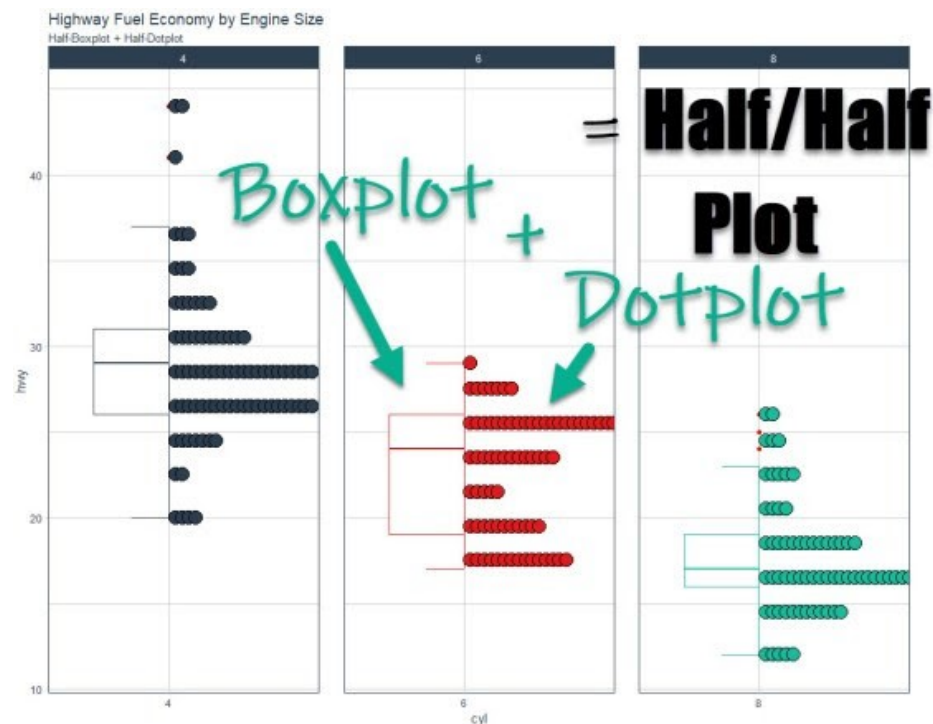


What is gghalves?

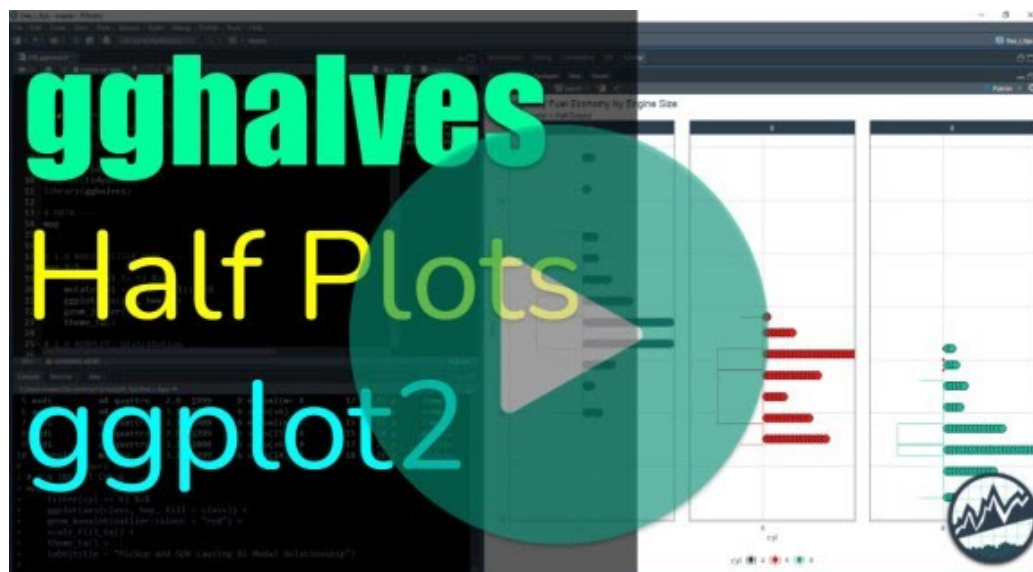
`gghalves` is a new R package that makes it easy to compose your own half-plots using `ggplot2`.



gghalves Video Tutorial

For those that prefer Full YouTube Video Tutorials.

Learn how to use `gghalves` in our free 8-minute YouTube video.

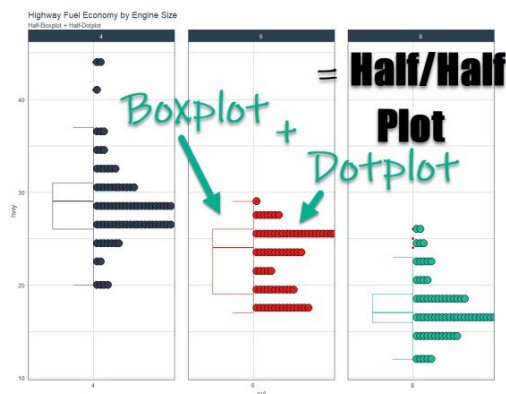


(Click image to play tutorial)

[Watch our full YouTube Tutorial](#)

What are Half Plots?

Combining two plots side-by-side.



Half/Half Plots are a way to showcase two plots side-by-side. Here's a common example:

1. Showing a **Boxplot** to identify outliers and quantiles
2. Showing a **Dotplot** to identify distribution

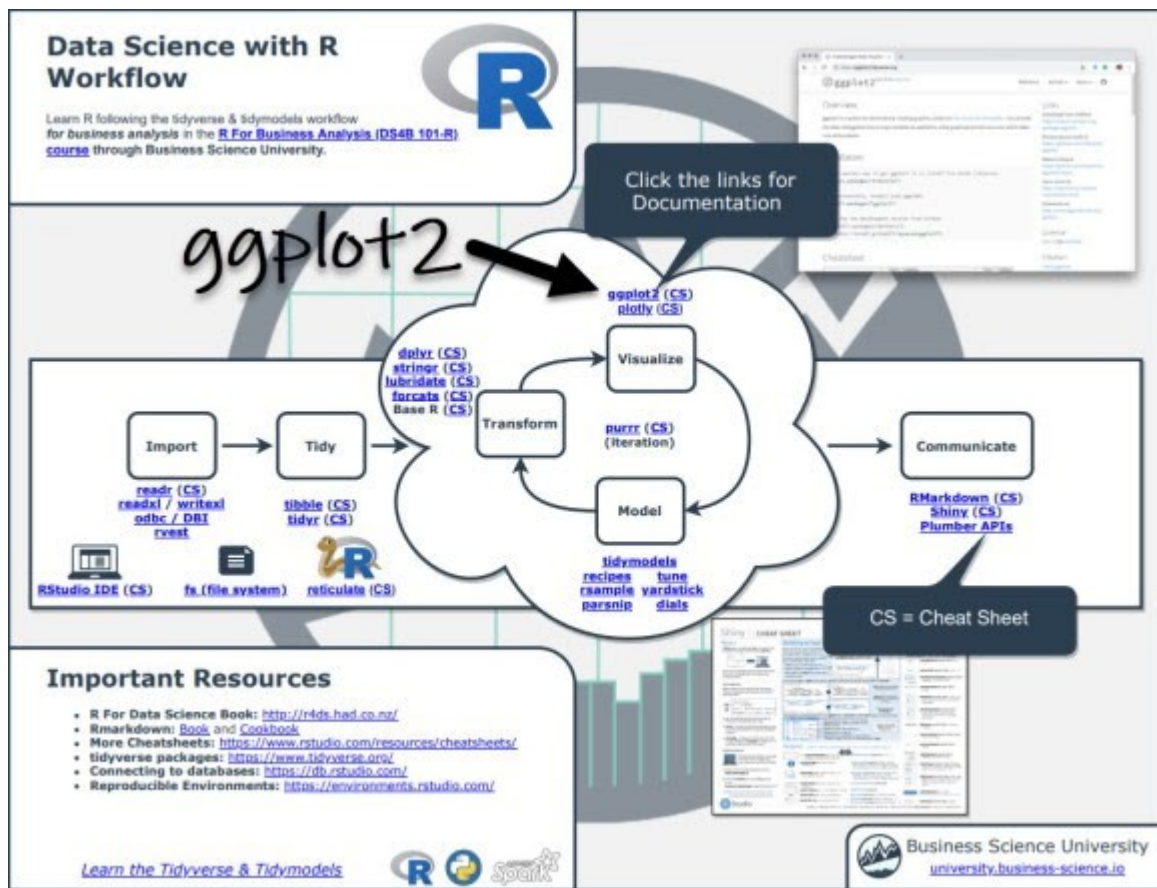
We can easily do this with a half-plot thanks to `gghalves`.

Before we get started, get the R Cheat Sheet

`gghalves` is great for making customized `ggplot2` plots. But, you'll still need to learn how to wrangle data with `dplyr` and visualize data with `ggplot2`. For those topics, I'll use the [Ultimate R Cheat Sheet](#) to refer to `dplyr` and `ggplot2` code in my workflow.

Quick Example:

[Download the Ultimate R Cheat Sheet](#) Then Click the "CS" next to "ggplot2" opens the Data Visualization with GGplot2 Cheat Sheet.



Now you're ready to quickly reference ggplot2 functions.

Onto the tutorial.

Data Visualization with ggplot2 : : CHEAT SHEET

Basics

ggplot2 is based on the **grammar of graphics**, the idea that you can build every graph from the same components: a **data set**, a **coordinate system**, and **geoms**—visual marks that represent data points.

To display values, map variables in the data to visual properties of the geom (aesthetics): line size, color, and x and y locations.

Complete the template below to build a graph.

```
ggplot(data = "data") +
  <GEOM_FUNCTION>(mapping = aes(<MAPPING>))
  stat = "stat", position = "position") +
  <COORDINATE_FUNCTION> +
  <FACET_FUNCTION> +
  <SCALE_FUNCTION> +
  <THEME_FUNCTION>
```

ggplot(data = mpg, aes(x = cty, y = hwy)) Begins a plot that you finish by adding layers to. Add one geom function per layer.

geom_poin(mapping = aes(x = cty, y = hwy)) Creates a complete plot with green data, geom, and mappings. Supplies many useful defaults.

last_plot() Returns the last plot

ggsave("plot.png", width = 5, height = 5) Saves last plot as 5" x 5" file named "plot.png" in working directory. Matches file type to file extension.

Geoms

Use a geom function to represent data points, use the geom's aesthetic properties to represent variables. Each function returns a layer.

GRAPHICAL PRIMITIVES

```
a = ggplot(mapping, aes(x = long, y = lat))
b = geom_blank()
c = geom_line(mapping = aes(x = long, y = lat))
d = geom_point(mapping = aes(x = long, y = lat))
e = geom_path(mapping = aes(x = long, y = lat))
f = geom_polygon(mapping = aes(x = long, y = lat))
g = geom_rect(mapping = aes(x = long, y = lat))
h = geom_ribbon(mapping = aes(x = long, y = lat))
i = geom_area(mapping = aes(x = long, y = lat))
j = geom_bar(mapping = aes(x = long, y = lat))
k = geom_histogram(mapping = aes(x = long, y = lat))
l = geom_density(mapping = aes(x = long, y = lat))
m = geom_violin(mapping = aes(x = long, y = lat))
n = geom_boxplot(mapping = aes(x = long, y = lat))
o = geom_facet(mapping = aes(x = long, y = lat))
p = geom_facet(mapping = aes(x = long, y = lat))
q = geom_facet(mapping = aes(x = long, y = lat))
r = geom_facet(mapping = aes(x = long, y = lat))
s = geom_facet(mapping = aes(x = long, y = lat))
t = geom_facet(mapping = aes(x = long, y = lat))
u = geom_facet(mapping = aes(x = long, y = lat))
v = geom_facet(mapping = aes(x = long, y = lat))
w = geom_facet(mapping = aes(x = long, y = lat))
x = geom_facet(mapping = aes(x = long, y = lat))
y = geom_facet(mapping = aes(x = long, y = lat))
z = geom_facet(mapping = aes(x = long, y = lat))
```

LINE SEGMENTS

```
a = geom_line(mapping = aes(x = long, y = lat))
b = geom_path(mapping = aes(x = long, y = lat))
c = geom_polygon(mapping = aes(x = long, y = lat))
d = geom_rect(mapping = aes(x = long, y = lat))
e = geom_ribbon(mapping = aes(x = long, y = lat))
f = geom_area(mapping = aes(x = long, y = lat))
g = geom_bar(mapping = aes(x = long, y = lat))
h = geom_histogram(mapping = aes(x = long, y = lat))
i = geom_density(mapping = aes(x = long, y = lat))
j = geom_violin(mapping = aes(x = long, y = lat))
k = geom_boxplot(mapping = aes(x = long, y = lat))
l = geom_facet(mapping = aes(x = long, y = lat))
m = geom_facet(mapping = aes(x = long, y = lat))
n = geom_facet(mapping = aes(x = long, y = lat))
o = geom_facet(mapping = aes(x = long, y = lat))
p = geom_facet(mapping = aes(x = long, y = lat))
q = geom_facet(mapping = aes(x = long, y = lat))
r = geom_facet(mapping = aes(x = long, y = lat))
s = geom_facet(mapping = aes(x = long, y = lat))
t = geom_facet(mapping = aes(x = long, y = lat))
u = geom_facet(mapping = aes(x = long, y = lat))
v = geom_facet(mapping = aes(x = long, y = lat))
w = geom_facet(mapping = aes(x = long, y = lat))
x = geom_facet(mapping = aes(x = long, y = lat))
y = geom_facet(mapping = aes(x = long, y = lat))
z = geom_facet(mapping = aes(x = long, y = lat))
```

ONE VARIABLE - continuous

```
a = ggplot(mapping, aes(x = long))
b = geom_area(mapping = aes(x = long))
c = geom_bar(mapping = aes(x = long))
d = geom_histogram(mapping = aes(x = long))
e = geom_density(mapping = aes(x = long))
f = geom_violin(mapping = aes(x = long))
g = geom_boxplot(mapping = aes(x = long))
h = geom_facet(mapping = aes(x = long))
i = geom_facet(mapping = aes(x = long))
j = geom_facet(mapping = aes(x = long))
k = geom_facet(mapping = aes(x = long))
l = geom_facet(mapping = aes(x = long))
m = geom_facet(mapping = aes(x = long))
n = geom_facet(mapping = aes(x = long))
o = geom_facet(mapping = aes(x = long))
p = geom_facet(mapping = aes(x = long))
q = geom_facet(mapping = aes(x = long))
r = geom_facet(mapping = aes(x = long))
s = geom_facet(mapping = aes(x = long))
t = geom_facet(mapping = aes(x = long))
u = geom_facet(mapping = aes(x = long))
v = geom_facet(mapping = aes(x = long))
w = geom_facet(mapping = aes(x = long))
x = geom_facet(mapping = aes(x = long))
y = geom_facet(mapping = aes(x = long))
z = geom_facet(mapping = aes(x = long))
```

TWO VARIABLES

continuous x, continuous y

```
a = ggplot(mapping, aes(x = long, y = lat))
b = geom_area(mapping = aes(x = long, y = lat))
c = geom_bar(mapping = aes(x = long, y = lat))
d = geom_histogram(mapping = aes(x = long, y = lat))
e = geom_density(mapping = aes(x = long, y = lat))
f = geom_violin(mapping = aes(x = long, y = lat))
g = geom_boxplot(mapping = aes(x = long, y = lat))
h = geom_facet(mapping = aes(x = long, y = lat))
i = geom_facet(mapping = aes(x = long, y = lat))
j = geom_facet(mapping = aes(x = long, y = lat))
k = geom_facet(mapping = aes(x = long, y = lat))
l = geom_facet(mapping = aes(x = long, y = lat))
m = geom_facet(mapping = aes(x = long, y = lat))
n = geom_facet(mapping = aes(x = long, y = lat))
o = geom_facet(mapping = aes(x = long, y = lat))
p = geom_facet(mapping = aes(x = long, y = lat))
q = geom_facet(mapping = aes(x = long, y = lat))
r = geom_facet(mapping = aes(x = long, y = lat))
s = geom_facet(mapping = aes(x = long, y = lat))
t = geom_facet(mapping = aes(x = long, y = lat))
u = geom_facet(mapping = aes(x = long, y = lat))
v = geom_facet(mapping = aes(x = long, y = lat))
w = geom_facet(mapping = aes(x = long, y = lat))
x = geom_facet(mapping = aes(x = long, y = lat))
y = geom_facet(mapping = aes(x = long, y = lat))
z = geom_facet(mapping = aes(x = long, y = lat))
```

continuous bivariate distribution

```
a = ggplot(mapping, aes(x = long, y = lat))
b = geom_area(mapping = aes(x = long, y = lat))
c = geom_bar(mapping = aes(x = long, y = lat))
d = geom_histogram(mapping = aes(x = long, y = lat))
e = geom_density(mapping = aes(x = long, y = lat))
f = geom_violin(mapping = aes(x = long, y = lat))
g = geom_boxplot(mapping = aes(x = long, y = lat))
h = geom_facet(mapping = aes(x = long, y = lat))
i = geom_facet(mapping = aes(x = long, y = lat))
j = geom_facet(mapping = aes(x = long, y = lat))
k = geom_facet(mapping = aes(x = long, y = lat))
l = geom_facet(mapping = aes(x = long, y = lat))
m = geom_facet(mapping = aes(x = long, y = lat))
n = geom_facet(mapping = aes(x = long, y = lat))
o = geom_facet(mapping = aes(x = long, y = lat))
p = geom_facet(mapping = aes(x = long, y = lat))
q = geom_facet(mapping = aes(x = long, y = lat))
r = geom_facet(mapping = aes(x = long, y = lat))
s = geom_facet(mapping = aes(x = long, y = lat))
t = geom_facet(mapping = aes(x = long, y = lat))
u = geom_facet(mapping = aes(x = long, y = lat))
v = geom_facet(mapping = aes(x = long, y = lat))
w = geom_facet(mapping = aes(x = long, y = lat))
x = geom_facet(mapping = aes(x = long, y = lat))
y = geom_facet(mapping = aes(x = long, y = lat))
z = geom_facet(mapping = aes(x = long, y = lat))
```

discrete x, continuous y

```
a = ggplot(mapping, aes(x = long, y = lat))
b = geom_area(mapping = aes(x = long, y = lat))
c = geom_bar(mapping = aes(x = long, y = lat))
d = geom_histogram(mapping = aes(x = long, y = lat))
e = geom_density(mapping = aes(x = long, y = lat))
f = geom_violin(mapping = aes(x = long, y = lat))
g = geom_boxplot(mapping = aes(x = long, y = lat))
h = geom_facet(mapping = aes(x = long, y = lat))
i = geom_facet(mapping = aes(x = long, y = lat))
j = geom_facet(mapping = aes(x = long, y = lat))
k = geom_facet(mapping = aes(x = long, y = lat))
l = geom_facet(mapping = aes(x = long, y = lat))
m = geom_facet(mapping = aes(x = long, y = lat))
n = geom_facet(mapping = aes(x = long, y = lat))
o = geom_facet(mapping = aes(x = long, y = lat))
p = geom_facet(mapping = aes(x = long, y = lat))
q = geom_facet(mapping = aes(x = long, y = lat))
r = geom_facet(mapping = aes(x = long, y = lat))
s = geom_facet(mapping = aes(x = long, y = lat))
t = geom_facet(mapping = aes(x = long, y = lat))
u = geom_facet(mapping = aes(x = long, y = lat))
v = geom_facet(mapping = aes(x = long, y = lat))
w = geom_facet(mapping = aes(x = long, y = lat))
x = geom_facet(mapping = aes(x = long, y = lat))
y = geom_facet(mapping = aes(x = long, y = lat))
z = geom_facet(mapping = aes(x = long, y = lat))
```

discrete x, discrete y

```
a = ggplot(mapping, aes(x = long, y = lat))
b = geom_area(mapping = aes(x = long, y = lat))
c = geom_bar(mapping = aes(x = long, y = lat))
d = geom_histogram(mapping = aes(x = long, y = lat))
e = geom_density(mapping = aes(x = long, y = lat))
f = geom_violin(mapping = aes(x = long, y = lat))
g = geom_boxplot(mapping = aes(x = long, y = lat))
h = geom_facet(mapping = aes(x = long, y = lat))
i = geom_facet(mapping = aes(x = long, y = lat))
j = geom_facet(mapping = aes(x = long, y = lat))
k = geom_facet(mapping = aes(x = long, y = lat))
l = geom_facet(mapping = aes(x = long, y = lat))
m = geom_facet(mapping = aes(x = long, y = lat))
n = geom_facet(mapping = aes(x = long, y = lat))
o = geom_facet(mapping = aes(x = long, y = lat))
p = geom_facet(mapping = aes(x = long, y = lat))
q = geom_facet(mapping = aes(x = long, y = lat))
r = geom_facet(mapping = aes(x = long, y = lat))
s = geom_facet(mapping = aes(x = long, y = lat))
t = geom_facet(mapping = aes(x = long, y = lat))
u = geom_facet(mapping = aes(x = long, y = lat))
v = geom_facet(mapping = aes(x = long, y = lat))
w = geom_facet(mapping = aes(x = long, y = lat))
x = geom_facet(mapping = aes(x = long, y = lat))
y = geom_facet(mapping = aes(x = long, y = lat))
z = geom_facet(mapping = aes(x = long, y = lat))
```

THREE VARIABLES

```
a = ggplot(mapping, aes(x = long, y = lat, color = "color"))
b = geom_area(mapping = aes(x = long, y = lat, color = "color"))
c = geom_bar(mapping = aes(x = long, y = lat, color = "color"))
d = geom_histogram(mapping = aes(x = long, y = lat, color = "color"))
e = geom_density(mapping = aes(x = long, y = lat, color = "color"))
f = geom_violin(mapping = aes(x = long, y = lat, color = "color"))
g = geom_boxplot(mapping = aes(x = long, y = lat, color = "color"))
h = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
i = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
j = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
k = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
l = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
m = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
n = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
o = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
p = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
q = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
r = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
s = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
t = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
u = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
v = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
w = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
x = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
y = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
z = geom_facet(mapping = aes(x = long, y = lat, color = "color"))
```

R Studio

RStudio is a trademark of RStudio, Inc. • CC BY-SA RStudio • info@rstudio.com • 844-484-1212 • rstudio.com • Learn more at <http://ggplot2.tidyverse.org> • ggplot2 3.1.0 • Updated 2018-12

How ggplot2 works

The `gghalves` package extends `ggplot2` by adding several new “geoms” (ggplot geometries) that allow us to add half plots. In this tutorial, we’ll cover:

- `geom_half_boxplot()`: For creating half-boxplots
- `geom_half_dotplot()`: For creating half-dotplots

Pro Tip:

Simply type “geom_half” in your R console and hit Tab to show all of the half plotting geoms available.

Load the Libraries and Data

First, run this code to:

1. **Load Libraries:** Load `gghalves`, `tidyverse` and `tidyquant`.
2. **Import Data:** We’re using the `mpg` dataset that comes with `ggplot2`.

```
0
1
2
3
4
5
6
7 # LIBRARIES ----
8
9 library(tidyverse)
10 library(tidyquant)
11 library(gghalves)
12
13 # DATA ----
14 mpg
15
```

[Get the code.](#)

Make the Half-Boxplot / Half-Dotplot

Next, we can combine a half-boxplot and half-dotplot. This has the advantage of showing:

- Quantiles and Outliers (Boxplot)
- Distribution (Dotplot)

Business Goal

Suppose we have a question:

What effect does **Engine Size (number of Cylinders)** have on **Vehicle Highway Fuel Economy (Highway MPG)**?

We can visualize this with `gghalves` by making half-plots of Cylinder vs Highway.

Half-Plot Visualization Code

Using the [Ultimate R Cheat Sheet](#), we can make a `ggplot` from the `ggplot2` data visualization cheat sheet. We’ll add `geom_half_boxplot()` and `geom_half_dotplot()` to make the half-plots of Cylinder vs Highway.


```

33 # 3.0 HALF-BOXPLOT / HALF-DOTPLOT ----
34 mpg %>%
35   filter(cyl != 5) %>%
36   mutate(cyl = factor(cyl)) %>%
37   ggplot(aes(cyl, hwy, color = cyl)) +
38
39   geom_half_boxplot(outlier.color = "red") +
40   geom_half_dotplot(
41     aes(fill = cyl),
42     dotsize = 0.75,
43     stackratio = 0.5,
44     color = "black"
45   ) +
46
47   facet_grid(cols = vars(cyl), scales = "free_x") +
48   scale_color_tq() +
49   scale_fill_tq() +
50   theme_tq() +
51   labs(
52     title = "Highway Fuel Economy by Engine Size",
53     subtitle = "Half-Boxplot + Half-Dotplot"
54   )
55

```

[Get the code.](#)

Half-Plot Visualization

Here is the visualization. We can explore to find an interesting relationship between Engine Size and Fuel Economy.



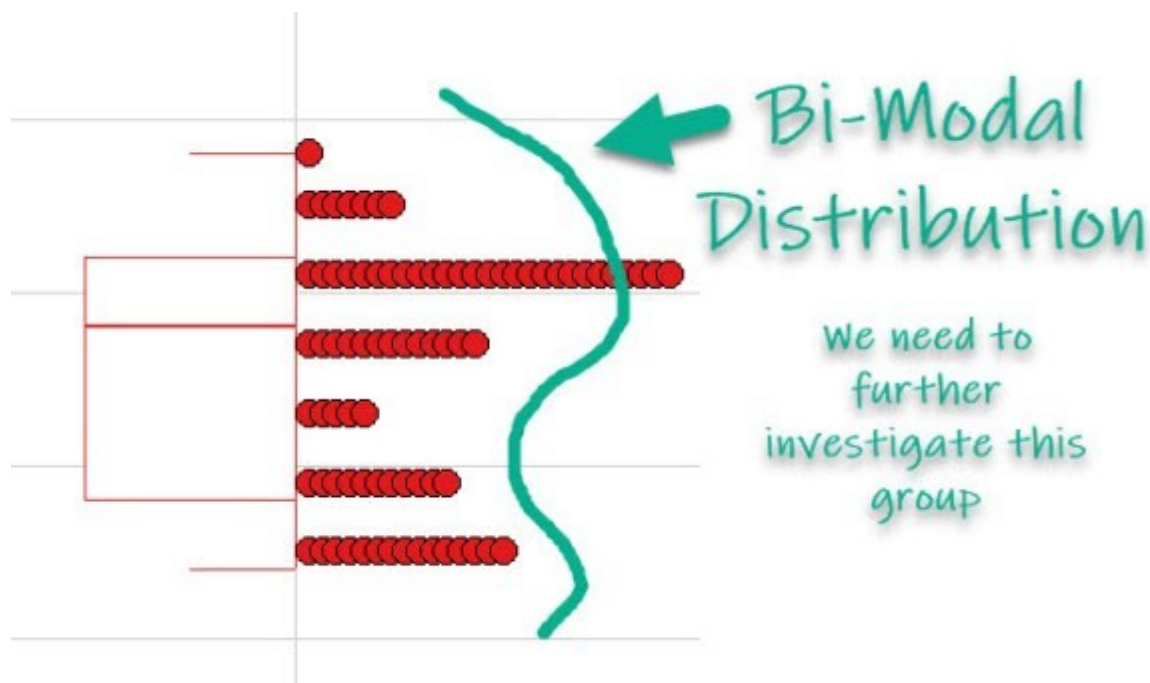
[Get the code.](#)

Insights: **Bimodal Distribution** of 6-Cylinder Engine Class

Generally speaking, fuel economy goes down as engine size increases. But, the 6-Cylinder engine has something unique going on that has been uncovered by the

```
gghalves::geom_half_dotplot().
```

The 6-Cylinder Engine class of car has a **bimodal distribution**, which is when there are two peaks. This generally indicates that there are two different populations within the group. We need to investigate with `ggplot2`.



[Get the code.](#)

Exploring the Bimodal Relationship

We can explore the 6 Cylinder Vehicle Class a bit further to identify the cause of the Bimodal Distribution. It looks like:

- **SUV and Pickup classes** have much lower fuel economy
- **Compact, Midsize, Minivan, and Subcompact** have much higher fuel economy

