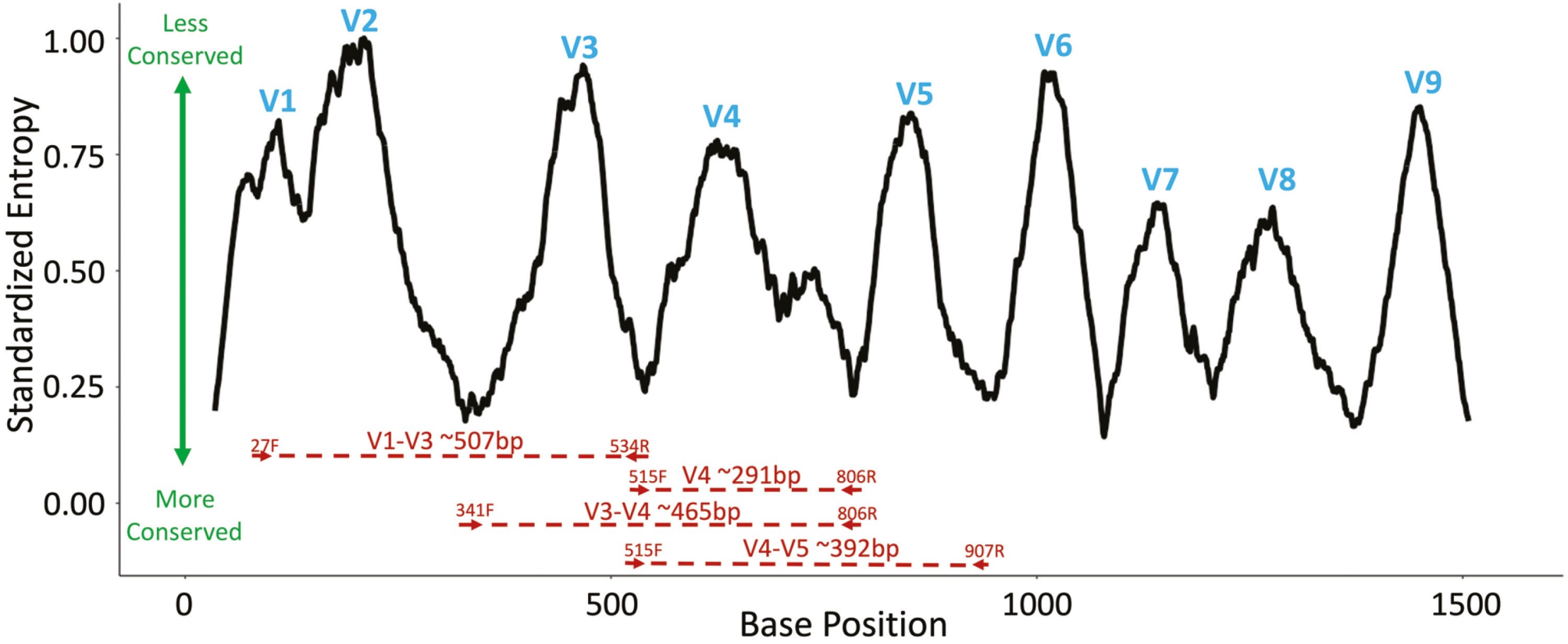
Why do we look at the 16S gene?

Why do we look at the 16S gene?

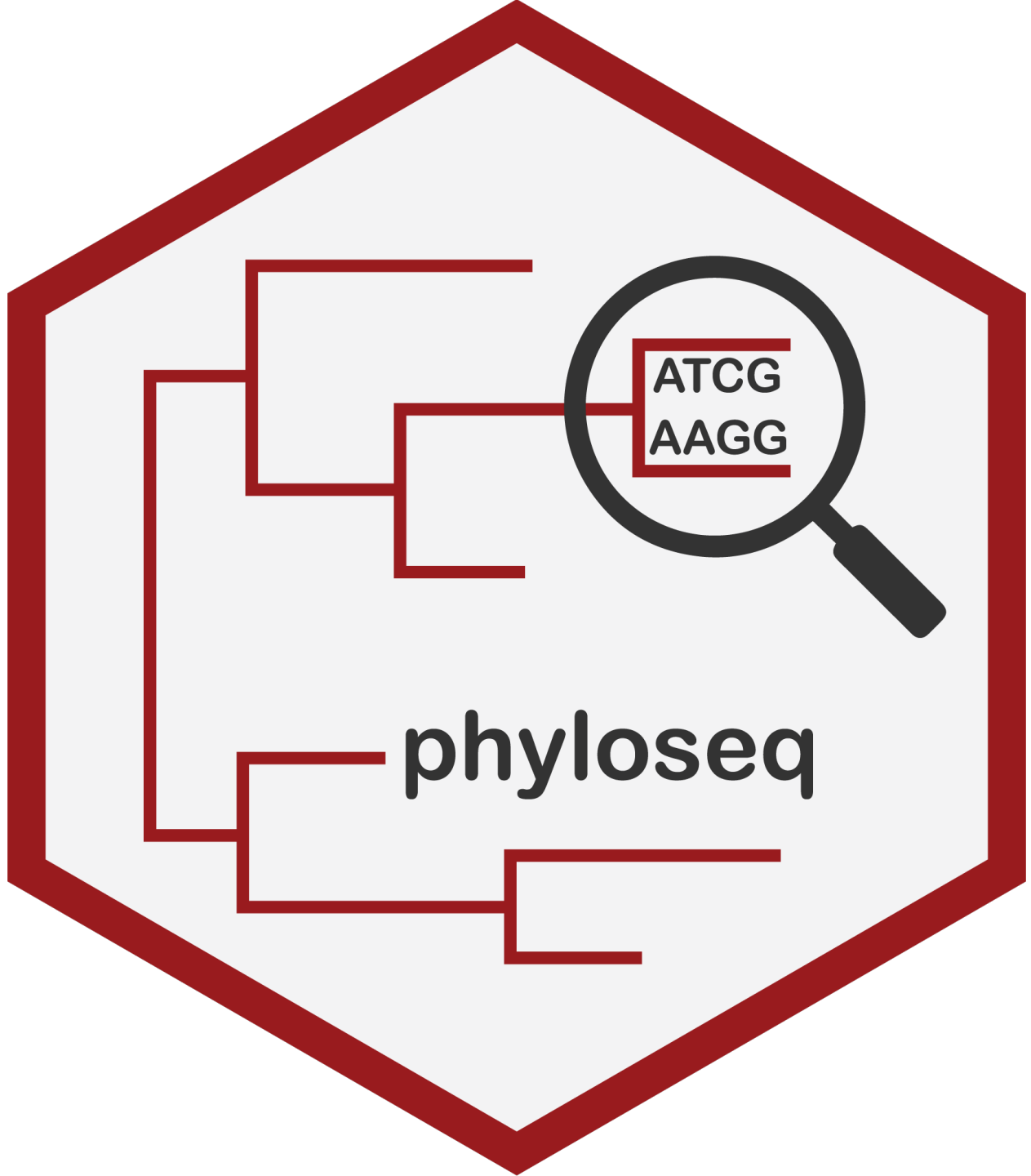


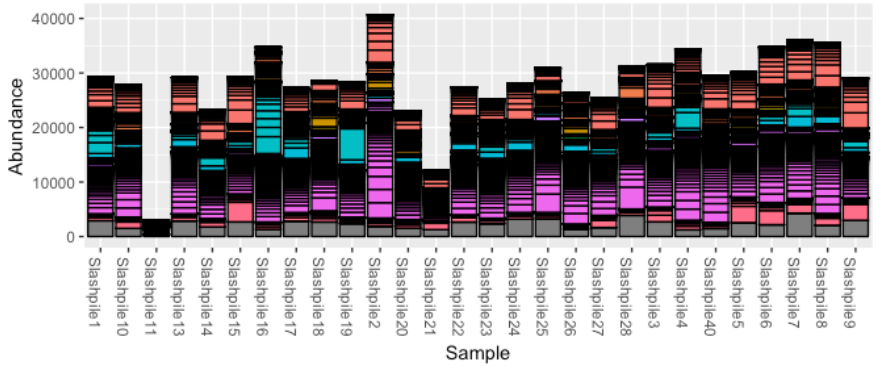
Time to make pretty things!

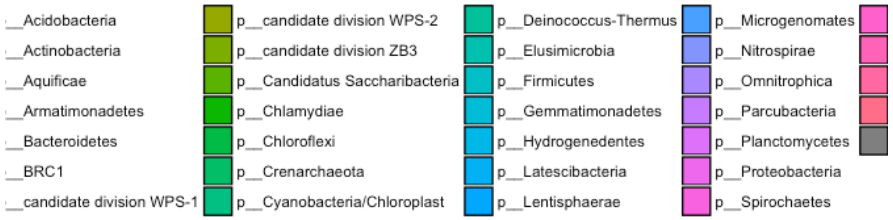
raw data

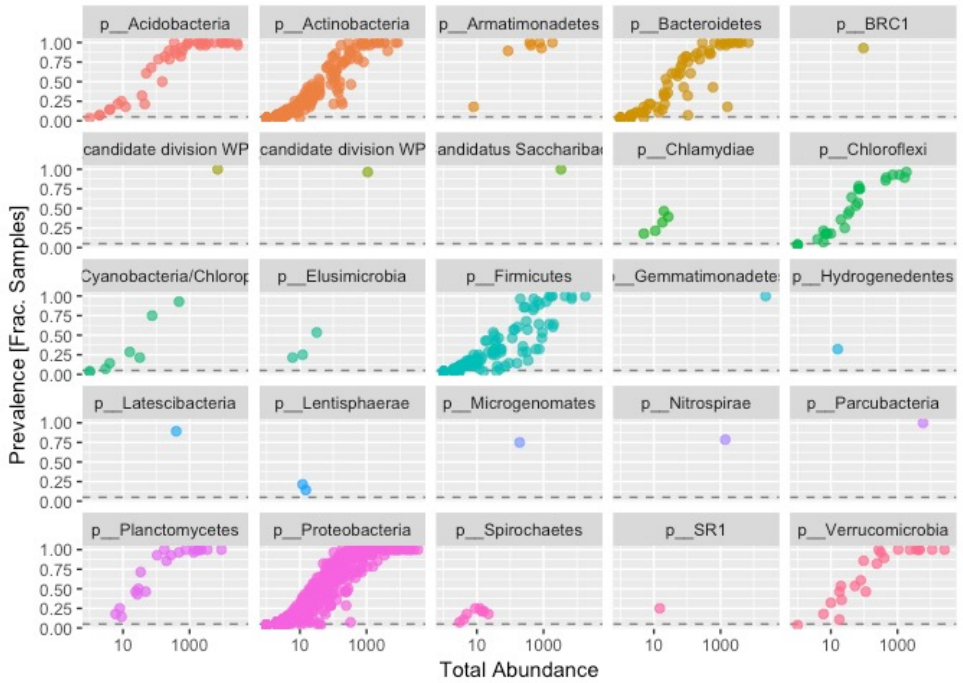
data processing / pipeline

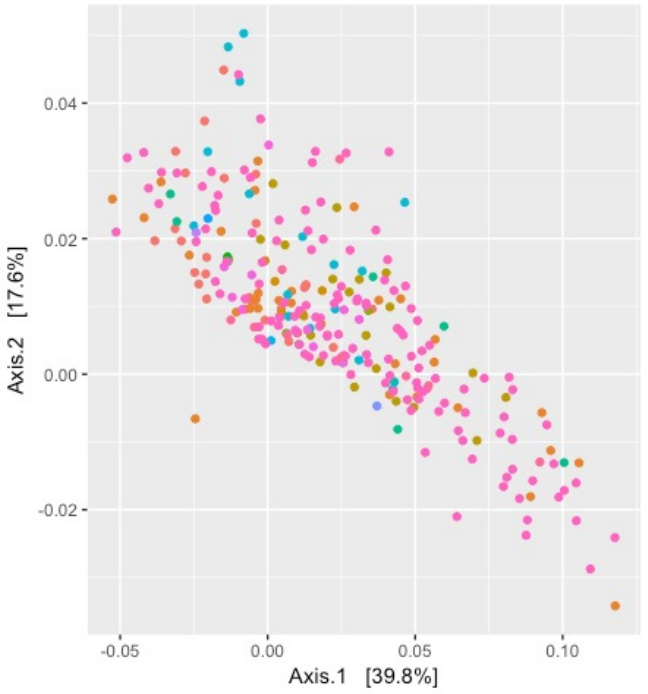
results & data visualization

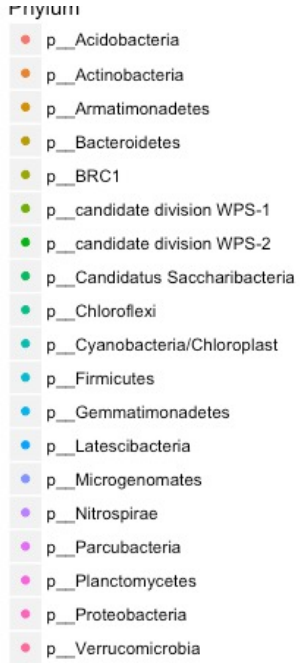

ATCG
AAGG
phyloseq

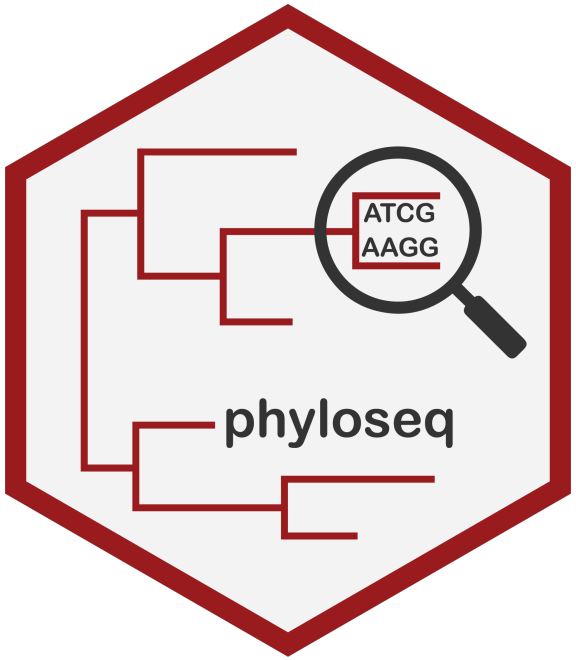

Abundance
40000
30000
20000
10000
0
Slashpile1
Slashpile10
Slashpile11
Slashpile13
Slashpile14
Slashpile15
Slashpile16
Slashpile17
Slashpile18
Slashpile19
Slashpile2
Slashpile20
Slashpile21
Slashpile22
Slashpile23
Slashpile24
Slashpile25
Slashpile26
Slashpile27
Slashpile28
Slashpile3
Slashpile4
Slashpile40
Slashpile5
Slashpile6
Slashpile7
Slashpile8
Slashpile9
Sample


__Acidobacteria
__Actinobacteria
__Aquificae
__Armatimonadetes
__Bacteroidetes
__BRC1
__candidate division WPS-1
p__candidate division WPS-2
p__candidate division ZB3
p__Candidatus Saccharibacteria
p__Chlamydiae
p__Chloroflexi
p__Crenarchaeota
p__Cyanobacteria/Chloroplast
p__Deinococcus-Thermus
p__Elusimicrobia
p__Firmicutes
p__Gemmatimonadetes
p__Hydrogenedentes
p__Latescibacteria
p__Lentisphaerae
p__Microgenomates
p__Nitrospirae
p__Omnitrophica
p__Parcubacteria
p__Planctomycetes
p__Proteobacteria
p__Spirochaetes


p__Acidobacteria
p__Actinobacteria
p__Armatimonadetes
p__Bacteroidetes
p__BRC1
candidate division WP
candidate division WP
andidatus Saccharibac
p__Chlamydiae
p__Chloroflexi
Cyanobacteria/Chlorop
p__Elusimicrobia
p__Firmicutes
__Gemmatimonadetes
p__Hydrogenedentes
p__Latescibacteria
p__Lentisphaerae
p__Microgenomates
p__Nitrospirae
p__Parcubacteria
p__Planctomycetes
p__Proteobacteria
p__Spirochaetes
p__SR1
p__Verrucomicrobia
Prevalence [Frac. Samples]
1.00
0.75
0.50
0.25
0.00
10
1000
Total Abundance

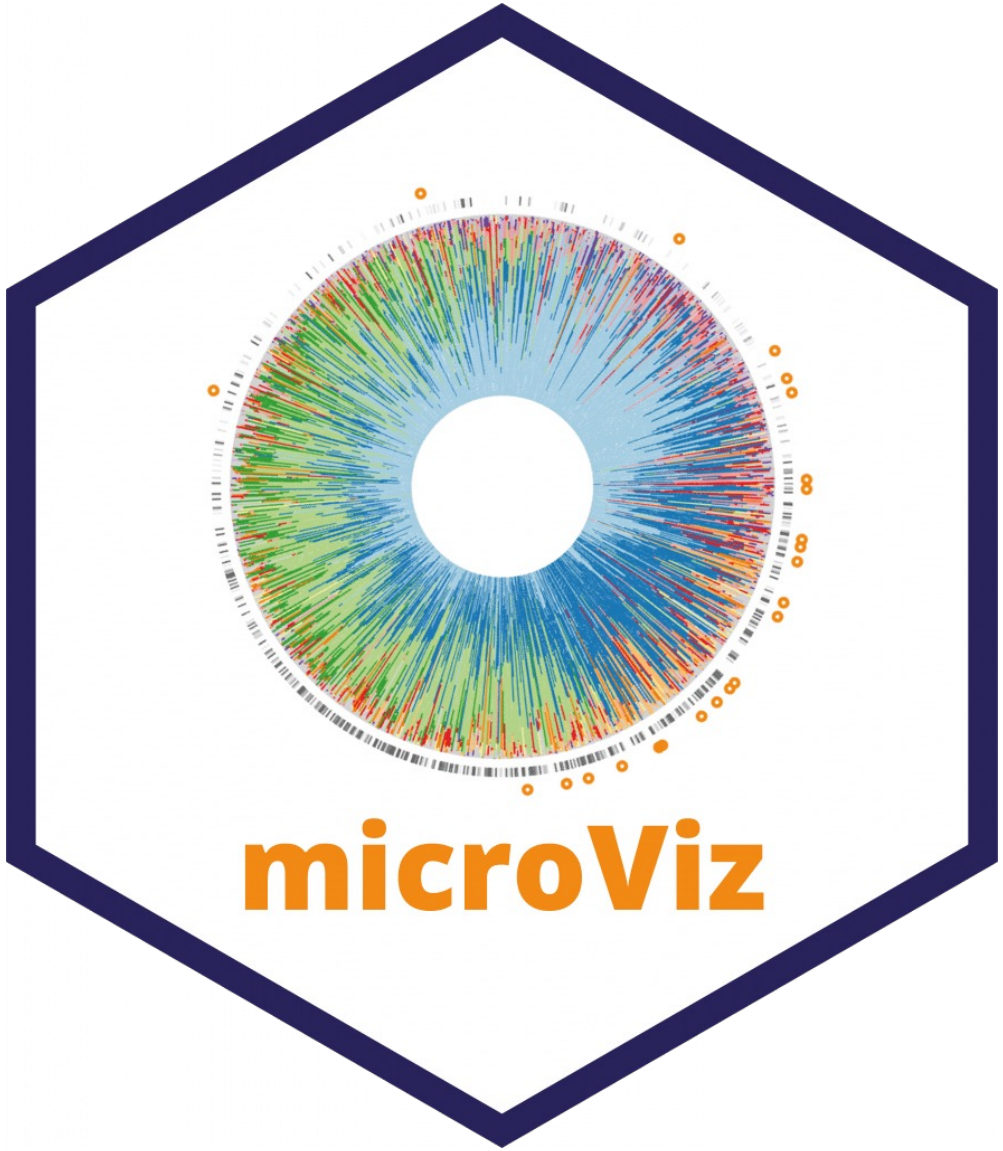

0.04
0.02
0.00
-0.02
Axis.2 [17.6%]
-0.05
0.00
0.05
0.10
Axis.1 [39.8%]


Phylum
p__Acidobacteria
p__Actinobacteria
p__Armatimonadetes
p__Bacteroidetes
p__BRC1
p__candidate division WPS-1
p__candidate division WPS-2
p__Candidatus Saccharibacteria
p__Chloroflexi
p__Cyanobacteria/Chloroplast
p__Firmicutes
p__Gemmatimonadetes
p__Latescibacteria
p__Microgenomates
p__Nitrospirae
p__Parcubacteria
p__Planctomycetes
p__Proteobacteria
p__Verrucomicrobia


ATCG
AAGG
phyloseq

https://ucdavis-bioinformatics-training.github.io/2017-September-Microbial-Community-Analysis-Workshop/friday/MCA_Workshop_R/phyloseq.html

https://david-barnett.github.io/microViz/



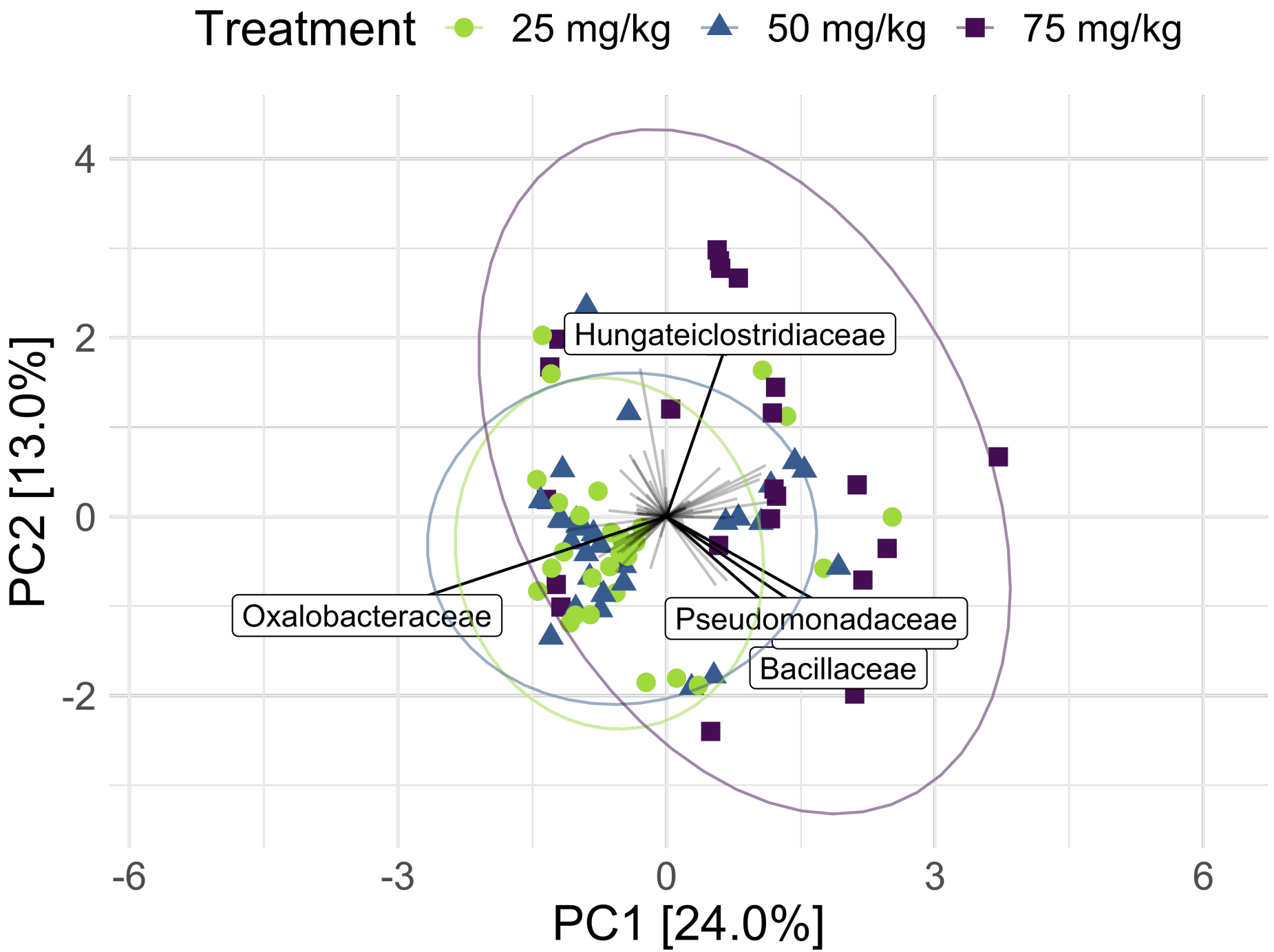
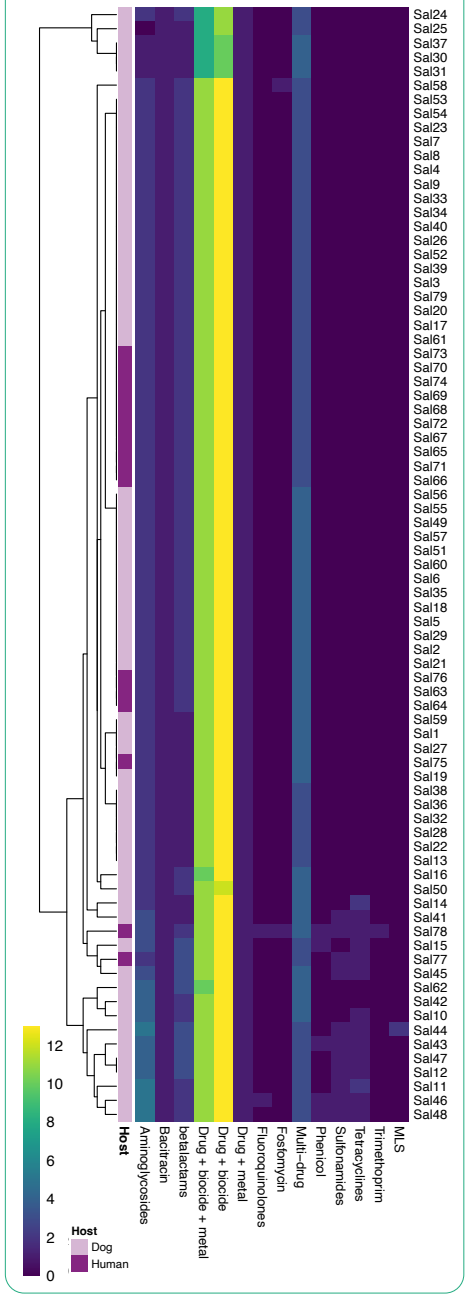
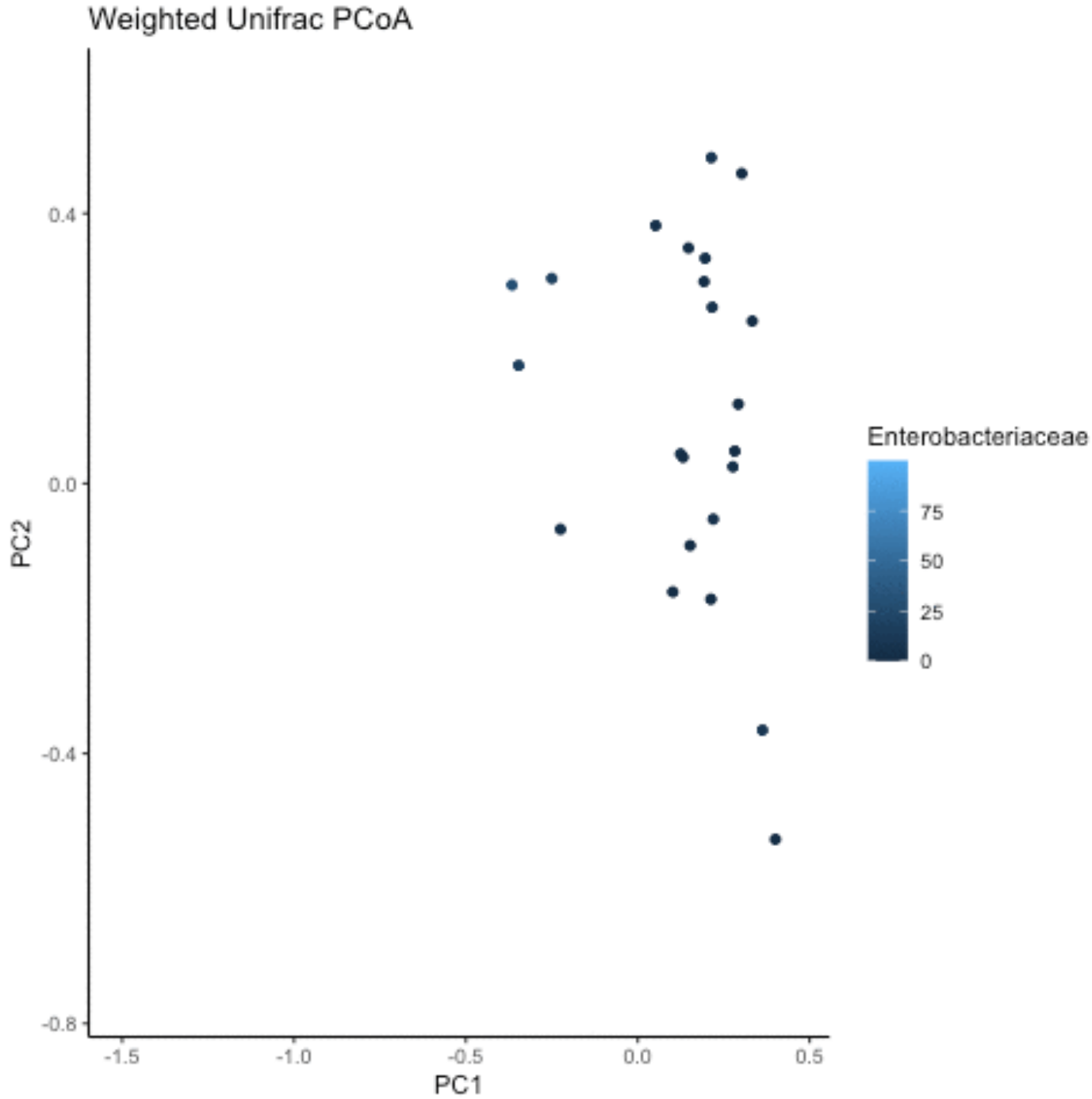


Figure 3. Number of AMRg per isolate by Drug Class and isolate host source. Individual isolates (rows) are arranged by hierarchical clustering. Host source is indicated by color ("Host" column).



Bonus

Bonus

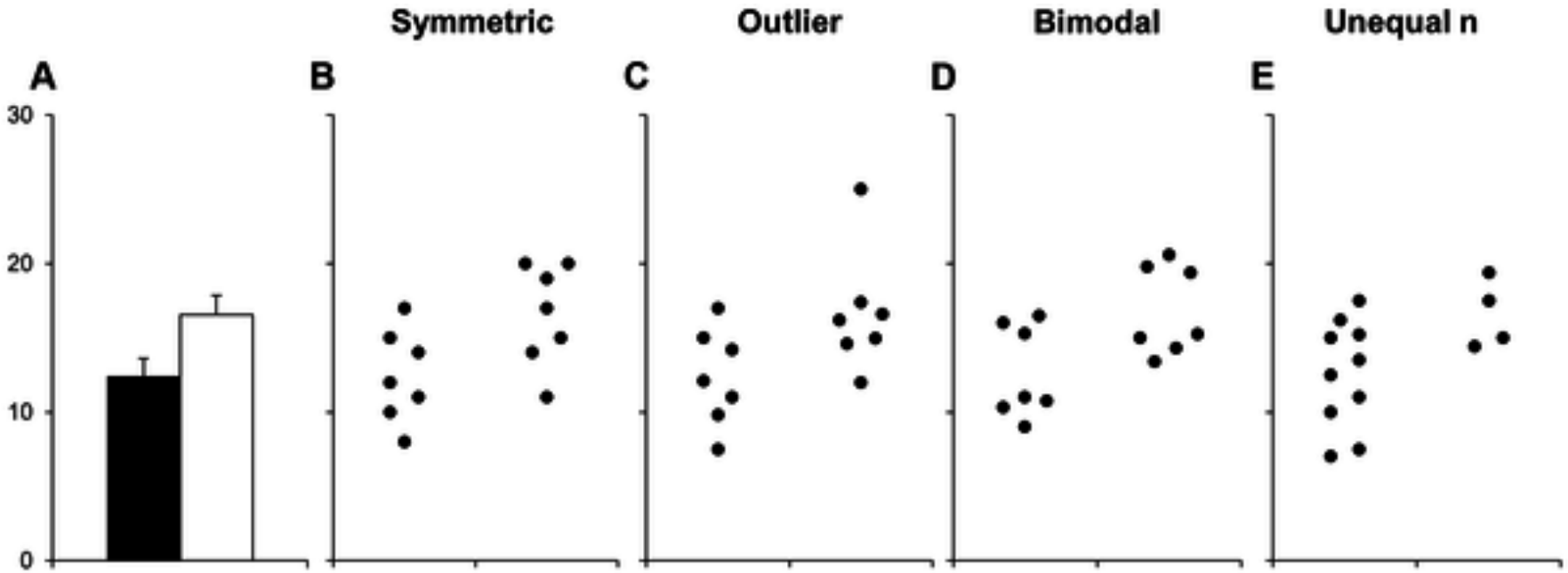

Weighted Unifrac PCoA
0.4
0.0
-0.4
-0.8
PC2
-1.5
-1.0
-0.5
0.0
0.5
PC1
Enterobacteriaceae
75
50
25
0

Food for Thought

Be truthful to your data

Be truthful to your data

Fig 1. Many different datasets can lead to the same bar graph.

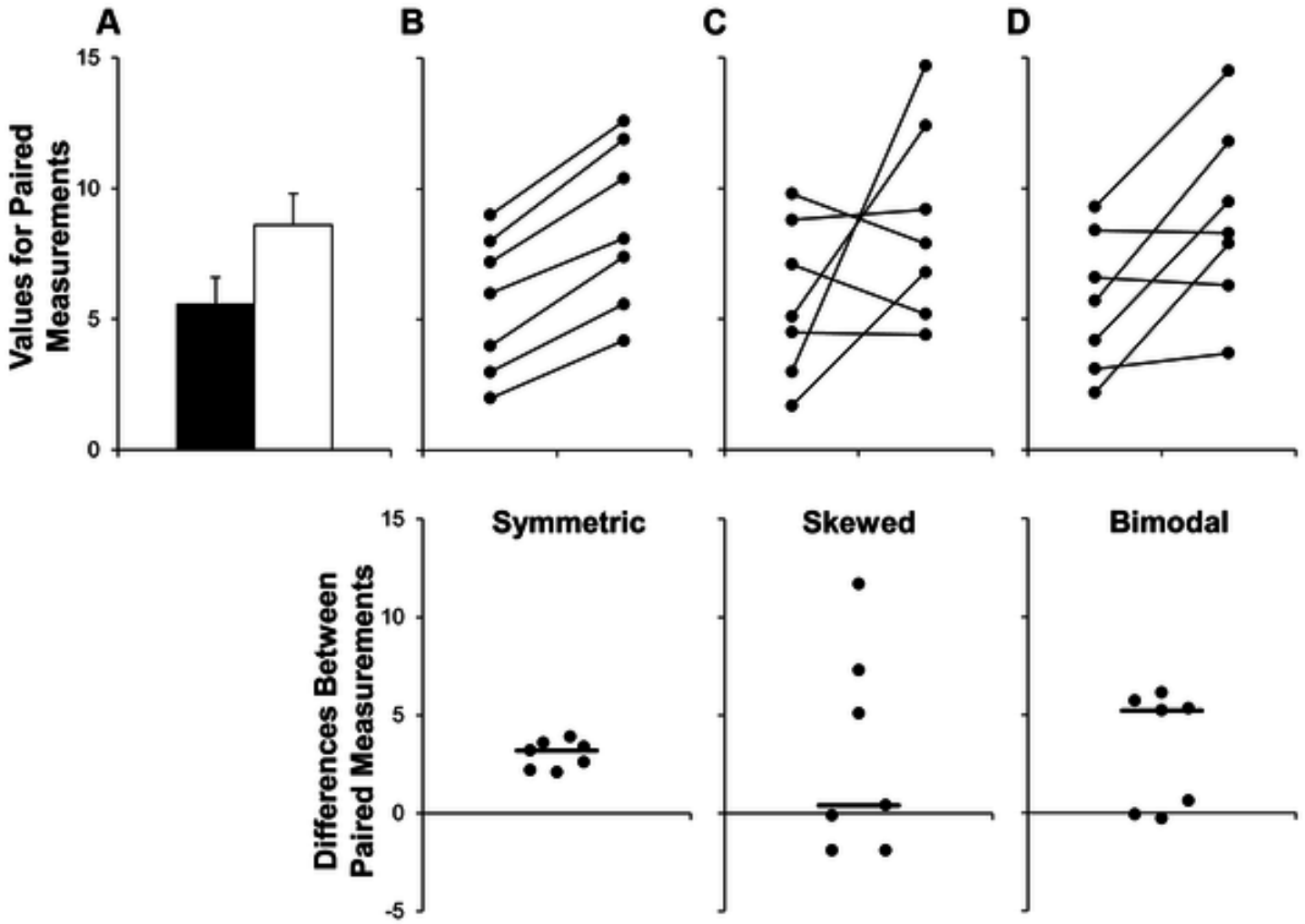


Test	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLOS Biology 13(4): e1002128. https://doi.org/10.1371/journal.pbio.1002128
https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128

Be truthful to your data

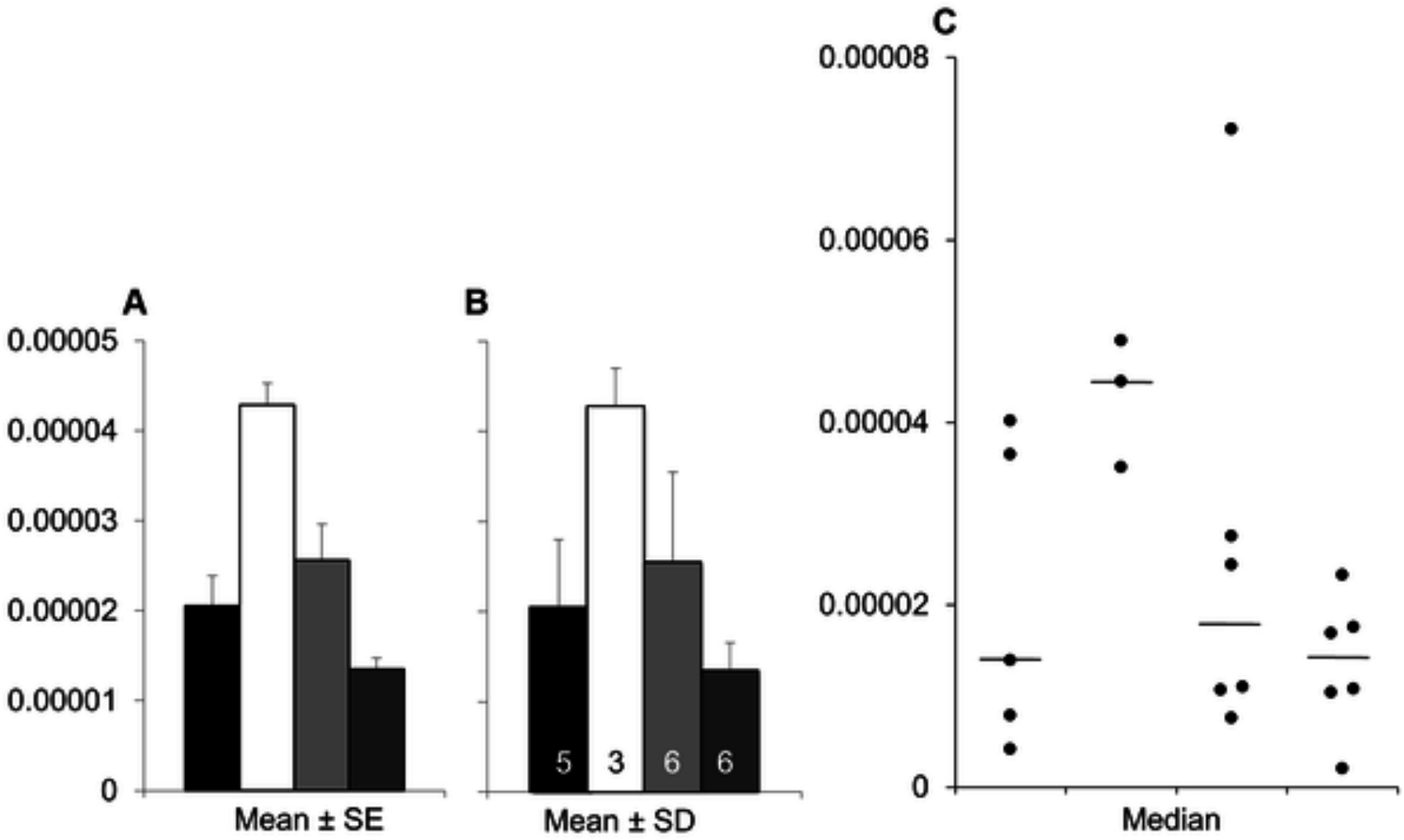
Fig 2. Additional problems with using bar graphs to show paired data.



Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLOS Biology 13(4): e1002128. https://doi.org/10.1371/journal.pbio.1002128
https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128

Be truthful to your data

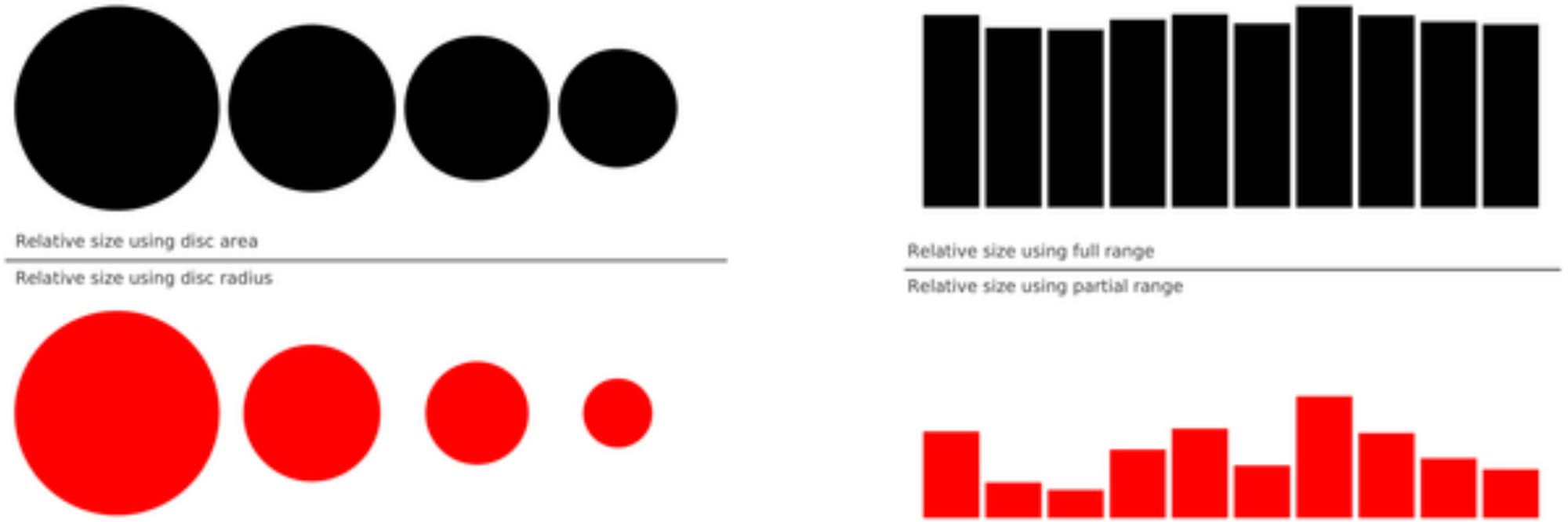
Fig 3. Bar graphs and scatterplots convey very different information.



Weissgerber TL, Milic NM, Winham SJ, Garovic VD (2015) Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLOS Biology 13(4): e1002128. https://doi.org/10.1371/journal.pbio.1002128
https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128

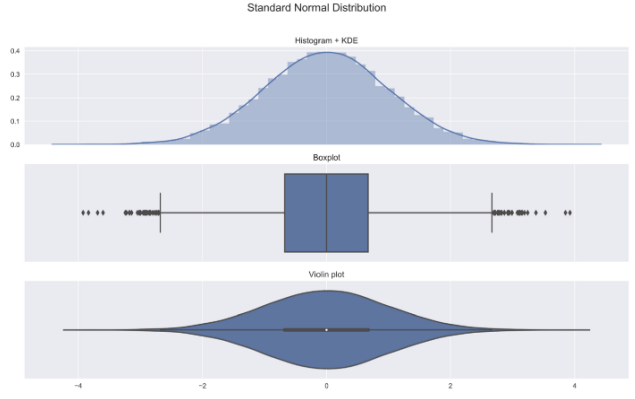
Be truthful to your data

Figure 6. Do not mislead the reader.



Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

Violin Plots



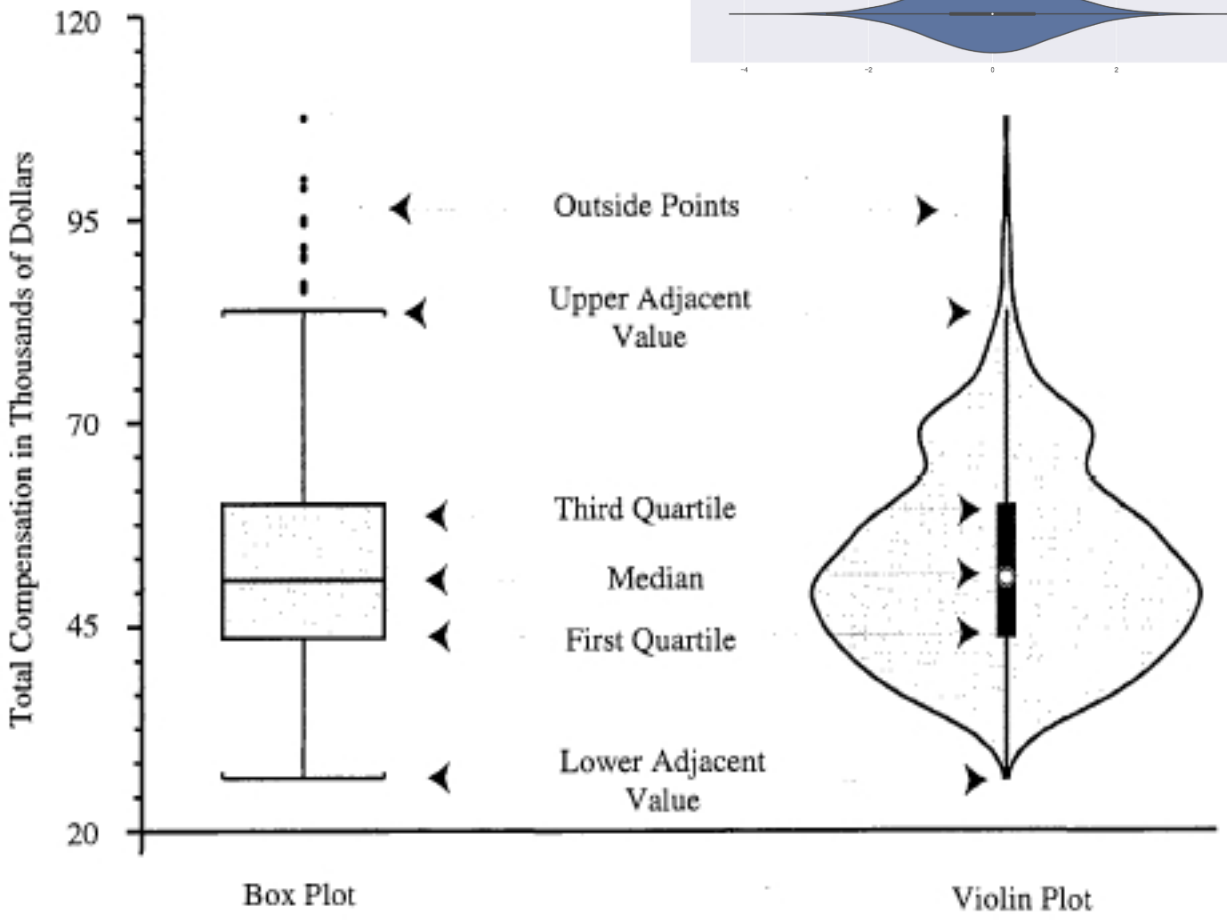
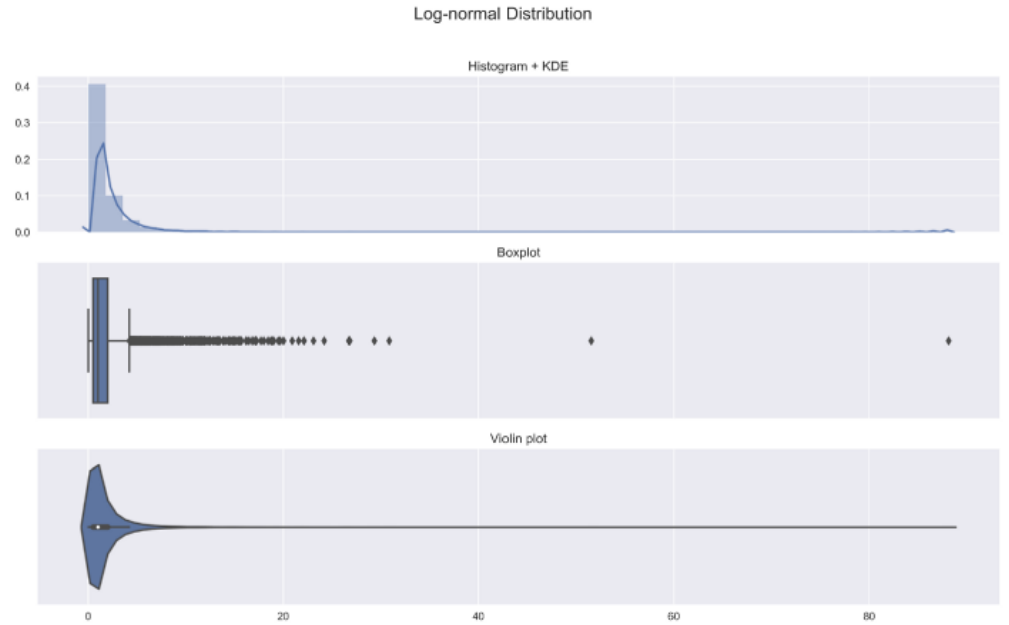
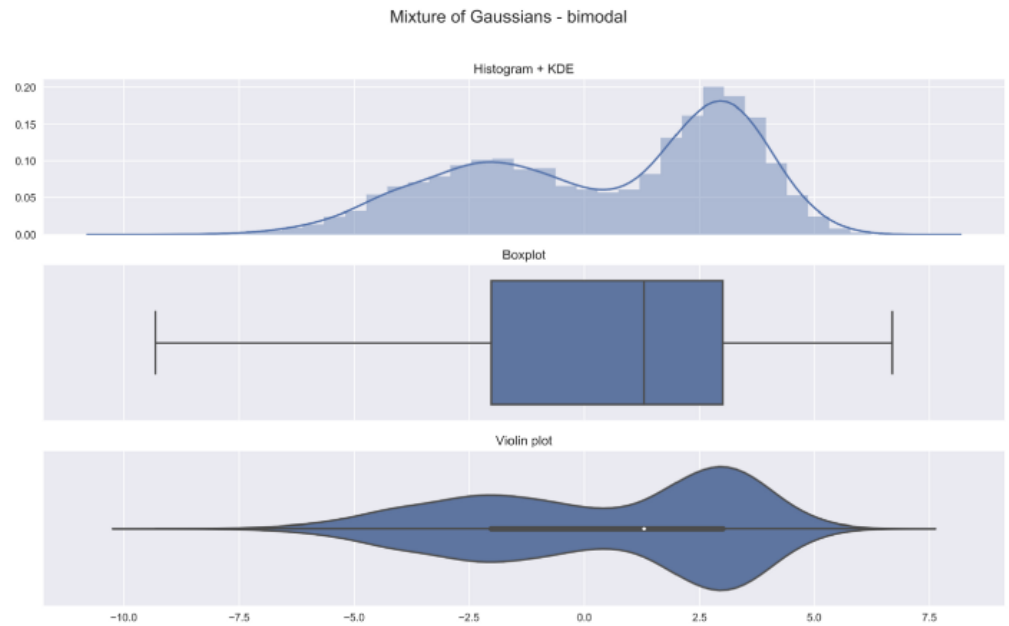


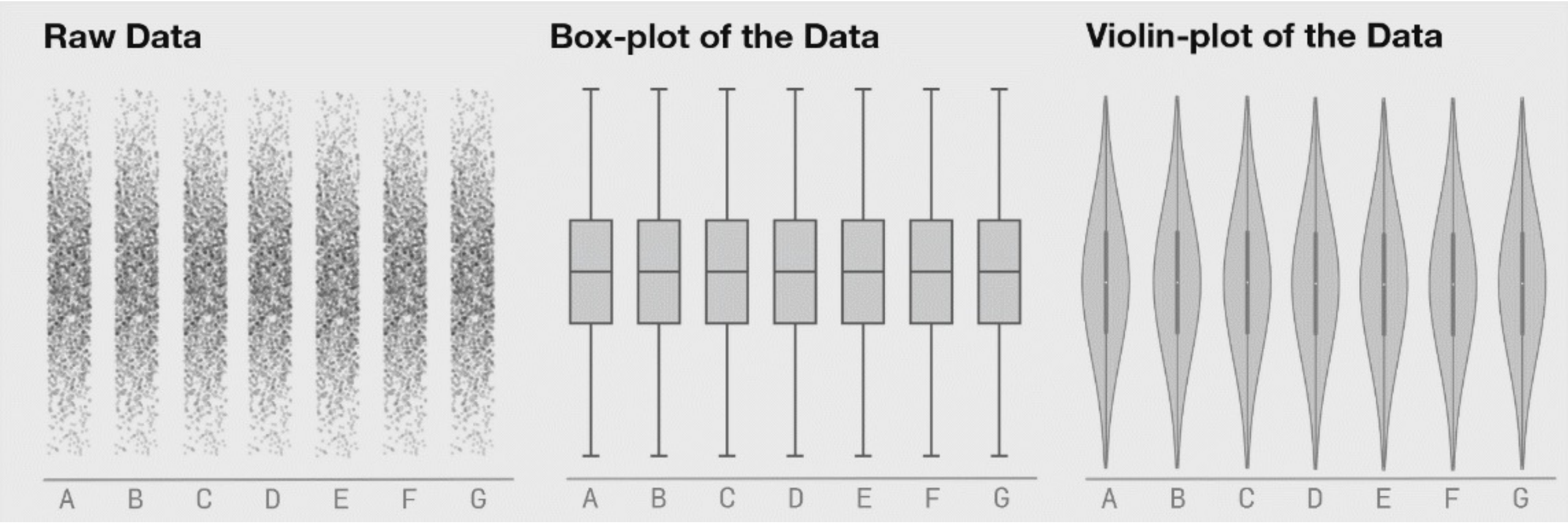
Figure 1. Common Components of Box Plot and Violin Plot. Total compensation for all academic ranks.





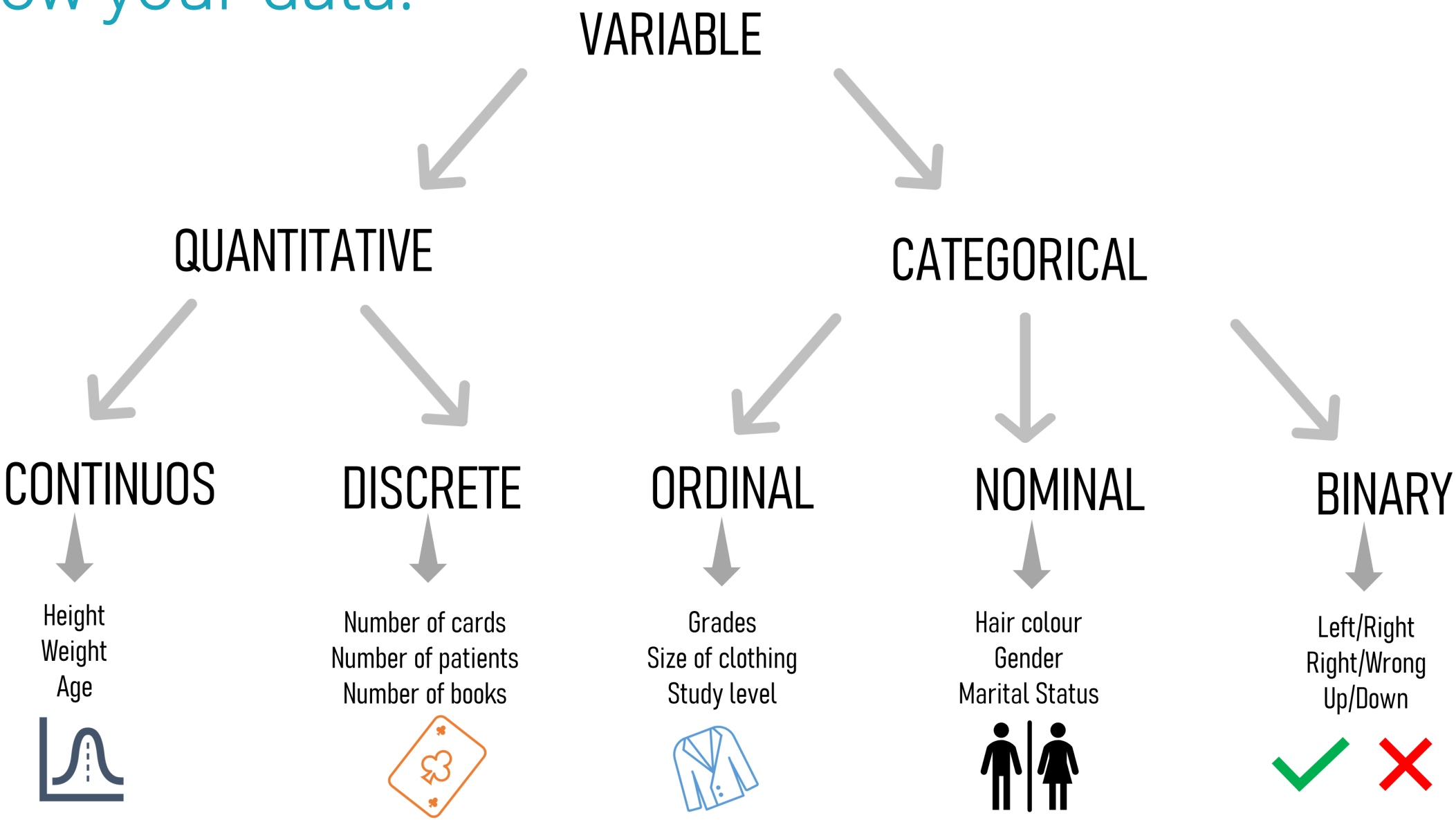
https://towardsdatascience.com/violin-plots-explained-fb1d115e023d

Violin Plots



https://towardsdatascience.com/violin-plots-explained-fb1d115e023d

Know your data!



https://towardsdatascience.com/statistical-testing-understanding-how-to-select-the-best-test-for-your-data-52141c305168

Know your data!

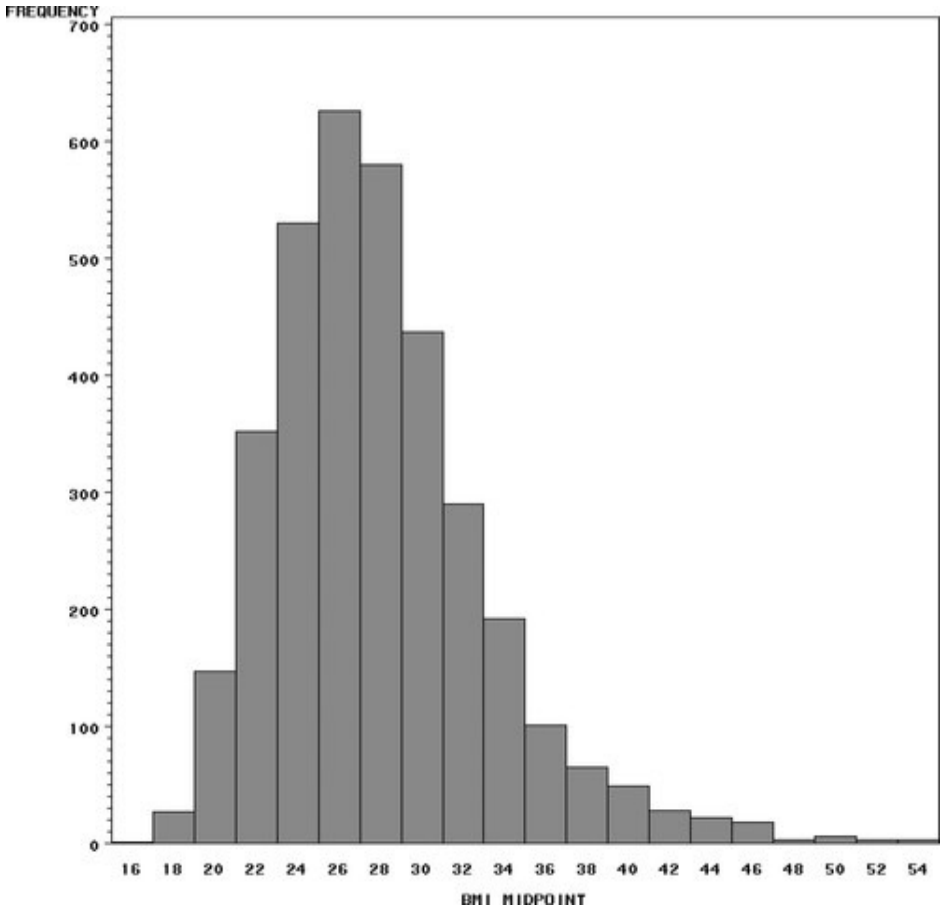


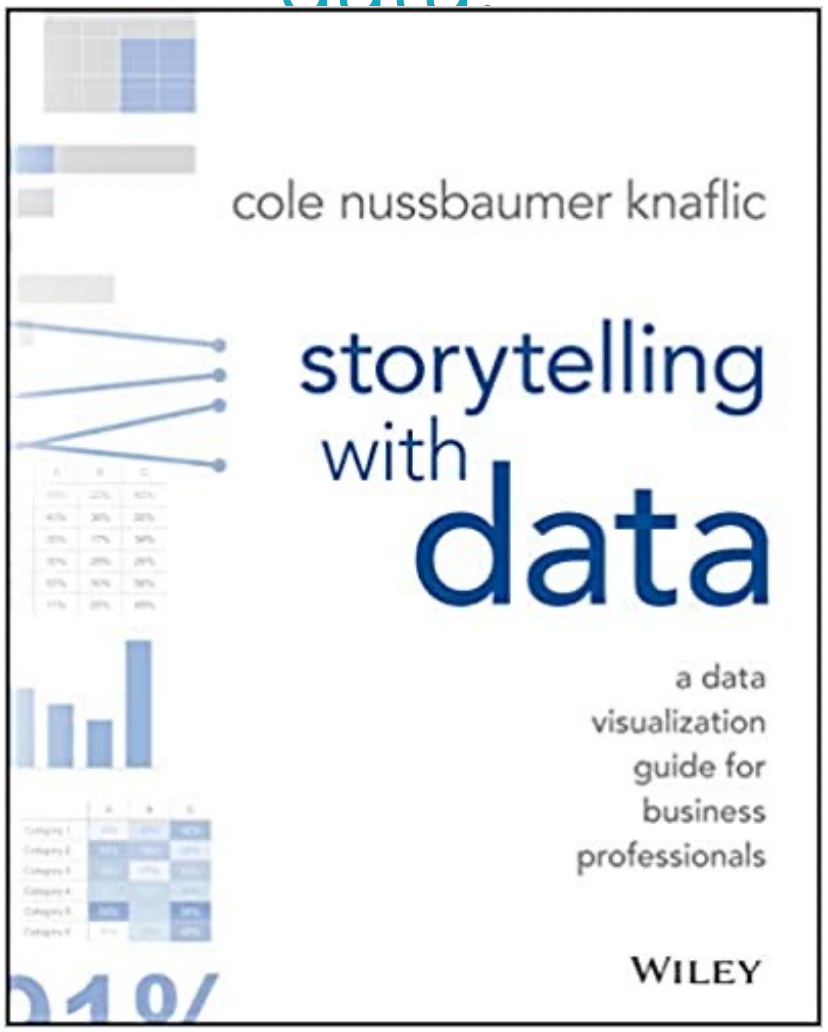
TABLE 2. Descriptive Statistics for BMI (kg/m^2) in Framingham Offspring Participants, Sixth Examination

Sample Size, n	3480	Variance	26.5
Sum	97 190.4	Standard deviation	5.1
Mean	27.9	Skewness	1.1
Mode	26.4	Kurtosis	2.0
Q1	24.4	Minimum	16.6
Q2, median	27.2	Maximum	54.3
Q3	30.6	Range: maximum–minimum	37.7
		IQR=Q3–Q1	6.2

Martin G. Larson. Circulation. Descriptive Statistics and Graphical Displays, Volume: 114, Issue: 1, Pages: 76-81, DOI: (10.1161/CIRCULATIONAHA.105.584474)

https://towardsdatascience.com/statistical-testing-understanding-how-to-select-the-best-test-for-your-data-52141c305168

Visualize your data!



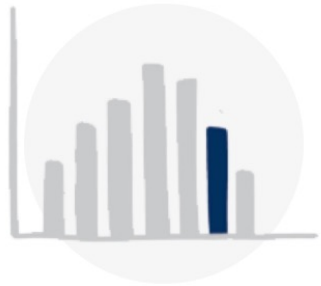
https://www.storytellingwithdata.com/

Visualize your data!

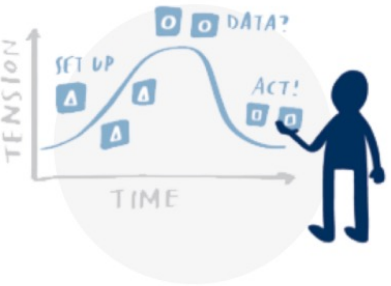
understand the **context**

choose an **effective visual**

eliminate **clutter**



focus **attention**

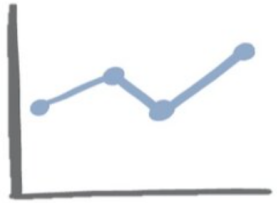


tell a **story**

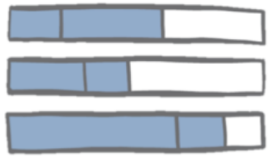
https://www.storytellingwithdata.com/

Visualize your data!

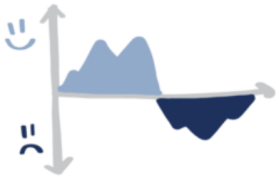
WHAT IS A LINE GRAPH?



WHAT IS A BAR CHART?

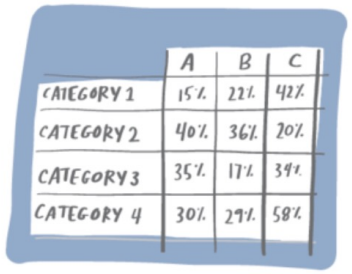


WHAT IS AN AREA GRAPH?



WHAT IS A WATERFALL?

WHAT IS A DATA TABLE?



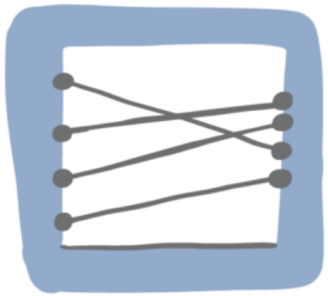
	A	B	C
CATEGORY 1	15%	22%	42%
CATEGORY 2	40%	36%	20%
CATEGORY 3	35%	17%	39%
CATEGORY 4	30%	29%	58%

WHAT IS A DOT PLOT?

WHAT IS A PIE CHART?

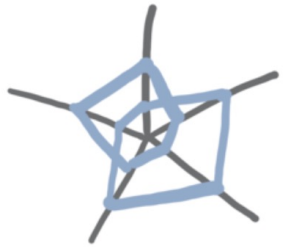
WHAT IS A SCATTERPLOT?

WHAT IS A SLOPEGRAPH?



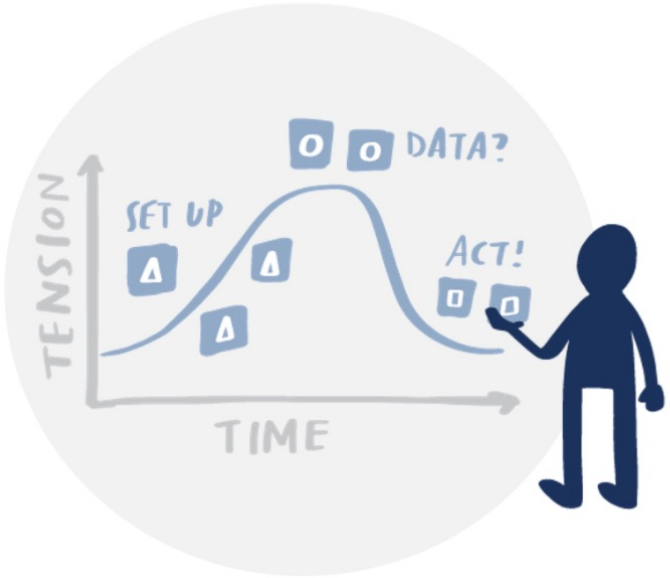
WHAT IS A BUBBLE CHART?

WHAT IS A SPIDER CHART?



https://www.storytellingwithdata.com/

Visualize your data!

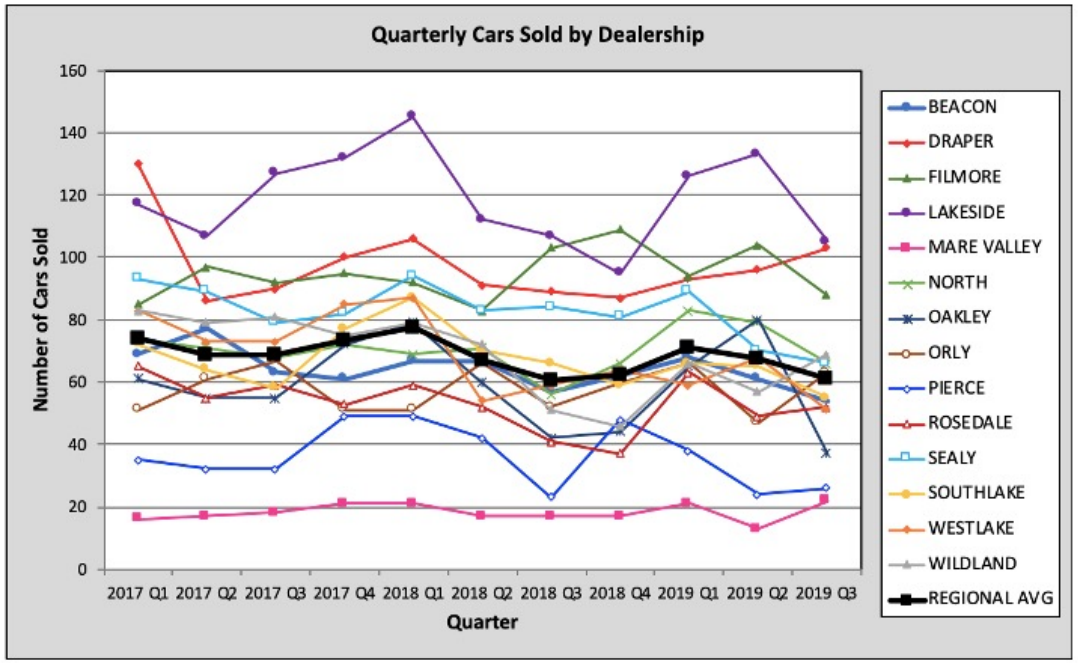


move beyond
simply showing data
with
SWD MAKEOVERS

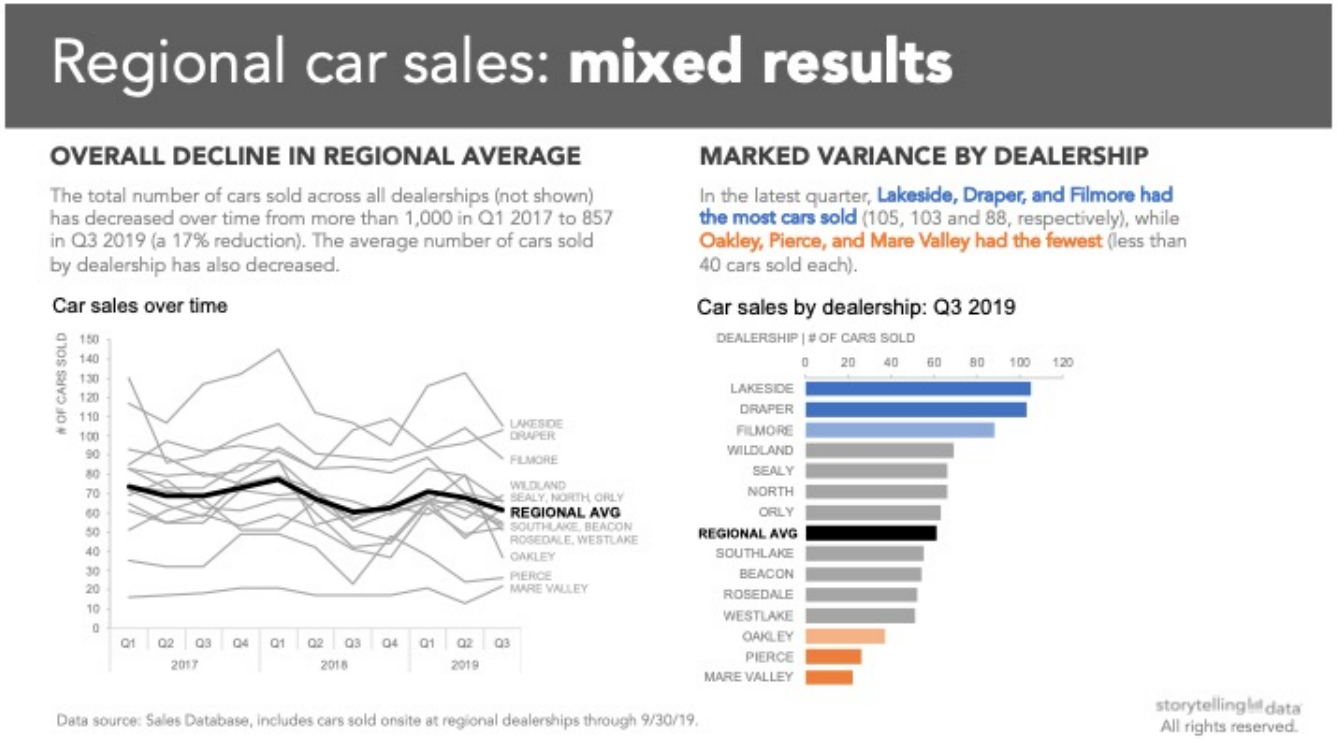
https://www.storytellingwithdata.com/

Visualize your data!

BEFORE: SHOWING DATA



AFTER: STORYTELLING WITH DATA



https://www.storytellingwithdata.com/

Visualize your data!

BEFORE: SHOWING DATA

Global Competitive Summary

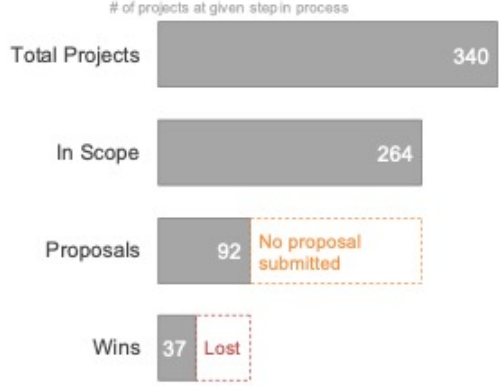
Global Competitive Summary (1/1/2015-12/31/2015)	
	Global
Total Projects	340
In Scope	264
Proposals	92
Wins	37

AFTER: STORYTELLING WITH DATA

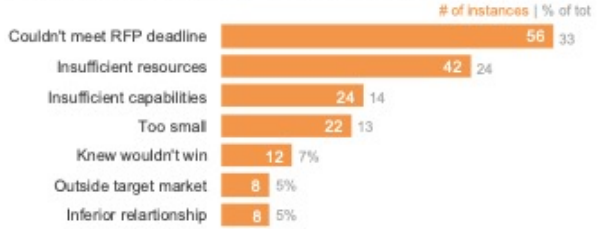
Opportunity: resolve timing, resource issues

Timing & resource issues are top reasons we aren't submitting/winning proposals.

Global project summary
January - December 2015



Why we **aren't submitting** proposals



Why we **aren't winning** proposals

https://www.storytellingwithdata.com/

Visualize your data!

Plotly

from Data to Viz

https://www.data-to-viz.com/

Visualize your data!

from Data to Viz

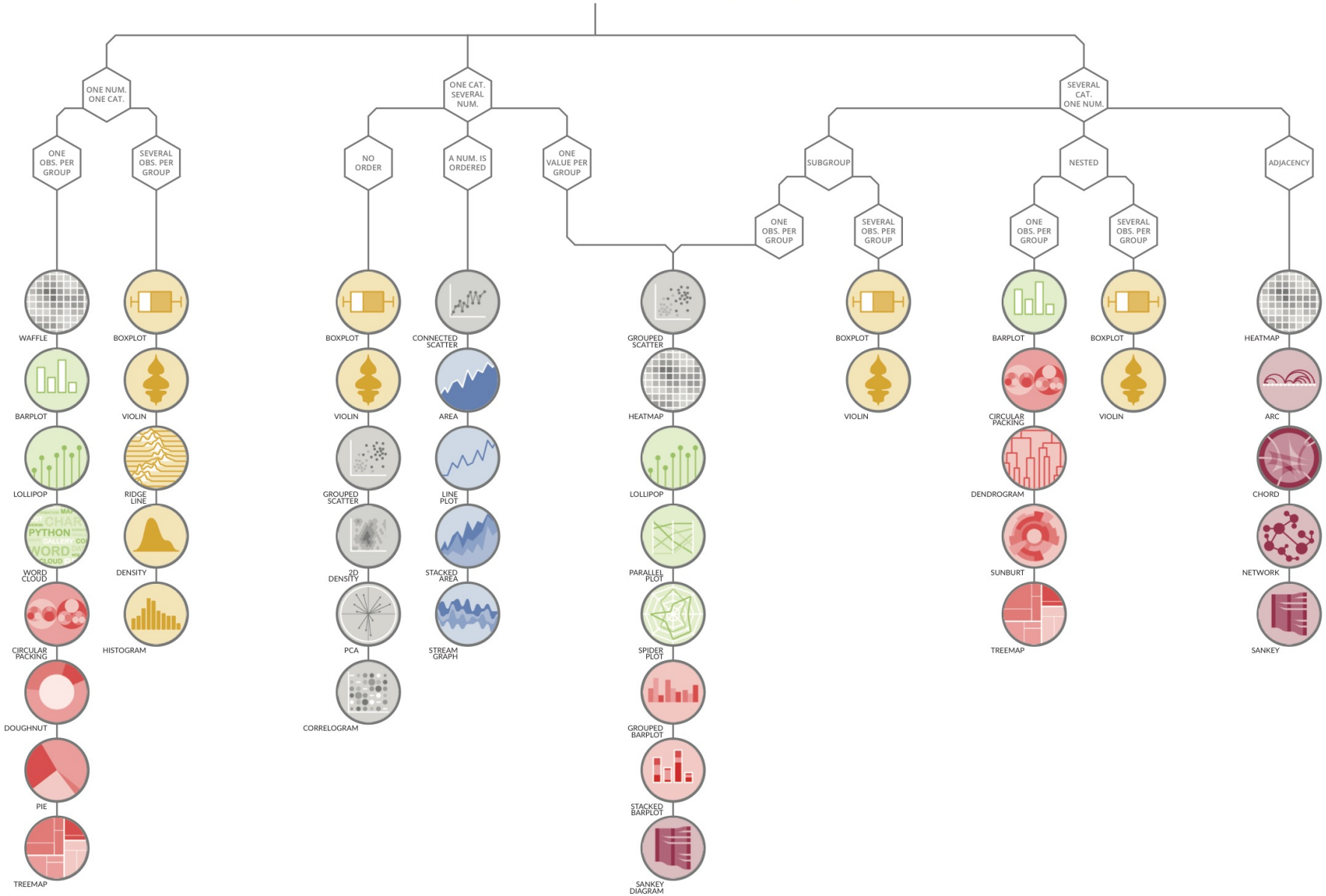
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

1. Identify what type of data you have.
2. Go to the corresponding decision tree and follow it down to a set of possible charts.
3. Choose the chart from the set that will suit your data and your needs best.

Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

data-to-viz.com

CATEGORIC AND NUMERIC



Visualize your data!

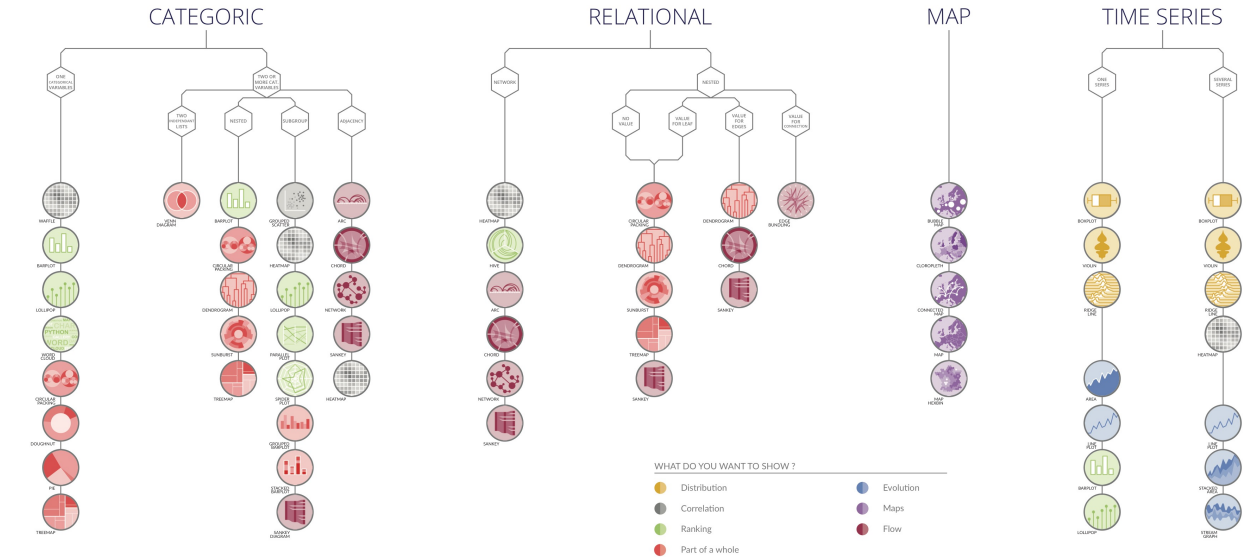
from Data to Viz

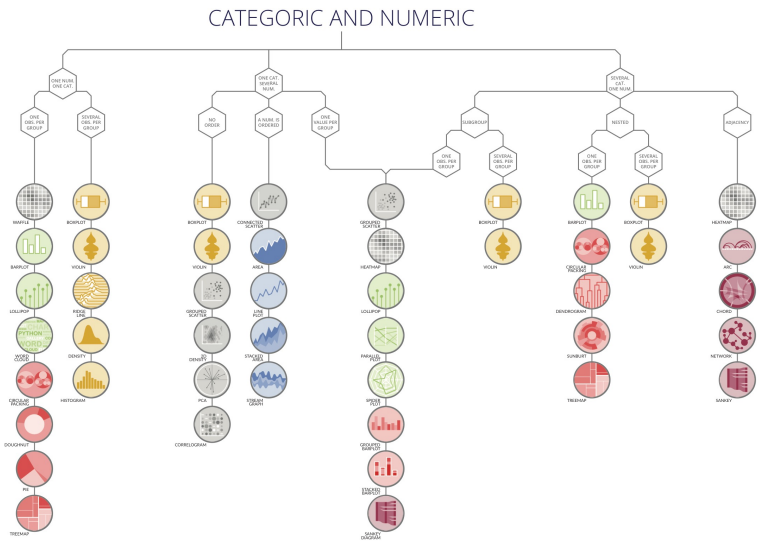
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

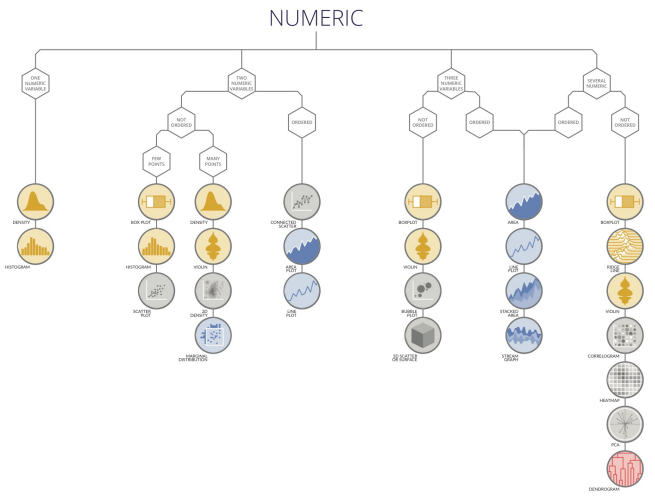
1. Identify what type of data you have.
2. Go to the corresponding decision tree and follow it down to a set of possible charts.
3. Choose the chart from the set that will suit your data and your needs best.

Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

data-to-viz.com

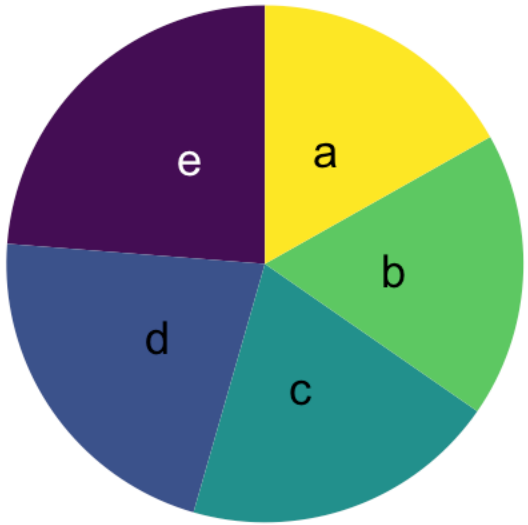


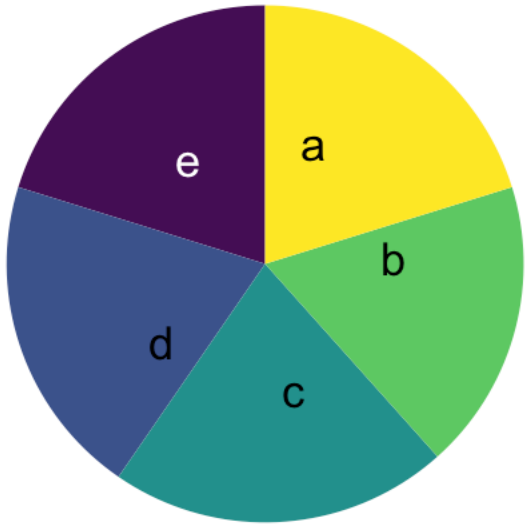


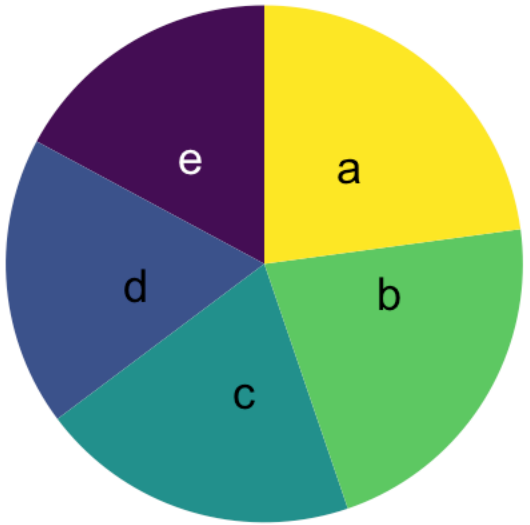


https://www.data-to-viz.com/

Pie charts are

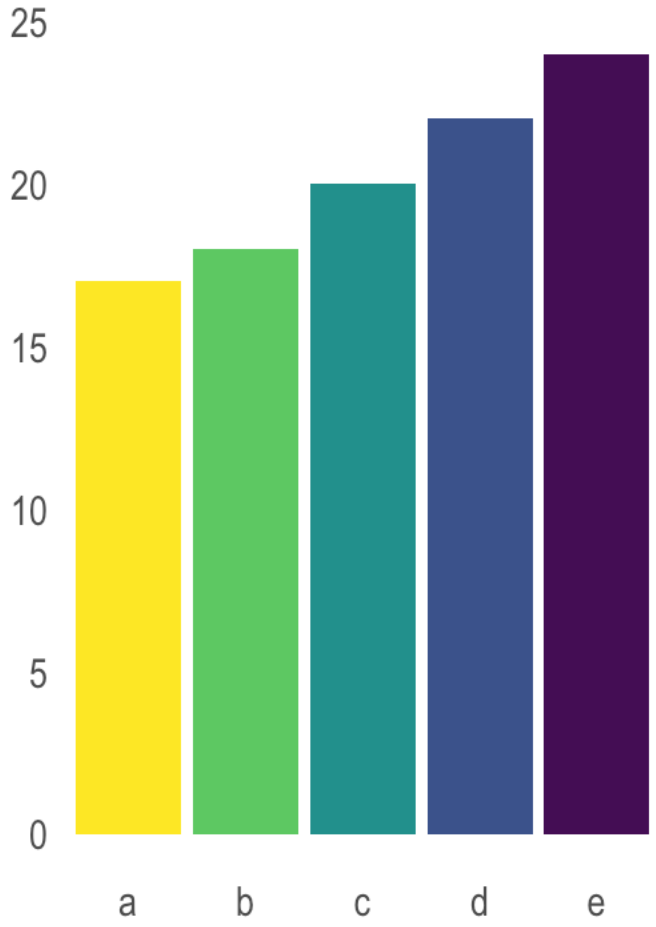


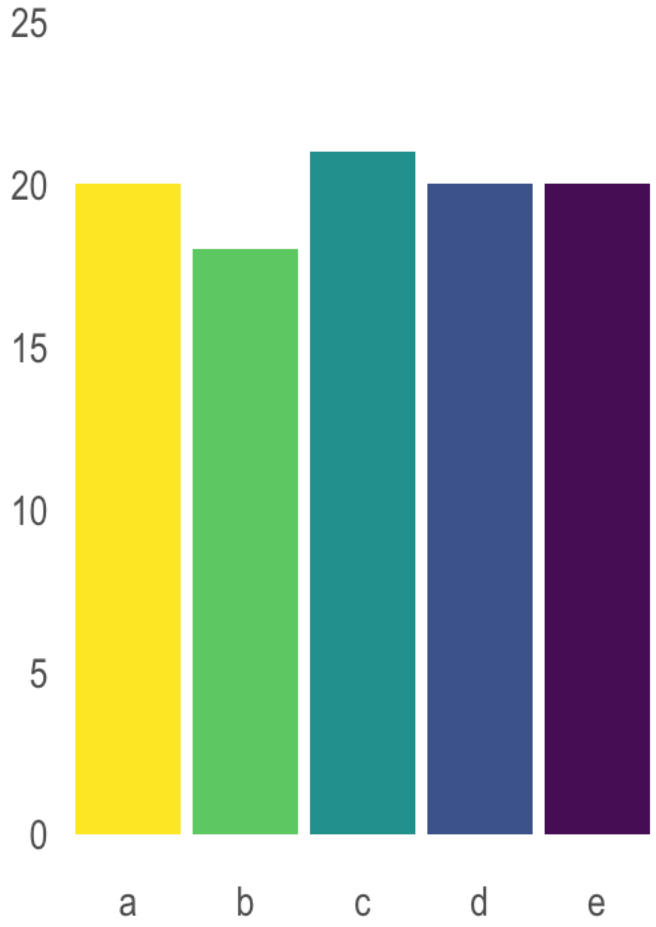


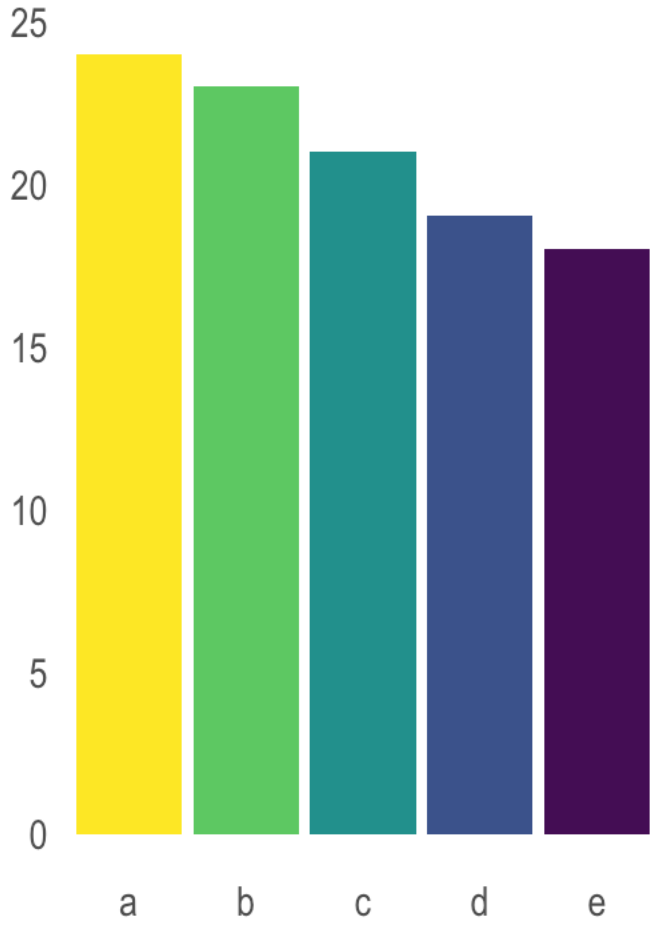


https://www.data-to-viz.com/caveat/pie.html

Pie charts are

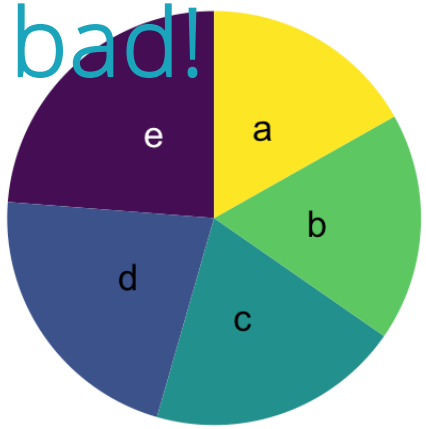


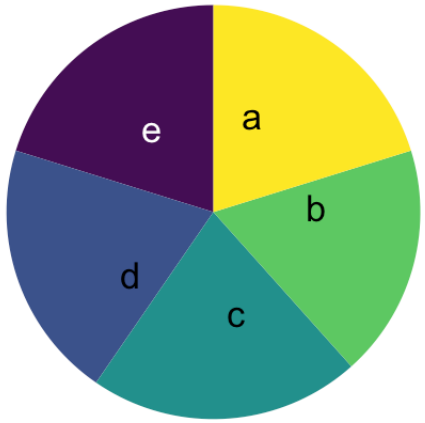


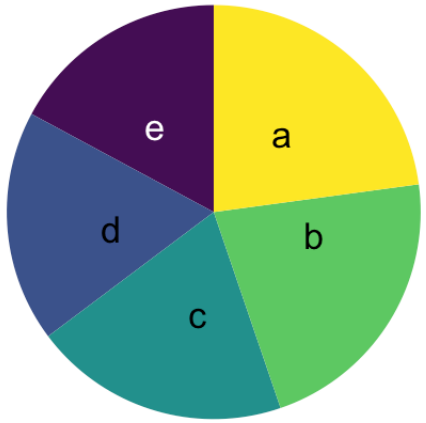


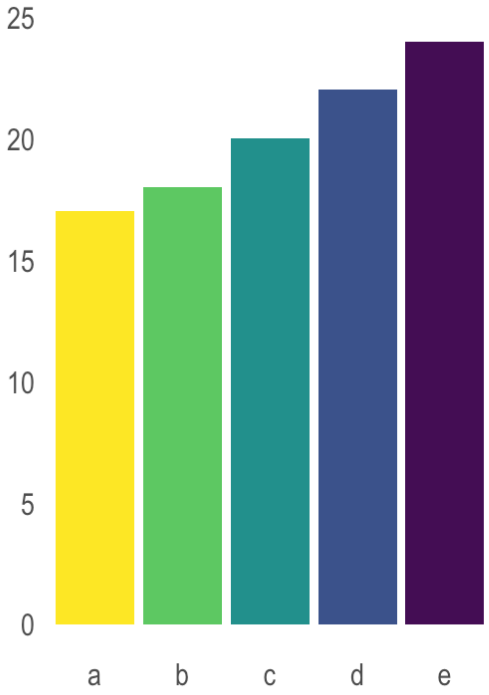
https://www.data-to-viz.com/caveat/pie.html

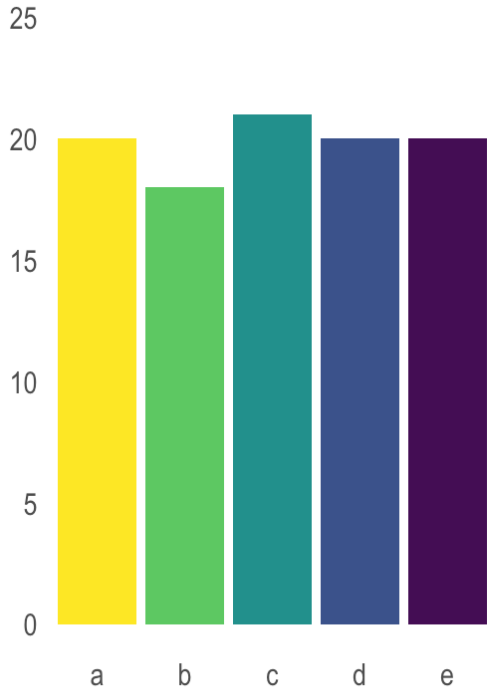
Pie charts are bad!

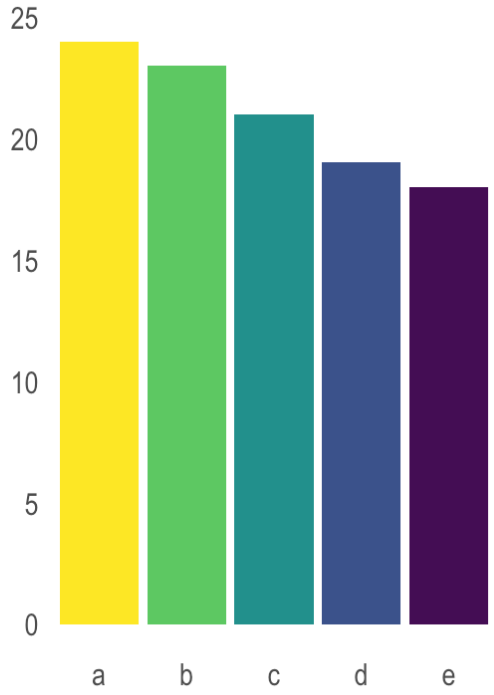








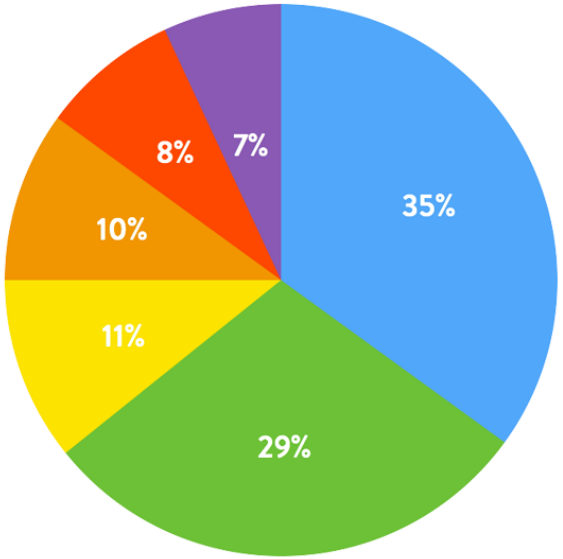


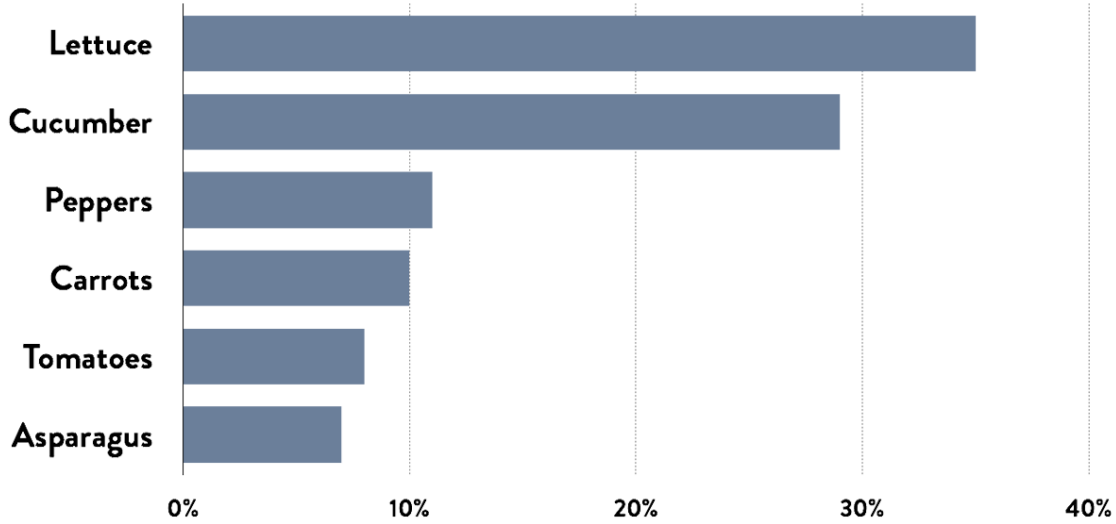


https://www.data-to-viz.com/caveat/pie.html

Pie charts are bad!

VEGETABLE	AMOUNT
Lettuce	35%
Cucumber	29%
Peppers	11%
Carrots	10%
Tomatoes	8%
Asparagus	7%





https://www.geckoboard.com/blog/pie-charts/

Best Practices for Making figures

1: Know Your Audience

2: Identify Your Message

3: Adapt the Figure to the Support Medium

4: Captions Are Not Optional

5: Do Not Trust the Defaults

6: Use Color Effectively

7: Do Not Mislead the Reader

8: Avoid “Chartjunk”

9: Message Trumps Beauty

10: Get the Right Tool

+

Consistence, order, simplicity!

OPEN ACCESS Freely available online

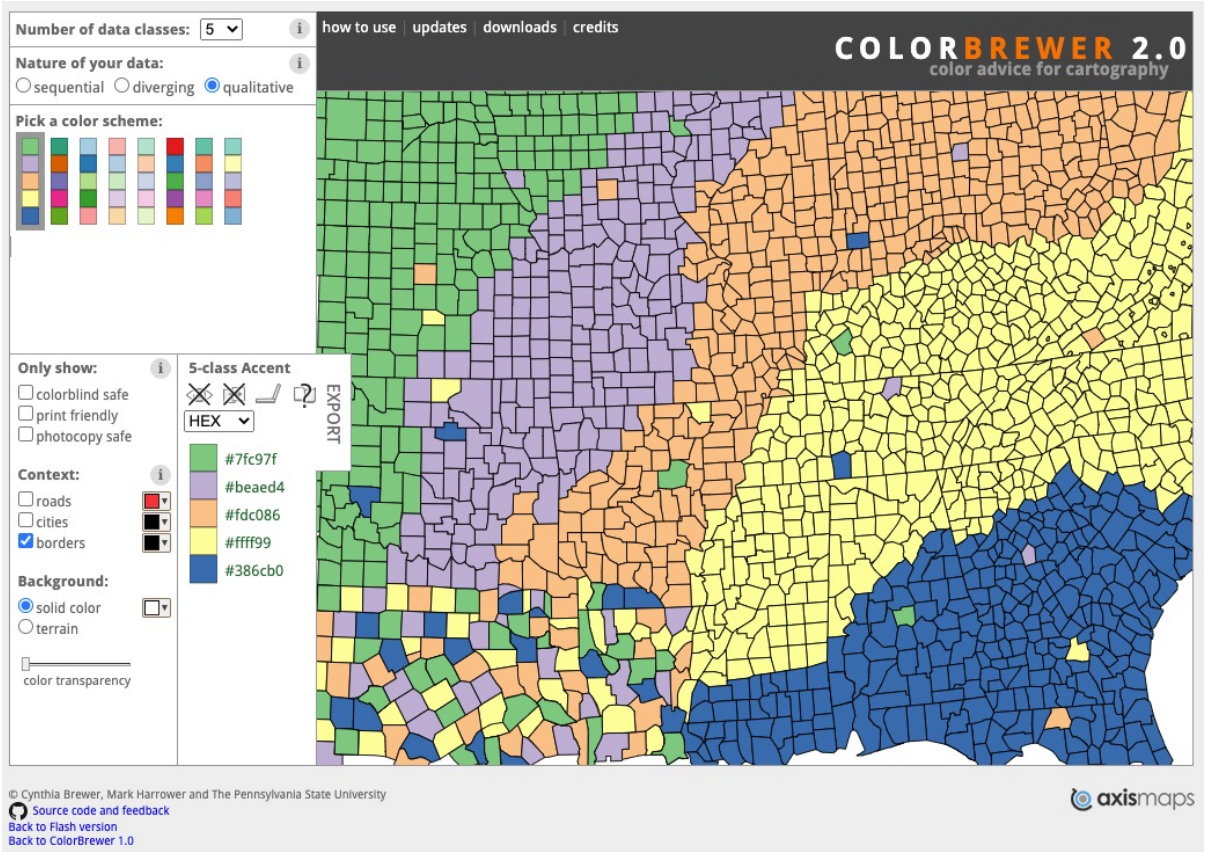
PLOS | COMPUTATIONAL BIOLOGY

Editorial

Ten Simple Rules for Better Figures

Nicolas P. Rougier[1,2,3]*, Michael Droettboom[4], Philip E. Bourne[5]

1 INRIA Bordeaux Sud-Ouest, Talence, France, **2** LaBRI, UMR 5800 CNRS, Talence, France, **3** Institute of Neurodegenerative Diseases, UMR 5293 CNRS, Bordeaux, France, **4** Space Telescope Science Institute, Baltimore, Maryland, United States of America, **5** Office of the Director, The National Institutes of Health, Bethesda, Maryland, United States of America



https://colorbrewer2.org/#type=qualitative&scheme=Accent&n=5

https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

Best Practices for Making figures

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

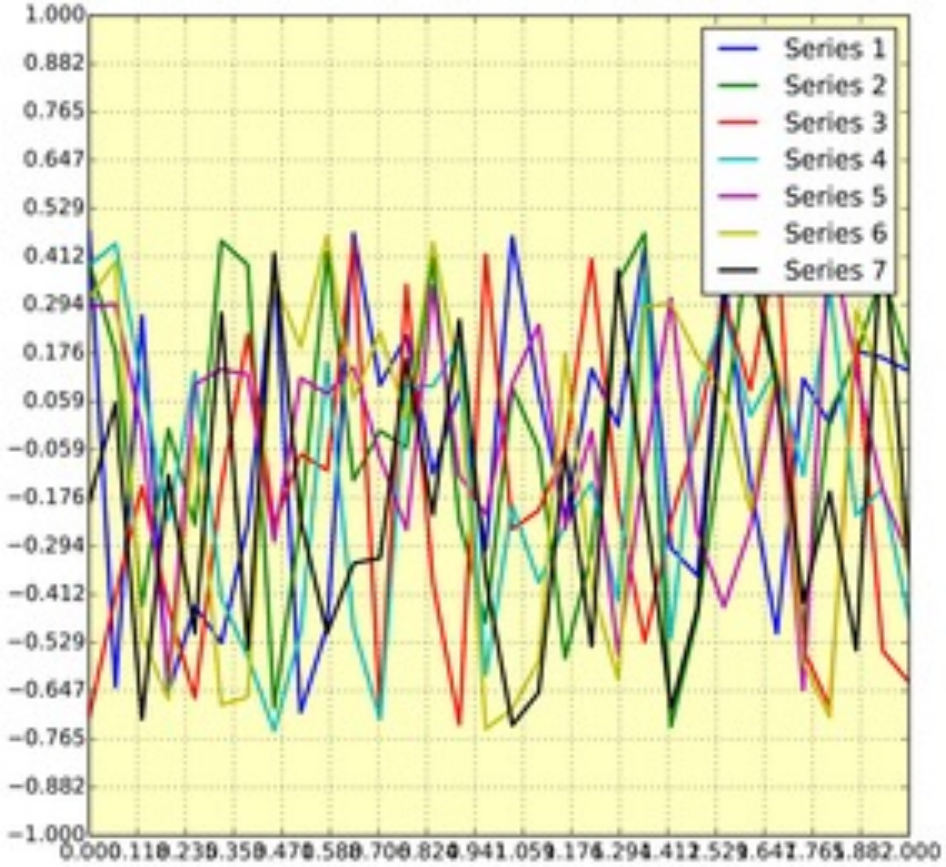
Ten Simple Rules for Better Figures

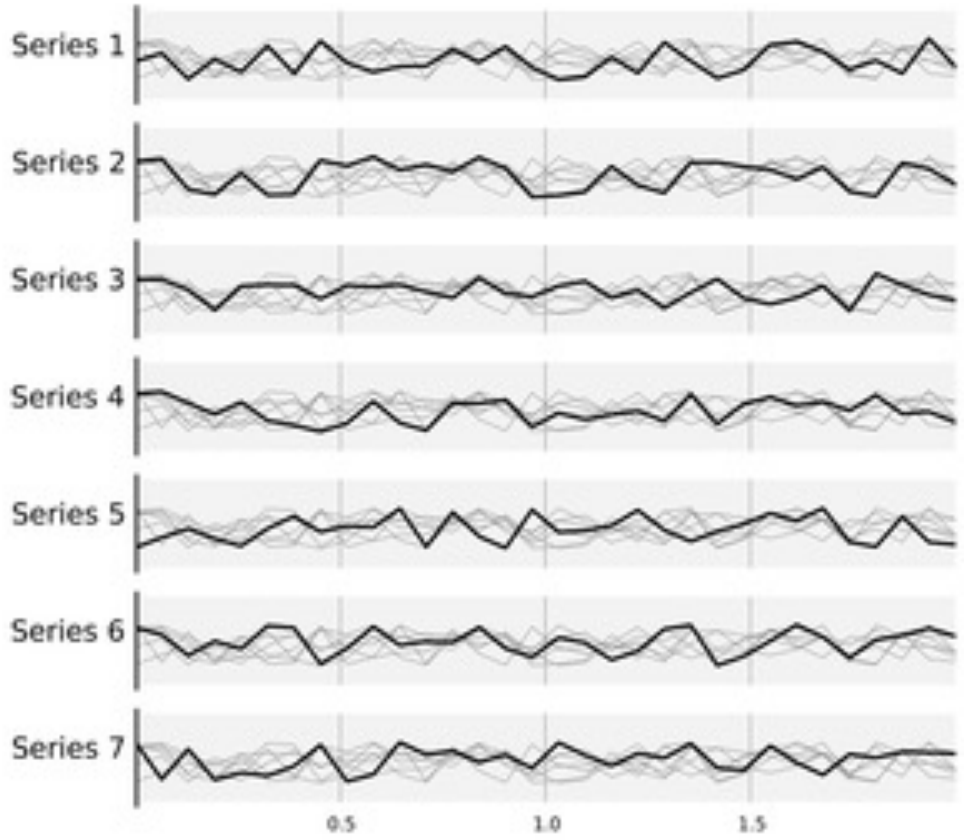
Nicolas P. Rougier[1,2,3]*, Michael Droettboom[4], Philip E. Bourne[5]

1 INRIA Bordeaux Sud-Ouest, Talence, France, 2 LaBRI, UMR 5800 CNRS, Talence, France, 3 Institute of Neurodegenerative Diseases, UMR 5293 CNRS, Bordeaux, France, 4 Space Telescope Science Institute, Baltimore, Maryland, United States of America, 5 Office of the Director, The National Institutes of Health, Bethesda, Maryland, United States of America

Avoid “Chartjunk”

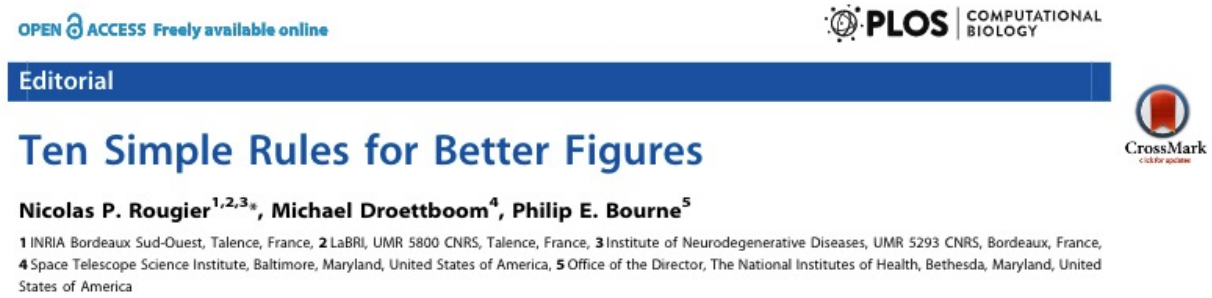
Figure 7. Avoid chartjunk.





Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833

Best Practices for Making figures



OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

Ten Simple Rules for Better Figures

Nicolas P. Rougier[1,2,3]*, Michael Droettboom[4], Philip E. Bourne[5]

1 INRIA Bordeaux Sud-Ouest, Talence, France, 2 LaBRI, UMR 5800 CNRS, Talence, France, 3 Institute of Neurodegenerative Diseases, UMR 5293 CNRS, Bordeaux, France, 4 Space Telescope Science Institute, Baltimore, Maryland, United States of America, 5 Office of the Director, The National Institutes of Health, Bethesda, Maryland, United States of America

Matplotlib (http://matplotlib.org/gallery.html).

R (ggplot2, ggpubr)

Inkscape (professional vector graphics editor).

TikZ and PGF are TeX packages for creating graphics programmatically.

(http://www.texample.net/tikz/examples/all/).

GIMP (GNU Image Manipulation Program) - retouching, image composition, and image authoring.

ImageMagick (images from the command line)

(http://www.fmwconcepts.com/imagemagick/index.php)

D3.js (interactive data-based graphical forms which run in web browsers)

http://github.com/mbostock/d3/wiki/Gallery.

Cytoscape (complex networks)

Circos (relationships or multilayered annotations of one or more scales)

Best Practices for Making figures

CHECKLIST:

- ☐ Figure width
- ☐ File type (tiff or EPS) and resolution
- ☐ Does the journal allow or encourage you to use color?
- ☐ If the use of color is encouraged, is the color space RGB or CYMK?
- ☐ Font size and type
- ☐ Line weight

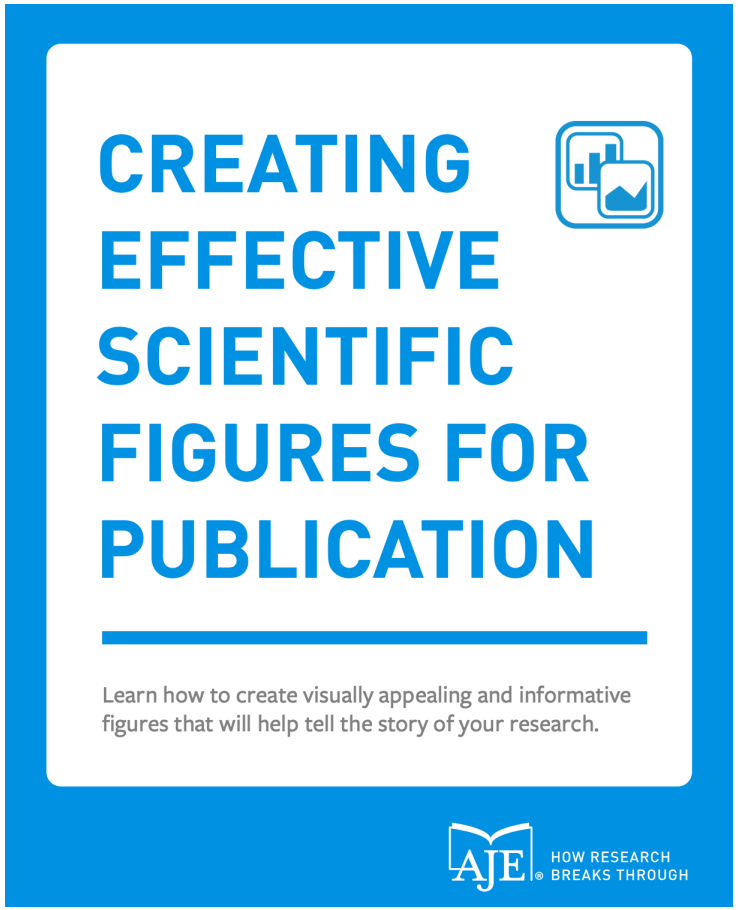
https://www.aje.com/dist/docs/Guide-Creating-Effective-Scientific-Figures-for-Publication.pdf

CREATING EFFECTIVE SCIENTIFIC FIGURES FOR PUBLICATION

Learn how to create visually appealing and informative figures that will help tell the story of your research.

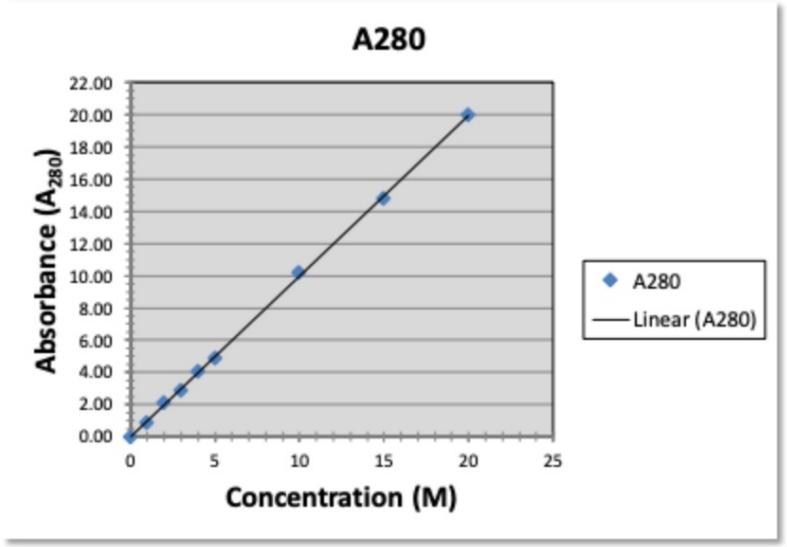
AJE® HOW RESEARCH BREAKS THROUGH

Best Practices for Making figures

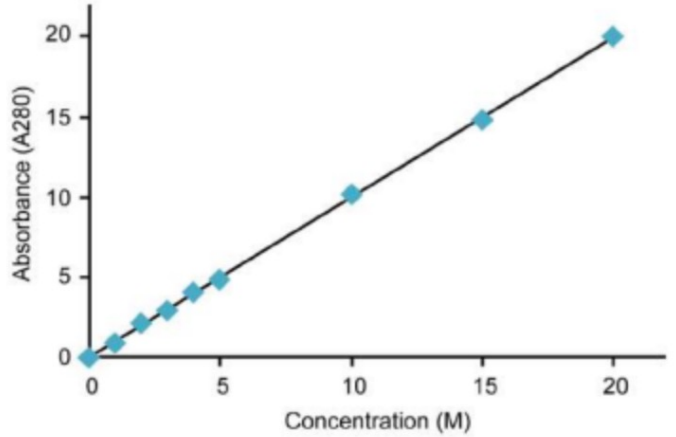


https://www.aje.com/dist/docs/Guide-Creating-Effective-Scientific-Figures-for-Publication.pdf

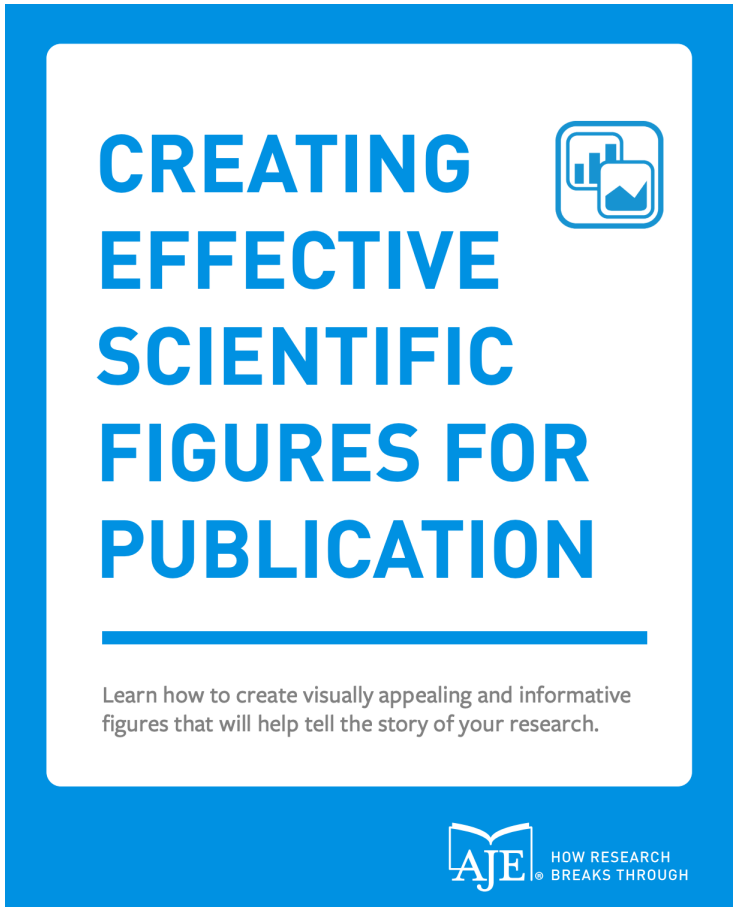
BEFORE (DEFAULT GRAPH):

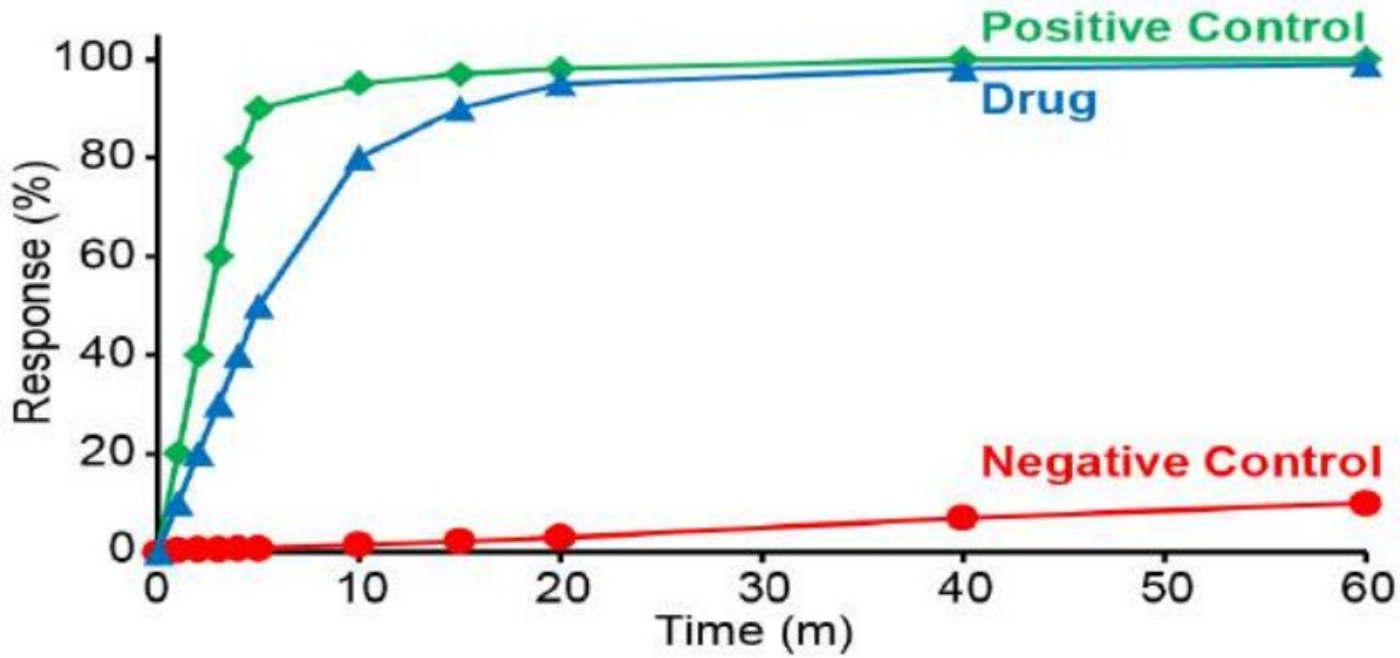


AFTER (EDITED GRAPH):



Best Practices for Making figures





https://www.aje.com/dist/docs/Guide-Creating-Effective-Scientific-Figures-for-Publication.pdf

Tips for good visualization

Source: https://bit.ly/2XXDAUv

Know your audience, understand what you're trying to convey, and how they'll interact with the information (Online? In a report? On a poster?):

- Do you want to explore the data (or let others explore)?
- Do you want to foster dialogue between two parties?
- Do you want to tell a compelling story?

Simplify, simplify. Eliminate anything you don't need, like:

- Data that aren't pertinent to your story (Note: this does not mean you get to cherry-pick which data looks best, but if certain variables are ancillary to the story, don't include them. Edit but don't skew, and be transparent about what you've excluded!)
- Make sure colors have a clear meaning and purpose. Don't go overboard.
- Tick marks and axis labels. Use as many as you need to make it clear, but don't go overboard.
- Excess numerical digits (your axis does not need to be labeled $1.894327849 B; $1.9 B will work just fine).
- Remove unnecessary grid lines, boxes, and tick marks.
- 3D charts or embellishments: they almost always make things more confusing.

Fancy software is not needed to make good visualizations (but it can help).

People should be able to understand the whole story by looking at the abstract, figures, and figure captions. All of the following should be good, descriptive, and BIG (it must be readable):

- Titles
- Captions
- Annotations
- Axes (with units!)
- Legends (if you have to have one)
- Scale bars (like a temperature gauge)
- Sources for your data, if it's from someone else

Label things directly wherever possible. Don't make the reader work to figure out what the orange star means.

Avoid uncommon acronyms wherever you can. If you have room to spell it out, and it's not something obvious like USAID, do it! This prevents the reader from having to do unnecessary work to search within the text to find what the acronym means.

Be consistent wherever you can:

- Axis limits: if you compare two plots, make sure the x-axis and y-axis have the same range.
- Have colors mean the same thing everywhere.
- Start axes at zero unless you have a good reason not to.

Order your variables in a meaningful way.

You can also group and transform data. Convert to millions; create a new ratio; calculate a percentage.

Don't mislead people with your representation.

- Don't cherry-pick data. If you exclude data, have a justifiable reason that is more than "It makes the figure look better."
- Cite where you got your information.
- Be clear if info is zero or if it's missing.
- Highlight any limitations of the data or uncertainty in their measurement, such as using error bars.

Sketch and try out different options. Every chart type, color, symbol, and placement has its own merits; see what works for your data and your story.

Most importantly: Good data visualizations will not save you from bad data.

Adapted from Geocenter (Laura Hughes).

@OpenAcademics