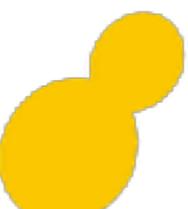


# Figure making in R

## A workshop for beginners

Jacob Steenwyk



@JLSteenwyk



jlsteenwyk.github.io

**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)
```

# 2-part workshop

## Part 1

Seminar-style presentation  
~.5 mins

## Part 2

Workshop tutorial  
~1.5 hours

# Workshop is made possible by...

Jacob L. Steenwyk  
@jlsteenwyk

Kate Snyder  
@KateTSnyder

Katrina Ngo  
@phd\_in\_trying

Anne Hatmaker  
@annemakerofhats

Jessie Perlmutter  
@jipermutter

Cait Kirby  
and GSA



IBA



AMERICAN  
SOCIETY FOR  
MICROBIOLOGY

# Goals

- Communication, things you've already known

# Goals

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

# Goals

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

# Images communicate observations

# Images communicate observations



Pech Merle Cave Painting

# Images communicate observations



Pech Merle Cave Painting

**Mammoth and a bison**

# Images communicate observations

**horns, eyes, color, tail**



Pech Merle Cave Painting

**Mammoth and a bison**



Prehistoric Spanish Cave Painting

**Bison**

# Images communicate observations

**horns, eyes, color, tail**



Pech Merle Cave Painting

**Mammoth and a bison**



Prehistoric Spanish Cave Painting

**Bison**

# Images 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**

# Images communicate observations

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

**horns, eyes, color, tail**



Pech Merle Cave Painting

**Mammoth and a bison**

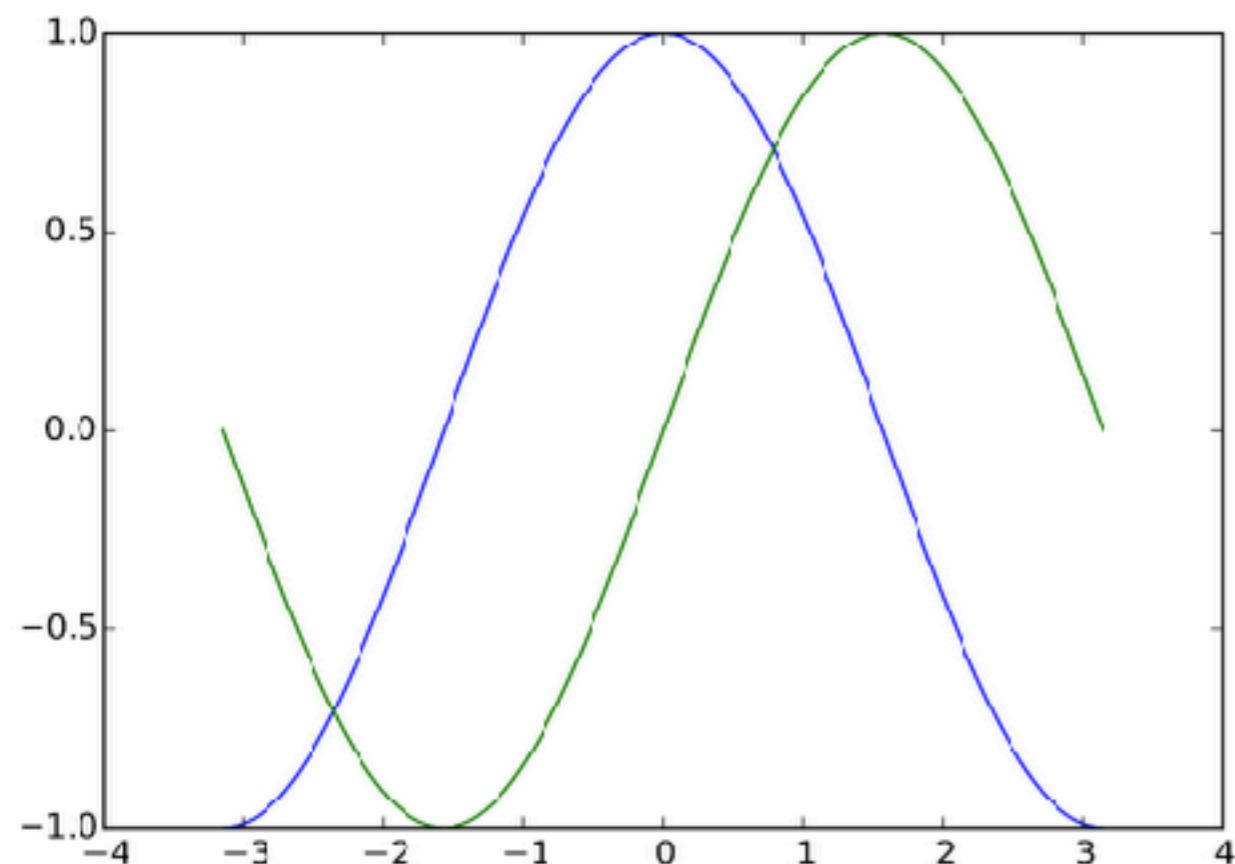
**Less assumptions made by the viewer**



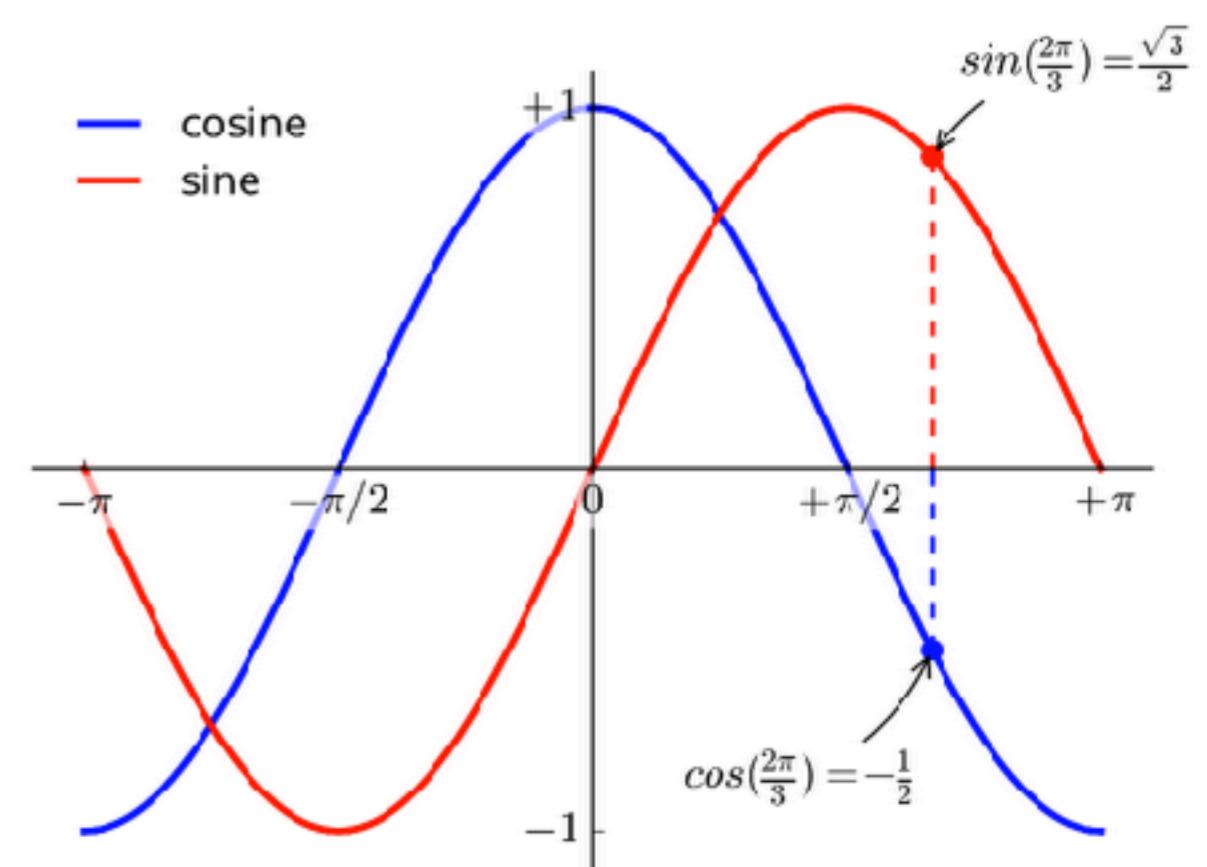
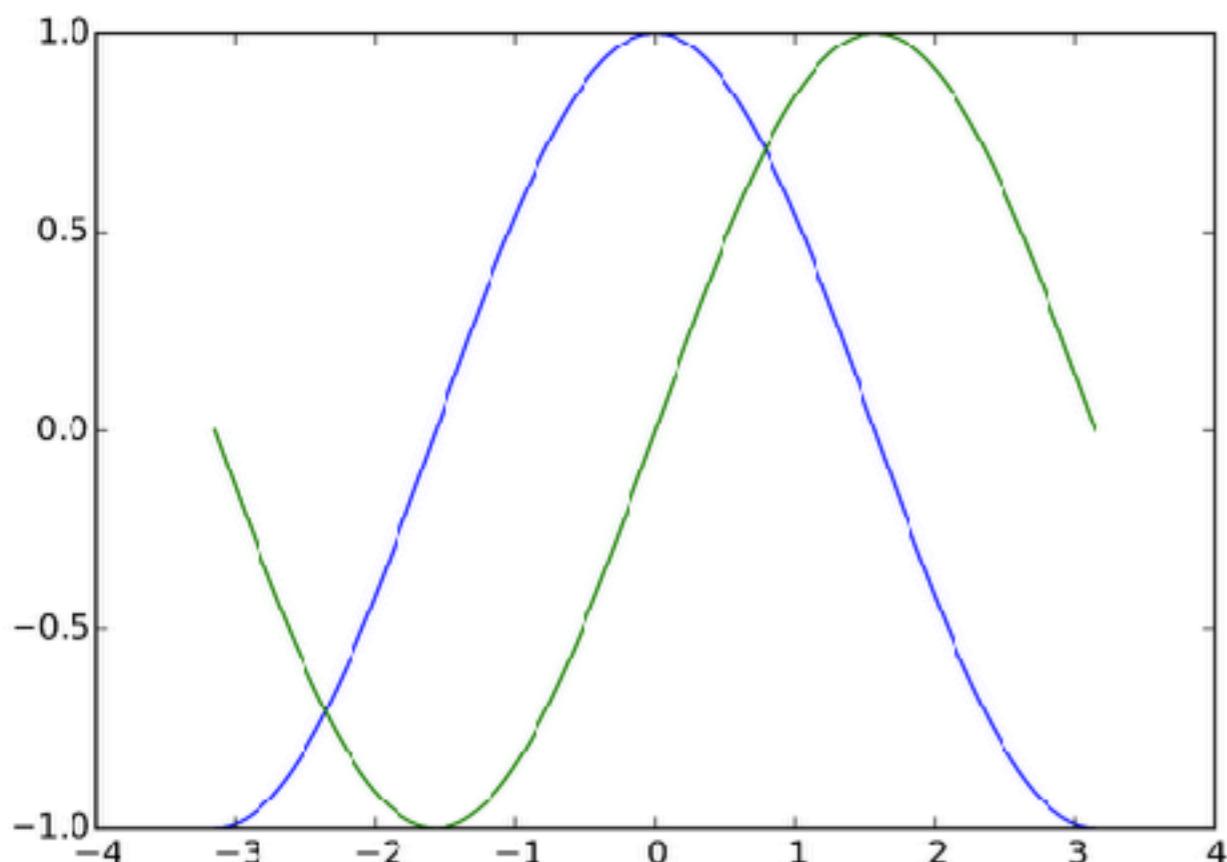
Prehistoric Spanish Cave Painting

**Bison**

# A simple example of the same phenomenon



# A simple example of the same phenomenon



# 5 Principles for Improving Figures

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

# 5 Principles for Improving Figures

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

# Softwares to plot data

# Softwares to plot data

**matplotlib**



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

# Softwares to plot data

**Seaborn**

**matplotlib**



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

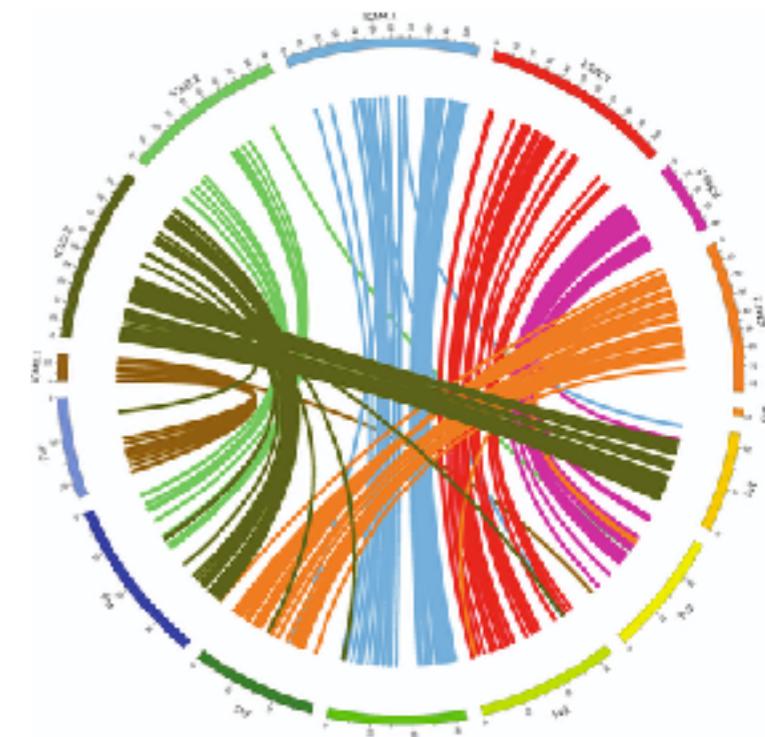
# Softwares to plot data

**Seaborn**

**matplotlib**



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



**Circos**

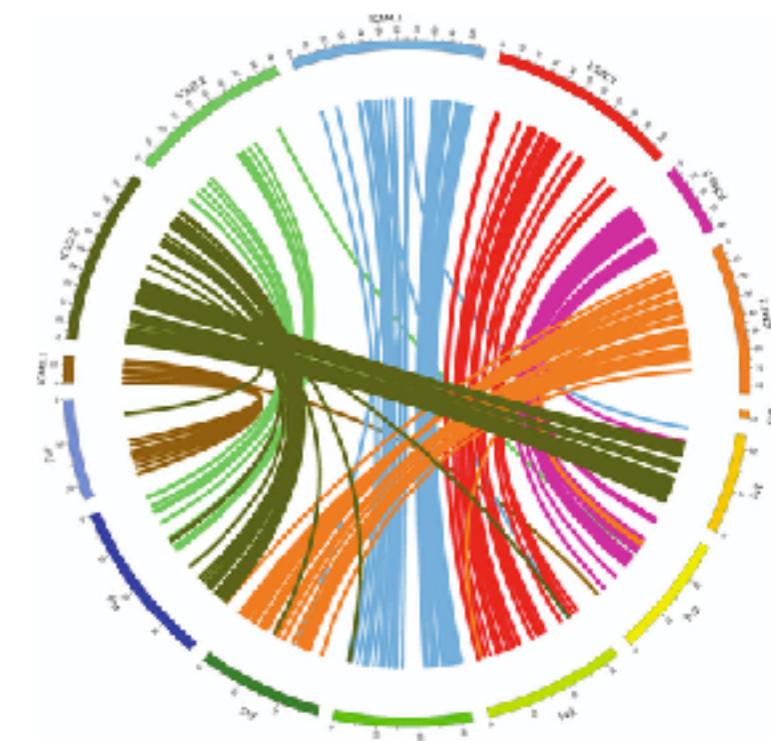
# Softwares to plot data

**Seaborn**

**matplotlib**

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

 **Cytoscape**



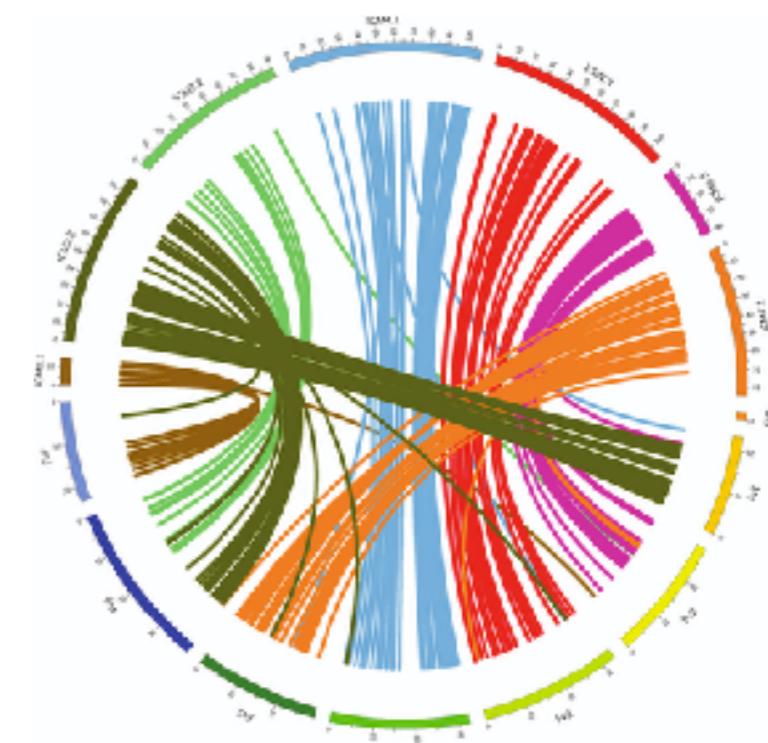
**Circos**

# Softwares to plot data

**Seaborn**

**matplotlib**

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

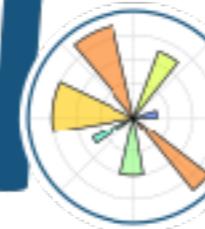


**Circos**

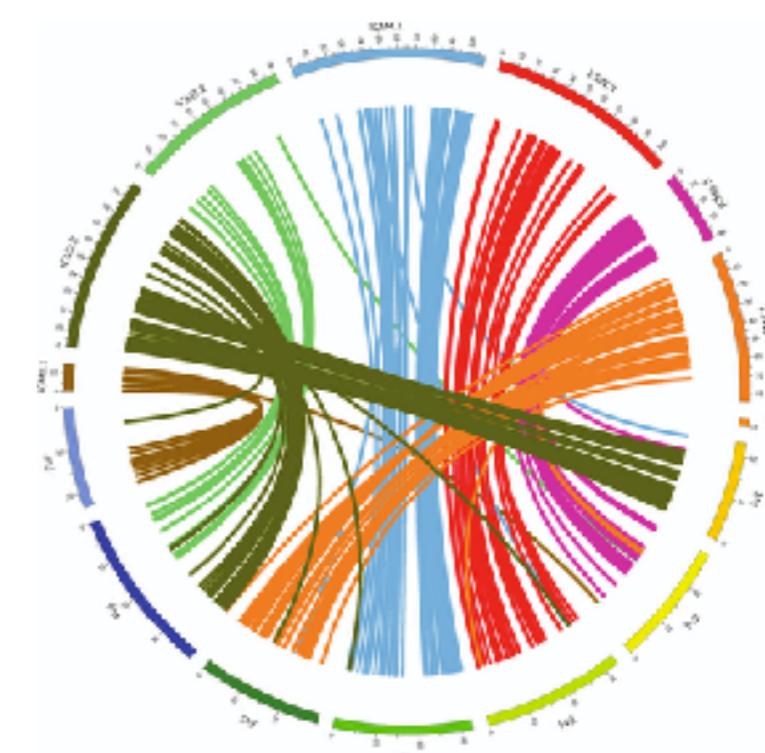
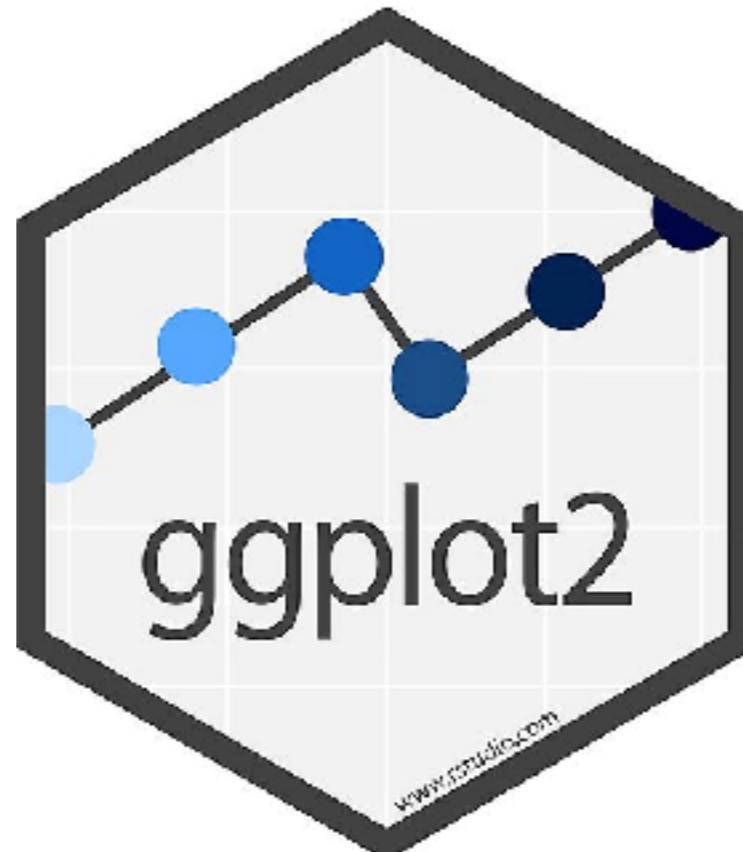
# Softwares to plot data

# Seaborn

# matplotlib



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



Circos

# Softwares to edit graphics



**GIMP**

# Softwares to edit graphics



**GIMP**



**INKSCAPE**

# Softwares to edit graphics



**GIMP**



**INKSCAPE**



# Softwares to edit graphics



**GIMP**



**INKSCAPE**



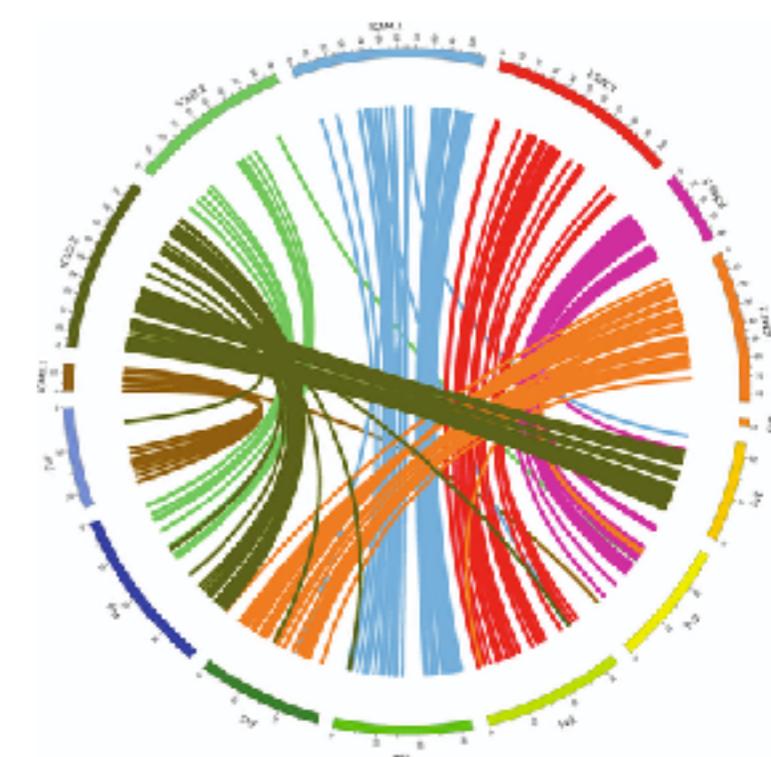
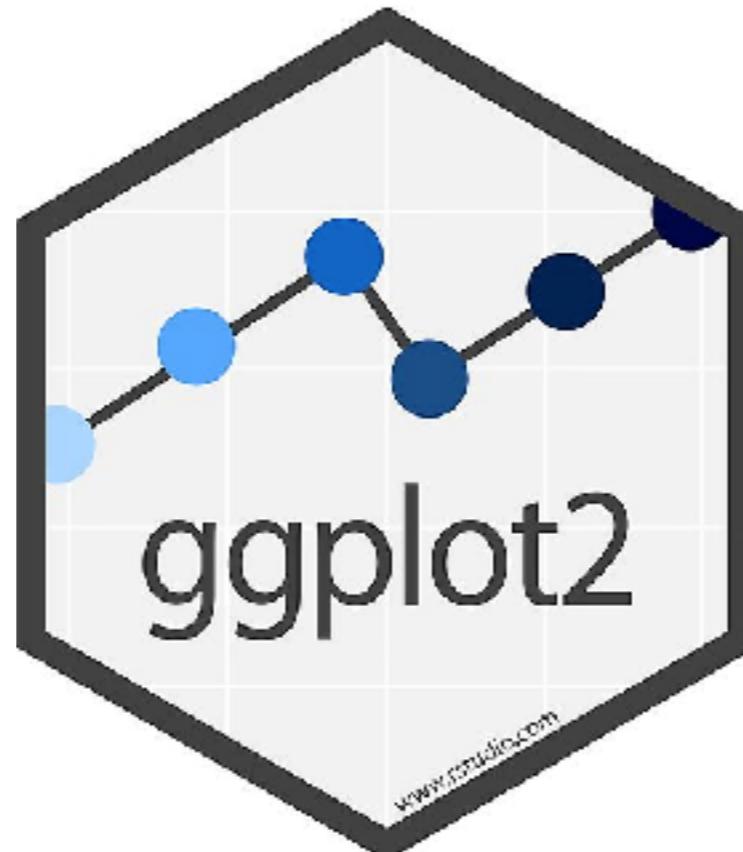
# How to choose?

**Seaborn**

**matplotlib**



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



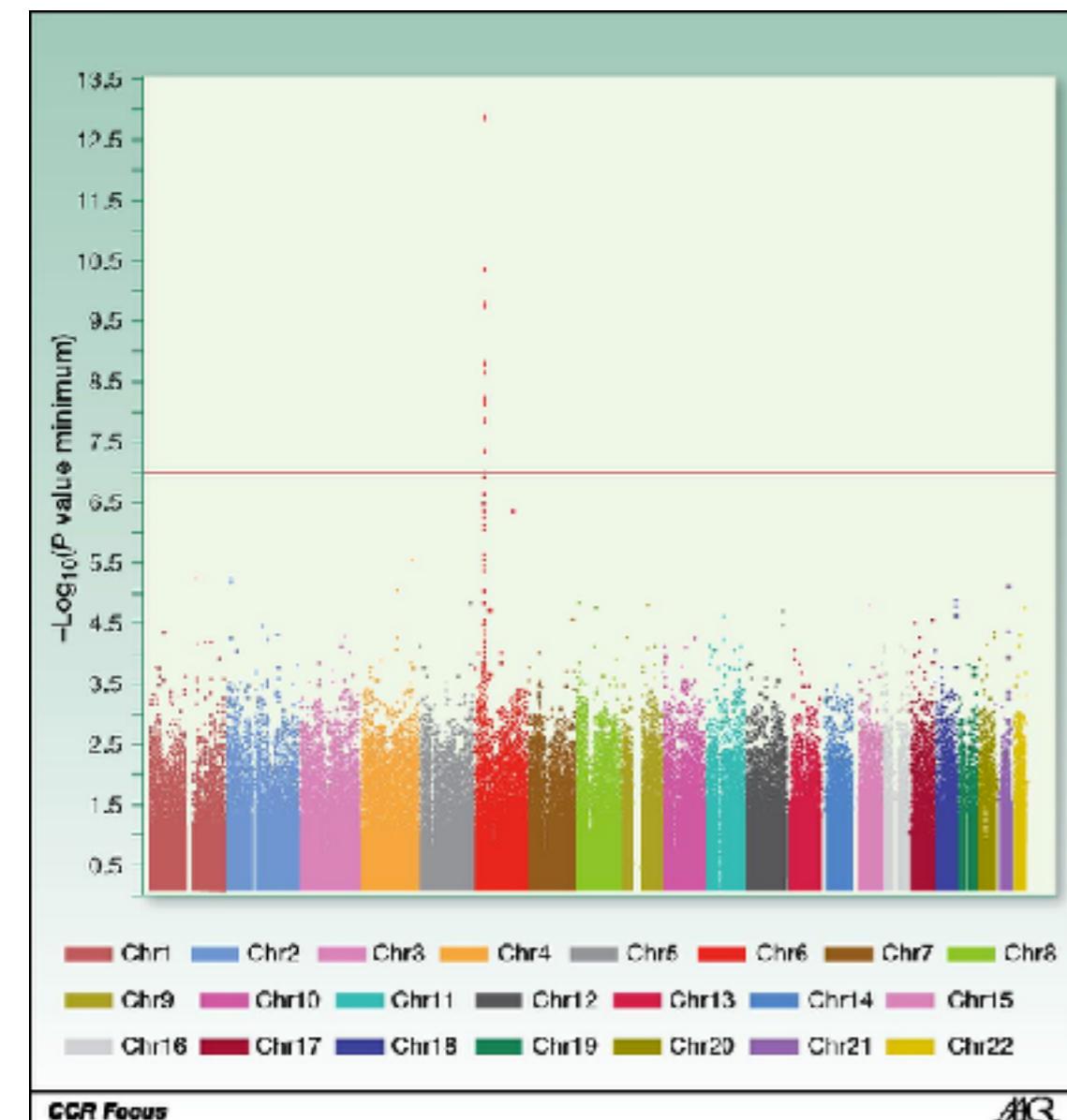
**Circos**

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

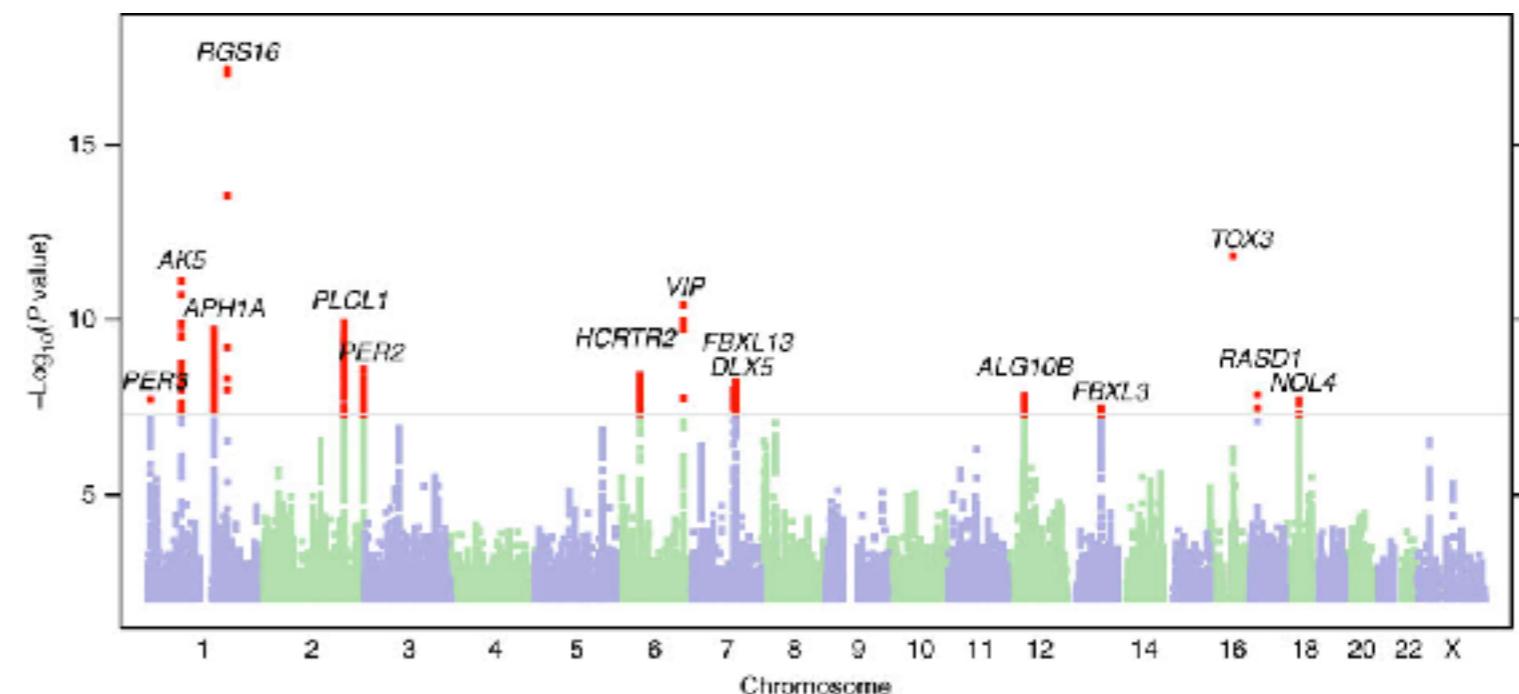
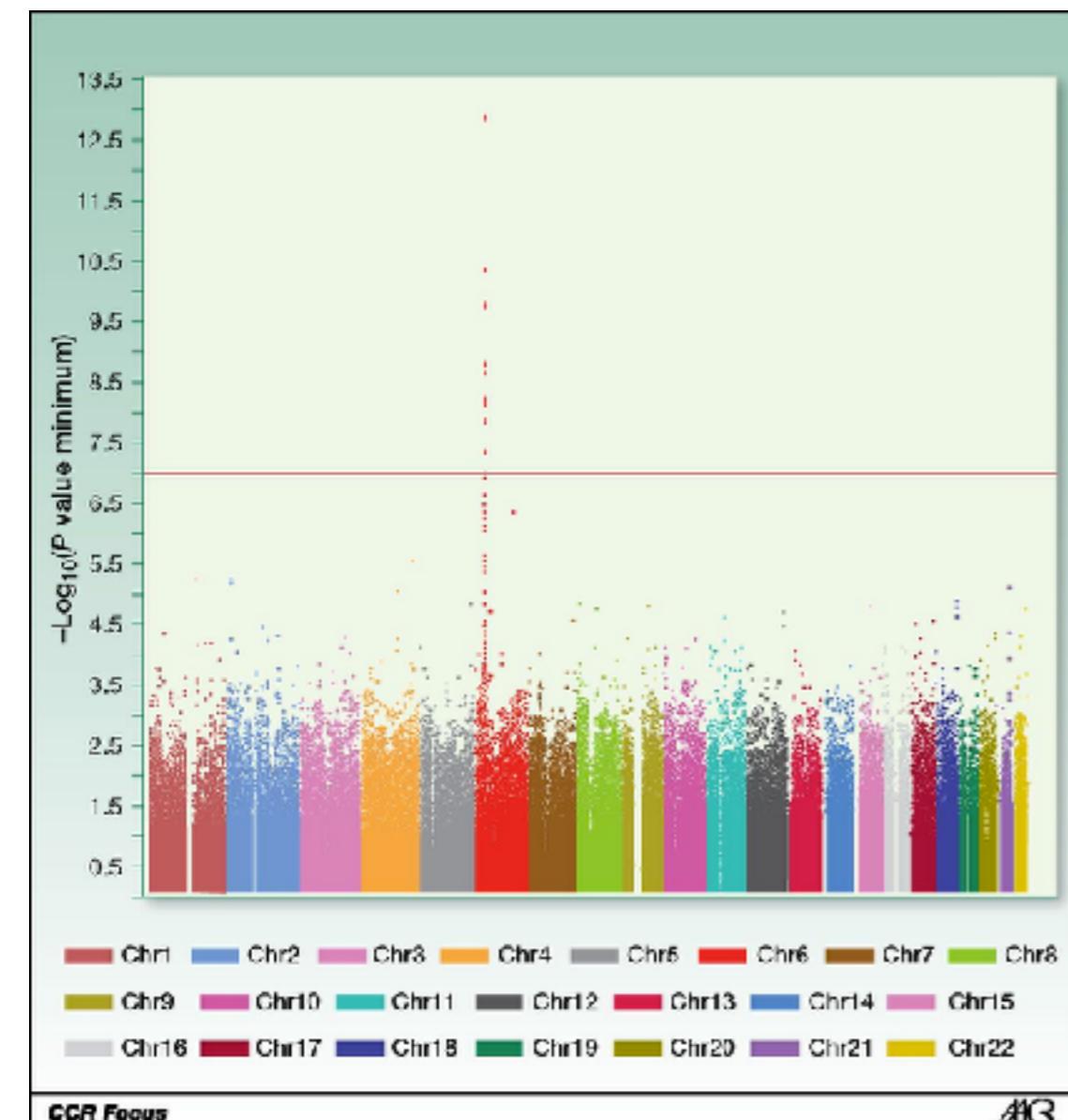
# 5 Principles for Improving Figures

- 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 100s of observations



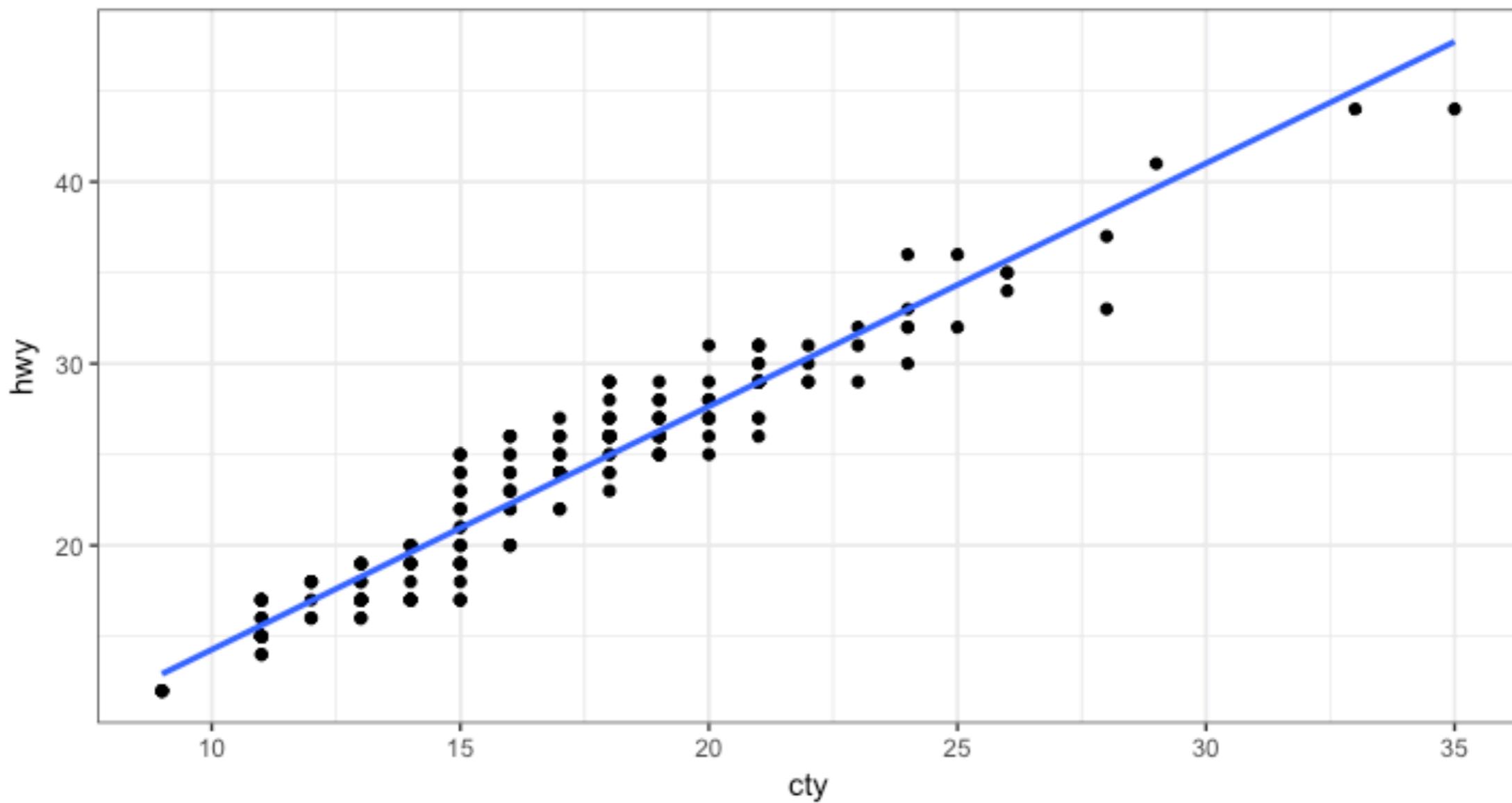
# Manhattan plots communicate 100s of observations



# Know your plots – correlations

Scatterplot with overlapping points

mpg: city vs highway mileage

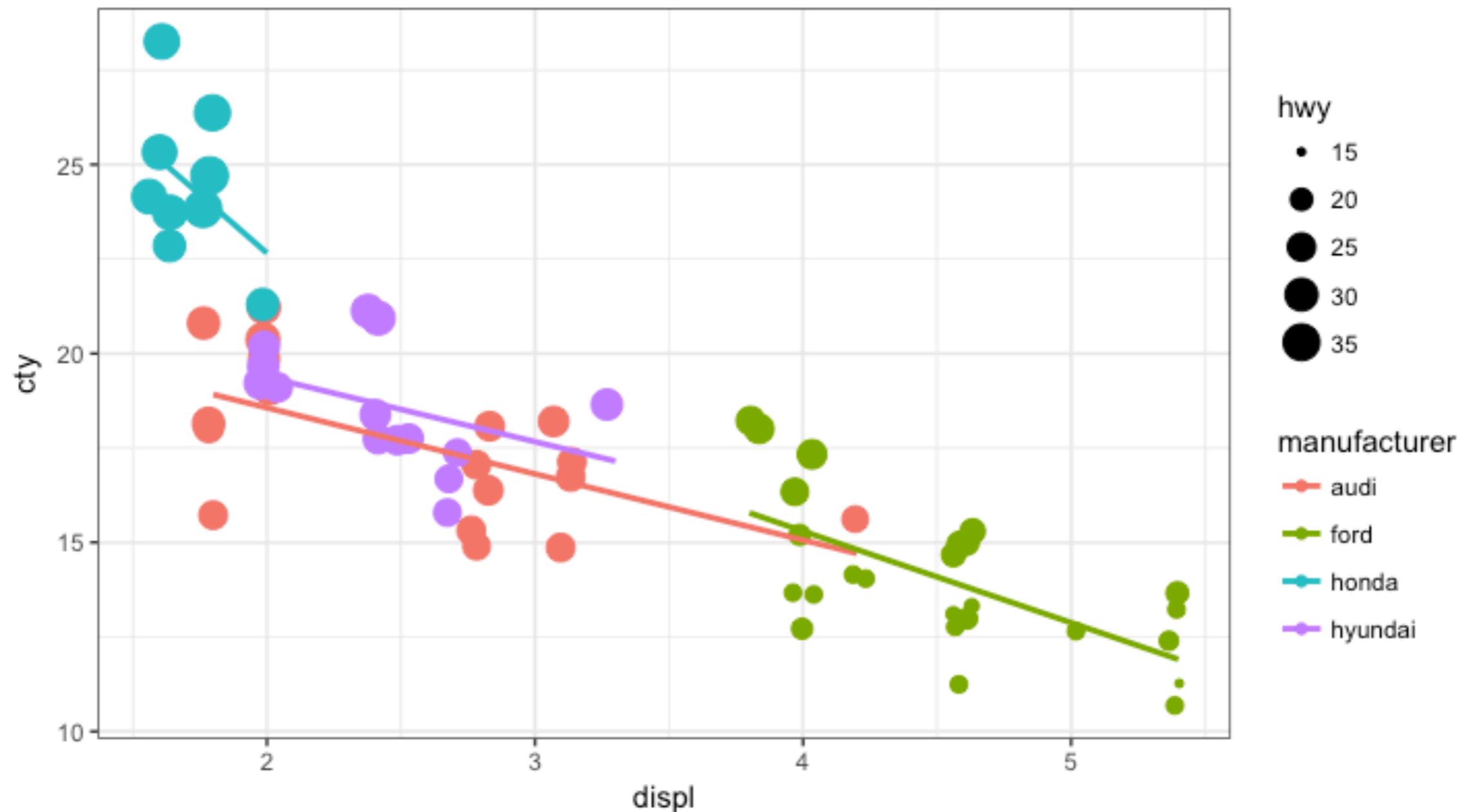


Source: midwest

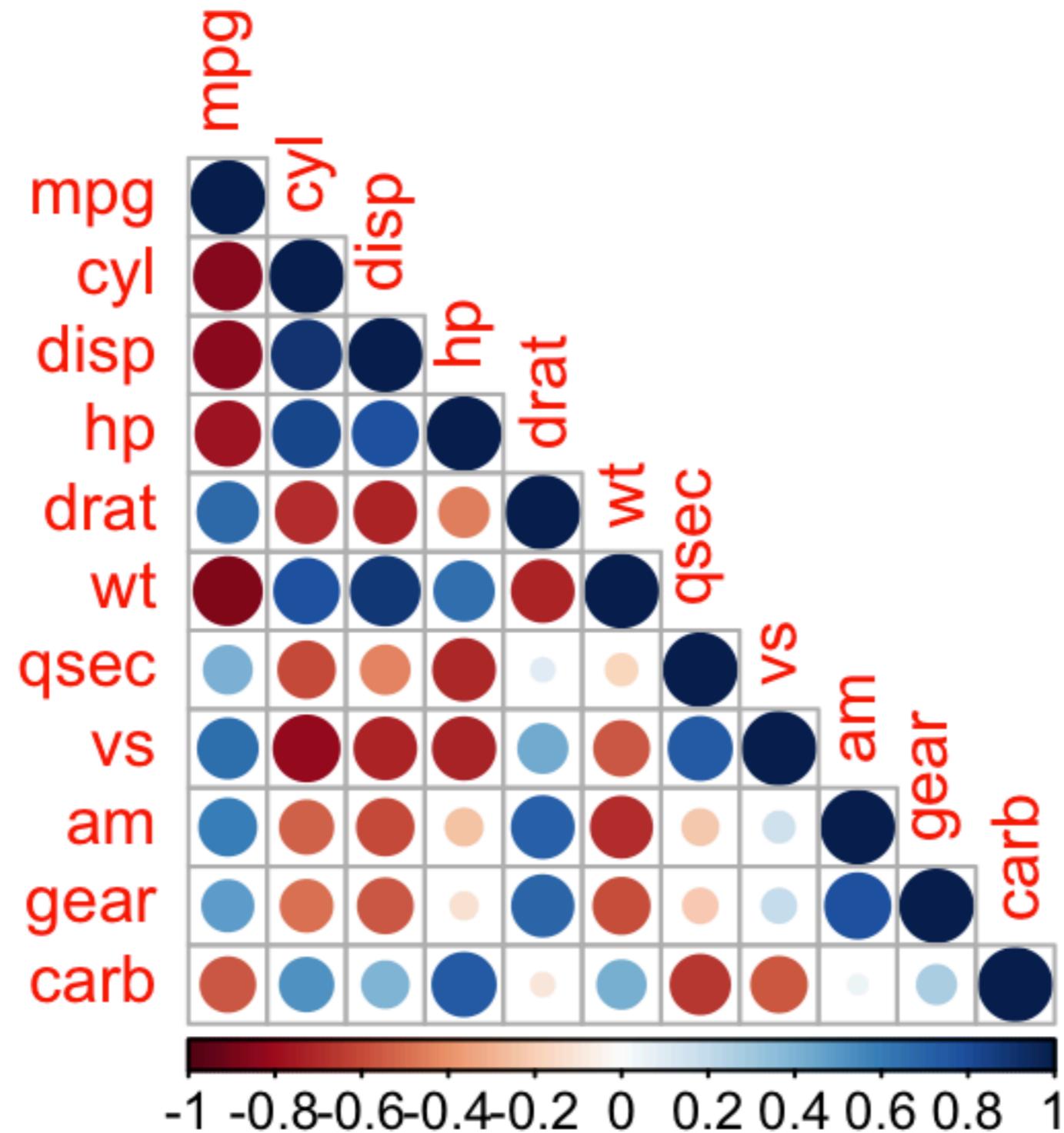
# Know your plots – correlations

## Bubble chart

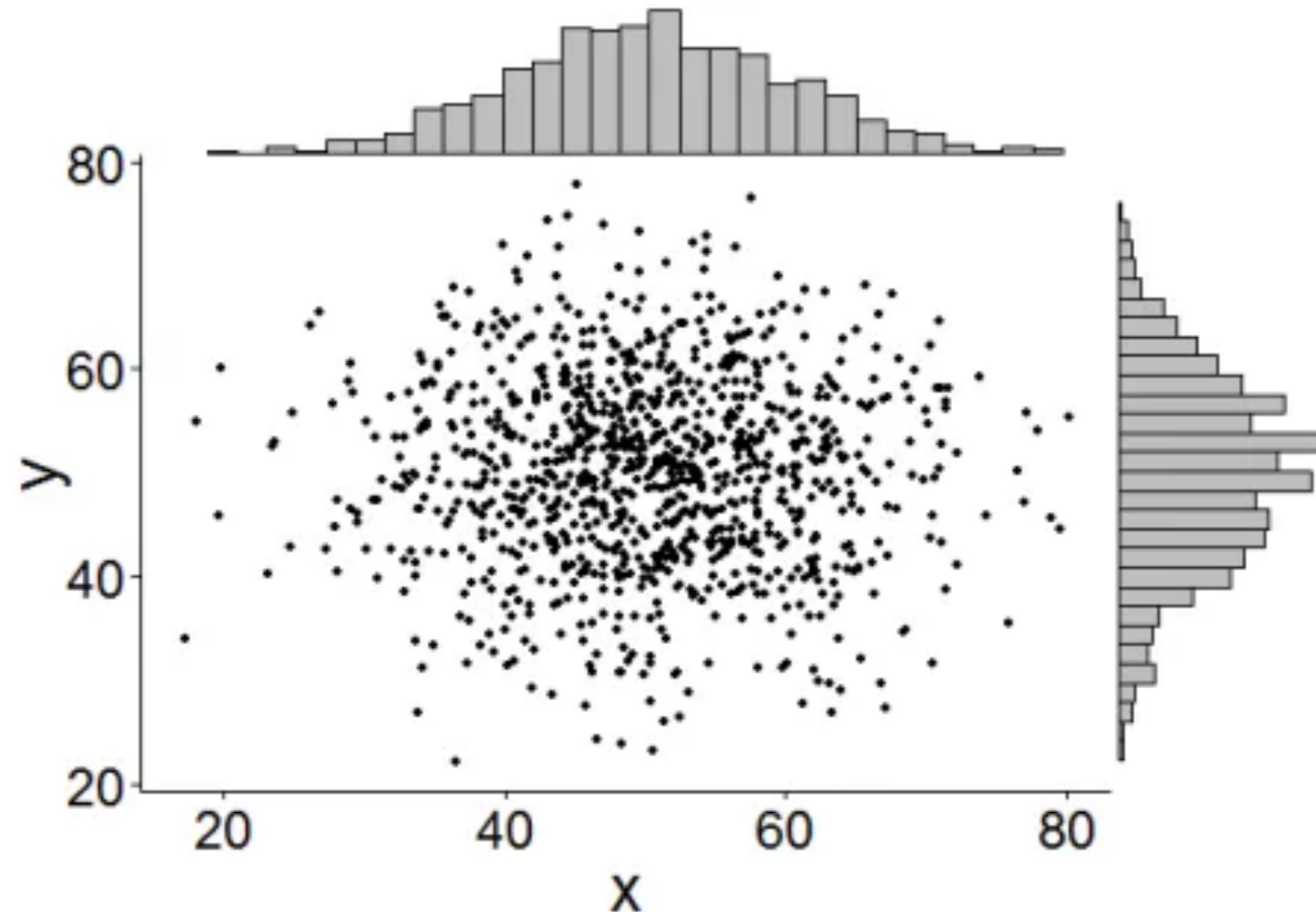
## mpg: Displacement vs City Mileage



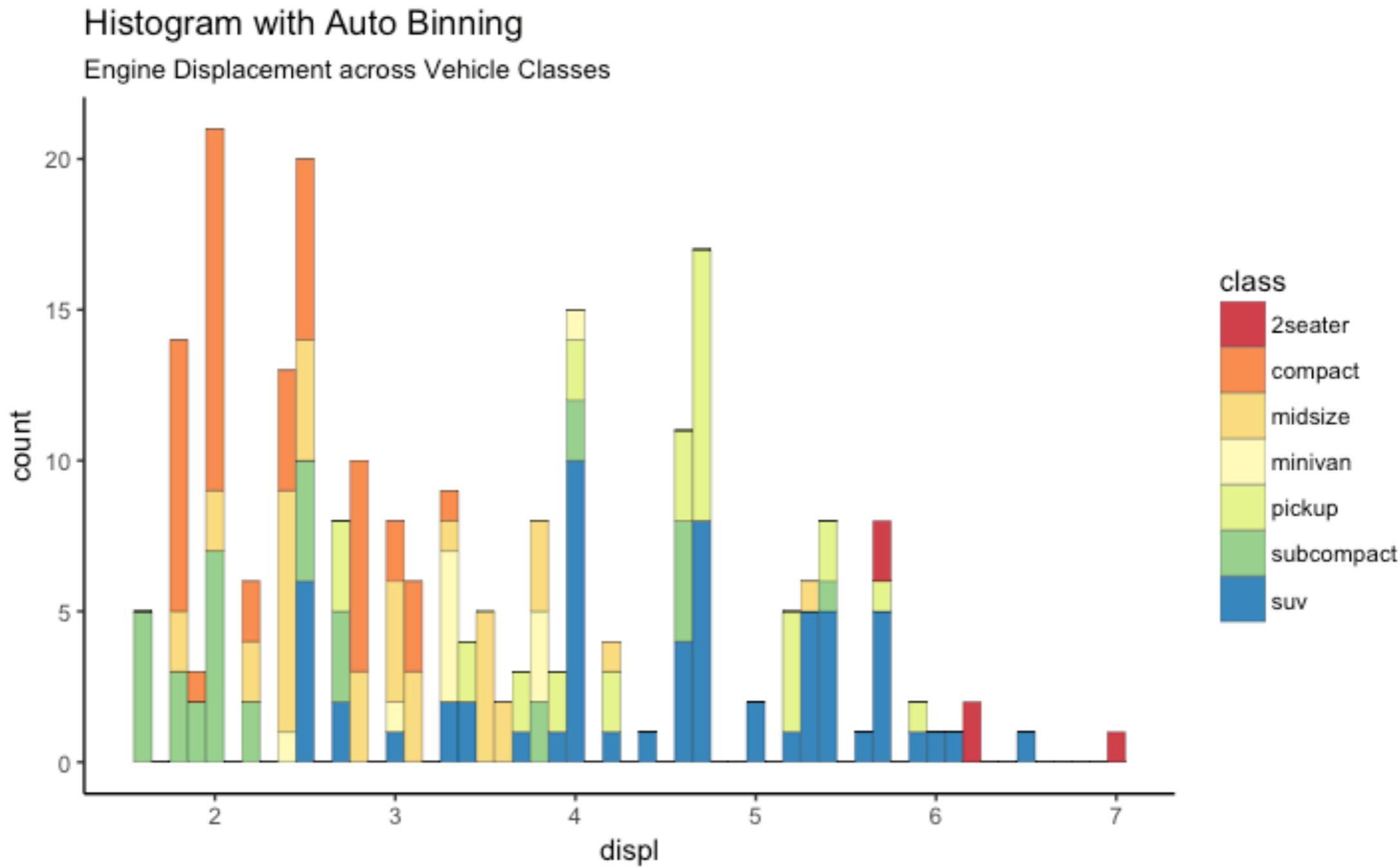
# Know your plots – correlations



# Know your plots – distributions



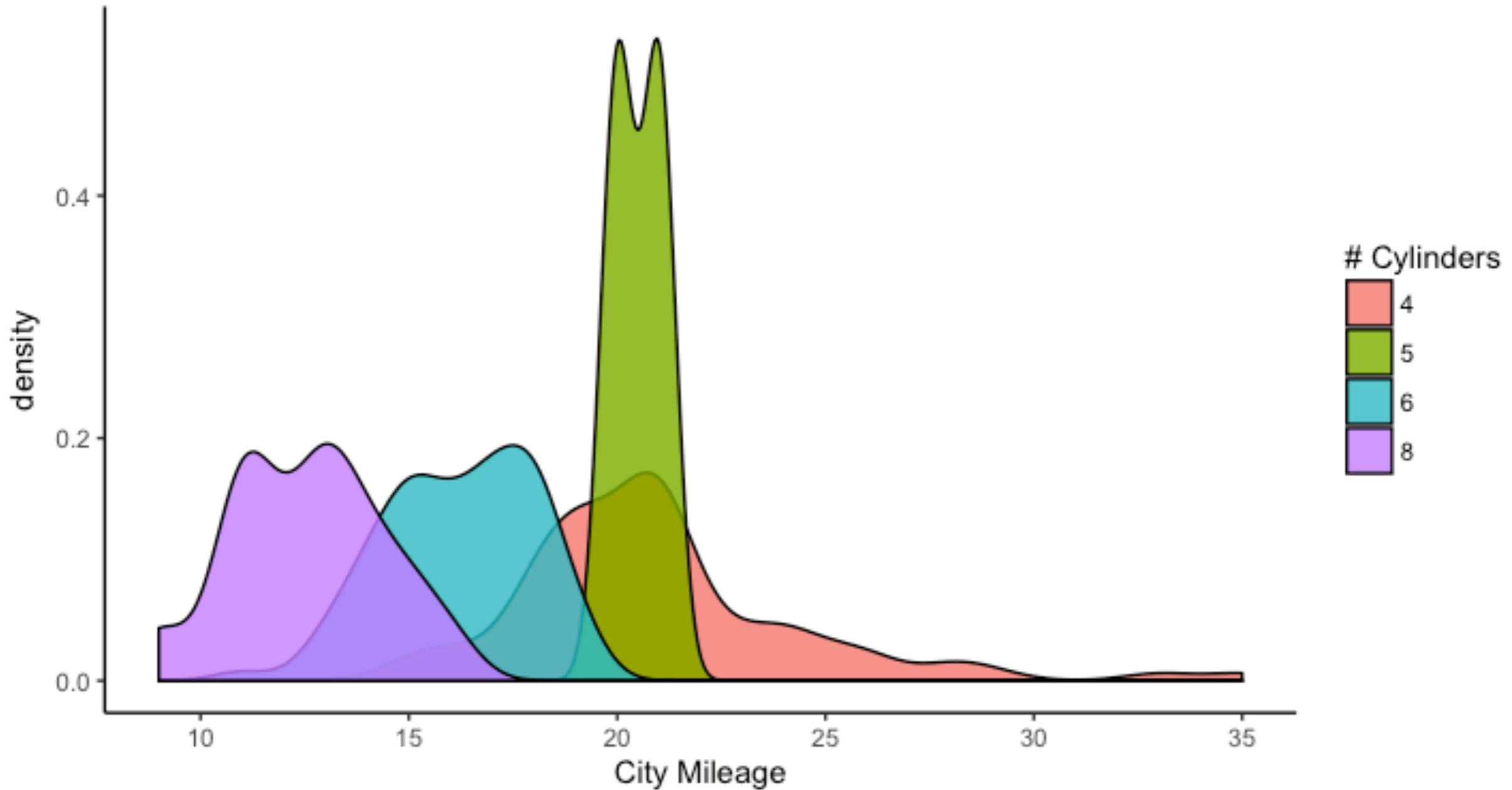
# Know your plots – distributions



# Know your plots – distributions

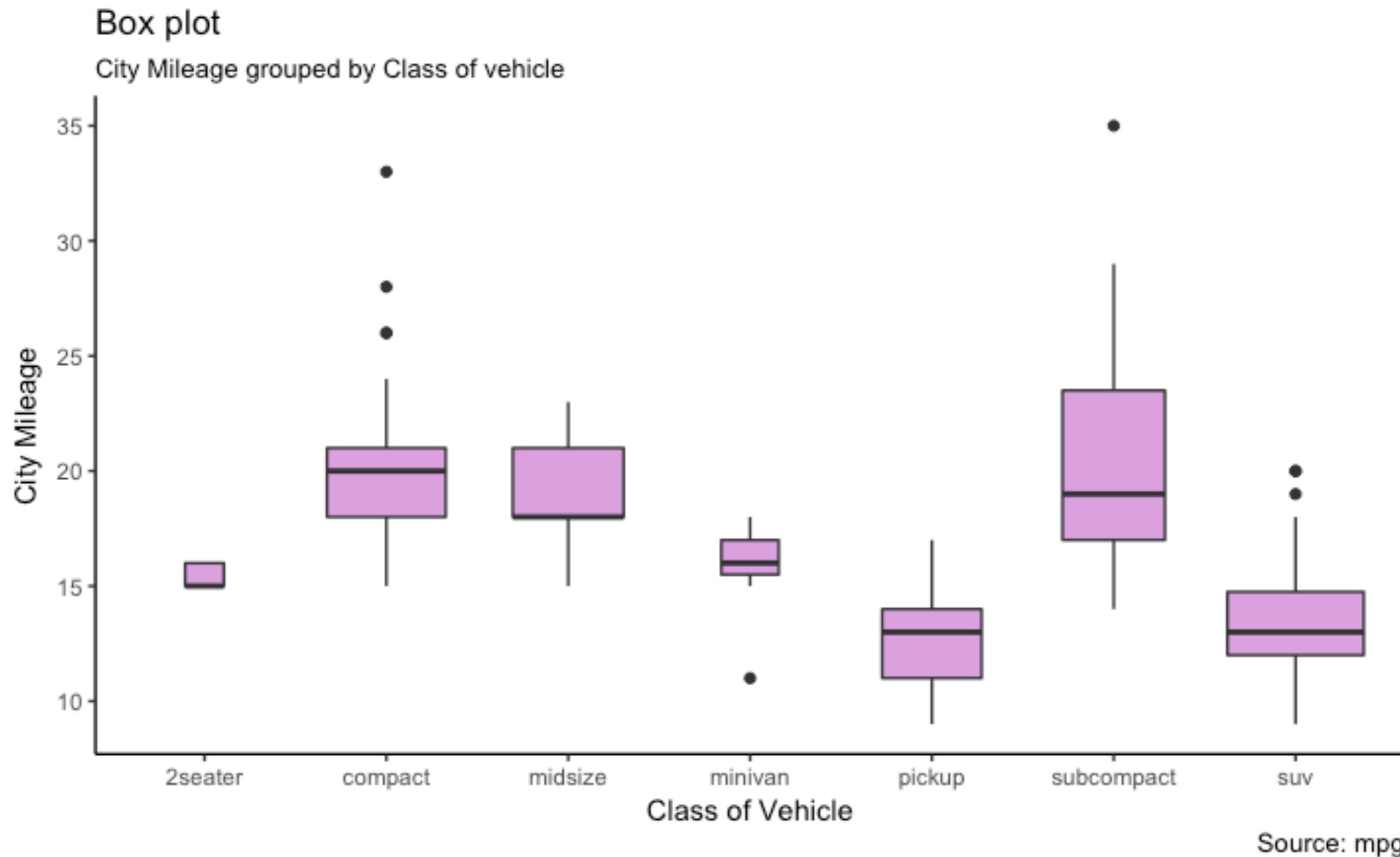
## Density plot

City Mileage Grouped by Number of cylinders

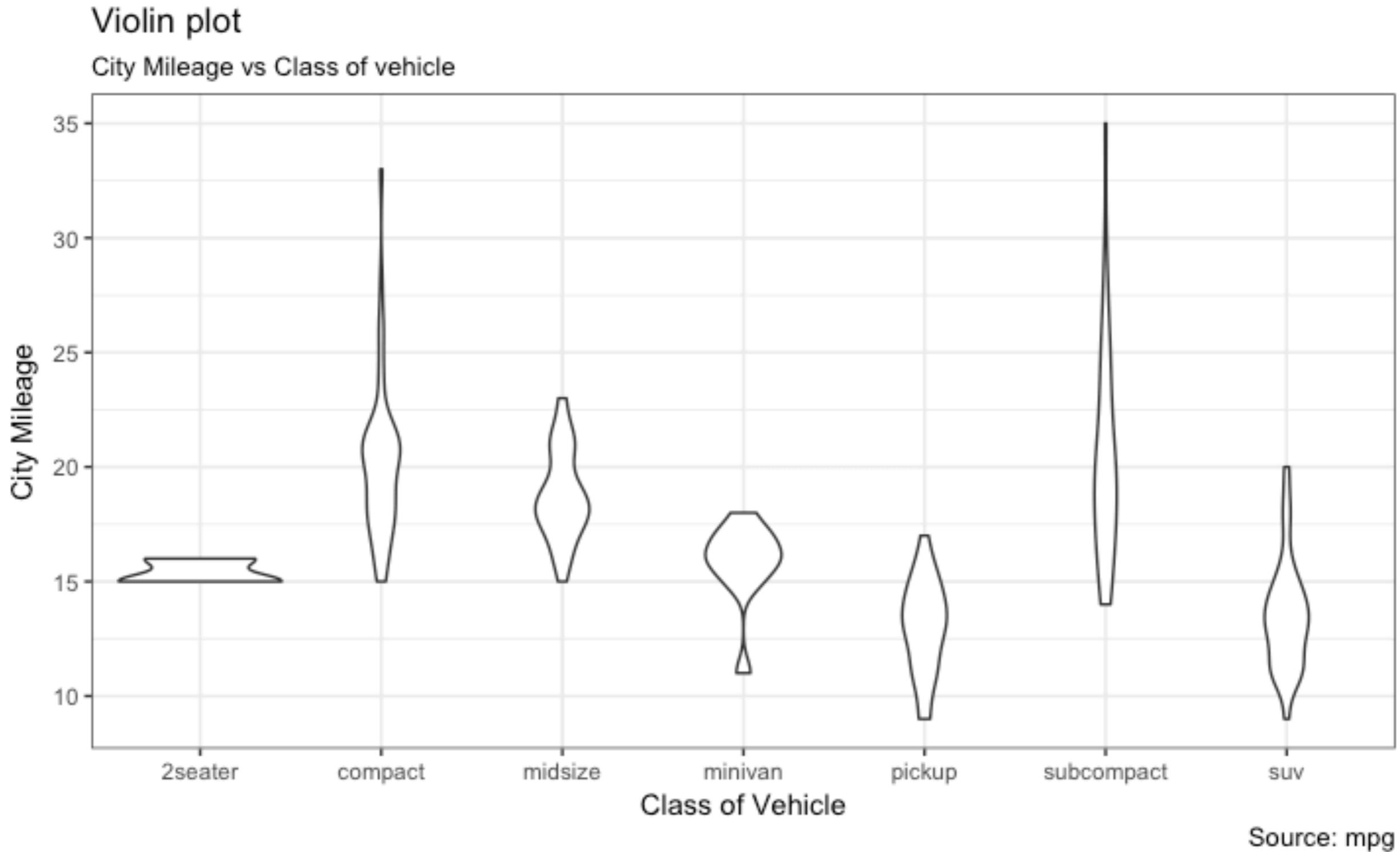


Source: mpg

# Know your plots – distributions

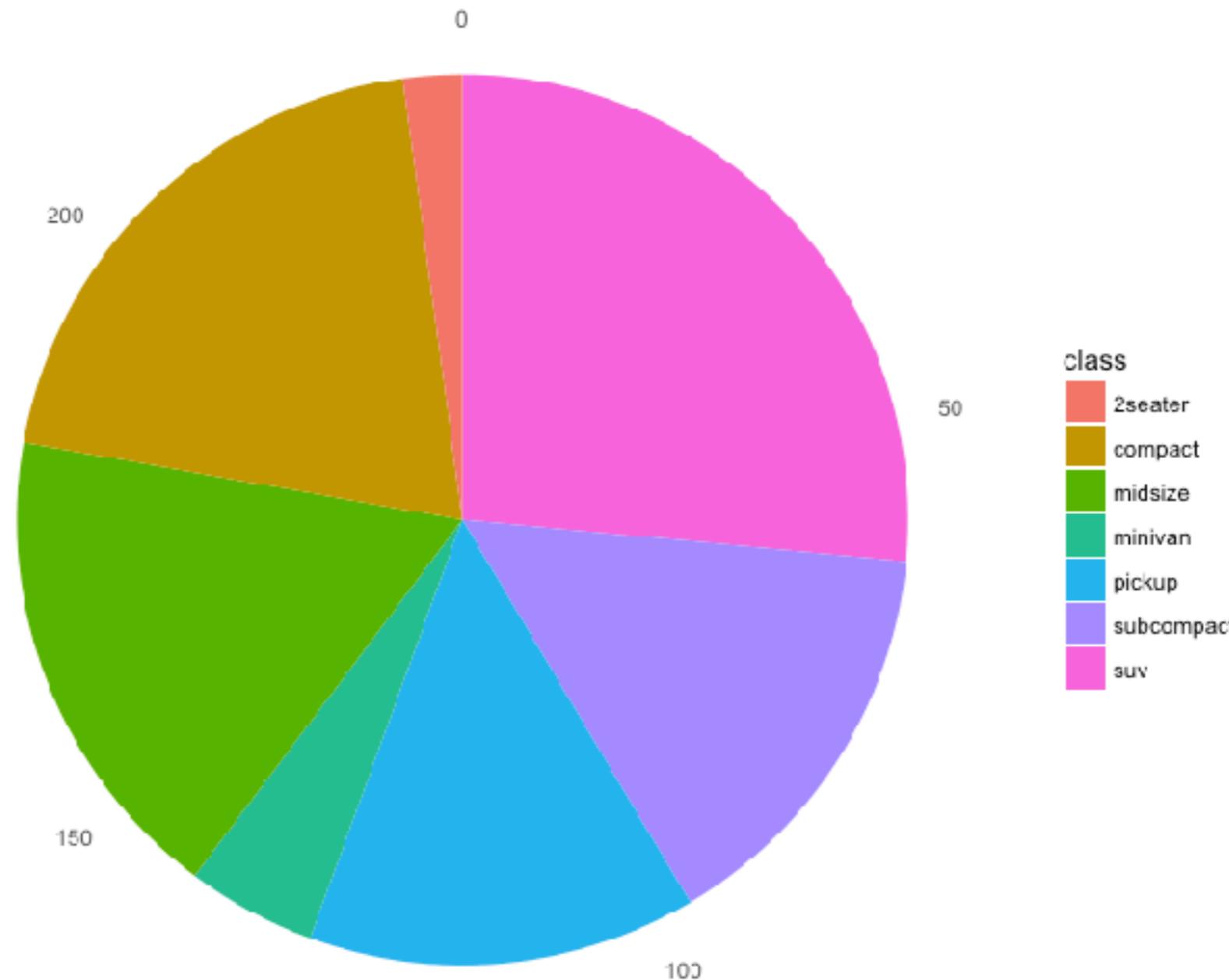


# Know your plots – distributions



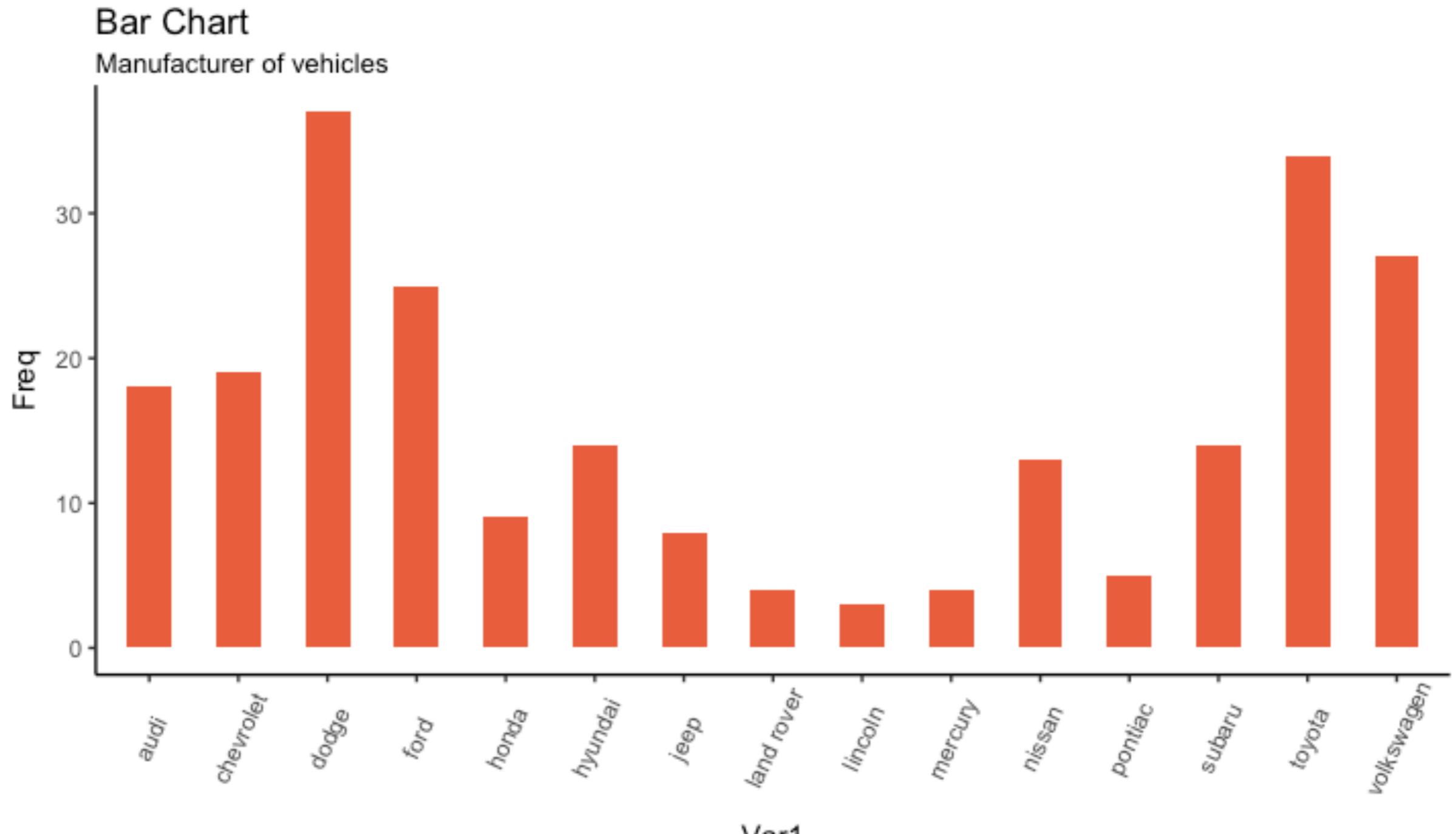
# Know your plots – composition

Pie Chart of class

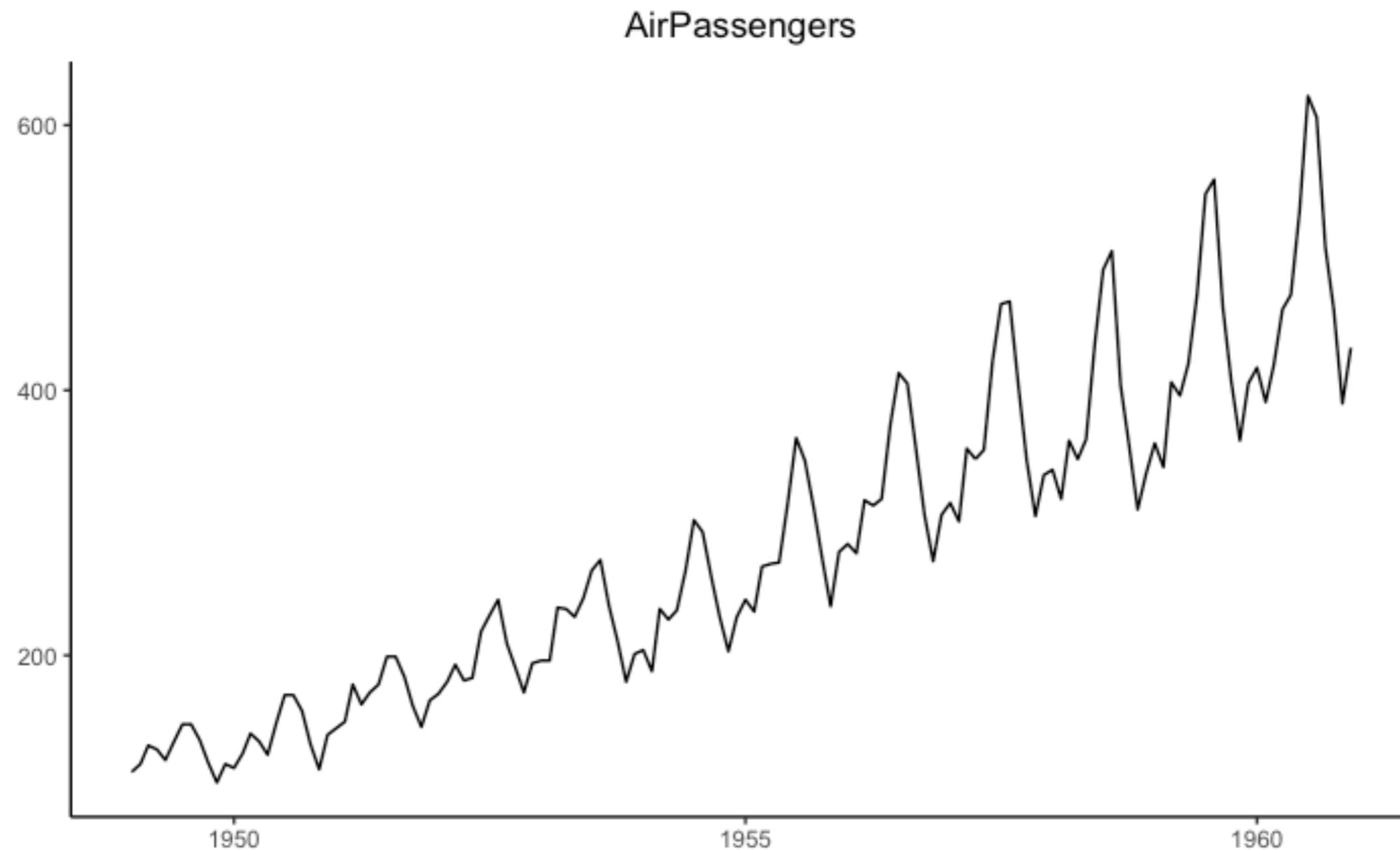


Source: mpg

# Know your plots – composition



# Know your plots – change

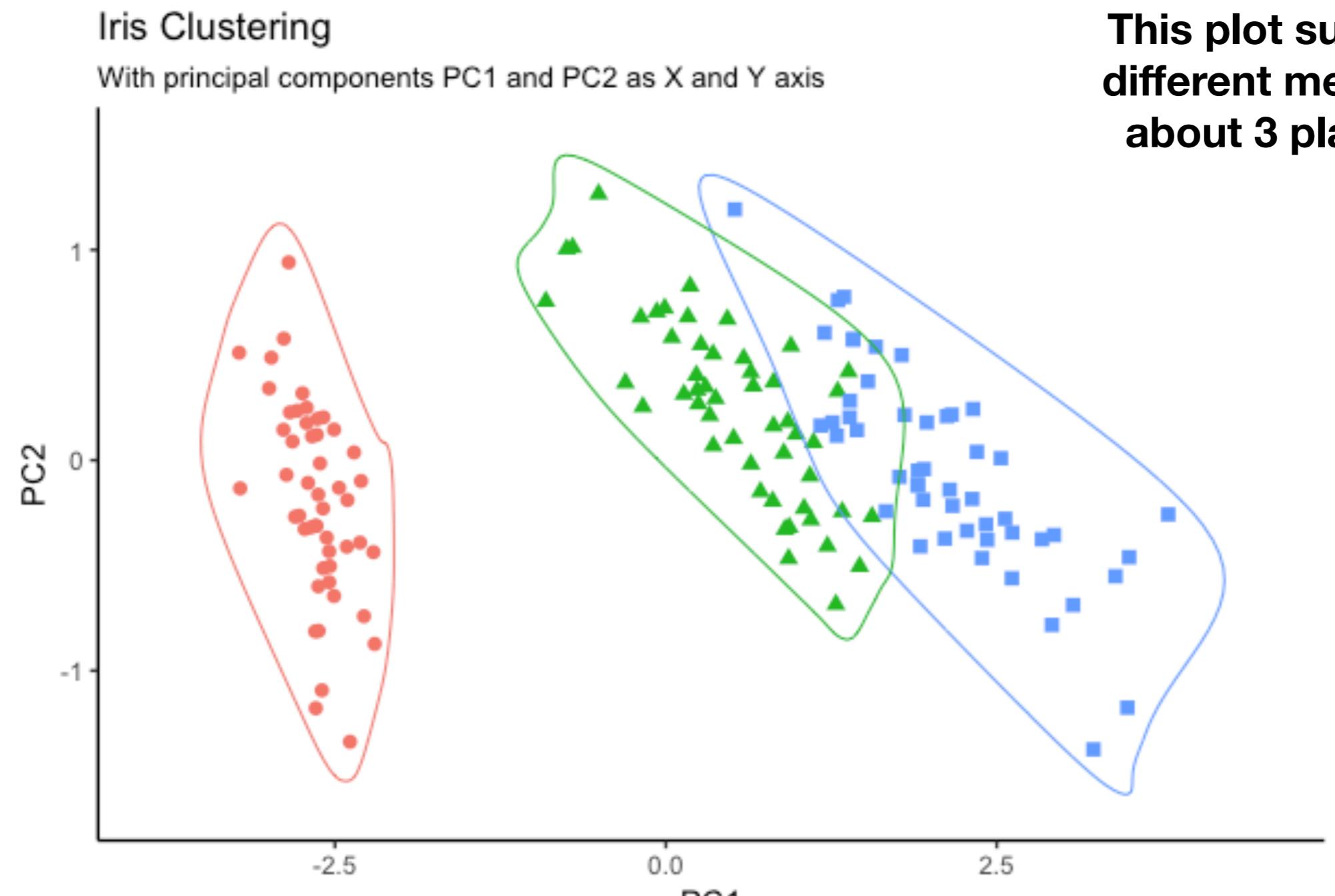


# Know your plots – multi-dimensional data

```
> iris
```

	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
...	...	...	...	...	...

# Know your plots – multi-dimensional data

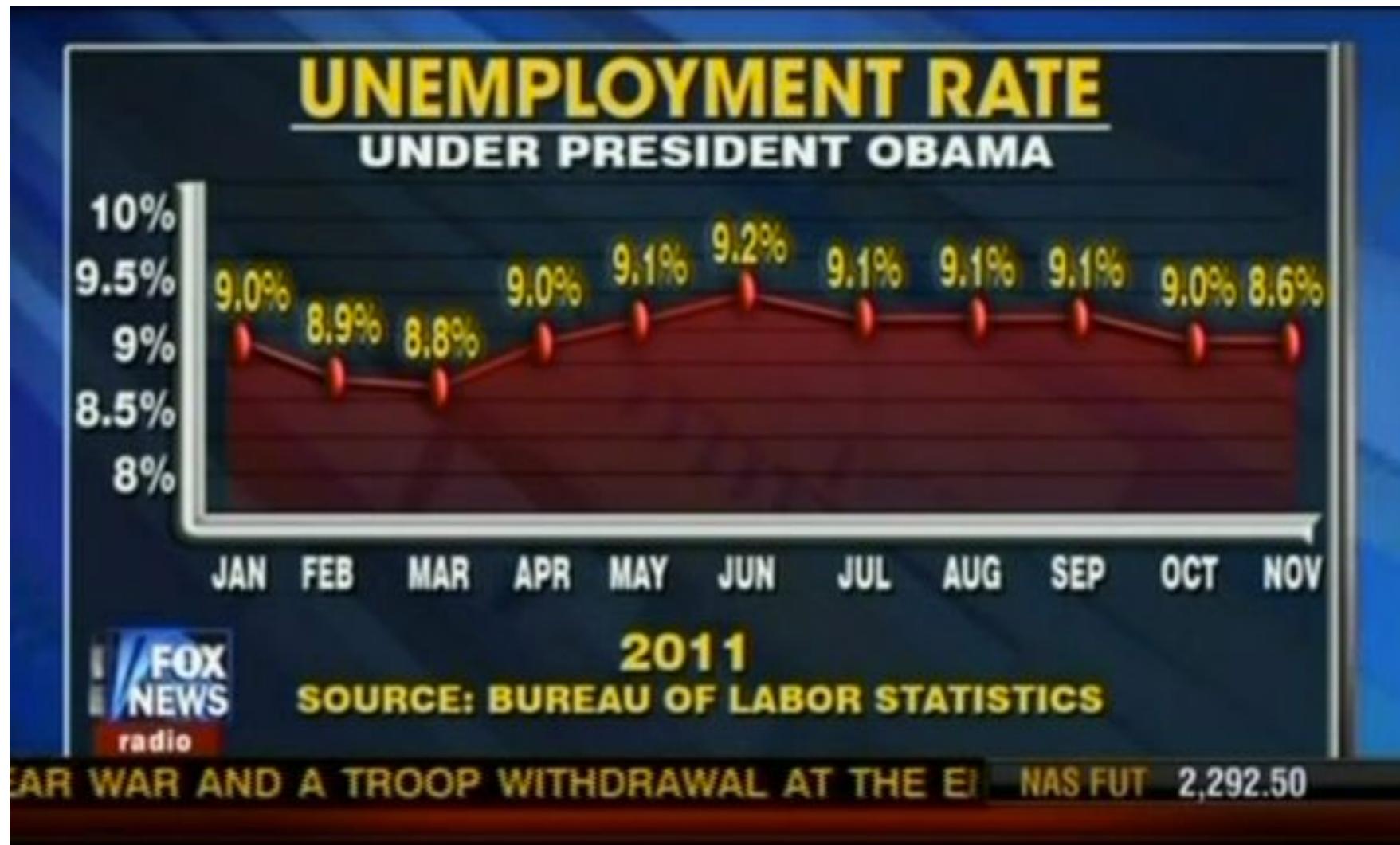


**This plot summarizes 4 different measurements about 3 plant species**

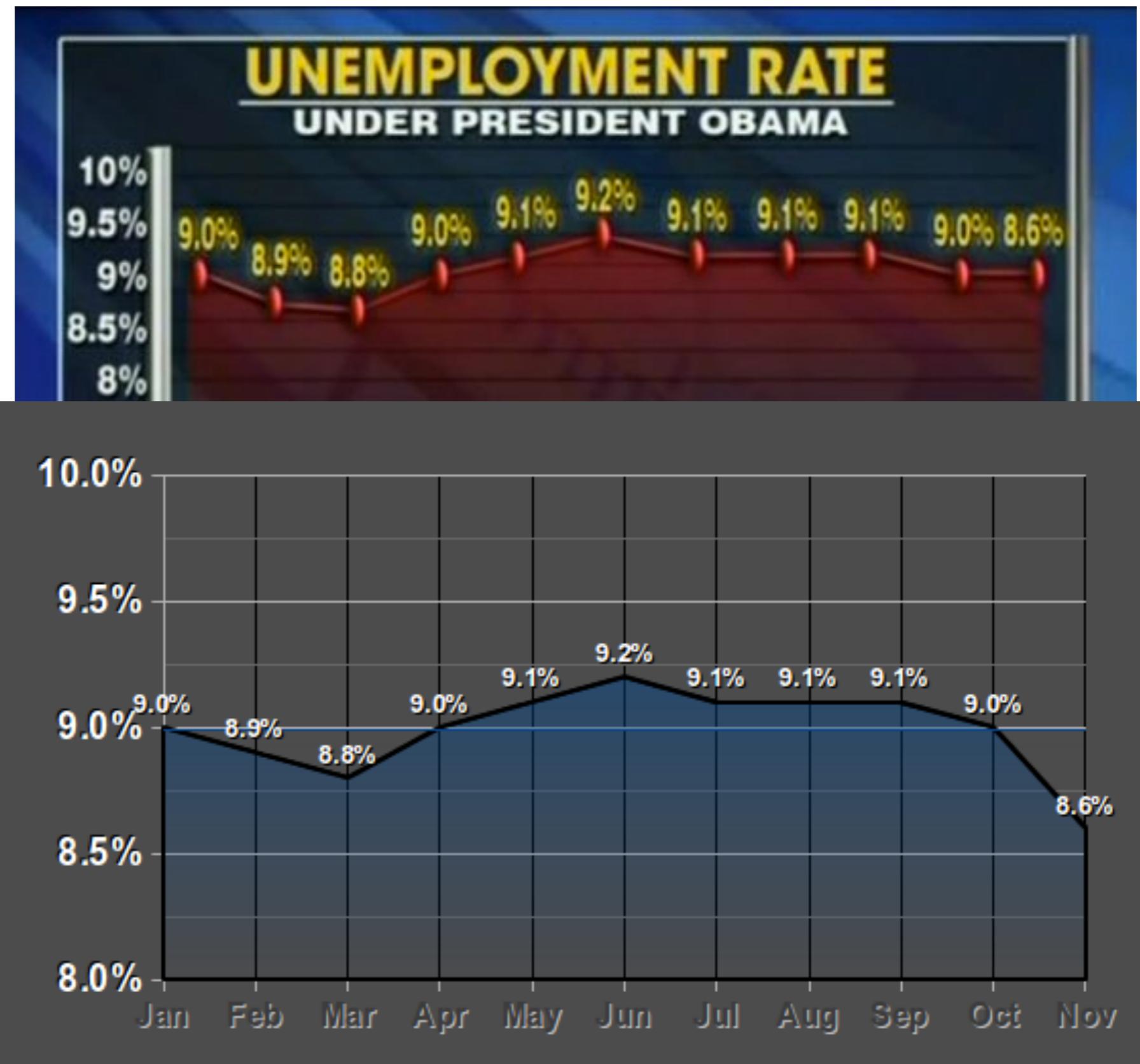
# 5 Principles for Improving Figures

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

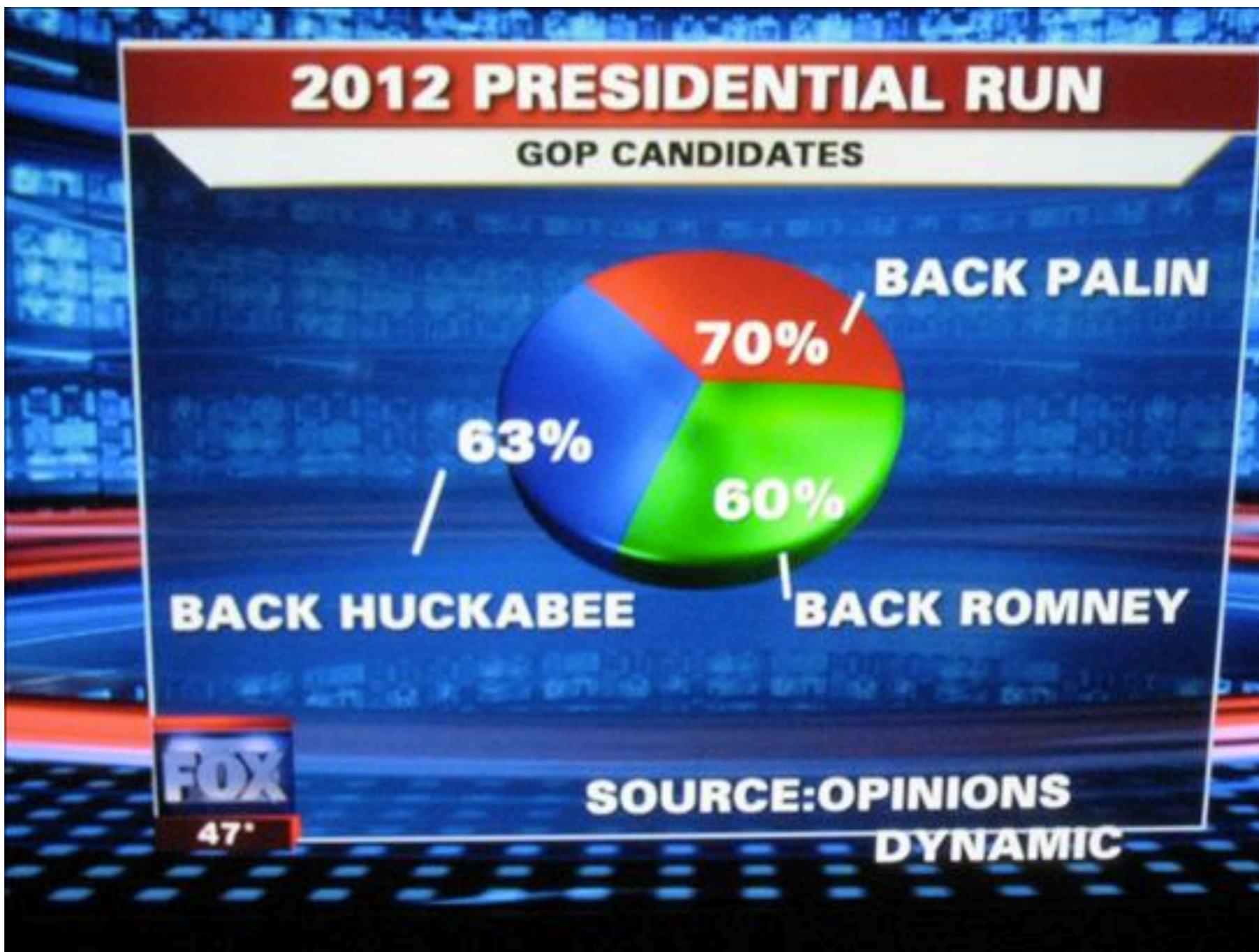
# Don't mislead the viewer



# Don't mislead the viewer

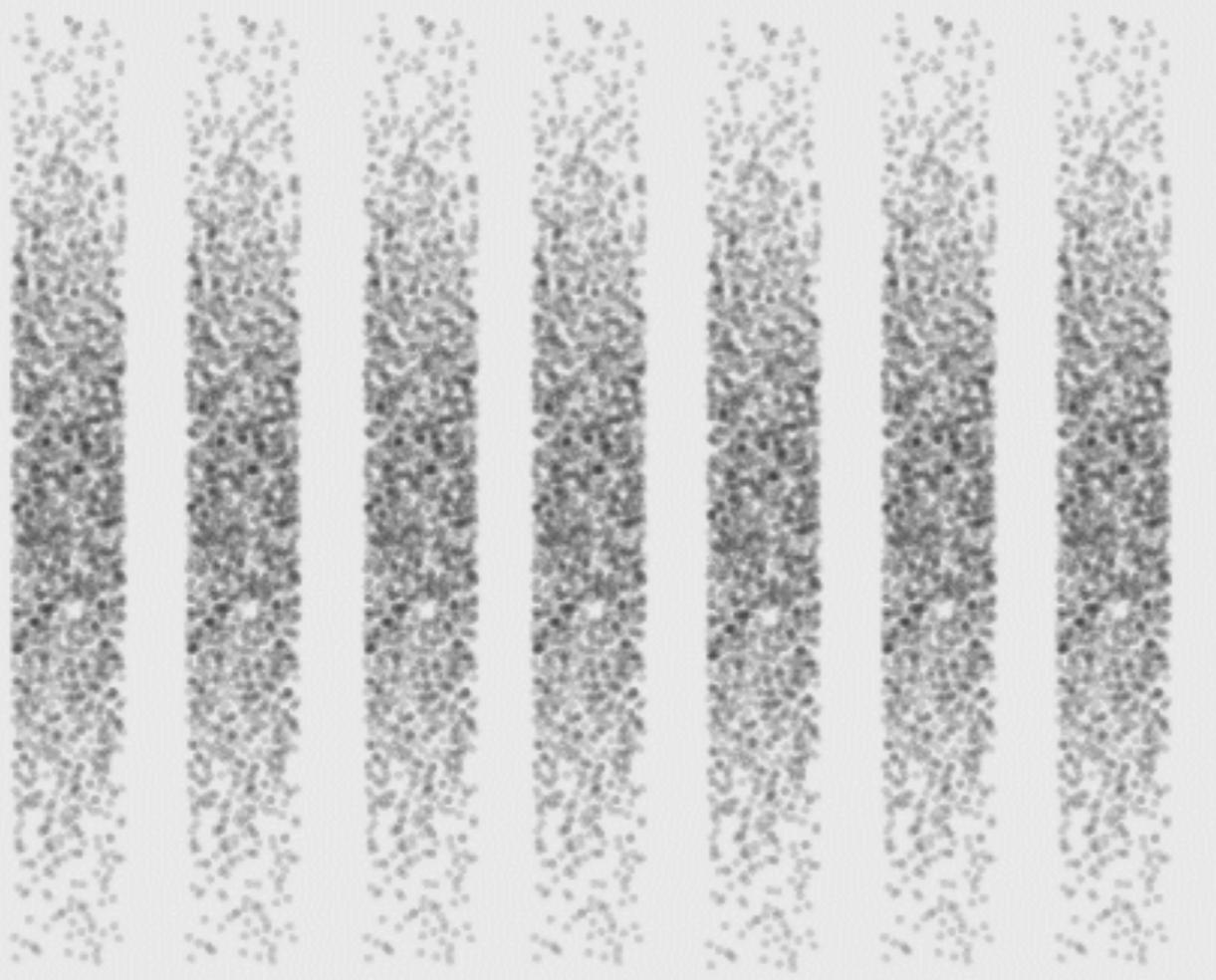


# Don't mislead the viewer

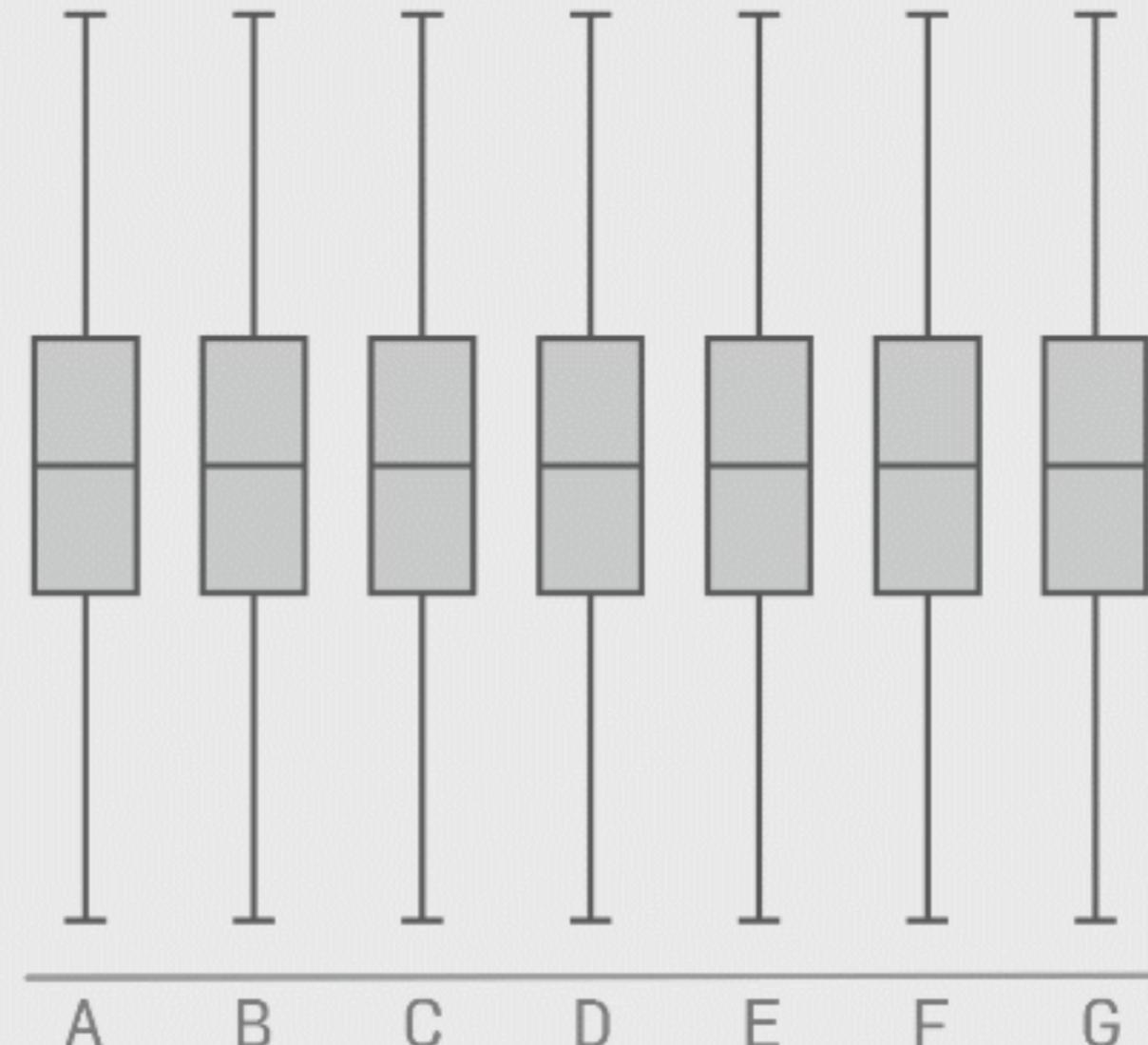


# Always be honest

**Raw Data**



**Box-plot of the Data**



# 5 Principles for Improving Figures

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

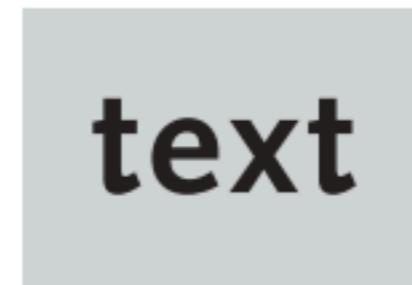
# Utilize contrast appropriately

Adequate readability due to high value contrast



# Utilize contrast appropriately

Adequate readability due to high value contrast



Inadequate readability due to low value contrast



# Utilize contrast appropriately

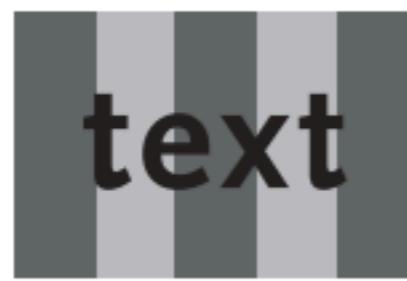
Adequate readability due to high value contrast



Inadequate readability due to low value contrast



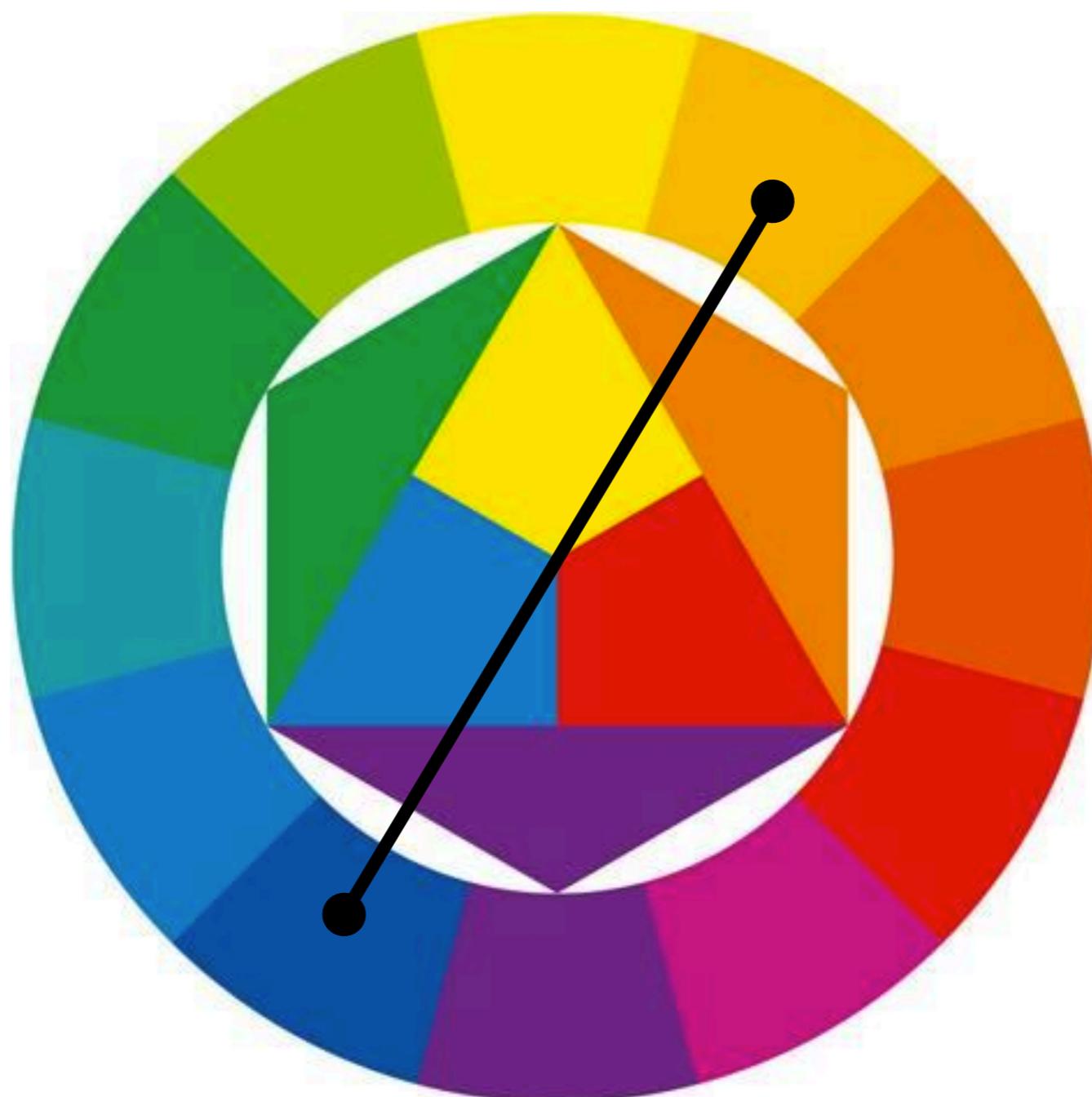
Inadequate readability due to patterned background



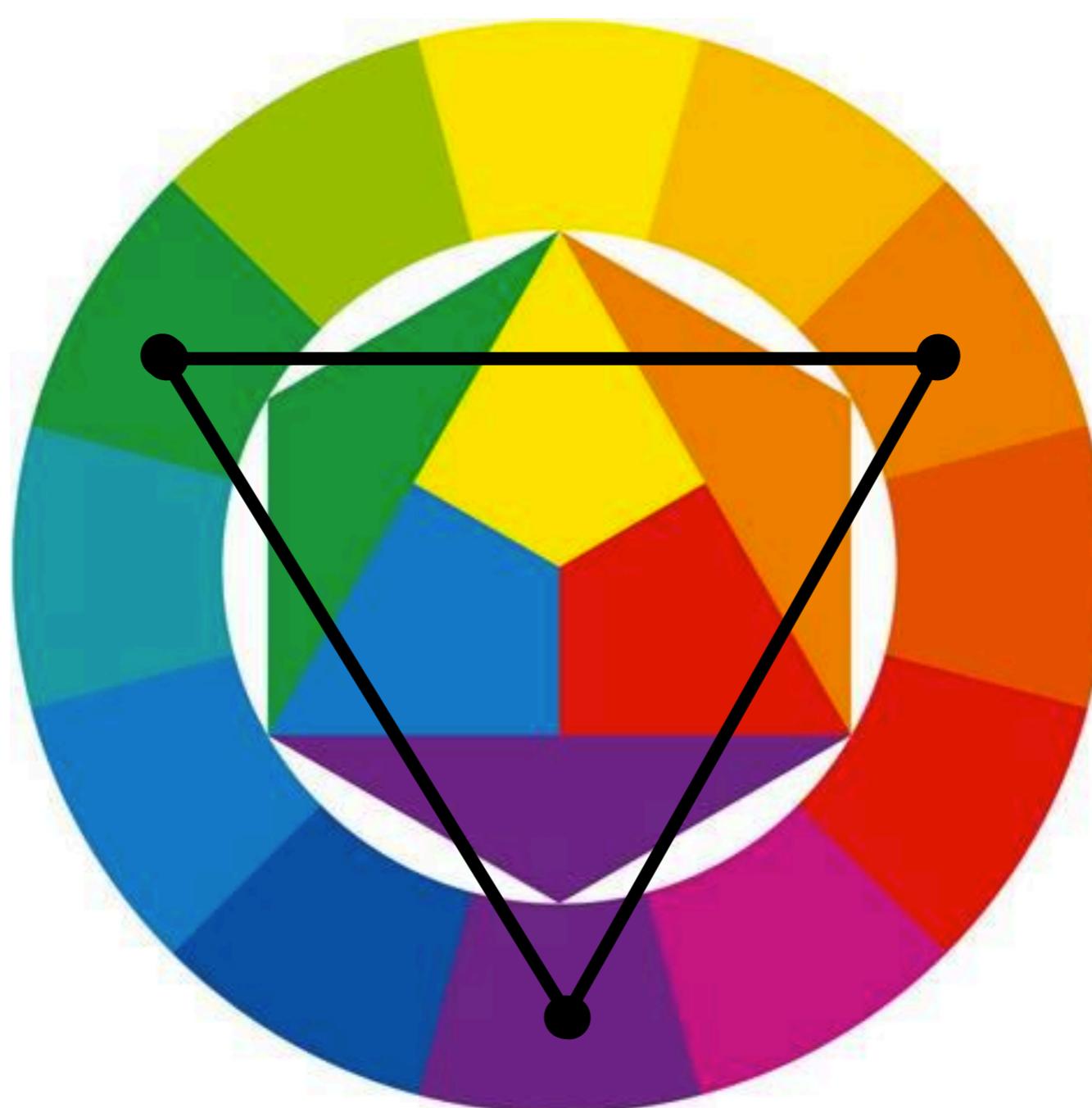
# Using color theory to pick color combinations



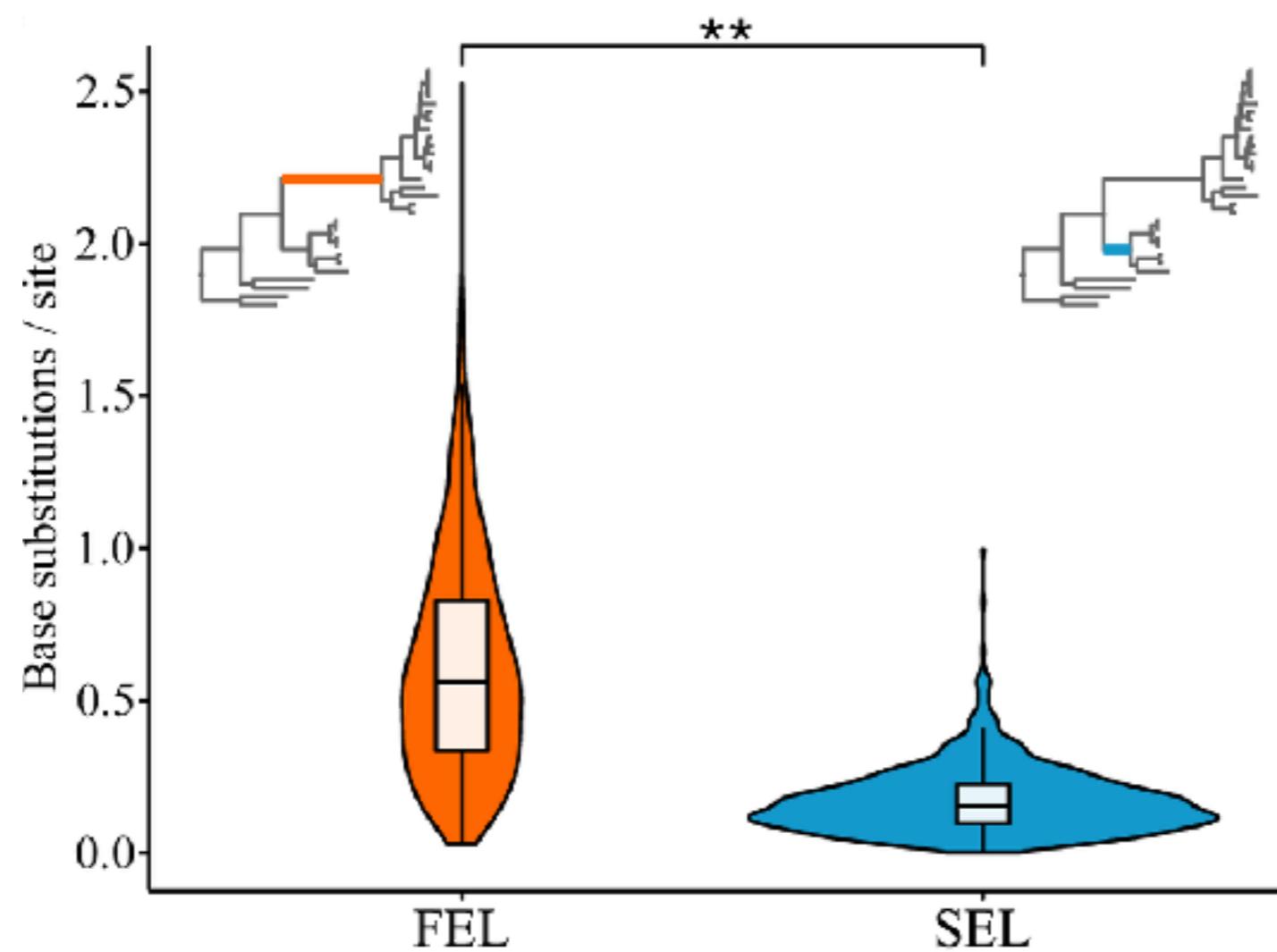
# Using color theory to pick color combinations



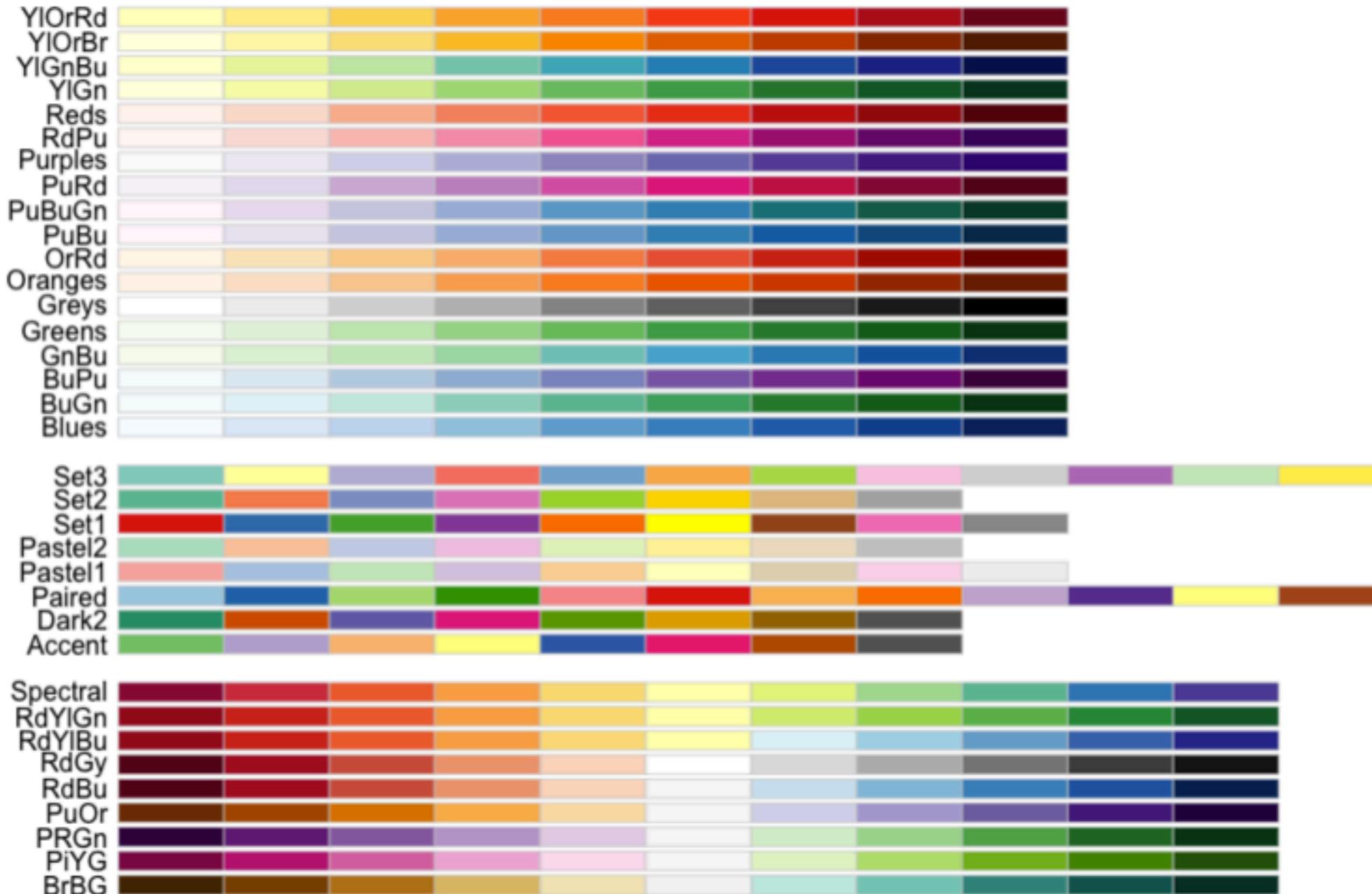
# Using color theory to pick color combinations



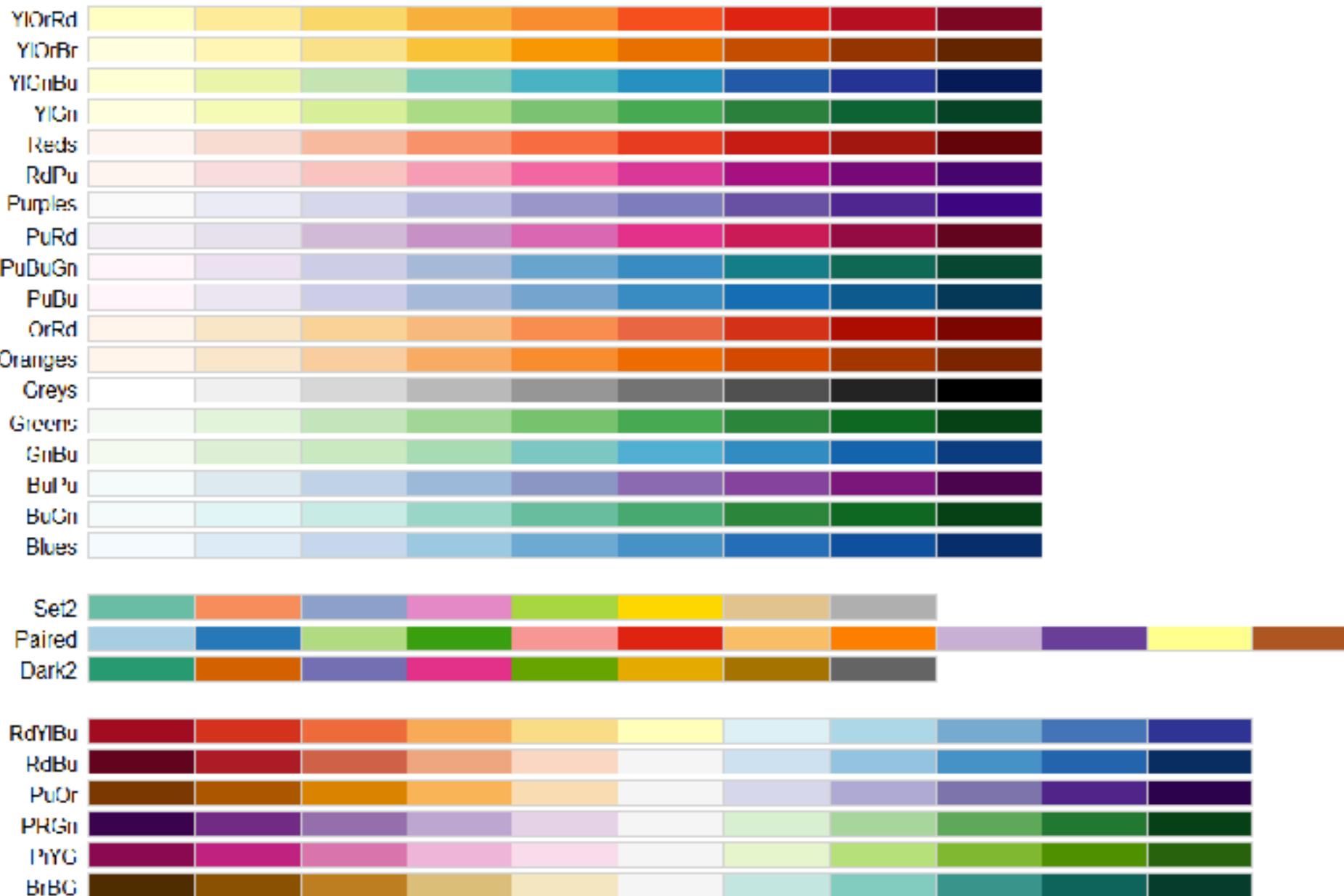
# Combining multiple figures can be ‘more’ honest



# Picking from palettes



# Ensure figures are color blind friendly



# 5 Principles for Improving 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

**pepto**  
BISMOL

[facebook.com/peptobismol](http://facebook.com/peptobismol)

**Who doesn't love grilling?**  
The most popular grilling occasions are Fourth of July, Memorial Day and Labor Day.  
45% of U.S. adults use the grill at least 12 times per year.  
82% of U.S. households own a grill or smoker.  
15,700,000 Grills were shipped in 2009.  
The most popular foods for grilling are:

- ④ CHICKEN
- ③ HOT DOGS
- ② STEAK

When we get together, we tend to eat more.<sup>4</sup>  
Meals eaten with one other person were 35% larger than those eaten alone.

**1 BURGERS**

**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 (or many) more 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>1</sup>

**TEXAS** 1M  
**MINNESOTA** 1M  
**ARIZONA** 1M  
**NEW ENGLAND** 1M  
**NEW YORK** 1M  
**IOWA** 1M

**Major League Eating champions are eating more these days.**  
The number of hot dogs eaten at the Nathan's Famous Hot Dog Eating Contest has been increasing since its inception in 1916.

**More Than 59 Million**  
Americans will attend a festival or fair this year.<sup>2</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
- Fajitas
- Chocolate chip cookies
- Ice cold soda
- Draft beer

**SOURCES**

- 1. USA Today, National Cultural Traveler, 2002 edition.
- 2. Department of Marketing and Nutritional Sciences, University of Illinois at Urbana-Champaign.
- 3. Board of Pepto & Barbecue Association.

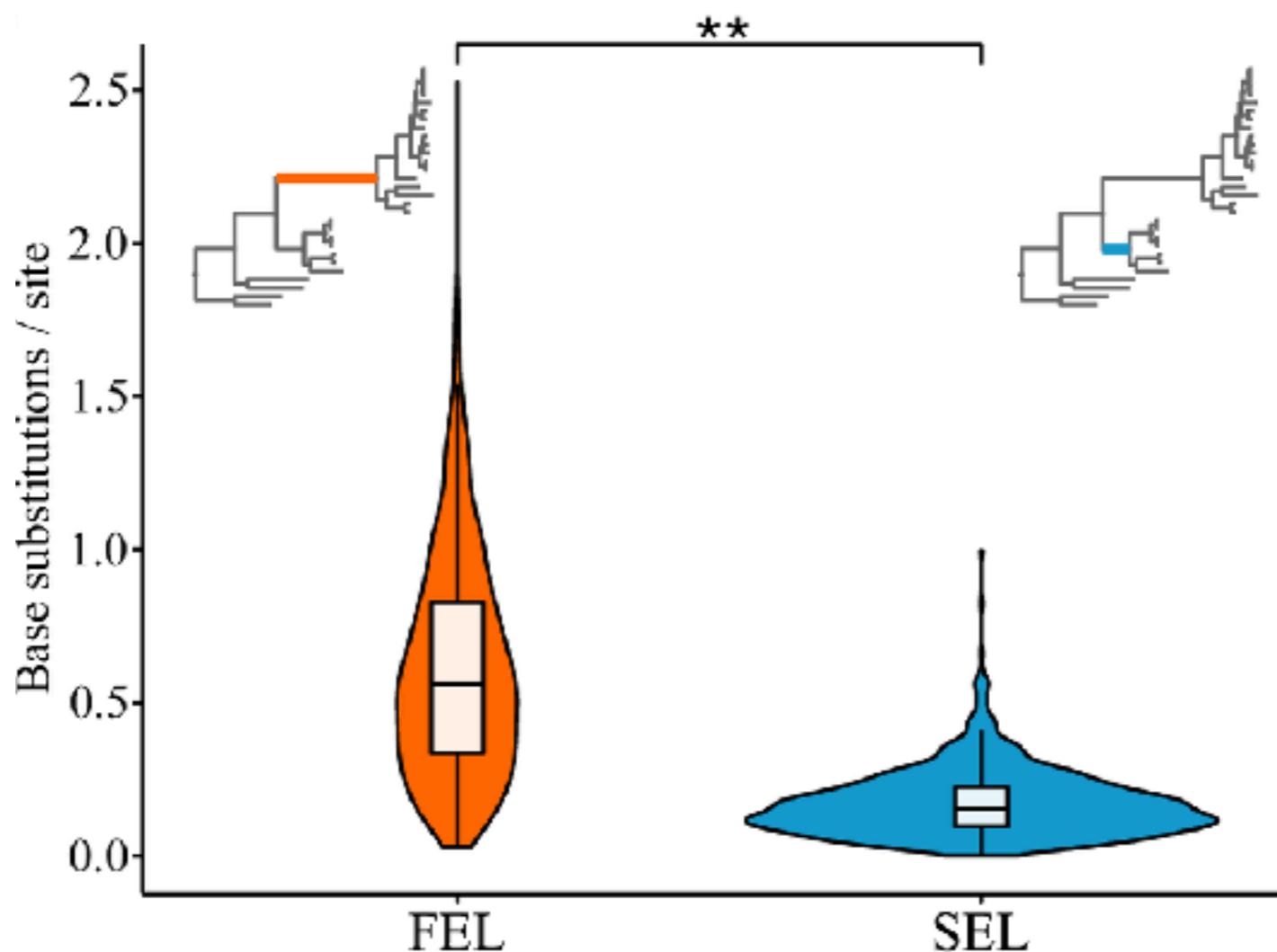
**EAT,  
DRINK  
& BE COVERED.**

© 2010 Pfizer Consumer Healthcare Inc. Pepto-Bismol® is a registered trademark of Pfizer Consumer Healthcare Inc. © 2010 Pfizer Consumer Healthcare Inc. All rights reserved. 1. Based on 2009 sales. 2. Based on 2009 sales. 3. Based on 2009 sales. 4. Based on 2009 sales.

# Balance beauty and information

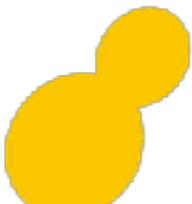


# Keep it simple



# 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

# Where to find the worksheet

1. [jlsteenwyk.github.io](https://jlsteenwyk.github.io)
2. Click the resources tab
3. Under R workshop