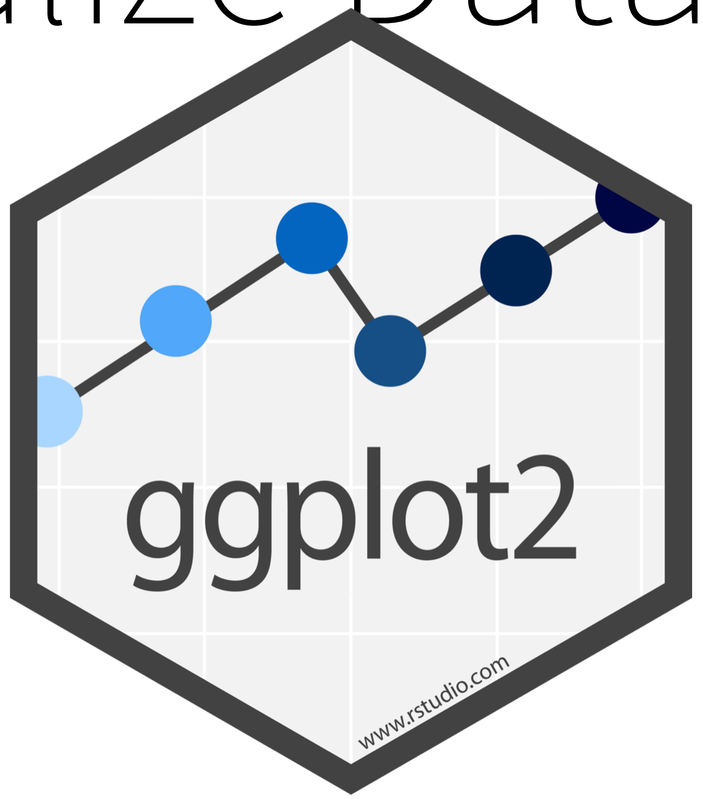
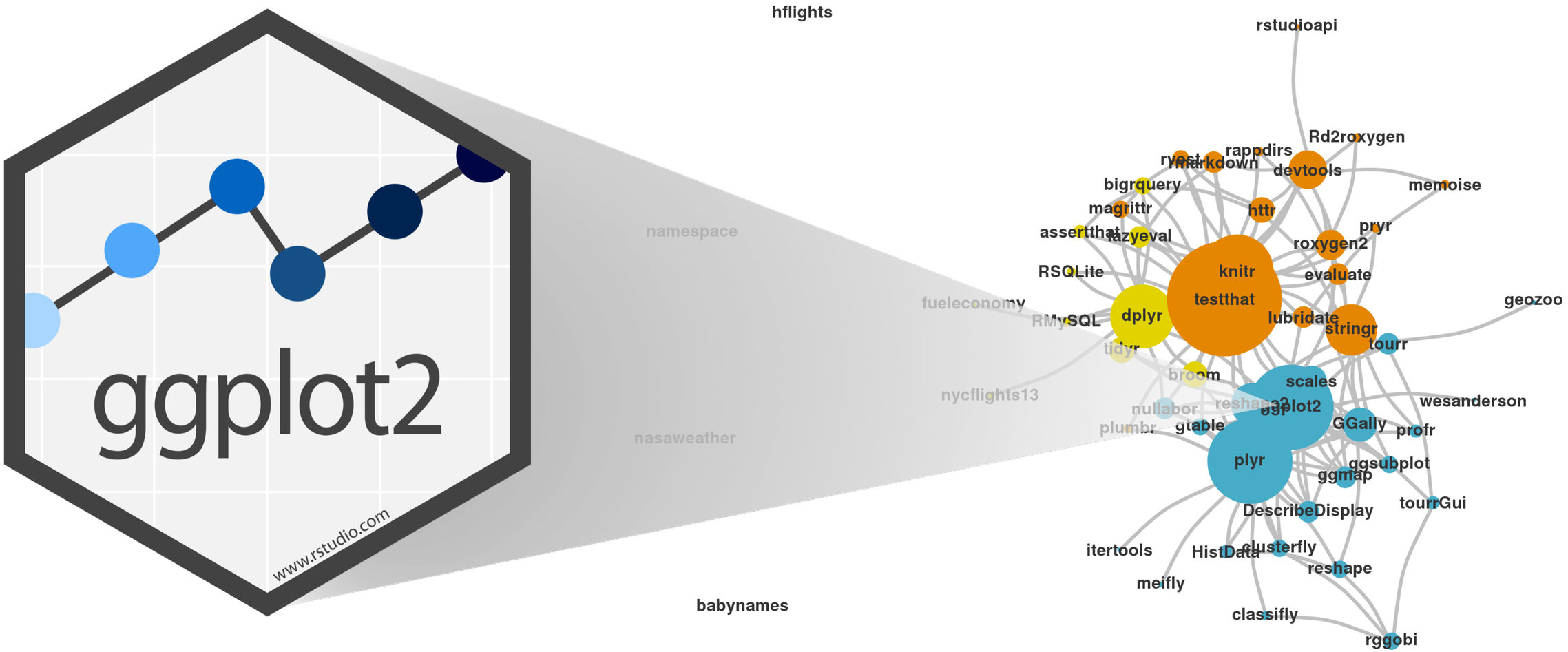
Visualize Data with



CC BY-SA RStudio

"The simple graph has brought more information to the data analyst’s mind than any other device. "

- John Tukey



CC BY-SA RStudio

Display by Adolfo Álvarez

Setup

The setup chunk is always run once before anything else

CC BY-SA RStudio

mpg

Fuel economy data for 38 models of car.

```
mpg
```

CC BY-SA RStudio

Quiz

Confer with your group.

What relationship do you expect to see between engine size (displ) and mileage (hwy)?

No peeking ahead!

CC BY-SA RStudio

Your Turn 1

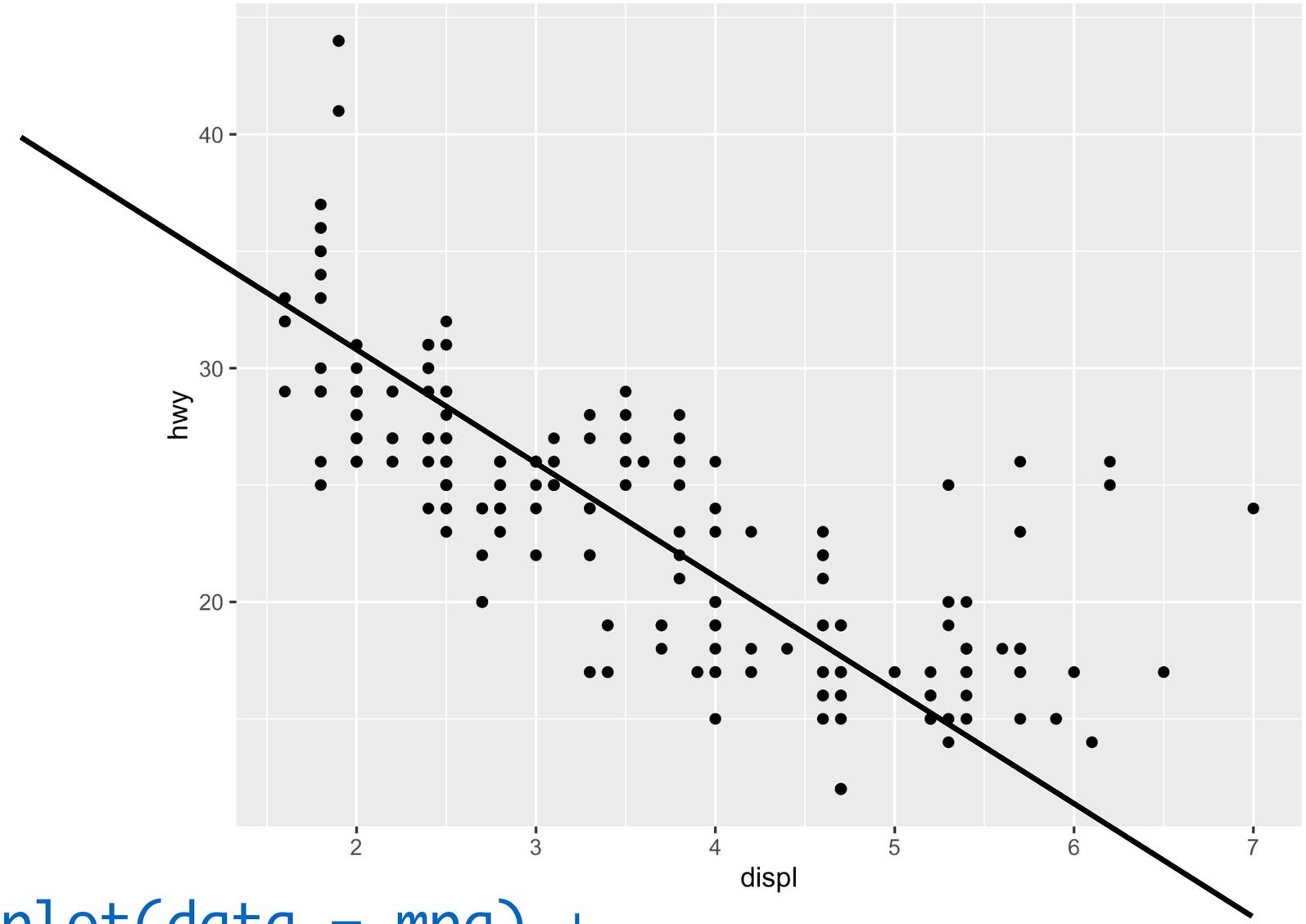
Run this code in your notebook to make a graph.

Pay strict attention to spelling, capitalization, and parentheses!

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```

02:00

CC BY-SA RStudio




```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```

CC BY-SA RStudio

1. "Initialize" a plot with ggplot()
2. Add layers with geom_ functions

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```

CC BY-SA RStudio

Pro tip: Always put the + at the end of a line, Never at the start

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```

CC BY-SA RStudio

data

+ before new line

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```

type of layer

aes()

x variable

y variable

CC BY-SA RStudio

A template

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

```
  geom_point(mapping = aes(x = displ, y = hwy))
```

CC BY-SA RStudio

A template

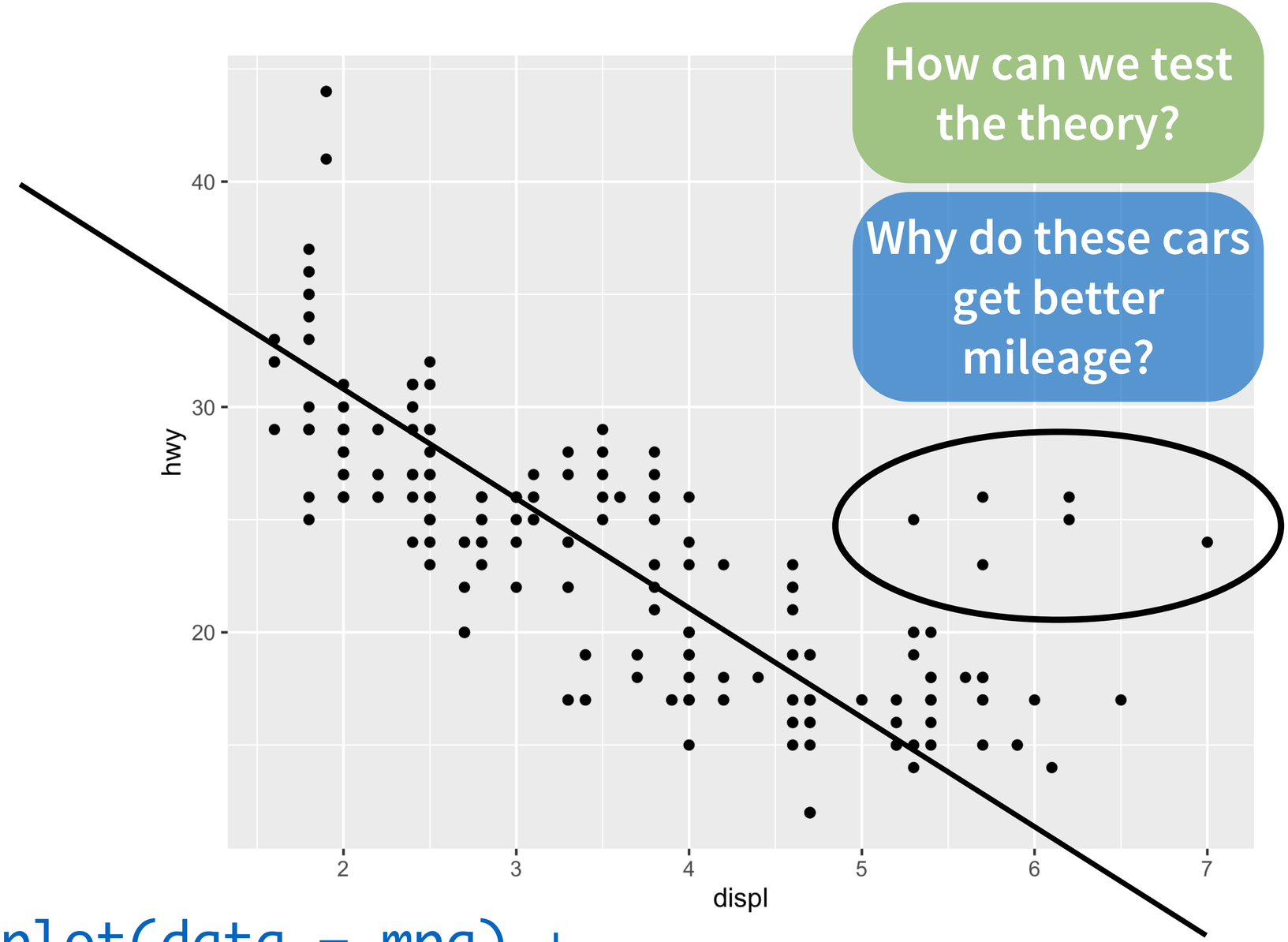
```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

CC BY-SA RStudio

Mappings

"The greatest value of a picture is when it forces us to notice what we never expected to see."

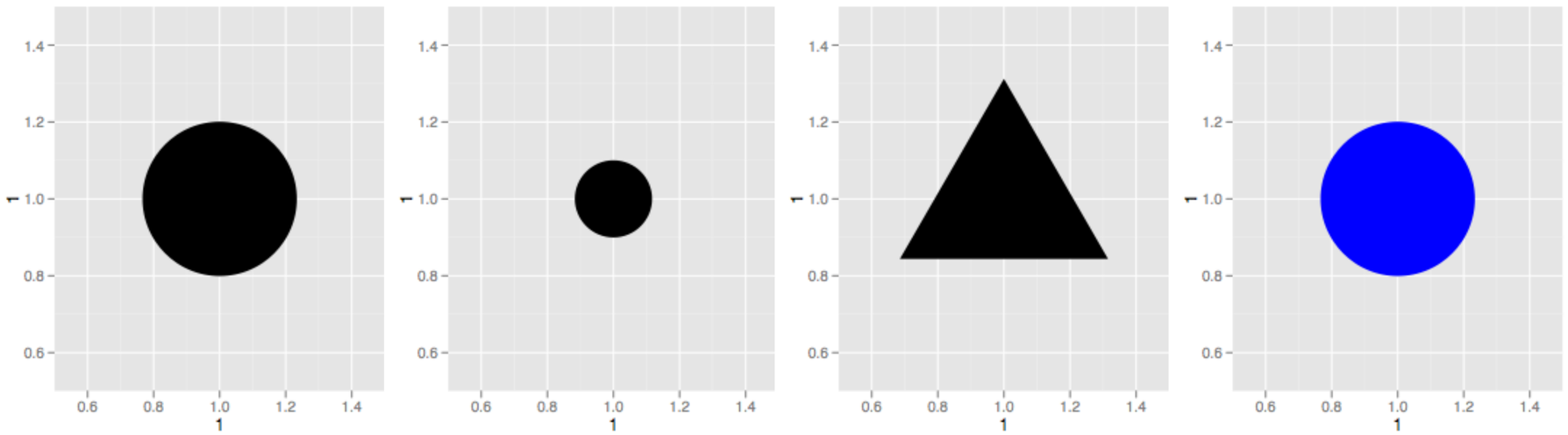
- John Tukey




```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))
```

CC BY-SA RStudio

Aesthetics



CC BY-SA RStudio

Visual Space		Data Space
color	⟷	class
Red	⟷	2seater
Brown	⟷	compact
Green	⟷	midsize
Aqua	⟷	minivan
Blue	⟷	pickup
Violet	⟷	subcompact
Pink	⟷	suv

CC BY-SA RStudio

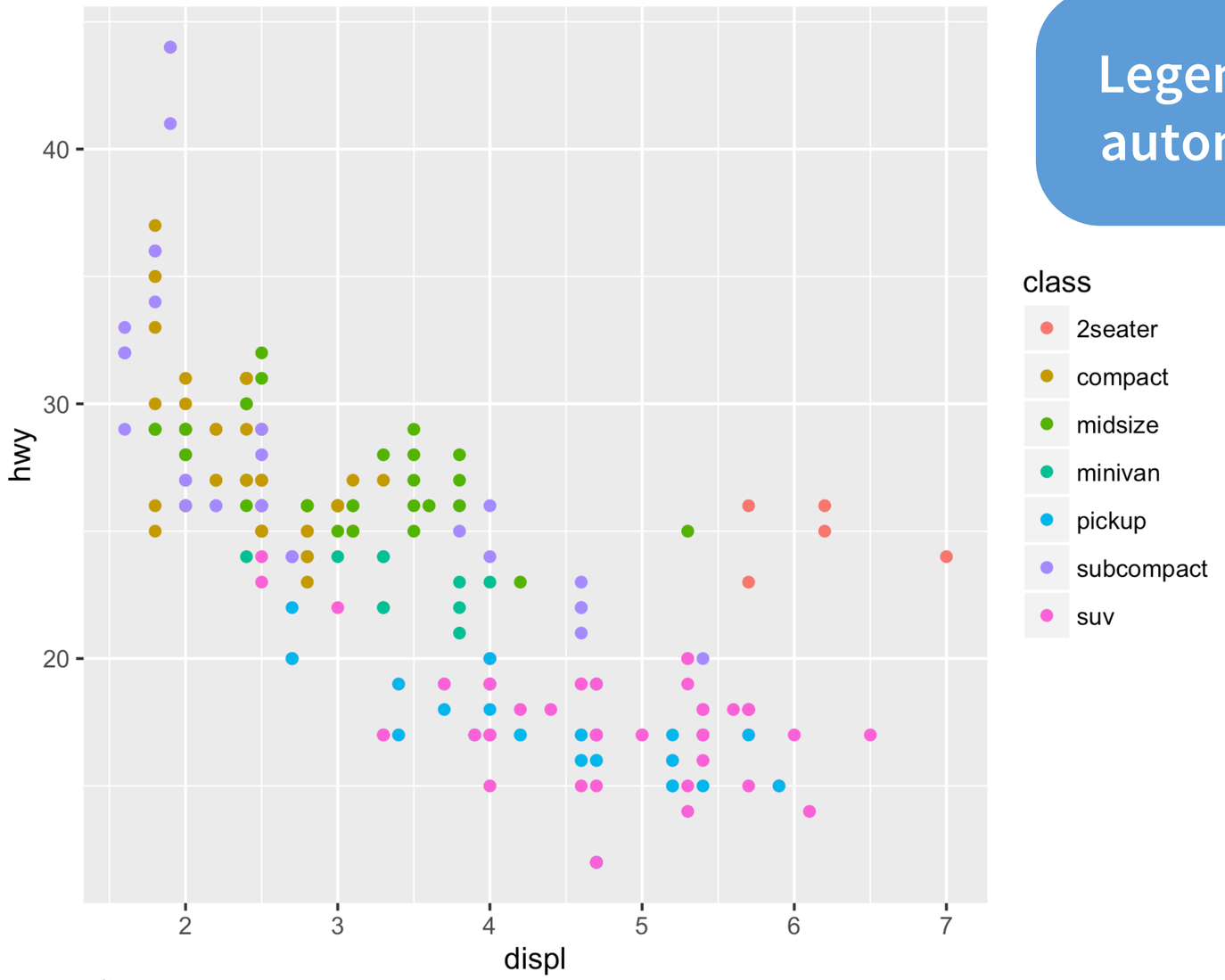
Aesthetics

aesthetic property

Variable to map it to

```
ggplot(mpg) + geom_point(aes(x = displ, y = hwy, color = class))
ggplot(mpg) + geom_point(aes(x = displ, y = hwy, size = class))
ggplot(mpg) + geom_point(aes(x = displ, y = hwy, shape = class))
ggplot(mpg) + geom_point(aes(x = displ, y = hwy, alpha = class))
```

CC BY-SA RStudio



Legend added automatically

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = class))
```

CC BY-SA RStudio

Your Turn 2

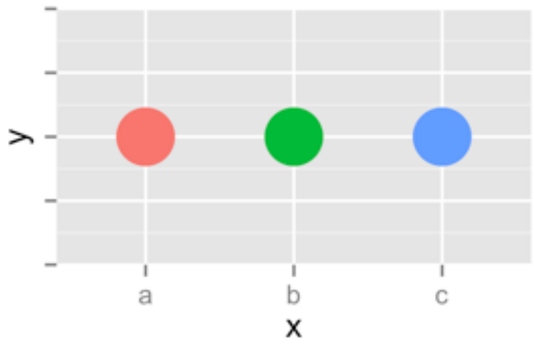
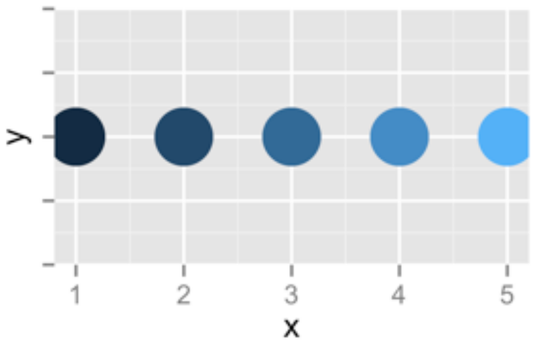
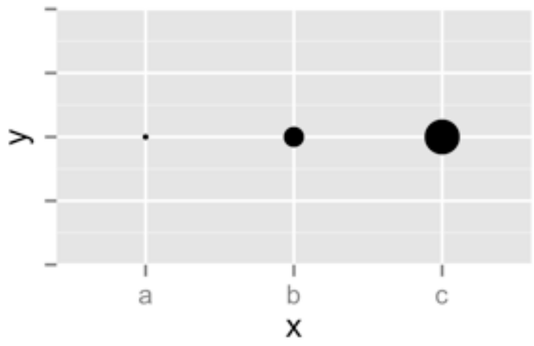
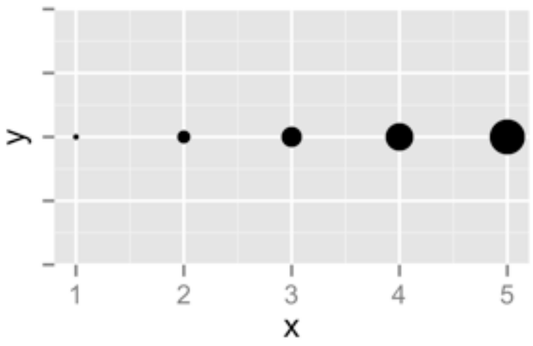
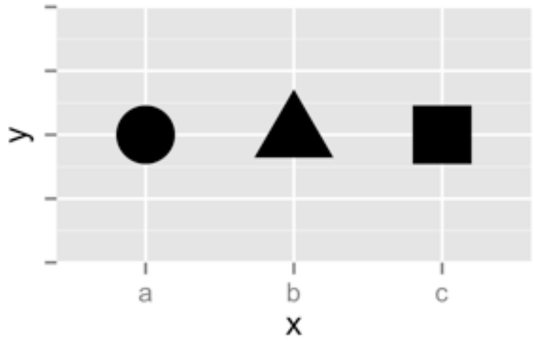
In the next chunk, add color, size, alpha, and shape aesthetics to your graph. Experiment.

Do different things happen when you map aesthetics to discrete and continuous variables?

What happens when you use more than one aesthetic?

05:00

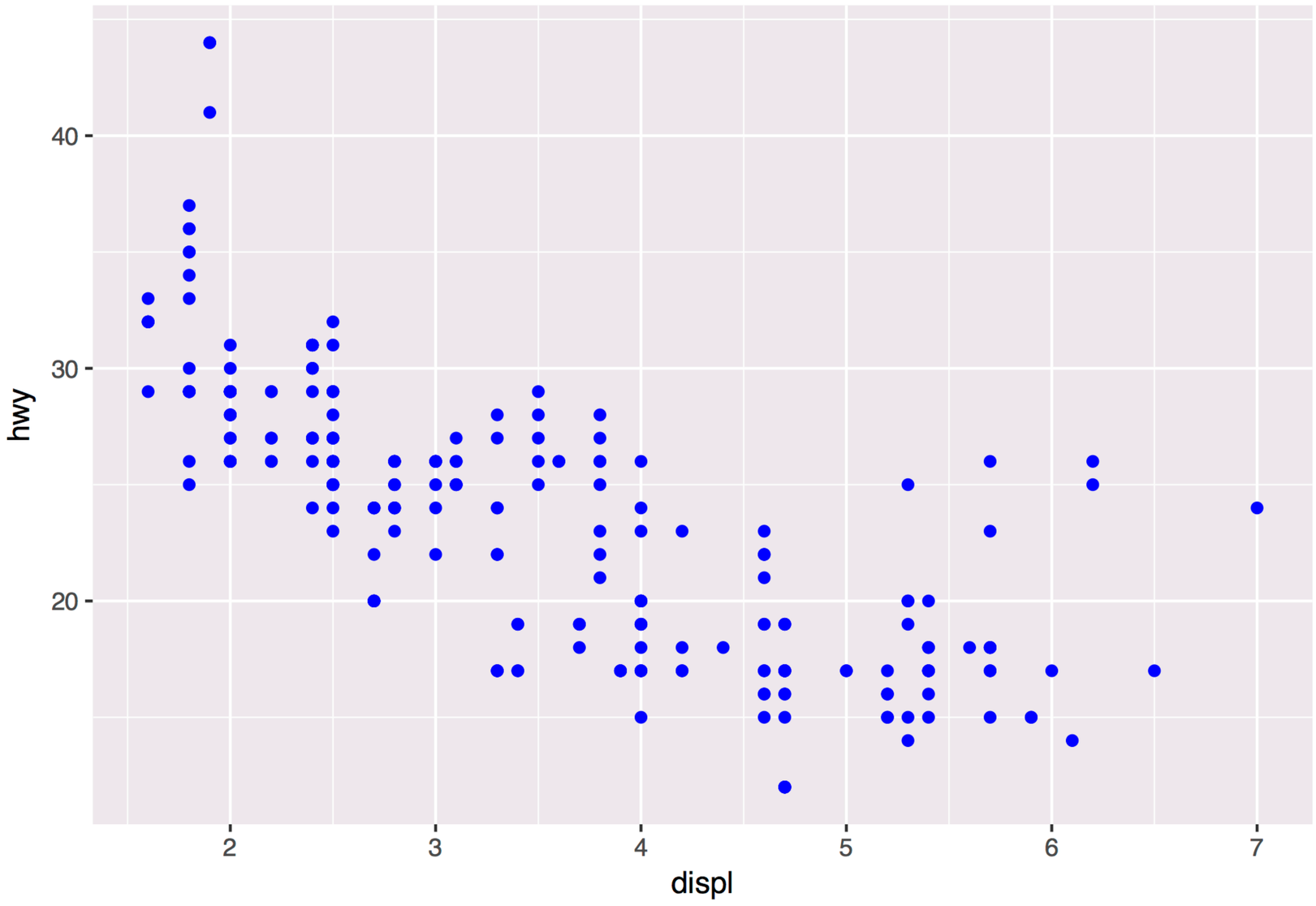
CC BY-SA RStudio

	Discrete	Continuous
Color		
Size		
Shape		

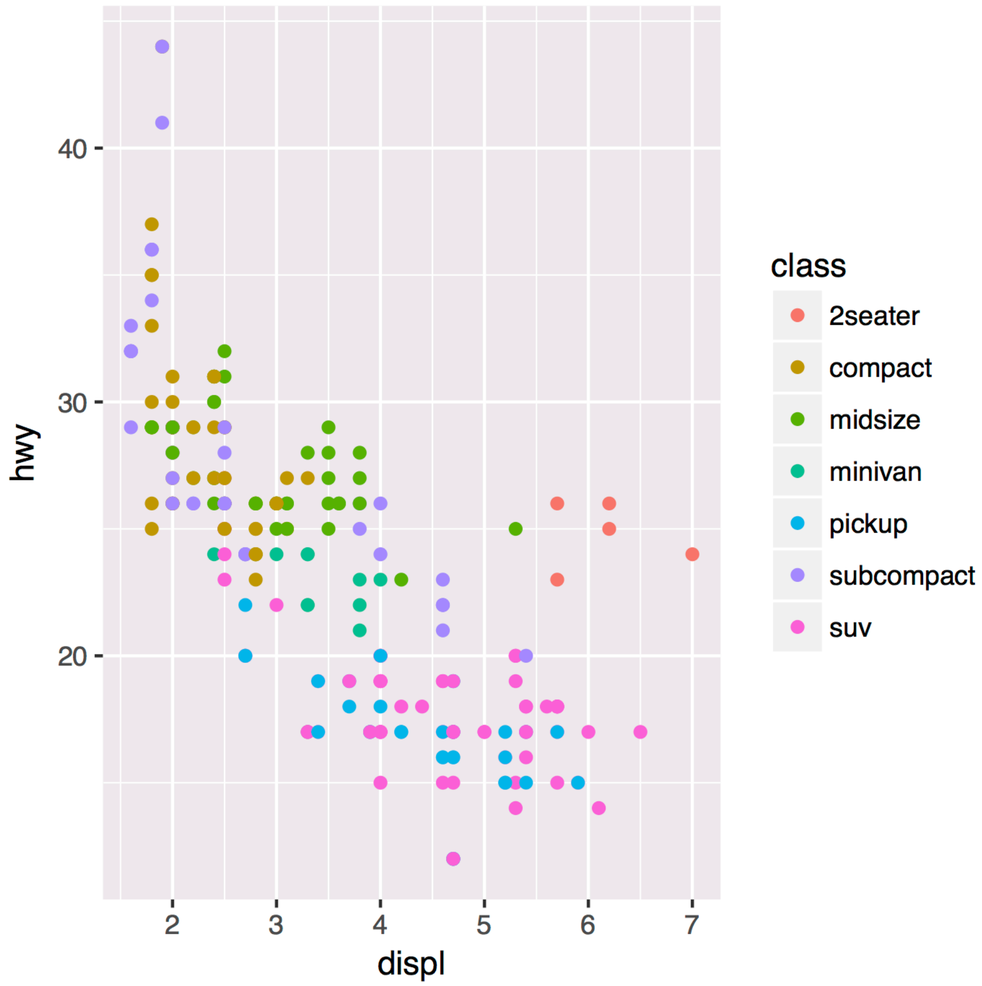
CC BY-SA RStudio

set vs. map

How would you make this plot?



CC BY-SA RStudio

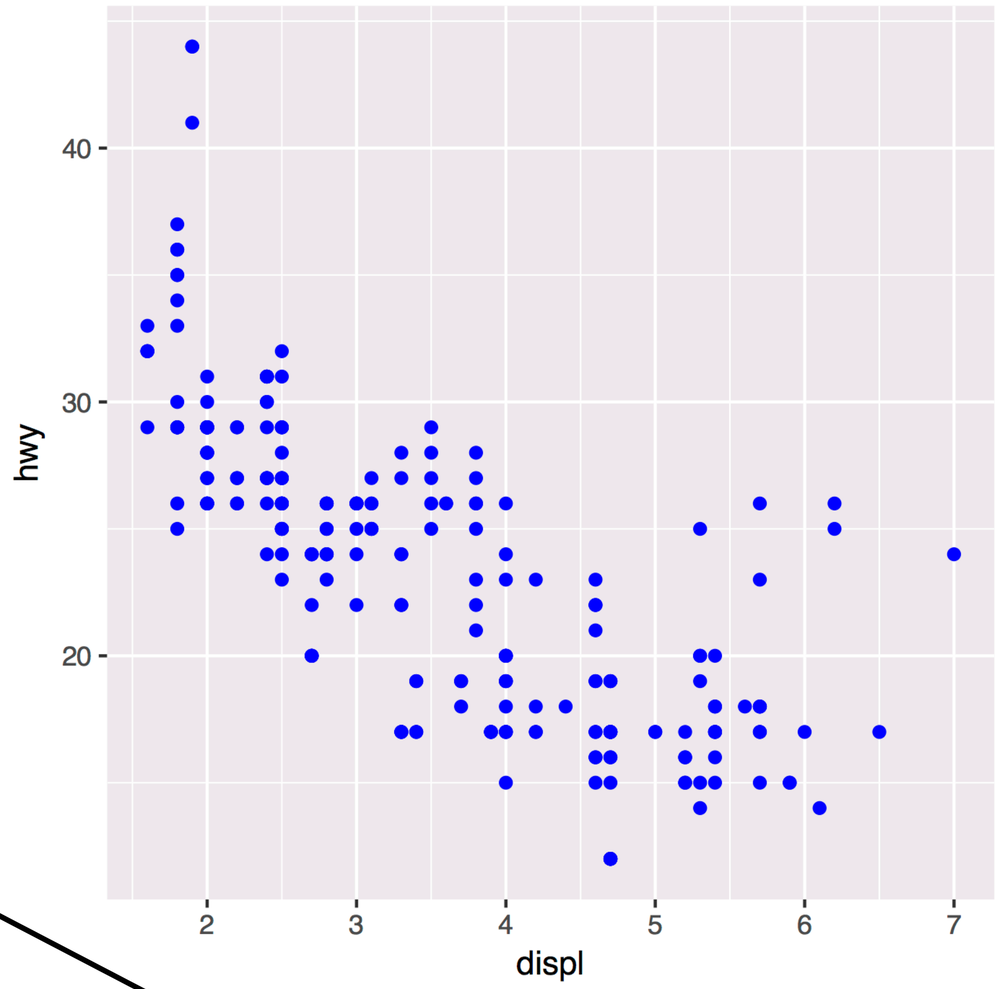


Inside of aes(): maps an aesthetic to a variable

```
ggplot(mpg) + geom_point(aes(x = displ, y = hwy, color = class))
```

CC BY-SA RStudio

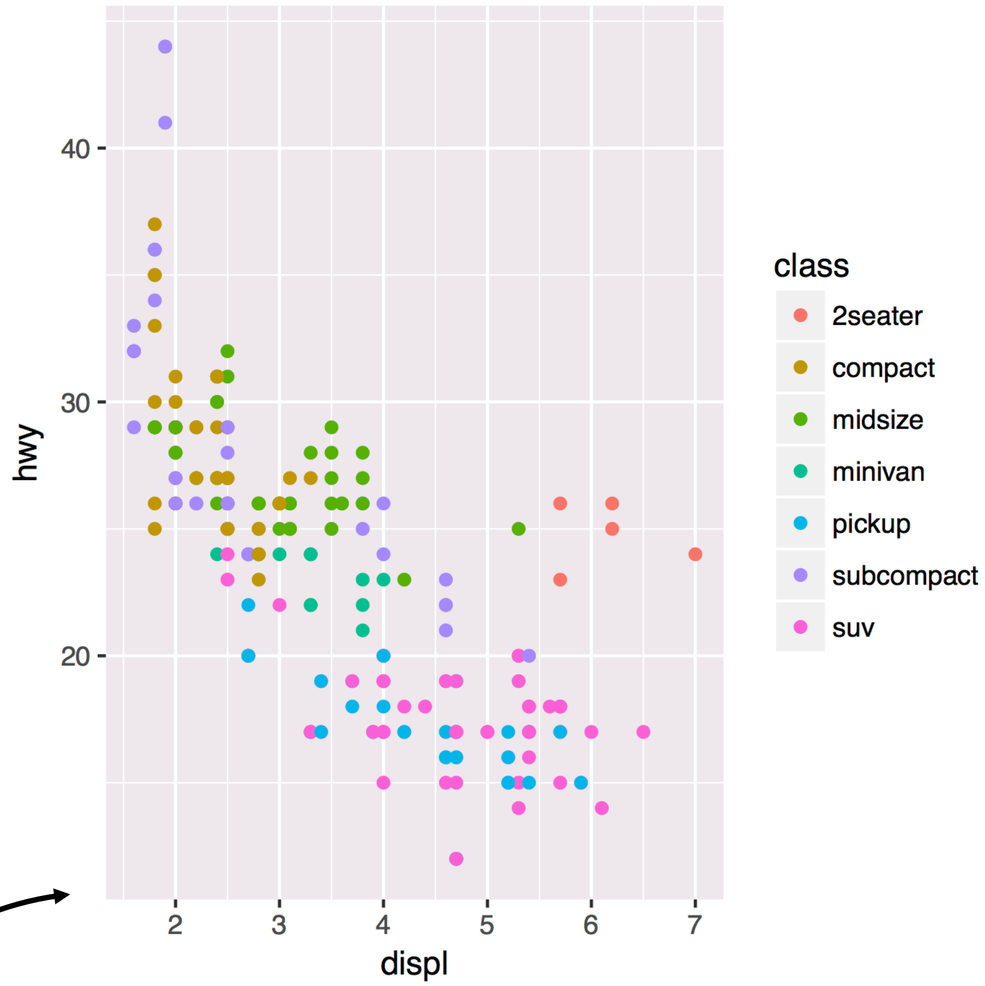
Outside of aes(): sets an aesthetic to a value

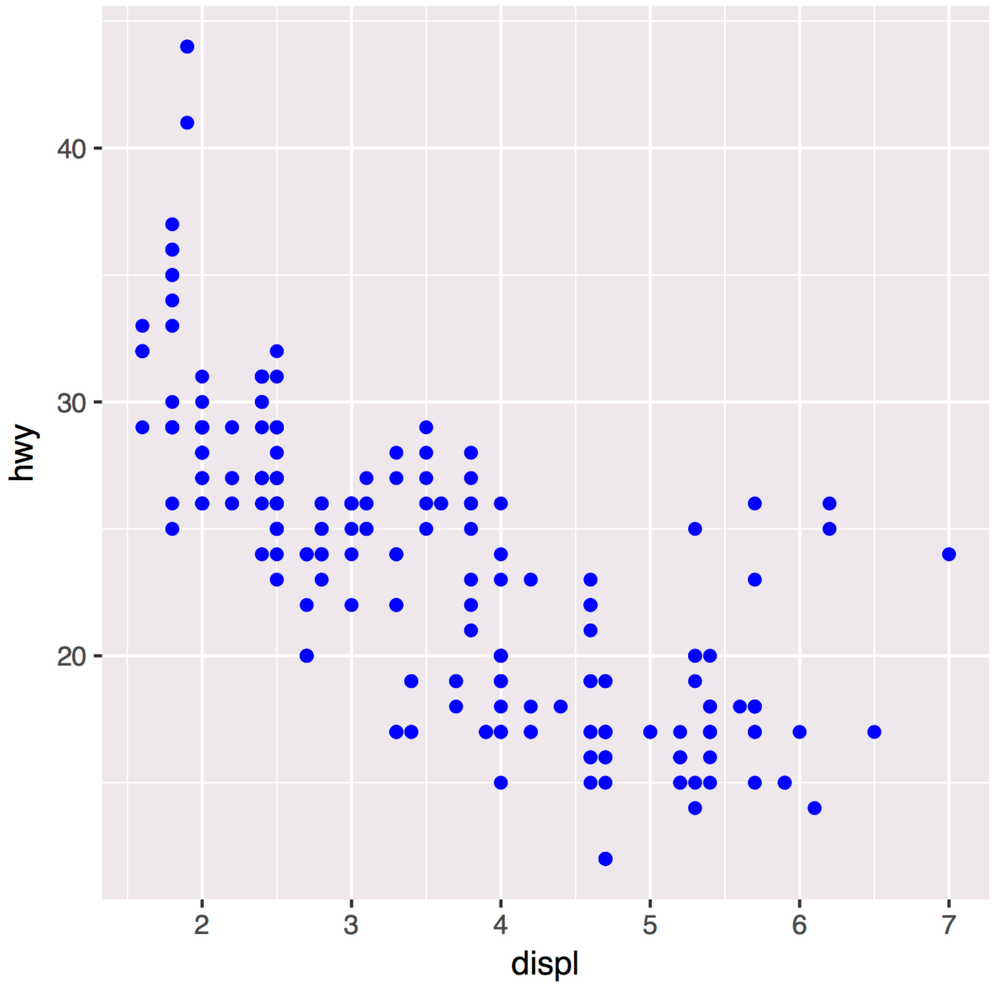



```
ggplot(mpg) + geom_point(aes(x = displ, y = hwy, color = class))
```

```
ggplot(mpg) + geom_point(aes(x = displ, y = hwy), color = "blue")
```

CC BY-SA RStudio

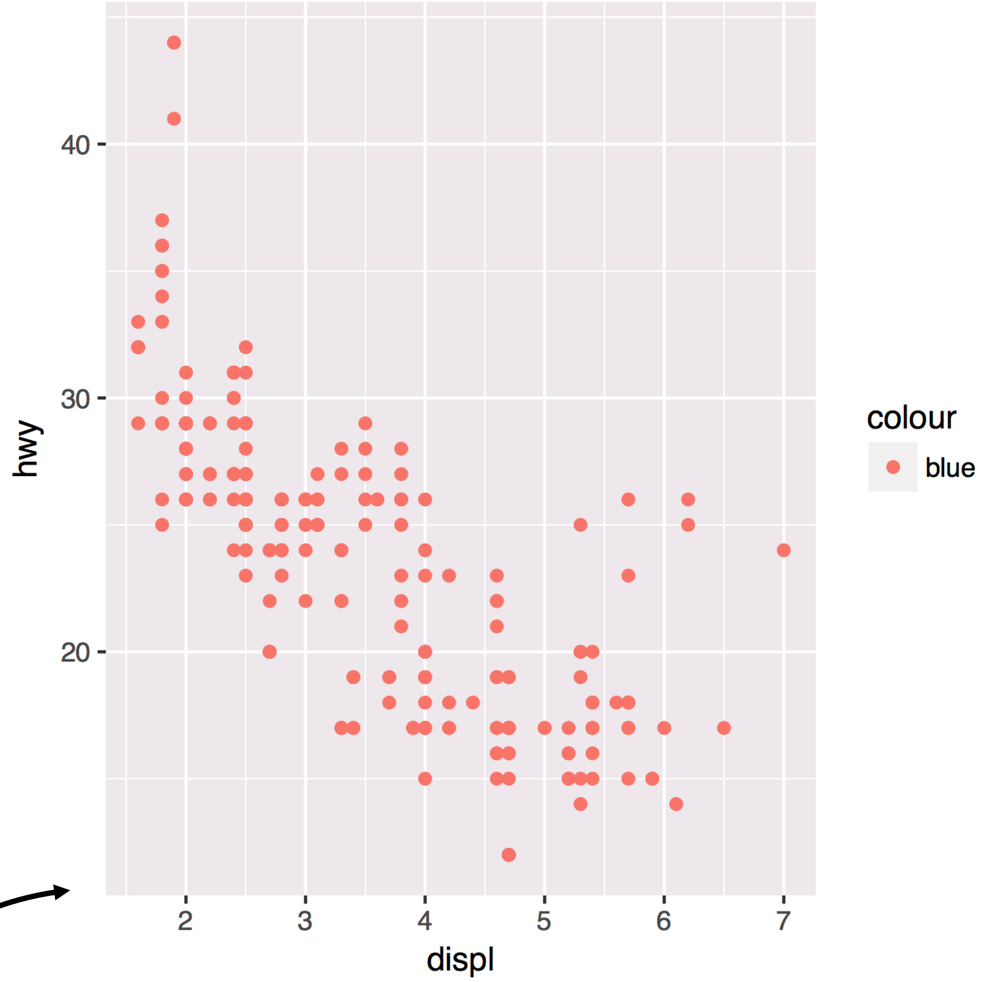


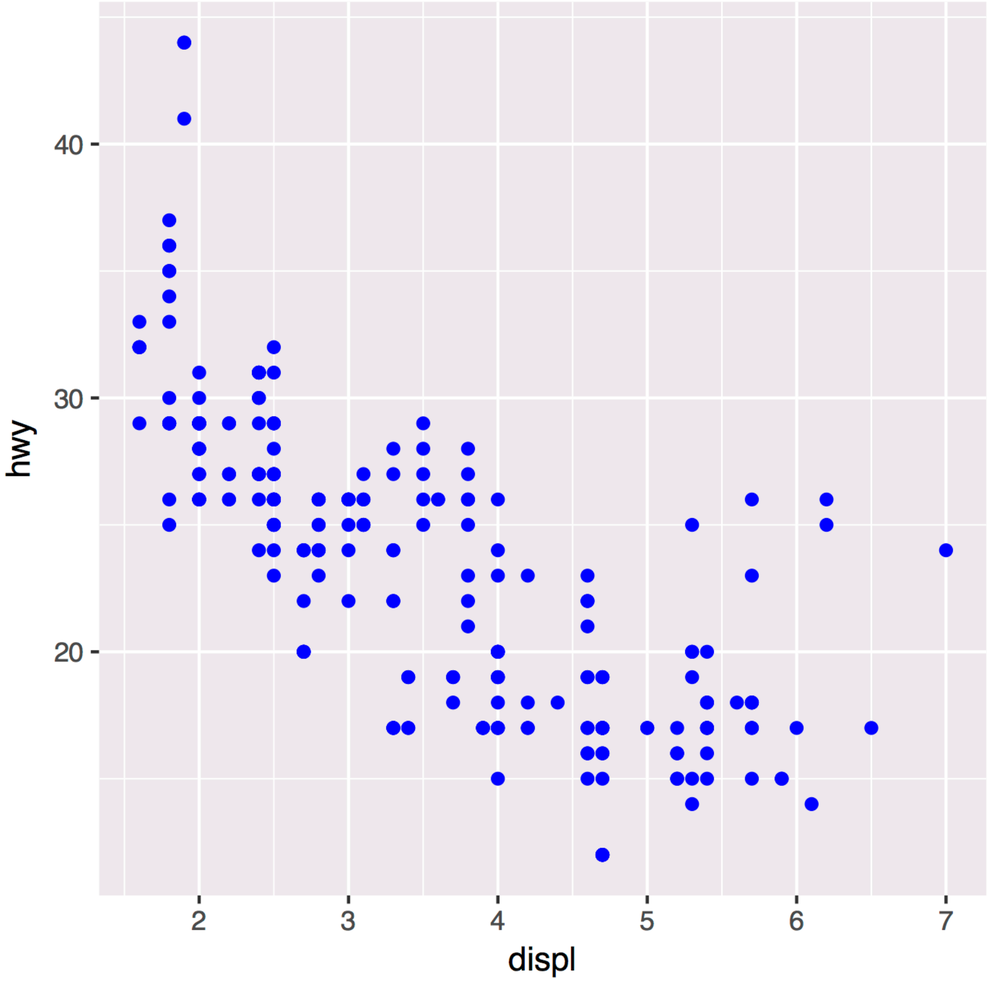



```
ggplot(mpg) + geom_point(aes(x = displ, y = hwy, color = class))
```

```
ggplot(mpg) + geom_point(aes(x = displ, y = hwy), color = "blue")
```

CC BY-SA RStudio






```
ggplot(mpg) + geom_point(aes(x = displ, y = hwy, color = "blue"))
```

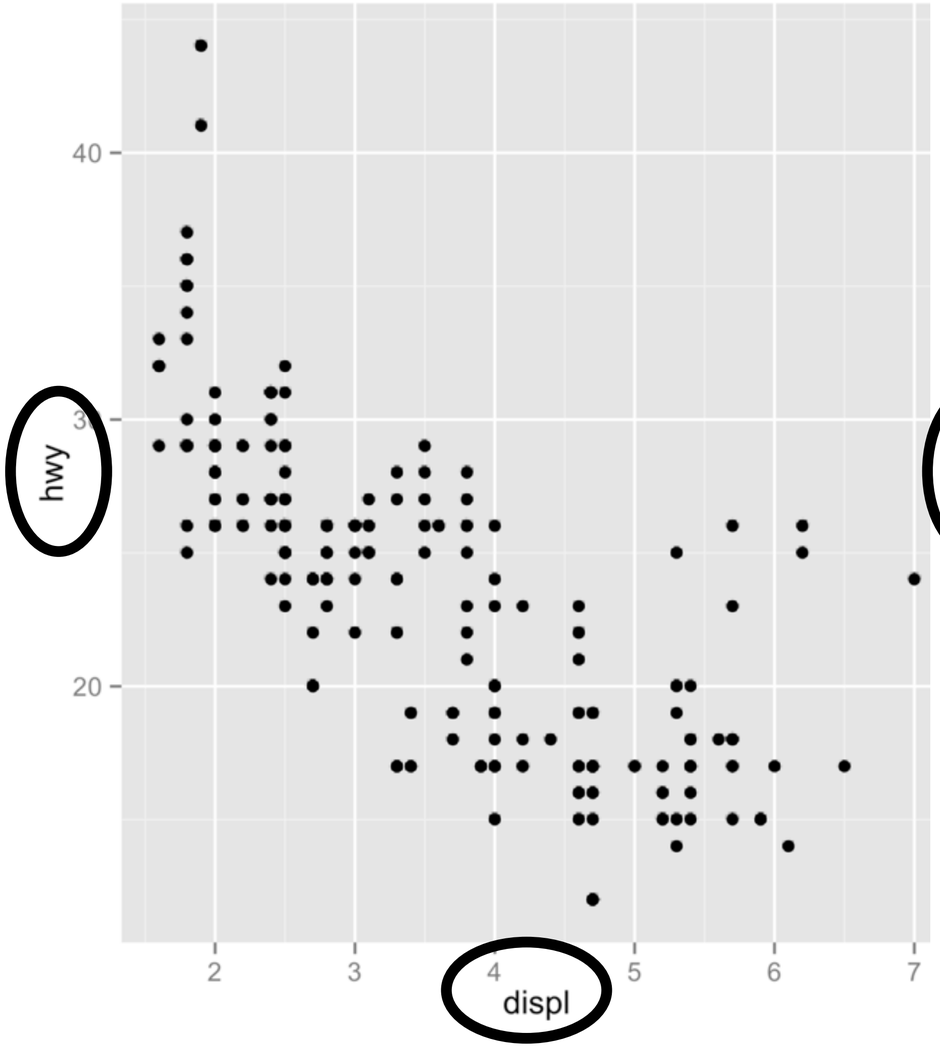
```
ggplot(mpg) + geom_point(aes(x = displ, y = hwy), color = "blue")
```

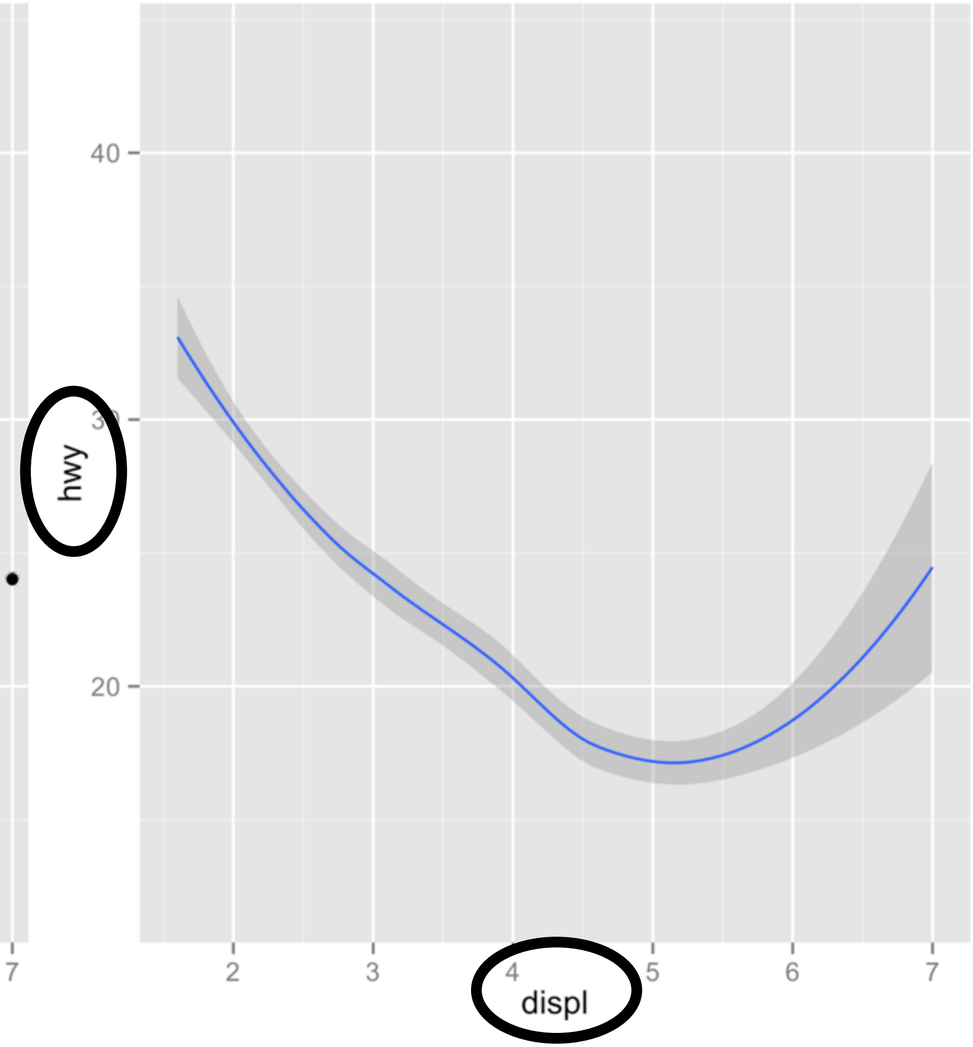
CC BY-SA RStudio

Geoms

How are these plots similar?

Same: x var , y var , data



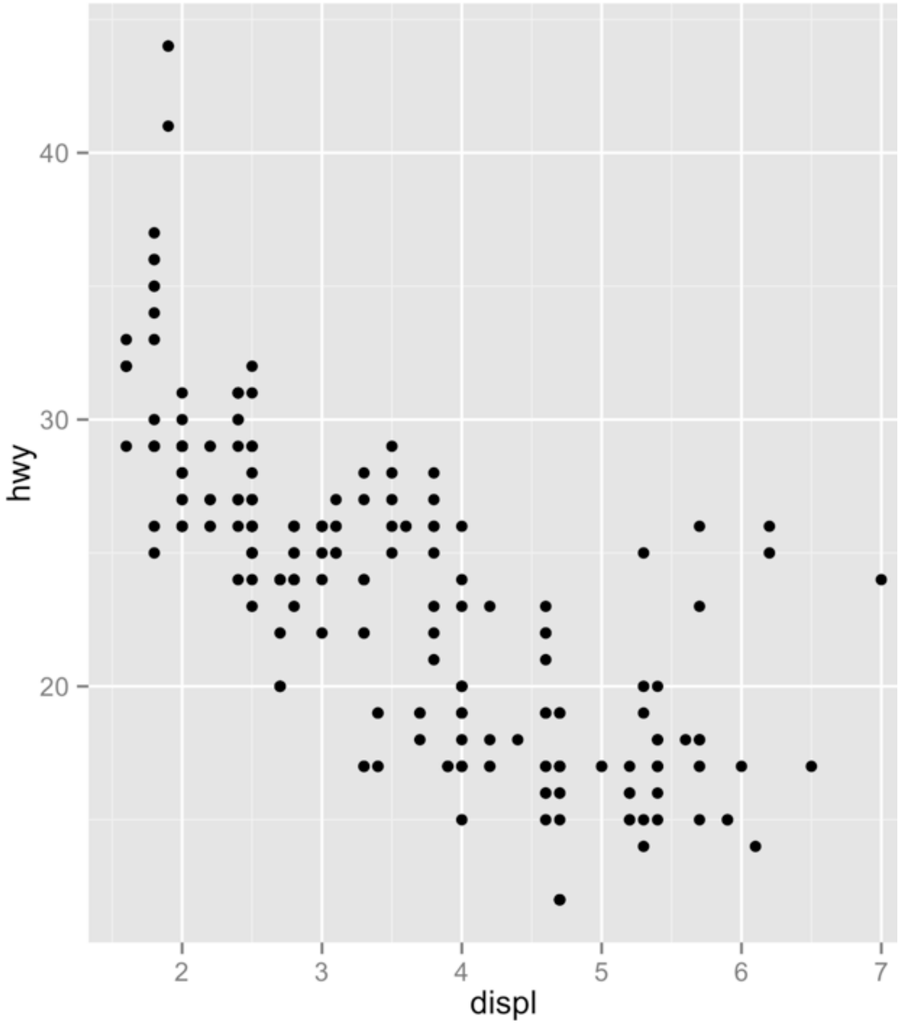


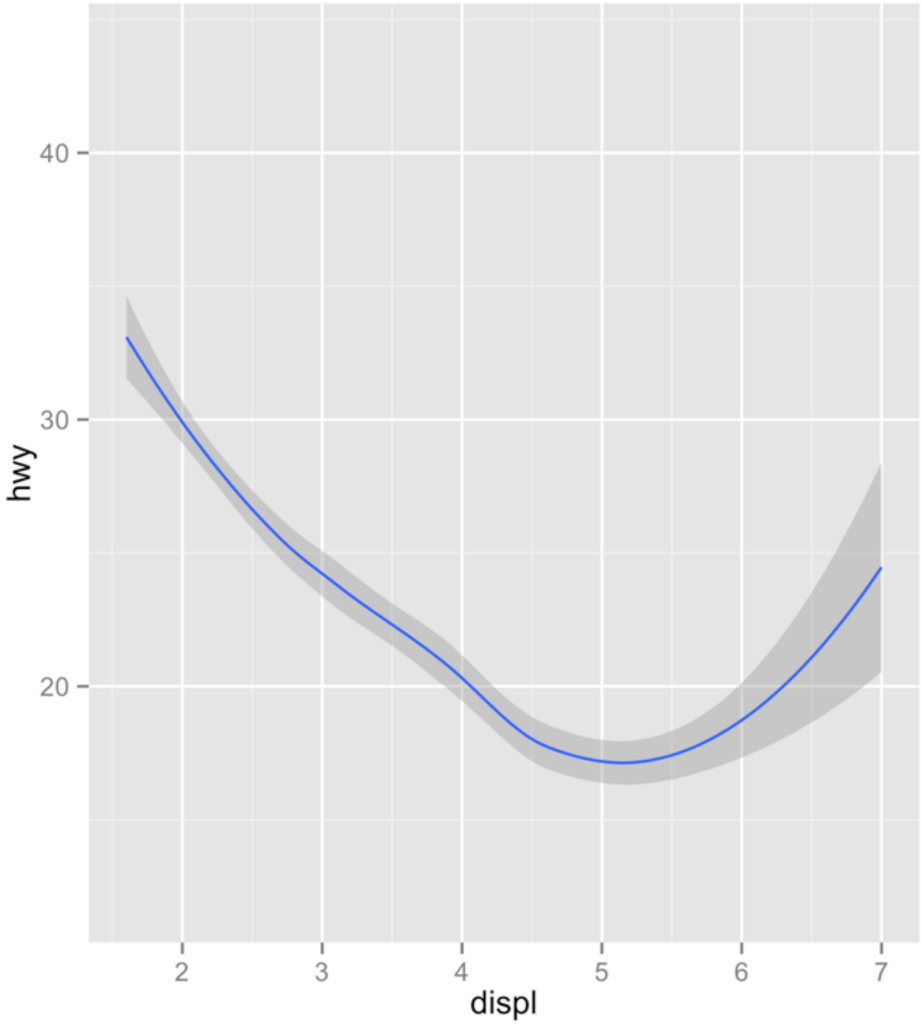
CC BY-SA RStudio

How are these plots different?

Different: geometric object (geom),

e.g. the visual object used to represent the data





CC BY-SA RStudio

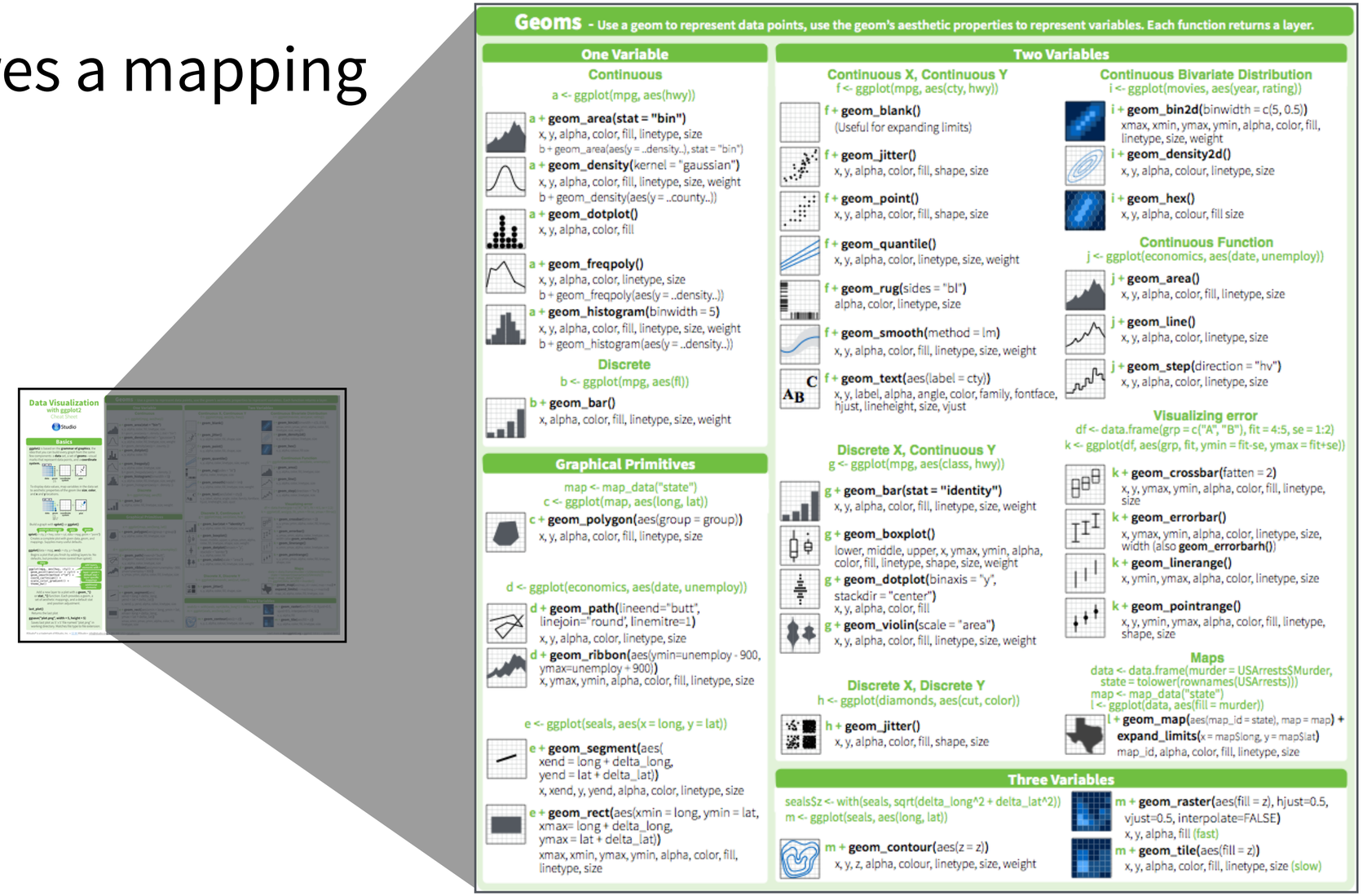
geoms

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

CC BY-SA RStudio

geom_ functions

Each requires a mapping argument.



CC BY-SA RStudio

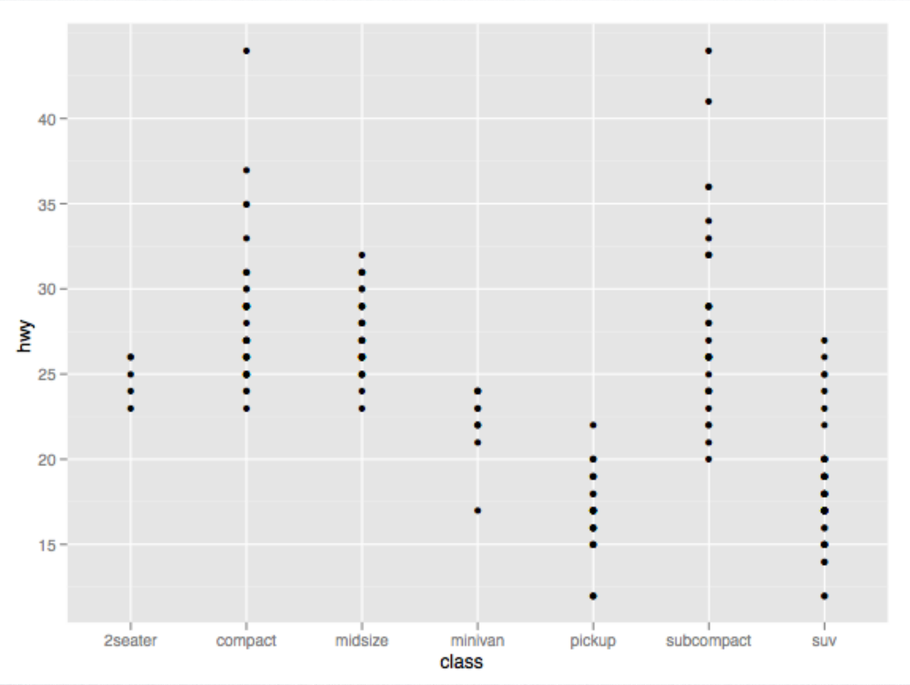
Your Turn

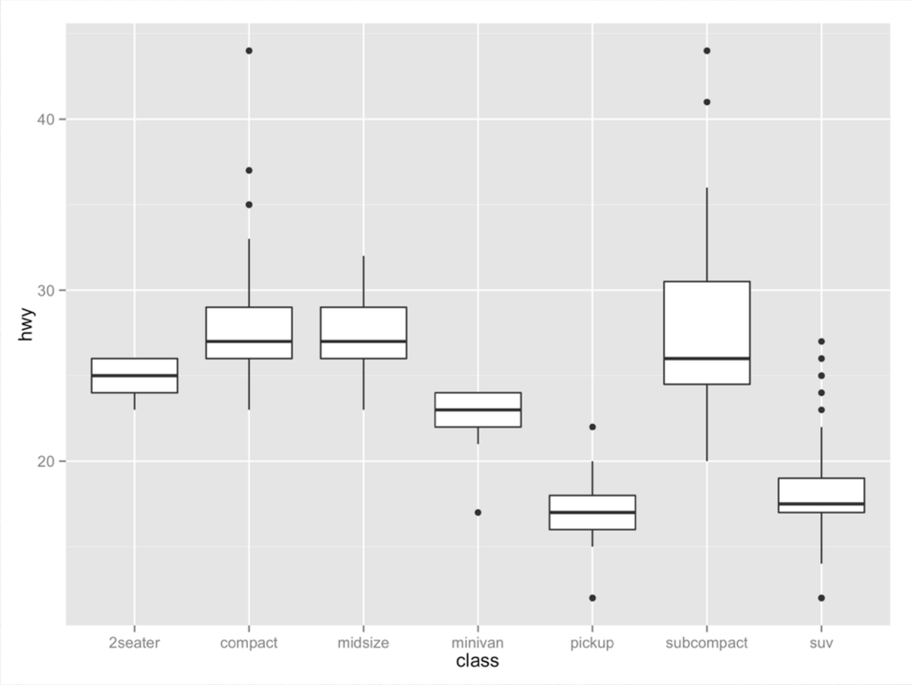
Pair up.

CC BY-SA RStudio

Your Turn 3

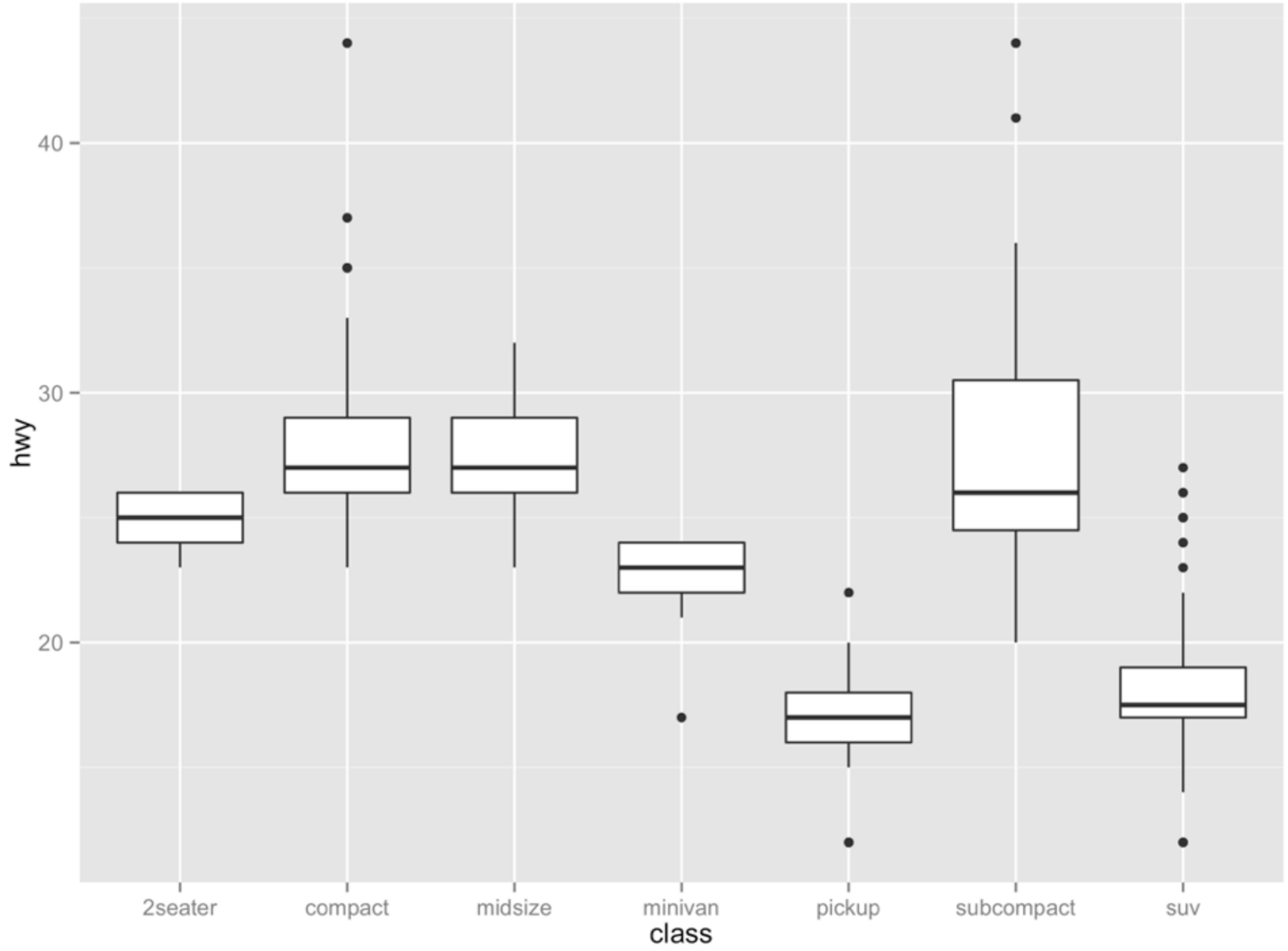
With your partner, decide how to replace this scatterplot with one that draws boxplots? Use the cheatsheet. Try your best guess.






```
ggplot(mpg) + geom_point(aes(class, hwy))
```

CC BY-SA RStudio

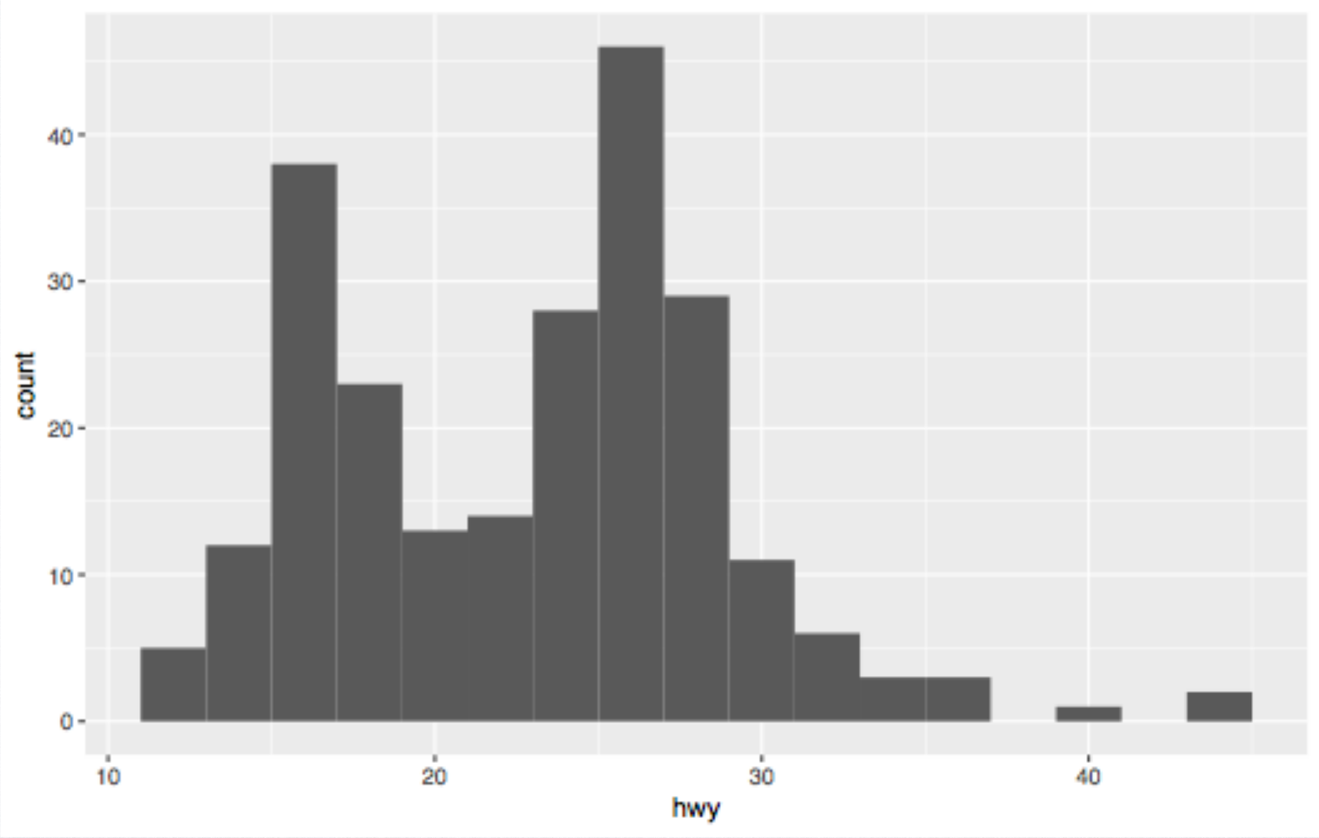



```
ggplot(data = mpg) +
  geom_boxplot(mapping = aes(x = class, y = hwy))
```

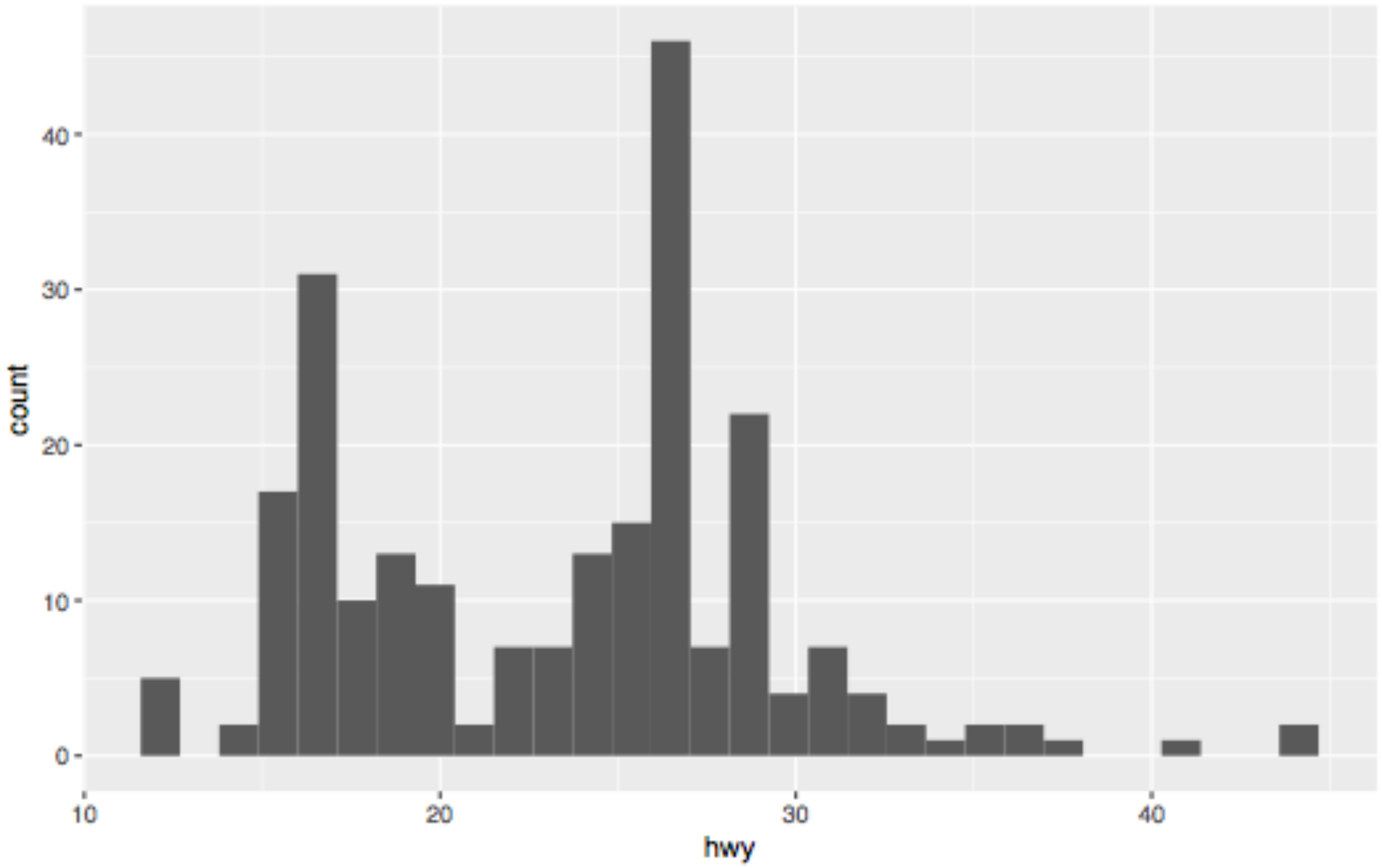
CC BY-SA RStudio

Your Turn 4

With your partner, make the histogram of **hwy** below. Use the cheatsheet. Hint: do not supply a **y** variable.

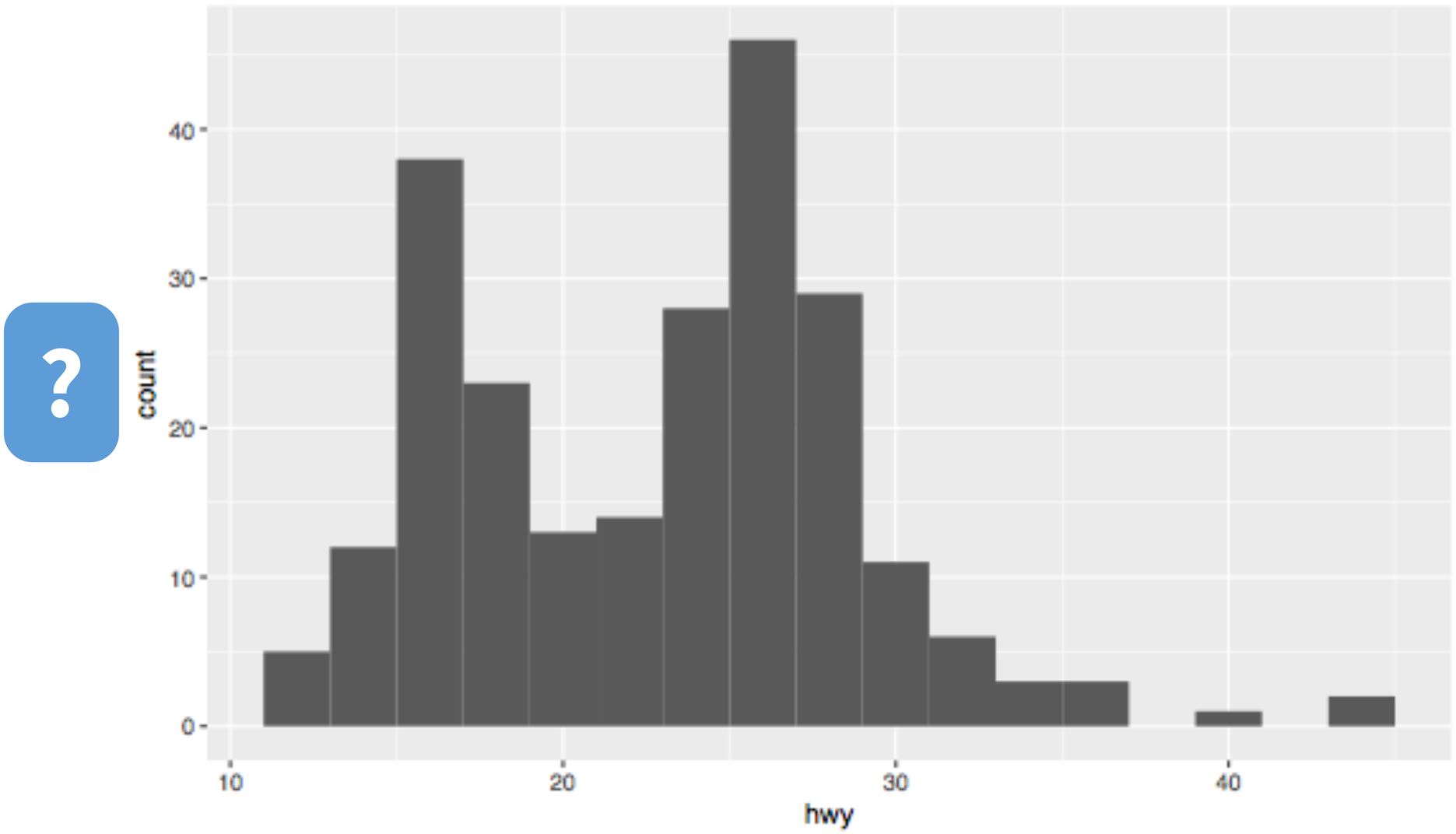


CC BY-SA RStudio




```
ggplot(data = mpg) +
  geom_histogram(mapping = aes(x = hwy))
```

CC BY-SA RStudio

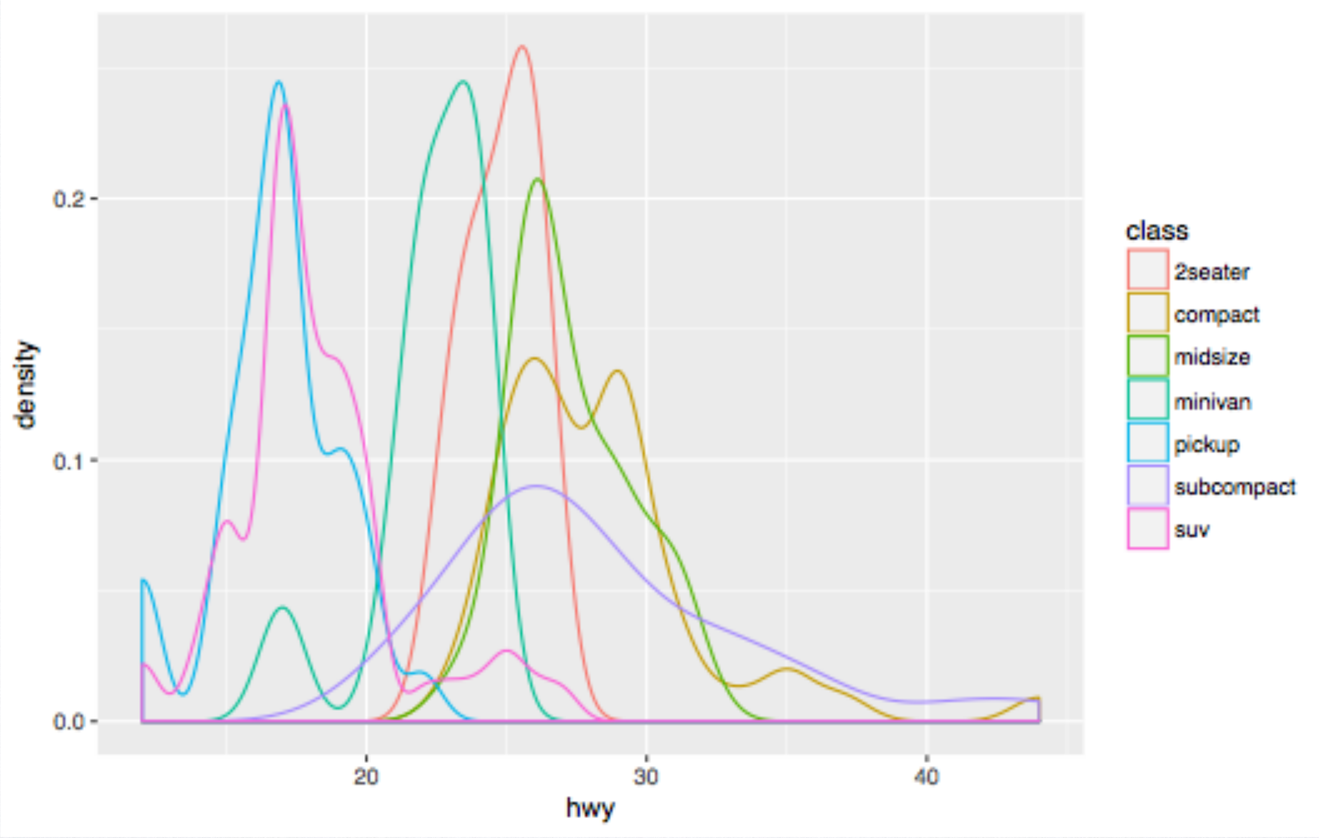



```
ggplot(data = mpg) +
  geom_histogram(mapping = aes(x = hwy), binwidth = 2)
```

CC BY-SA RStudio

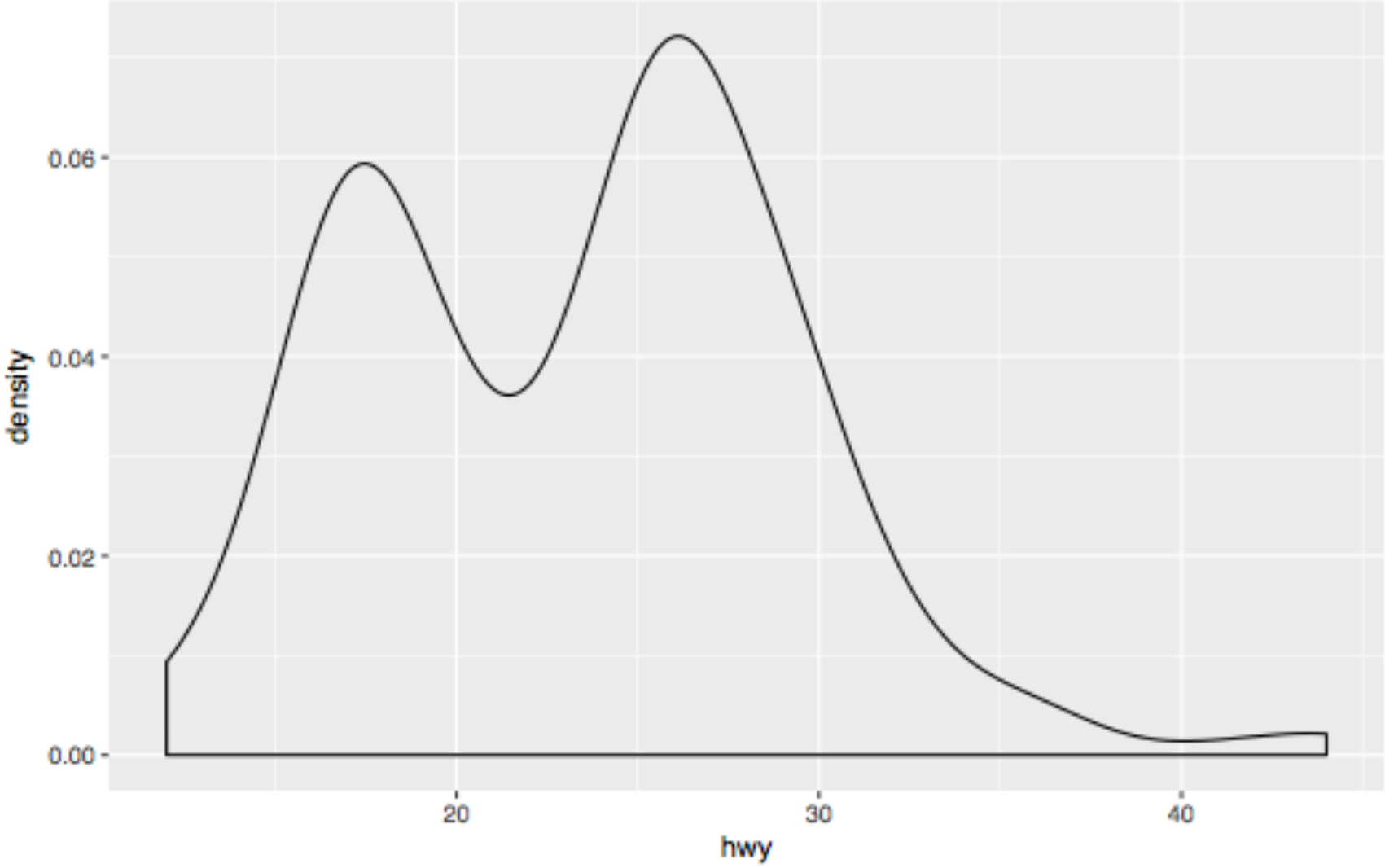
Your Turn 5

With your partner, make the density plot of **hwy** colored by **class** below. Use the cheatsheet. Try your best guess.



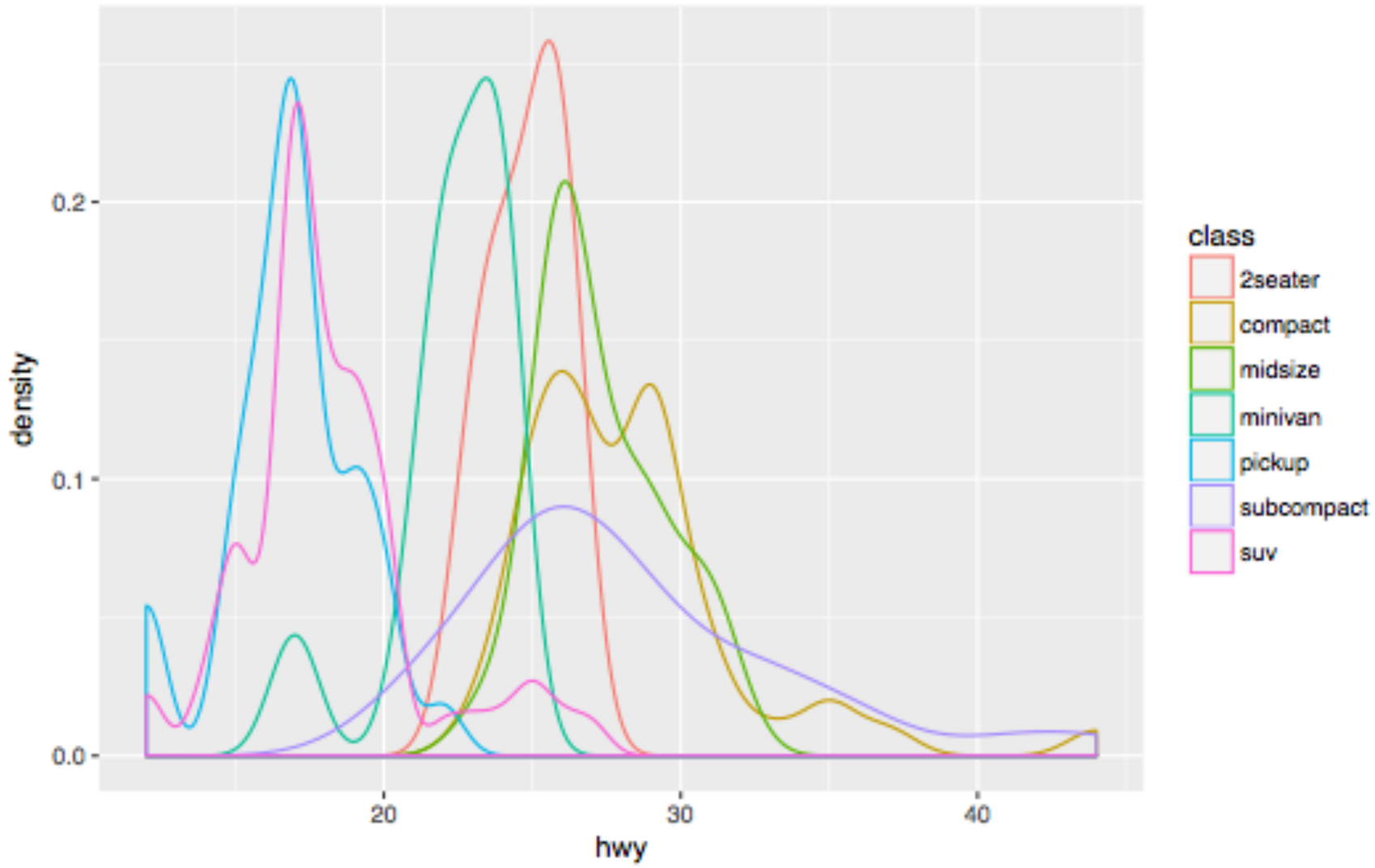
02:00

CC BY-SA RStudio



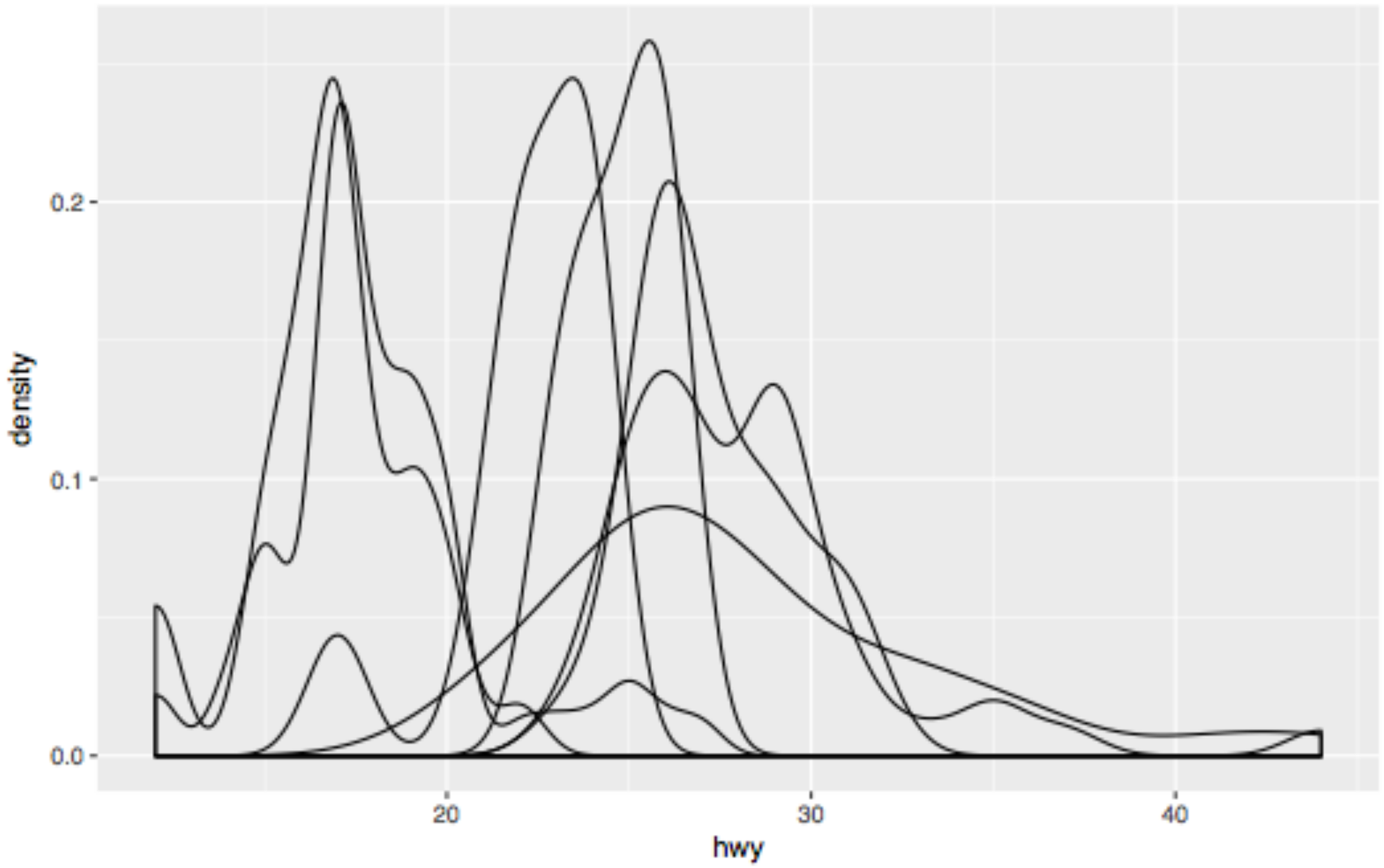

```
ggplot(data = mpg) +
  geom_density(mapping = aes(x = hwy))
```

CC BY-SA RStudio




```
ggplot(data = mpg) +
  geom_density(mapping = aes(x = hwy, color = class))
```

CC BY-SA RStudio

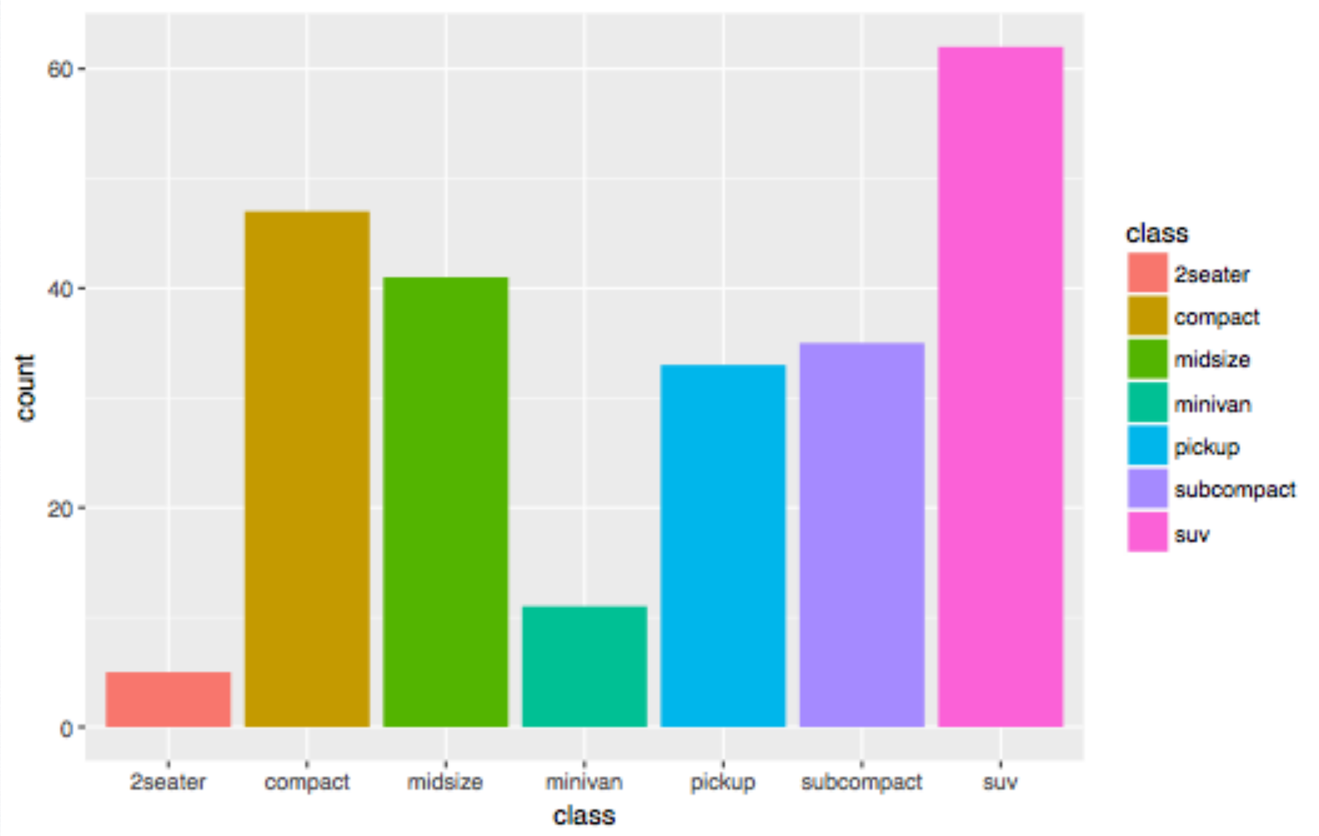



```
ggplot(data = mpg) +
  geom_density(mapping = aes(x = hwy, group = class))
```

CC BY-SA RStudio

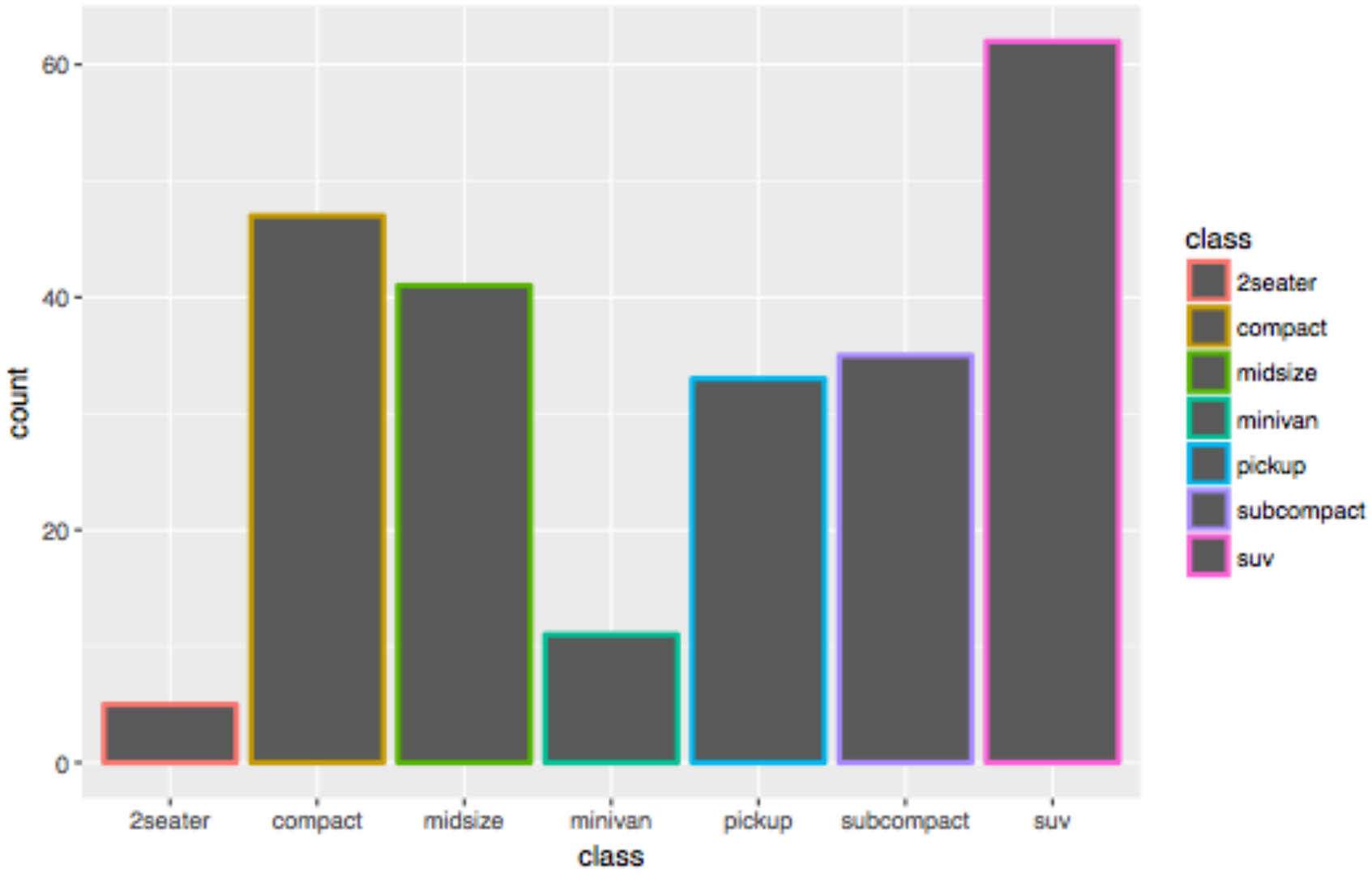
Your Turn 6

With your partner, make the bar chart of **hwy** colored by **class** below. Use the cheatsheet. Try your best guess.



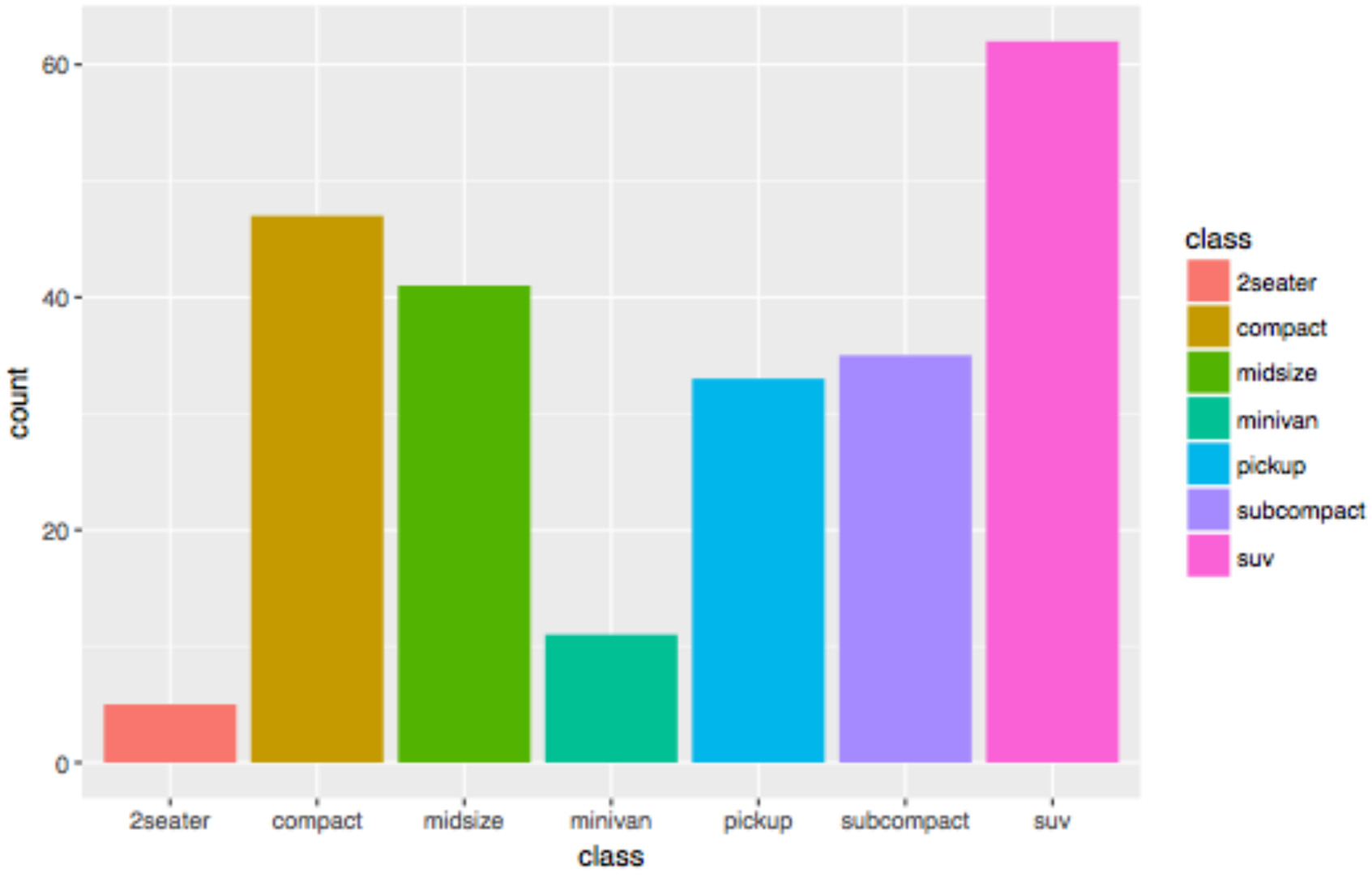
02:00

CC BY-SA RStudio




```
ggplot(data = mpg) +
  geom_bar(mapping = aes(x = class, color = class))
```

CC BY-SA RStudio




```
ggplot(data = mpg) +
  geom_bar(mapping = aes(x = class, fill = class))
```

CC BY-SA RStudio

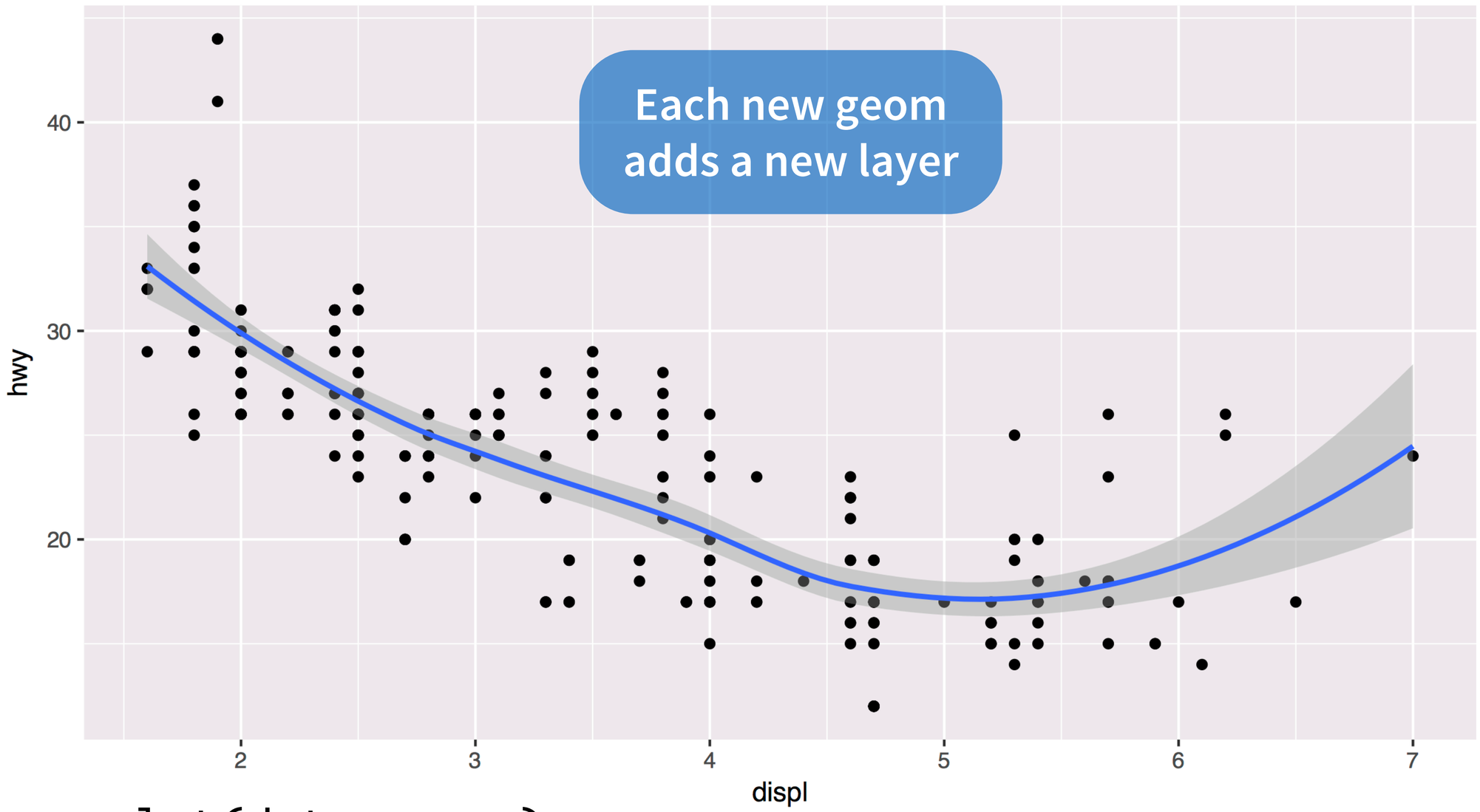
Your Turn 7

With your partner, predict what this code will do.

Then run it.

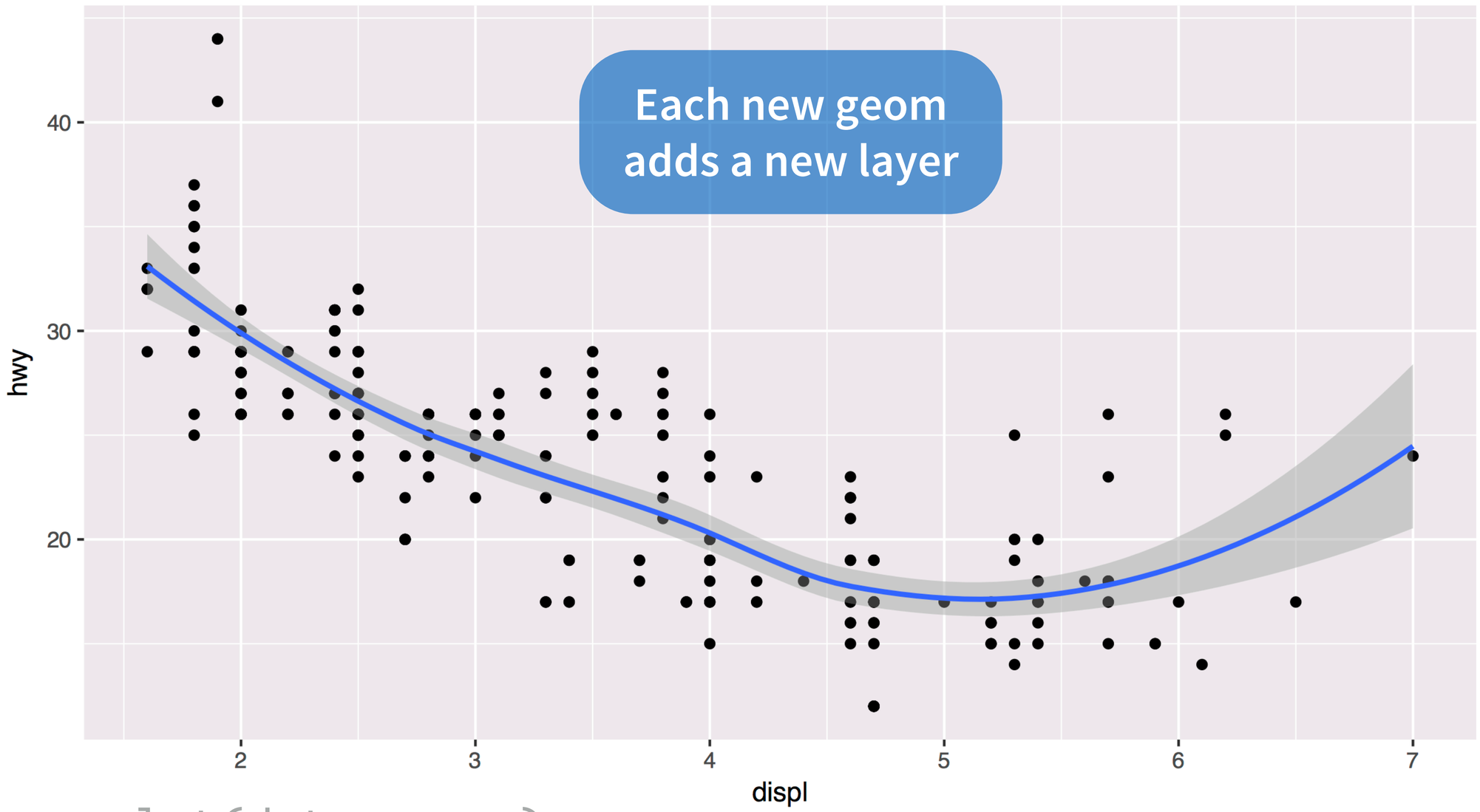
```
ggplot(mpg) +
  geom_point(aes(displ, hwy)) +
  geom_smooth(aes(displ, hwy))
```

CC BY-SA RStudio




```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  geom_smooth(mapping = aes(x = displ, y = hwy))
```

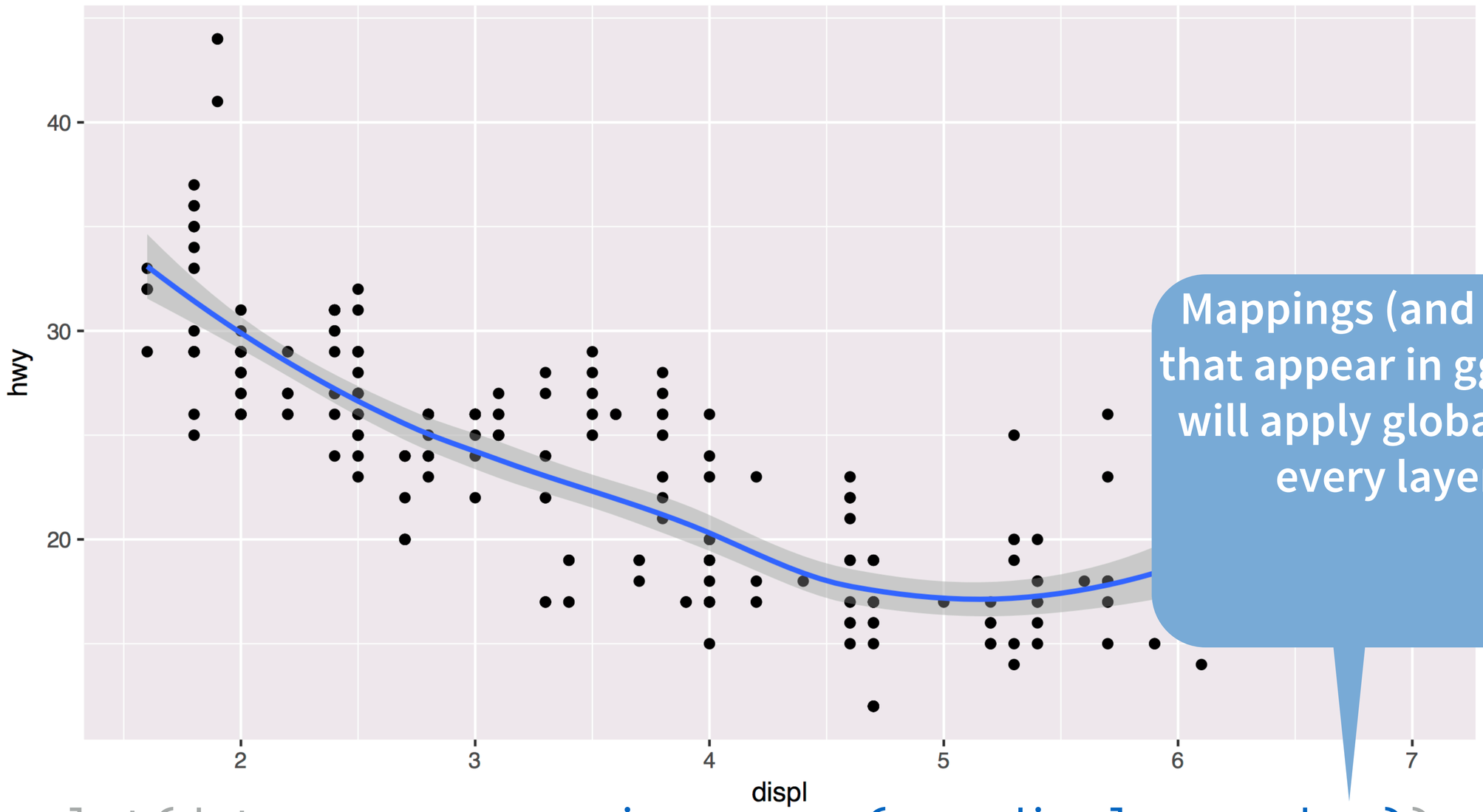
CC BY-SA RStudio




```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  geom_smooth(mapping = aes(x = displ, y = hwy))
```

CC BY-SA RStudio

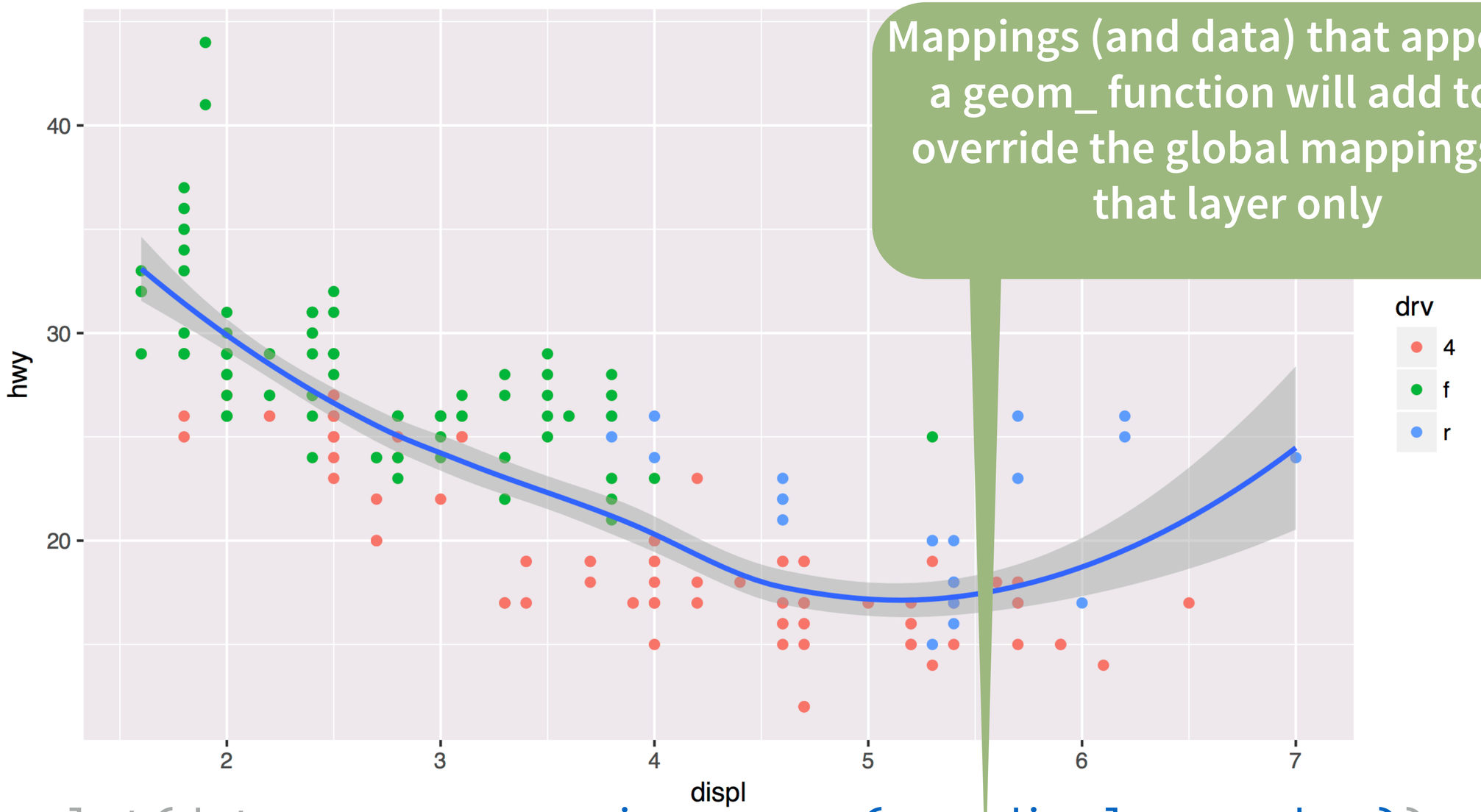
global vs. local



Mappings (and data) that appear in ggplot() will apply globally to every layer

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_point() +
  geom_smooth()
```

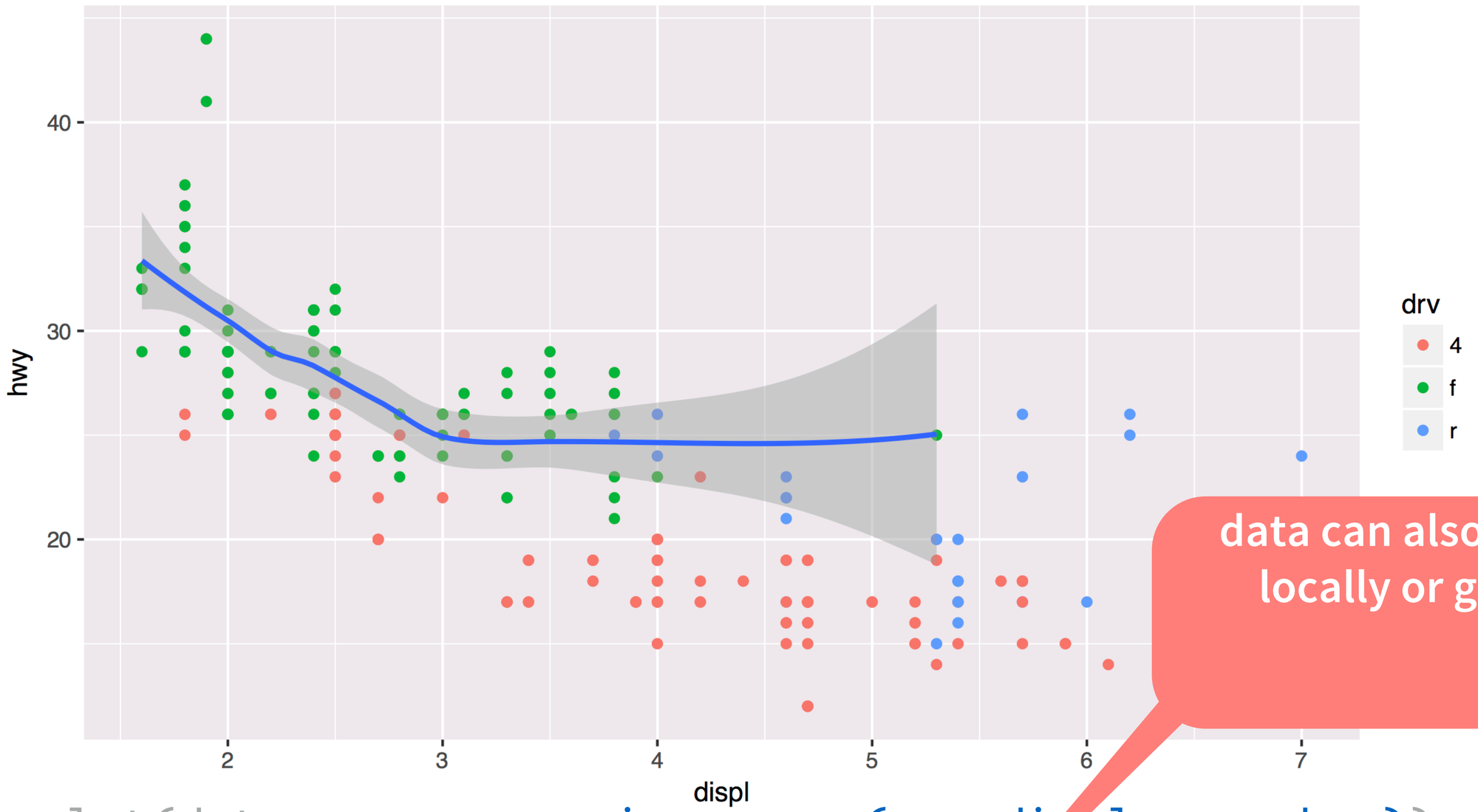
CC BY-SA RStudio



Mappings (and data) that appear in a geom_ function will add to or override the global mappings for that layer only

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_point(mapping = aes(color = drv)) +
  geom_smooth()
```

CC BY-SA RStudio



data can also be set locally or globally

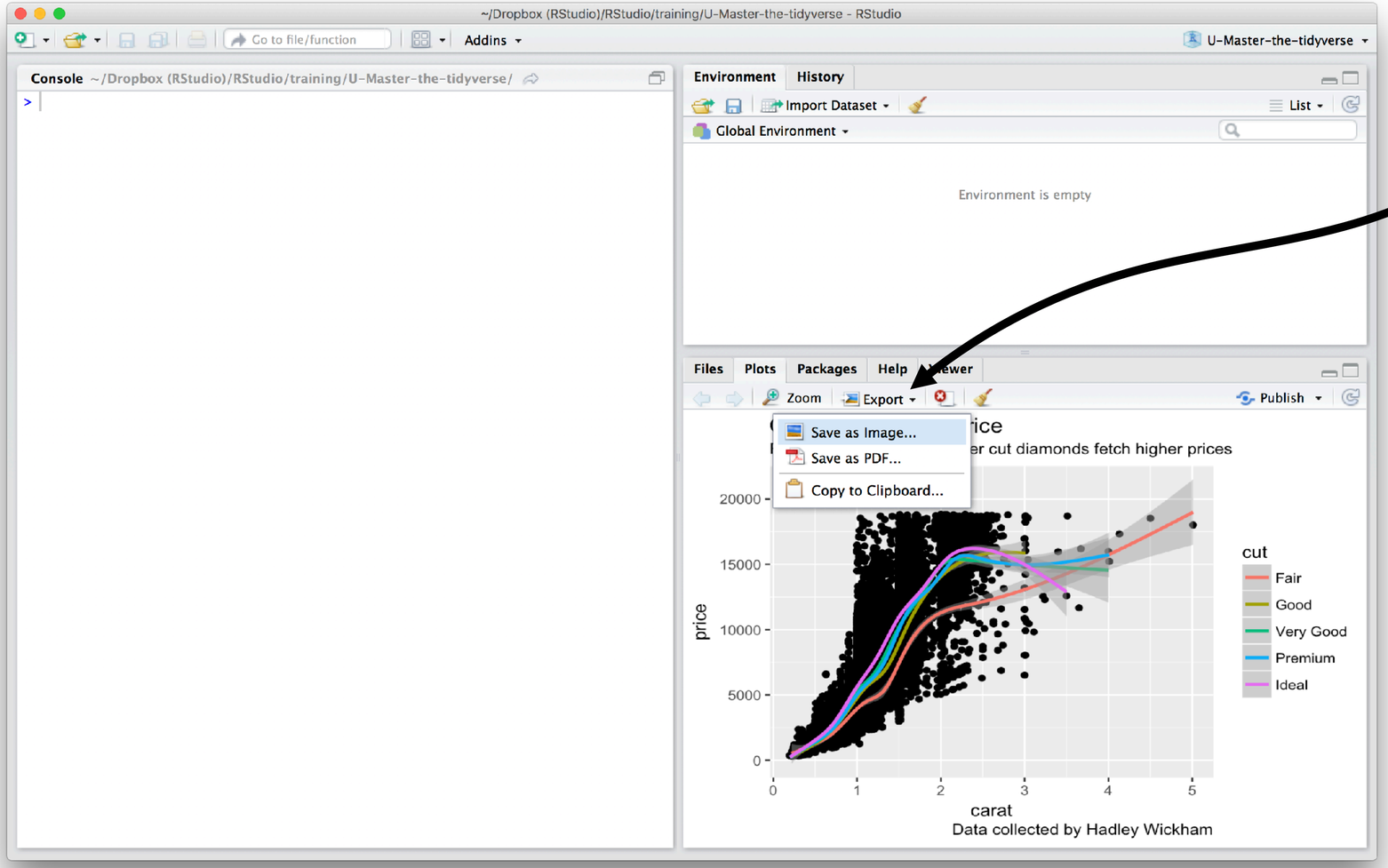
```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_point(mapping = aes(color = drv)) +
  geom_smooth(data = filter(mpg, drv == "f"))
```

CC BY-SA RStudio

Saving graphs

Manually saving plots

Save plots manually with the export menu



CC BY-SA RStudio

Your Turn 8

What does this command return?

```
getwd()
```

CC BY-SA RStudio

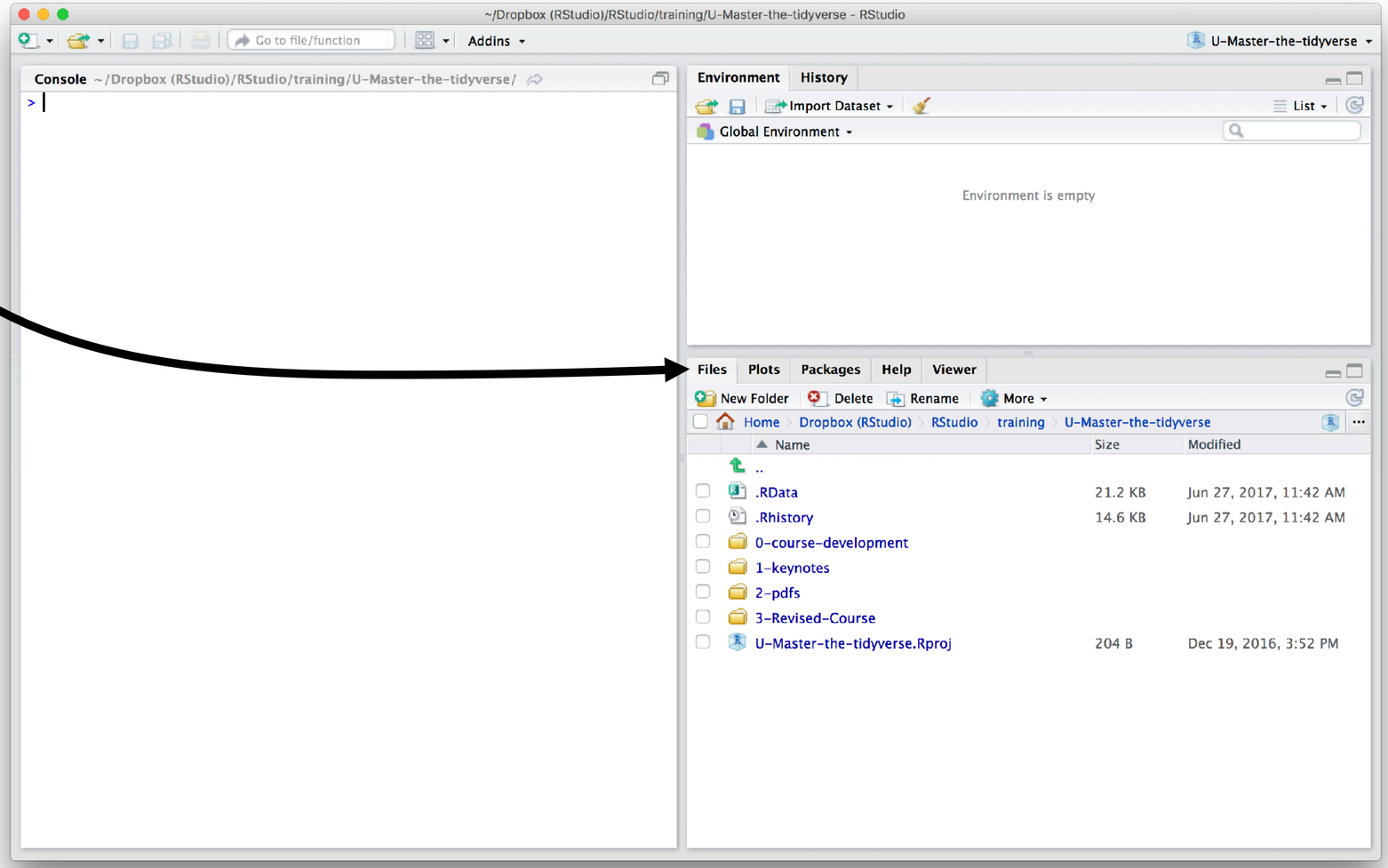
Working directory

R associates itself with a folder (i.e. directory) on your computer.

- This folder is known as your "working directory"
- When you save files, R will save them here
- When you load files, R will look for them here

CC BY-SA RStudio

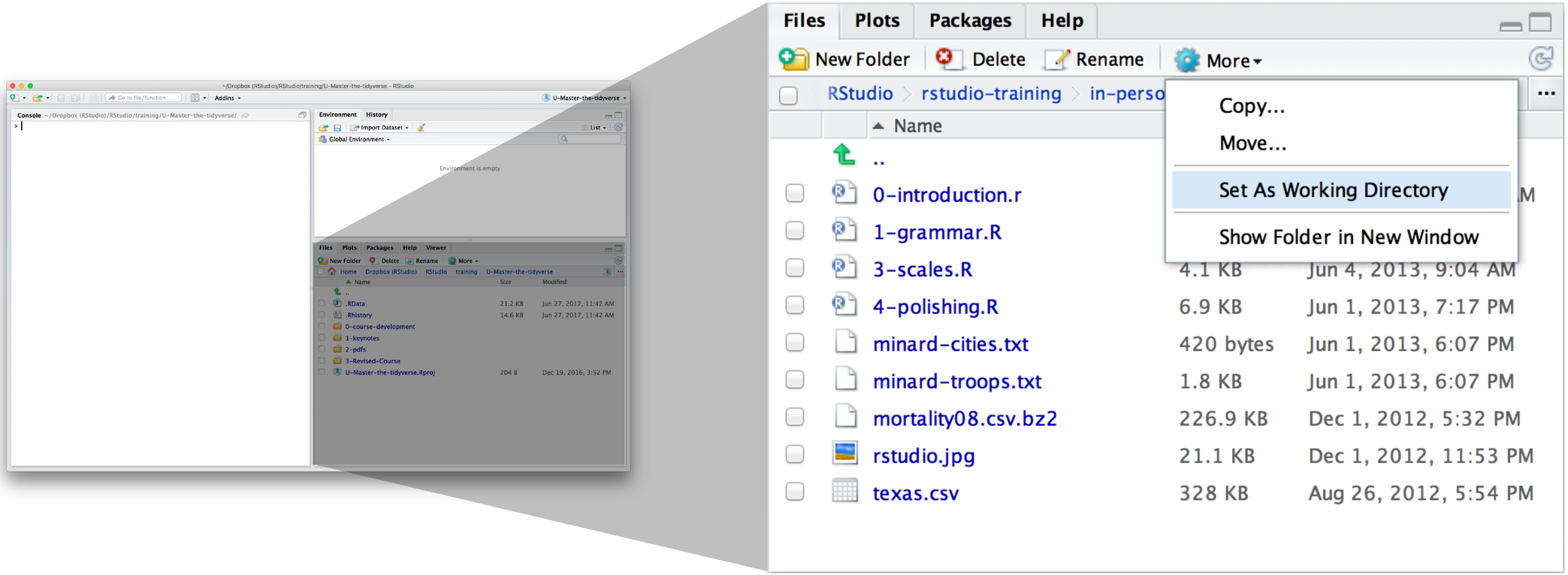
The files pane of the IDE displays the contents of your working directory



CC BY-SA RStudio

Changing the Working directory

Navigate in the files pane to a new directory. Click More > Set As Working Directory



CC BY-SA RStudio

Saving plots

ggsave() saves the last plot.
Uses size on screen:

```
ggsave("my-plot.pdf")
ggsave("my-plot.png")
```

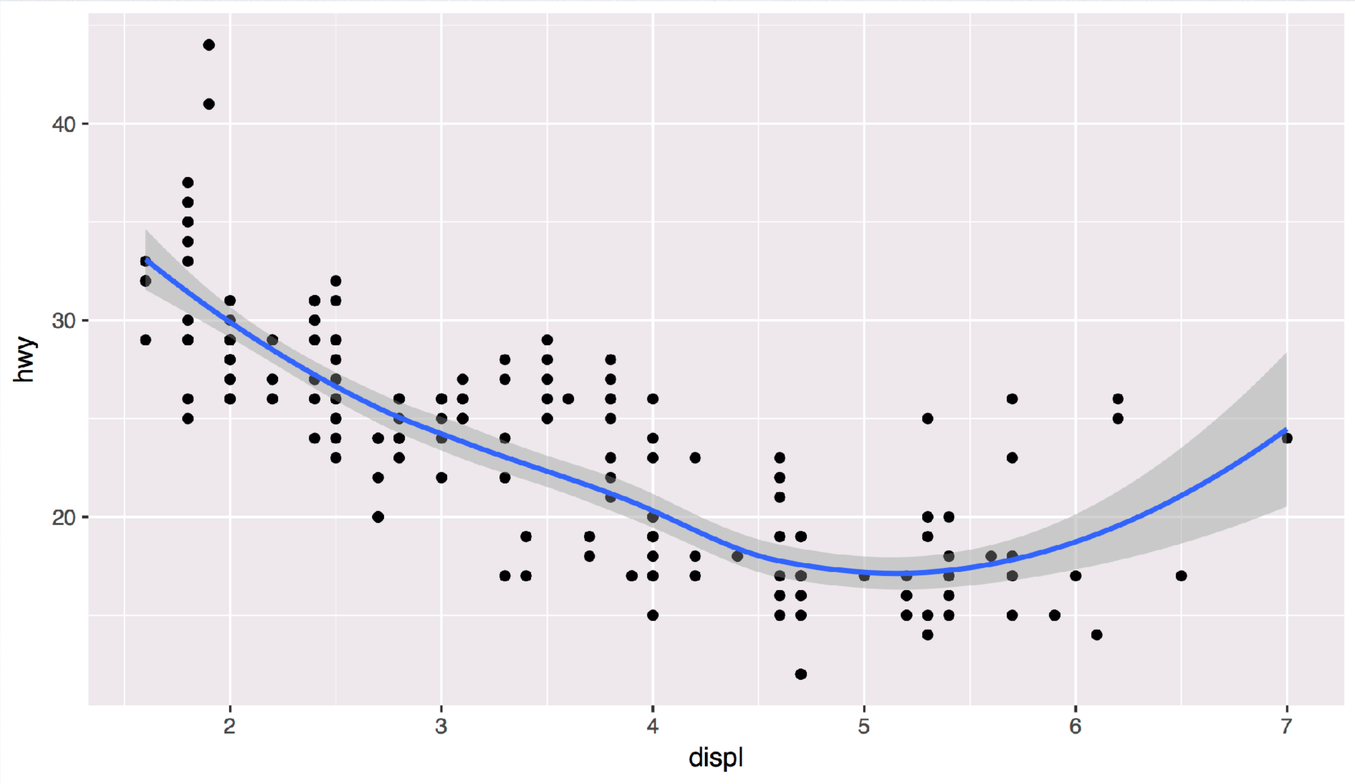
Specify size in inches

```
ggsave("my-plot.pdf", width = 6, height = 6)
```

CC BY-SA RStudio

Your Turn 9

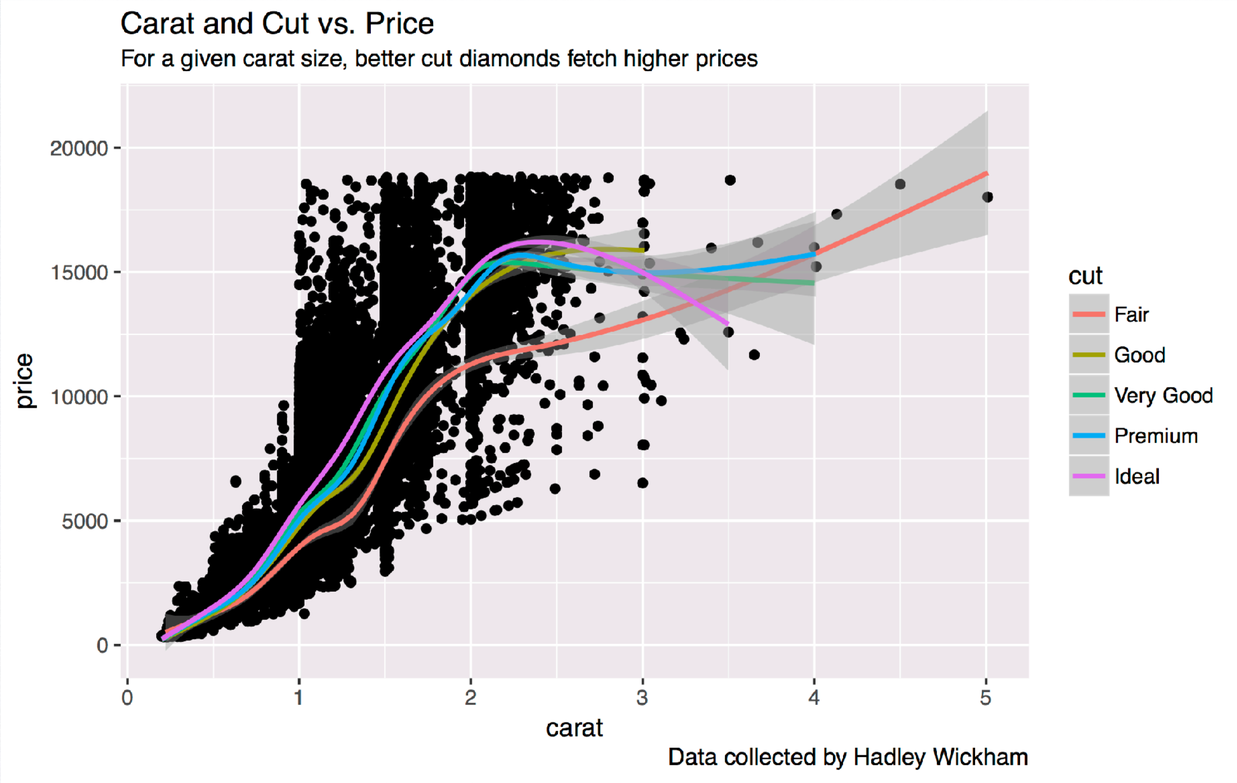
Save your last plot and then locate it in your files pane. (You may have to refresh the files list).



CC BY-SA RStudio

Your turn

Save your last plot and then locate it in your files pane.

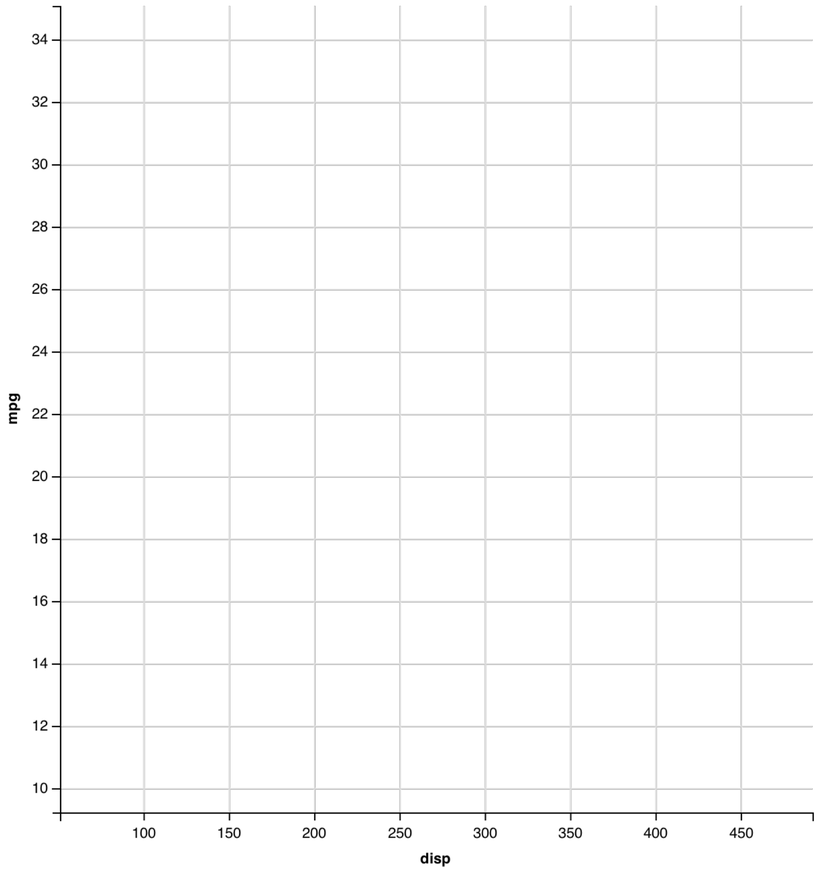


CC BY-SA RStudio

Grammar of Graphics

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

data **geom**

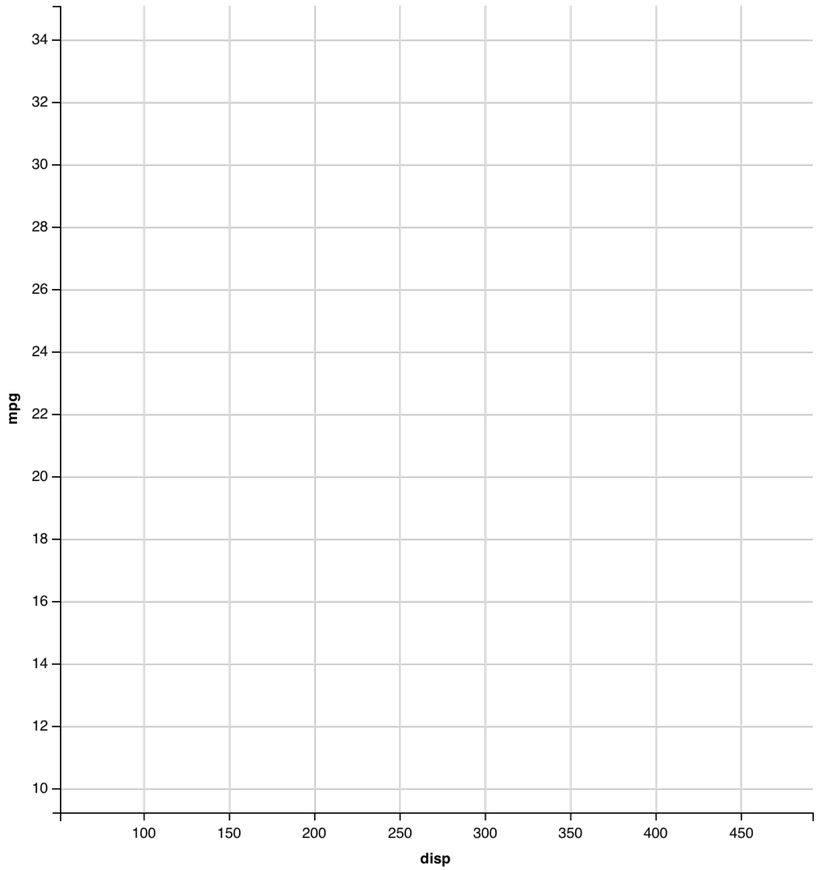


CC BY-SA RStudio

mappings

fill

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1



data geom

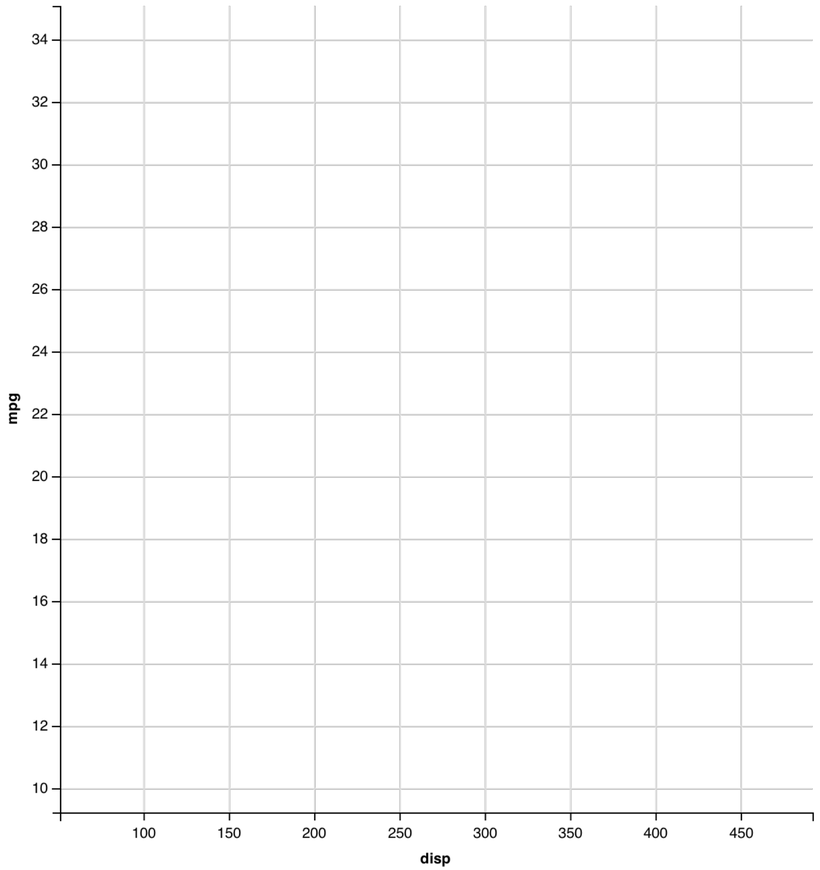
CC BY-SA RStudio

mappings

shape fill

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

data geom



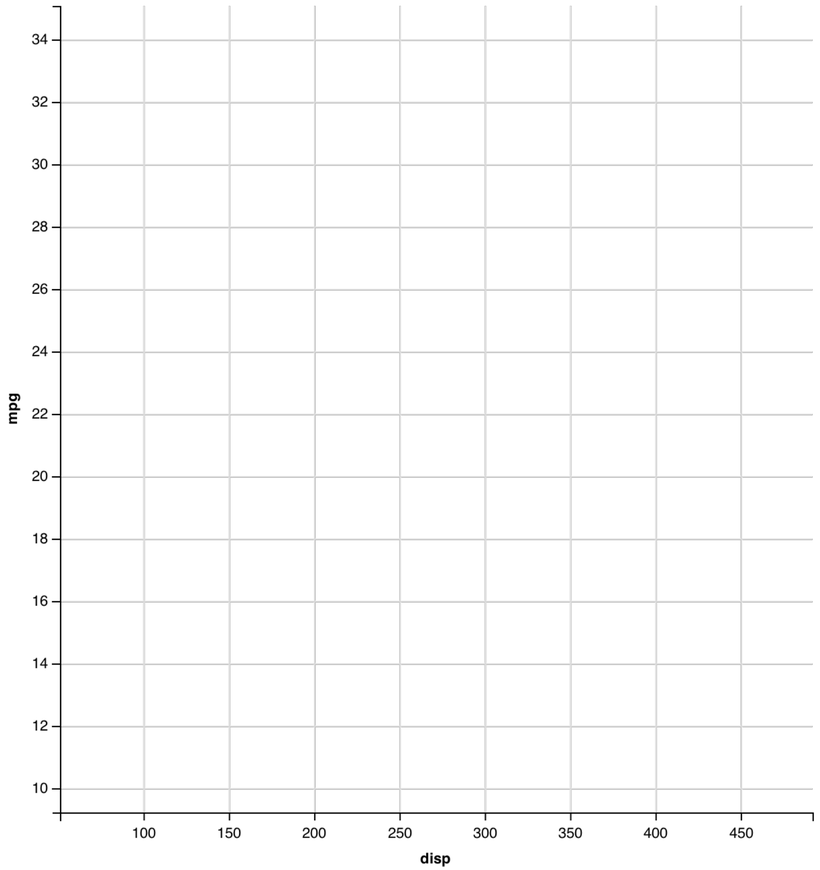
CC BY-SA RStudio

mappings

shape x fill

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

data geom

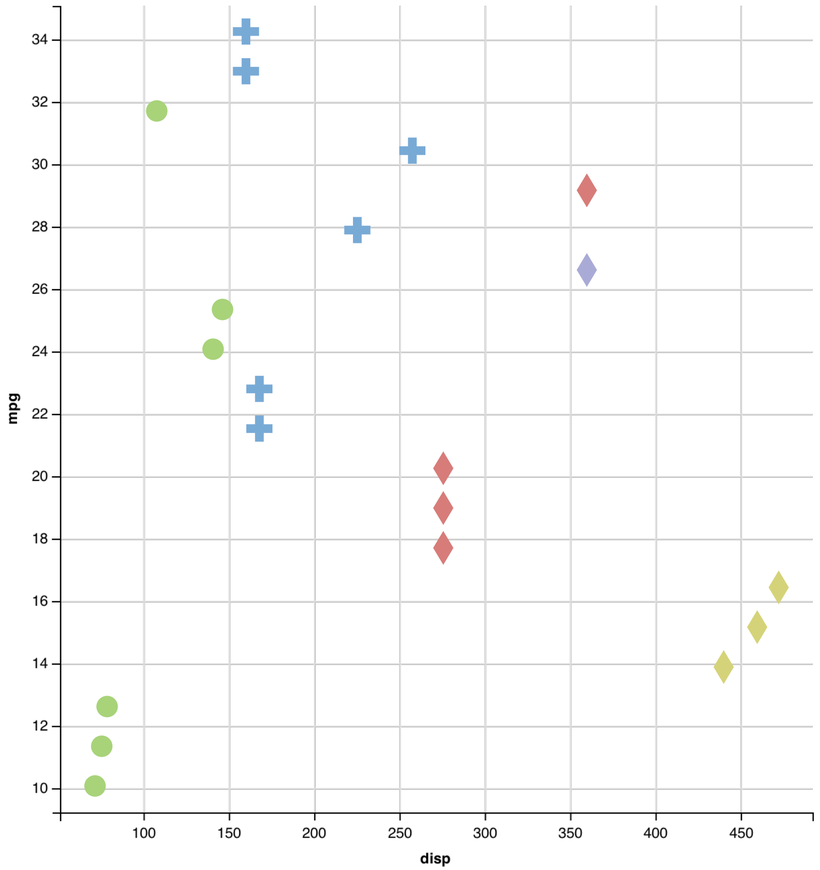


CC BY-SA RStudio

mappings

y	shape	x	fill
mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

data **geom**



CC BY-SA RStudio

mappings

y shape x fill

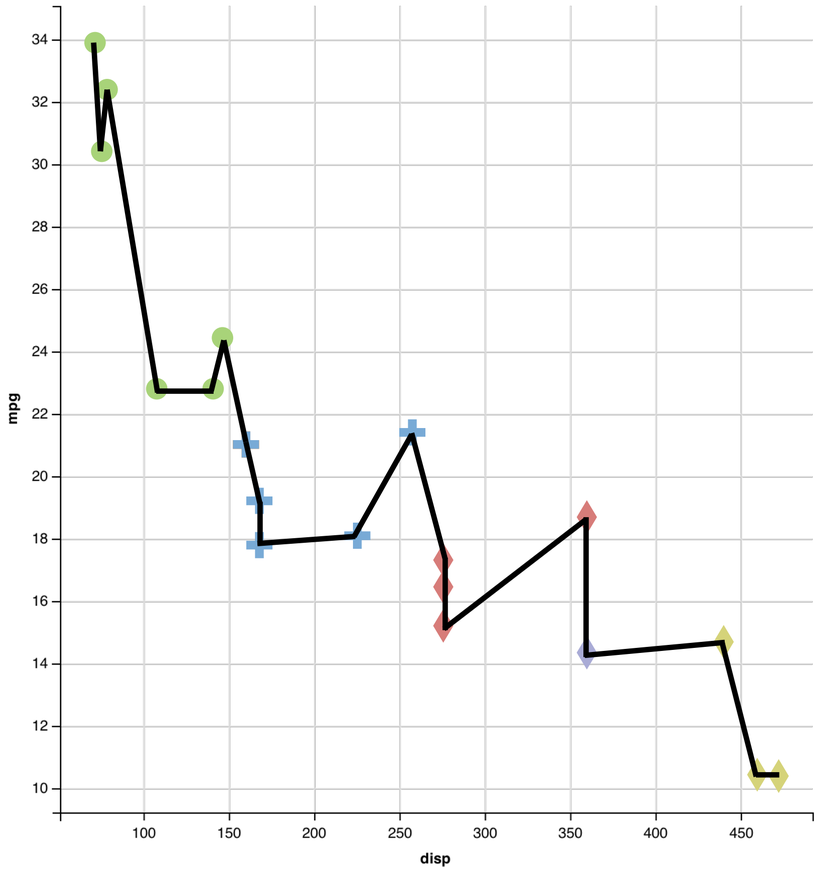
mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

data

geom

points

lines



CC BY-SA RStudio

mappings

y x

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

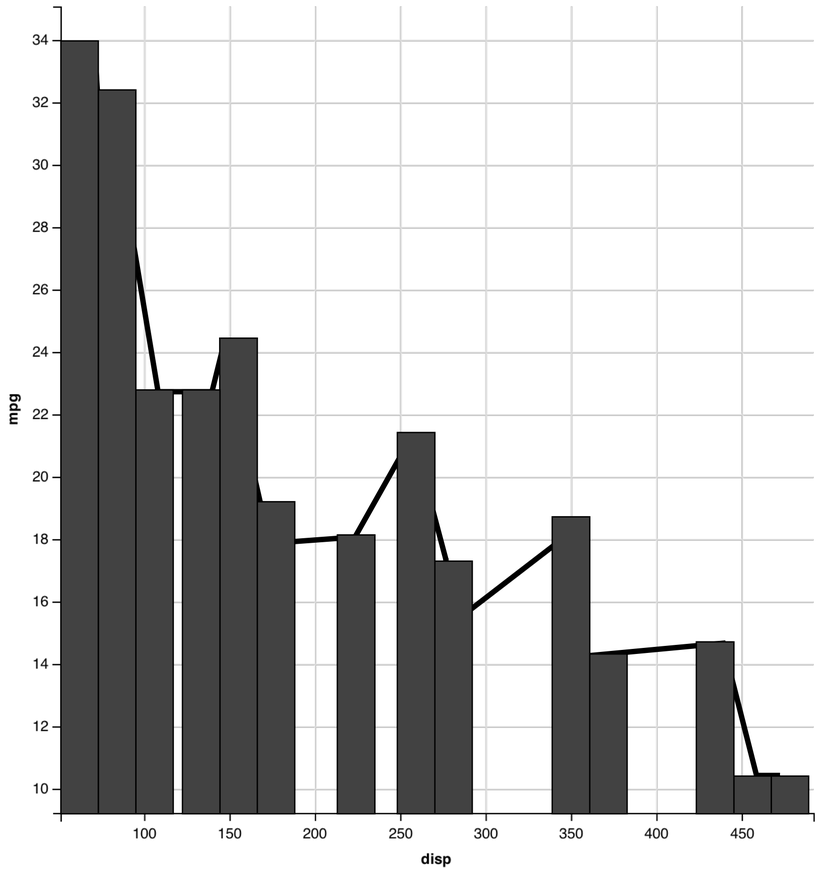
data

geom

points

lines

bars



CC BY-SA RStudio

mappings

y fill

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

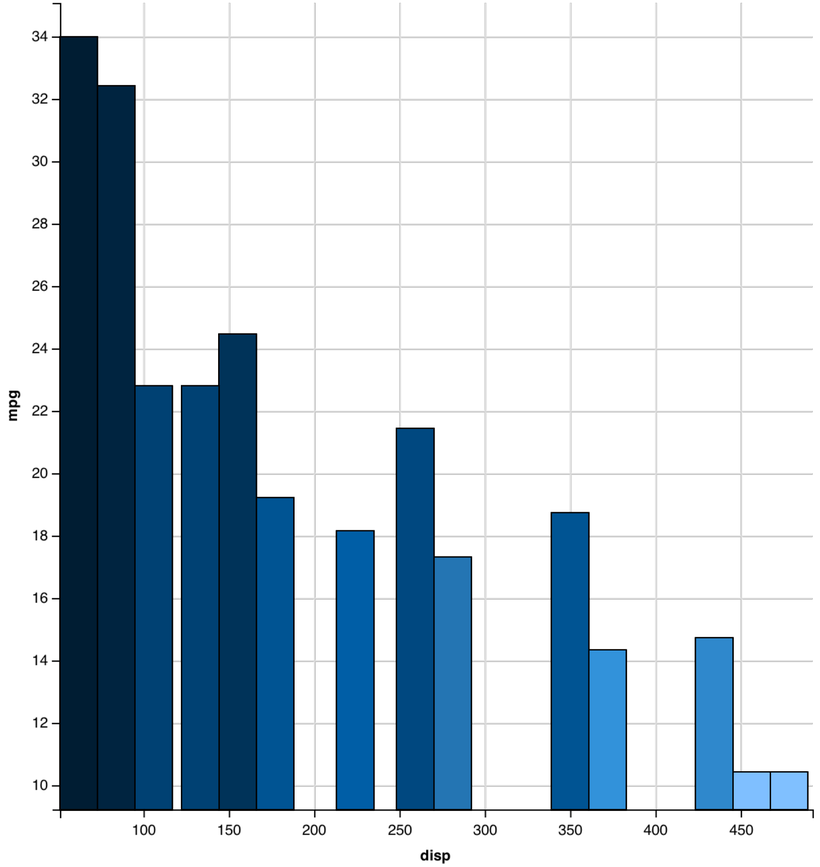
data

geom

points

lines

bars



CC BY-SA RStudio

To make a graph

[template]

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

CC BY-SA RStudio

To make a graph

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

data

1. Pick a **data** set

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

CC BY-SA RStudio

To make a graph

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

data **geom**

1. Pick a **data** set

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

2. Choose a **geom** to display cases

CC BY-SA RStudio

To make a graph

mappings

fill

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

data **geom**

1. Pick a **data** set

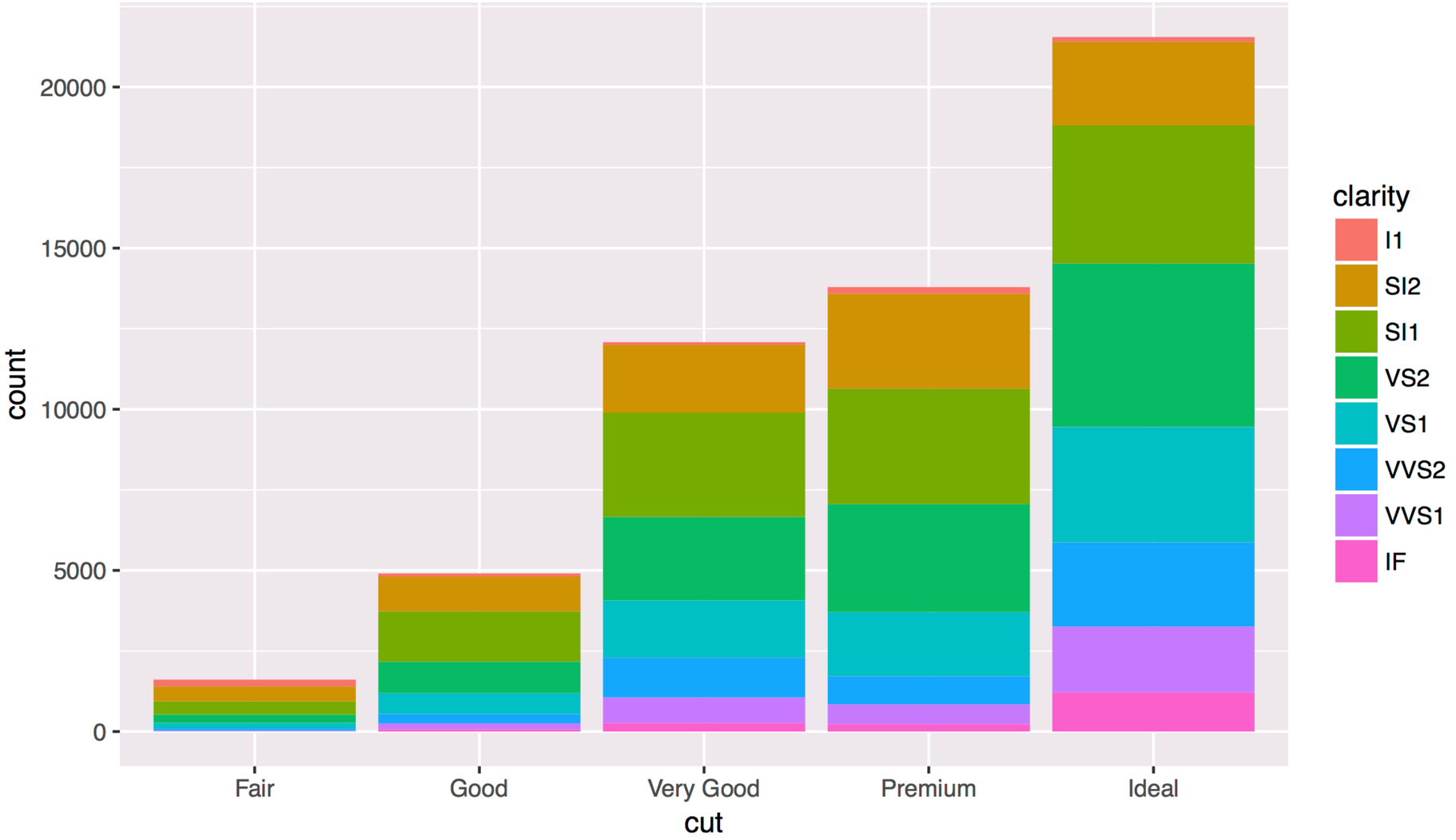
```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

2. Choose a **geom** to display cases

3. **Map** aesthetic properties to variables

CC BY-SA RStudio

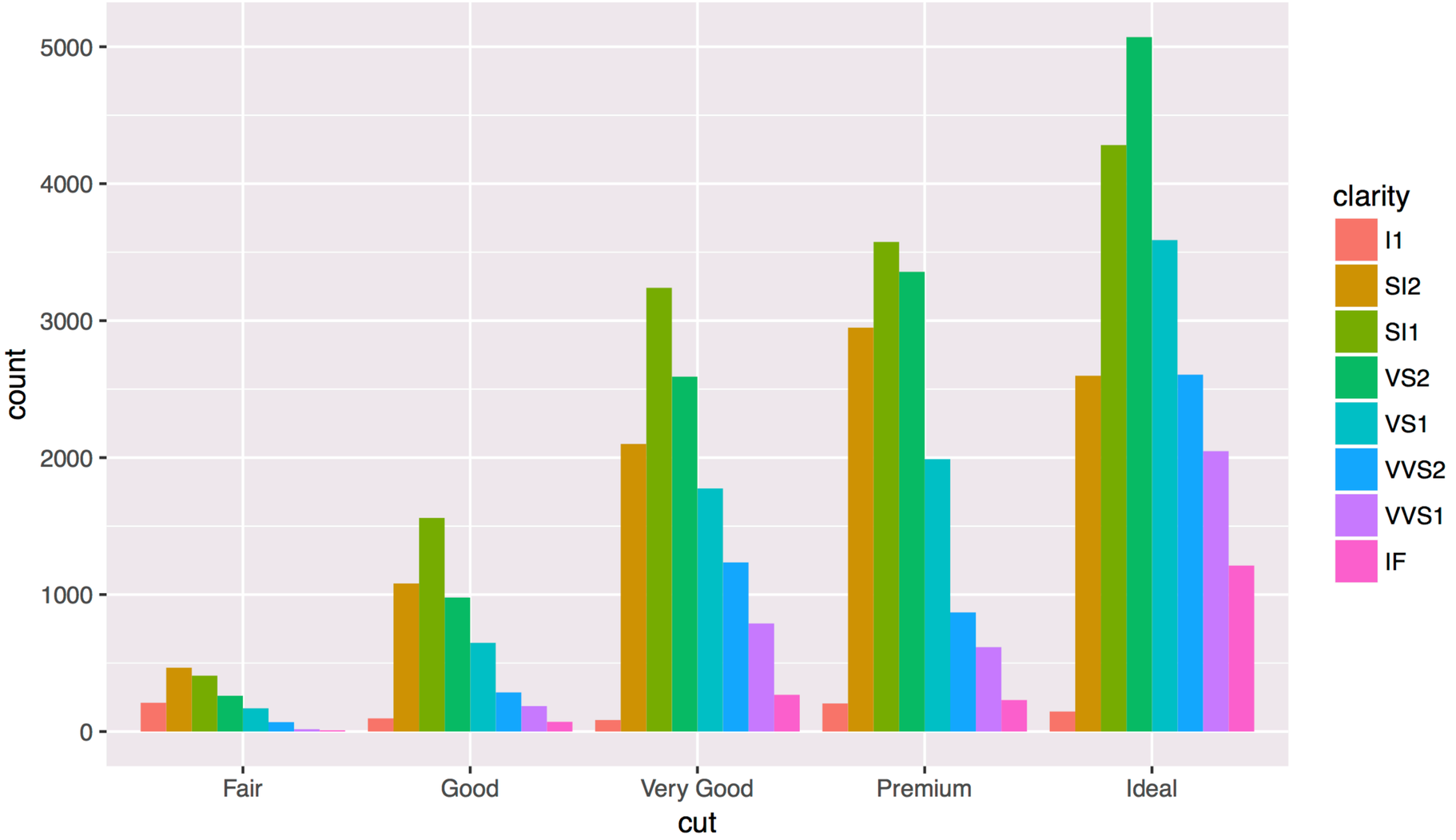
What else?



CC BY-SA RStudio

Position Adjustments

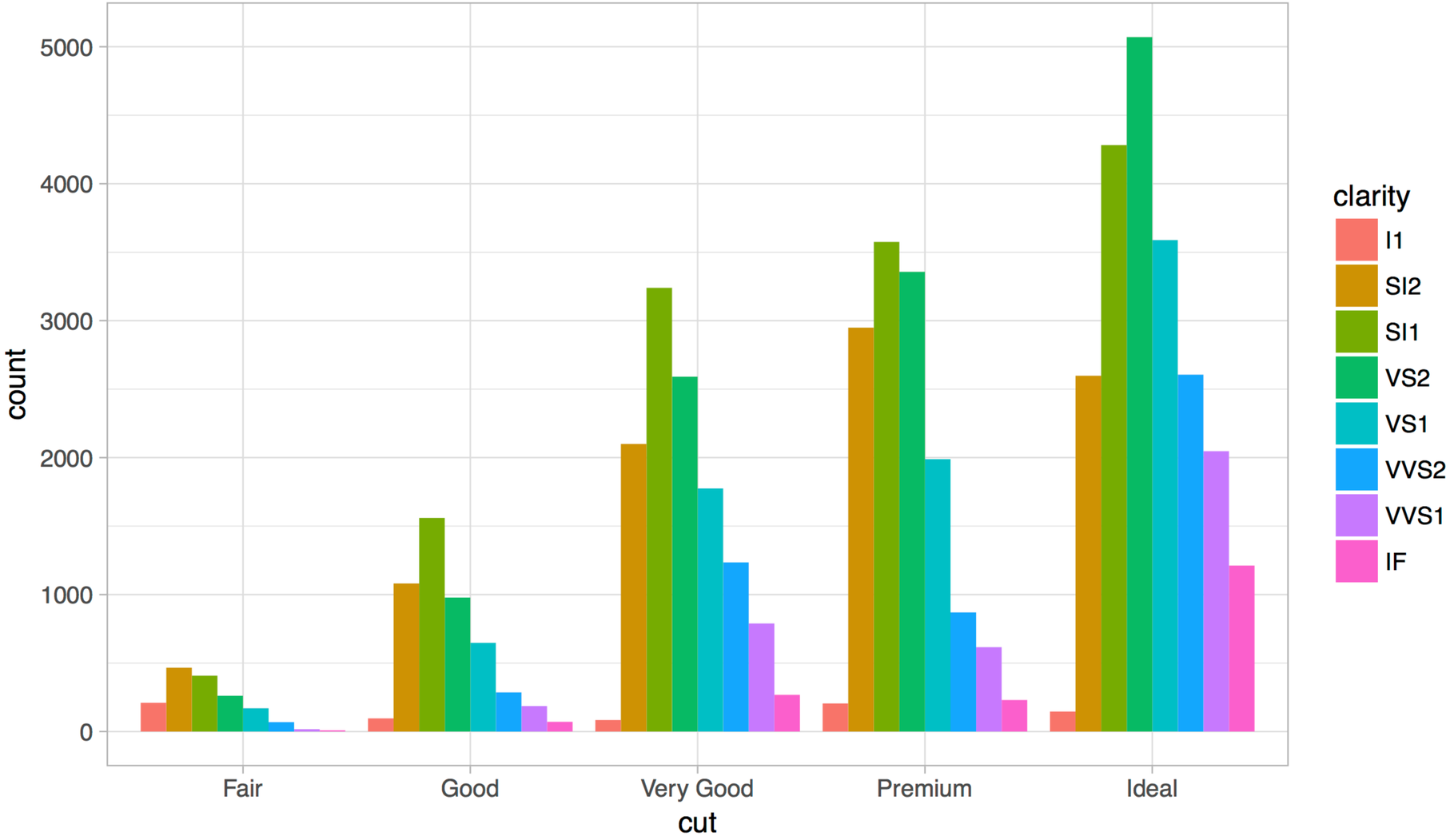
How overlapping objects are arranged



CC BY-SA RStudio

Themes

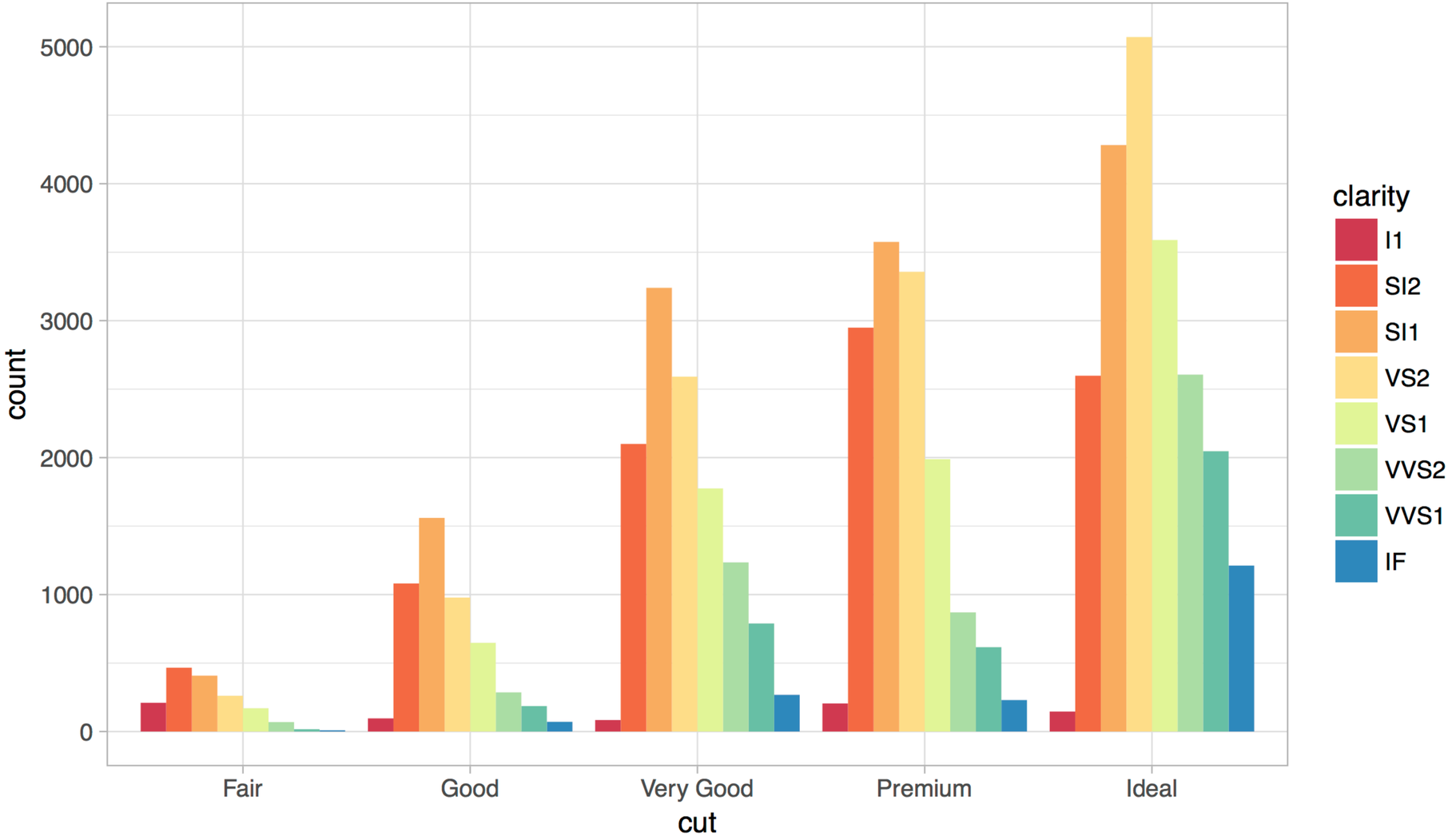
Visual appearance of non-data elements



CC BY-SA RStudio

Scales

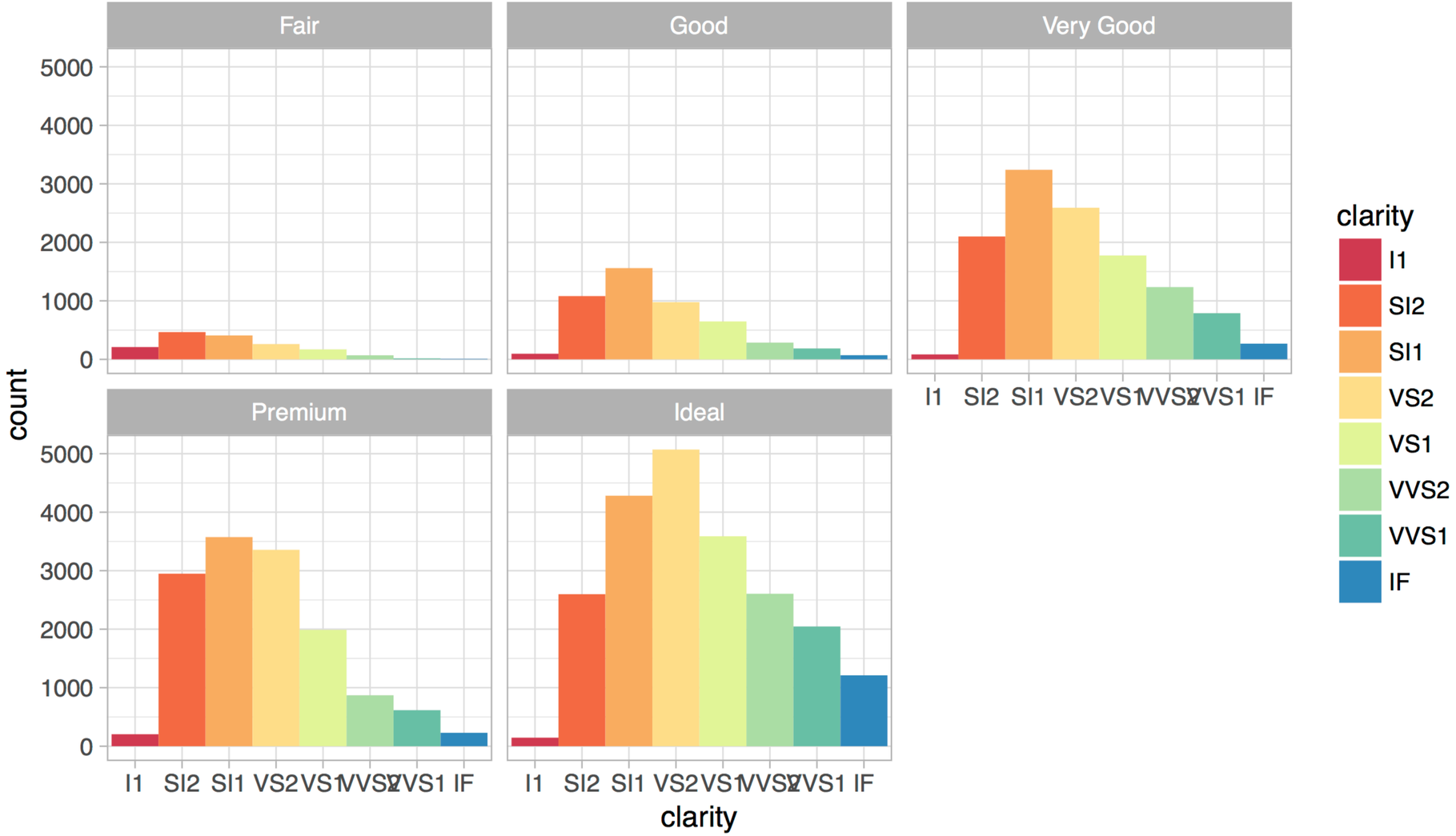
Customize color scales, other mappings



CC BY-SA RStudio

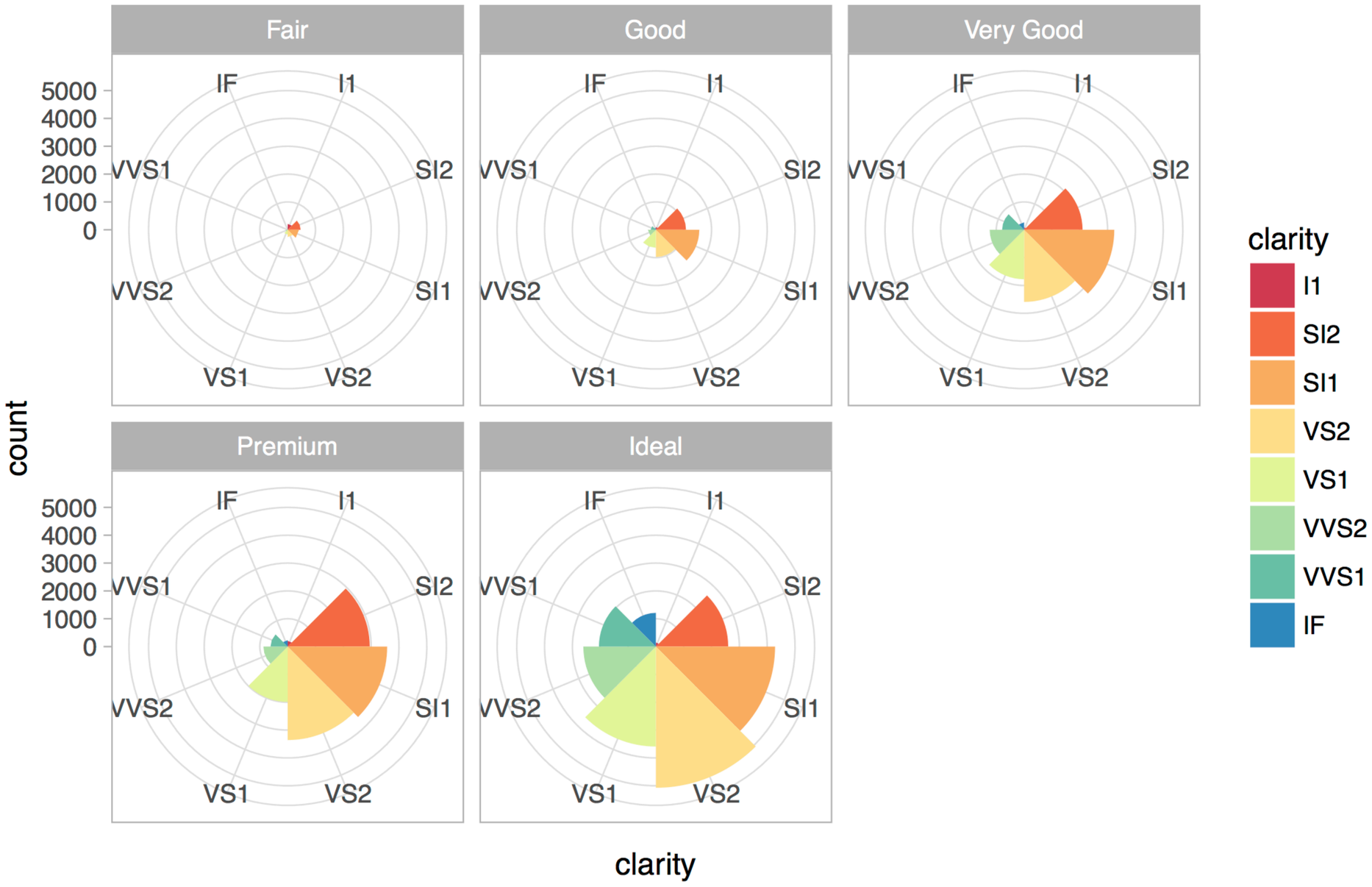
Facets

Subplots that display subsets of the data.



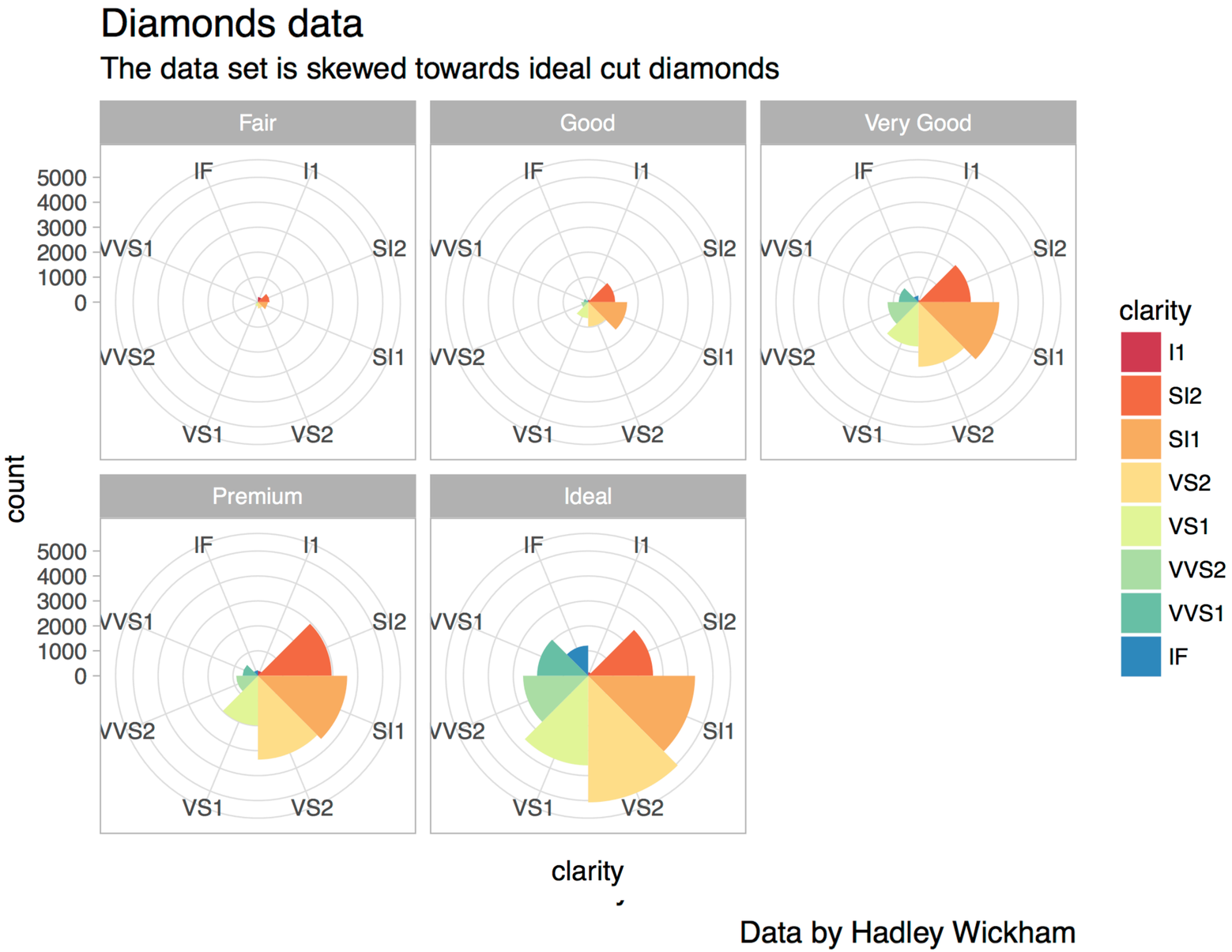
CC BY-SA RStudio

Coordinate systems



CC BY-SA RStudio

Titles and captions



CC BY-SA RStudio

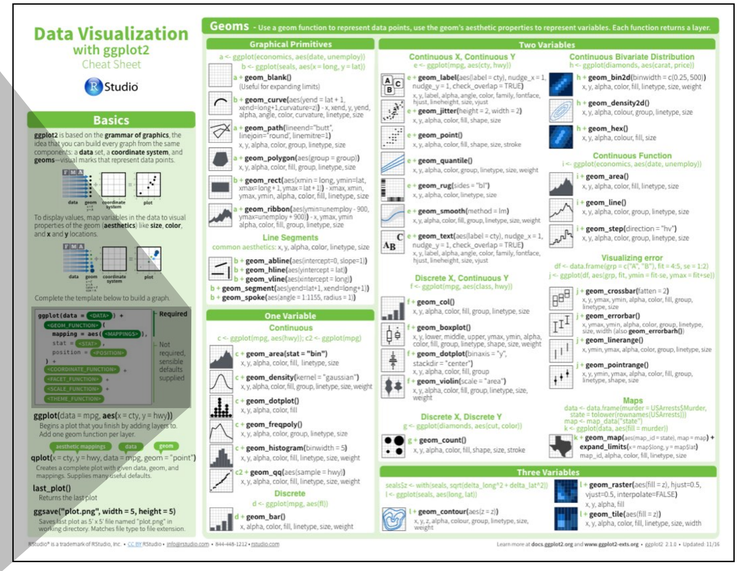
A ggplot2 template

Make any plot by filling in the parameters of this template

```
ggplot(data = <DATA>) +
  <GEOM_FUNCTION> (
    mapping = aes(<MAPPINGS>),
    stat = <STAT>,
    position = <POSITION>
  ) +
  <COORDINATE_FUNCTION> +
  <FACET_FUNCTION> +
  <SCALE_FUNCTION> +
  <THEME_FUNCTION>
```

Required

Not required, sensible defaults supplied



CC BY-SA RStudio

ggplot2.tidyverse.org

Create Elegant Data Visualisati ×

Garrett

ggplot2.tidyverse.org

ggplot2 part of the tidyverse

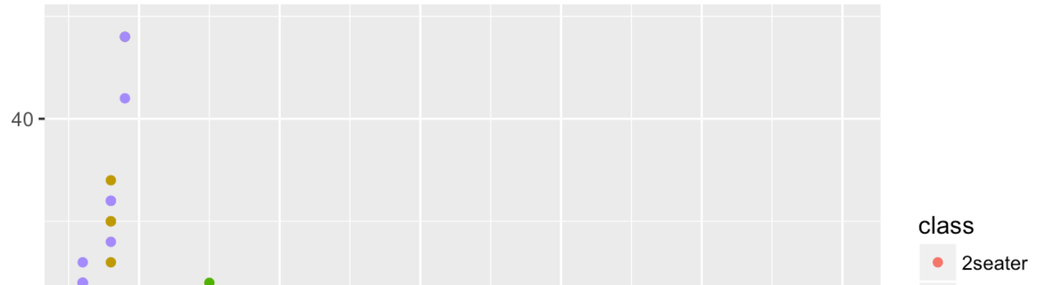
Reference Articles News

Usage

It's hard to succinctly describe how ggplot2 works because it embodies a deep philosophy of visualisation. However, in most cases you start with `ggplot()`, supply a dataset and aesthetic mapping (with `aes()`). You then add on layers (like `geom_point()` or `geom_histogram()`), scales (like `scale_colour_brewer()`), faceting specifications (like `facet_wrap()`) and coordinate systems (like `coord_flip()`).

```
library(ggplot2)

ggplot(mpg, aes(displ, hwy, colour = class)) +
  geom_point()
```



Links

Download from CRAN at https://cran.r-project.org/package=ggplot2

Browse source code at https://github.com/tidyverse/ggplot2

Report a bug at https://github.com/tidyverse/ggplot2/issues

Learn more at http://r4ds.had.co.nz/data-visualisation.html

License

GPL-2 | file LICENSE

Developers

Hadley Wickham

Author, maintainer

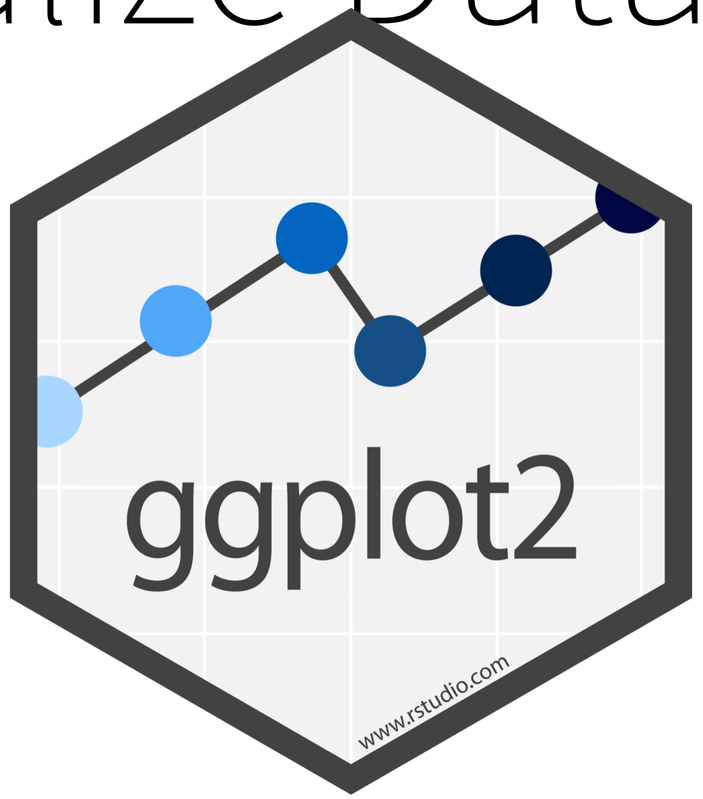
CC BY-SA RStudio

Your Turn

Open **02-Transform-Data.Rmd**.

CC BY-SA RStudio

Visualize Data with



CC BY-SA RStudio