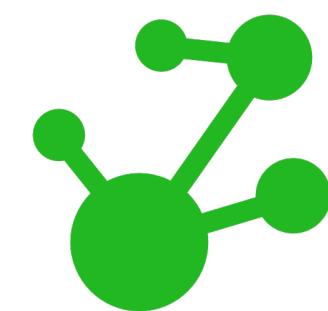
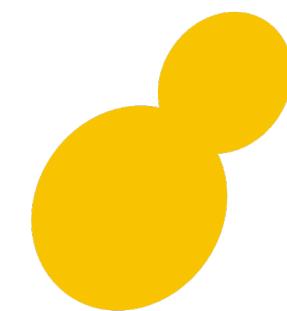
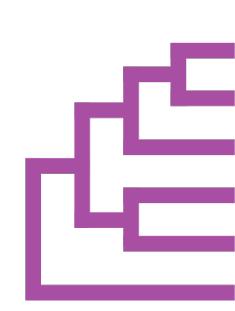
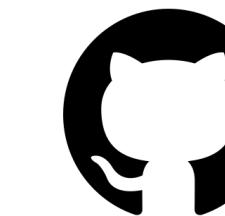


# Colorblind-friendly and Publication-Quality Data Visualization in R

Jacob L. Steenwyk



@JLSteenwyk



[jlsteenwyk.com](http://jlsteenwyk.com)

**1) Do you have a computer?**

**2) Did you download R?**

**3) Install these packages**

```
> install.packages(c("ggplot2", "ggfortify",
"reshape2", "ggExtra","RColorBrewer") ,dep=T)
```

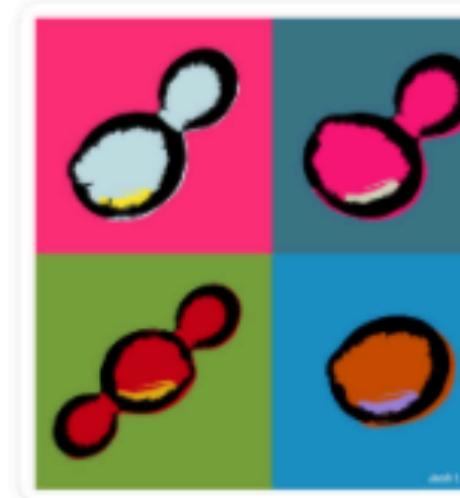
# *Art for Earth*



Purchase with purpose: 100% of profits go toward global conservation efforts

---

Shop by product type,  
conservation status, or  
buy a sticker of [Sciart](#)  
logo!



[Vinyl Stickers](#)



[Poster Prints](#)



[Camper Mugs](#)



[Endangered Animals](#)



[No current concern](#)

Have a question? Check out the [Frequently Asked Questions \(FAQ\)](#) section or get in touch via [twitter](#)!



@JLSteenwyk

<https://www.etsy.com/shop/JLSteenwyk>

# *Art for Earth*

## Using art to raise awareness and immortalize endangered species



African wild dog (*Lycaon pictus*)

- Status: Endangered
- Population: 1,409

Blue whale (*Balaenoptera musculus*)

- Status: Endangered
- Population: 10,000 - 25,000

Galápagos penguin (*Spheniscus mendiculus*)

- Status: Endangered
- Population: fewer than 2,000



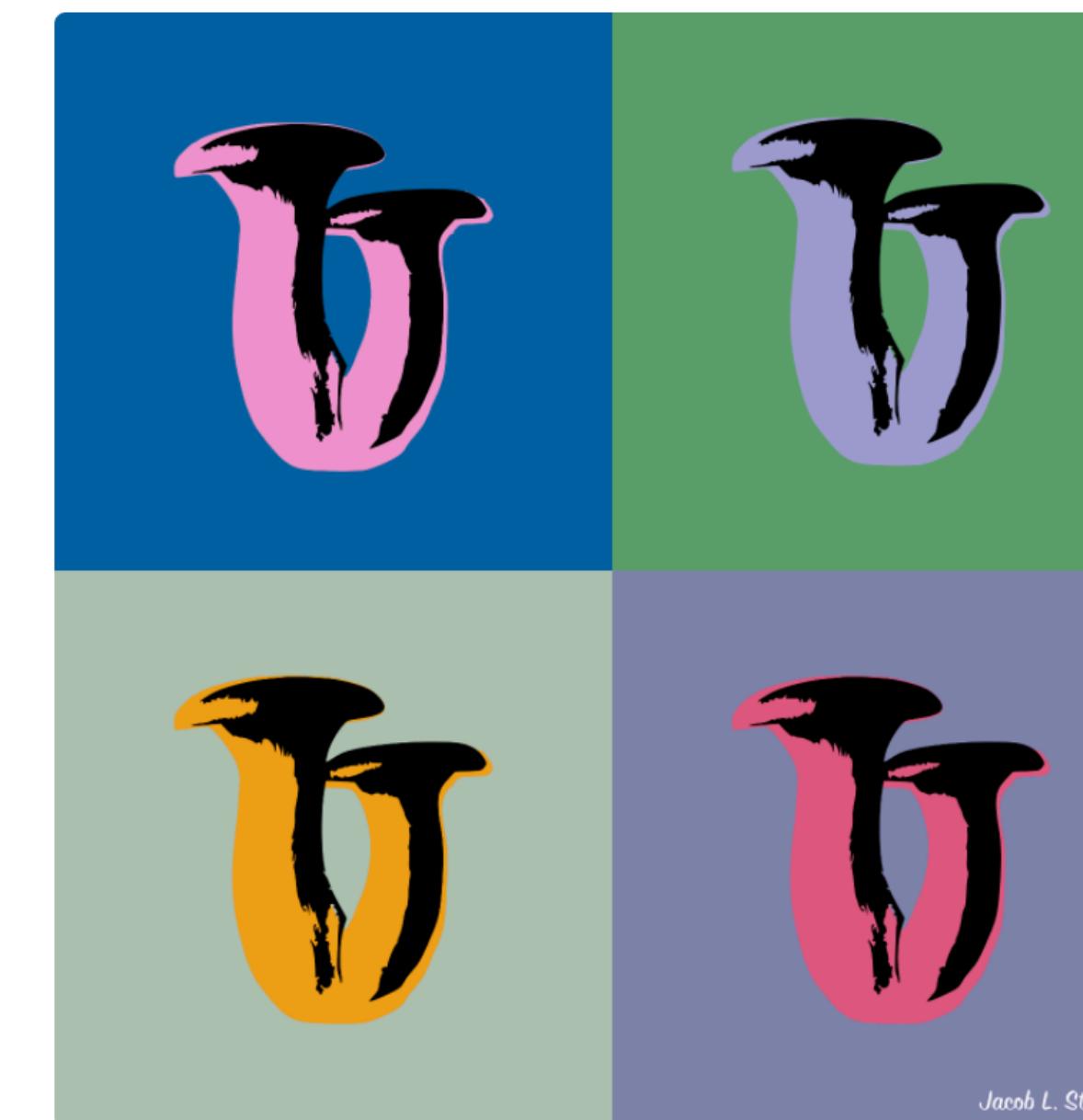
@JLSteenwyk

# Art for Earth

Using art to raise awareness and immortalize  
endangered species or species I love



Jacob L. Steenwyk



Jacob L. Steenwyk



Jacob L. Steenwyk

## Fly agaric (*Amanita muscaria*)

- \$8.99 vinyl sticker (FREE shipping)
- \$22.99+ poster print (FREE shipping)
- \$28.99 camper mug (FREE shipping)

## Oyster mushroom (*Pleurotus ostreatus*)

- \$8.99 vinyl sticker (FREE shipping)
- \$22.99+ poster print (FREE shipping)
- \$28.99 camper mug (FREE shipping)

## Morel mushroom (*Morchella esculenta*)

- \$8.99 vinyl sticker (FREE shipping)
- \$22.99+ poster print (FREE shipping)
- \$28.99 camper mug (FREE shipping)



@JLSteenwyk

<https://www.etsy.com/shop/JLSteenwyk>

# Featured on *Yeast* magazine

Received: 29 July 2020 | Accepted: 20 August 2020  
DOI: 10.1002/yea.3518

**SPECIAL ISSUE ARTICLE**

**Yeast** WILEY

## A portrait of budding yeasts: A symbol of the arts, sciences and a whole greater than the sum of its parts

Jacob L. Steenwyk<sup>1,2</sup> 

<sup>1</sup>Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA  
<sup>2</sup>Early Career Leadership Program Communication and Outreach Subcommittee, Genetics Society of America, Rockville, MD, USA

**Correspondence:** Jacob L. Steenwyk, Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA.  
Email: jacob.steenwyk@vanderbilt.edu

**Funding Information:** Vanderbilt University; Howard Hughes Medical Institute

**KEY WORDS:** art, budding yeast, cell cycle, Merian, naturalist, non-conventional yeasts, sciart, science, STEAM, Warhol

### 1 | INTRODUCTION

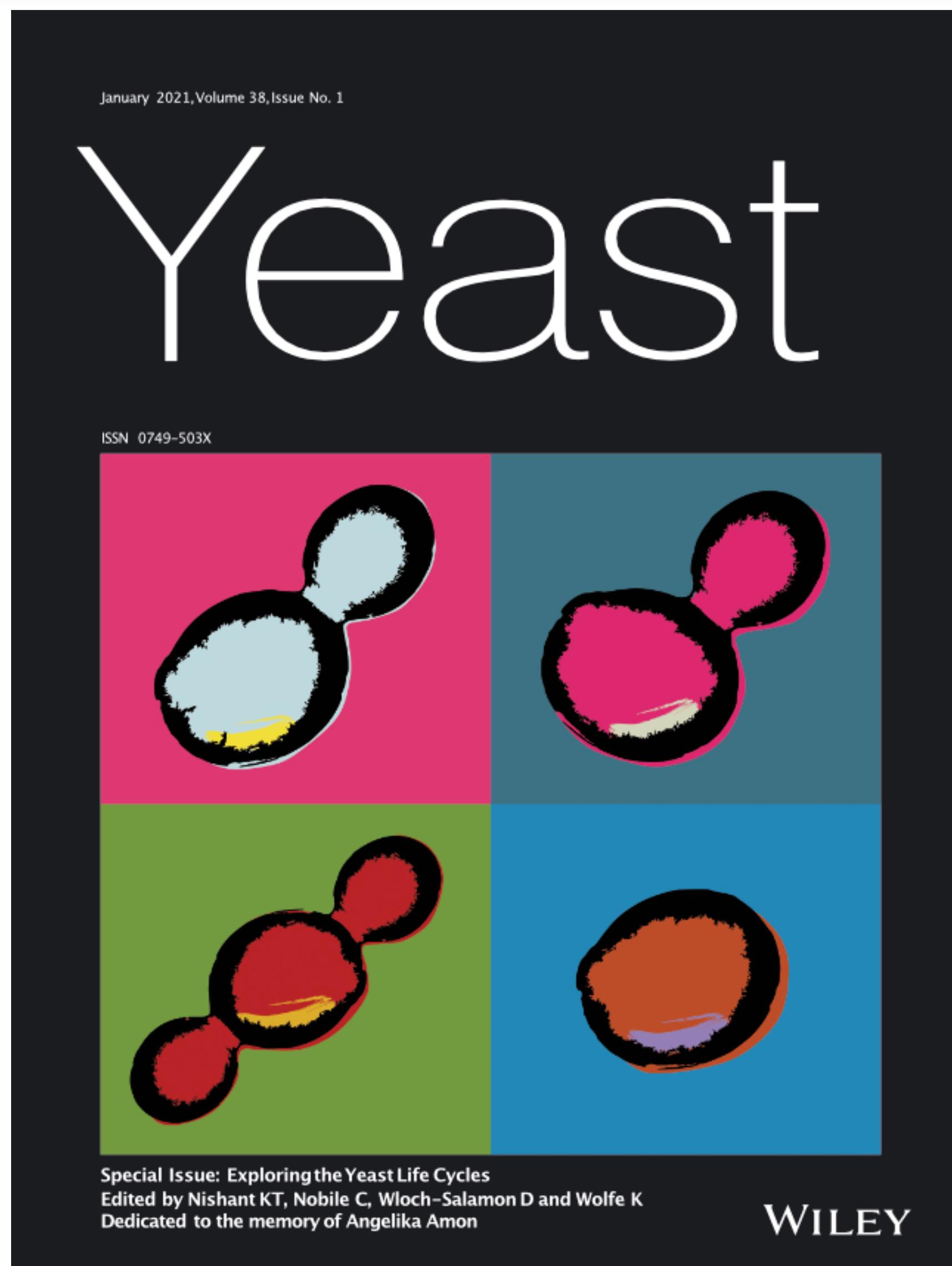
In the year 1660, 13-year-old Maria Sibylla Merian roamed the gardens and countryside of Germany taking detailed notes about caterpillars, moths, butterflies and their interactions with host plants, accompanying her notes were elaborate multimedia depictions of insect and plant life cycles (Figure 1). Merian's efforts in documenting interspecies relationships are regarded as early contributions to modern natural history and ecology, although the term 'ecology' was coined approximately two centuries later (Etheridge, 2011a, 2011b; Pieters & Winthagen, 1999). Her influence can be seen in the work of naturalists such as John James Audubon (Etheridge, 2015; Palmeri, 2017). Merian's success in part stems from her ability to use art to bolster her science and vice versa.

Merian is one of many scientists and artists who blended the arts and sciences over the centuries. In fact, scientist-artist polymaths like Aristotle and Leonardo da Vinci were more commonplace in part because of the common goal science and art share: interpreting and representing the natural world. The 'great divide' of the arts and sciences in Western cultures is thought to have started in the 19th century, coinciding with the term 'scientist' being coined (Braund & Reiss, 2019; Sumner, 1959; Zhu & Goyal, 2019). The division became reinforced. Schools for arts and sciences were separated as unfounded claims about brain differentiation formulated (Zhu & Goyal, 2019). For example, the right and left brain hemispheres were thought to be individually responsible for arts and science learning, respectively (Sperry, 1968). However, evidence from cognitive scientists favours a holistic view of the brain wherein a wide range of stimulation (e.g., arts and sciences) improves broad brain function and critical thinking skills (Braund & Reiss, 2019; Howes, Kaneva, Swanson, & Williams, 2013).

Today, the benefits of a holistic view of the arts and the sciences have been recognized by numerous institutions. For example, Science, Technology, Engineering, Arts and Mathematics (STEAM) inspired curriculum is used to help students build skills for broad problem solving in K-12 schools (Kim & Park, 2012; Peppler, 2013; Sochacka, Guyotte, & Walther, 2016). In higher education, artists, designers, researchers and inventors have formed forward-thinking coalitions such as the Center for Art, Science & Technology at Massachusetts Institute of Technology (<https://arts.mit.edu/cast/>) and ArtLab at Vanderbilt University (<https://artlabvanderbilt.com/>) to reunite the arts and sciences. These initiatives and many others have used the arts as an effective form of communication between scientists and the broader community (Ilingworth, 2017), ultimately helping disseminate major scientific findings across society.

Perhaps one of the most important and recent scientific findings in the field of biological sciences is our understanding of the cellular life cycle. Seminal discoveries that unraveled the controls of the life cycle were made studying the model unipolar budding yeast *Saccharomyces cerevisiae* (Hartwell, Culotti, Pringle, & Reid, 1974). Comparative studies of *S. cerevisiae*, the fission yeast (*Schizosaccharomyces pombe*) and animals revealed striking similarities suggesting the life cycle is evolutionarily stable (Breedon & Nasmyth, 1987). Exploiting these similarities has enabled yeasts to be powerful models for cancer biology research and the development of anticancer therapeutics (Gao, Chen, & Huang, 2014; Guaragna et al., 2014; Schwartz & Dickson, 2009). However, examination of non-conventional yeasts and their life cycles can provide novel insights important to the fields of cell biology, evolutionary biology and more. For example, species of the budding yeast genus *Hanseniaspora* have lost numerous cell cycle control genes, including *MAD1*, *MAD2* and *RAD9*, and components of the Anaphase Promoting Complex and display atypical bipolar budding patterns (Steenwyk

54 | © 2020 John Wiley & Sons, Ltd. [wileyonlinelibrary.com/journal/yea](http://wileyonlinelibrary.com/journal/yea) [Yeast. 2021;38:54–56.](http://doi.org/10.1002/yea.3518)



# Fungi, the Dr. Jekyll and Mr. Hyde Kingdom



# Fungi, the Dr. Jekyll and Mr. Hyde Kingdom



- Estimated 2-5 million fungal species

# Fungi, the Dr. Jekyll and Mr. Hyde Kingdom



- Estimated 2-5 million fungal species
  - The good
  - Used in food production
  - Used for industrial fermentation processes, including production of vitamins & pigments
  - Outstanding model organisms

# Fungi, the Dr. Jekyll and Mr. Hyde Kingdom



- Estimated 2-5 million fungal species

## The good

- Used in food production
- Used for industrial fermentation processes, including production of vitamins & pigments
- Outstanding model organisms

## The bad

- ~8,000 plant pathogens; ~300 human pathogens
- Global warming is expanding the niche of fungal pathogens
- The rise of new multi-drug resistant pathogens like *Candida auris*



@JLSteenwyk

# Software engineering for the life sciences

The screenshot shows a GitHub profile for the user **JLSteenwyk**. The profile features a large circular profile picture with a pixelated, abstract design. Below the picture, the user's name is displayed as **Jacob L. Steenwyk** and their GitHub handle as **JLSteenwyk**. A brief bio indicates they are an Omics & Software Eng at HHMI, an awardee of the LSRF & Berkeley Science Fellow at UC-Berkeley, and a previous HHMI Gilliam Fellow at Vanderbilt.

The top navigation bar includes links for Overview, Repositories (42), Projects, Packages, Stars (16), and a search bar with a magnifying glass icon. To the right of the search bar are icons for creating a new repository, issues, pull requests, and notifications, along with a profile picture for JLSteenwyk.

The main content area is divided into sections: **Pinned** (containing five projects) and **Contribution activity**.

**Pinned Projects:**

- ClipKIT** (Public): a multiple sequence alignment-trimming algorithm for accurate phylogenomic inference. Python, 63 stars, 4 forks.
- PhyKIT** (Public): a UNIX shell toolkit for processing and analyzing multiple sequence alignments and phylogenies. Python, 59 stars, 7 forks.
- BioKIT** (Public): a versatile toolkit for processing and analyzing diverse types of sequence data. Python, 21 stars, 10 forks.
- orthosnap** (Public): a tree splitting and pruning algorithm for retrieving single-copy orthologs from gene family trees. Python, 23 stars, 1 fork.
- orthofisher** (Public): a broadly applicable tool for automated gene identification and retrieval. Python, 28 stars, 1 fork.
- ggpubfigs** (Public): colorblind friendly color palettes and ggplot2 graphic system extensions for publication-quality scientific figures. R, 105 stars, 9 forks.

**Contribution activity:** Shows a heatmap of contributions over the last year, with the most active period being March. The heatmap uses a color scale from light gray (Less) to dark green (More). The x-axis represents months from November to October, and the y-axis represents days of the week (Mon, Wed, Fri).

**Contribution settings** dropdown is set to **2024**.

**Contribution activity** section shows the date as **November 2024**.

**Followers:** 90 followers · 14 following.

**Contributors:** University of California, Berkeley, Berkeley, CA, jlsteenwyk@gmail.com, jlsteenwyk.com, @JLSteenwyk, https://genomely.substack.com.

**Twitter:**  @JLSteenwyk

# Two-part workshop

## Part 1

Seminar-style presentation  
~5 mins

## Part 2

Workshop tutorial  
~1.5 hours



@JLSteenwyk

# Two-part workshop

## Part 1

Seminar-style presentation  
~5 mins

## Part 2

Workshop tutorial  
~1.5 hours



@JLSteenwyk

# Two-part workshop

- Communication, things you've already known



@JLSteenwyk

# Two-part workshop

- Communication, things you've already known
  - Enable you to make Figures using R and ggplot2



@JLSteenwyk

# Two-part workshop

- Communication, things you've already known
  - Enable you to make Figures using R and ggplot2
- We will *not* cover everything



@JLSteenwyk

# Figures communicate observations



# Figures communicate observations



Pech Merle Cave Painting



@JLSteenwyk

# Figures communicate observations



Pech Merle Cave Painting

**Mammoth and a bison**



@JLSteenwyk

# Figures communicate observations

**horns, eyes, color, tail**



Pech Merle Cave Painting

**Mammoth and a bison**



Prehistoric Spanish Cave Painting

**Bison**



@JLSteenwyk

# Figures communicate observations

**horns, eyes, color, tail**



Pech Merle Cave Painting

**Mammoth and a bison**



Prehistoric Spanish Cave Painting

**Bison**



@JLSteenwyk

# Figures communicate observations

**horns, eyes, color, tail**



Pech Merle Cave Painting

**Mammoth and a bison**

**Less assumptions made by the viewer**



Prehistoric Spanish Cave Painting

**Bison**



@JLSteenwyk

# Figures communicate observations

**\*\*better communicates a collection of observations**

**horns, eyes, color, tail**



Pech Merle Cave Painting

**Mammoth and a bison**

**Less assumptions made by the viewer**



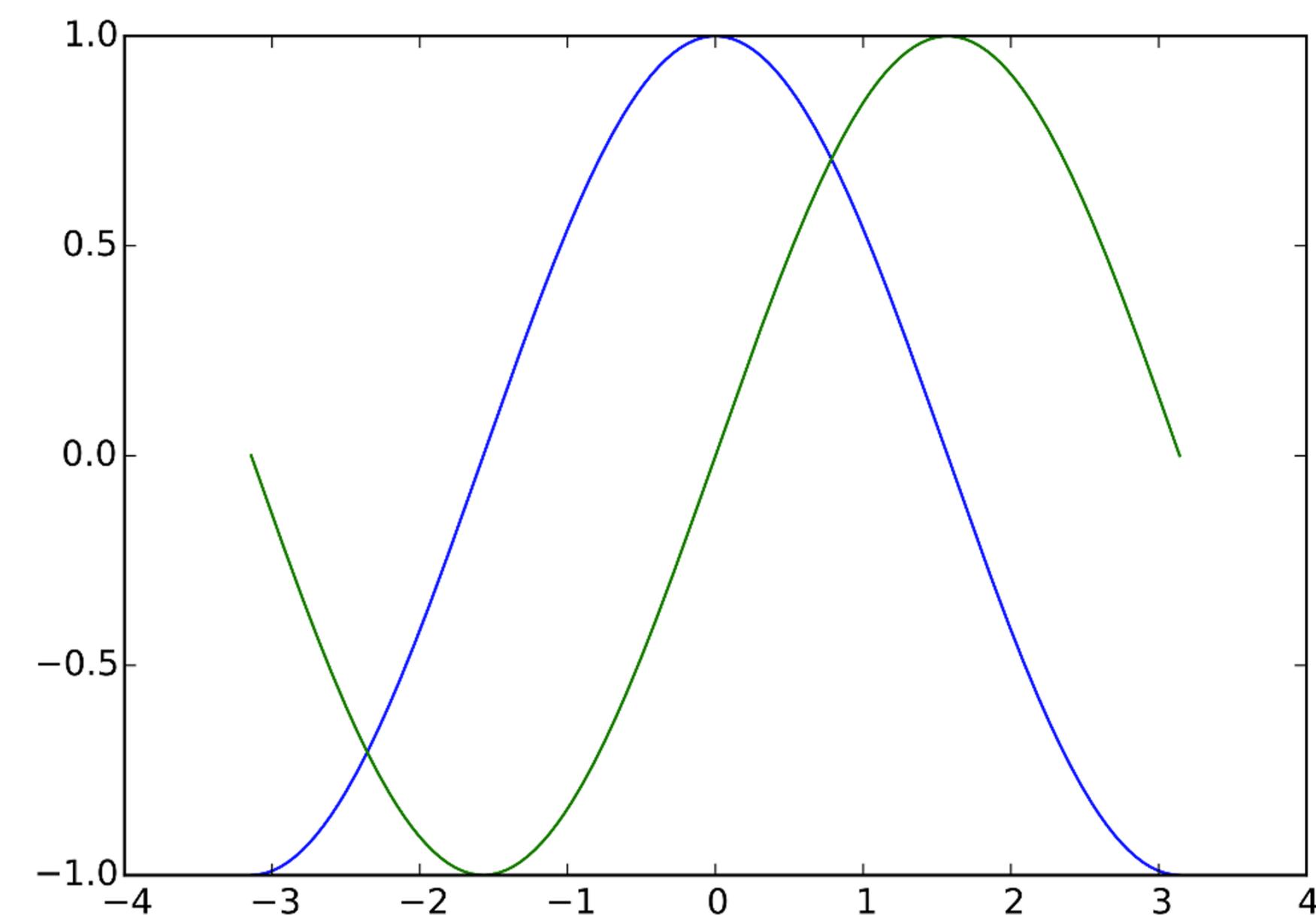
Prehistoric Spanish Cave Painting

**Bison**

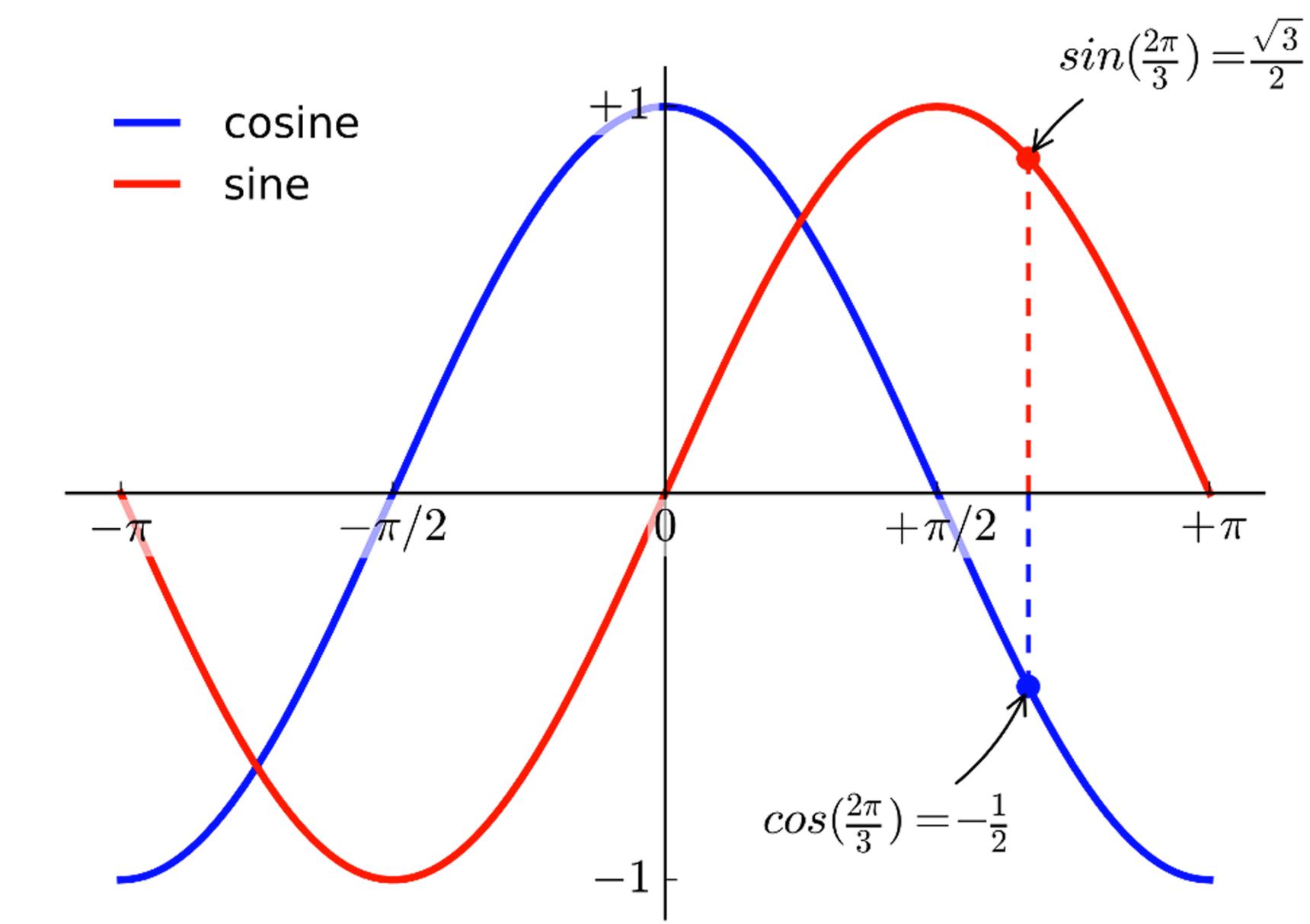
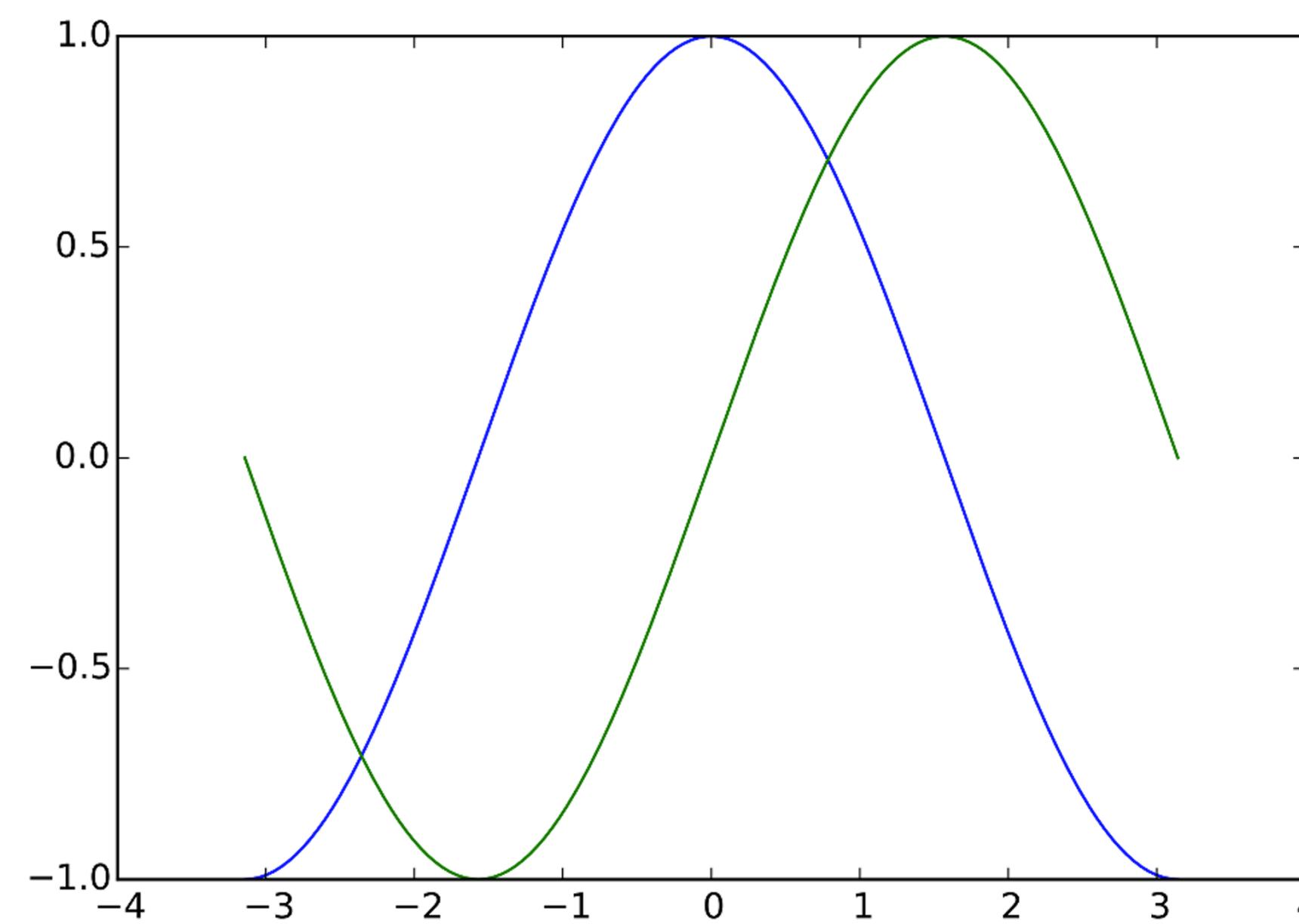


@JLSteenwyk

# A scientific example of the same phenomenon



# A scientific example of the same phenomenon



@JLSteenwyk

# Five principles to improve figures

- 1) Know your tools
- 2) Identify your message
- 3) Be honest about your data
- 4) Use color effectively
- 5) Message before beauty

# Five principles to improve figures

- 1) Know your tools
- 2) Identify your message
- 3) Be honest about your data
- 4) Use color effectively
- 5) Message before beauty

# Tools to plot data



@JLSteenwyk

# Tools to plot data

matplotlib



<http://matplotlib.org/gallery.html>



@JLSteenwyk

# Tools to plot data

**Seaborn**

**matplotlib**



<http://matplotlib.org/gallery.html>



@JLSteenwyk

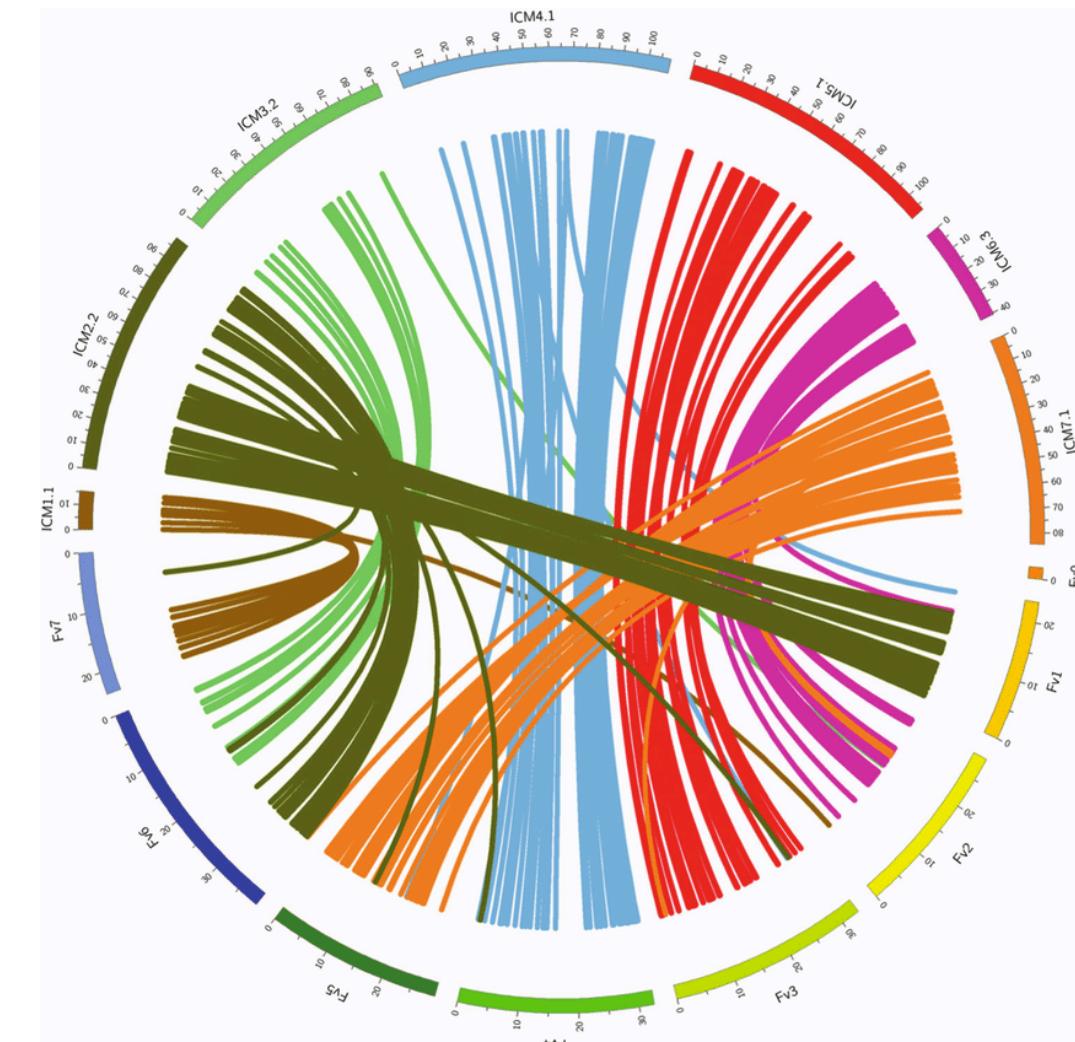
# Tools to plot data

**Seaborn**

**matplotlib**



<http://matplotlib.org/gallery.html>



**Circos**



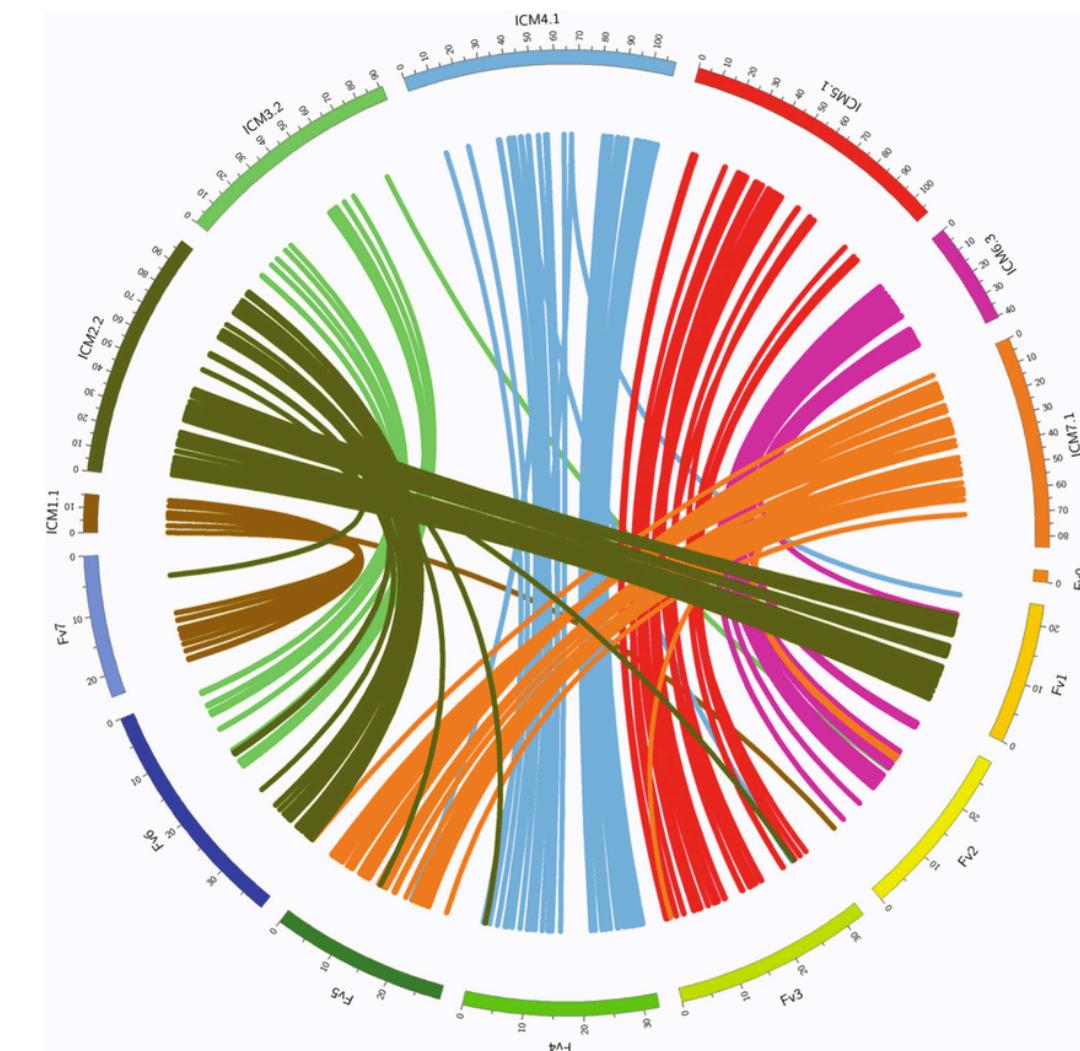
@JLSteenwyk

# Tools to plot data

**Seaborn**

**matplotlib**

<http://matplotlib.org/gallery.html>



**Circos**



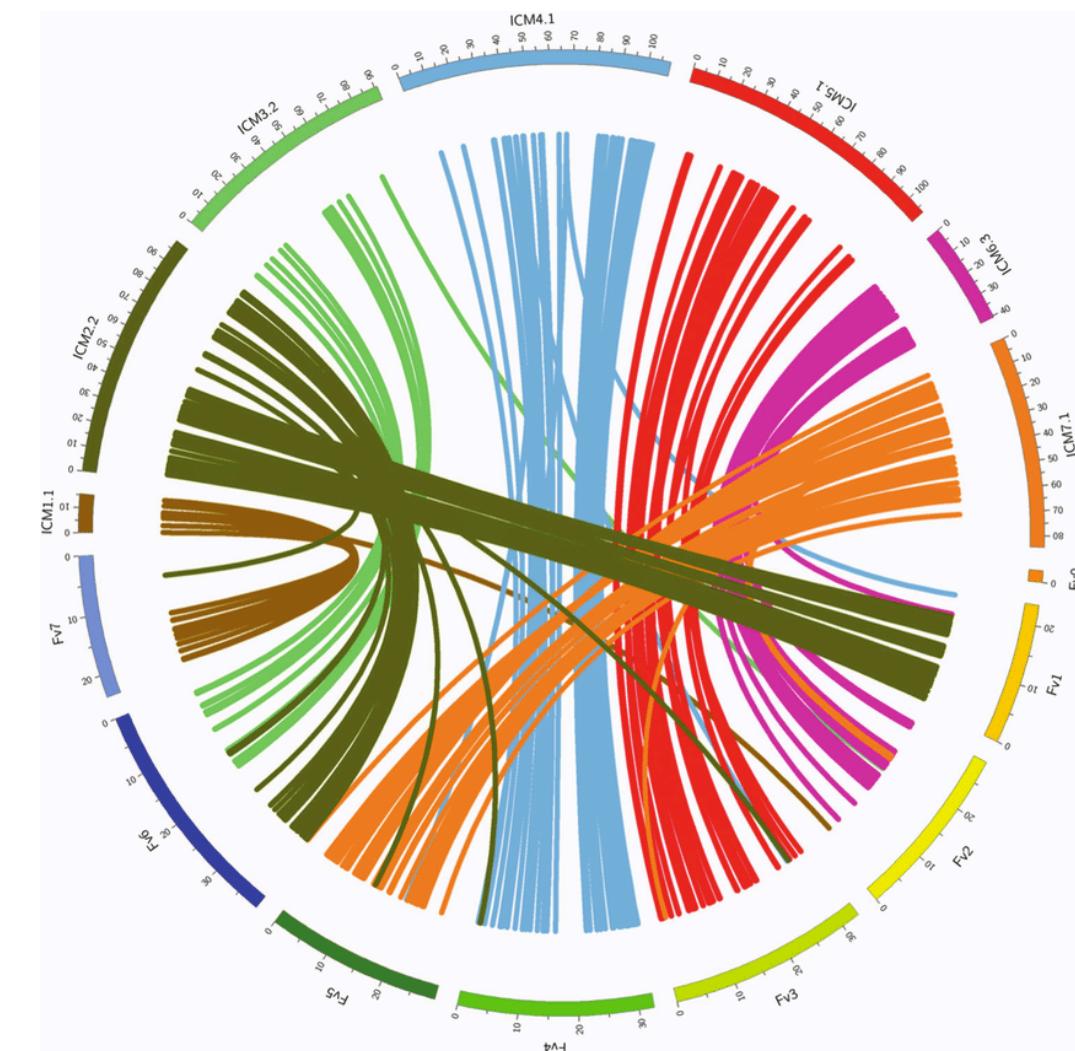
@JLSteenwyk

# Tools to plot data

**Seaborn**

**matplotlib**

<http://matplotlib.org/gallery.html>



**Circos**



@JLSteenwyk

# Tools to plot data

**Seaborn**

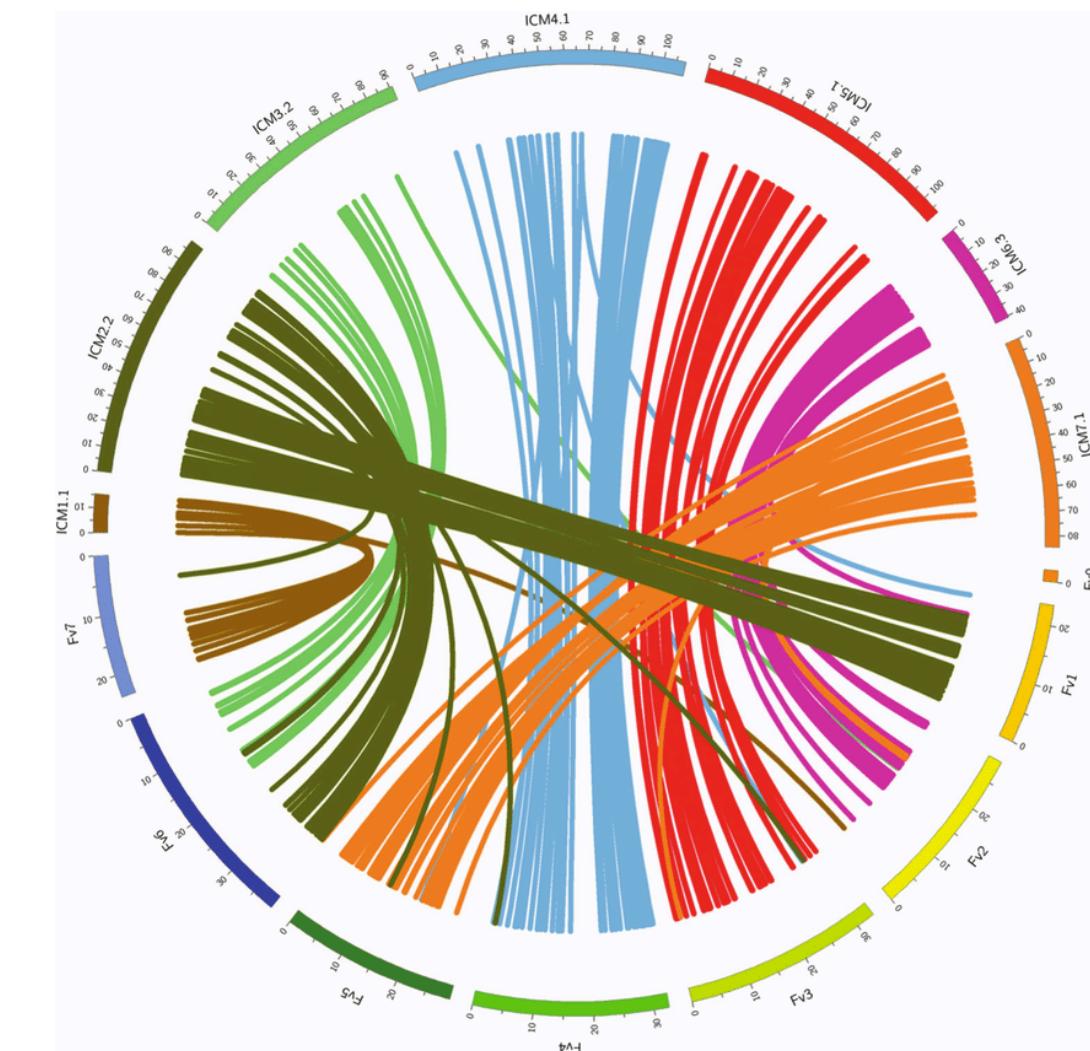
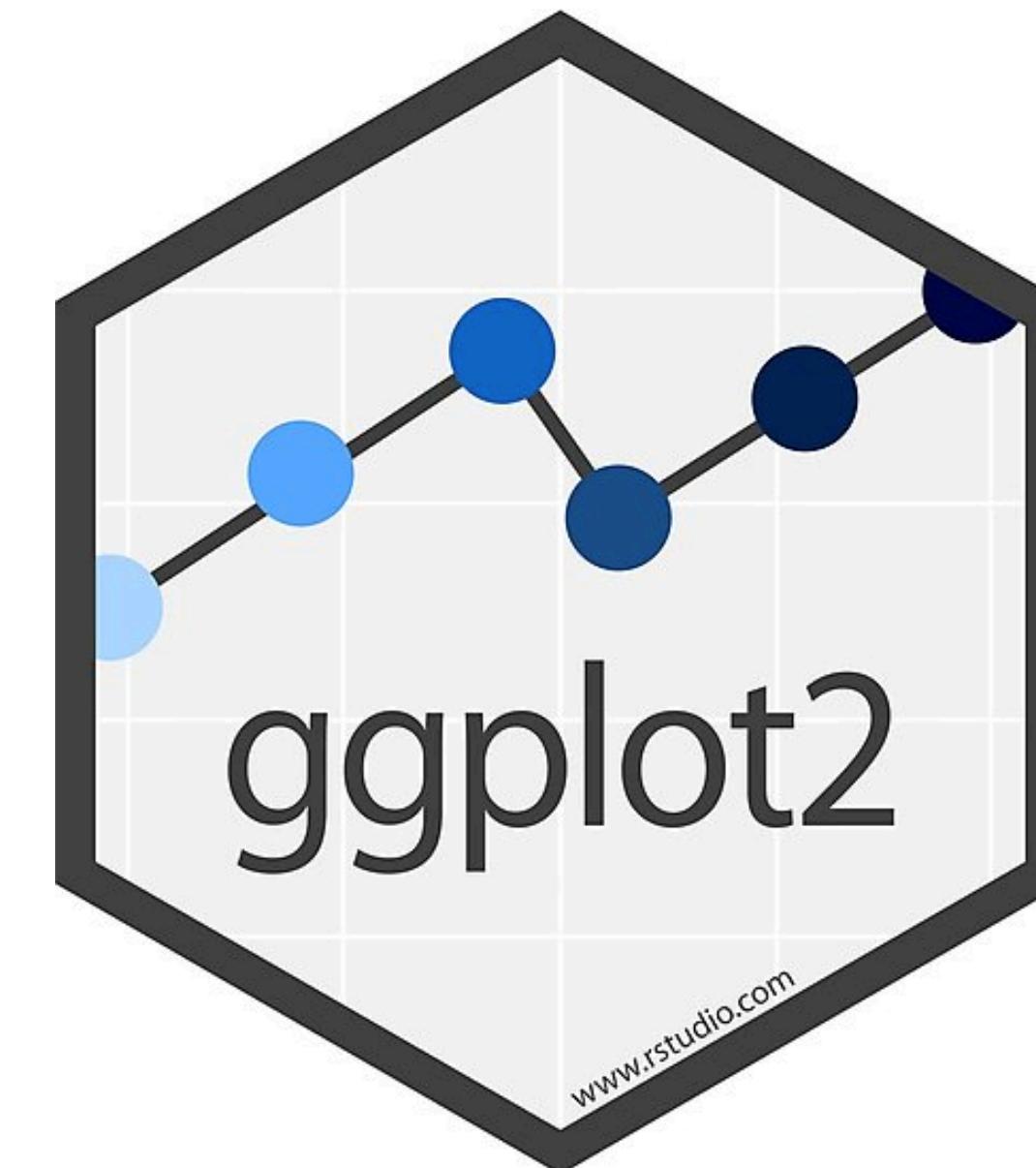
**matplotlib**



<http://matplotlib.org/gallery.html>



**Cytoscape**

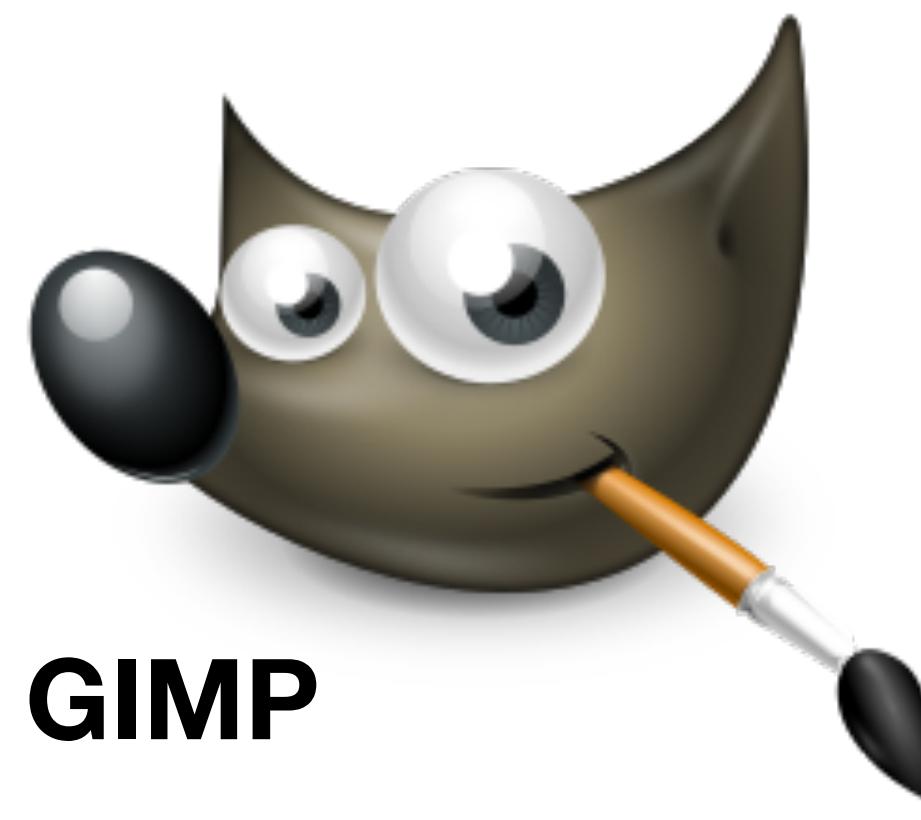


**Circos**

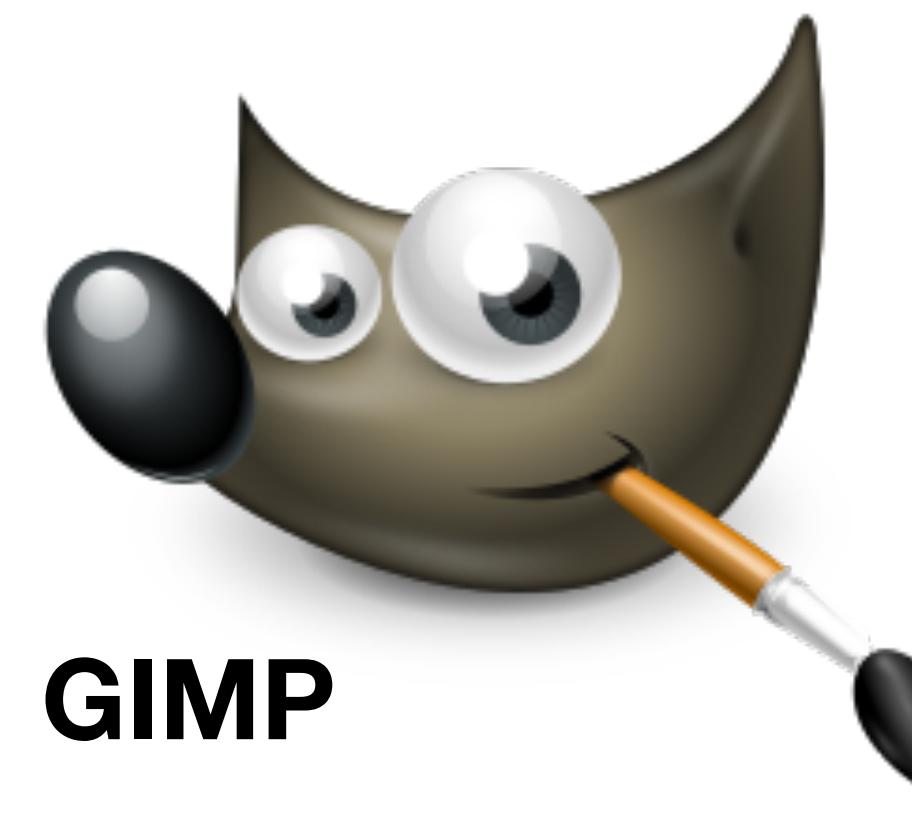


@JLSteenwyk

# Tools to edit graphics



# Tools to edit graphics

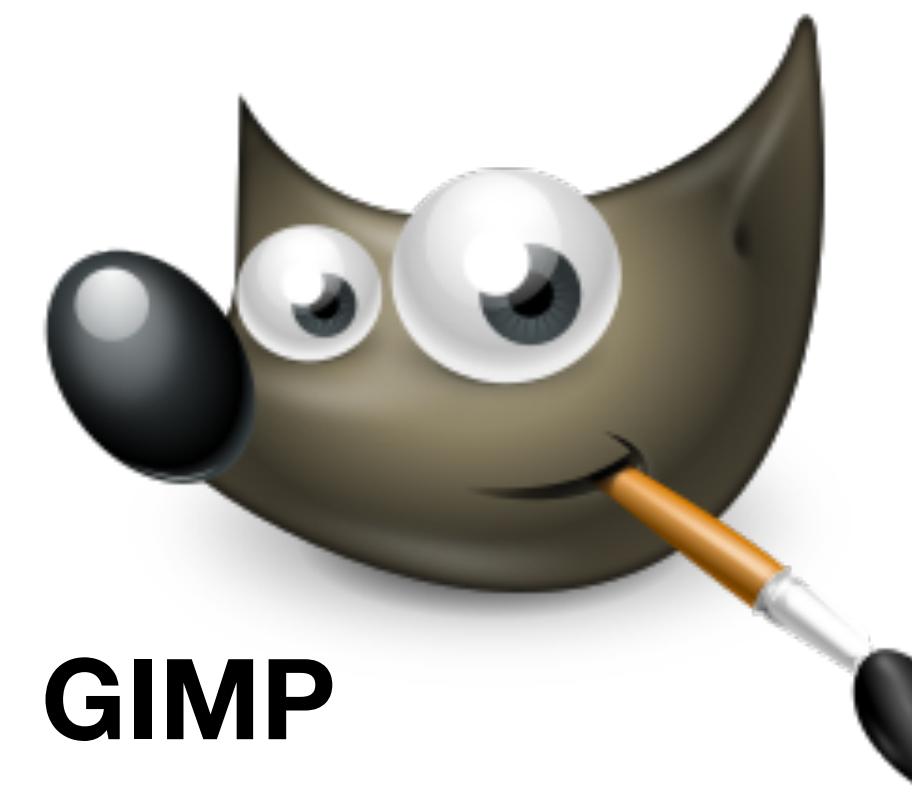


**GIMP**



**INKSCAPE**

# Tools to edit graphics



GIMP

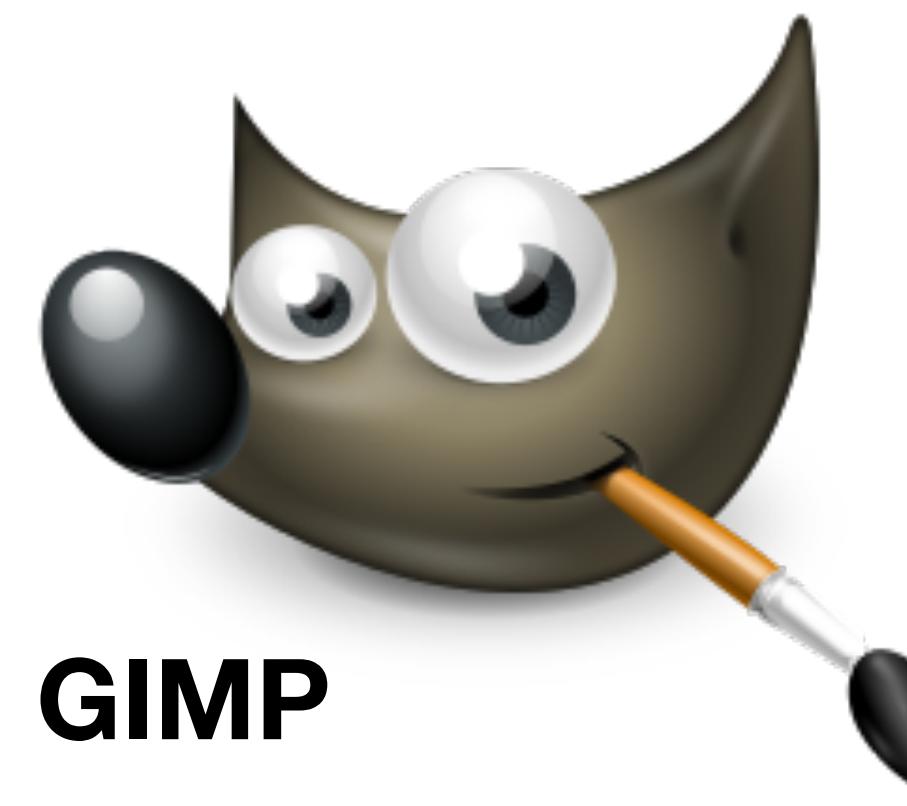


INKSCAPE



@JLSteenwyk

# Tools to edit graphics



GIMP



INKSCAPE



@JLSteenwyk

# How to choose?

**Seaborn**

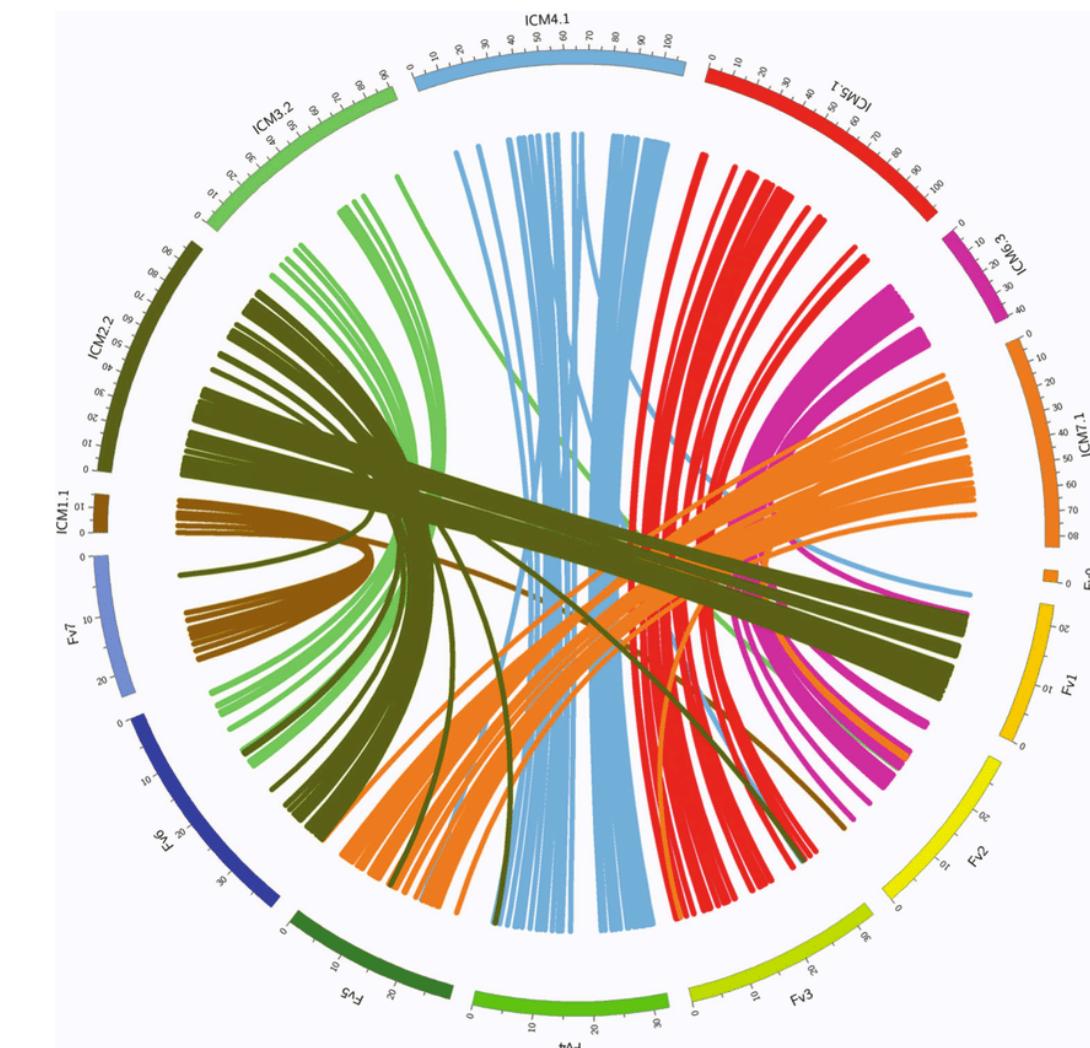
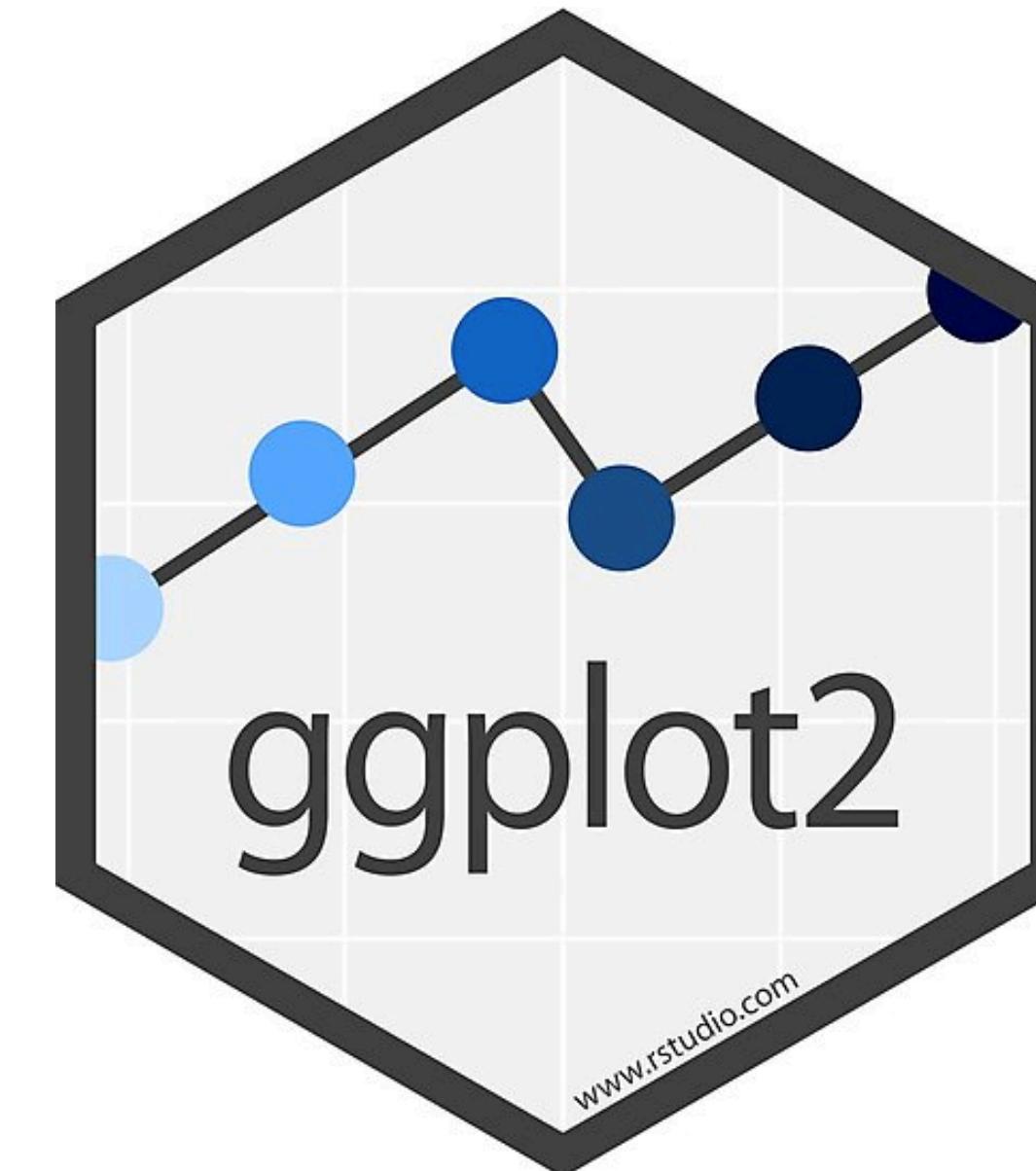
**matplotlib**



<http://matplotlib.org/gallery.html>



**Cytoscape**



**Circos**



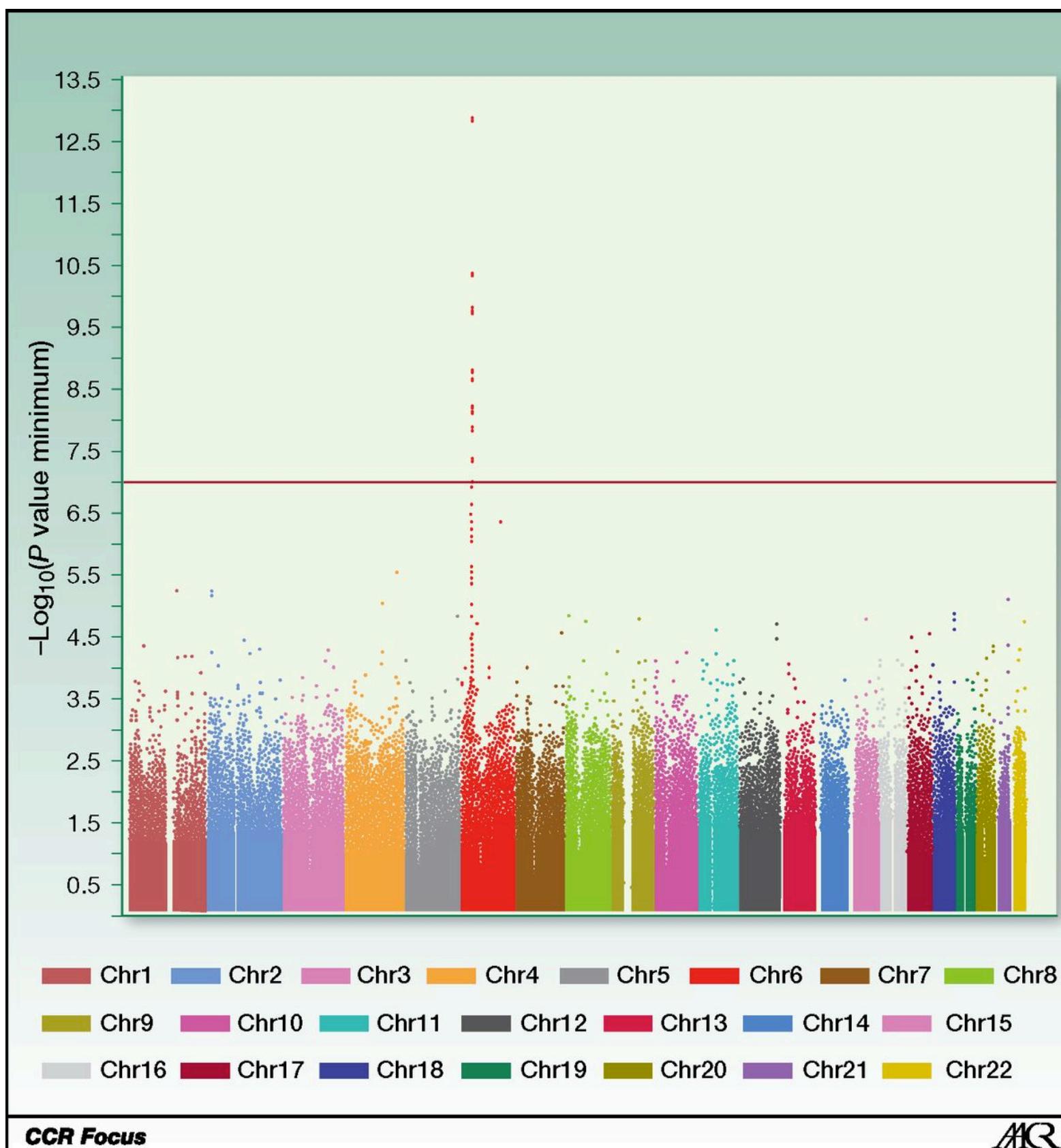
@JLSteenwyk

**Picking a software may be  
guided by the message  
you want to convey and  
the underlying data**

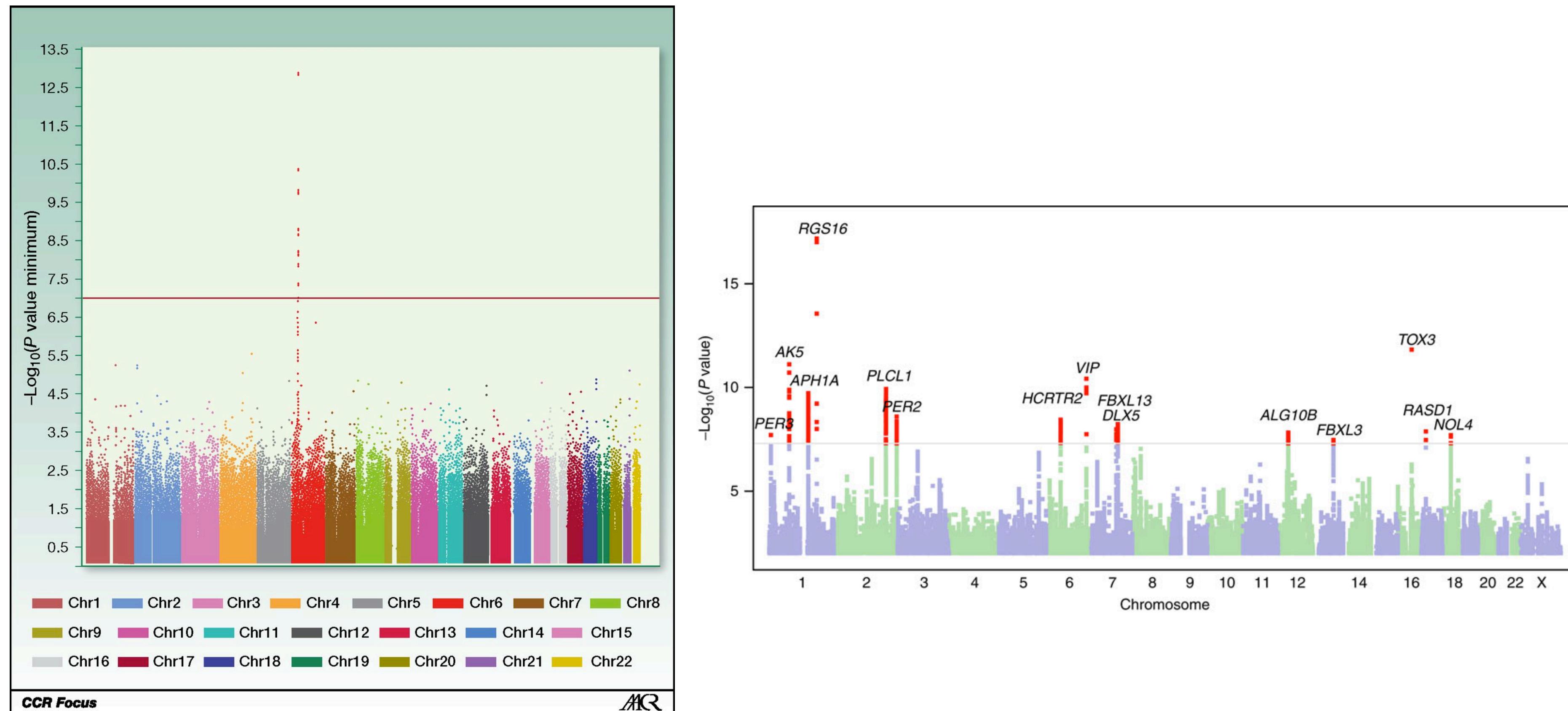
# Five principles to improve figures

- 1) Know your tools
- 2) Identify your message
- 3) Be honest about your data
- 4) Use color effectively
- 5) Message before beauty

# Manhattan plots communicate thousands of observations



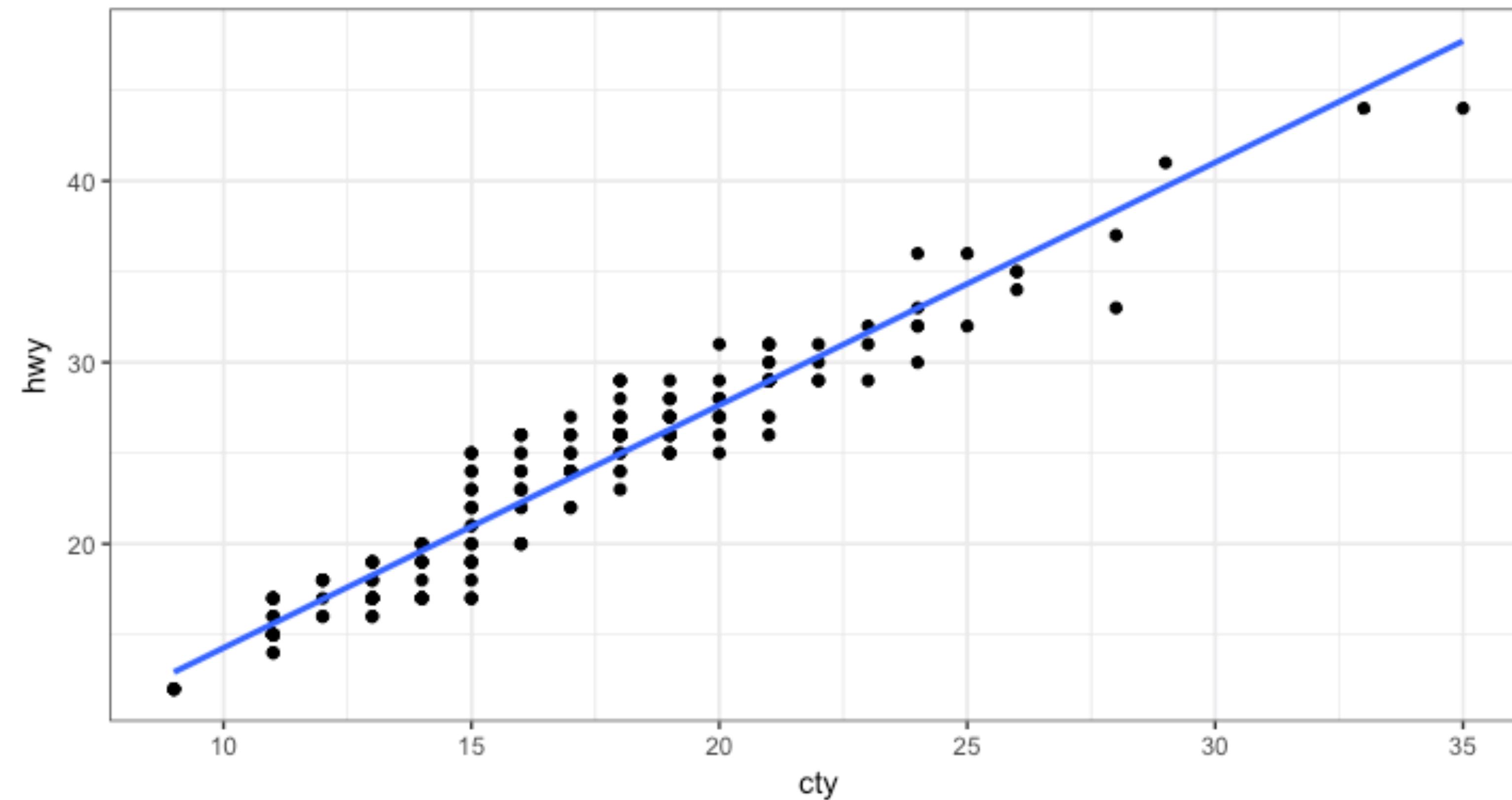
# Manhattan plots communicate thousands of observations



# Know your toolkit of plots – correlations

Scatterplot with overlapping points

mpg: city vs highway mileage

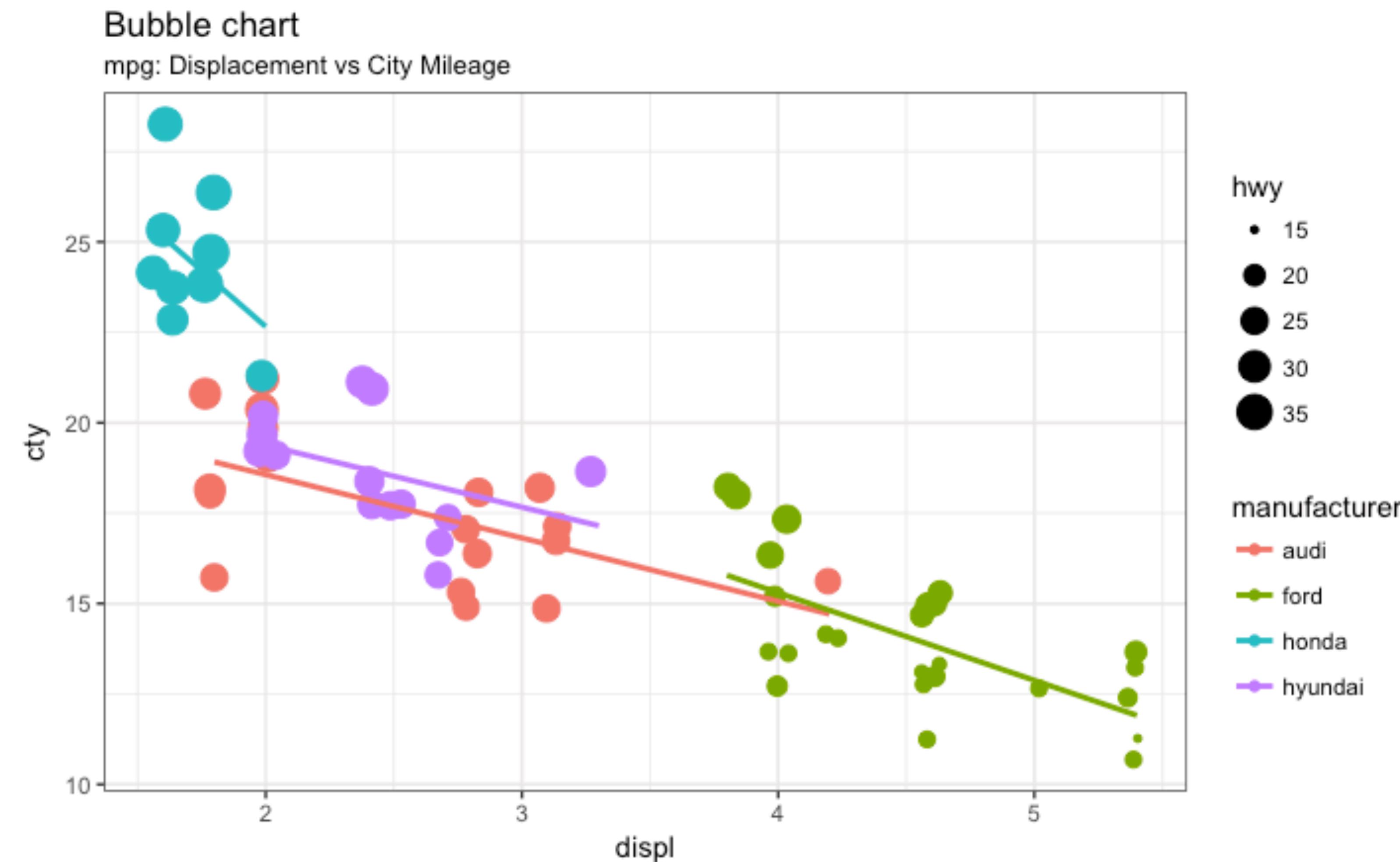


Source: midwest



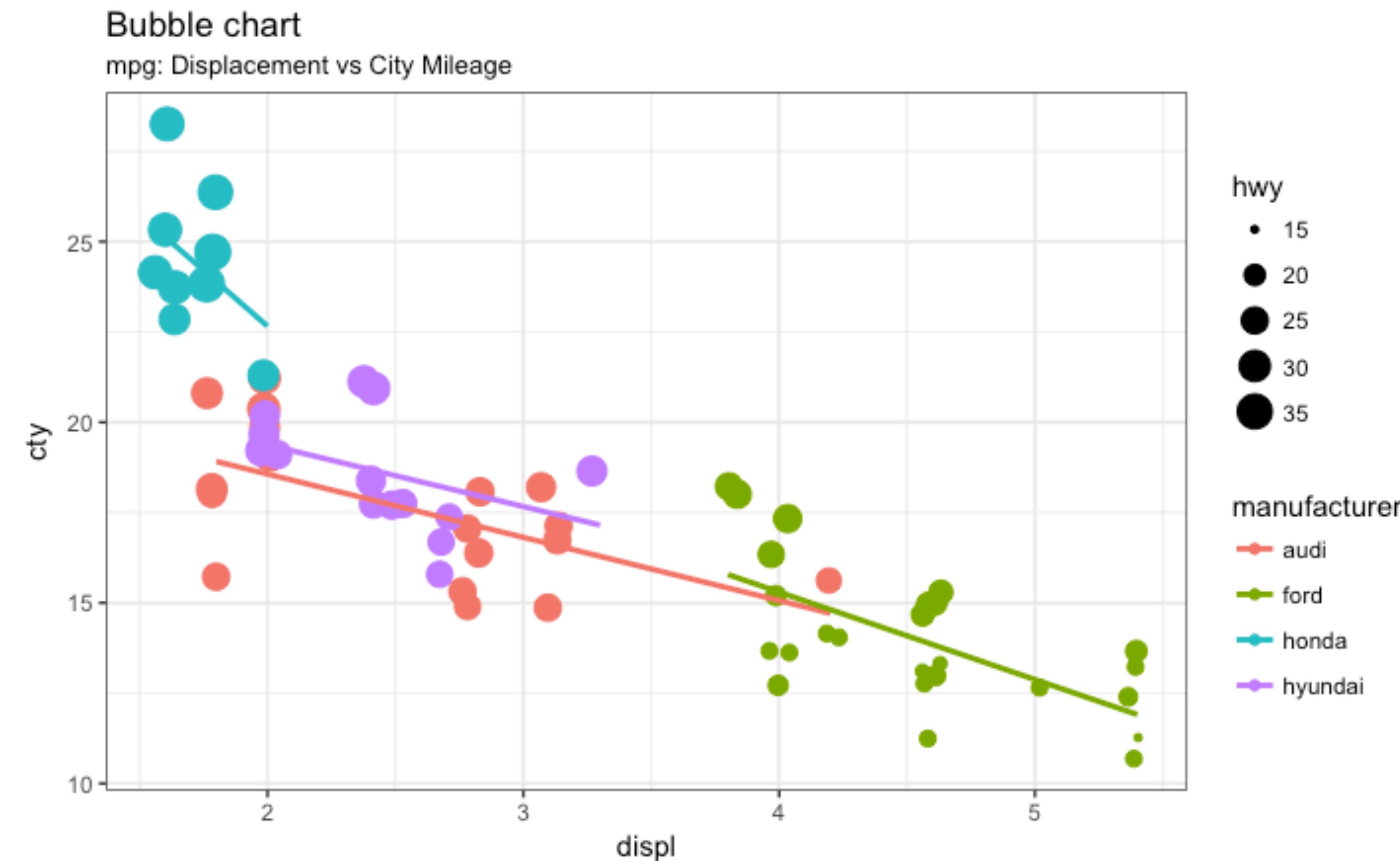
@JLSteenwyk

# Know your toolkit of plots – correlations



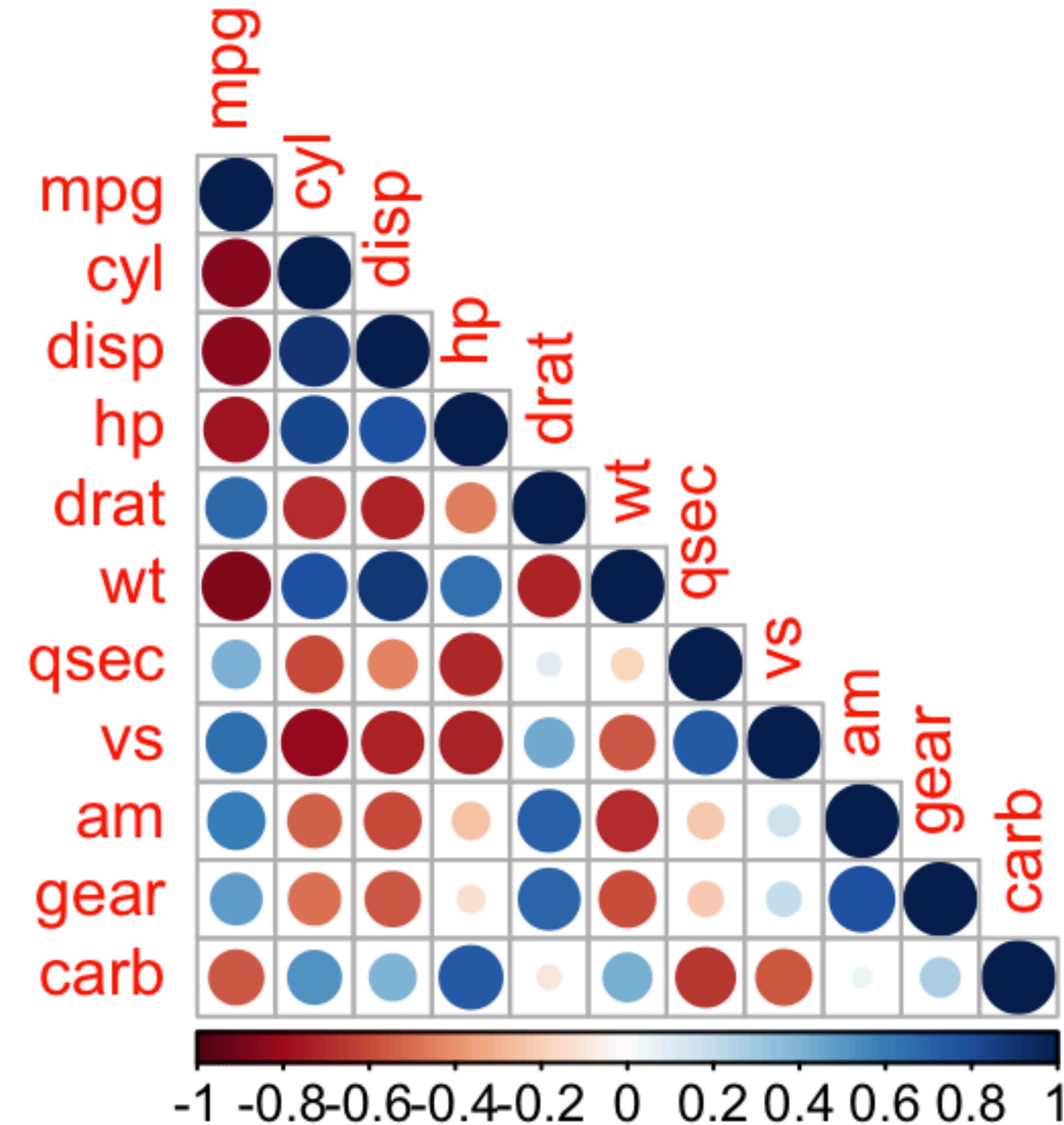
@JLSteenwyk

# Know your toolkit of plots – correlations

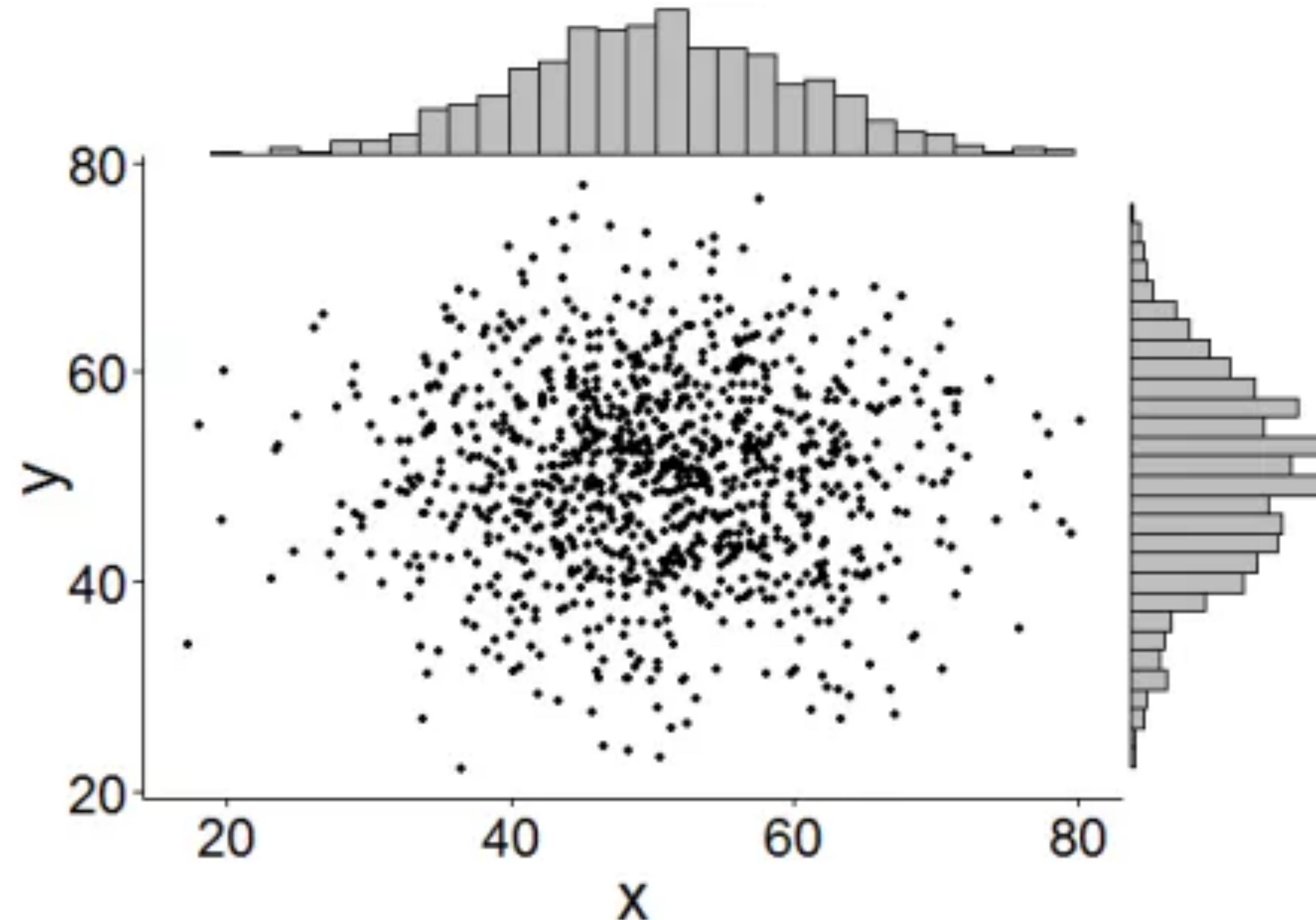


@JLSteenwyk

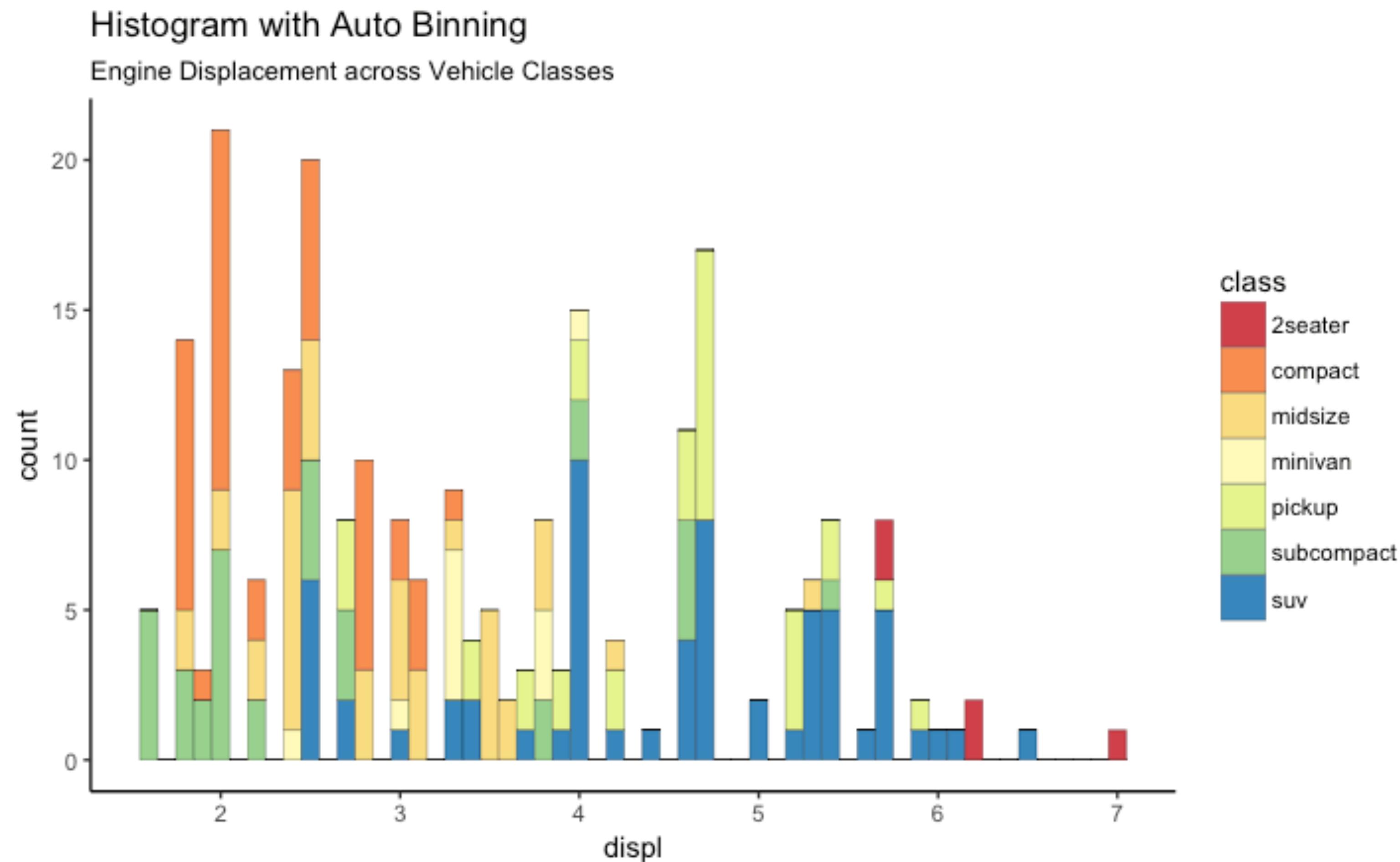
# Know your toolkit of plots – correlations



# Know your toolkit of plots – distributions

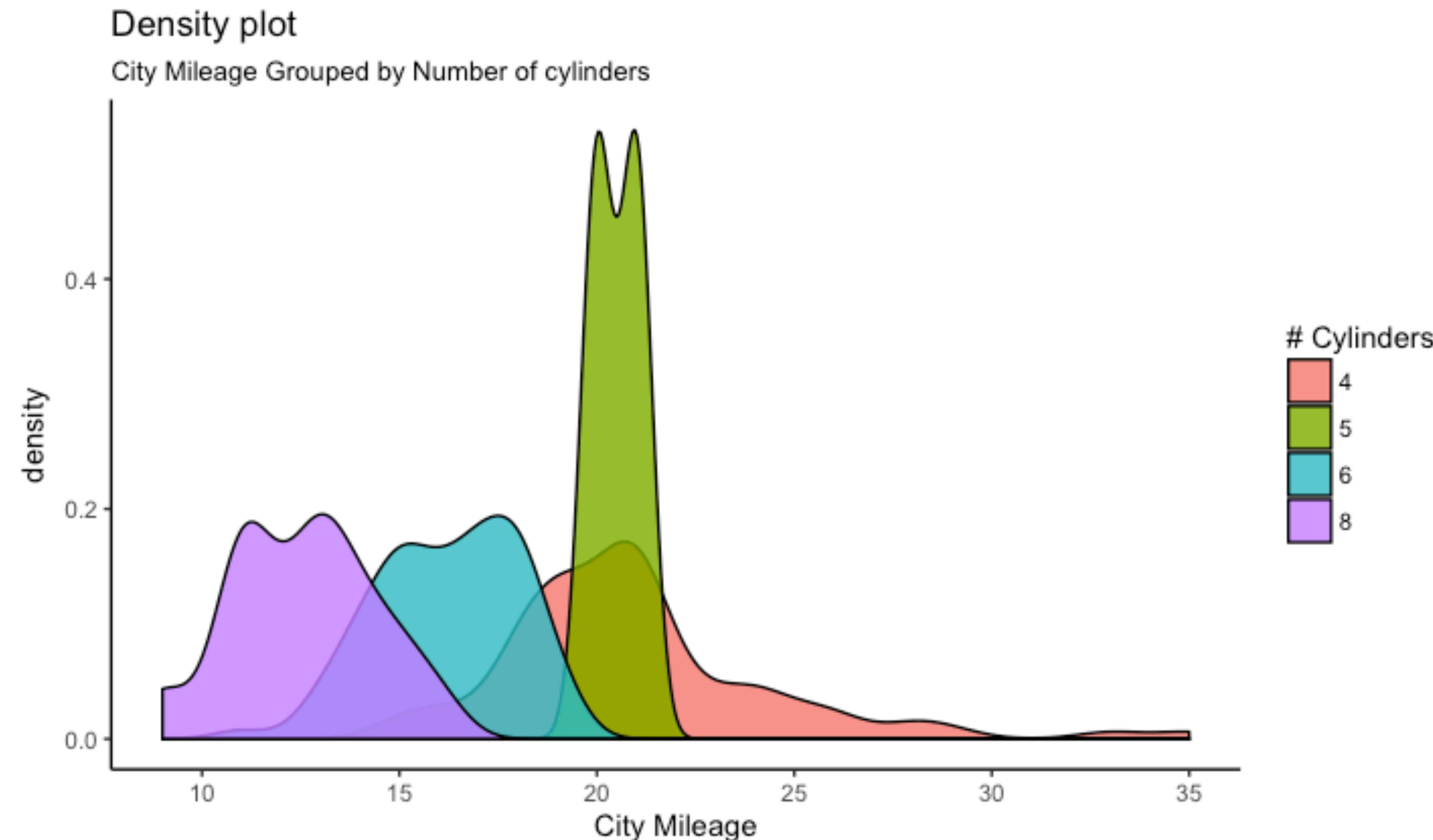


# Know your toolkit of plots – distributions



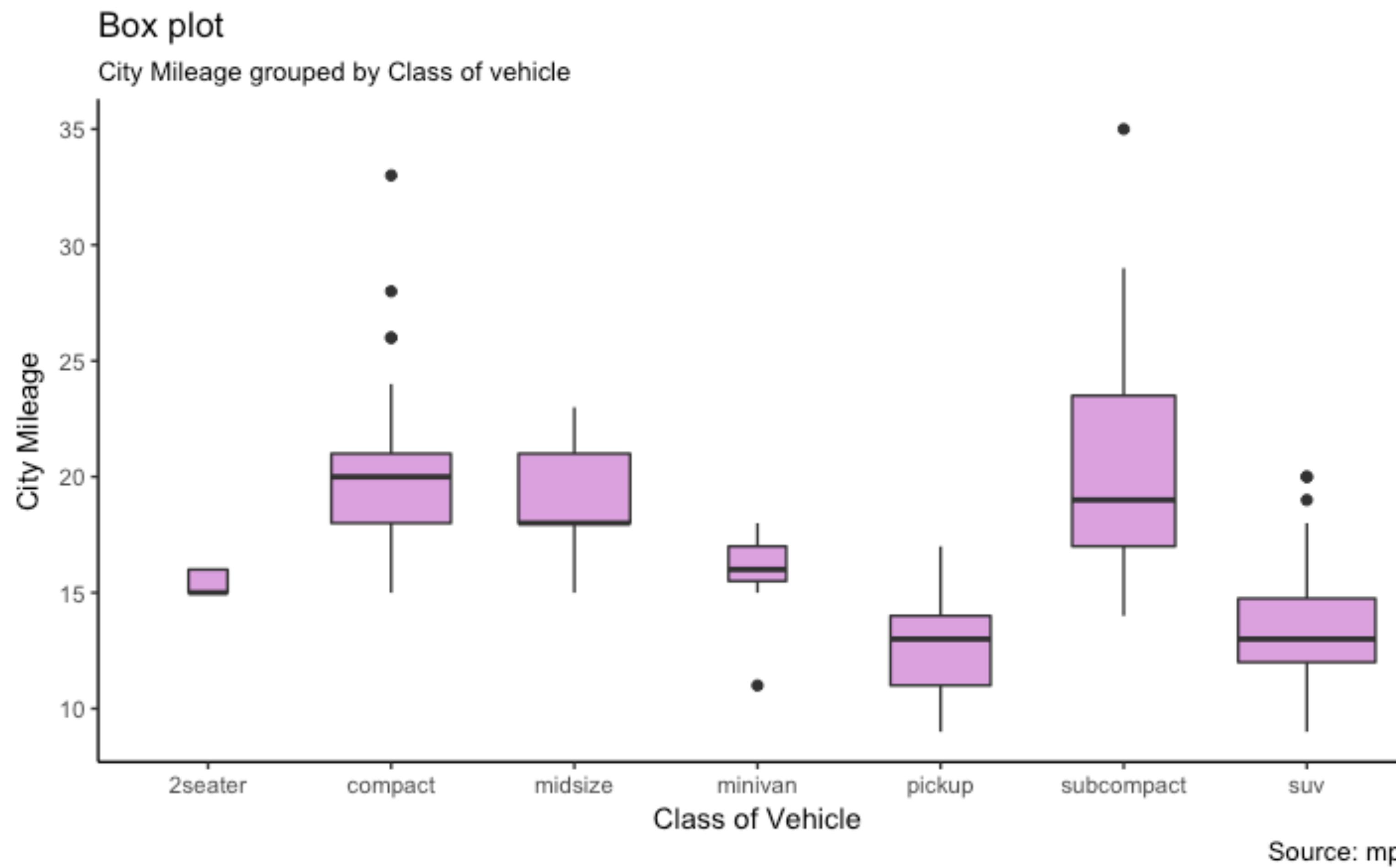
@JLSteenwyk

# Know your toolkit of plots – distributions

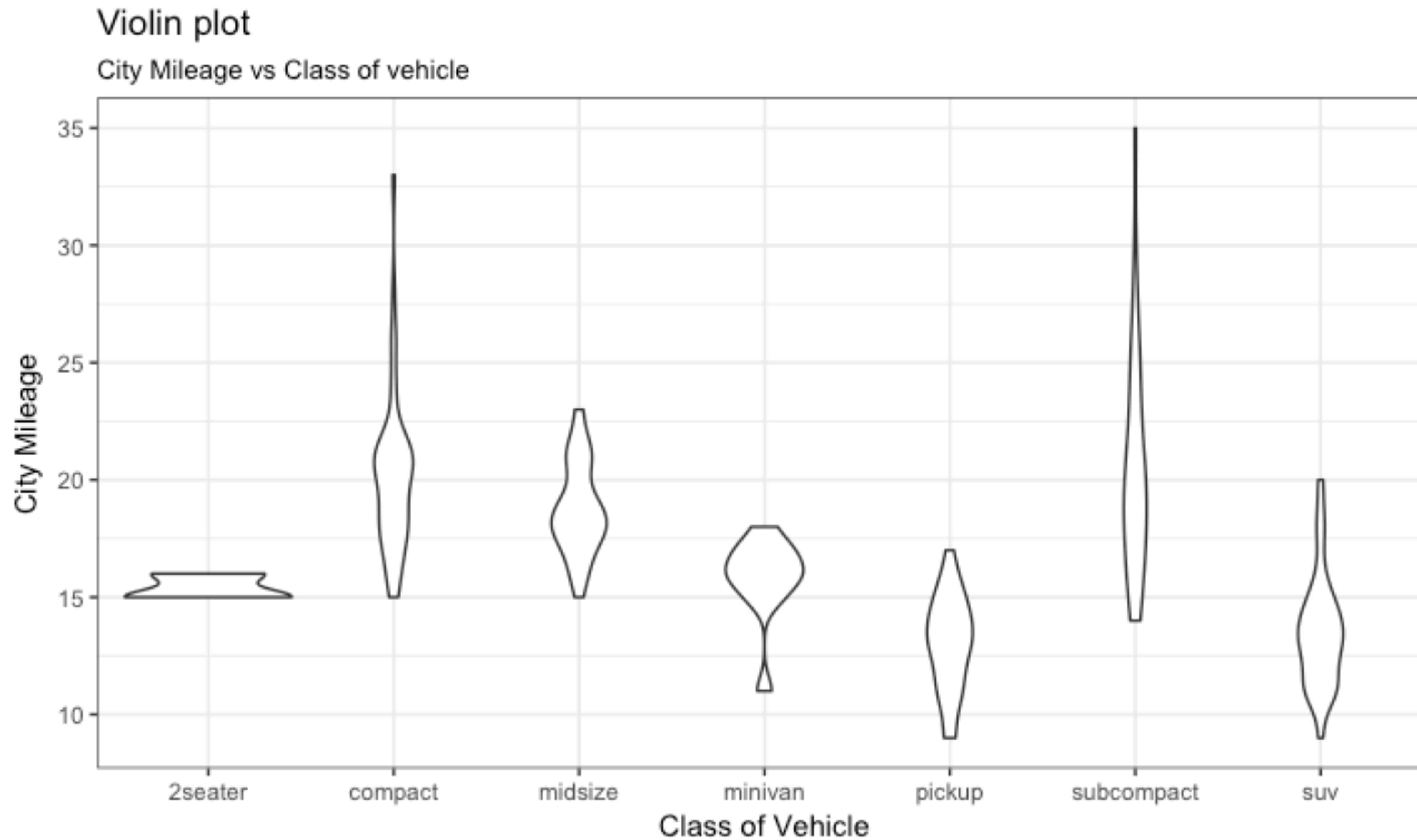


Source: mpg

# Know your toolkit of plots – distributions

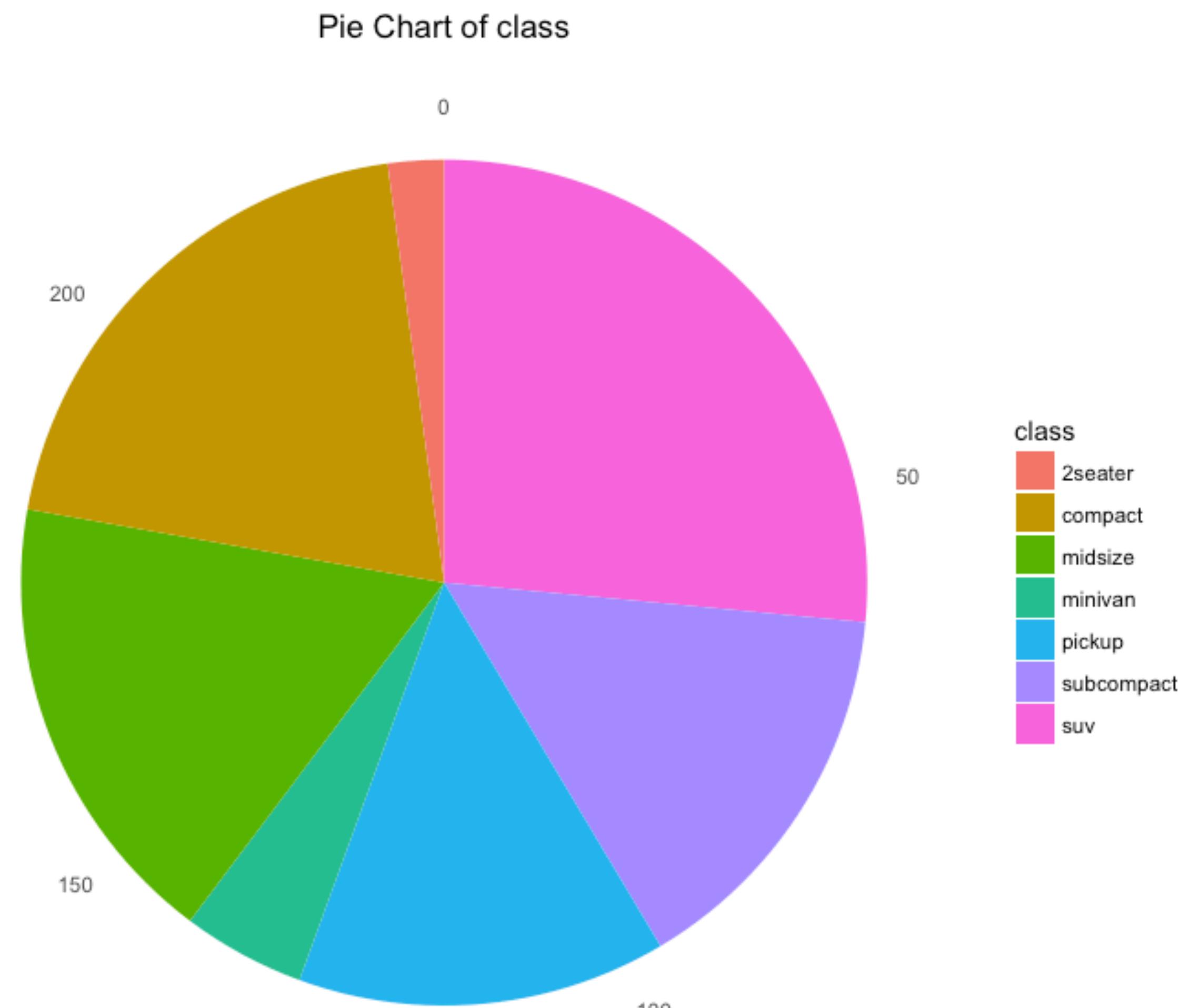


# Know your toolkit of plots – distributions



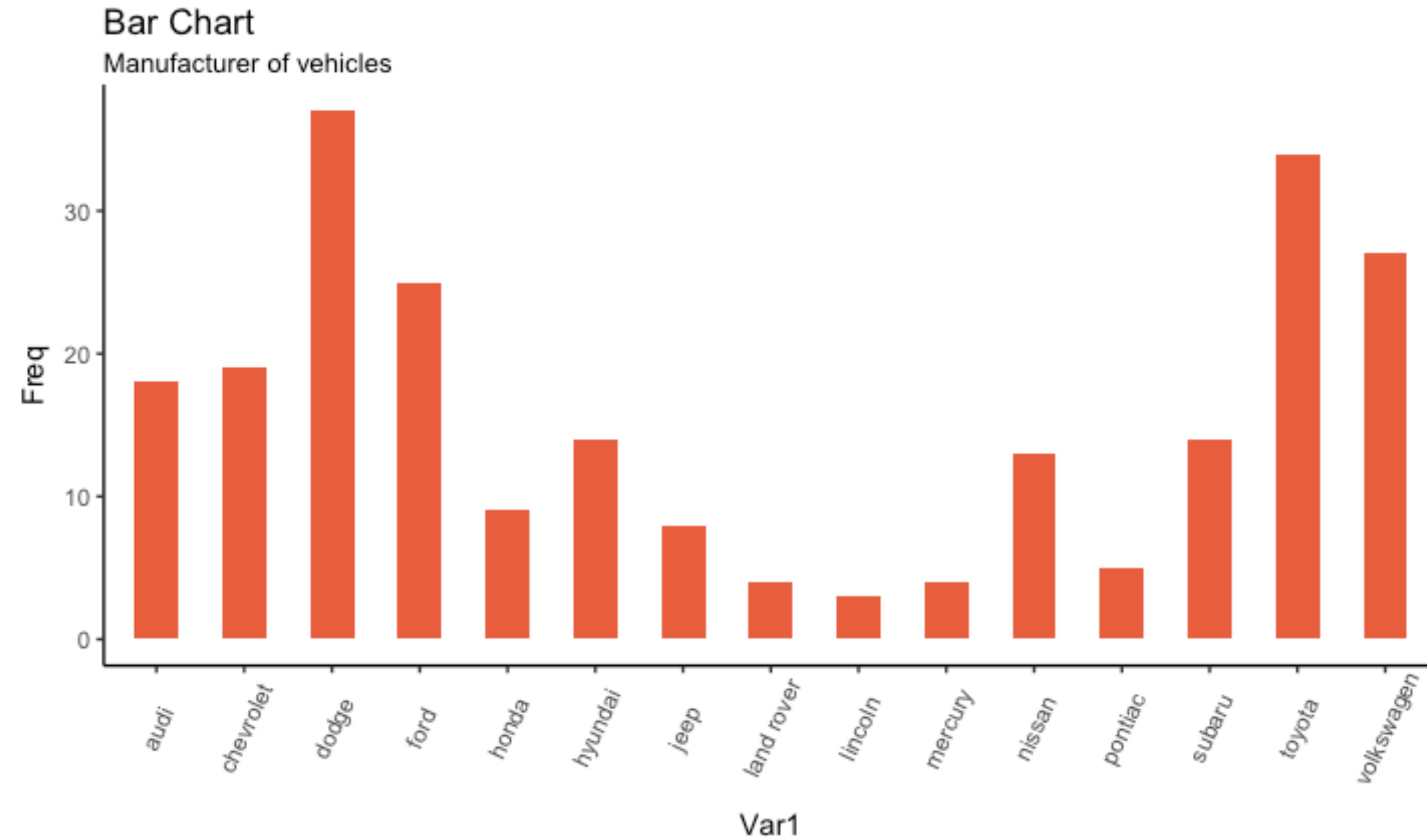
Source: mpg

# Know your toolkit of plots – composition



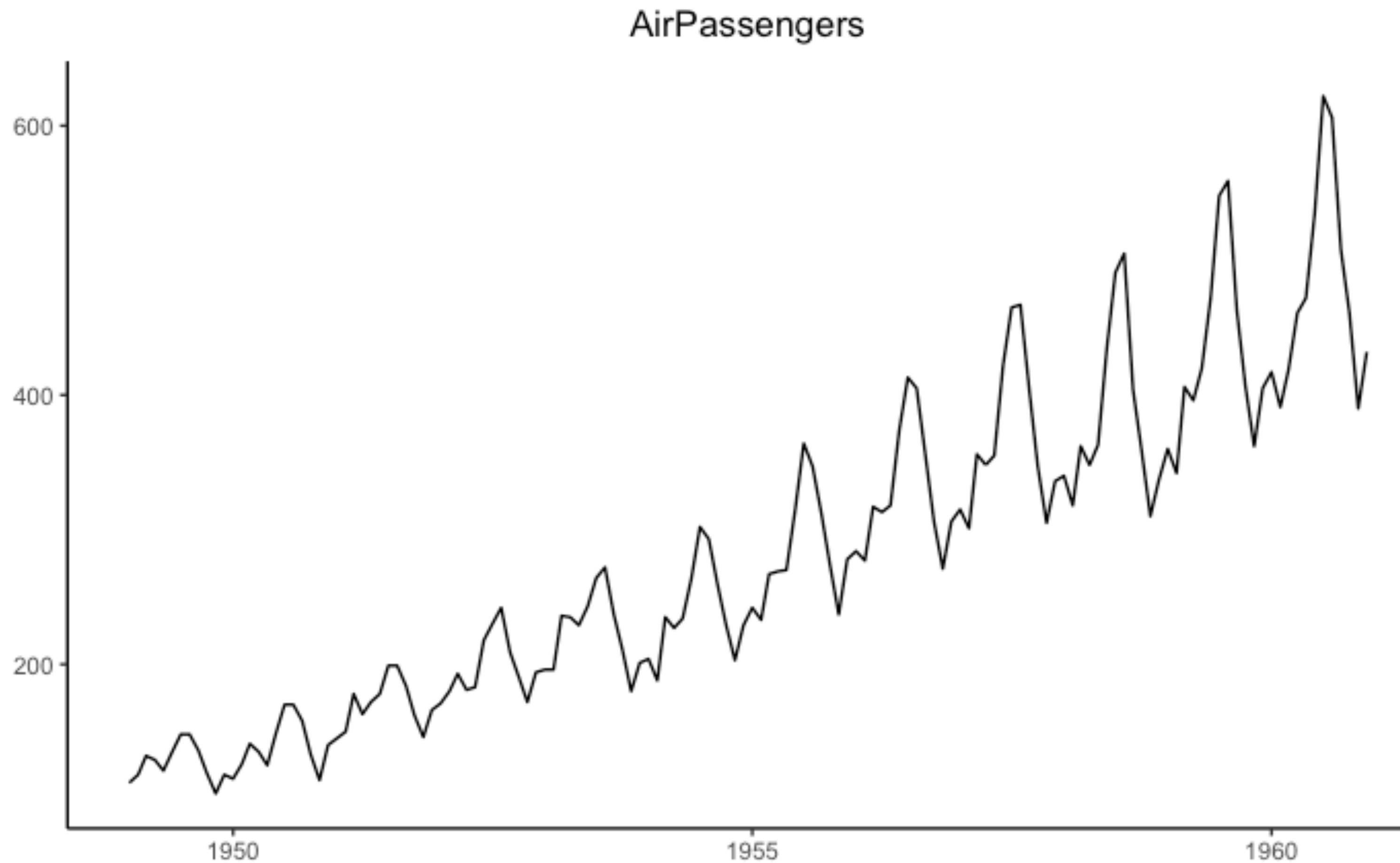
Source: mpg

# Know your toolkit of plots – composition



@JLSteenwyk

# Know your toolkit of plots – time series



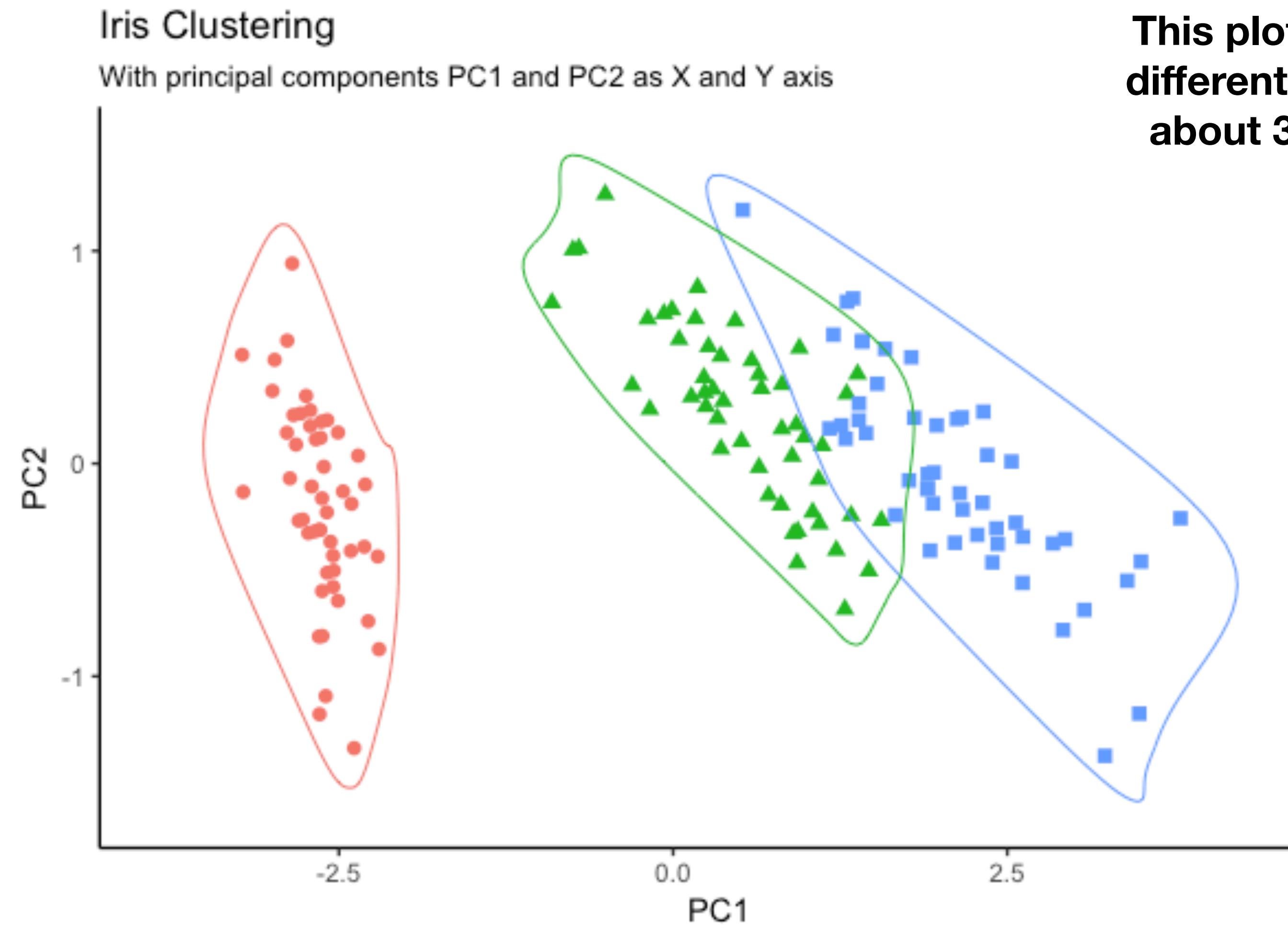
@JLSteenwyk

# Multidimensional data

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3.0	1.4	0.1	setosa
14	4.3	3.0	1.1	0.1	setosa
15	5.8	4.0	1.2	0.2	setosa
16	5.7	4.4	1.5	0.4	setosa
17	5.4	3.9	1.3	0.4	setosa
18	5.1	3.5	1.4	0.3	setosa
19	5.7	3.8	1.7	0.3	setosa
...	...	...	...	...	...



# Multidimensional data



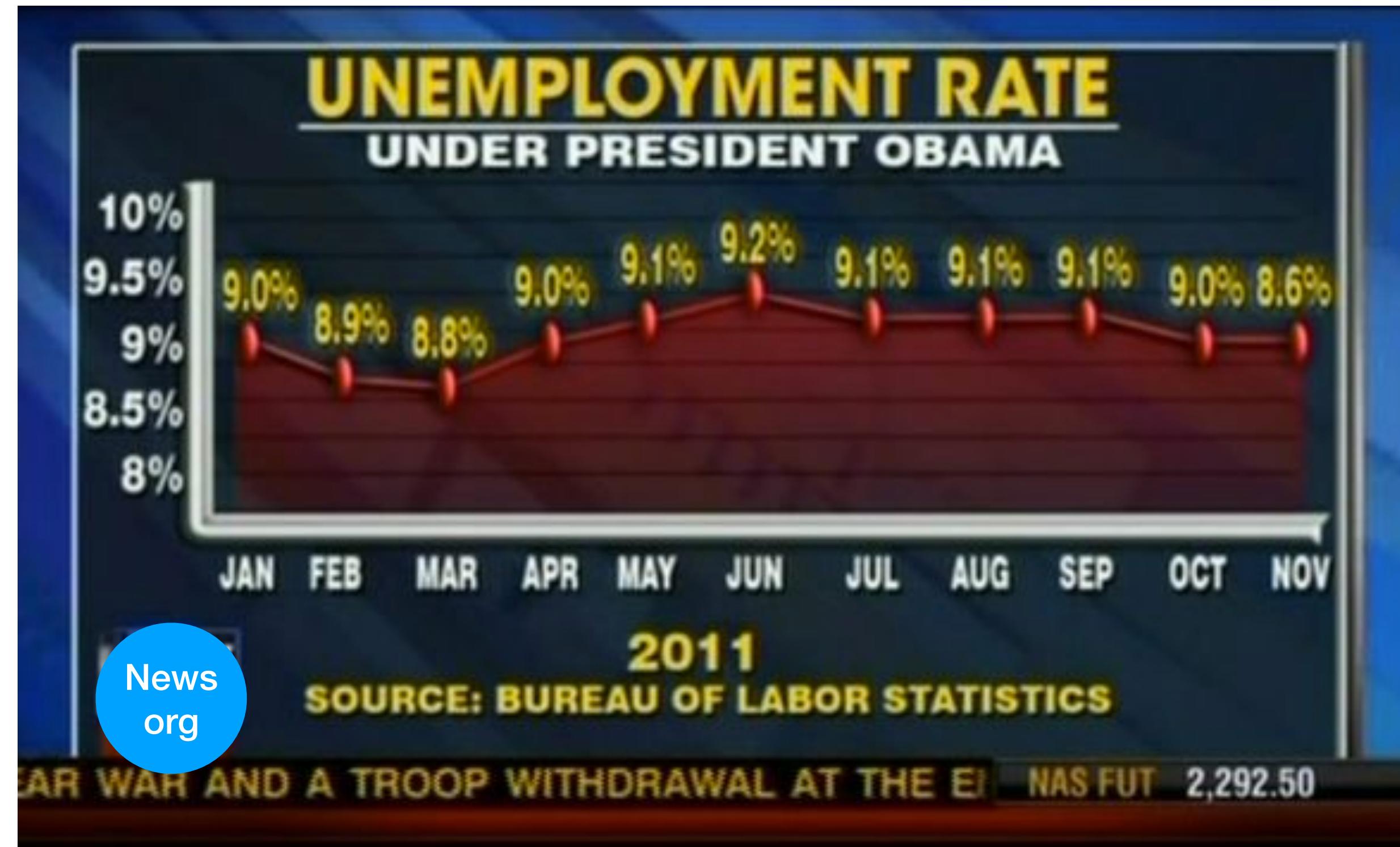
**This plot summarizes 4 different measurements about 3 plant species**

- Species
- setosa
  - versicolor
  - virginica

# Five principles to improve figures

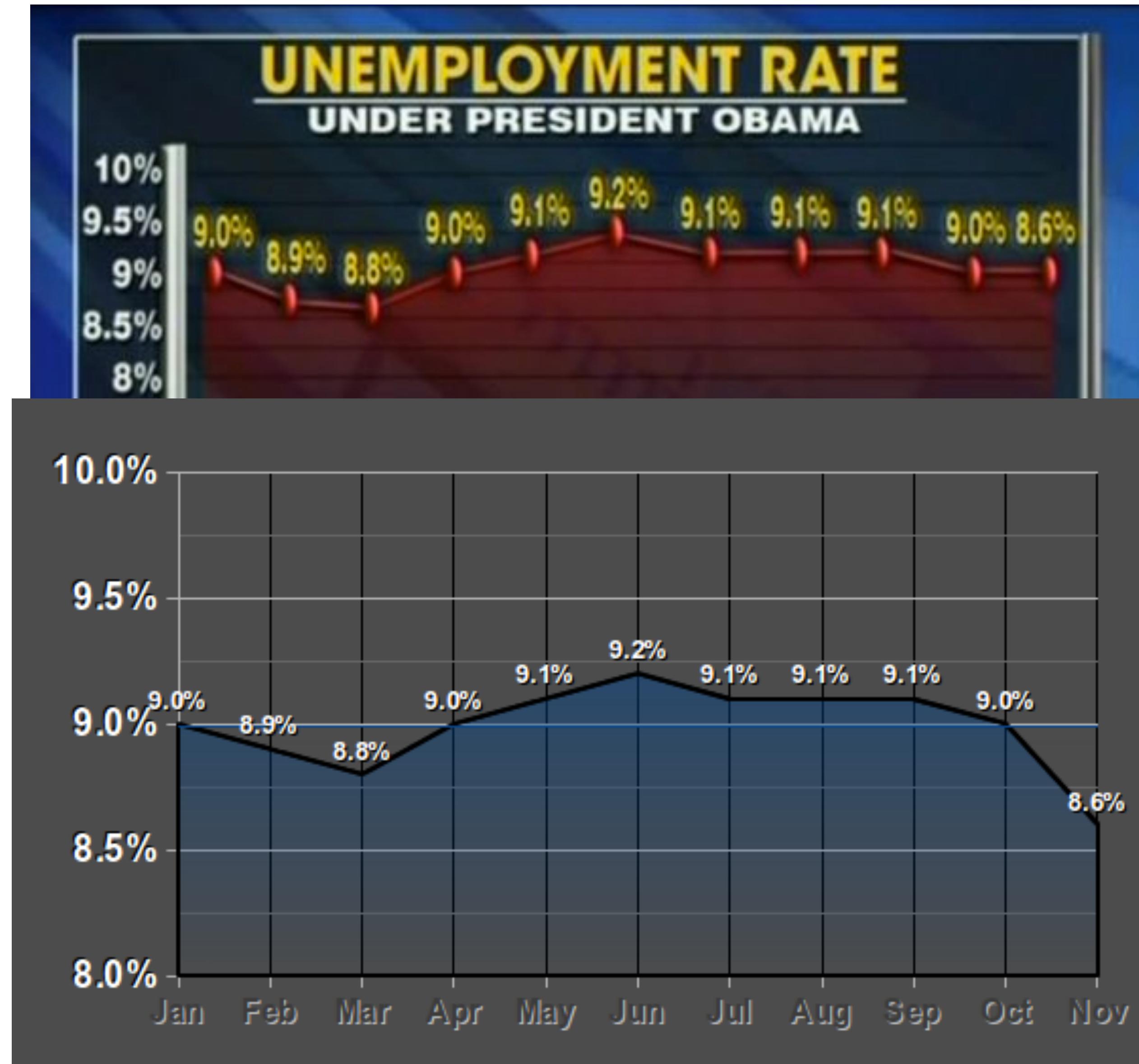
- 1) Know your tools
- 2) Identify your message
- 3) Be honest about your data
- 4) Use color effectively
- 5) Message before beauty

# Don't mislead the viewer



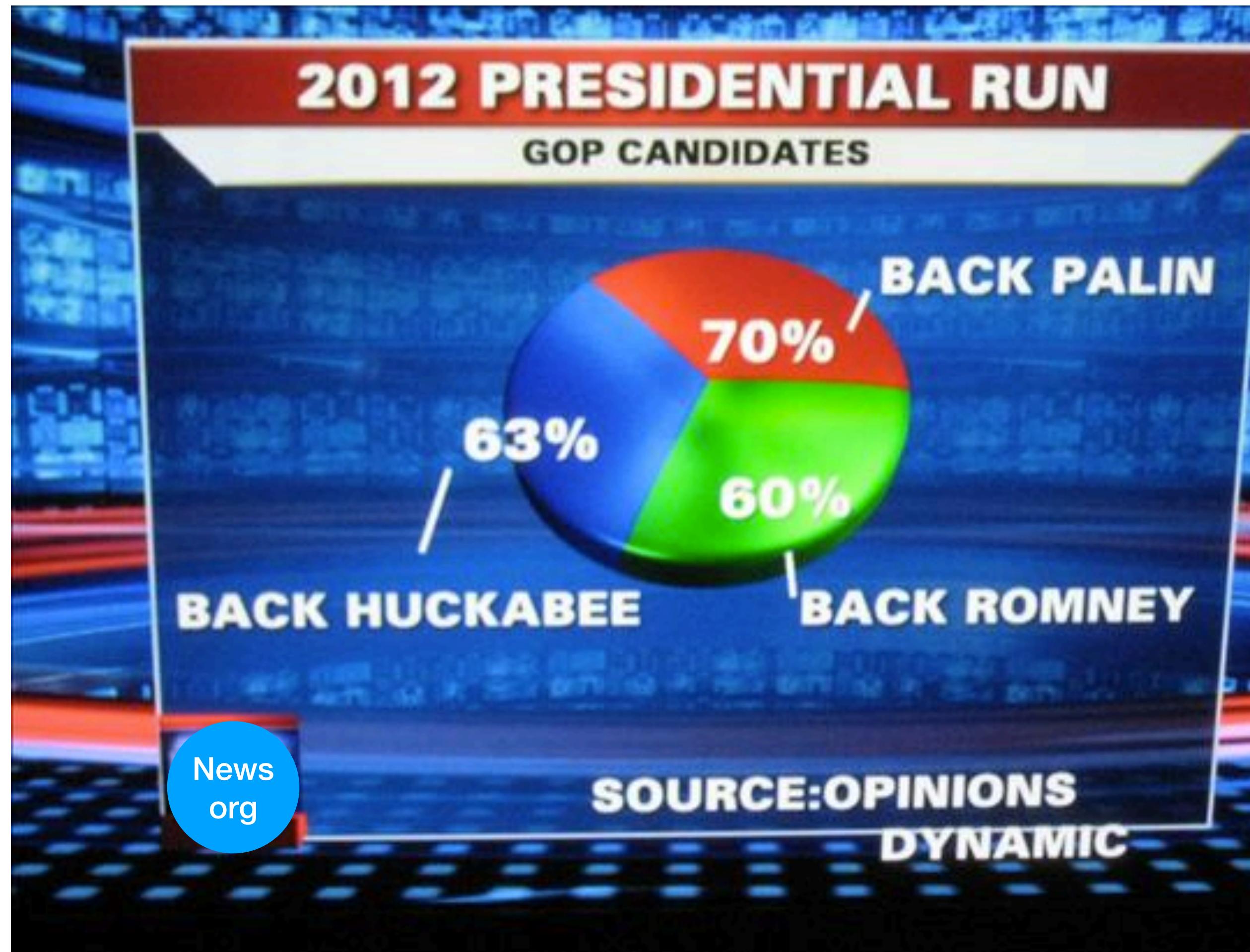
@JLSteenwyk

# Don't mislead the viewer

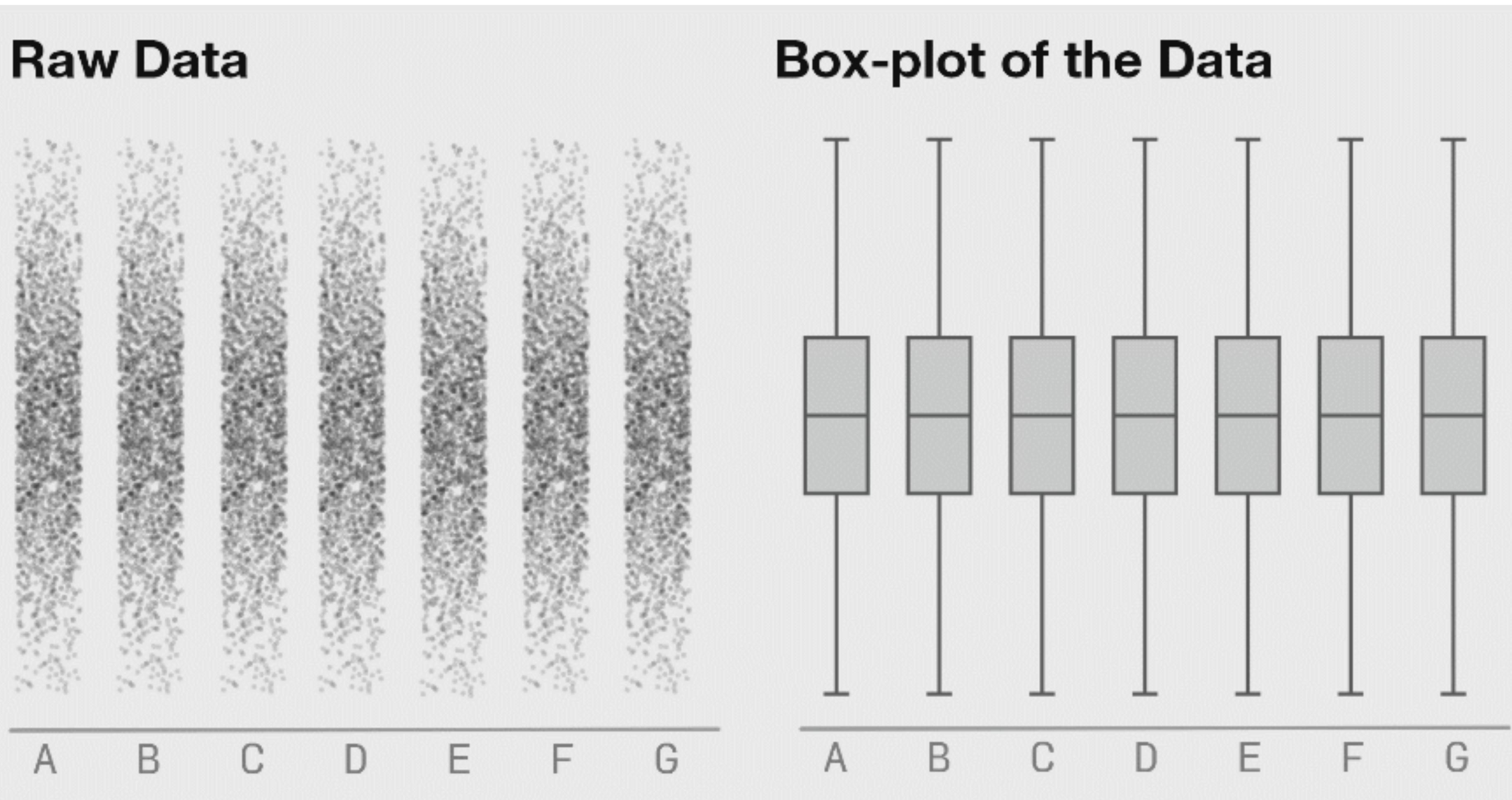


@JLSteenwyk

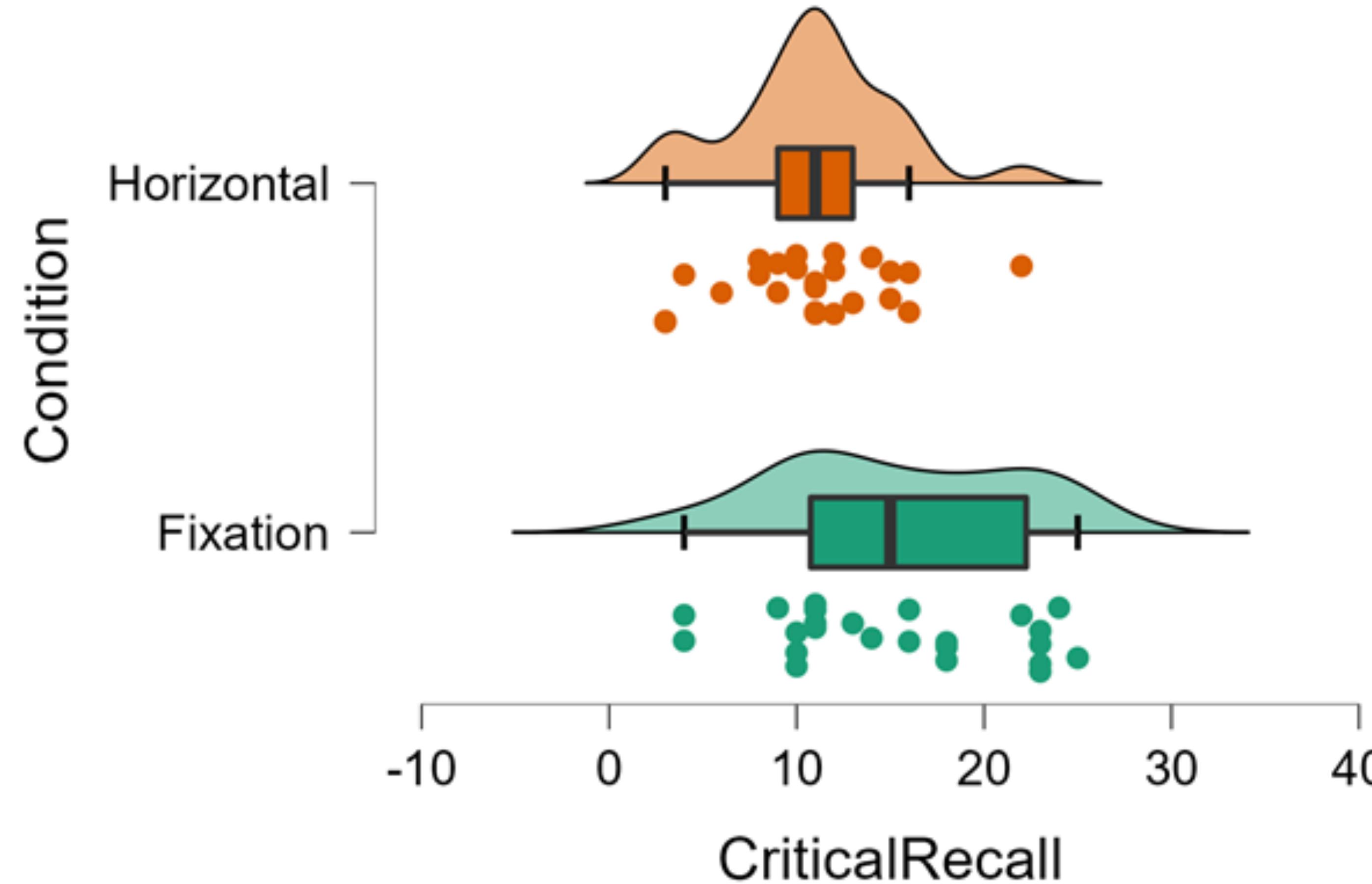
# Don't mislead the viewer



# Be transparent about the data



# Be transparent about the data



@JLSteenwyk

# Five principles to improve figures

- 1) Know your tools
- 2) Identify your message
- 3) Be honest about your data
- 4) Use color effectively
- 5) Message before beauty

# Utilize contrast appropriately

Adequate readability due to high value contrast



@JLSteenwyk

# Utilize contrast appropriately

Adequate readability due to high value contrast



Inadequate readability due to low value contrast



@JLSteenwyk

# Utilize contrast appropriately

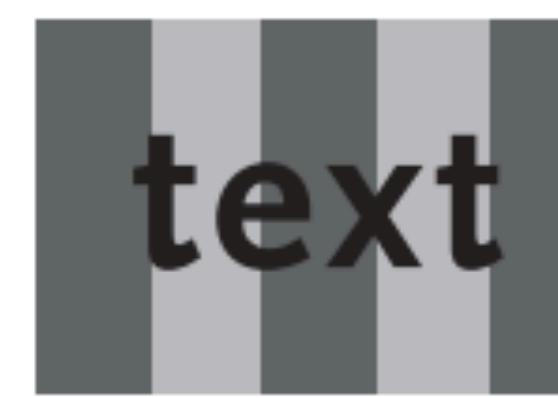
Adequate readability due to high value contrast



Inadequate readability due to low value contrast



Inadequate readability due to patterned background

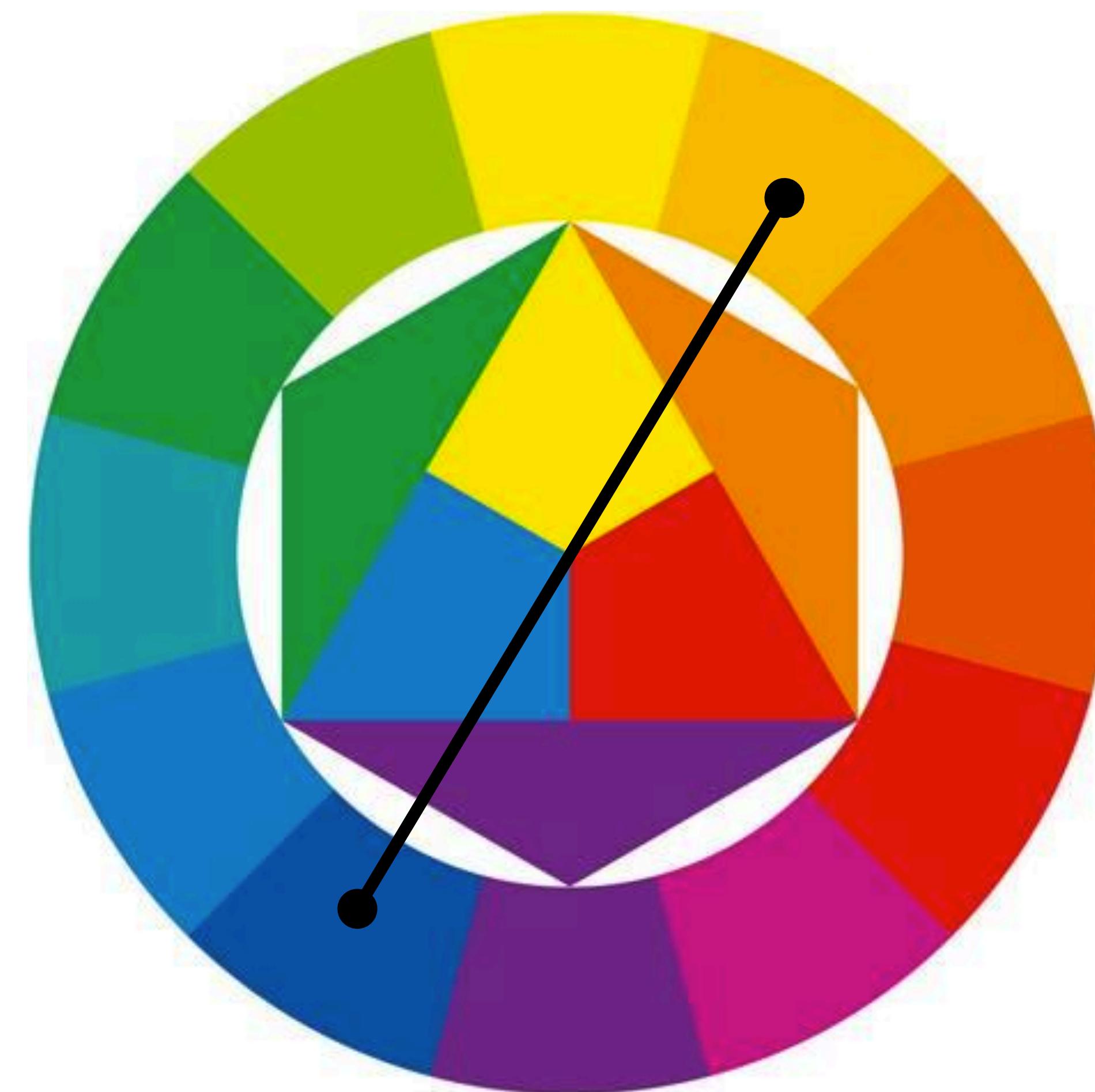


# Using color theory to pick combinations of colors



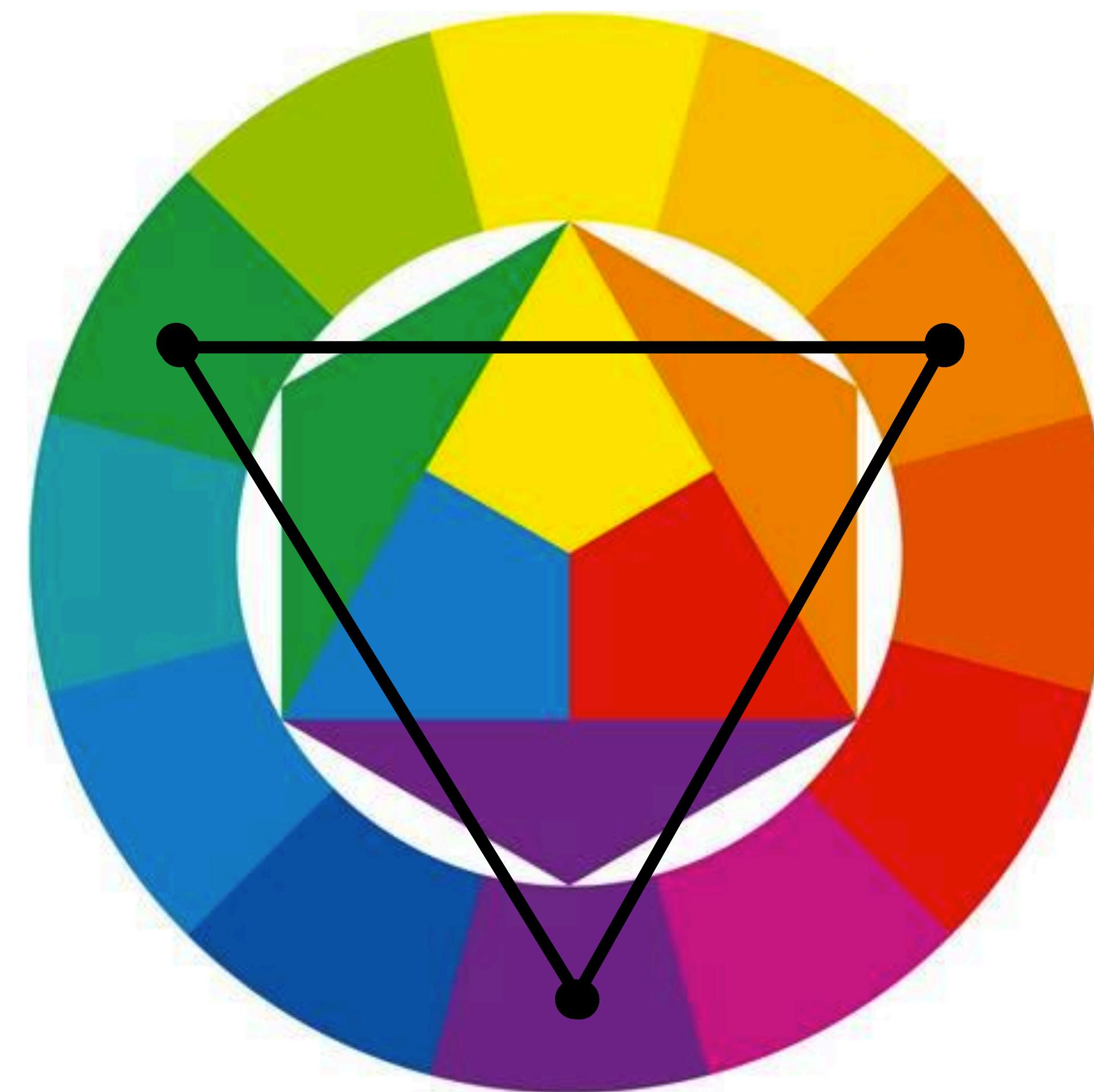
@JLSteenwyk

# Using color theory to pick combinations of colors



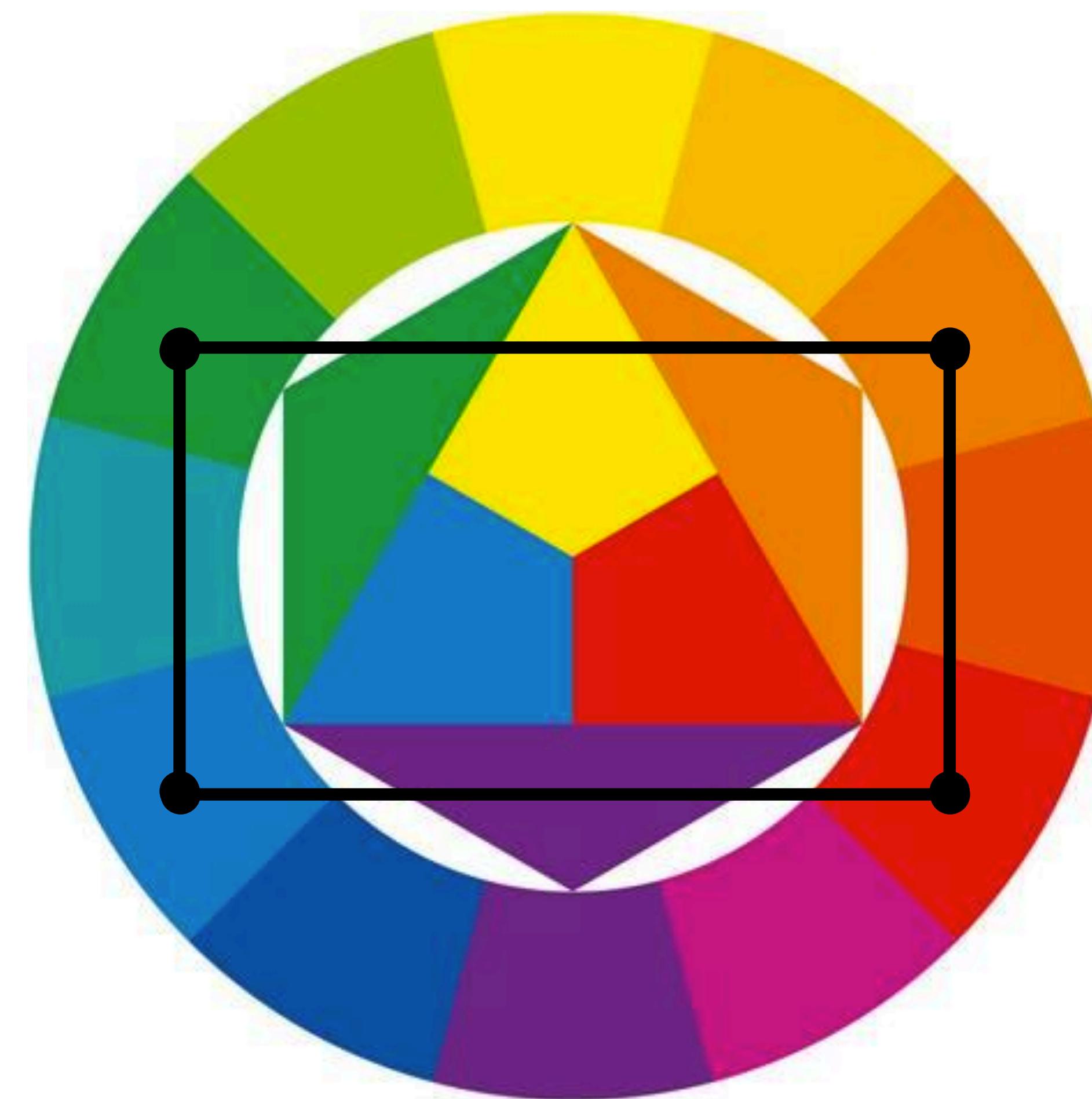
@JLSteenwyk

# Using color theory to pick combinations of colors

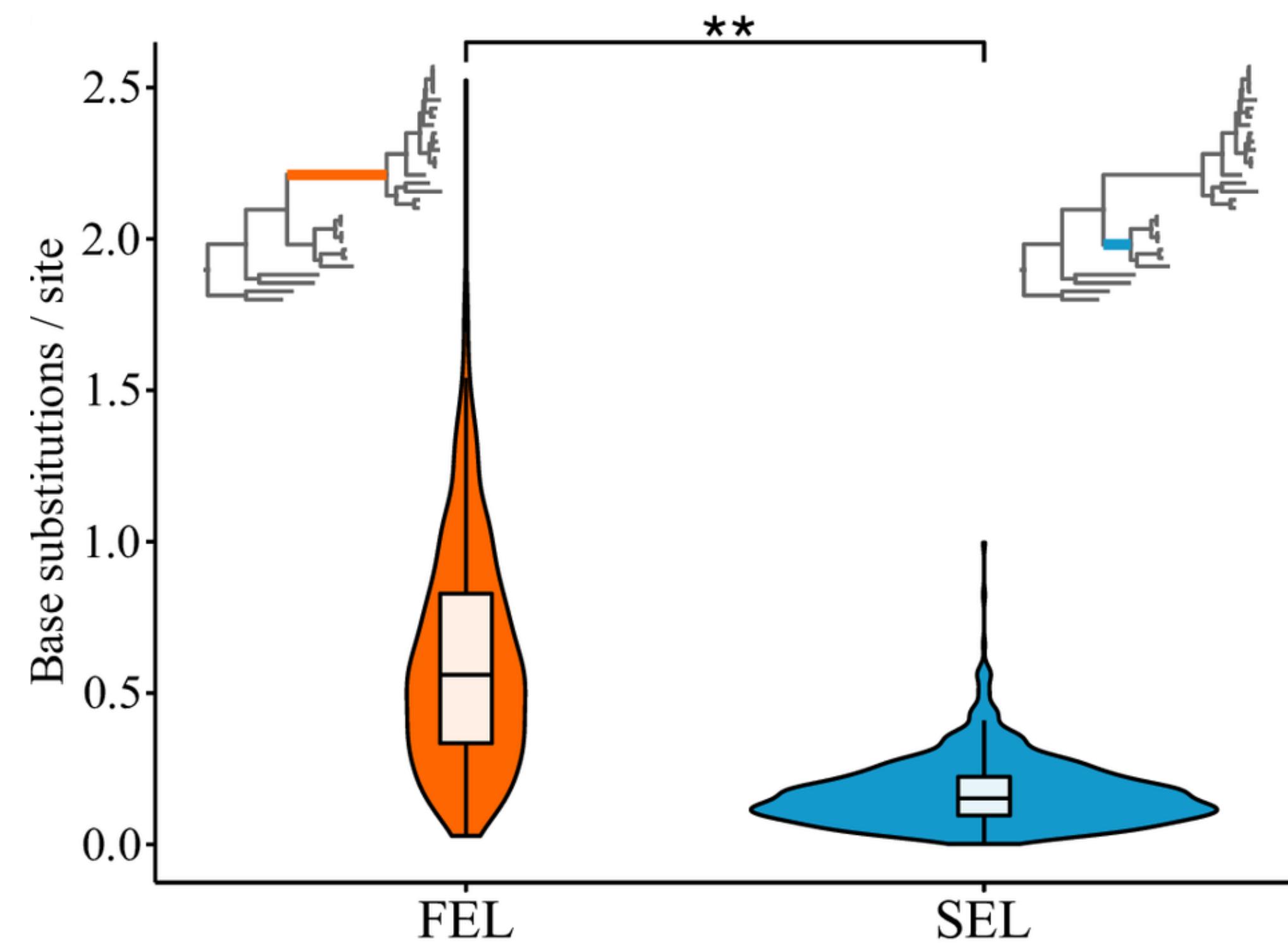


@JLSteenwyk

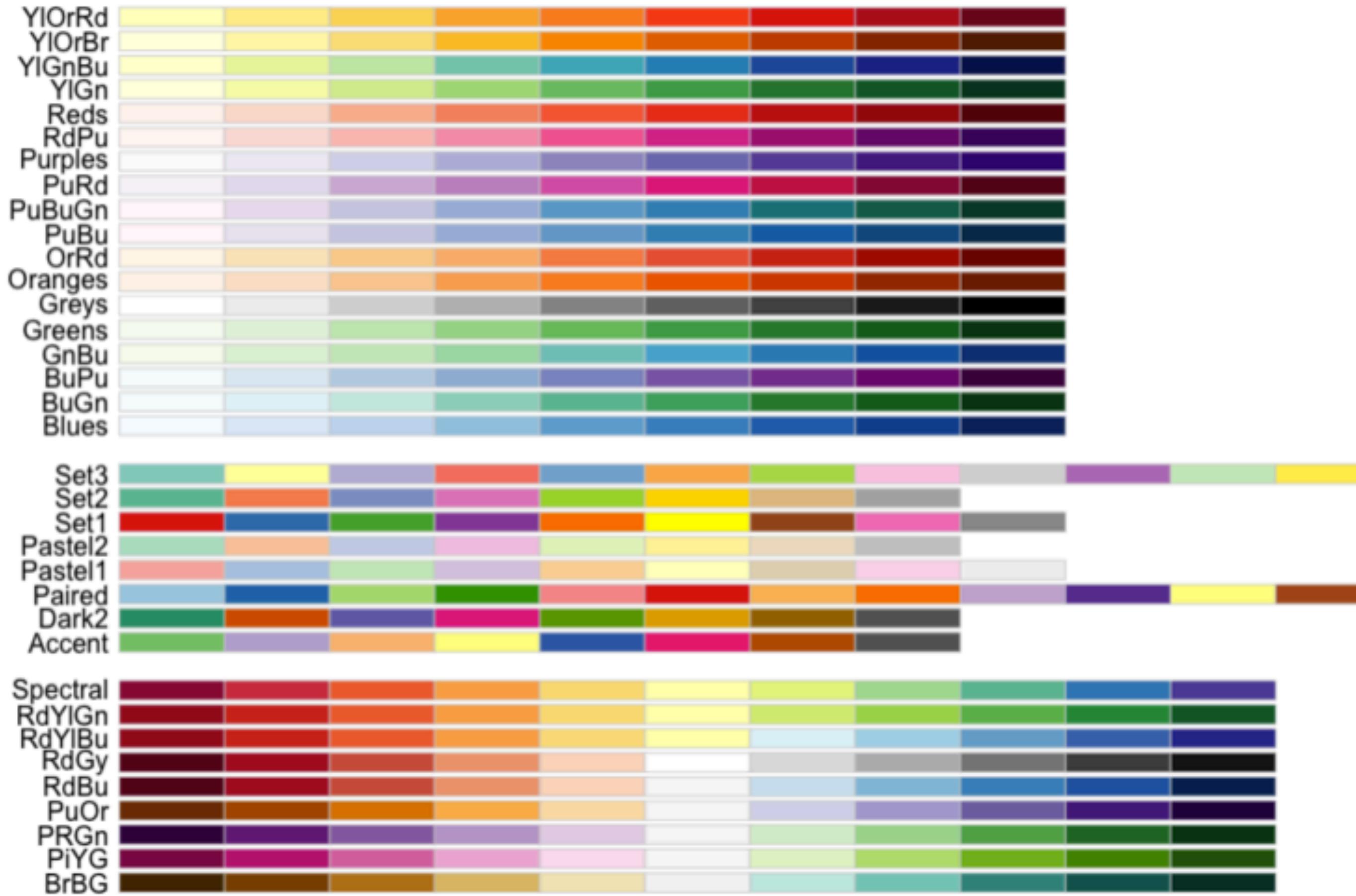
# Using color theory to pick combinations of colors



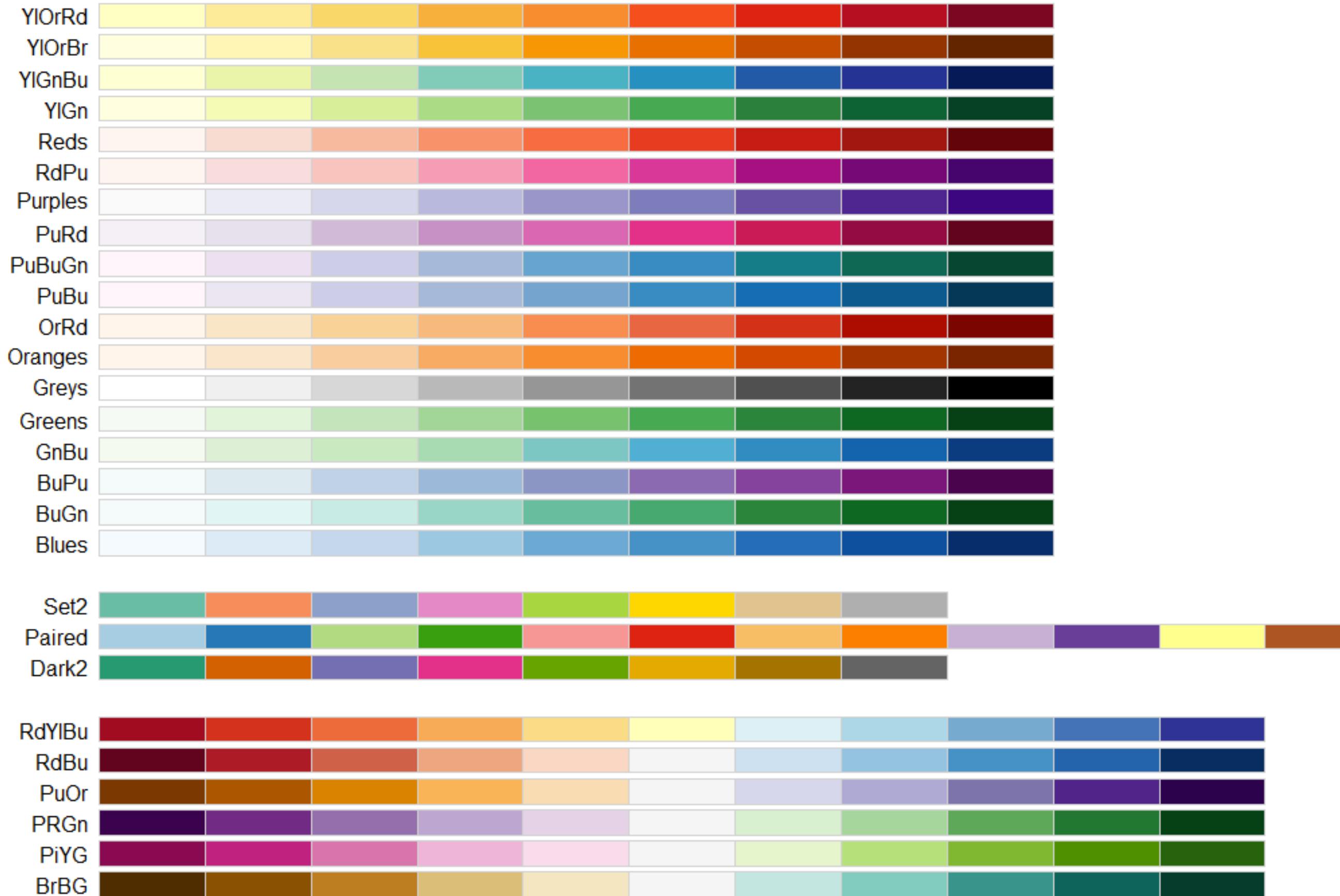
# Combining figure types (like the raincloud plot)



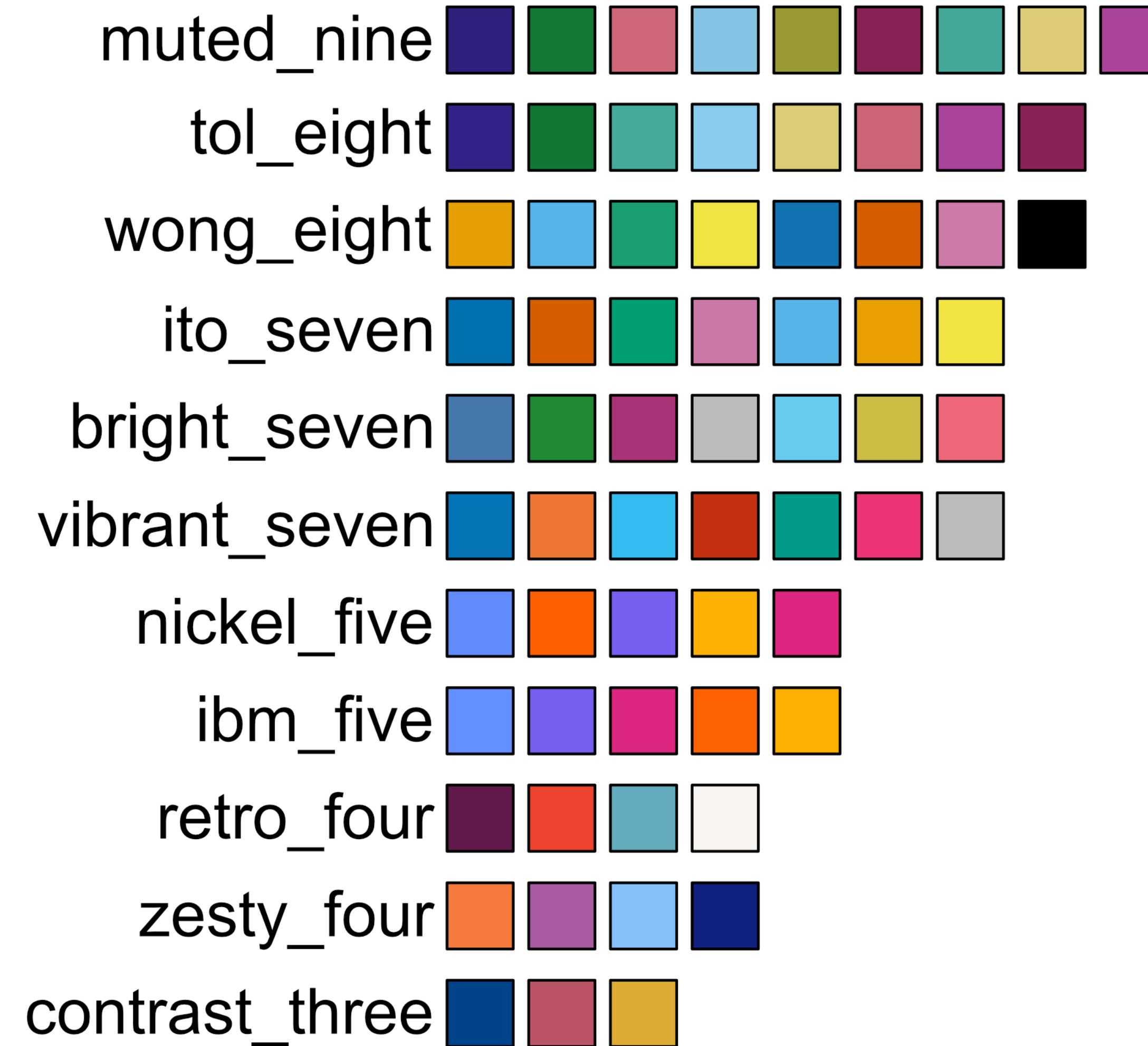
# Picking color palettes



# Colorblind-friendly color palettes



# Colorblind-friendly color palettes



@JLSteenwyk

# Five principles to improve figures

- 1) Know your tools
- 2) Identify your message
- 3) Be honest about your data
- 4) Use color effectively
- 5) Message before beauty

# When beautification goes wrong

**facebook.com/peptobismol**

**Who doesn't love grilling?**<sup>1</sup>

The most popular grilling occasions are Fourth of July, Memorial Day and Labor Day.

45% of Grill owners use the grill at least 1-2 times per week during peak summer months.

82% of U.S. households own a grill or smoker.

16,700,000 Grills were shipped in 2009

The most popular foods for grilling are:

**1 BURGERS**

**When we get together, we tend to eat more.<sup>2</sup>**

Meals eaten with one other person were 33% larger than those eaten alone.

**Major League Eating champions are eating more these days.<sup>6</sup>**

The number of hot dogs eaten at the Nathan's Famous Hot Dog Eating Contest has been trending upwards since its inception in 1916.

**It's Summertime!**

That means food festivals, amusement parks, baseball games and eating contests. Americans love to celebrate the warm weather by getting together and enjoying food with friends. When the enjoyment of summer leads to eating just one too many favorite foods, you may experience stomach troubles. Eat, drink and be covered\*.

**Record crowds are lining up for all that fried goodness.**

Big Tex, the Texas State Fair attracts more than 3 mil. attendees every year.<sup>5</sup>

**TEXAS** 3 M  
**MINNESOTA** 1.7 M  
**ARIZONA** 1.3 M  
**NEW ENGLAND** 1.2 M  
**NEW YORK** 1.1 M  
**IOWA** 1 M

**Over 14 Major League Baseball stadiums have all-you-can-eat options.<sup>4</sup>**

Some of the tastier menu items include:

- Smoked chicken barbecue sandwich
- Buffalo chicken wings
- All beef franks
- Baked potato salad
- Dry roasted salted peanuts
- Popcorn
- Chocolate chip cookies
- Ice cold soda
- Draft beer

**More Than 59 Million**  
Americans will attend a festival or fair this year.<sup>1</sup>

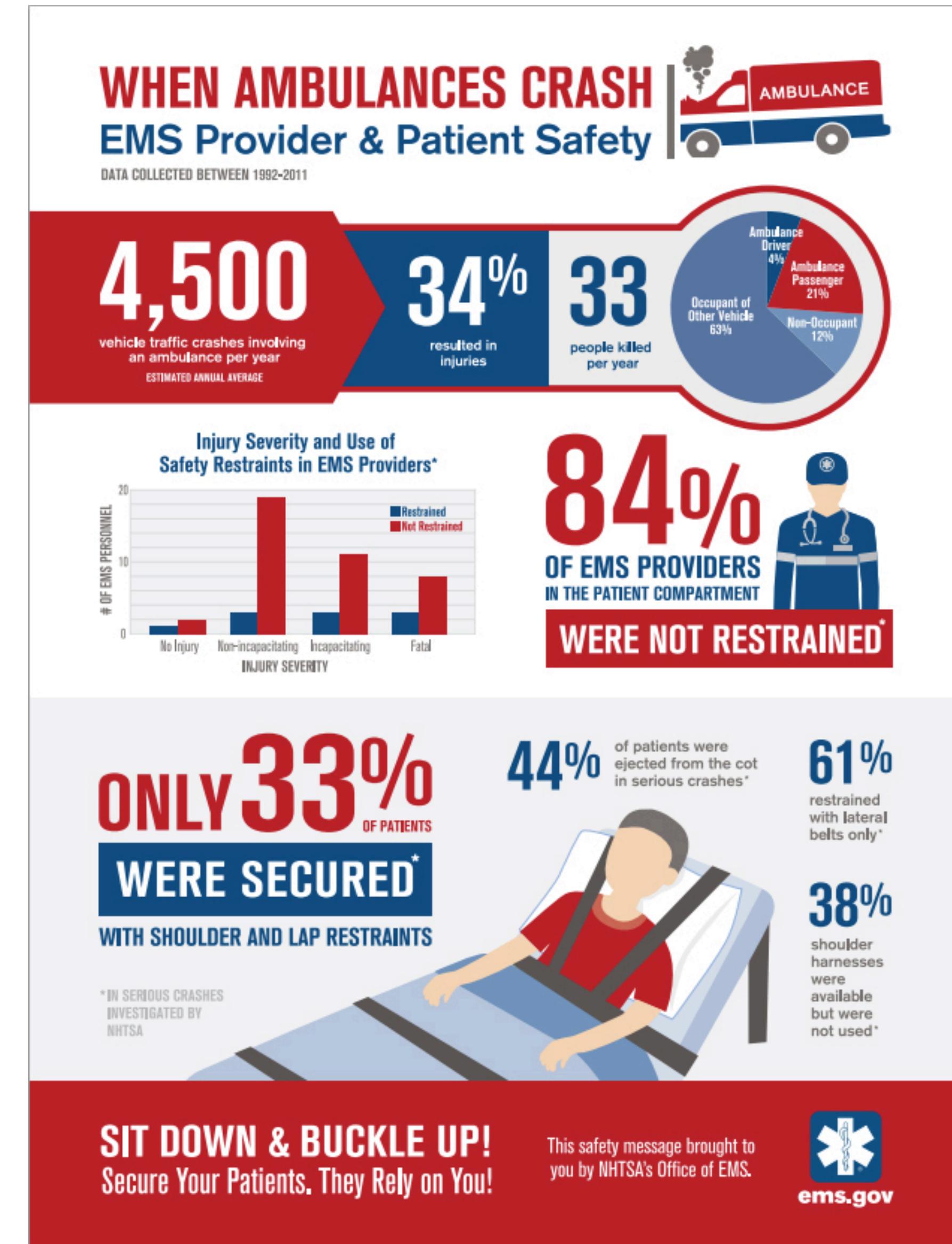
**SOURCES**

<sup>1</sup> TIA's The Historic/Cultural Traveler, 2003 Edition  
<sup>2</sup> Departments of Marketing and Nutritional Science, University of Illinois at Urbana-Champaign  
<sup>3</sup> Hearth, Patio & Barbecue Association  
<sup>4</sup> MLB Advanced Media  
<sup>5</sup> [http://en.wikipedia.org/wiki/State\\_fair](http://en.wikipedia.org/wiki/State_fair)  
<sup>6</sup> Major League Eating

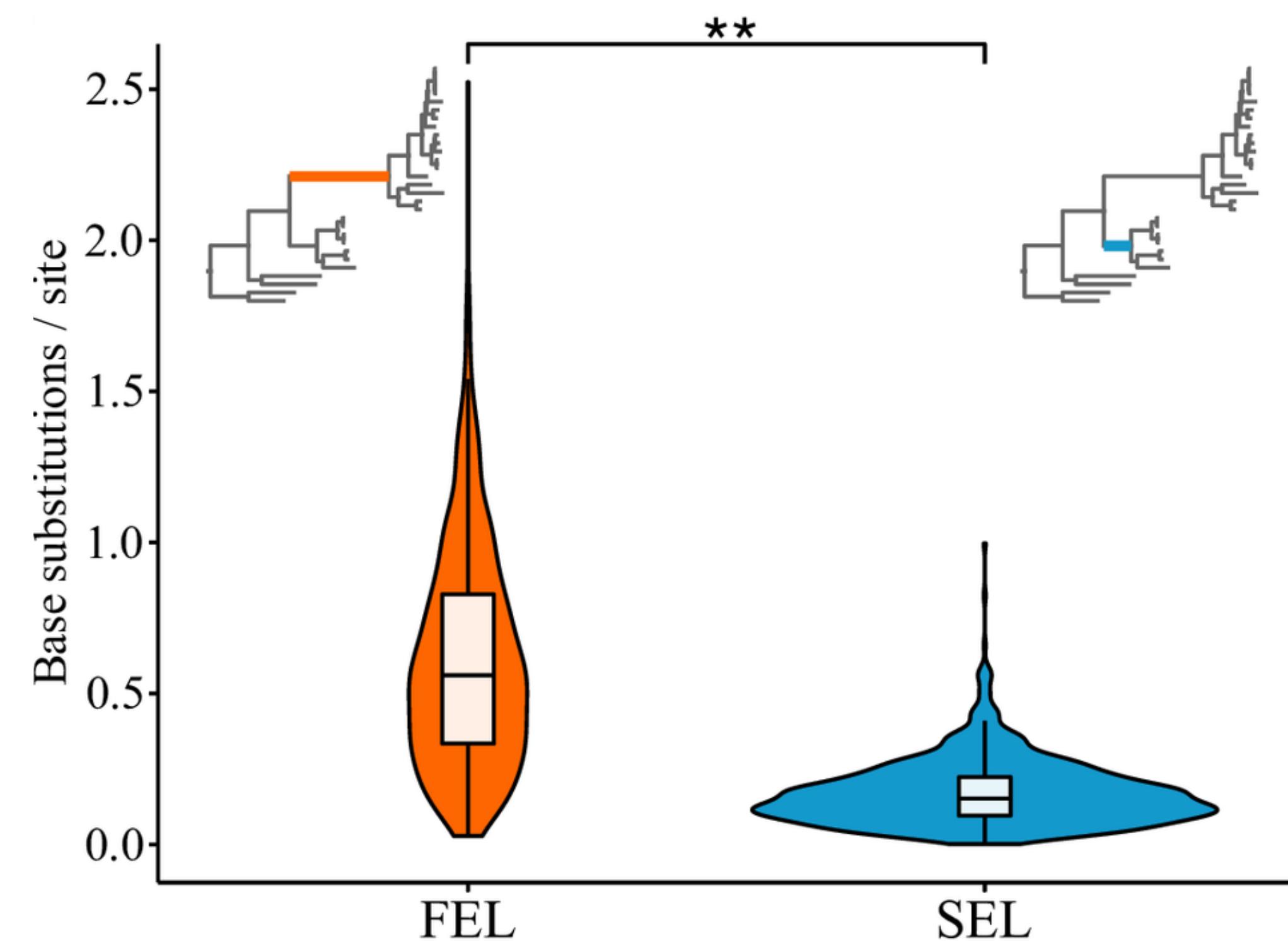
**Eat, drink & be covered**

\*Use as directed.

# Less can be more

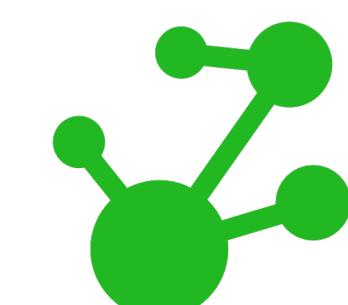
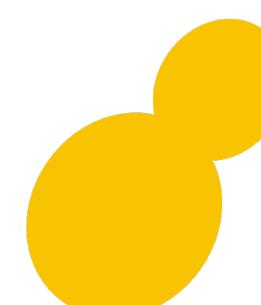


# Keep it simple

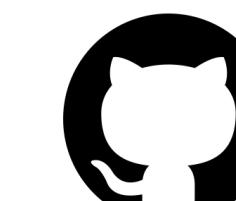


# Workshop portion

Jacob Steenwyk



@JLSteenwyk



jlsteenwyk.github.io

# Goals

- Basic data examination
- Learn to plot different numbers dimensions in a data set
- Provide the ‘know how’ to create figures for your data



@JLSteenwyk

# Where to find the worksheet

1. [jlsteenwyk.com](http://jlsteenwyk.com)

2. Click the resources tab

3. Under R workshop

## Bioinformatics tutorials (written by Jacob L. Steenwyk)

### Phylogenetics/Phylogenomics

- Five-step phylogenomics
- Phylogenomic sensitivity analysis and detecting incongruence
- Inferring a phylogeny, from sequences to tree
- Trim multiple sequence alignments
- Evaluate the accuracy of a multiple sequence alignment
- Calculate gene-gene covariation
- Summarize information content in phylogenomic data
- Identifying signatures of rapid radiation
- Prune phylogenies using a GUI

### UNIX

- Key resources
  - Common `tmux` commands and explanations thereof
- Challenges
  - Math and string replacements - Beginner
  - Looping through multiple gff files - Beginner
  - If statements - Beginner

### Making figures in R (Beginner)

- 2024 Version
  - Presentation
  - Worksheet
  - Answer sheet
- 2018 Version
  - Presentation
  - Worksheet
  - Answer sheet



@JLSteenwyk