

**R4DS**

Cohort 4

Wed 6:00 – 7:00 US Central

Twitter: @Rspjut

# 5-MINUTE ICE BREAKER

What's your favorite (or most recent or most anticipated) road trip?

# AGENDA

- 5-Minute Ice breaker
- Quick Housekeeping Reminders
- Last Week...
- Data Visualization
- Exercises
- Statistical Transformation
- Getting Help
- Next Week

# QUICK HOUSEKEEPING REMINDERS

- Video camera is optional, but encouraged.
- I purposely err on the side of going fast. Slowing me down does not hurt my feelings.
- Take time to learn the theory (Grammar of Graphics, Tidy Data whitepaper, Relational Database theory, etc.).
- Please do the chapter exercises. Second-best learning opportunity!
- Please plan on teaching one of the lessons. Best learning opportunity!

# LAST WEEK...

The following code draws a graph of sample data that comes with the Tidyverse package.

To make this work, you need to:

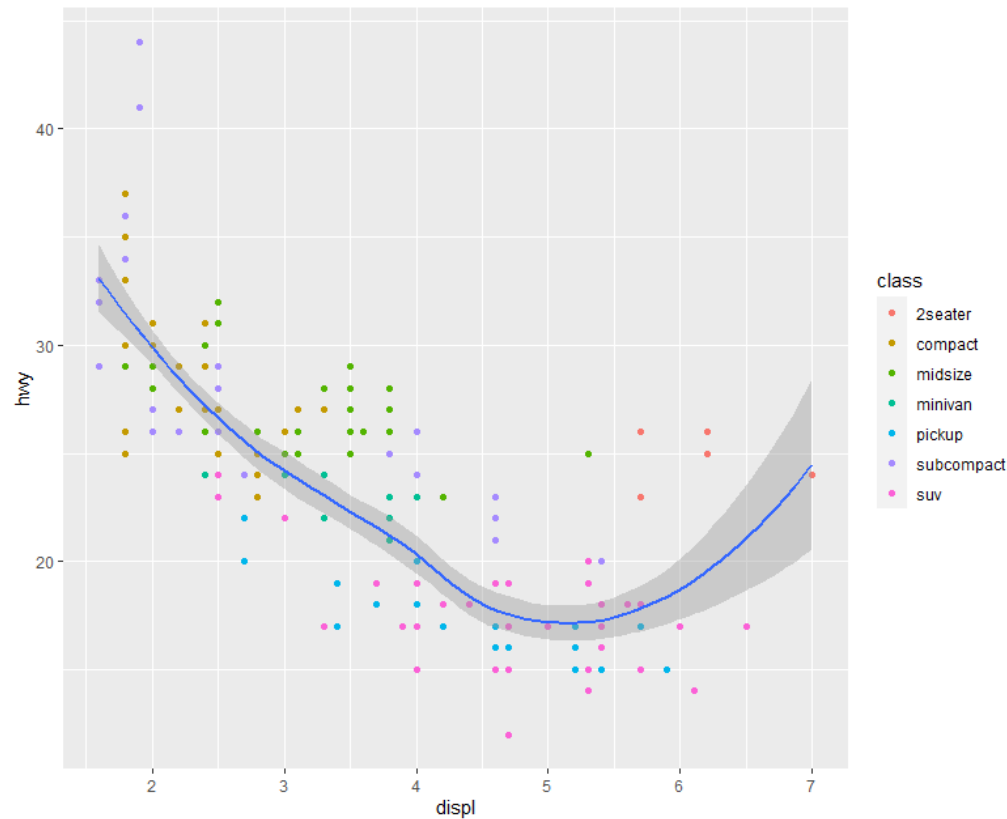
Action	Frequency	Code
Install the Tidyverse package	One-time action	<code>install.packages("tidyverse")</code>
Load the Tidyverse package	Each new R session	<code>library(tidyverse)</code>

Type this into the Source pane. Case/capitalization matters!

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

# LAST WEEK...

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



# CHAPTER 3 — DATA VISUALIZATION

- Data source is the `mpg` data set
- Draw this week's (simplified) plot

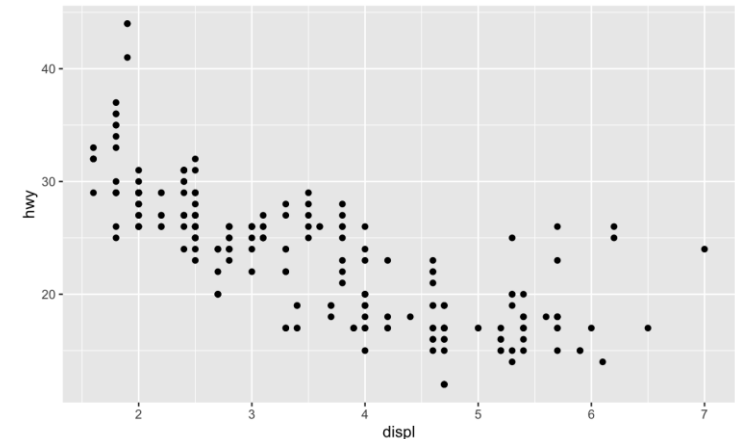
Last Week:

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

This Week:

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

- Note the differences
  - `geom_smooth` is gone
  - `mapping = aes(color = class)` is missing from `geom_point`
  - `mapping = aes(x = displ, y = hwy)` is with `geom_point`



# GGPLOT(DATASET) + GEOM()

*With ggplot2, you begin a plot with the function ggplot(). ggplot() creates a coordinate system that you can add layers to. The first argument of ggplot() is the dataset to use in the graph. So ggplot(data = mpg) creates an empty graph, but it's not very interesting so I'm not going to show it here.*

*You complete your graph by adding one or more layers to ggplot(). The function geom\_point() adds a layer of points to your plot, which creates a scatterplot. ggplot2 comes with many geom functions that each add a different type of layer to a plot. You'll learn a whole bunch of them throughout this chapter.*

*-Wickham and Grolemund, 3.2*

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

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

## LAYERS!



# GGPLOT(DATASET) + GEOM(AESTHETIC)

*An aesthetic is a visual property of the objects in your plot. Aesthetics include things like the size, the shape, or the color of your points. You can display a point in different ways by changing the values of its aesthetic properties.*

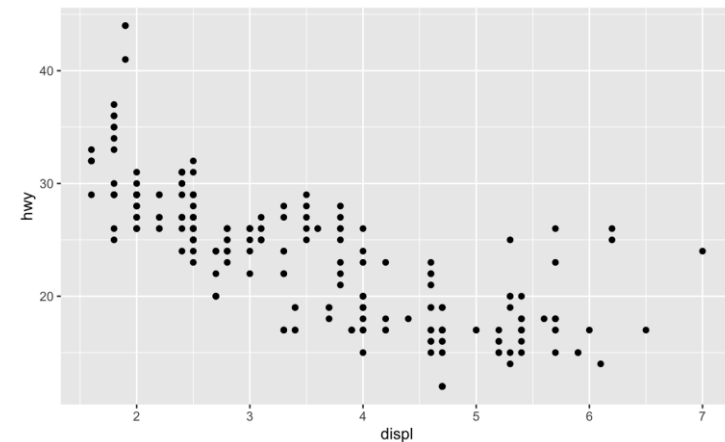
*-Wickham and Grolemund, 3.3*

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

## Aesthetics

`geom_point()` understands the following aesthetics (required aesthetics are in bold):

- **x**
- **y**
- alpha
- colour
- fill
- group
- shape
- size
- stroke



This plot uses defaults for aesthetics

# PRACTICE: GGLOT(DATASET) + GEOM(AESTHETIC)

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

Adjust the code so that the car's class (a variable/column in the dataset) dictates the color of the point

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

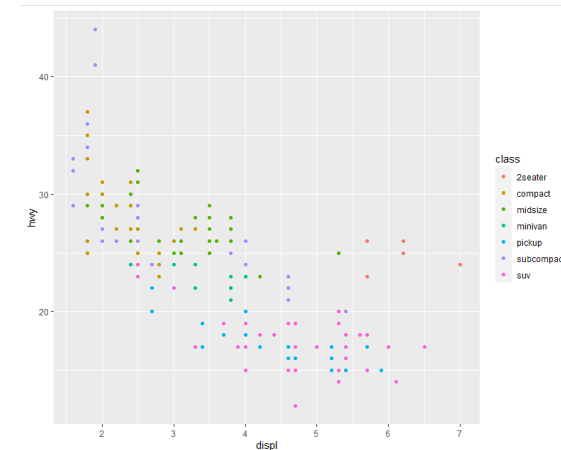
Adjust the code so that the car's fuel type dictates the shape of the point

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

# INSIDE AES() VS OUTSIDE AES()

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

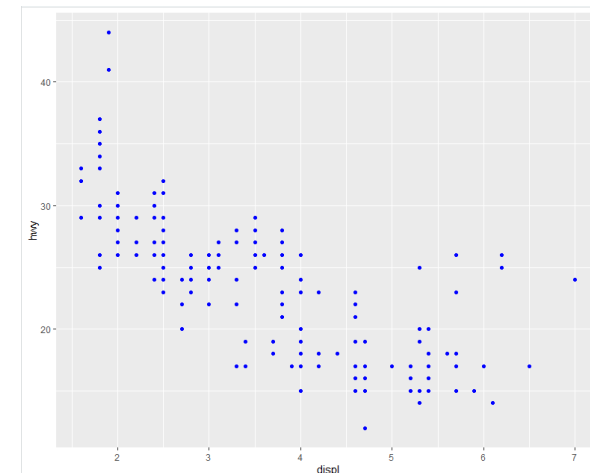
color = class inside the aes()  
argument for the aes()



If you want the aesthetic (i.e., color, size, alpha) to be dictated by a column in your data, then the code goes inside the aes().

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

color = class outside the aes()  
argument for geom\_point



If you want the aesthetic (i.e., color, size, alpha) to be a characteristic of the whole graph layer, then the code goes outside the aes().

# MULTIPLE LAYERS

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

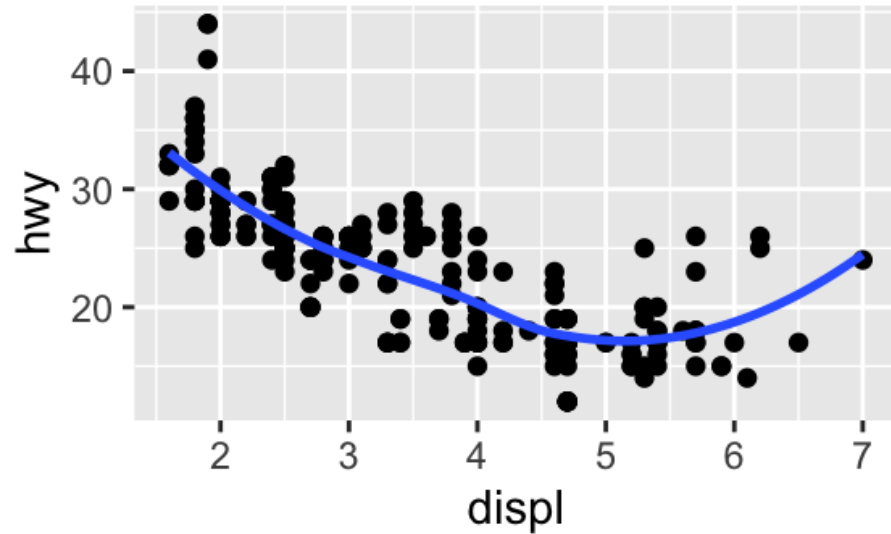
After each plus sign (+) is a new layer.

# GLOBAL AESTHETICS

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

If a mapping will apply to all layers, they can go in the `ggplot()` section.

# EXERCISES



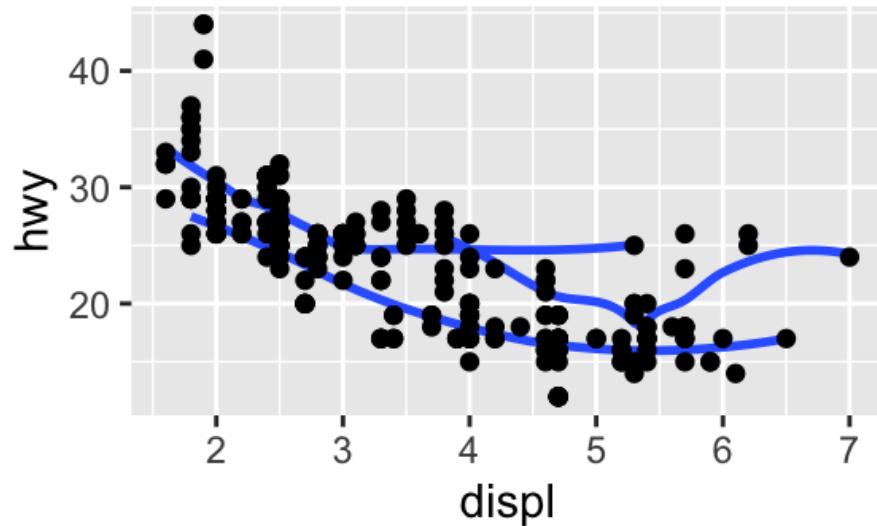
HINTS:

```
line = geom_smooth()  
remove error bands with se = FALSE
```

ANSWER:

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

# EXERCISES



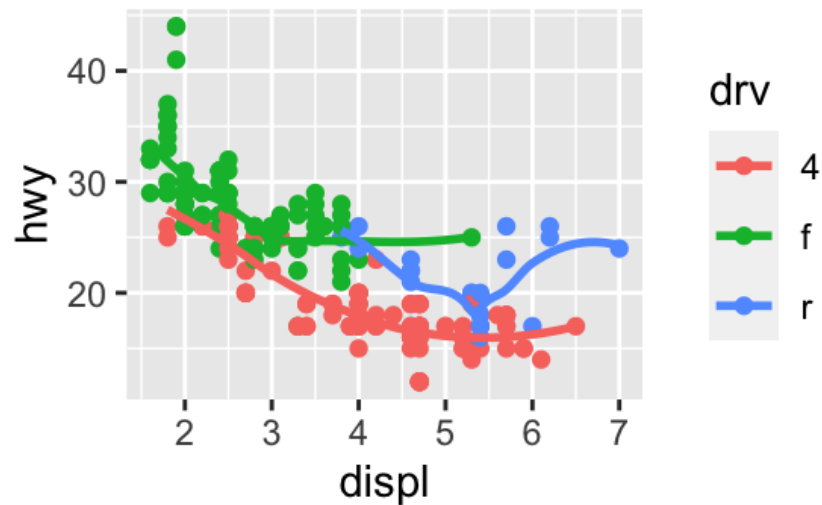
HINTS:

- `line = geom_smooth()`
- remove error bands with `se = FALSE`
- make different lines using `group = drv`
- different groupings are for lines so it's part of `geom_smooth()`
- grouping is mapped to a variable so it's inside the aesthetic

ANSWER:

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

# EXERCISES



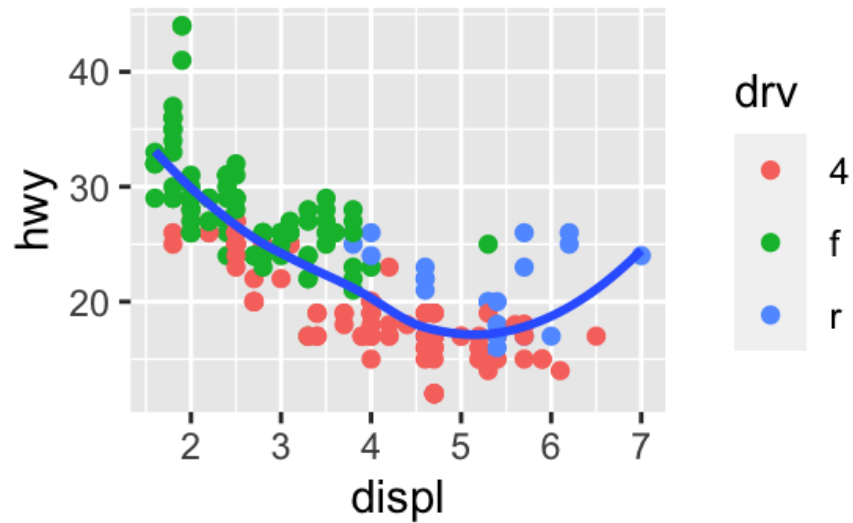
HINTS:

- `line = geom_smooth()`
- remove error bands with `se = FALSE`
- make different colors using `color = drv`
- grouping is mapped to a variable so it's in the aesthetic
- color aesthetic is global to capture point and line (`geom_smooth`)

ANSWER:

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

# EXERCISES



HINTS:

- `line = geom_smooth()`
- remove error bands with `se = FALSE`
- make different colors using `color = drv`
- grouping is mapped to a variable so it's in the aesthetic
- color aesthetic is only for points, not for the line

ANSWER:

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



# STATISTICAL TRANSFORMATION

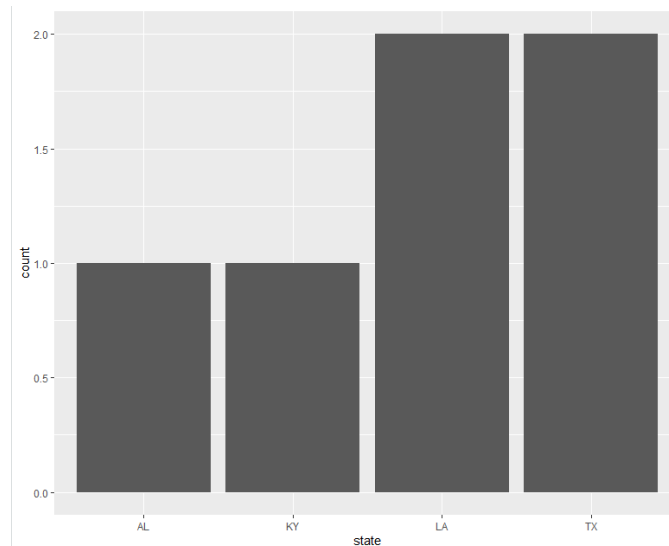
*The algorithm used to calculate new values for a graph is called a stat.*

*-Wickham and Grolemund, 3.7*

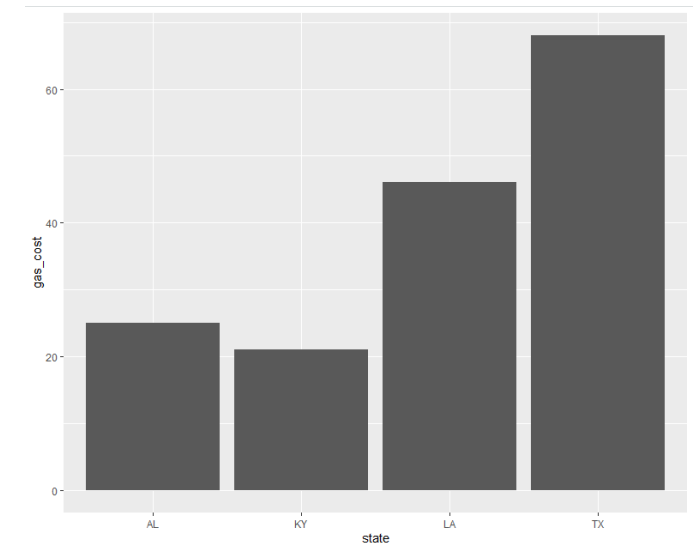
Data set = travel

	stop_num	state	gas_cost
1	1	TX	33
2	2	TX	35
3	3	LA	27
4	4	LA	19
5	5	AL	25
6	6	KY	21

```
ggplot(travel, aes(state)) +  
geom_bar(stat = "count")
```



```
ggplot(travel, aes(state, gas_cost)) +  
geom_bar(stat = "identity")
```



# GETTING HELP

- Ask questions during our call
- Google
- Stack Overflow
- Slack
- Office Hours [r4ds.io/calendar](https://r4ds.io/calendar)
- Twitter [#rstats](https://twitter.com/rstats)
- r4ds answer keys: Jeff Arnold (preferred) or Bryan Shalloway (also good)
- Cheatsheets

# NEXT WEEK...

- Chapter 4: Workflow Basics
- Chapter 5: Data Transformation
- Look over the `nycflights13` data

```
library(tidyverse)
library(nycflights13)
?nycflights13
```
- Volunteers?

