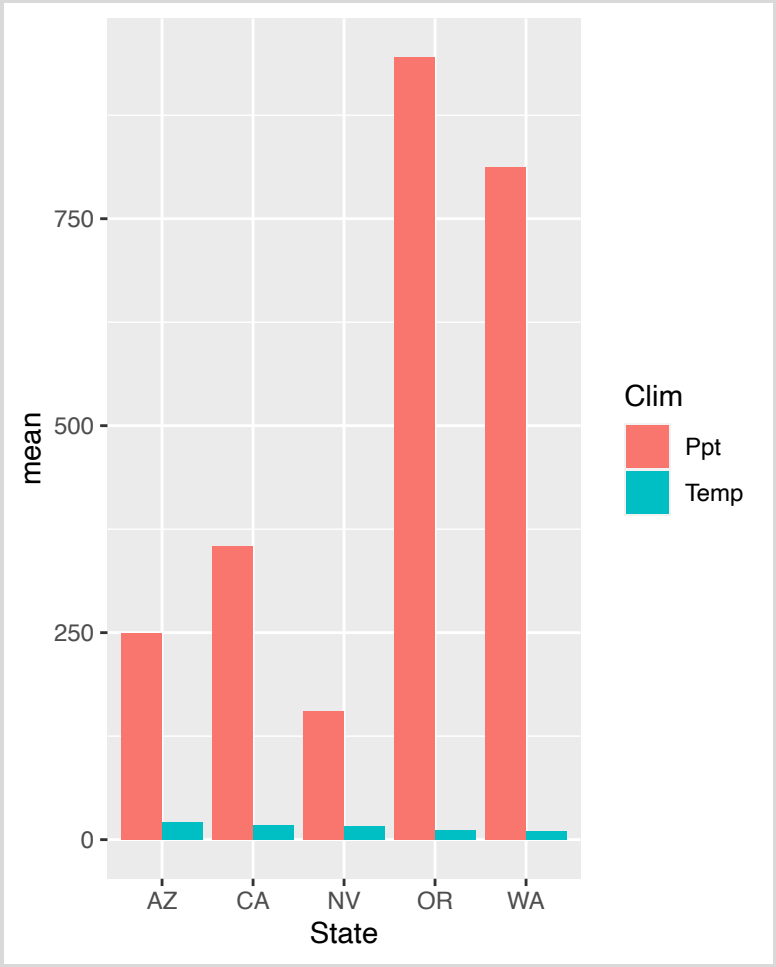
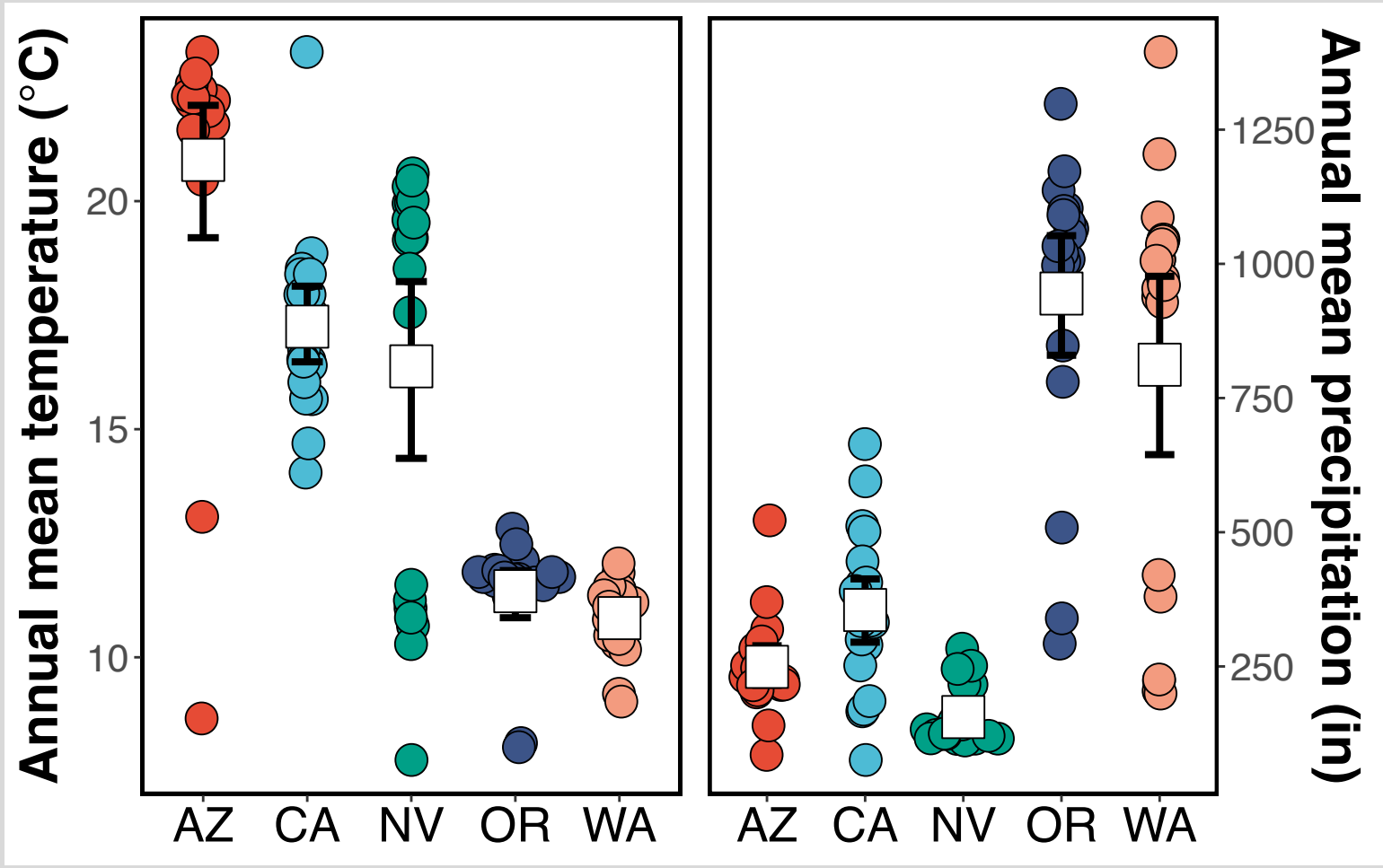
A guide to okay figures from an idiot

Why this is bad



and this is okay



https://github.com/trevorfaske/DataVisualization

Intro_ggplot2.ipynb	edit README	yesterday
Intro_ggplot2.pdf	add changed ipynb and readme	yesterday
Intro_ggplot2_follow.R	add changed ipynb and readme	yesterday
README.md	edit README	yesterday
city_df.csv	add changed ipynb and readme	yesterday

README.md

Data Visualizaiton

Disclaimer: this a continual work in progress

Data visualization is a key component to any scientific journal or popular science article. Being able to tell a compelling story using just the data at hand should be the goal of any figure.

All of the code generating figures will be done in **R** using the package **ggplot2**. While I hope to incorporate some philosphy of good data visualization practices, a majority of these modules will be a *how to* on generating various types of figures within ggplot.

Resources

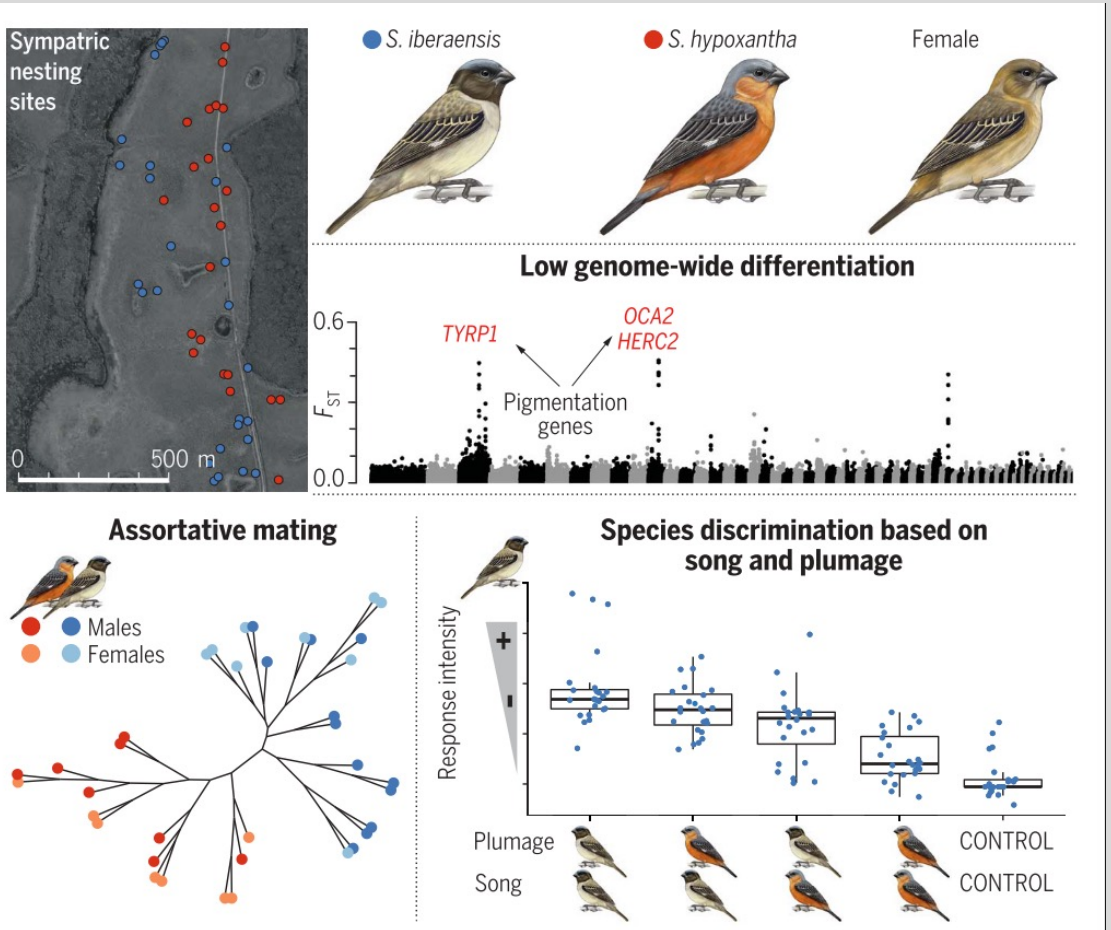
- Rougier, N. P., Droettboom, M., & Bourne, P. E. (2014). Ten Simple Rules for Better Figures. *PLoS Computational Biology*, *10*(9), 1–7. https://doi.org/10.1371/journal.pcbi.1003833
- https://robjhyndman.com/hyndsight/graphics/
- Nature collections: Visual strategies for biological data (pdf)
- https://clauswilke.com/dataviz/

General figure tips

Know your audience

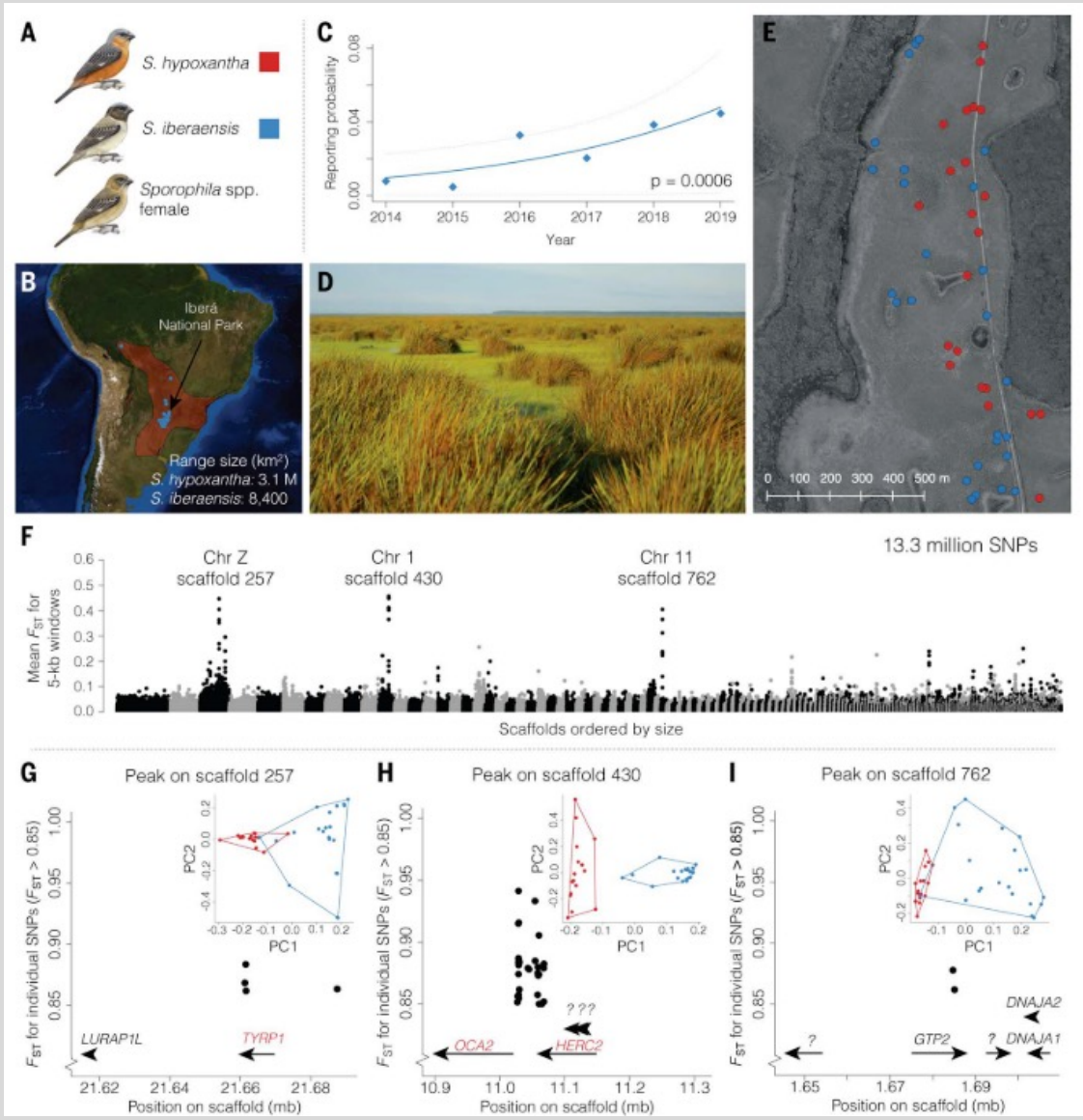
Turbek et al. (2021) *Science*

Broad



Novel mating signals restrict gene flow between co-occurring bird species. *Sporophila iberaensis* was first observed in 2001 and has a breeding range contained entirely within that of *S. hypoxantha*. Despite extremely low genomic differentiation, both species mate assortatively. Genetic differentiation is concentrated near genes known to be involved in plumage coloration. Field experiments show that both song and plumage are used to recognize sexual competitors.

Specific



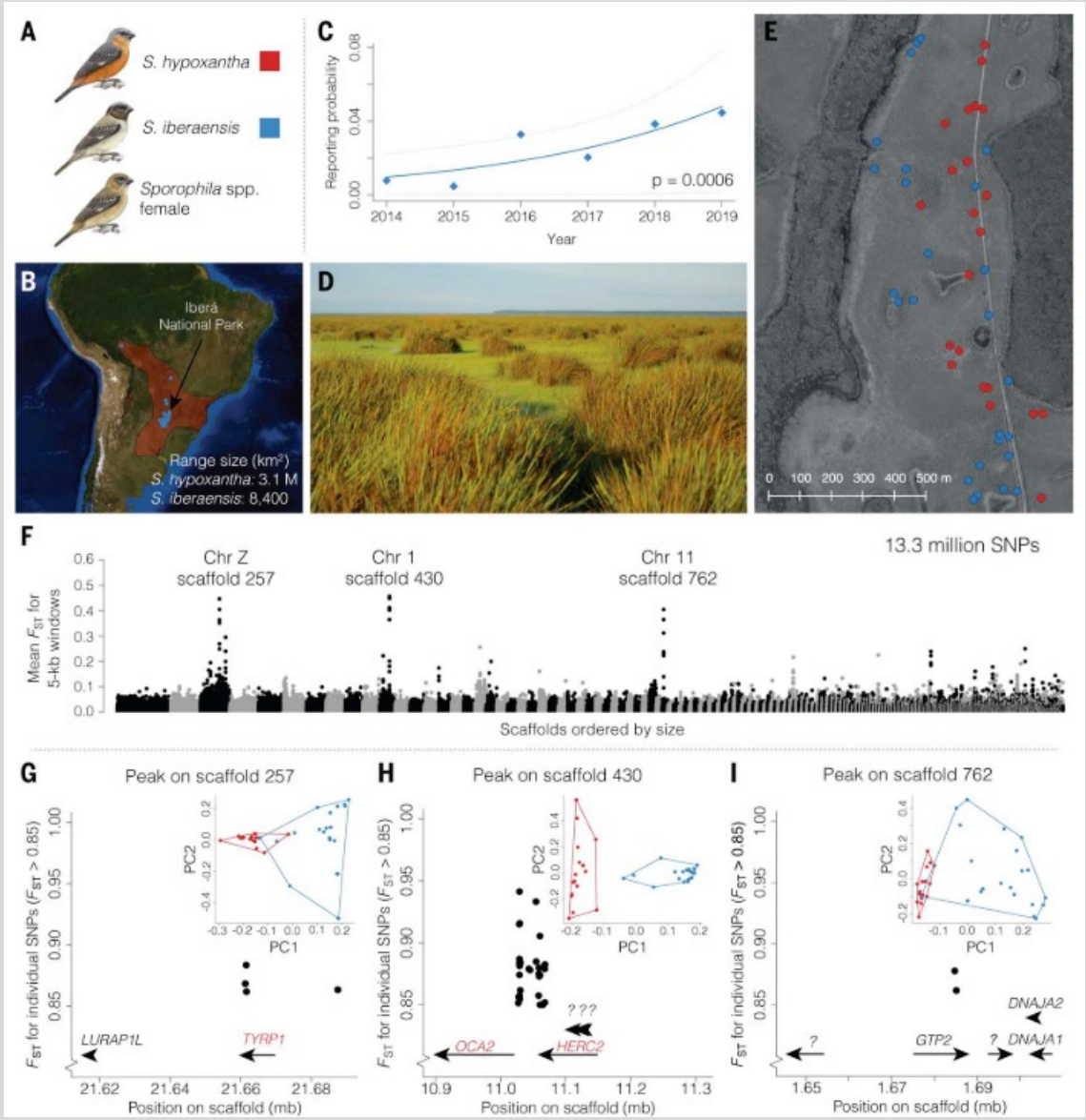
Know your medium

Talk figures are different from paper figures

Talk figures are different from paper figures

Paper

- Complicated
- Multi-panel
- Can be aided by caption
- Reader has prolonged time to process info



Turbek et al. (2021) *Science*

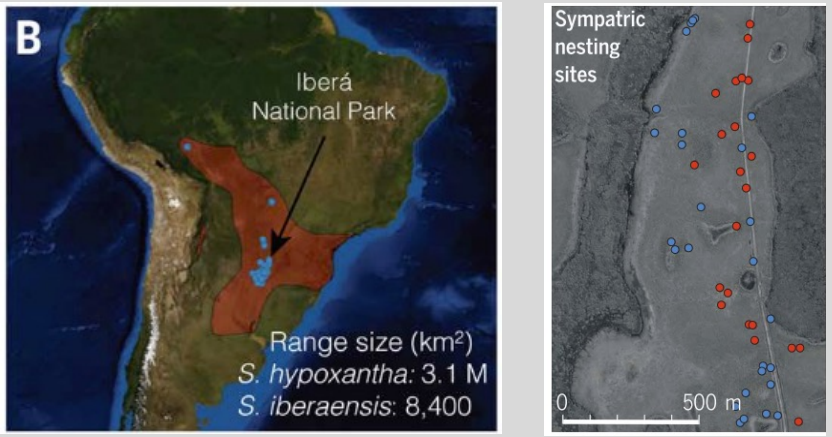
Talk figures are different from paper figures

Talk

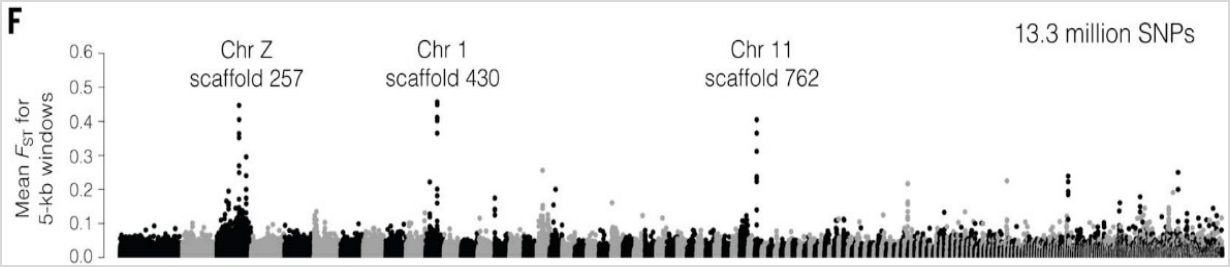
- Simple
- Reader has seconds to understand info
- Can be aided by animation (arrows, boxes, text)
- No space/color limit
- Multiple slides
- Only highlighting parts of interest

Slide 1

Slide 2

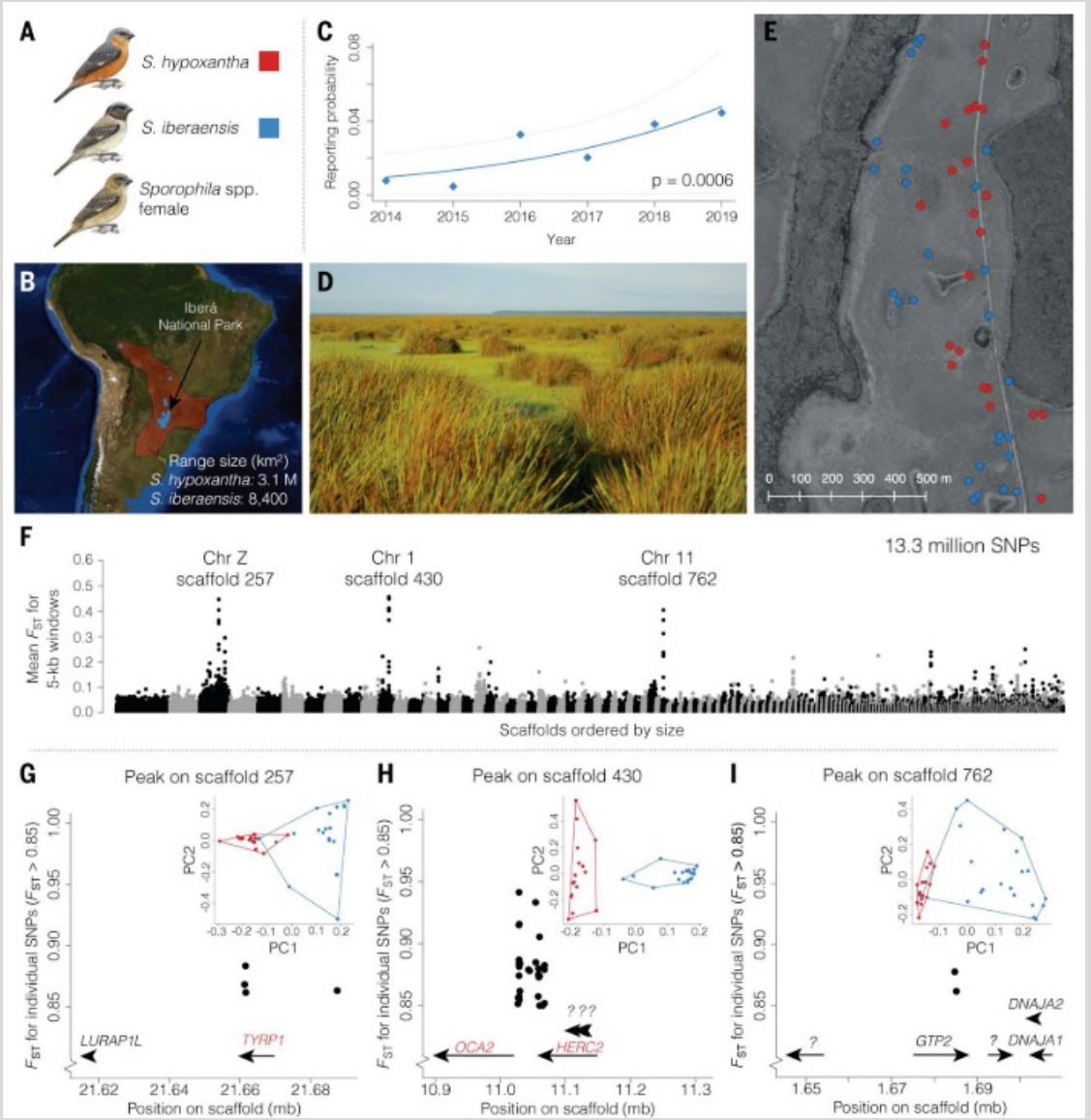


Slide 3



Turbek et al. (2021) *Science*

Can a reader understand what's happening without reading the paper

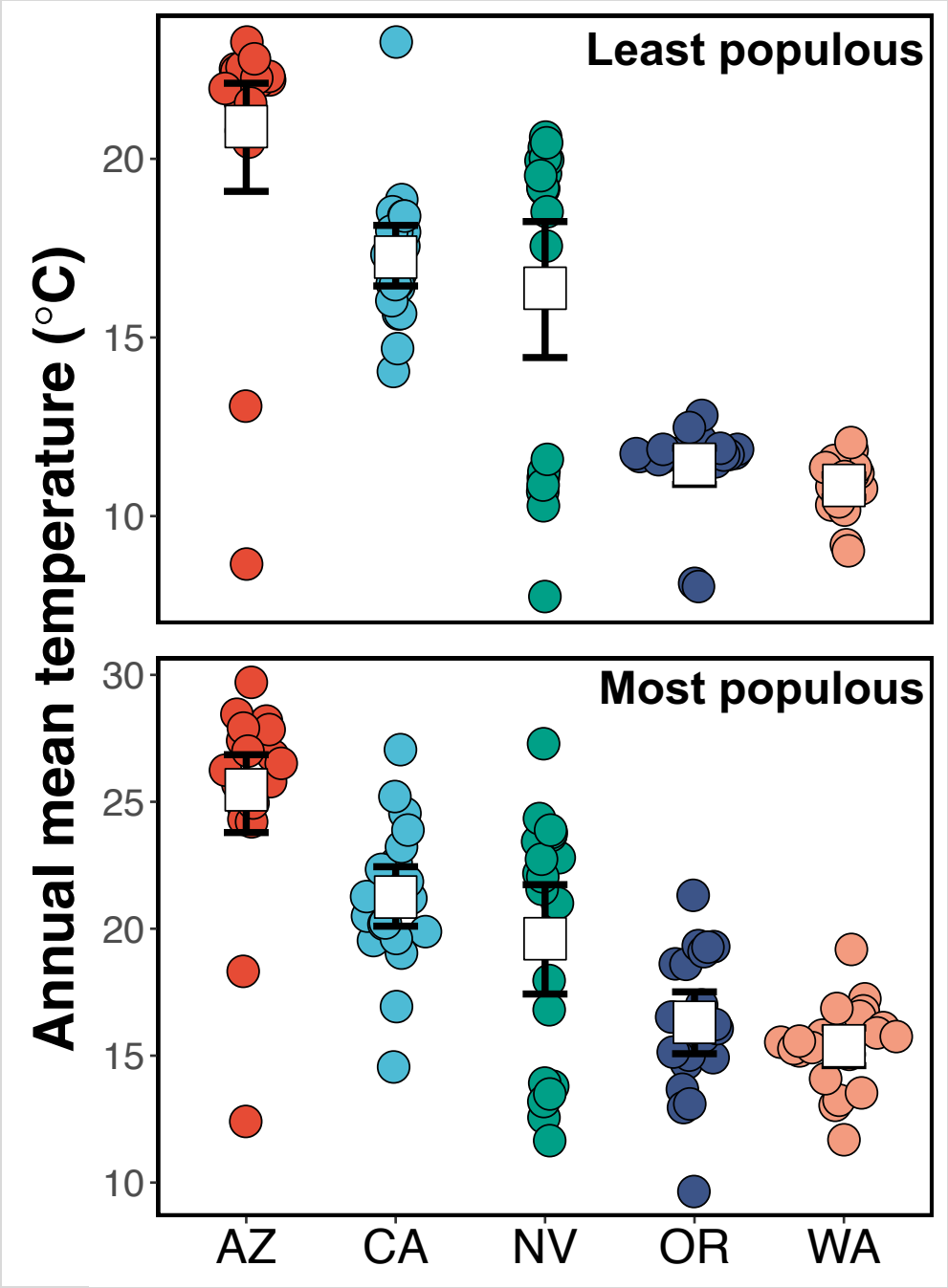


Turbek et al. (2021) *Science*

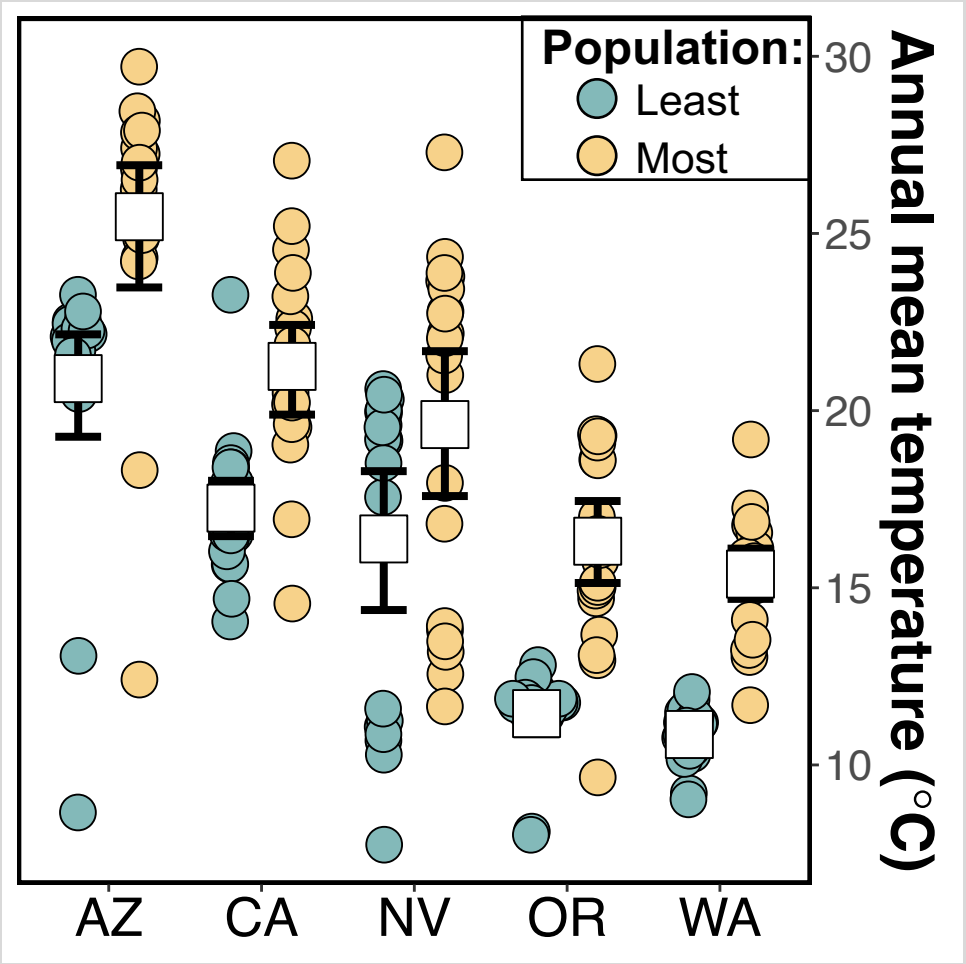
Identify your message

- What is it you want to show?
- Make sure comparing the right thing
- Don’t use unnecessary or redundant figures
- Ask: Does this add to the story or add confusion
- When in doubt, SUPPLEMENT (figure/table)

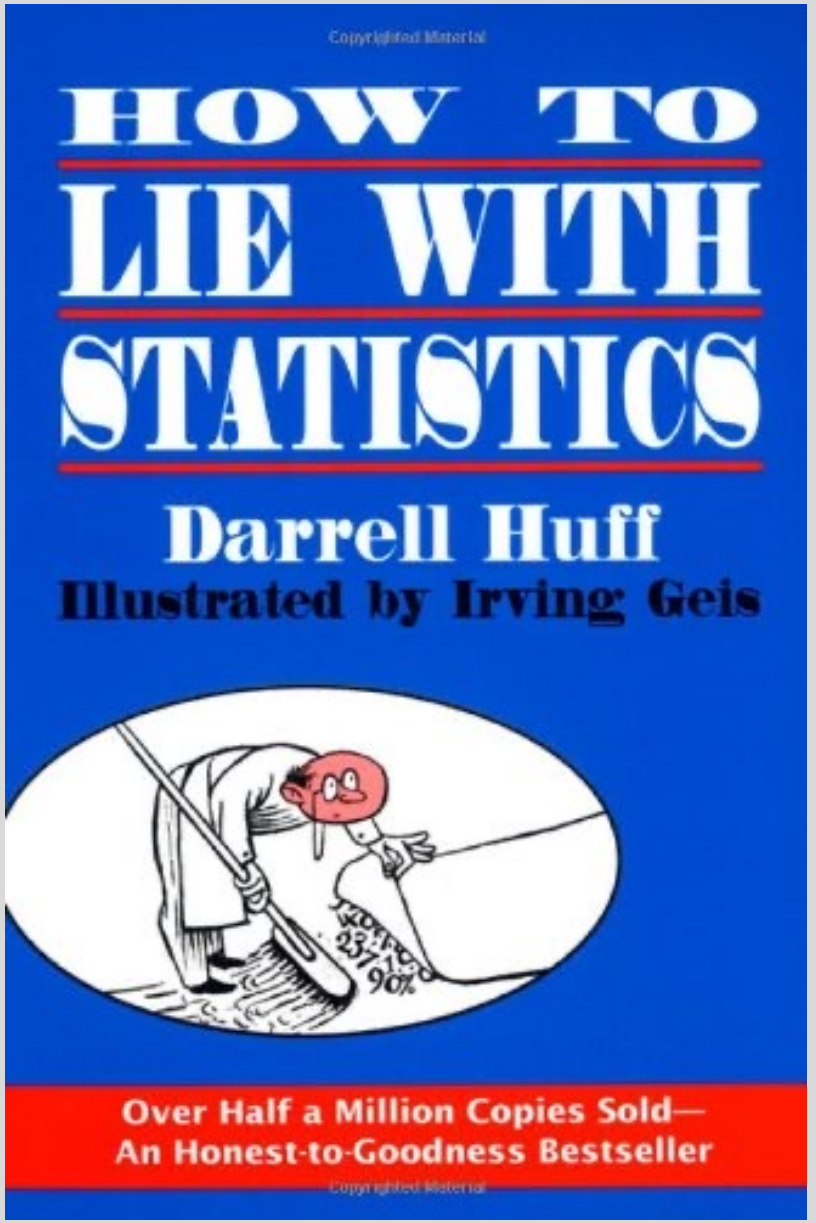
Comparison between states



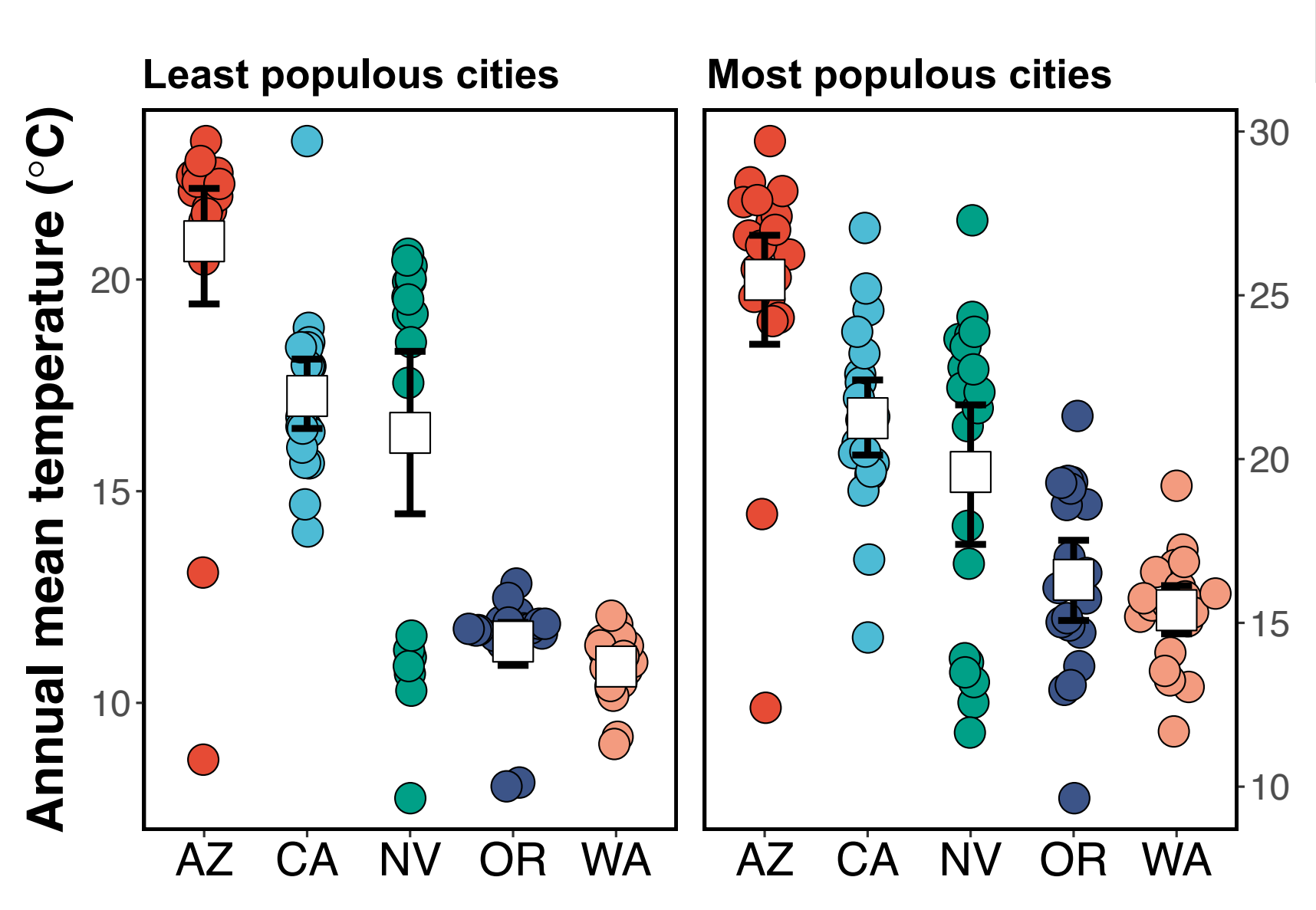
Comparison between city populous size



Don’t be misleading

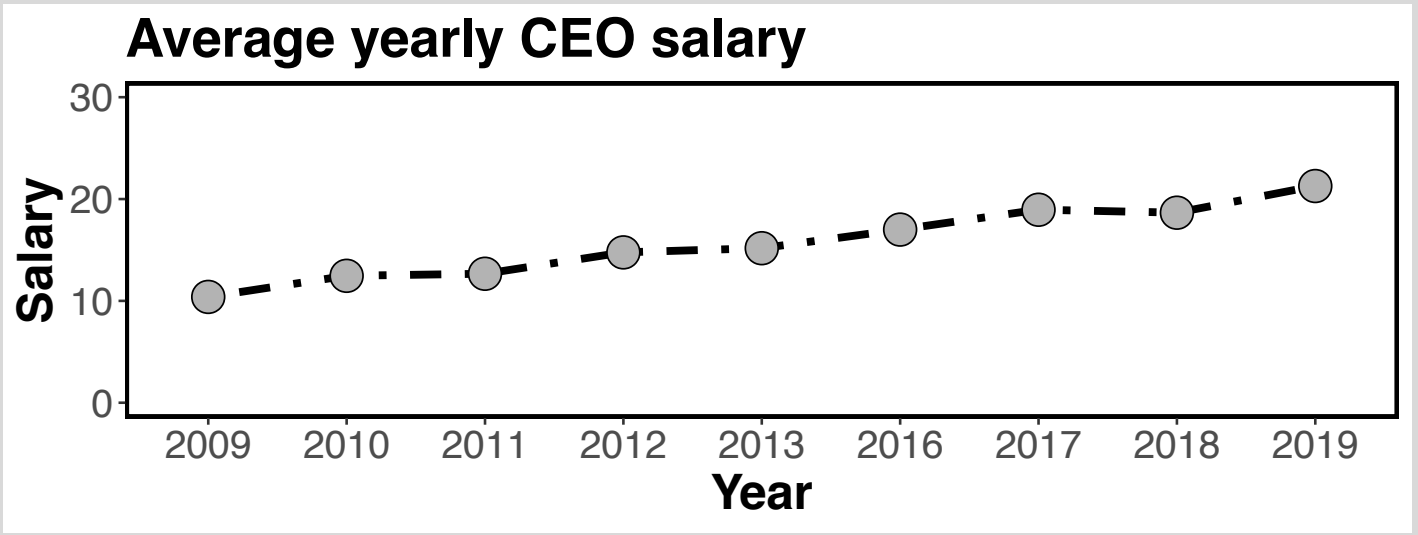


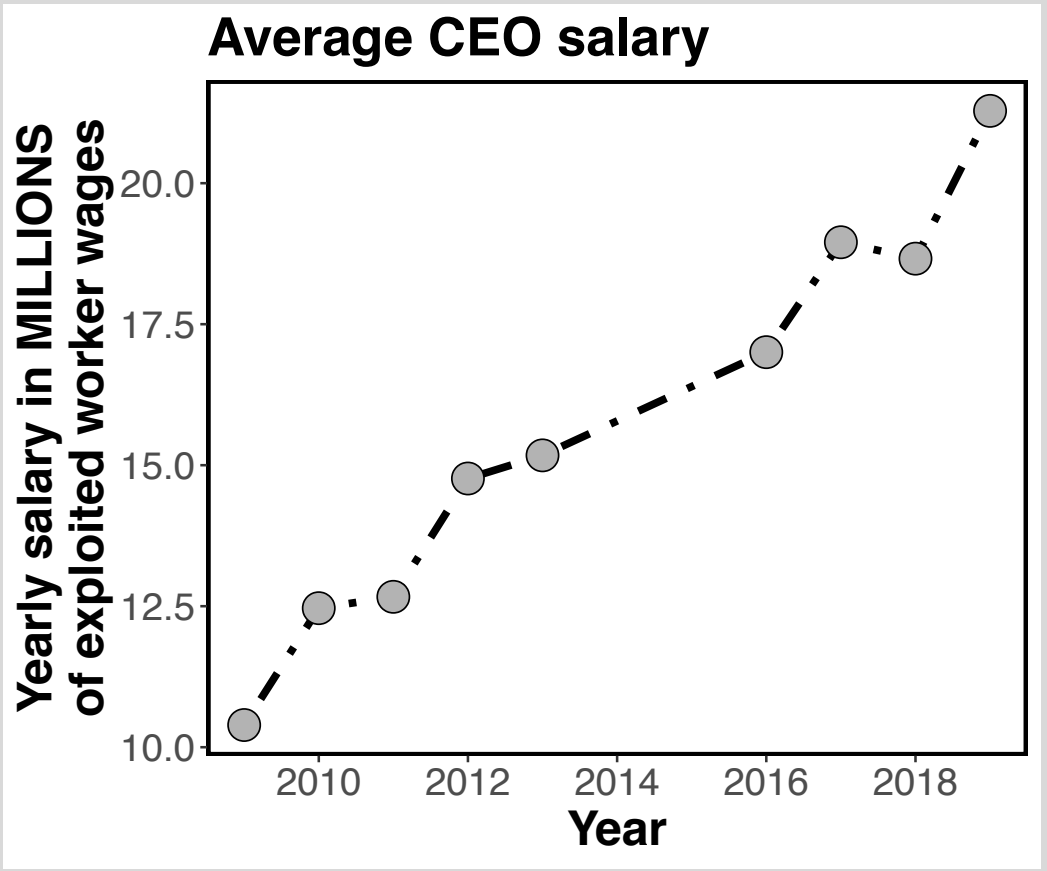
Why is this misleading?



Look at me, I'm a nice guy who cares about the environment

JUST KIDDING, I want to cultivate Mars and back a coup in Boliva for battery juice

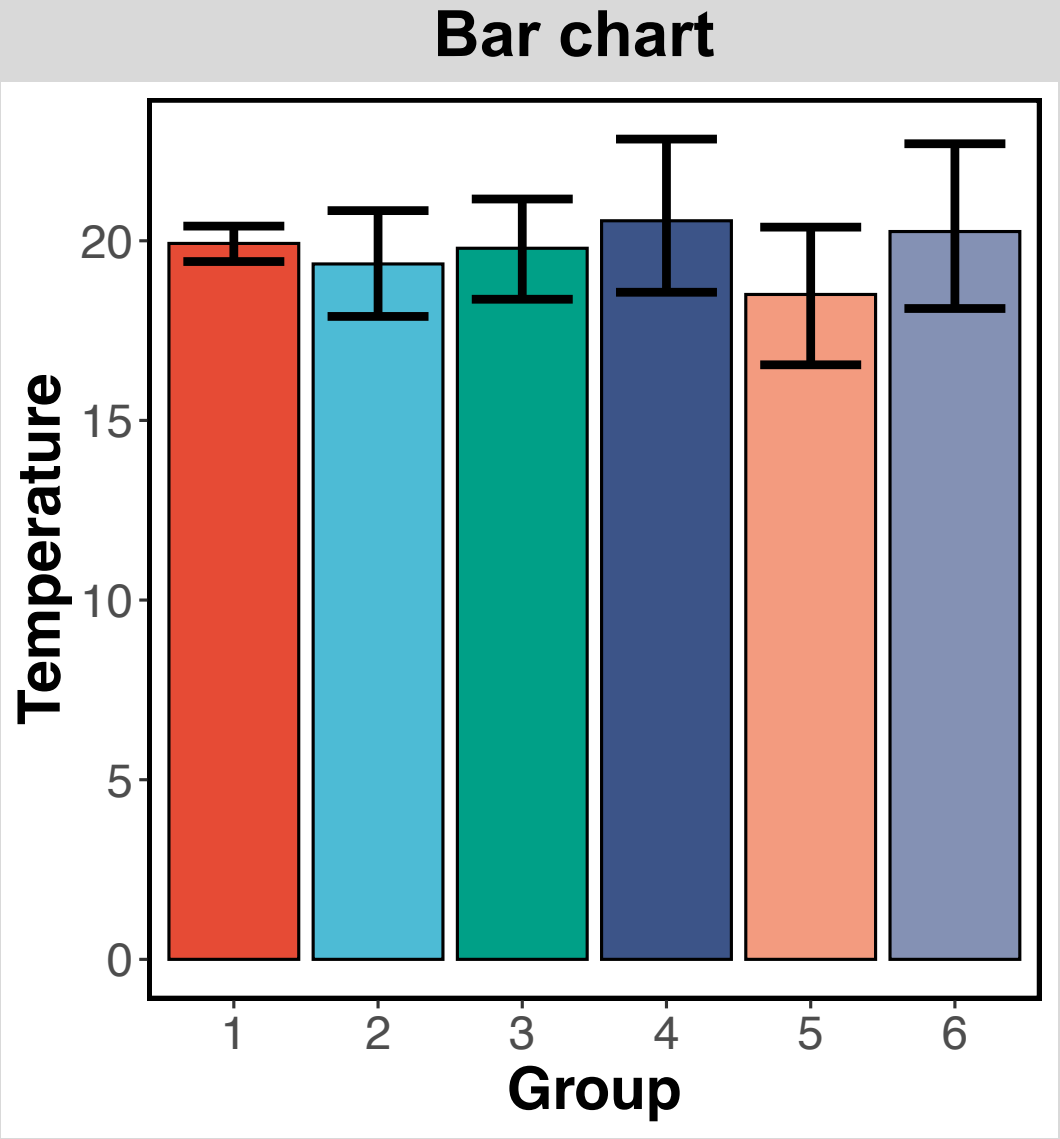




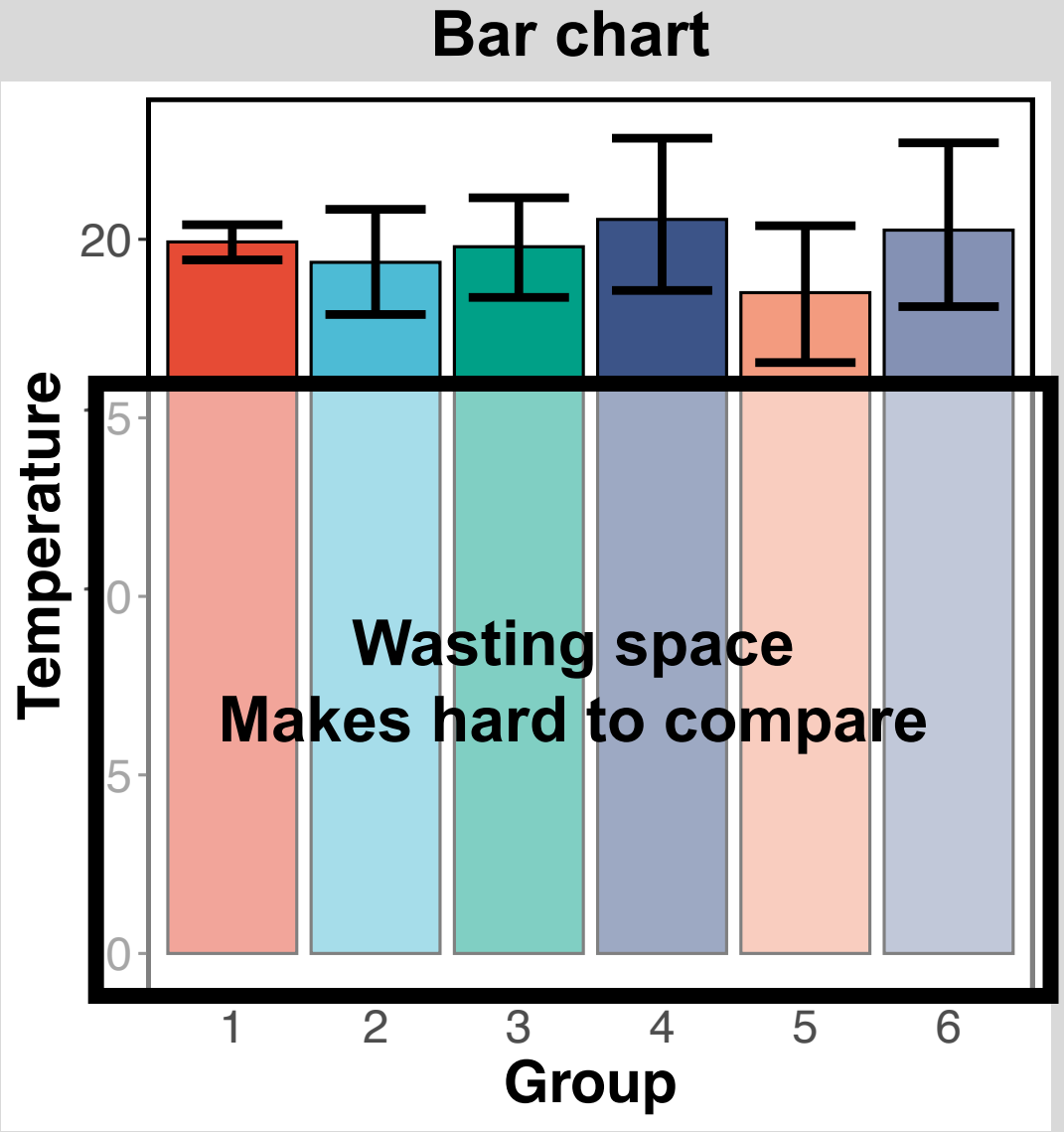
Common ways to be misleading

- Scales
- Leaving out data/outliers
- Not stating what stats are in figure (mean, median, sd, 95% ci, etc.)
- Jittering points
- Sample sizes
- Summarizing data is lying, do it with caution

Misleading continued: Show the data!

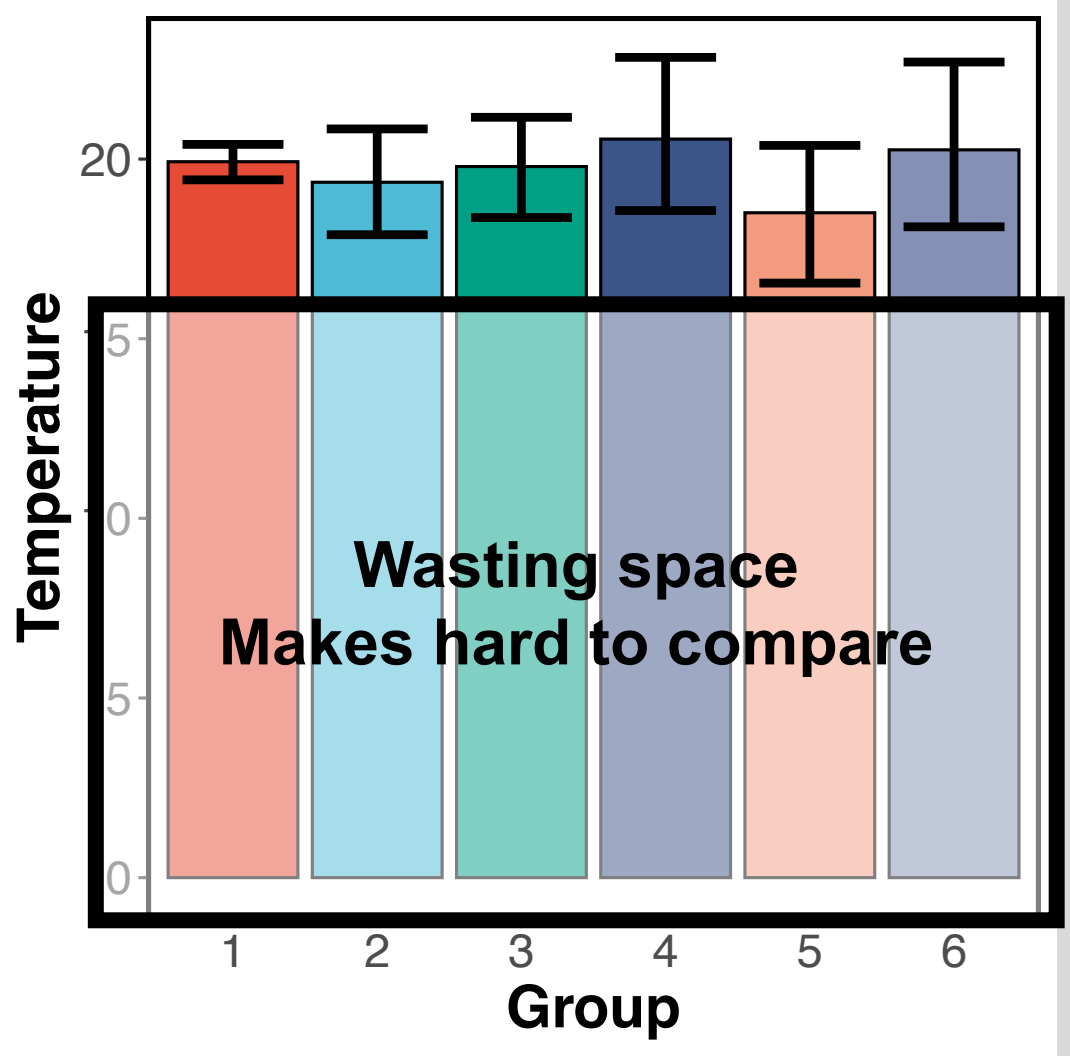


Misleading continued: Show the data!

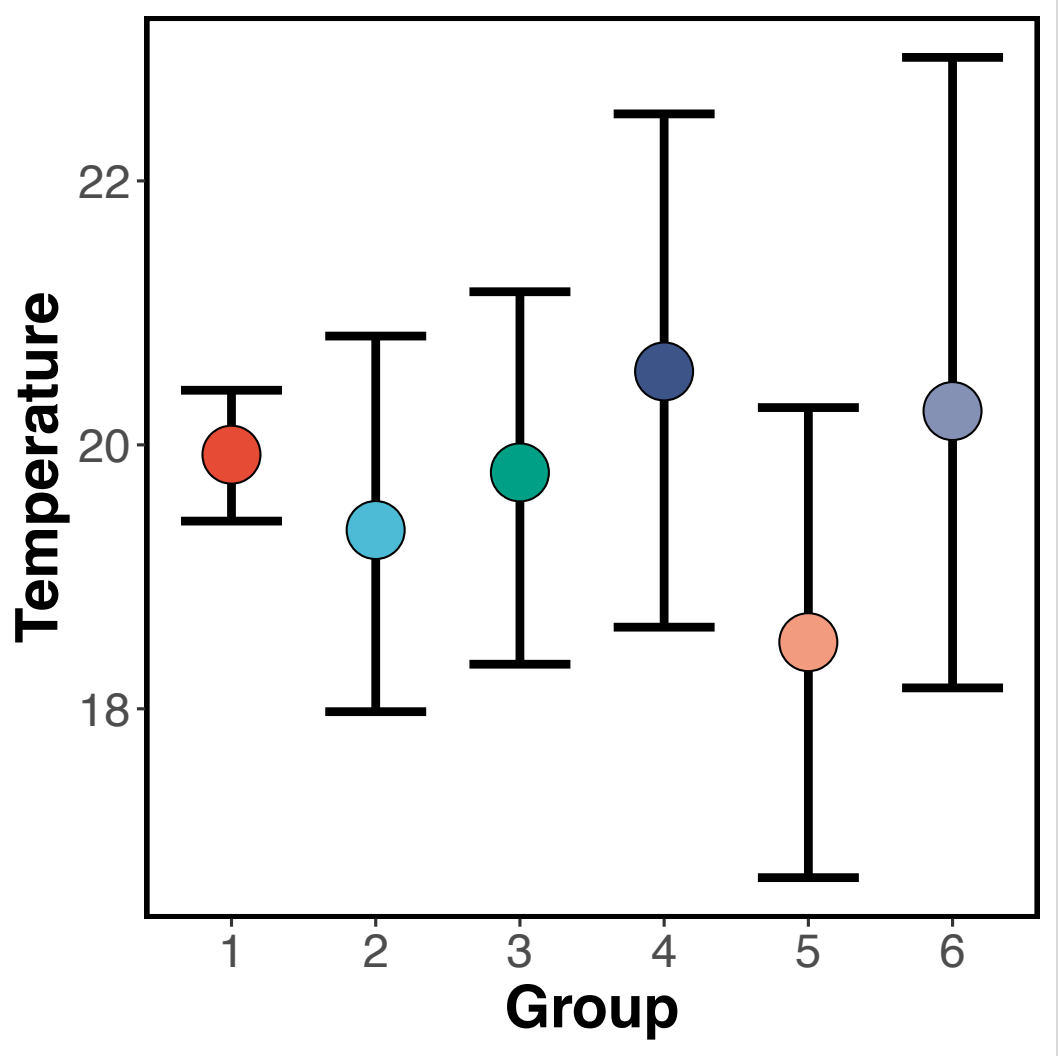


Misleading continued: Show the data!

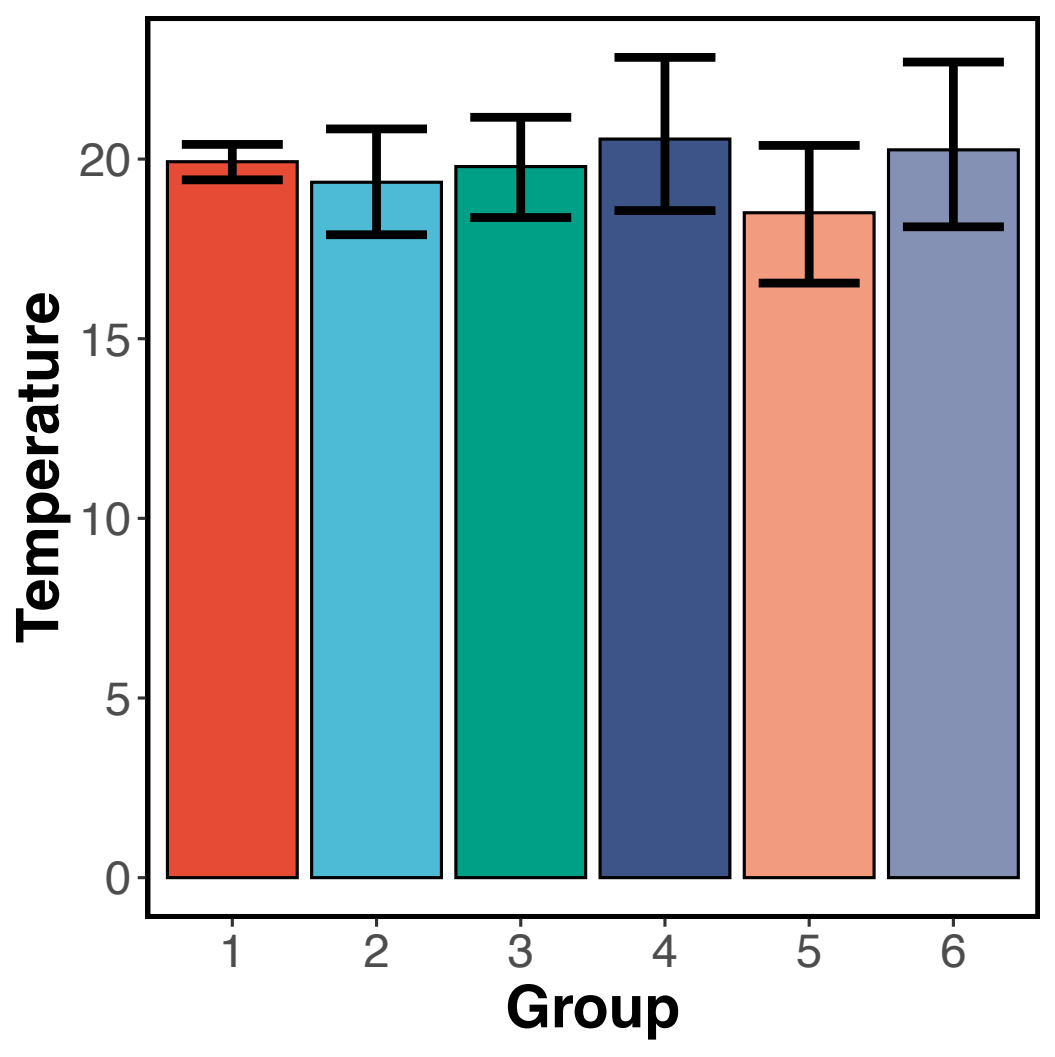
Bar chart

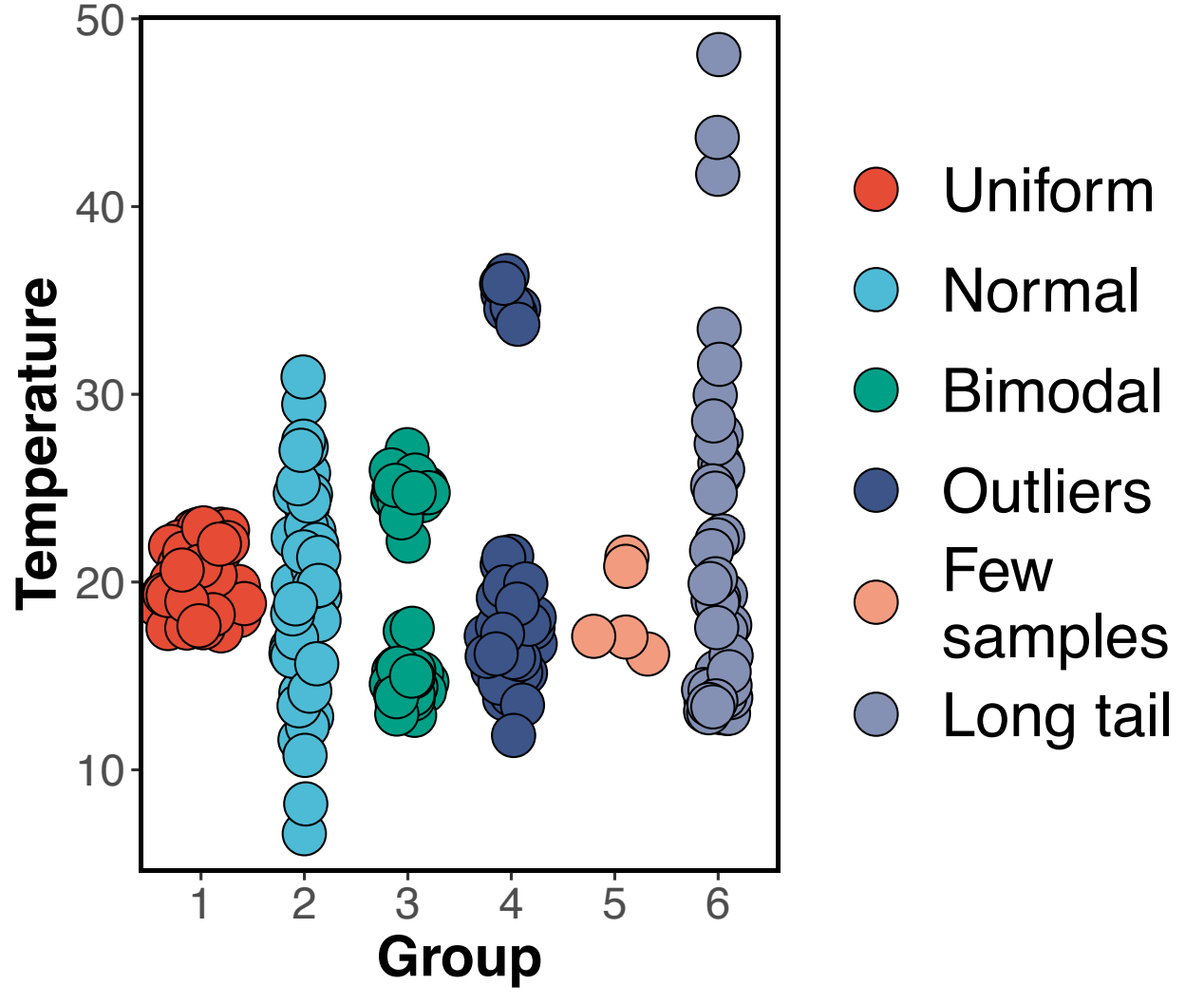


Point and error

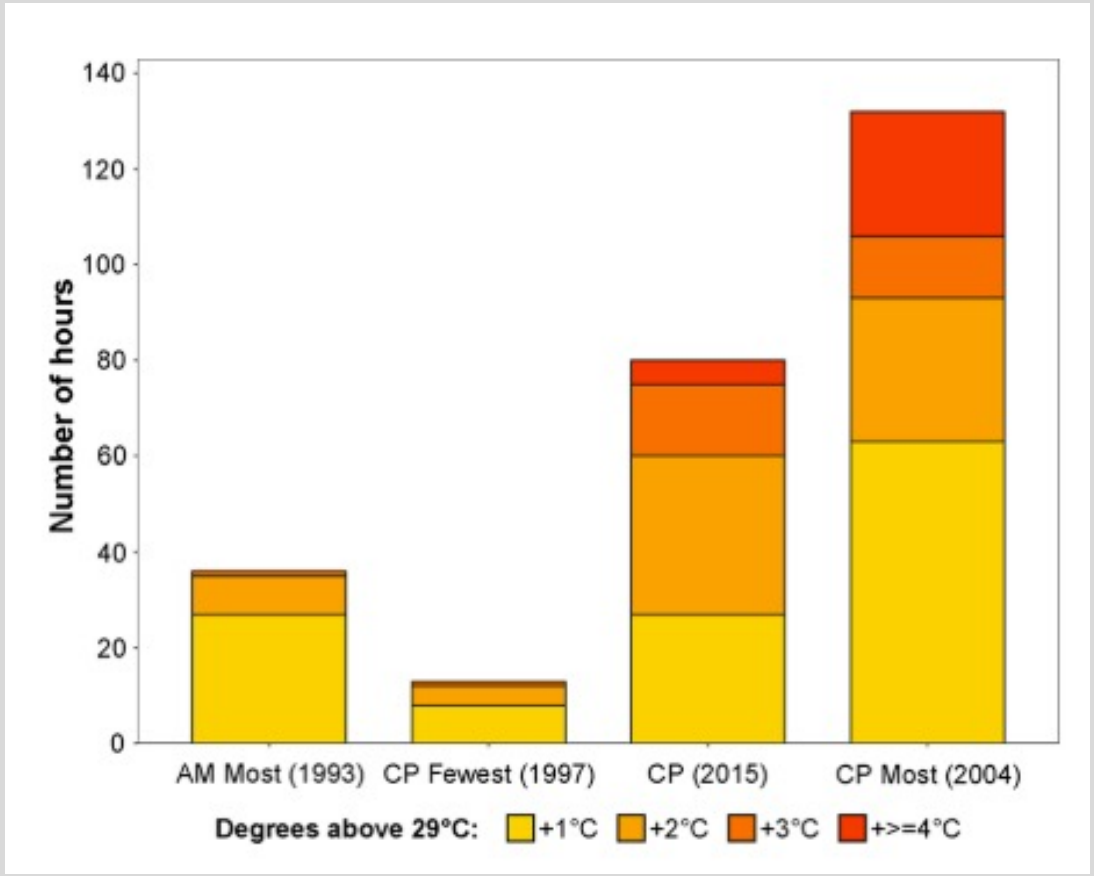


This is the exact same data…..

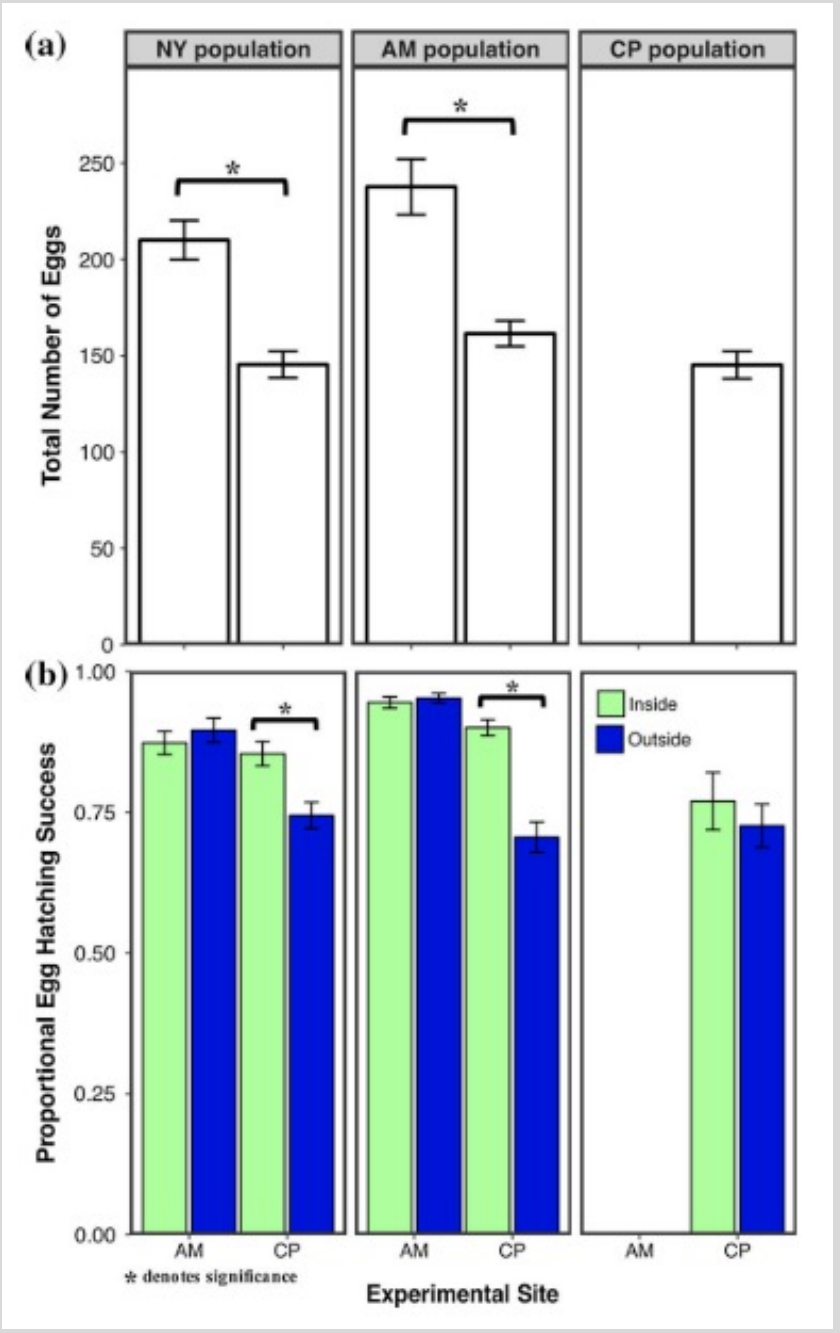




Don’t feel bad about past / current figures!



Faske et al. (2019) *obviously not Science*



More specific tips

Colors

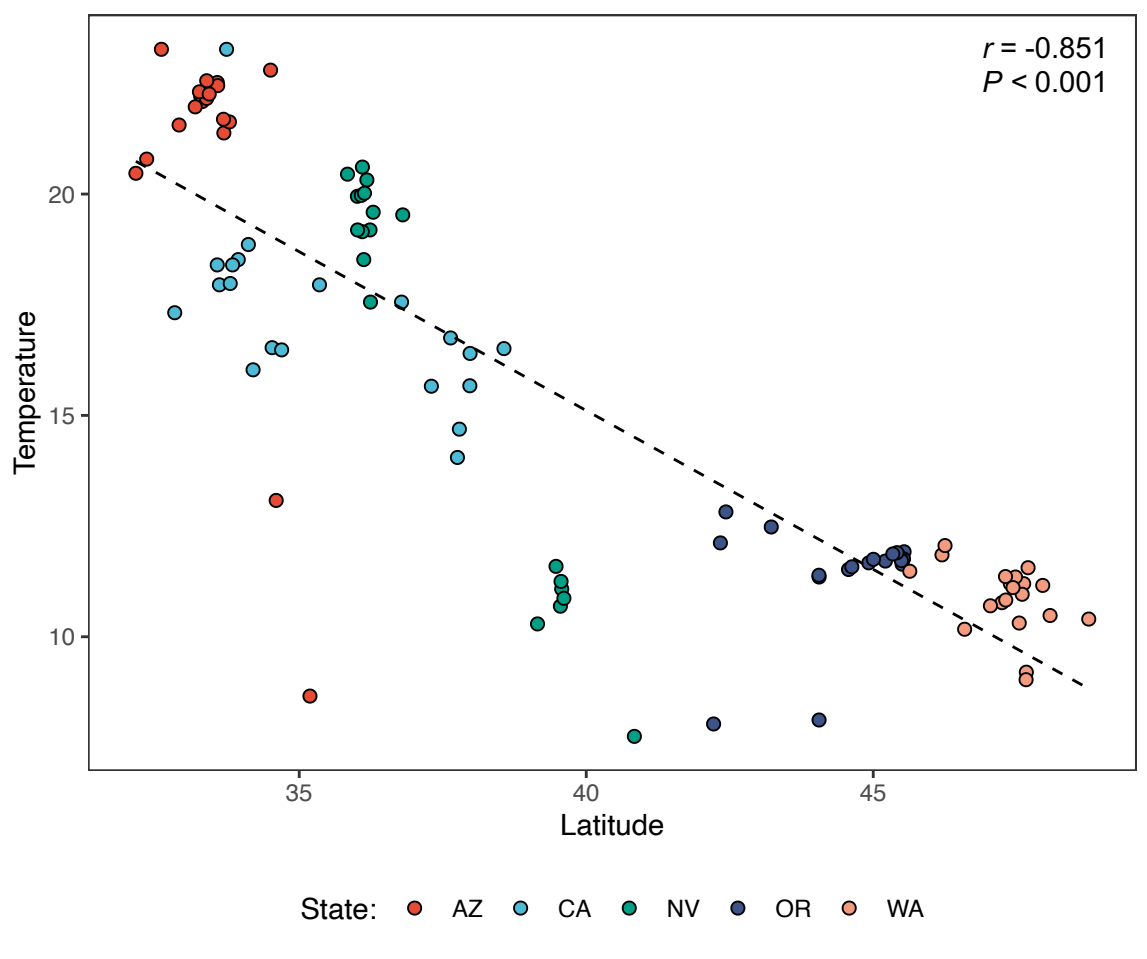
- Color blind-friendly (https://colororacle.org/)
- Ask: Can this be black and white? ($$$)
- Have colors make sense (Temp: hot = red, cold = blue)
- CONSISTANCY!!!
  - If species A is a blue square in Figure 1, species A should be a blue square in Figure 6

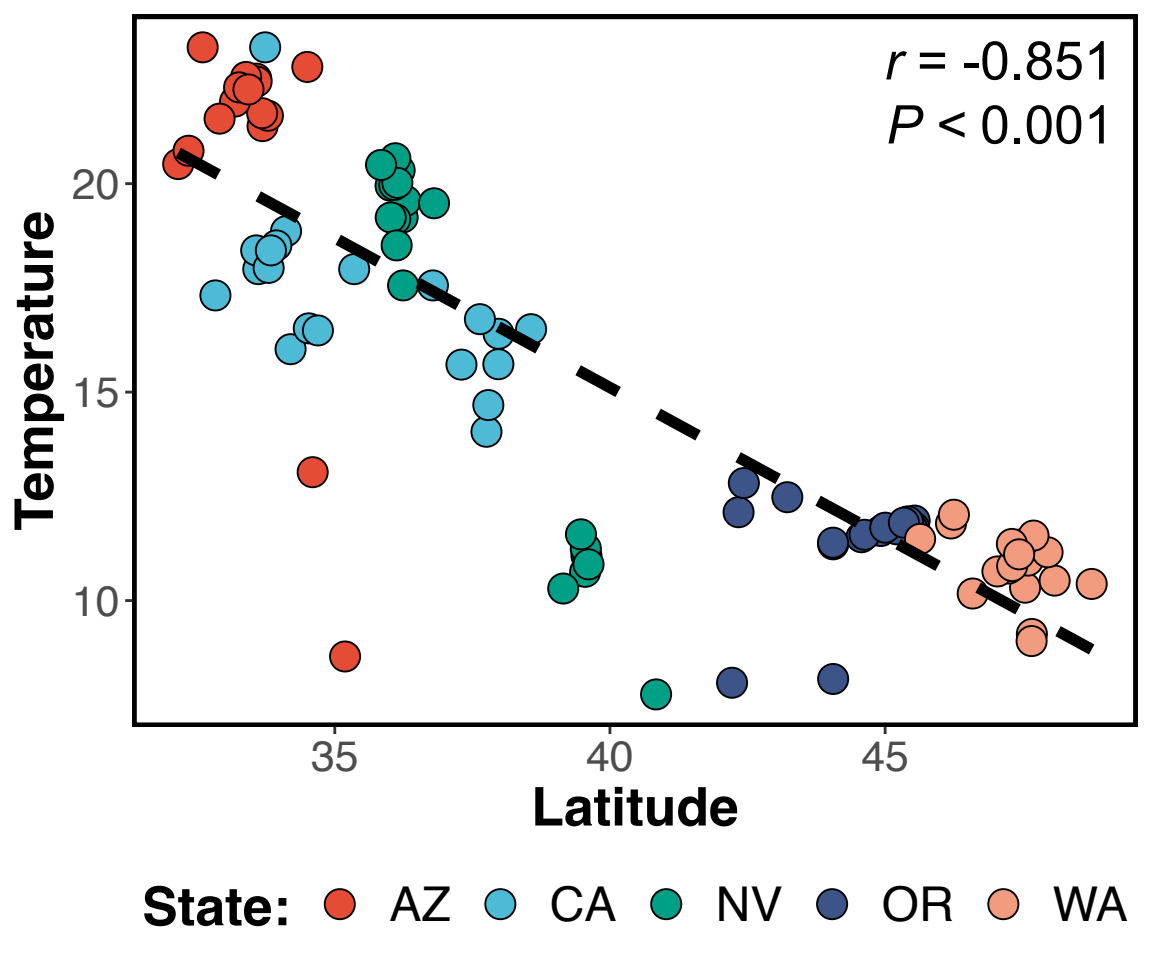
Color palettes

Just a few, not an exhaustive list:

- ggsci (https://cran.r-project.org/web/packages/ggsci/vignettes/ggsci.html)
- Wes Anderson (https://github.com/karthik/wesanderson)
- National Parks (https://github.com/katiejolly/nationalparkcolors)
- ColorBrewer2 (https://colorbrewer2.org/)
- Viridis (https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html)
- Paletton (https://paletton.com/)

Make text/labels/points annoyingly big





Typography/font: serif not ADA approved

Everything in this talk has been in **Arial**
Also, default font in ggplot2

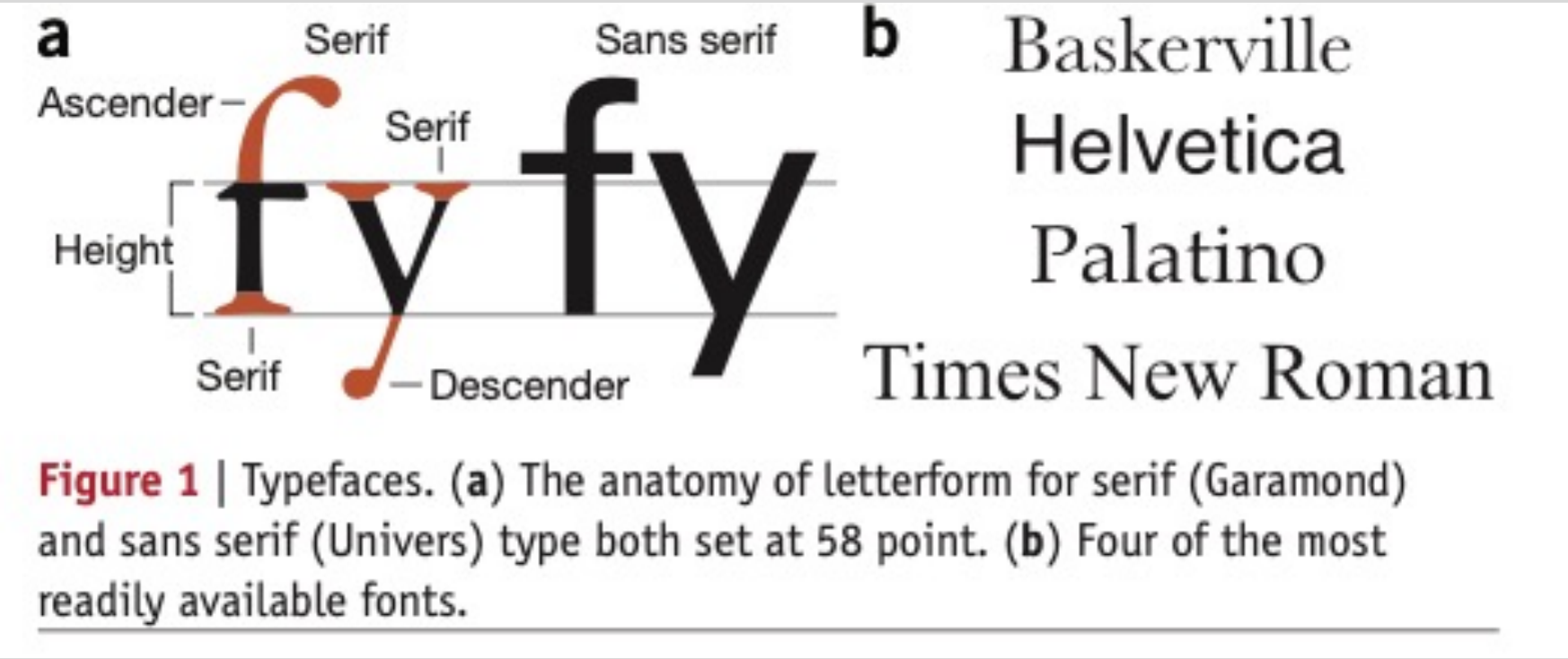
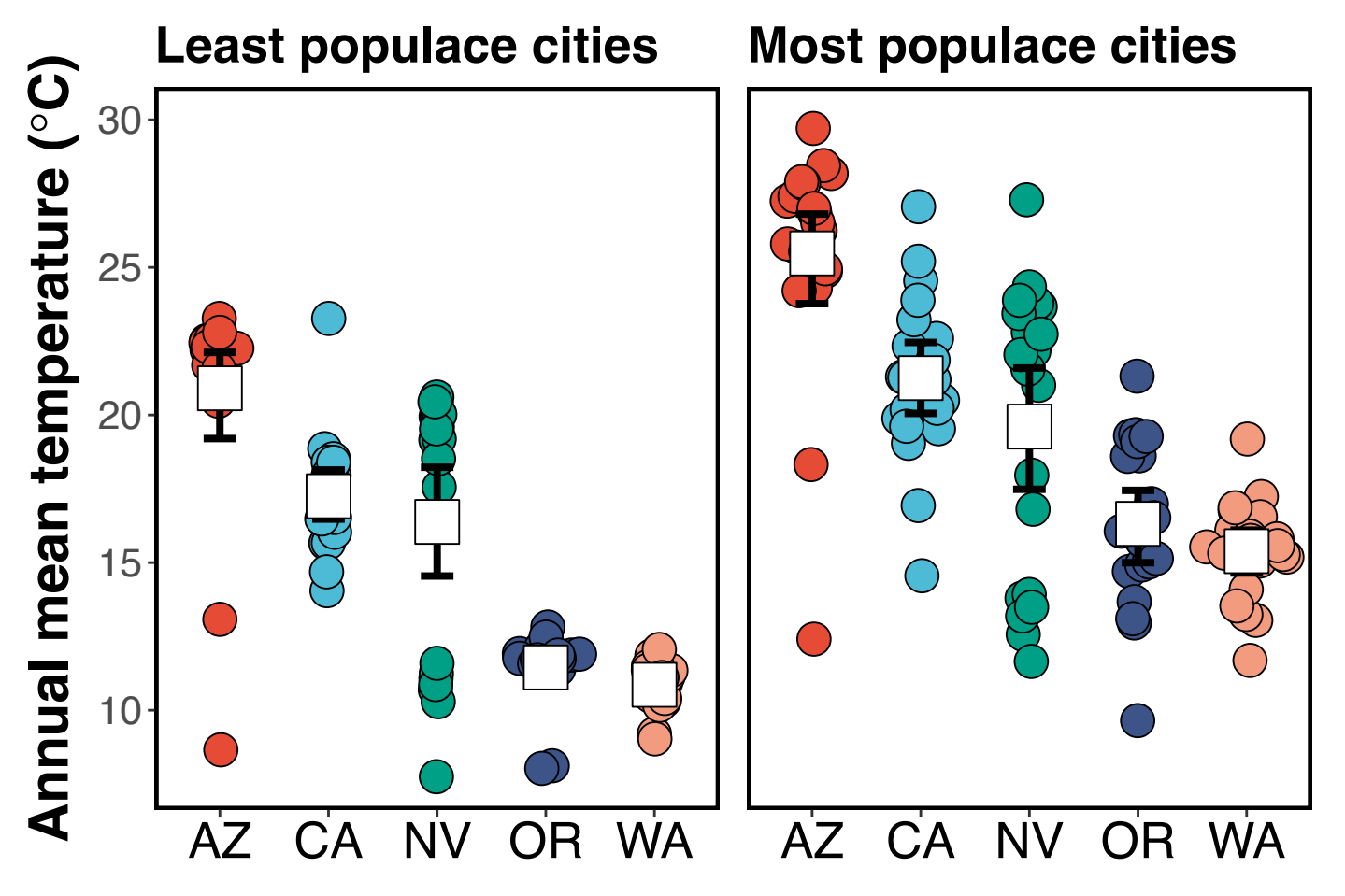
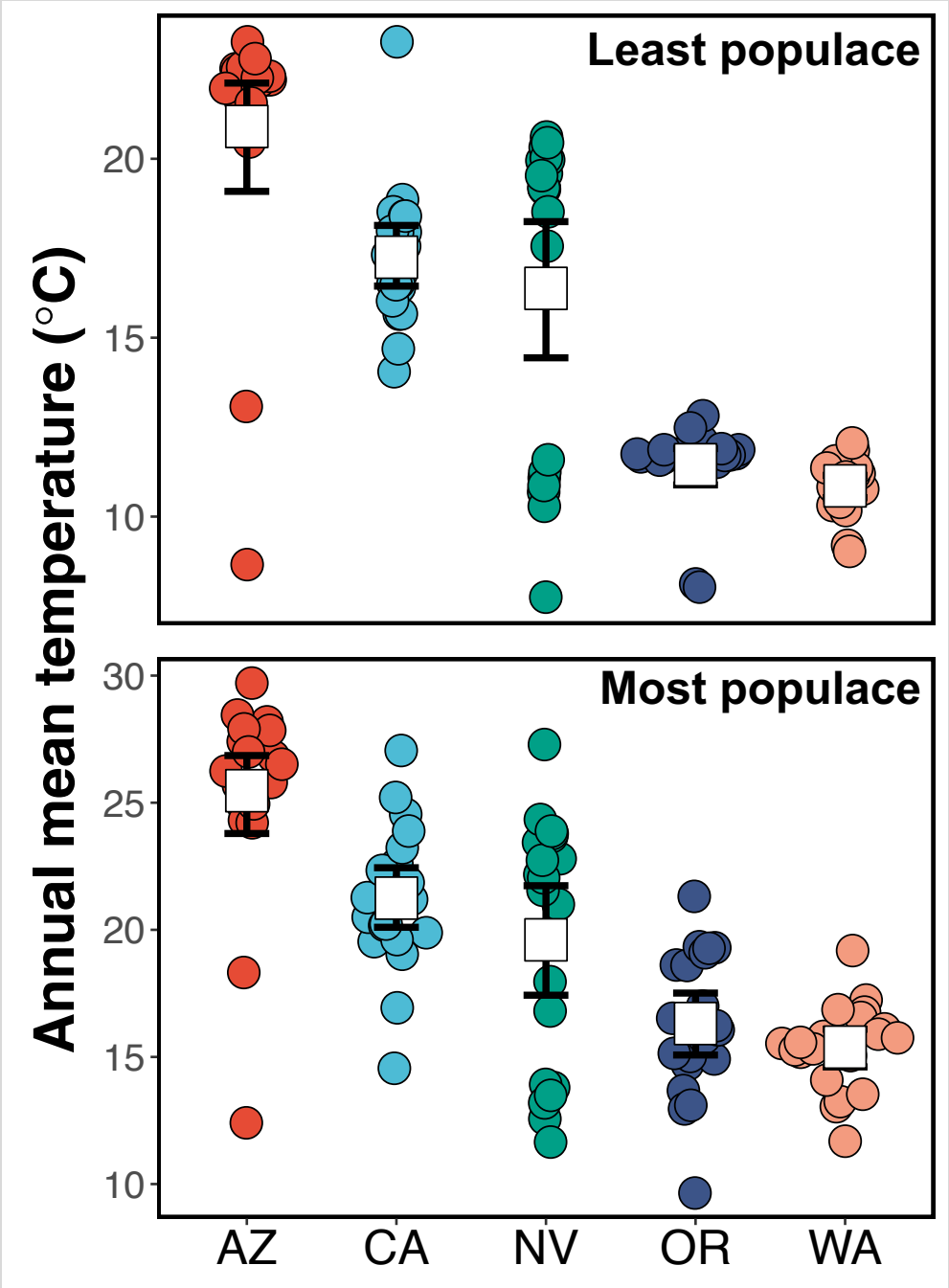


Figure 1 | Typefaces. (**a**) The anatomy of letterform for serif (Garamond) and sans serif (Univers) type both set at 58 point. (**b**) Four of the most readily available fonts.

Nature collections: Visual strategies for biological data (pdf)

Consolidate axes

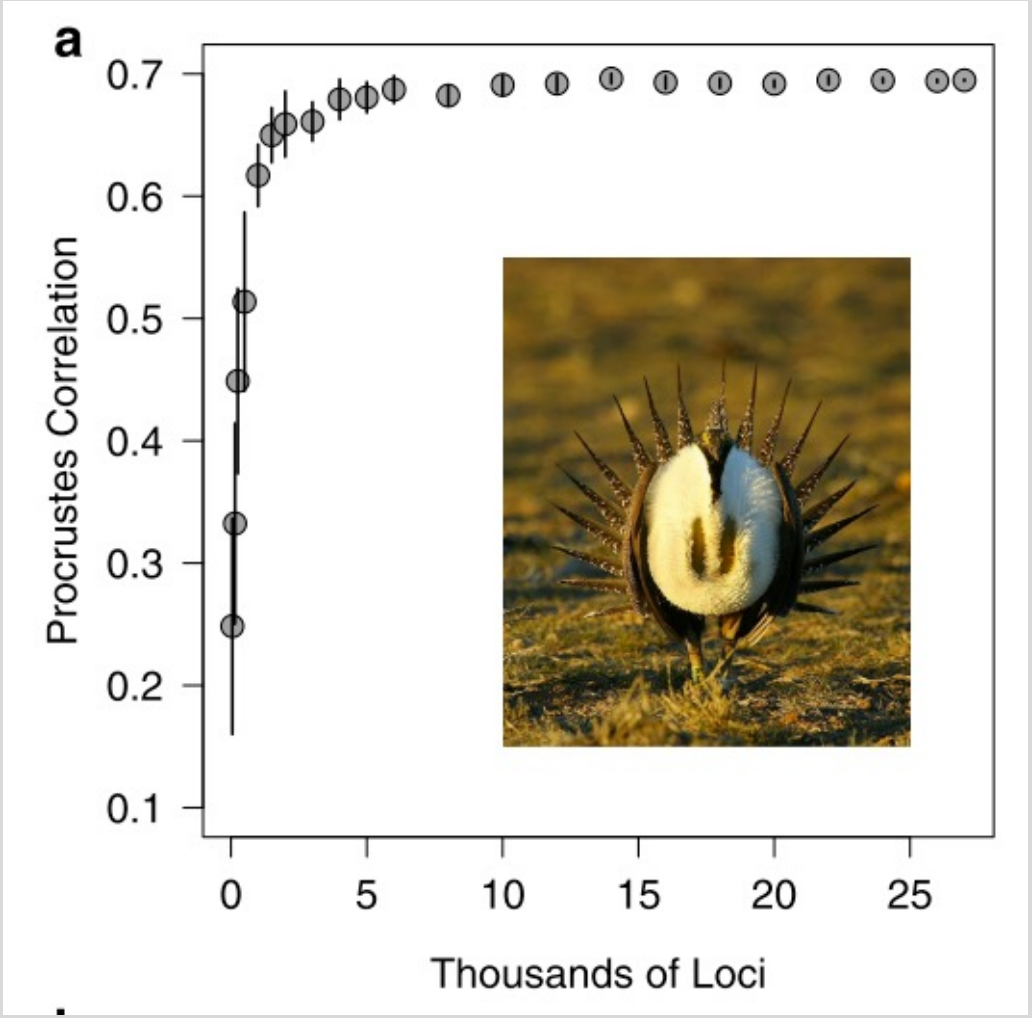




Lastly, from the wisdom of Josh Jahner… Add a photo of your study organism!!!

If you study sage-grouse and you don't put a picture, what is even the point of science!

Also, helps with **negative space**



Jahner et al. (2016) *should be Science*