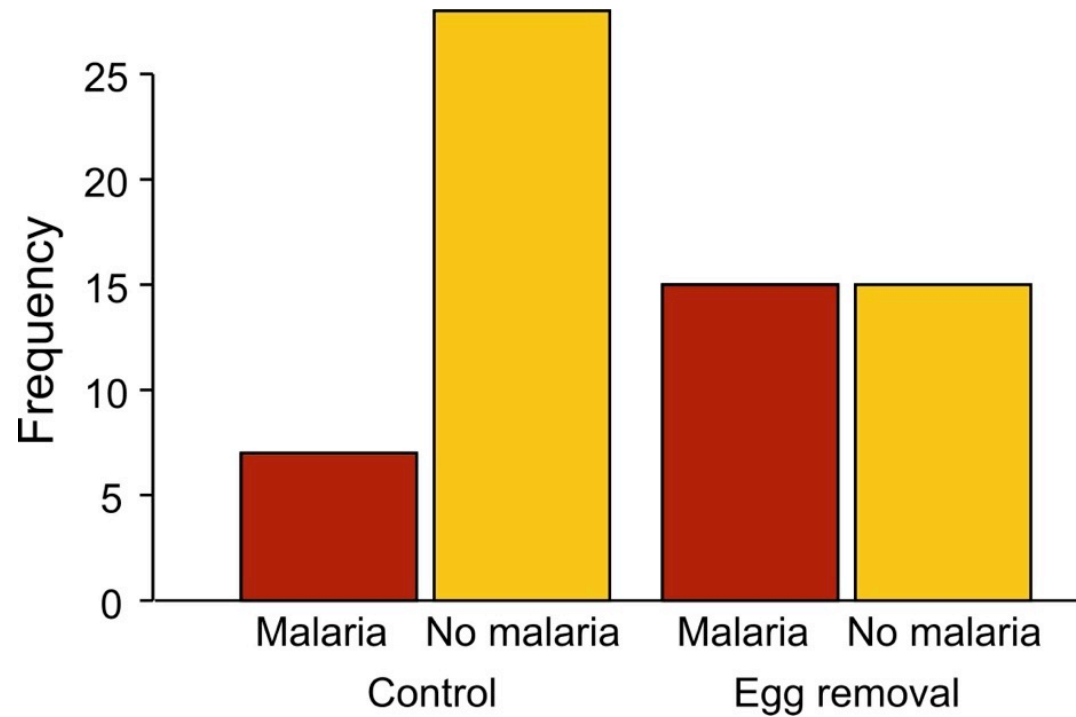
How to display association between numerical and categorical variables? Strip chart

- Displays **differences between groups**

- Shows the data points

- Non-zero baseline often ok (goal is association not magnitude or frequency)

- Points fill the space available
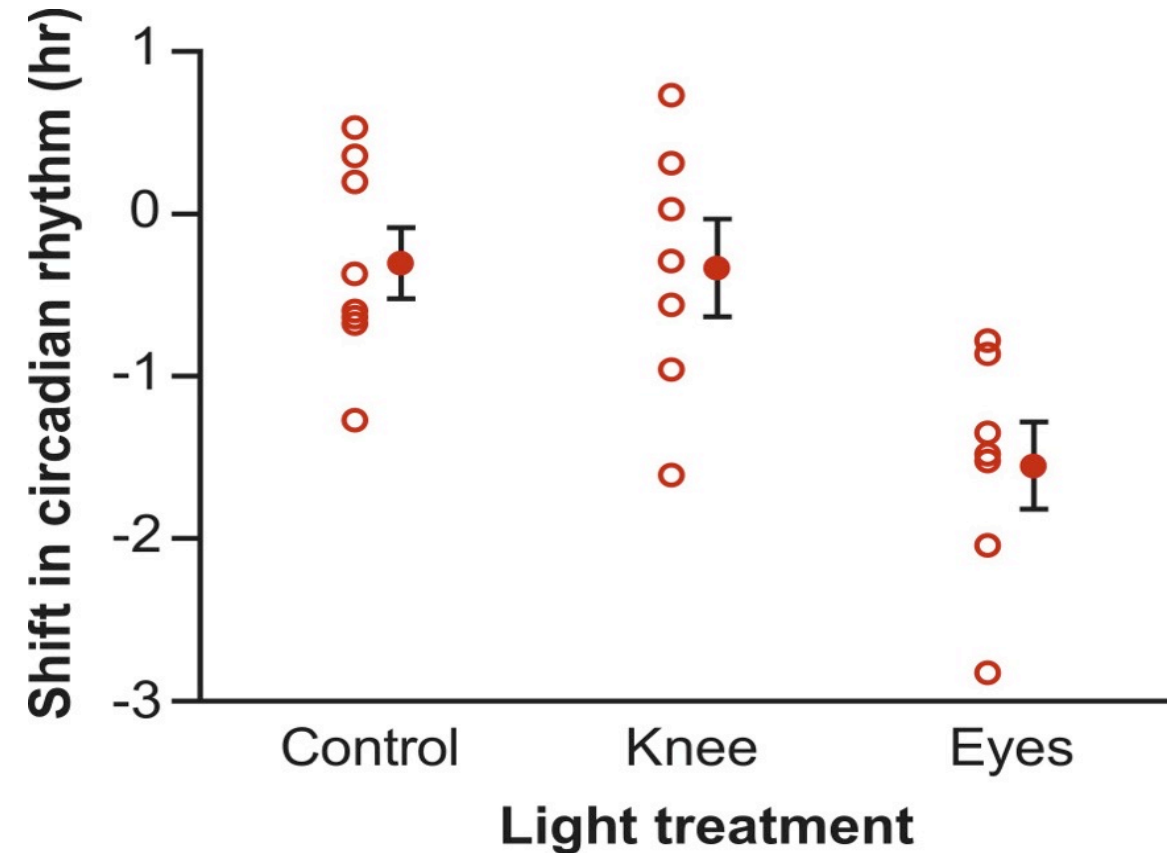
Phase shift in the circadian rhythm of melatonin production in 22 subjects given alternative light treatments (open circles). Group means ± 1 SE also shown.

How to display association between numerical and categorical variables?

- Strip chart vs multiple histograms
- Too many data points for a strip chart.
- Stack histograms vertically to best compare
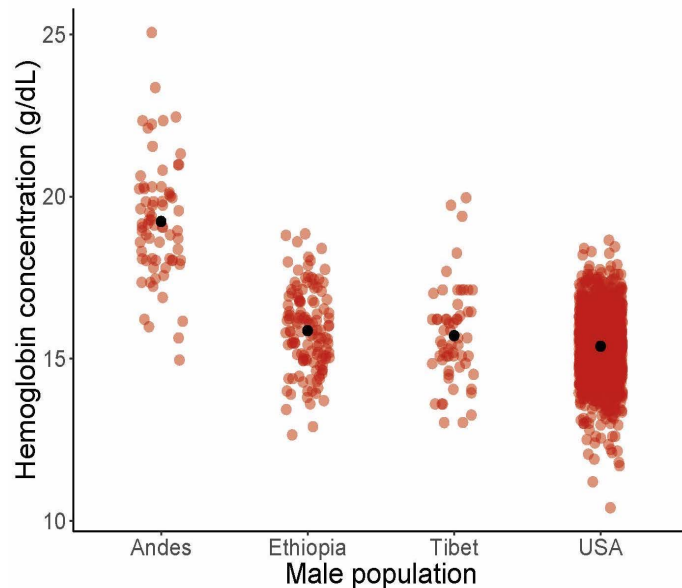
Hemoglobin concentration in blood of males living at high elevation compared to sea-level USA control
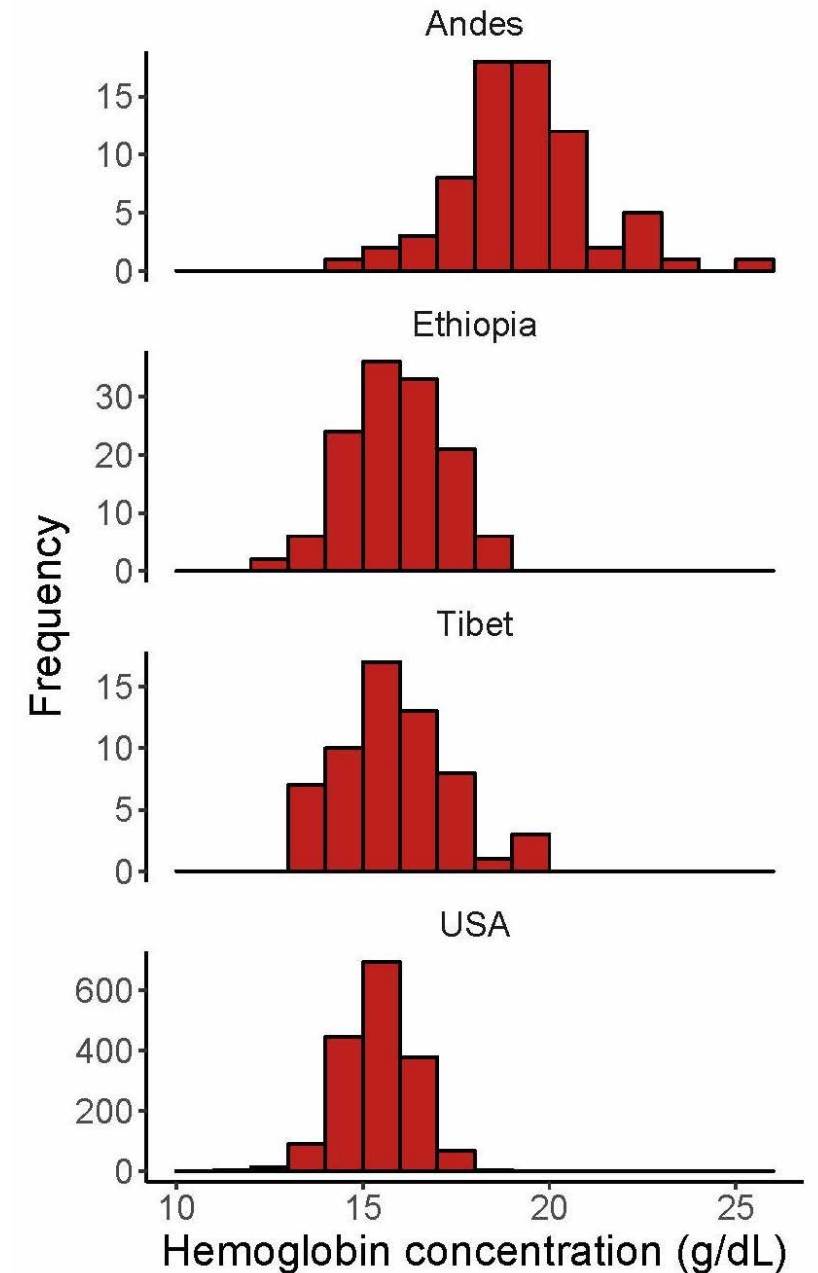
Discussion:
Which is more successful at conveying a pattern? Why?
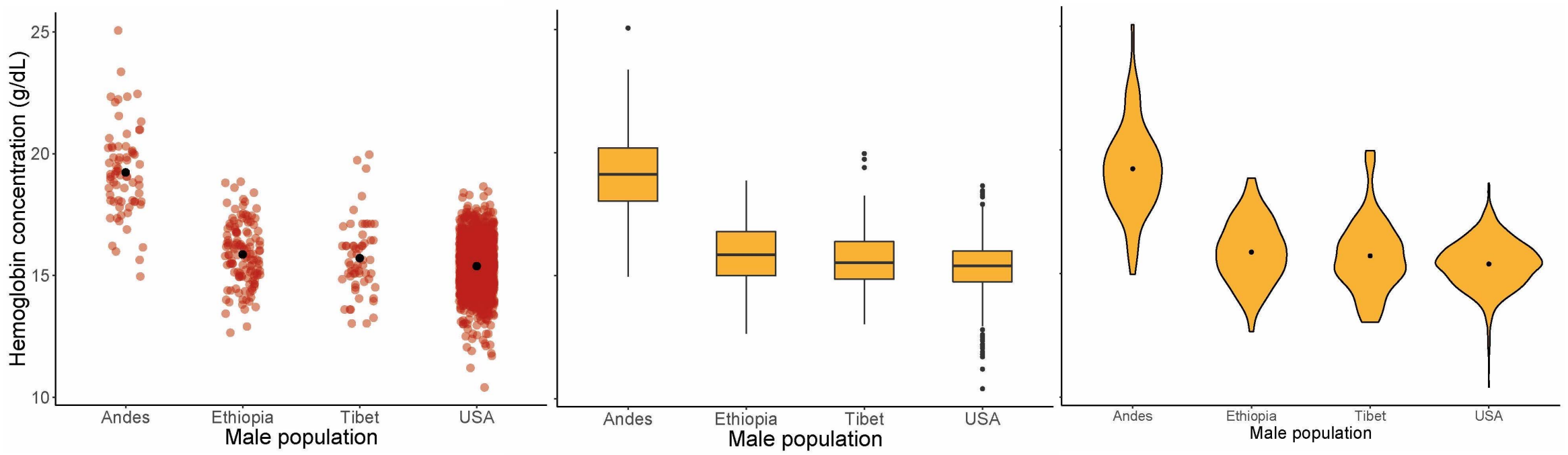
Strip chart | Box plot | Violin plot

How to display association between numerical and categorical variables? Box plot

- I use these a lot with LME models that are not predictive

- You can change order of X variable (default is alphabetical)

- **Question:** Does Y variable differ among the X variable categories?
  - **NOT** predicting Y variable, just asking if different
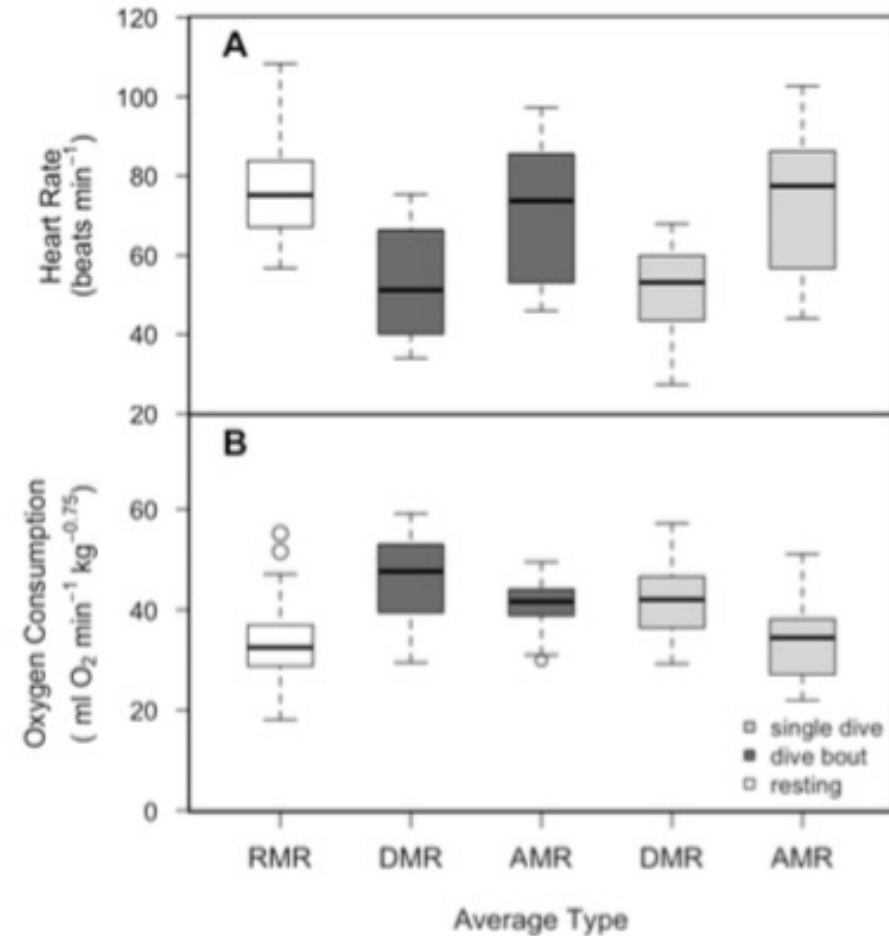
Boxplot example (with LME model)

Does the duration of an attempted prey capture (APC) differ between successful vs. un successful attempts?

Is this different if I look at each type of acceleration??

Fig 5. Analysis of successful prey captures by accelerometers. The accelerometer was not able to distinguish between successful and unsuccessful attempted prey captures (APC) on any axis of acceleration. Successful events included video where prey was captured, and unsuccessful events included events where the fur seal chased and missed an attempted capture attempt of the prey on video. Neither mean event duration (A, D, G), or integral area under the APC (B, E, H) significantly varied between successful or unsuccessful APC. The number of peaks in surge marginally differed between types of APC (C), but not for sway (F) or heave (I). Results are from Function 1 optimized with generic parameters (0.1 variance and 5 s). Each data point represents an individual APC.

Boxplot example (with LME model)

- Also looked at other acceleration metrics

Fig 5. Analysis of successful prey captures by accelerometers. The accelerometer was not able to distinguish between successful and unsuccessful attempted prey captures (APC) on any axis of acceleration. Successful events included video where prey was captured, and unsuccessful events included events where the fur seal chased and missed an attempted capture attempt of the prey on video. Neither mean event duration (A, D, G), or integral area under the APC (B, E, H) significantly varied between successful or unsuccessful APC. The number of peaks in surge marginally differed between types of APC (C), but not for sway (F) or heave (I). Results are from Function 1 optimized with generic parameters (0.1 variance and 5 s). Each data point represents an individual APC.

doi:10.1371/journal.pone.0128789.g005

Linear mixed-effects model: **not** predictive

- Numerical Y, **Categorical X**
- Animal random factor
- Fixed factor
- N=7 Steller sea lions

- Goal: **Is there a difference** in metabolic rate among **treatements (categorical X)**
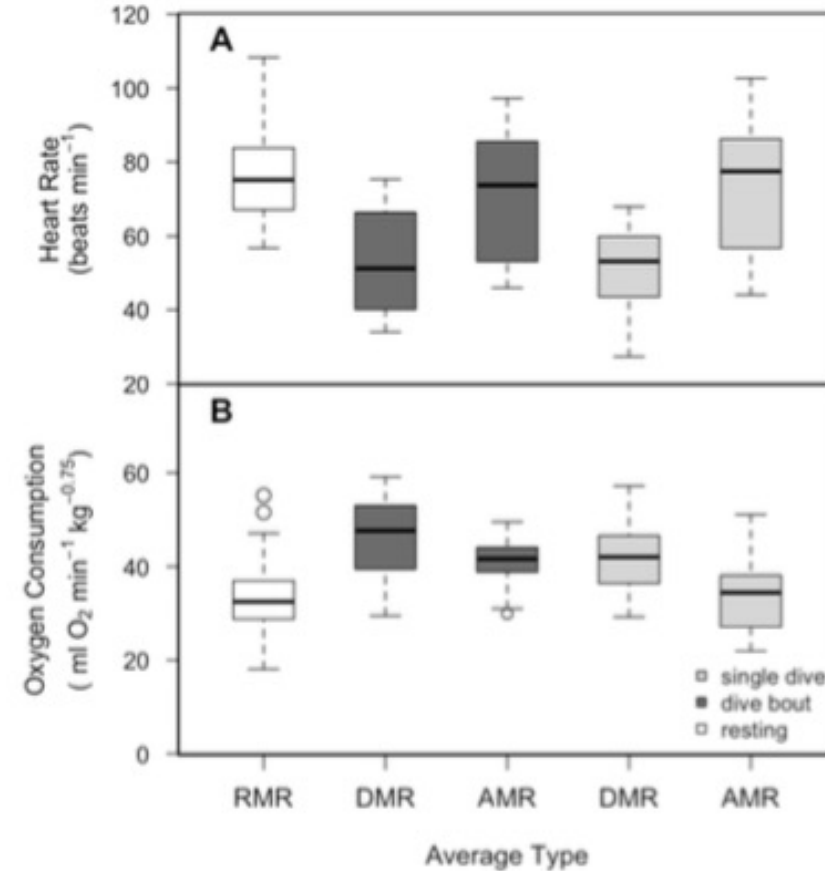
Figure A3. 2 Box-and-whiskers plot showing minimum, median, and maximum heart rate (f_h) and oxygen consumption $\dot{V}_{O_2}$ for animals resting in water (white), averaged over dive bouts (dark grey), or averaged over single dives (light grey). Water composite includes data when animals were resting in a swim mill and resting at the surface of open water. Diving metabolic rate (DMR) includes data averaged over single dives or several dives in dive bout. Average metabolic rate (AMR) includes data averaged over a complete dive cycle (dive +surface interval) or entire dive bout (multiple dives and surface intervals).

Linear mixed-effects model: **predictive**

- Numerical X and Y
- Animal random factor
- Fixed factor –3 level treatment
  - Fasted
  - Fed 4 kg
  - Fed 6 kg
- N=7 Steller sea lions

- Goal: **predictive** equation for metabolic rate
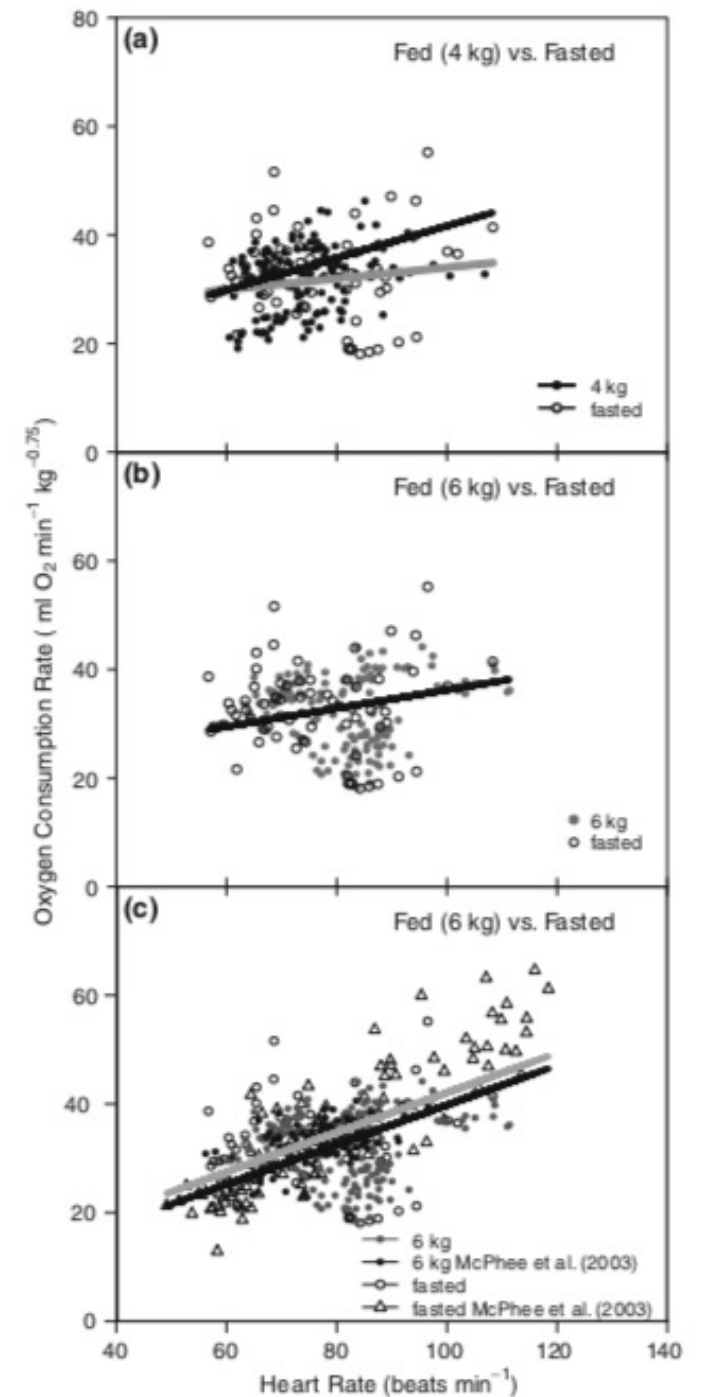
How to show repeated measures data Interaction plots
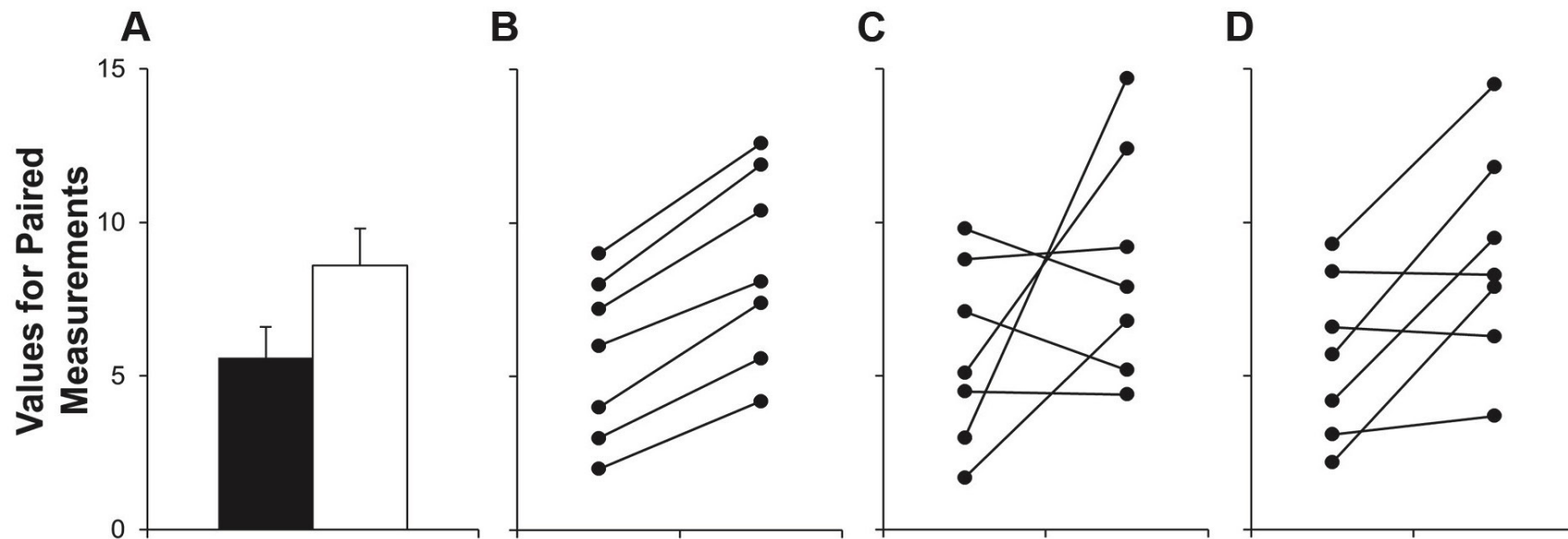
I use this more for exploratory, early LME plots than final plots in published papers

How to show repeated measures data
Strip charts of differences
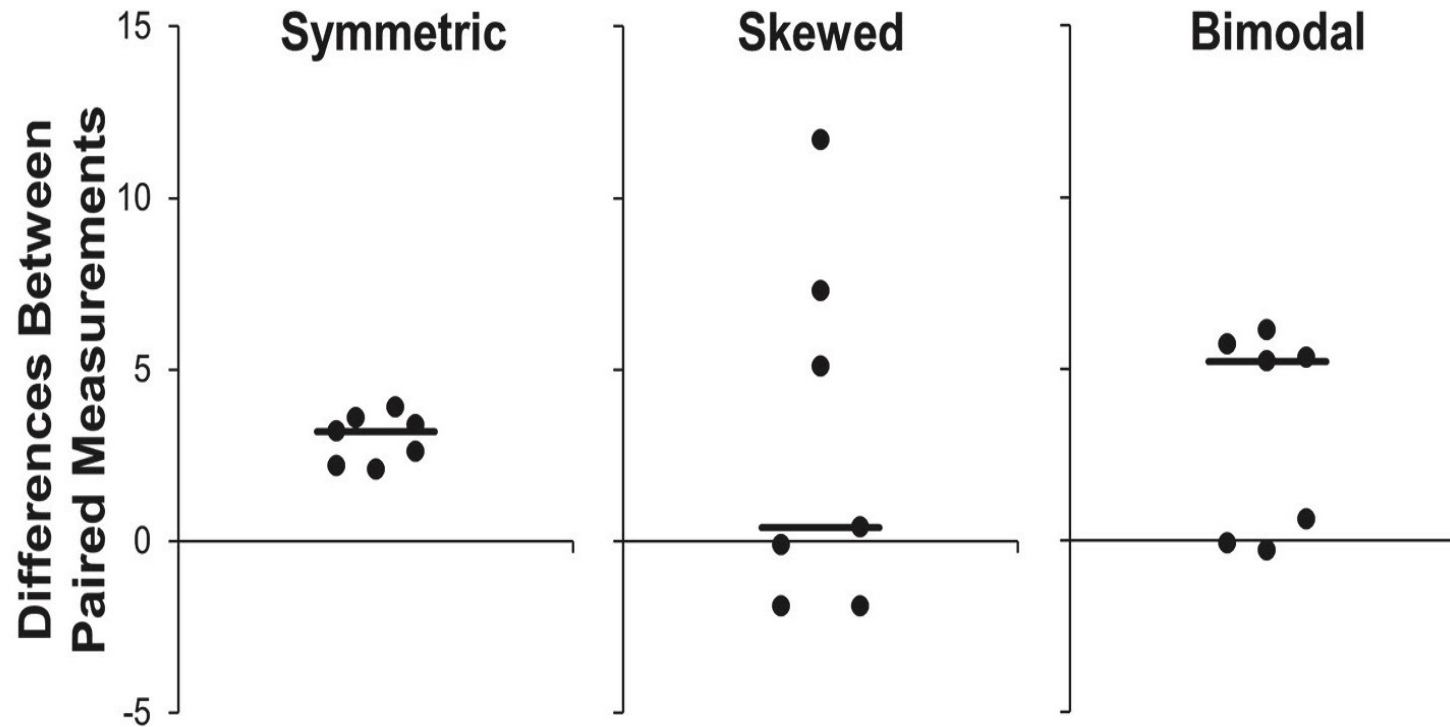
Weissgerber et al. (2015) beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol*. DOI:10.1371/journal.pbio.1002128

Avoid 3D graphs to display frequencies

- 3D makes it more difficult to see area, and hence the frequencies
- A graph that is meaningful only with numbers added is a failure to communicate clearly and effectively

BUT, I often add numbers above bar plots, histograms etc when showing supervisor or presentations (when reader doesn't actually have paper in front of them)

Cawley S, et al. (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116:499-509, Figure

Distribution of All TFBS Regions

Pseudogene/ambiguous 17%

5' to known gene 22%

Within or 3' flanking to a known gene 36%

Novel 24%

866 Total TFBS Regions
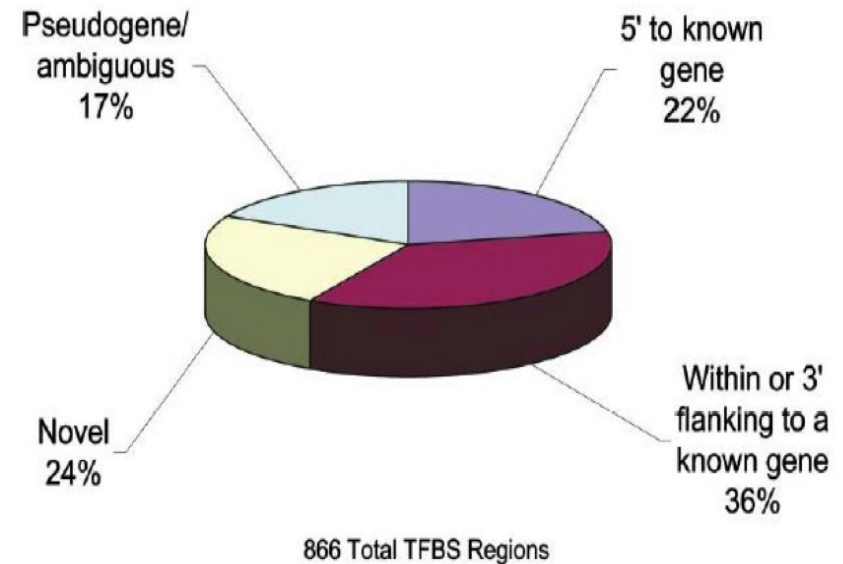
Design graphs to show important trends not plot all of your raw data

Instead: Make use of supplemental material of thesis appendix

Figure 3. Percent of PM$_{2.5}$ composition by component for yearly, winter, and summer averages, by region.

Bell ML, et al. (2007) Spatial and temporal variation in PM$_{2.5}$ chemical composition in the United States for health effects studies. Environmental Health Perspectives 115:989-995

What was the point of 9 pie charts?

- The main patterns of interest were the change in NO_3 and SO_4 between winter and summer, and the consistency of change between geographic regions. This is **not** easy to see ("Huh?" not "Oh!")

- **Try it later:** Design a graph to show the change from summer to winter in NO_3 and SO_4, rather than try to display everything.

Ratio Data (any proportion or %)

- **Problems:**Ratios less than 1 are sandwiched between 0 and 1, distorting magnitudes

- **Possible solutions:** Use log of ratio so that ratios above and below 1 have the same scale.
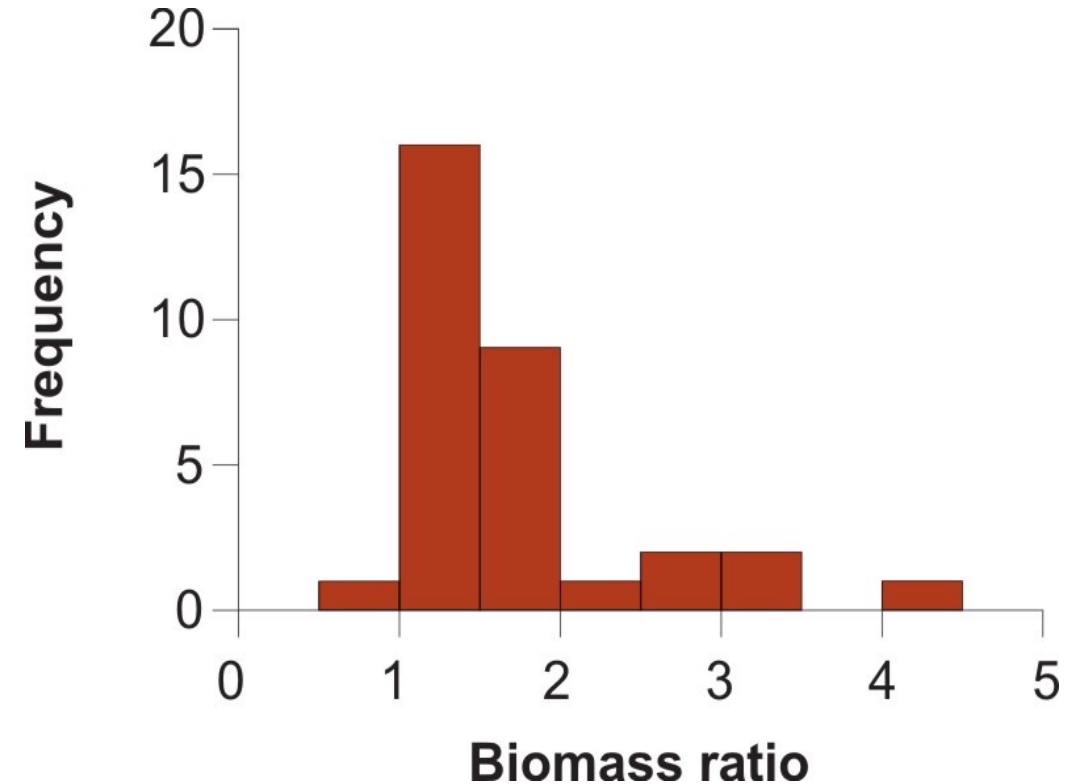
Biomass ratio is the total mass of all marine plants and animals per unit area of marine reserve divided by the same quantity in the unprotected control. N = 32 pairs (reserve and control). Data from Halpern (2003)

Whitlock, M. C. and D. Schluter. 2020. The analysis of biological data. 3rd Ed. Macmillan

Choosing colours for graphs

- Colour blindness (don't mix red and green)
- Publication costs for colour? (minimal online now)
- Most people still print articles in B&W so if use colour make sure it's readable in B/W greyscale
- **Keep colours and symbols for same treatment consistent throughout a manuscript**

R Workshop Tips

Point type cheat sheet (pch) for Base R

R color cheatsheet

Finding a good color scheme for presenting data can be challenging. This color cheatsheet will help!

R uses hexadecimal to represent colors
Hexadecimal is a base-16 number system used to describe color. Red, green, and blue are each represented by two characters (#rrggbb). Each character has 16 possible symbols: 0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F:

"00" can be interpreted as 0.0 and "FF" as 1.0
i.e., red= #FF0000 , black=#000000, white = #FFFFFF

Two additional characters (with the same scale) can be added to the end to describe transparency (#rrggbbaa)

R has 657 built in color names
To see a list of names:
colors()
These colors are displayed on P. 3.

Example:
peachpuff4

R translates various color models to hex, e.g.:
- RGB (red, green, blue): The default intensity scale in R ranges from 0-1; but another commonly used scale is 0-255. This is obtained in R using maxColorValue=255. *alpha* is an optional argument for transparency, with the same intensity scale.
  rgb(r, g, b, maxColorValue=255, alpha=255)
- HSV (hue, saturation, value): values range from 0-1, with optional alpha argument
  hsv(h, s, v, alpha)
- HCL (hue, chroma, luminance): hue describes the color and ranges from 0-360; 0 = red, 120 = green, blue = 240, etc. Range of chroma and luminance depend on hue and each other
  hcl(h, c, l, alpha)

A few notes on HSV/HLC
HSV is a better model for how humans perceive color. HCL can be thought of as a perceptually based version of the HSV model....blah blah blah...

Without delving into color theory: color schemes based on HSV/HLC models generally just look good.
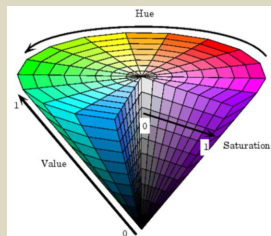
R can translate colors to rgb (this is handy for matching colors in other programs)
col2rgb(c("#FF0000", "blue"))

R Color Palettes
This is for all of you who don't know anything about color theory, and don't care but want some nice colors on your map or figure....NOW!

TIP: When it comes to selecting a color palette, **DO NOT** try to handpick individual colors! You will waste a lot of time and the result will probably not be all that great. R has some good packages for color palettes. Here are some of the options

Packages: grDevices and colorRamps
grDevices comes with the base installation and colorRamps must be installed. Each palette's function has an argument for the number of colors and transparency (*alpha*):

grDevices palettes
cm.colors
topo.colors
terrain.colors
heat.colors
rainbow
see P. 4 for options

heat.colors(4, alpha=1)
> #FF0000FF" "#FF8000FF" "#FFFF00FF" "#FFFF80FF"

For the rainbow palette you can also select start/end color (red = 0, yellow = *1/6*, green = *2/6*, cyan = *3/6*, blue = *4/6* and magenta = *5/6*) and saturation (s) and value (v):
rainbow(n, s = 1, v = 1, start = 0, end = max(1, n - 1)/n, alpha = 1)

Package: RcolorBrewer
This function has an argument for the number of colors and the color palette (see P. 4 for options).
brewer.pal(4, "Set3")
> "#8DD3C7" "#FFFFB3" "#BEBADA" "#FB8072"

To view colorbrewer palettes in R: display.brewer.all(5)
There is also a very nice interactive viewer:
http://colorbrewer2.org/

My Recommendation
Package: colorspace
These color palettes are based on HCL and HSV color models. The results can be very aesthetically pleasing. There are some default palettes:

colorspace default palettes
diverge_hcl
diverge_hsl
terrain_hcl
sequential_hcl
rainbow_hcl

rainbow_hcl(4)
"#E495A5" "#ABB065" "#39BEB1" "#ACA4E2"

However, all palettes are fully customizable:
diverge_hcl(7, h = c(246, 40), c = 96, l = c(65, 90))
Choosing the values *would* be daunting. But there are some recommended palettes in the colorspace documentation. There is also an interactive tool that can be used to obtain a customized palette. To start the tool:
pal <- choose_palette()
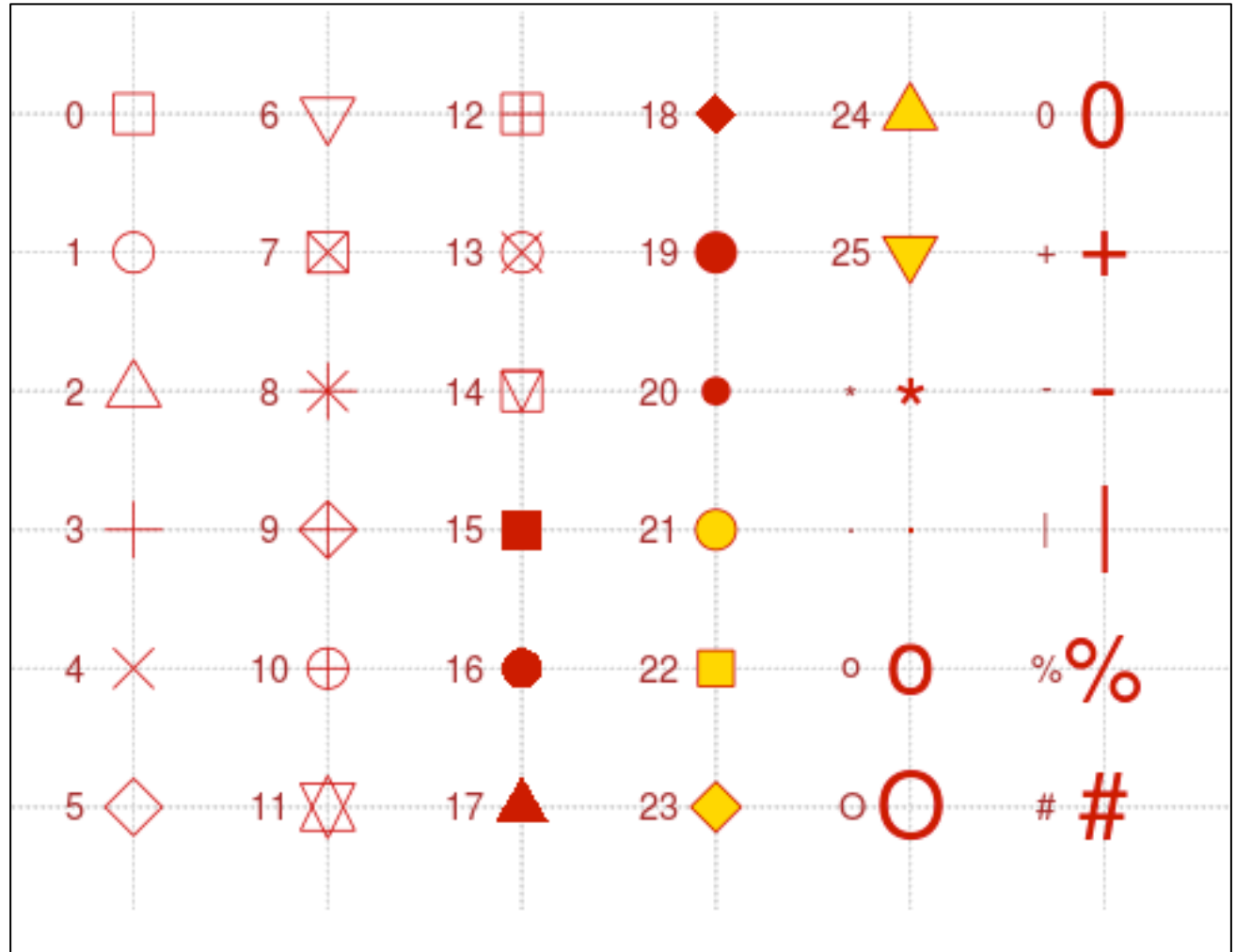
Tables

- General rule of thumb is don't show the exact same data in **both** a table and Figure

- For some kinds of data, they may be the best way to communicate results to a wider audience.

- Use tables to illuminate patterns and show details of numbers you don't want to write fully in text of manuscript

Tables

- Like graphs, tables are used to compare measurements between groups and expose relationships between variables.

- Put tables for storing numbers into online Appendix or Supplement

- Make your tables so that they cause the viewer to go "Oh!" and not "Huh?"

Choosing a graph vs table

- Do you want readers to take away a picture of the overall trend or exact numerical values?

- Graphs show trends and patterns

- Tables show exact numbers→more difficult for our eyes to see patterns in tables

Gastel, B., & Day, R. A. (2016). *How to write and publish a scientific paper* (Eighth ed.). Greenwood, an imprint of ABC-CLIO, LLC.

Discussion: Which drug is more effective? Table or Graph best?

Table 17.1. Effect of streptomycin, isoniazid, and streptomycin plus isoniazid on *Mycobacterium tuberculosis*[a]

Treatment[b]	Percentage of negative cultures at:			
	2 wk	4 wk	6 wk	8 wk
Streptomycin	5	10	15	20
Isoniazid	8	12	15	15
Streptomycin+Isoniazid	30	60	80	100

[a]The patient population, now somewhat less so, was described in a preceding paper (61).
[b]Highest quality available from our supplier (Town Pharmacy, Podunk, IA).
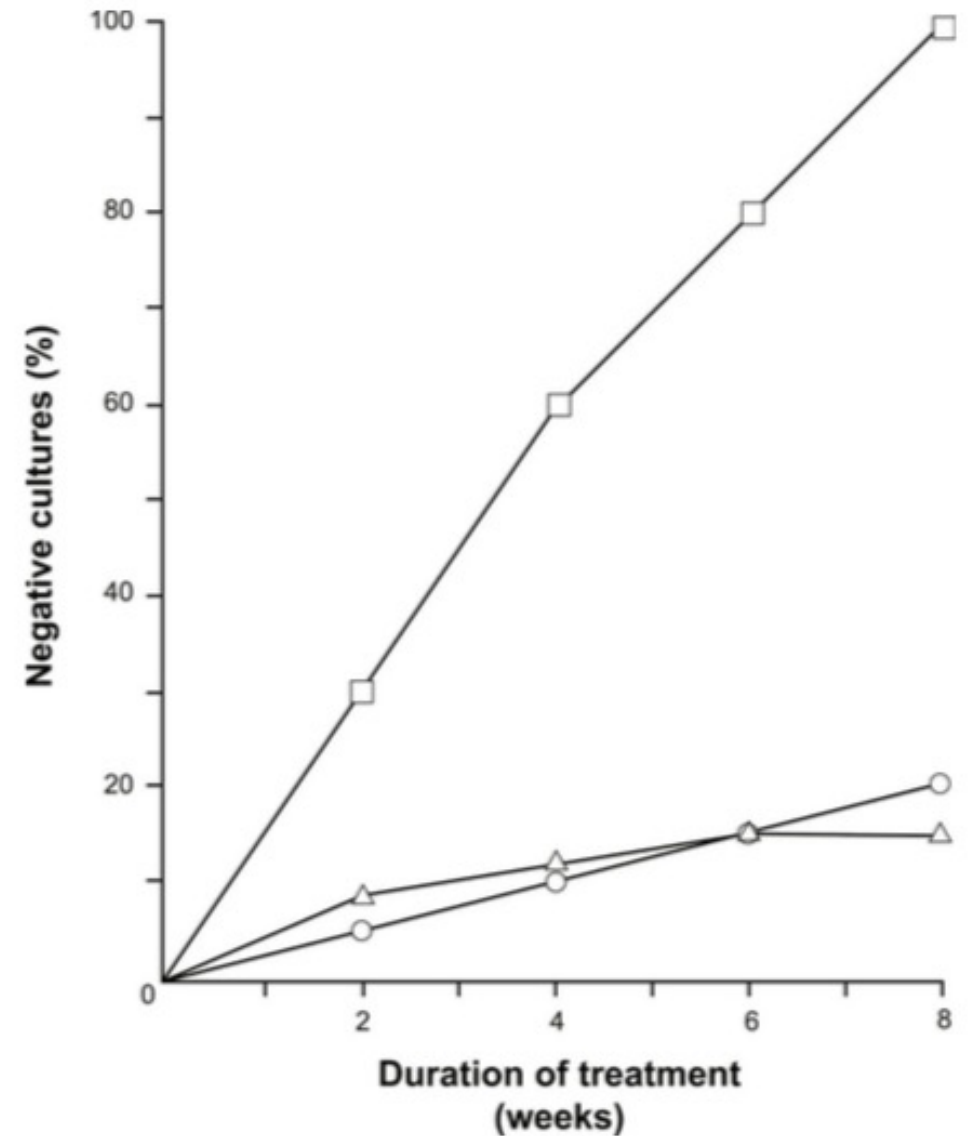
Fig. 17.1. Effect of streptomycin (○) , Isoniazid (△) , and combination of both (☐) on treating tuberculosis

Gastel, B., & Day, R. A. (2016). *How to write and publish a scientific paper* (Eighth ed.). Greenwood, an imprint of ABC-CLIO, LLC.

What do you think of this graph?
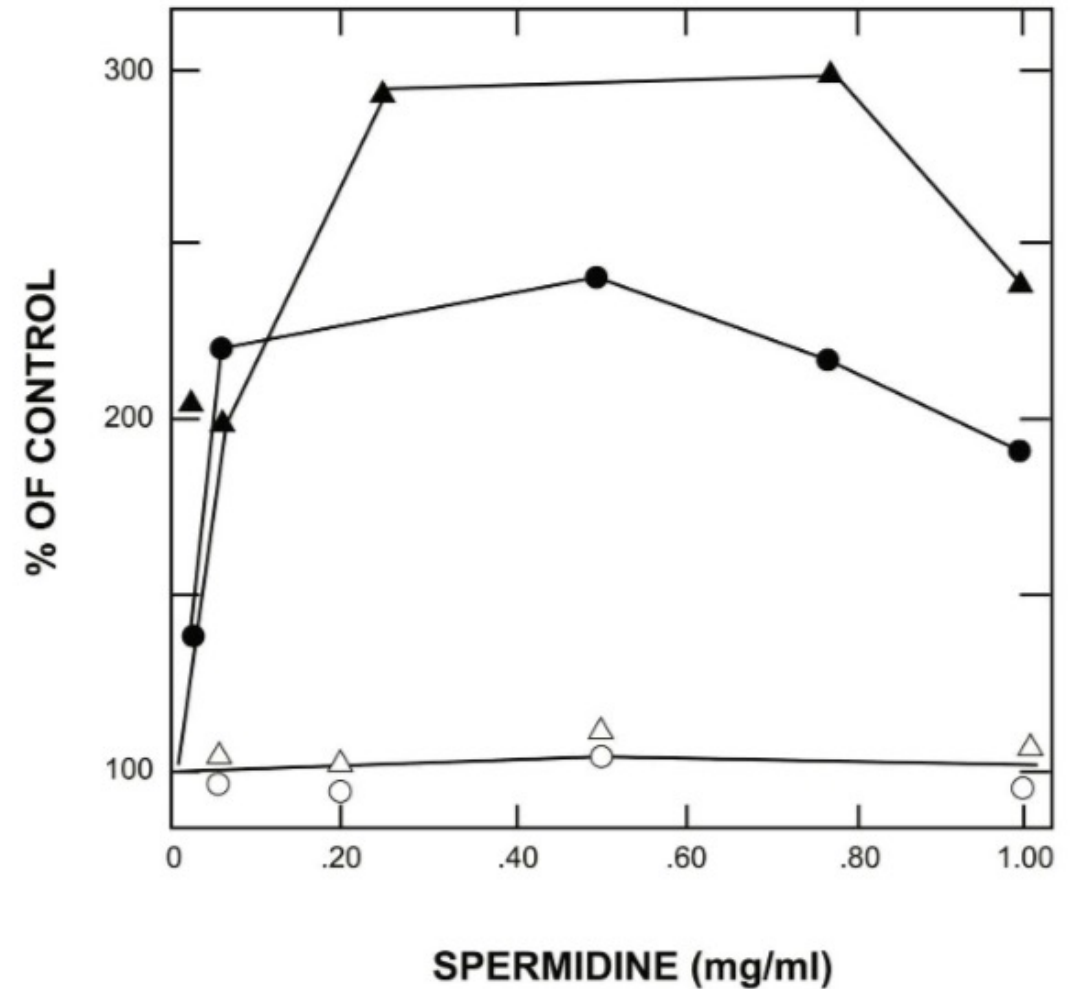
Figure 17.3. Effect of spermidine on the transformation of B. subtilis BR 151. Competent cells were incubated for 40 min with spermidine prior to the addition of 5 μg of donor DNA per ml (●) or 0.5 μg of donor DNA per ml (▲). DNA samples of 5 μg (○) or 0.5 μg per ml (△) were incubated for 20 min prior to the addition of cells. (Redrawn from Fig. 1 in Clark PO and Leach FR. Stimulation of *Bacillus subtilis* transformation by spermidine. Mol. Gen. Genet. 178:21–25, 1980. © by Springer-Verlag 1980. With permission of Springer.)

Gastel, B., & Day, R. A. (2016). *How to write and publish a scientific paper* (Eighth ed.). Greenwood, an imprint of ABC-CLIO, LLC.

Improving Tables Examples

Table 2.5-1 Inbreeding coefficient (*F*) of Spanish Habsburg kings and queens and survival of their progeny.

King/Queen	F	Pregnan-cies	Miscarriages & stillbirths	Neonatal deaths	Later deaths	Survivors to age 10	Survival (total)	Survival (postnatal)
Ferdinand of Aragon								
Elizabeth of Castile	0.039	7	2	0	0	5	0.714	1.000
Philip I								
Joanna I	0.037	6	0	0	0	6	1.000	1.000
Charles I								
Isabella of Portugal	0.123	7	1	1	2	3	0.429	0.600
Philip II								
Elizabeth of Valois	0.008	4	1	1	0	2	0.500	1.000
Anna of Austria	0.218	6	1	0	4	1	0.167	0.200
Philip III								
Margaret of Austria	0.115	8	0	0	3	5	0.625	0.625
Philip IV								
Elizabeth of Bourbon	0.050	7	0	3	2	2	0.286	0.500
Mariana of Austria	0.254	6	0	1	3	2	0.333	0.400

Source: Data are from Alvarez et al. (2009).

Critiquing my own table from PhD thesis

Table 3. Summary of accelerometer error metrics relative to video data calculated on Australian fur seals.

Acceleration	Function 1 Parameters	Averaged over all animals (%) Detection (± SD)		FP rate (± SD)		Precision (± SD)		Total APC
X Surge	Generic	96.3	(3.2)	48.1	(15.3)	51.9	(15.3)	327
	Animal-specific	93.9	(3.0)	42.1	(15.0)	57.9	(15.0)	306
Y Sway	Generic	96.8	(3.2)	58.6	(5.8)	41.4	(5.8)	402
	Animal-specific	91.3	(7.1)	45.1	(18.5)	54.9	(18.5)	361
Z Heave	Generic	82.3	(10.8)	32.8	(13.9)	67.2	(13.9)	262
	Animal-specific	81.7	(9.4)	26.1	(16.1)	73.9	(16.1)	252

Error metrics included detection, false positive rate (FP rate), and precision as defined in the abbreviations. Animal-specific parameters were the parameters of minimum interval and variance threshold that yielded the greatest detection per animal per acceleration axis (Fig 4). Generic parameters (same for all animal) were set at 0.1variance threshold with 5 second minimum interval for all animals and acceleration axis. Data are from the Random Testing subset (n = 97 dives).

- Undefined abbreviation: APC
- Function 1 is not defined

Outline

- Peer-feedback survey on Discussion Moderation and Presenting
- Why make graphs?
- Principles of effective display
- Types of graphs to achieve these principles
- Why some graphs fail, and what can be done
- Tables
- **Interactive Graphs**
- **Ggplot**
- **Motion (video data)**
- **Wrap-up**

Interactive plots

- **Interactive plots**
- Graphs that allow the viewer to manipulate the image.
  - **Example:** Dive Move() for dive depth dataloggers on animals
- Line plot with basic "hover"
  - https://rstudio.github.io/dygraphs/

- Interactive scatter plot:
- https://cengel.github.io/R-data-viz/interactive-graphs.html#plotly

Data that moves: video data

Video graphs for displaying data are becoming more common
- http://www.r-graph-gallery.com/3-r-animated-cube/
- https://ebird.org/science/status-and-trends/

Measles outbreak simulation
- https://www.theguardian.com/society/ng-interactive/2015/feb/05/-sp-watch-how-measles-outbreak-spreads-when-kids-get-vaccinated

The Atlantic Slave Trade in Two Minutes
- http://www.slate.com/articles/life/the_history_of_american_slavery/2015/06/animated_interactive_of_the_history_of_the_atlantic_slave_trade.html?wpsrc=sh_all_dt_fb_top

- The fallen of WWII: https://vimeo.com/128373915

Discussion paper next week

- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54: 187–211

- Download from "**Handouts**" tab on course web site.

- Next Tuesday
  - Presenter: Leah Wessler
  - Moderator: Shawn Nirody

- ****Please add your email to Google Sheets sign up****

- Presenters and moderators are responsible for communicating with each other prior

Assignment #1 Due 28 Jan @ 9pm

- Submit via Canvas only
- See header for Script
- See Rubric on Canvas

- Basic idea:
- Find a poor graph drawn from data and published by your thesis supervisor.
- Analyze it: why does it not succeed?
- Improve the graph using R.
- Analyze it: why does it succeed?
- Include R code **with meaningful annotations and R script header*

Optional Thesis Writing Resource

- Tips for writing the thesis as a whole
- Unique in describing thesis introductory & discussion chapter
- Ask me in workshop if you want an e-copy of scanned book (3rd ed)
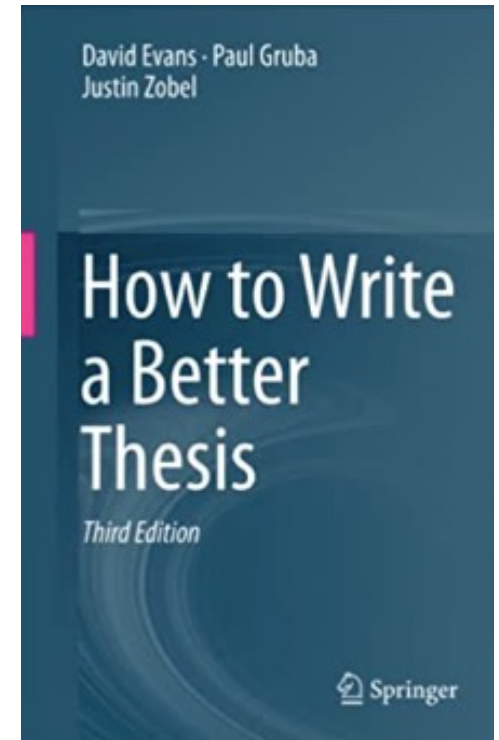
Evans, David, Paul Gruba, and Justin Zobel. *How to write a better thesis*. Melbourne Univ. Publishing, 2011.
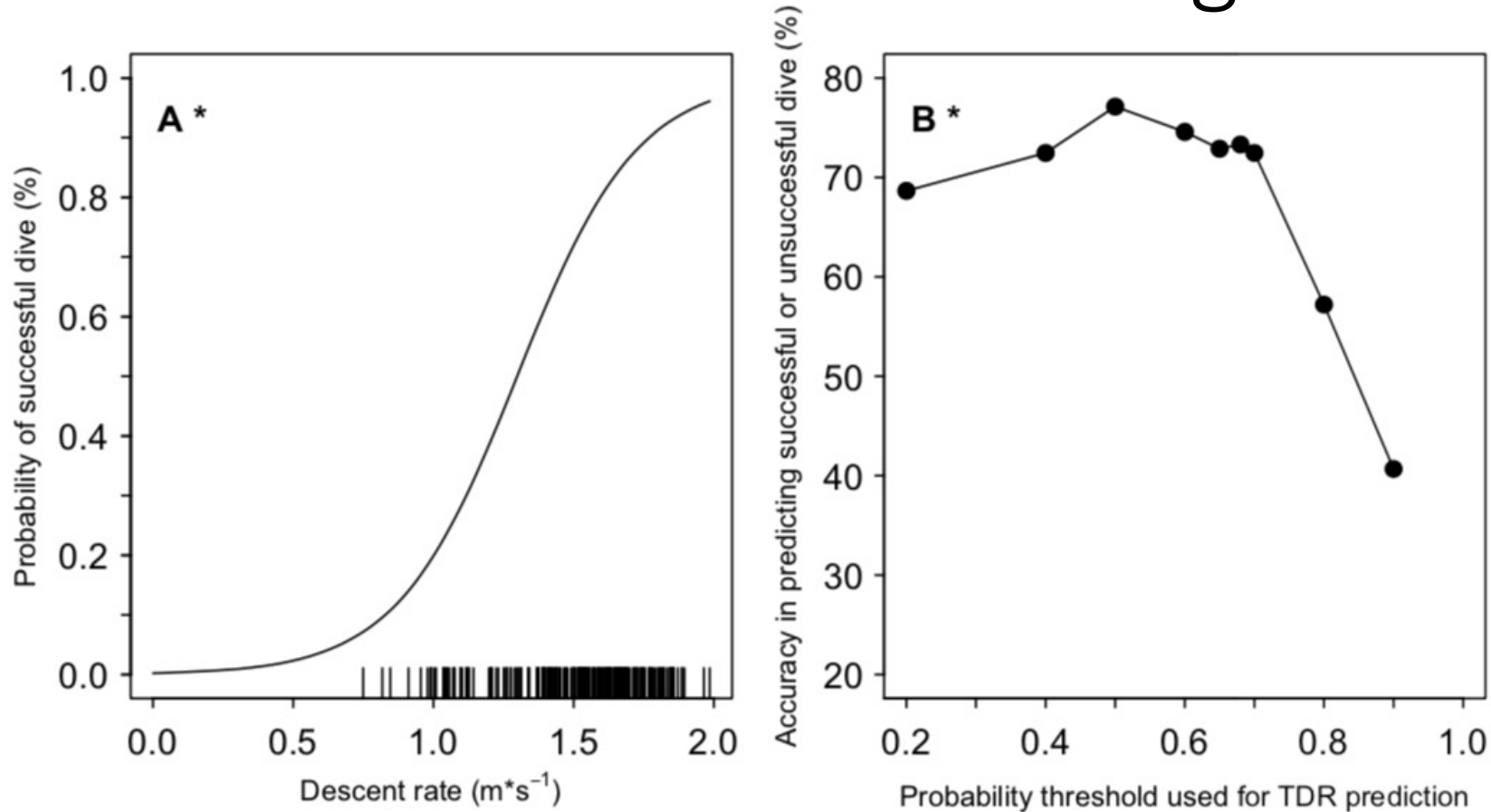
More example graphs
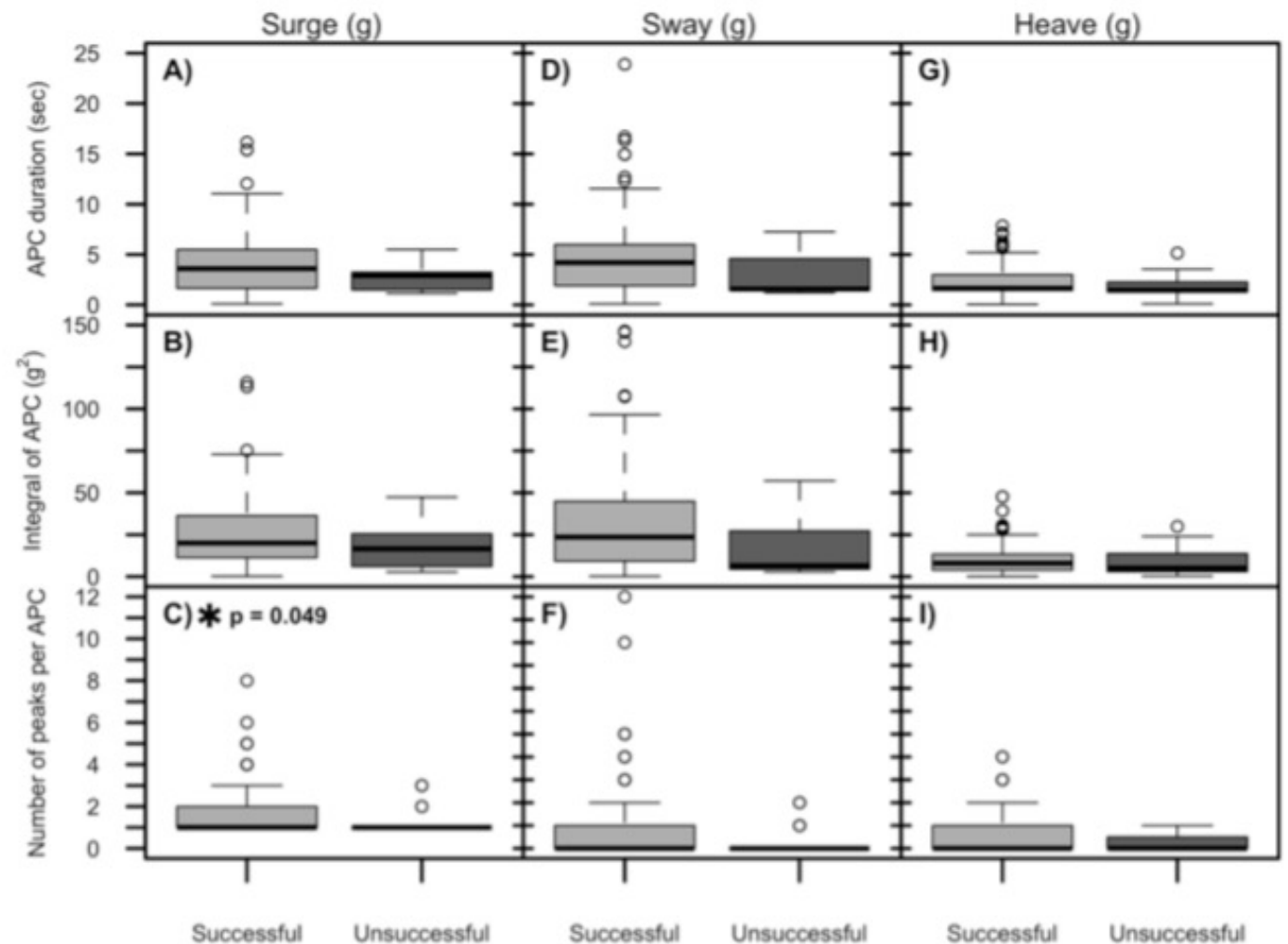
Beth's examples of most used graphs

- Scatter plot with a model line (LME, both X and Y numerical)
- Box-n-whiskers (LME with categorical X variable, numerical Y)
- GLMM Graph example
- GAMM Graph example

Generalized Linear Mixed Model Figure Example

Fig. 2. Probability of a successful dive in response to dive characteristics and accuracy of the GLMM relative to animal-borne video. (A) The most parsimonious model included descent rate as a predictor variable on the training subset (Table 2). Successful dives had at least one successful attempted prey capture (APC) per dive. Distributions are indicated with a rug plots. (B) Accuracy was calculated as the percent of dives correctly predicted as either successful or unsuccessful on the testing subset of dives (Table S2).

Lme4() or mgcs()

Fig 5. Analysis of successful prey captures by accelerometers. The accelerometer was not able to distinguish between successful and unsuccessful attempted prey captures (APC) on any axis of acceleration. Successful events included video where prey was captured, and unsuccessful events included events where the fur seal chased and missed an attempted capture attempt of the prey on video. Neither mean event duration (A, D, G), or integral area under the APC (B, E, H) significantly varied between successful or unsuccessful APC. The number of peaks in surge marginally differed between types of APC (C), but not for sway (F) or heave (I). Results are from Function 1 optimized with generic parameters (0.1 variance and 5 s). Each data point represents an individual APC.