

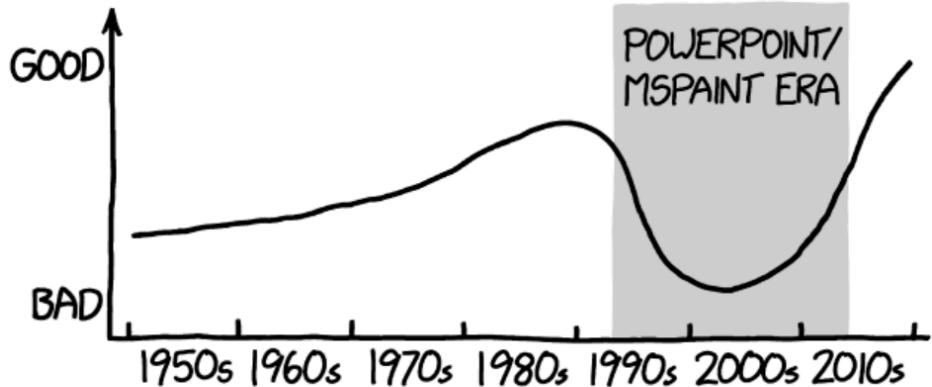
Visualizing Data

Biology 683

Lecture 2

Heath Blackmon

GENERAL QUALITY OF CHARTS AND
GRAPHS IN SCIENTIFIC PAPERS



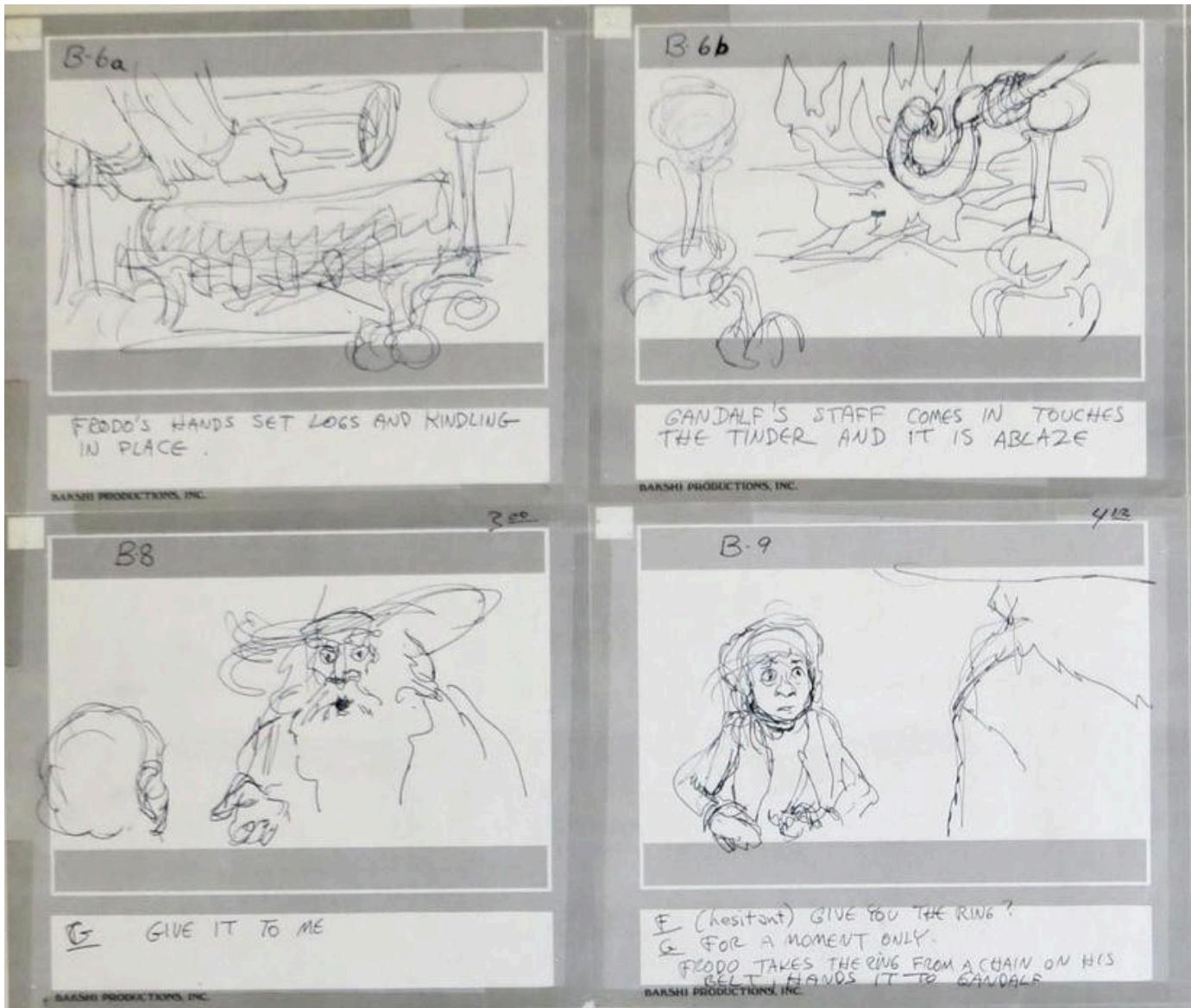
Last week

- What are some causes of the reproducibility crisis?
- Problems with eugenics?
- R basics

Plan for today

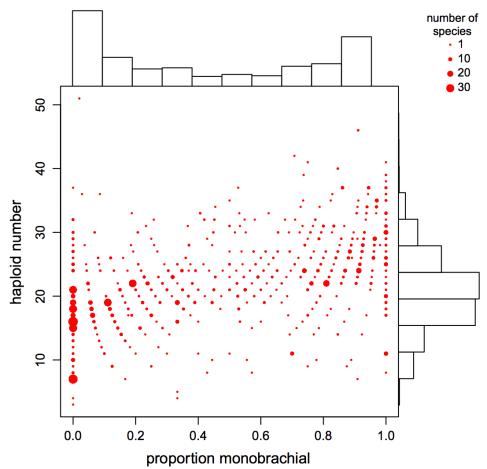
1. Importance of figures
2. General rules for making plots
3. Programs for plotting
4. Resources for plotting

Importance of figures

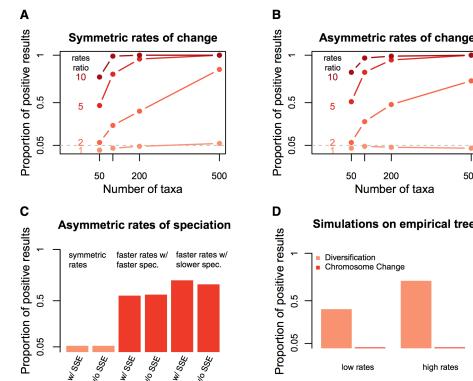
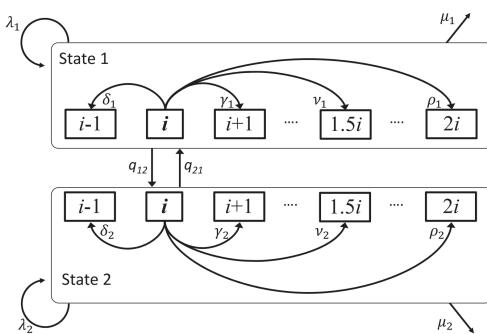


Serve a purpose

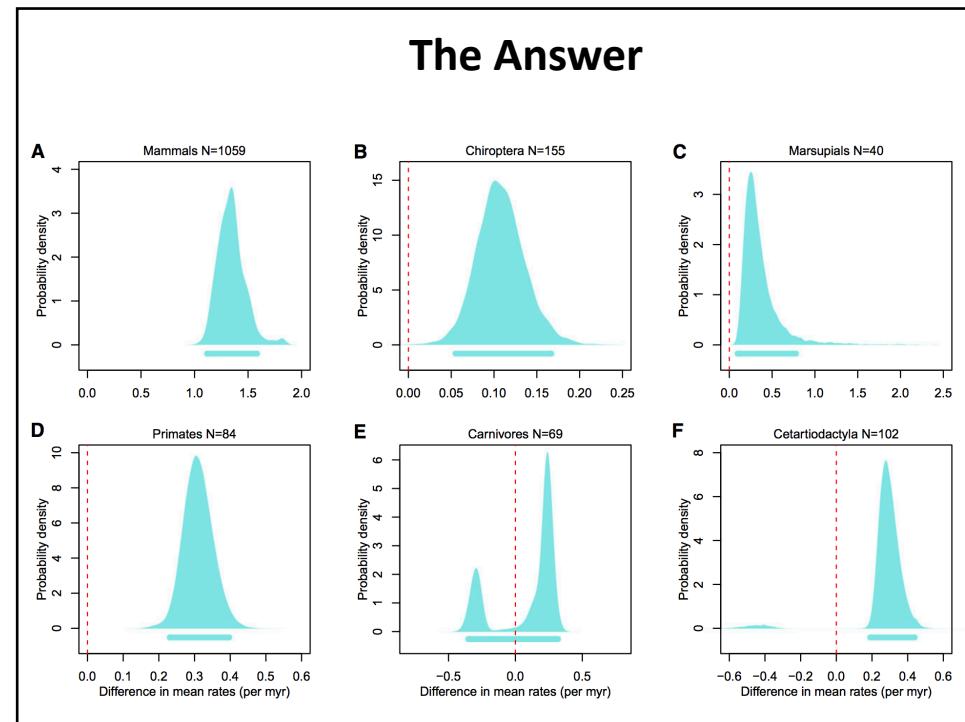
The Problem/Question



New tool and it works



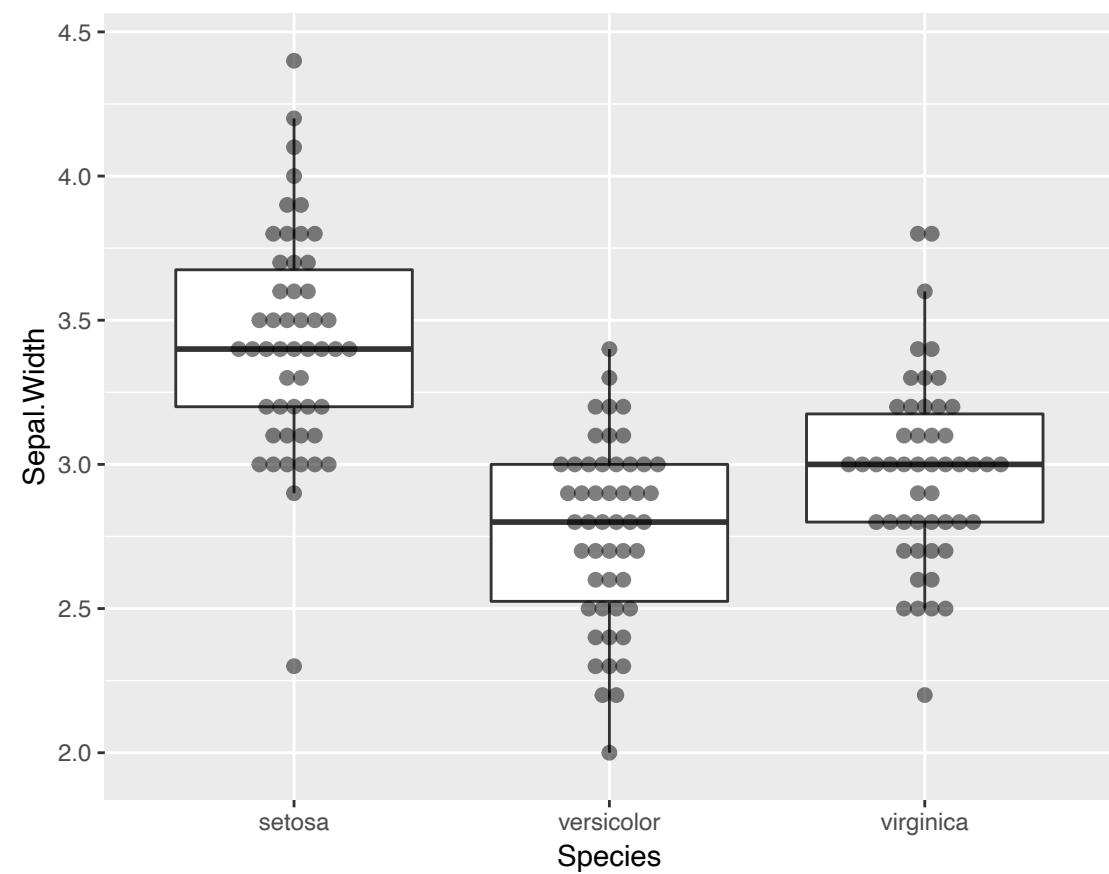
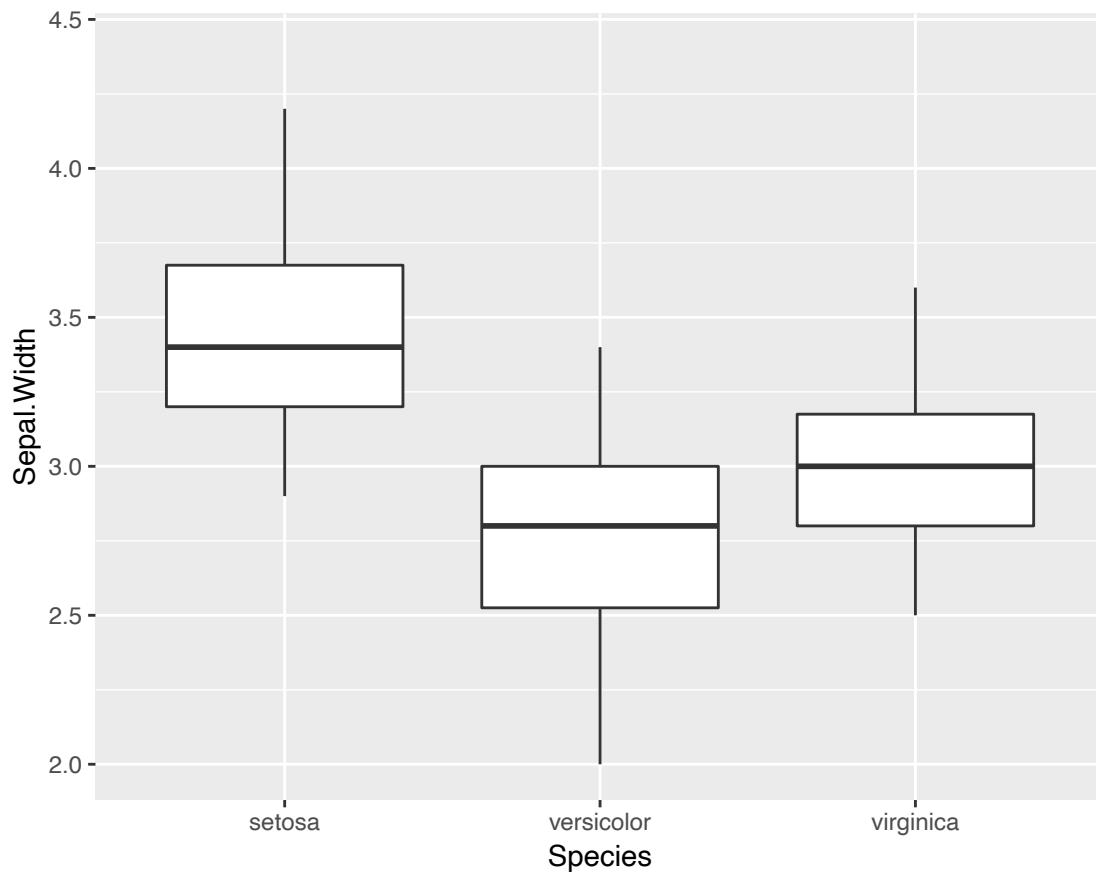
The Answer



Rules for plots

1. Show the data

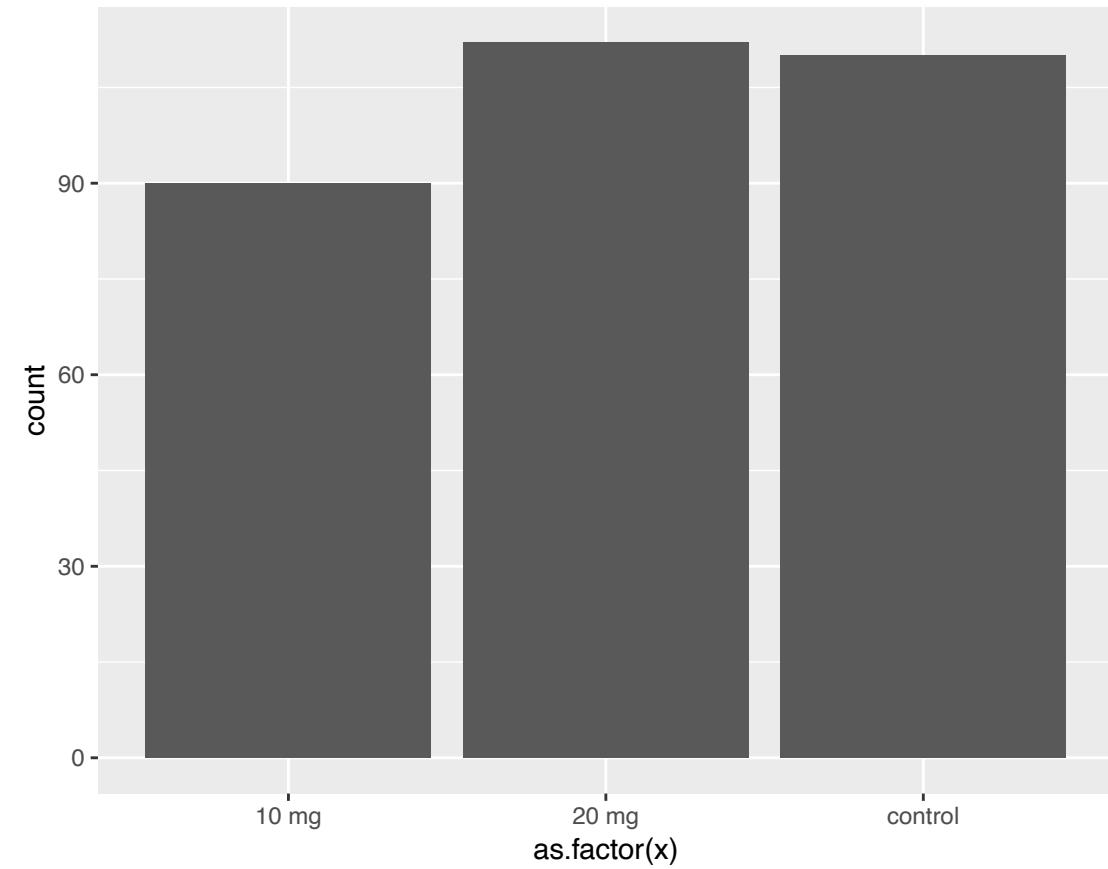
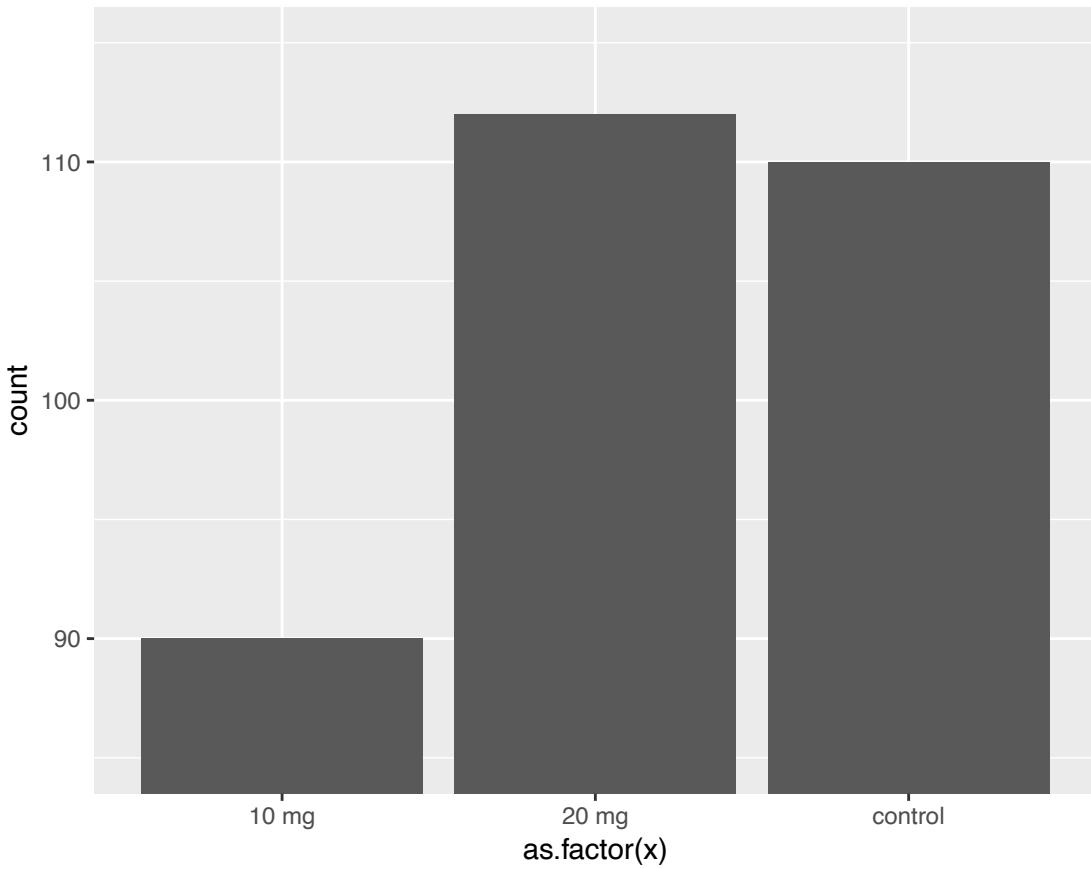
Show the data



Rules for plots

1. Show the data
2. Avoid distorting data

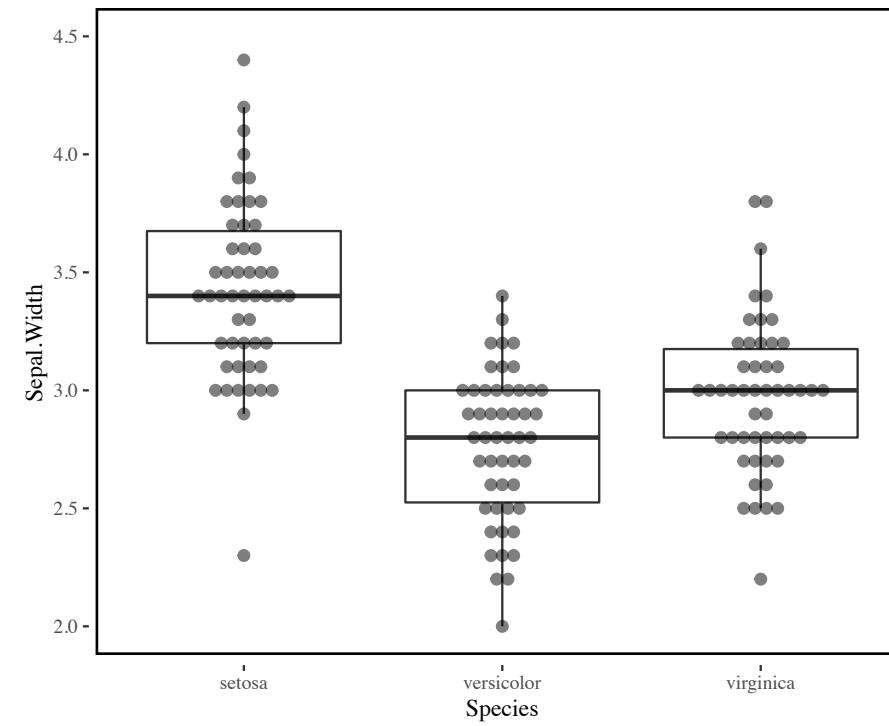
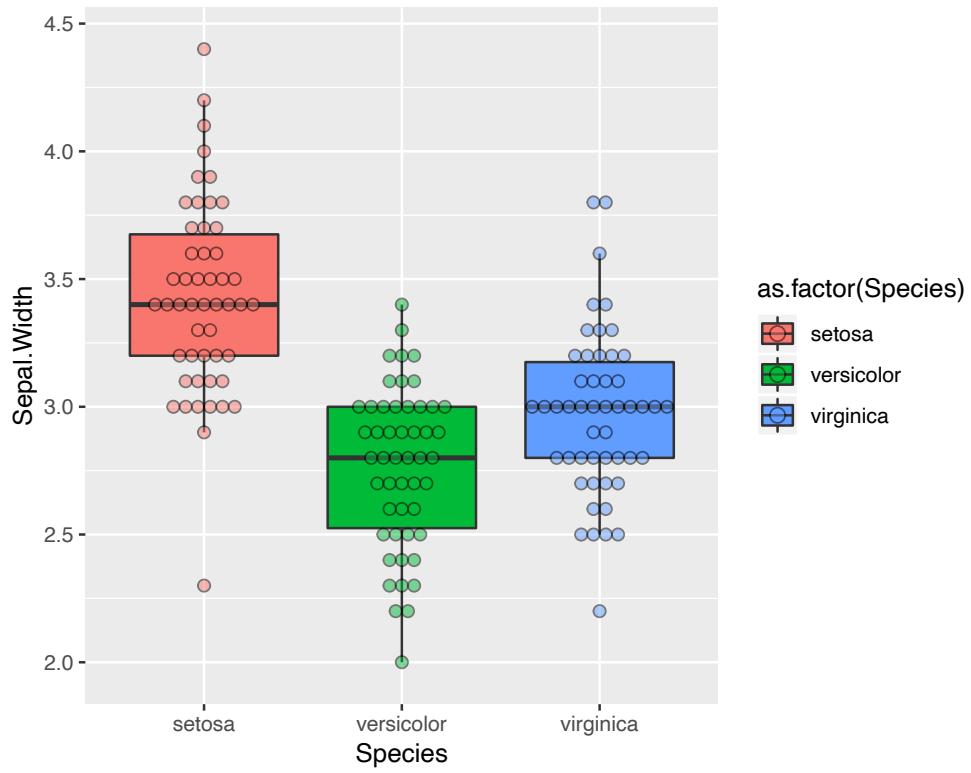
Avoid distorting data



Rules for plots

1. Show the data
2. Avoid distorting data
3. Avoid chart junk

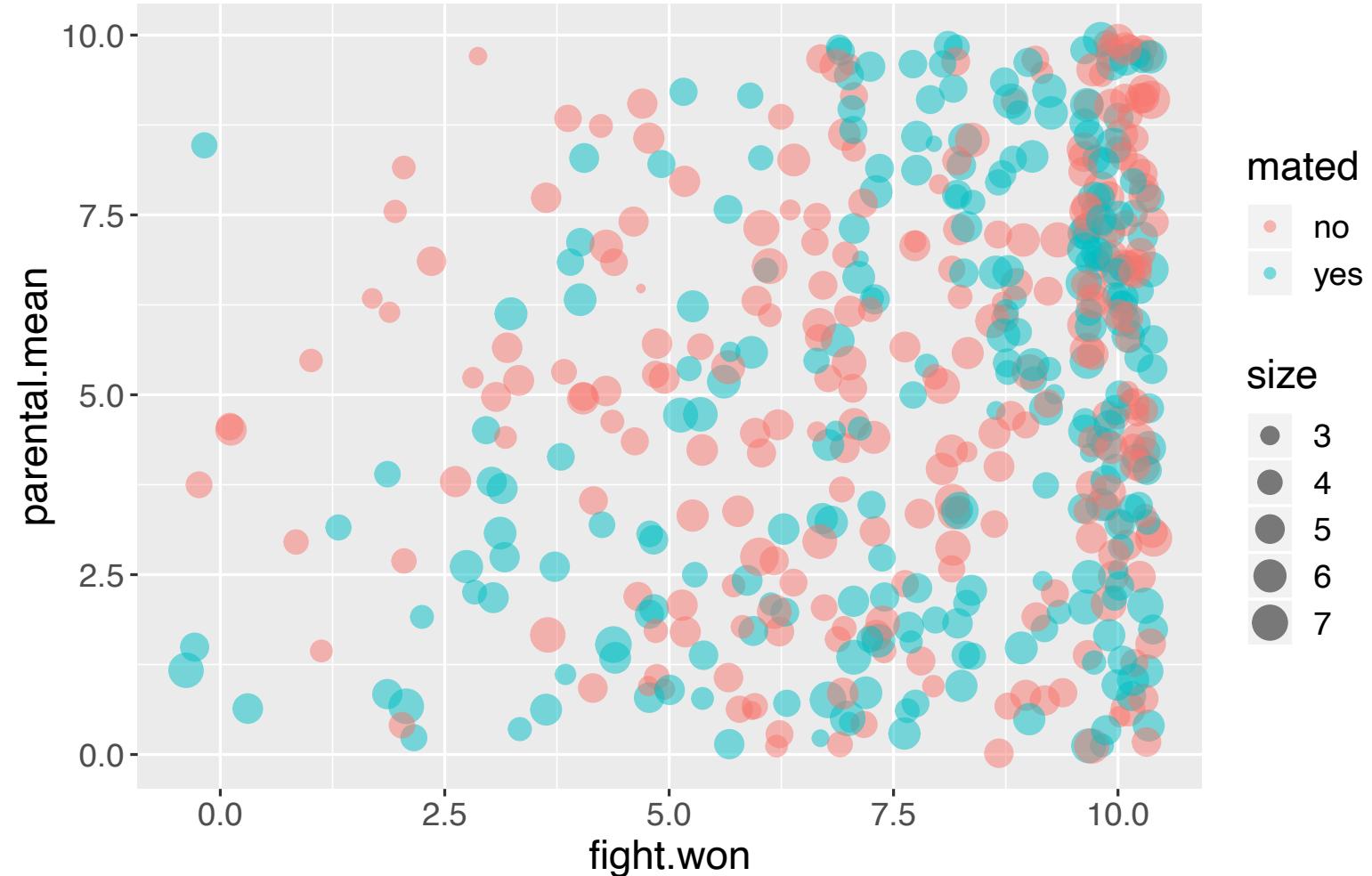
Avoid chart junk



Rules for plots

1. Show the data
2. Avoid distorting data
3. Avoid chart junk
4. Maximize data:ink ratio maximize information

Maximize data:ink ratio maximize information



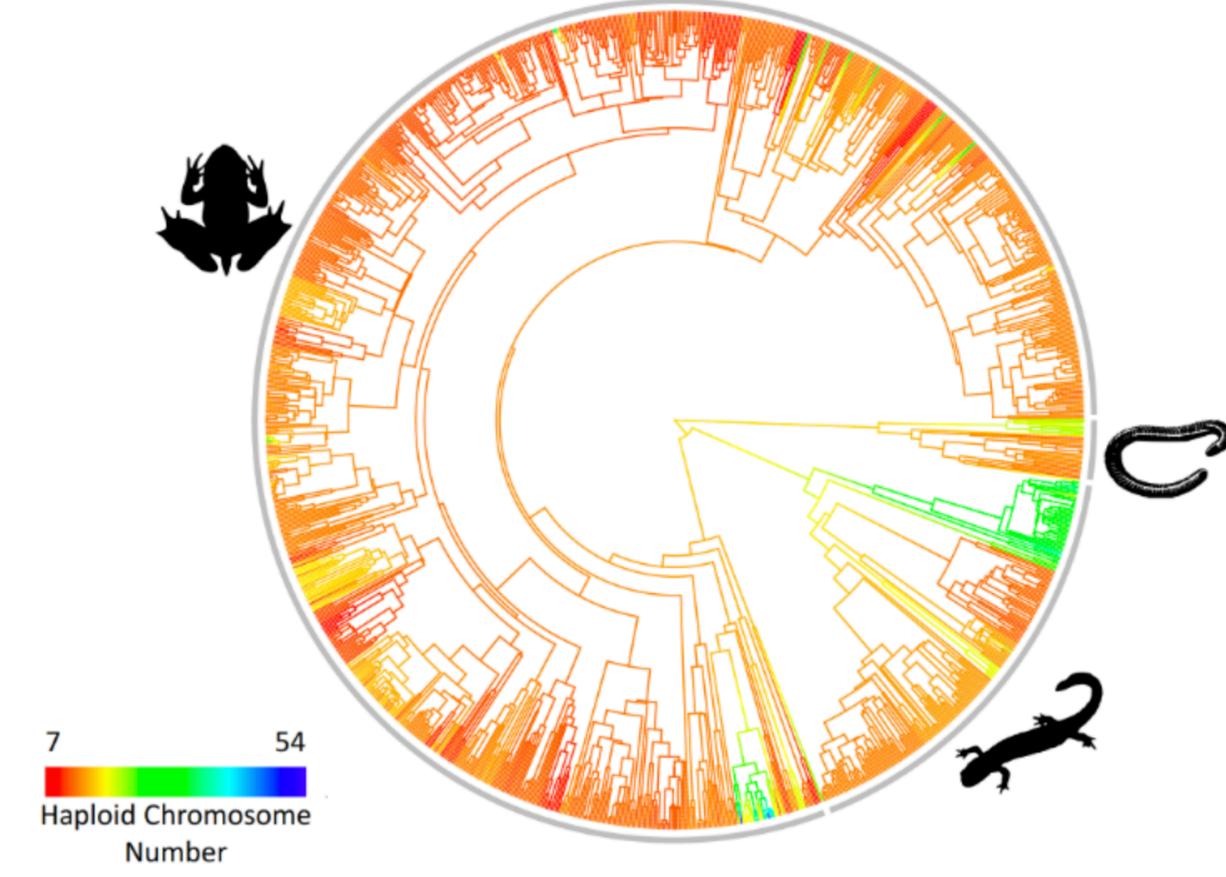
Rules for plots

1. Show the data
2. Avoid distorting data
3. Avoid chart junk
4. Maximize data:ink ratio maximize information

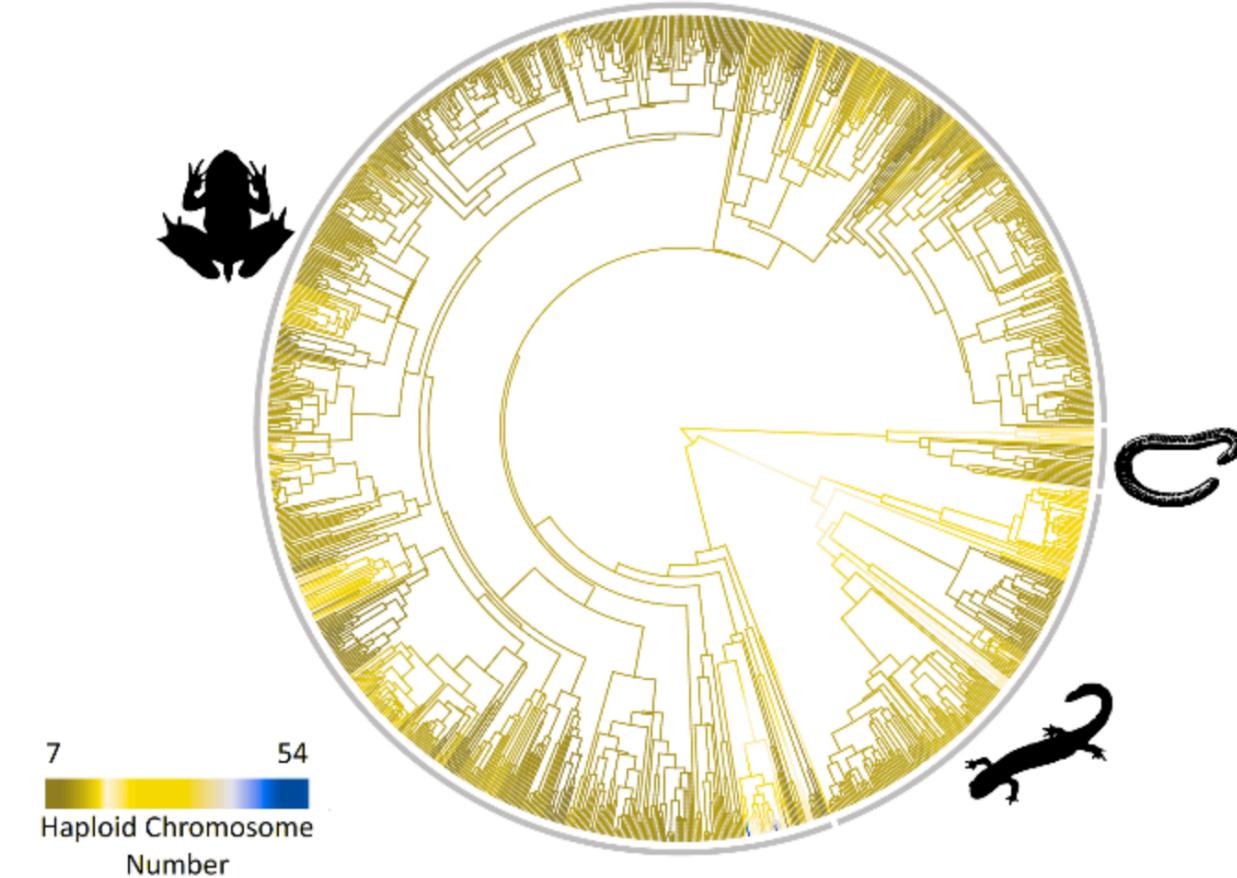
Rules for plots

1. Show the data
2. Avoid distorting data
3. Avoid chart junk
4. Maximize data:ink ratio maximize information
5. Make it accessible to all (5% color blind + BW prints).

Make it accessible to all



Make it accessible to all



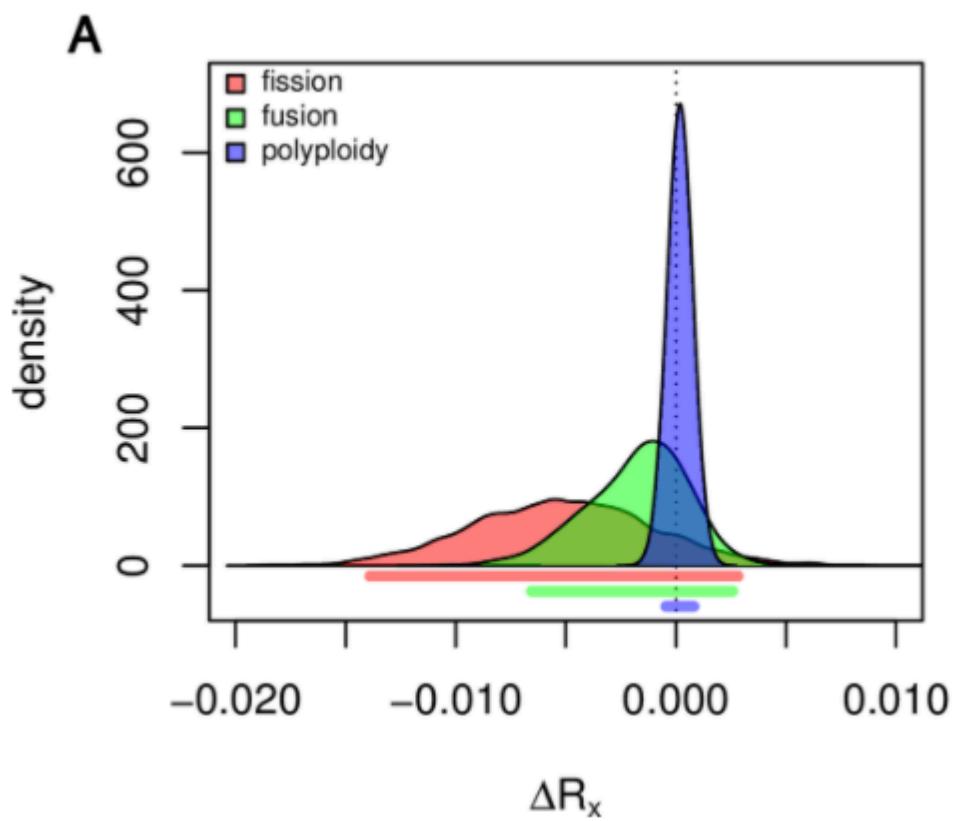
Use the viridis color palette

[Color blind simulator](#)

Rules for plots

1. Show the data
2. Avoid distorting data
3. Avoid chart junk
4. Maximize data:ink ratio maximize information
5. Make it accessible to all (5% color blind + BW prints).
6. Axes and legends that are informative and useful

Axes and legends



Use the right tool

Plotting continuous or discrete data

Matplotlib: is a python plotting library,

R: endless packages for plotting though base and ggplot2 are the most popular.

Inkscape: free vector graphics editor. Can read PDF and let you tweak things you couldn't quite fix in your script generated version.

Circos: created plots that compare genomes but can show connections between any large datasets.

Excel: it is possible but you have very limited options and control

Adobe Illustrator: Similar to inkscape particularly useful for building complex multipart figures

Image Manipulation

GIMP: Think of it as free photoshop. Steep learning curve but can do just about whatever you want

ImageJ: Similar to GIMP but with many packages for analyses of specific image types.

Photoshop: Image manipulation steep learning curve very powerful and definitely worth learning if your research includes figures of FISH results.

General Editing

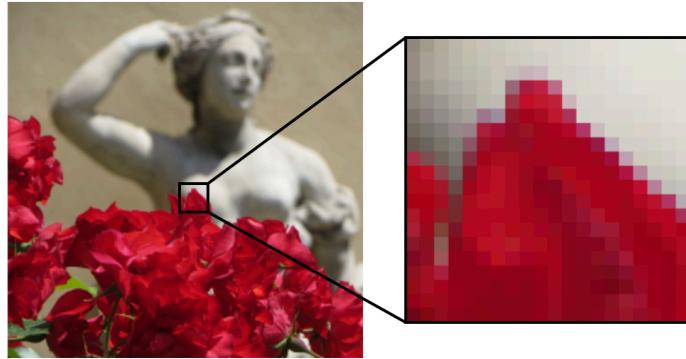
Adobe acrobat: often the easiest way to tweak something small you couldn't get exactly right in your script.

Powerpoint: not elegant but often the easiest way to combine different elements into a single plot.

Bold = Free

Types of images

Raster images



jpg, tiff, png, bmp, raw

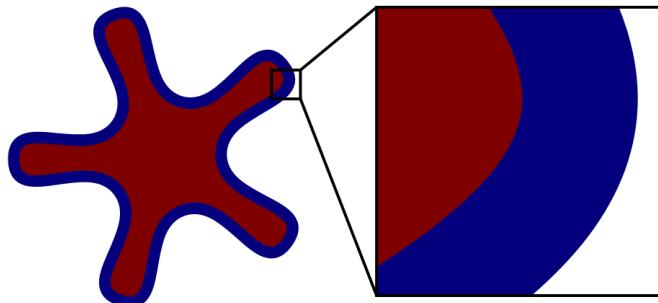
Example of journal requirements

Photos (edit in Adobe photoshop; gimp)

Plots (R; edit in inkscape; Adobe illustrator)

Complex plots (PPT; Adobe illustrator)

Vector images



eps, pdf, svg

Resources

[R graph gallery](#)

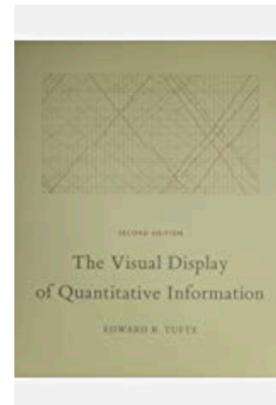
[Python graph gallery](#)

[Color Brewer](#)

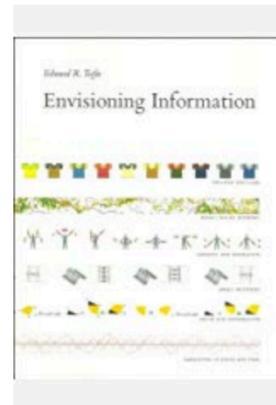
[Plotly](#)

[Data Visualization Catalog](#)

Philosophy of Data Visualization



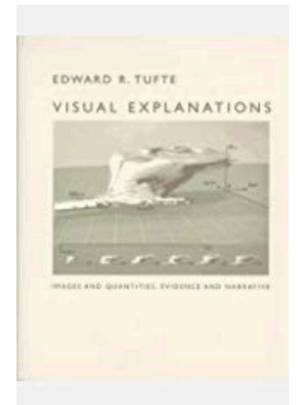
Hardcover
\$34.58



Hardcover
\$36.13



Hardcover
\$41.60



Hardcover
\$35.11

Steps in making a great figure

- 1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.

Steps in making a great figure

- 1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.

Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.

Steps in making a great figure

- 1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.
- 2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.

Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.

Steps in making a great figure

- 1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.
- 2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.

Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.

CLADE	CHROM#	#SP	SCS
ORTHOPTERA	10 - 16		XO
BLATTODEA	7 - 12		XY
Phas.			Parth
"			
"			

9 groups (7-50) 5-6+ 3 states + missing data

Steps in making a great figure

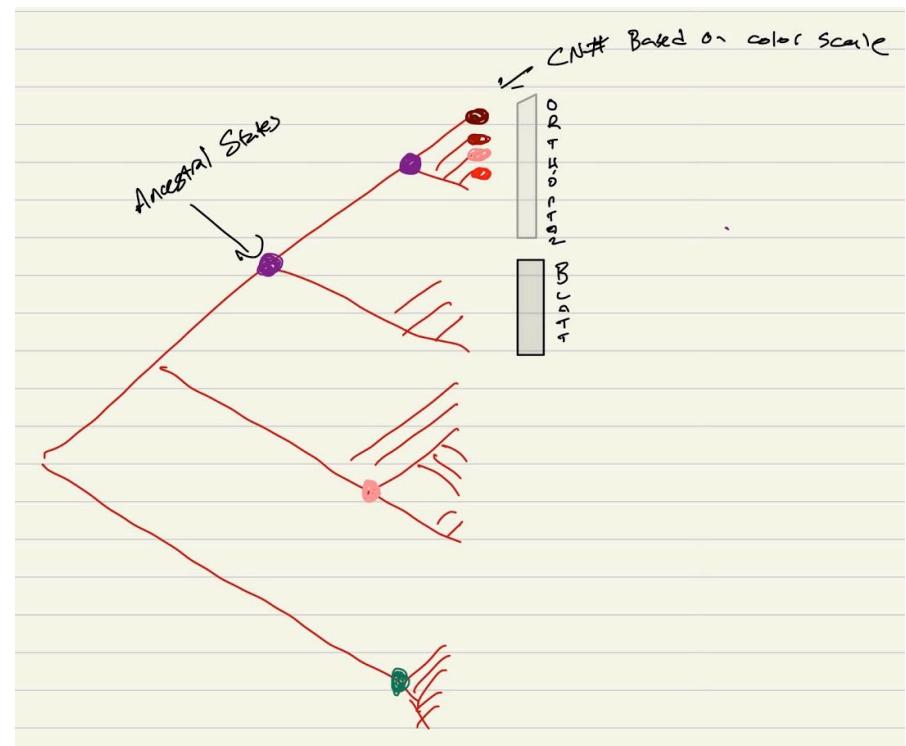
- 1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.
- 2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.
- 3) Sketch out what you think it will look like. Don't waste time figuring out how to make the perfect plot in your tool until you have settled on a best approach. Check out other papers and the graph gallery for ideas.

Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.

Steps in making a great figure

- 1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.
- 2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.
- 3) Sketch out what you think it will look like. Don't waste time figuring out how to make the perfect plot in your tool until you have settled on a best approach. Check out other papers and the graph gallery for ideas.

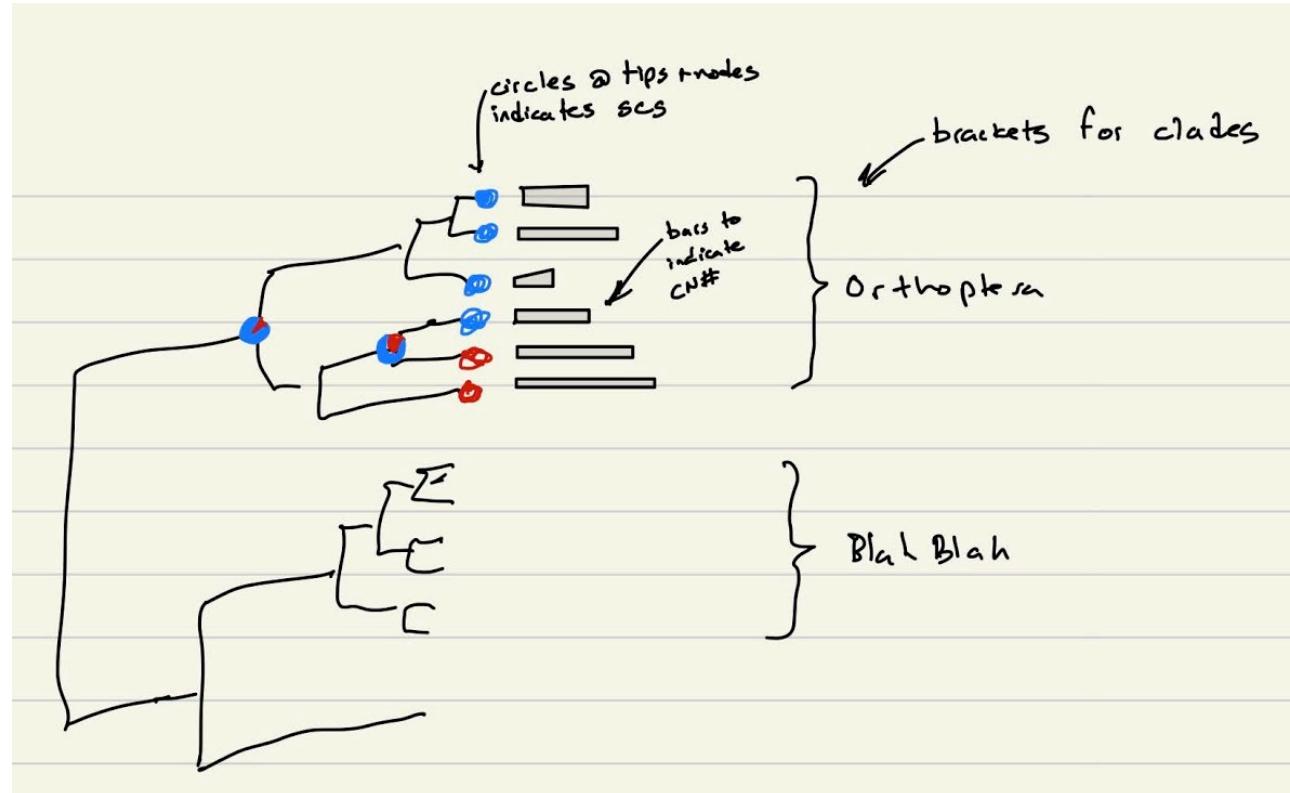
Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.



Steps in making a great figure

- 1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.
- 2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.
- 3) Sketch out what you think it will look like. Don't waste time figuring out how to make the perfect plot in your tool until you have settled on a best approach. Check out other papers and the graph gallery for ideas.

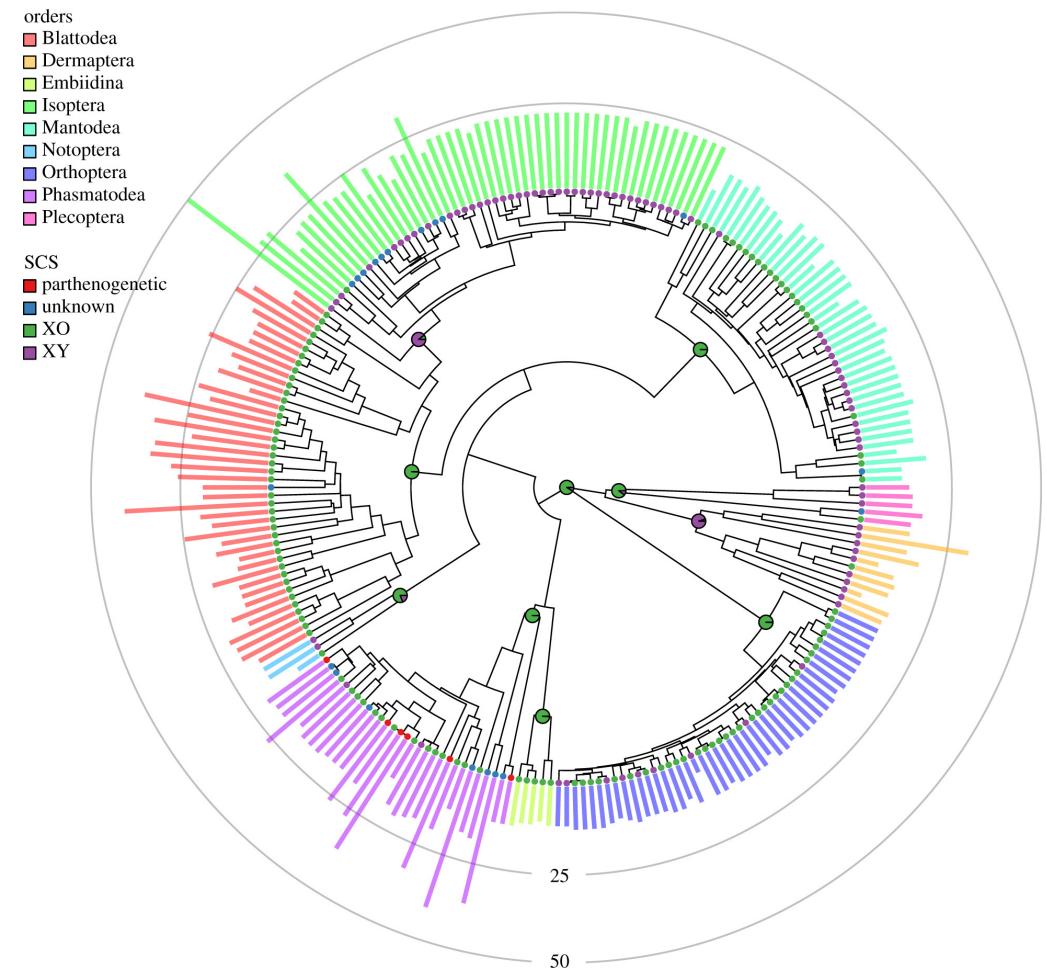
Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.



Steps in making a great figure

- 1) Figure out the purpose of the figure. Usually you will have a sentence in the paper or a point you want to make in a talk.
- 2) Make a list of the data that will need to be included. Is it continuous, discrete, or more complex.
- 3) Sketch out what you think it will look like. Don't waste time figuring out how to make the perfect plot in your tool until you have settled on a best approach. Check out other papers and the graph gallery for ideas.

Haploid chromosome number ranged from 7-50 across sample Polyneoptera species and an XO sex chromosome system was reconstructed as the most probable ancestral state for most orders.



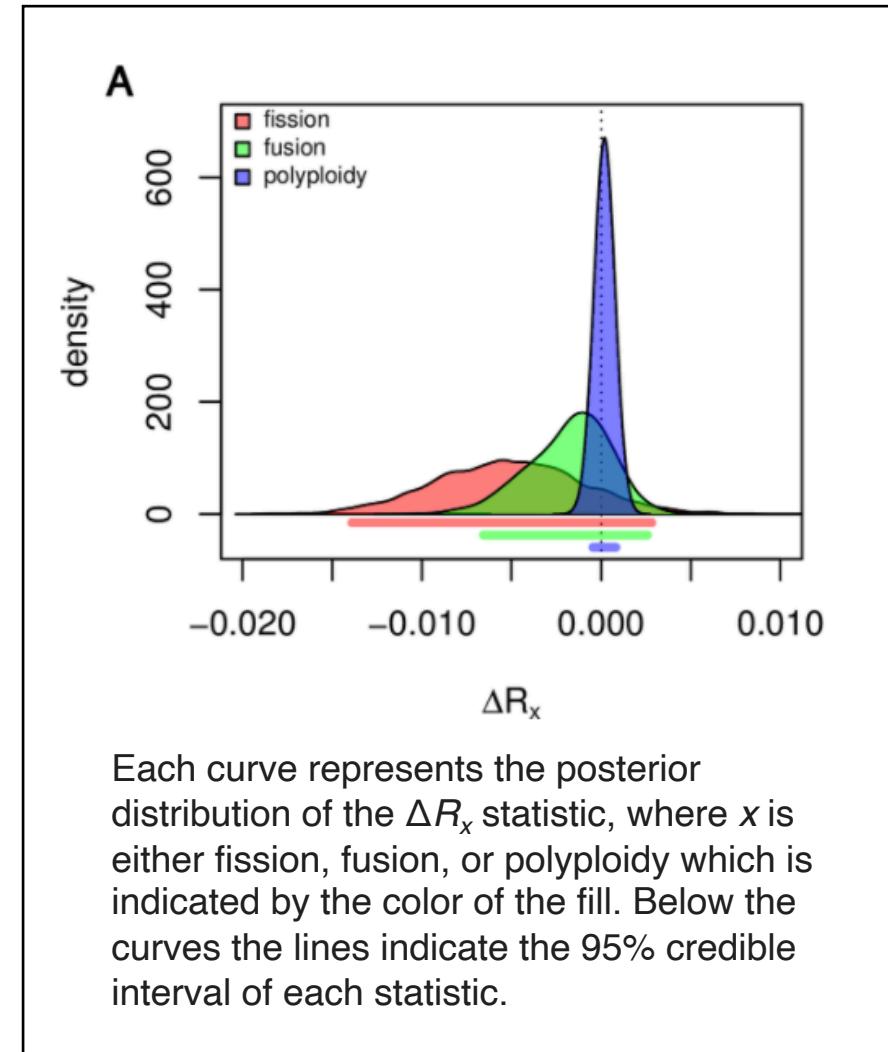
How I make a plot

1) Decide why you are making the plot. Often times you will have a sentence in your paper that demands a plot. On almost every project I've ever been involved in we have much data that is never included in a figure in the final paper.

The ΔR_x statistics for fusions, fissions and polyploidy had credible intervals that overlapped zero ([Fig 2A](#)).

	fission	fusion	polyploidy
1	1.924933e-02	6.691240e-03	1.194787e-02
2	8.055805e-03	-2.395435e-03	1.453424e-02
3	3.010314e-02	2.857327e-02	9.351597e-03
4	-7.035151e-03	2.745761e-02	8.345796e-03
5	5.011324e-04	-1.635201e-02	-1.280441e-02
6	2.711203e-02	-2.817951e-02	5.322630e-02
7	-4.188582e-03	2.202932e-02	-1.179885e-04
8	3.483755e-02	1.055559e-02	-1.843438e-02
9	3.550188e-02	-3.272404e-04	5.312943e-02
10	-1.145644e-02	2.068717e-03	2.317548e-02
11	1.620784e-02	1.637362e-02	1.869329e-03
12	-4.366205e-02	2.012690e-02	-2.899442e-02
13	3.398371e-03	2.127199e-03	-5.653567e-03
14	2.778317e-02	1.356907e-02	4.306500e-03
15	-6.903131e-03	5.692979e-02	3.769614e-03
16	2.218237e-02	8.972304e-05	1.469798e-02

10,000 rows



Plotting in this class

Over the course of the semester we will plot using several approaches (base, ggplot, custom packages)

Thursday

Intro to base plotting

Plot1: a single continuous

Plot2: 2 continuous variables (predictor and response)

Plot3: 1 continuous variable 1 binary discrete

Plot4: 1 continuous variable measured under 3 conditions

Functions

Base functions

plot

hist

barplot

boxplot

polygon

density

abline

lines

arrows

many associated arguments

xlab

ylab

main

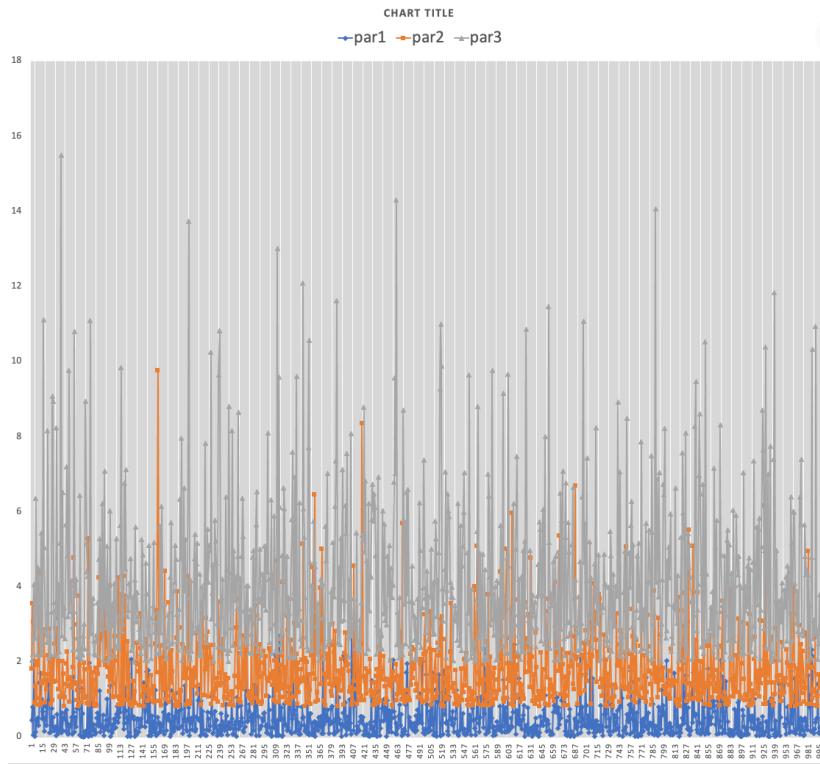
lty

lwd

pch

Homework 2

Download data from course website and make a better plot in R.



Make a better plot and save with the following attributes:

4"x4"

PDF – vector art

Follow the advice discussed in class.

Submit via blackboard

	par1	par2	par3
1	0.4402419375	1.8030852	2.045419
2	0.7925005285	3.5431847	2.209240
3	0.0469341415	3.0718043	2.606824
4	2.0172616494	2.9845387	4.064599
5	0.0072098206	0.9044033	2.626368
6	0.1238258580	1.3351645	6.353806
7	0.4408873200	0.8609138	3.433532
.....			
999	0.492969406	2.2664010	5.417992
1000	0.143511845	0.8638361	2.924785

Here is the text that we want to support with a figure of this data:

We estimated the rate parameters in our model using a Bayesian MCMC approach. Each estimate is based on 1000 random draws from the posterior distribution. We found that the 95% HPD of parameters one and three did not overlap.