CHRISTOPHER PRENER, PHD
FALL, 2017

WEEK 02
LECTURE 02B

QUANTITATIVE ANALYSIS

THE GRAMMAR OF GRAPHICS

AGENDA

1. Statistical Transformations

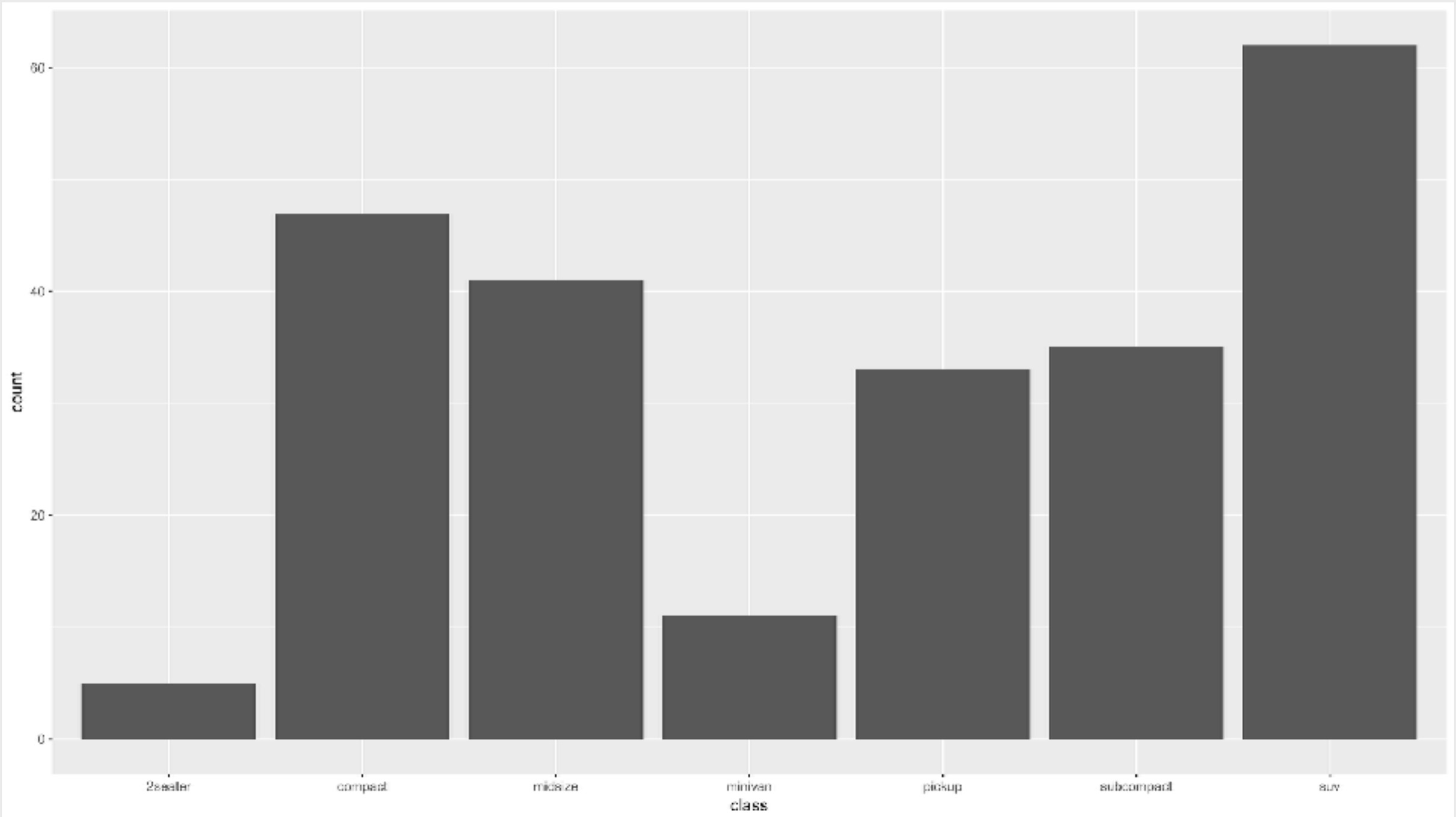
2. Aesthetic Adjustments

3. Coordinate Systems

4. The Grammar of Graphics

1 STATISTICAL TRANSFORMATIONS

BASIC BAR PLOT



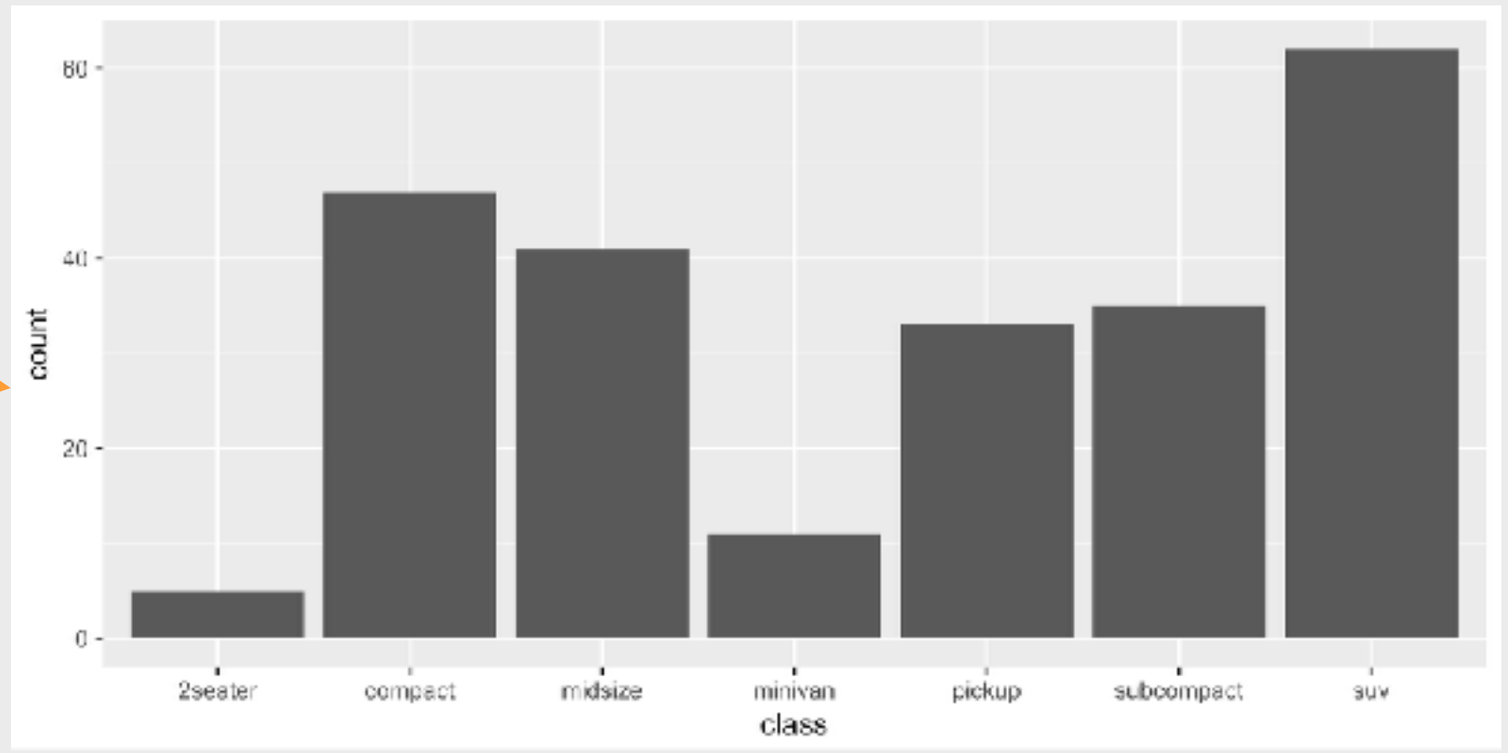
UNDER THE HOOD

	manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
1	audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
2	audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
3	audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	compact
4	audi	a4	2.0	2008	4	auto(av)	f	21	30	p	compact
5	audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
6	audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact
7	audi	a4	3.1	2008	6	auto(av)	f	18	27	p	compact

stat_count()

	class	count
1	2seater	5
2	compact	47
3	midsize	41
4	minivan	11
5	pickup	33
6	subcompact	35
7	suv	62

geom_bar()



BAR PLOTS

```
ggplot2::ggplot(data = dataFrame) +
   geom_bar(mapping = aes(aesthetic))
```

Example – the mpg data from ggplot2:

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

BAR PLOTS

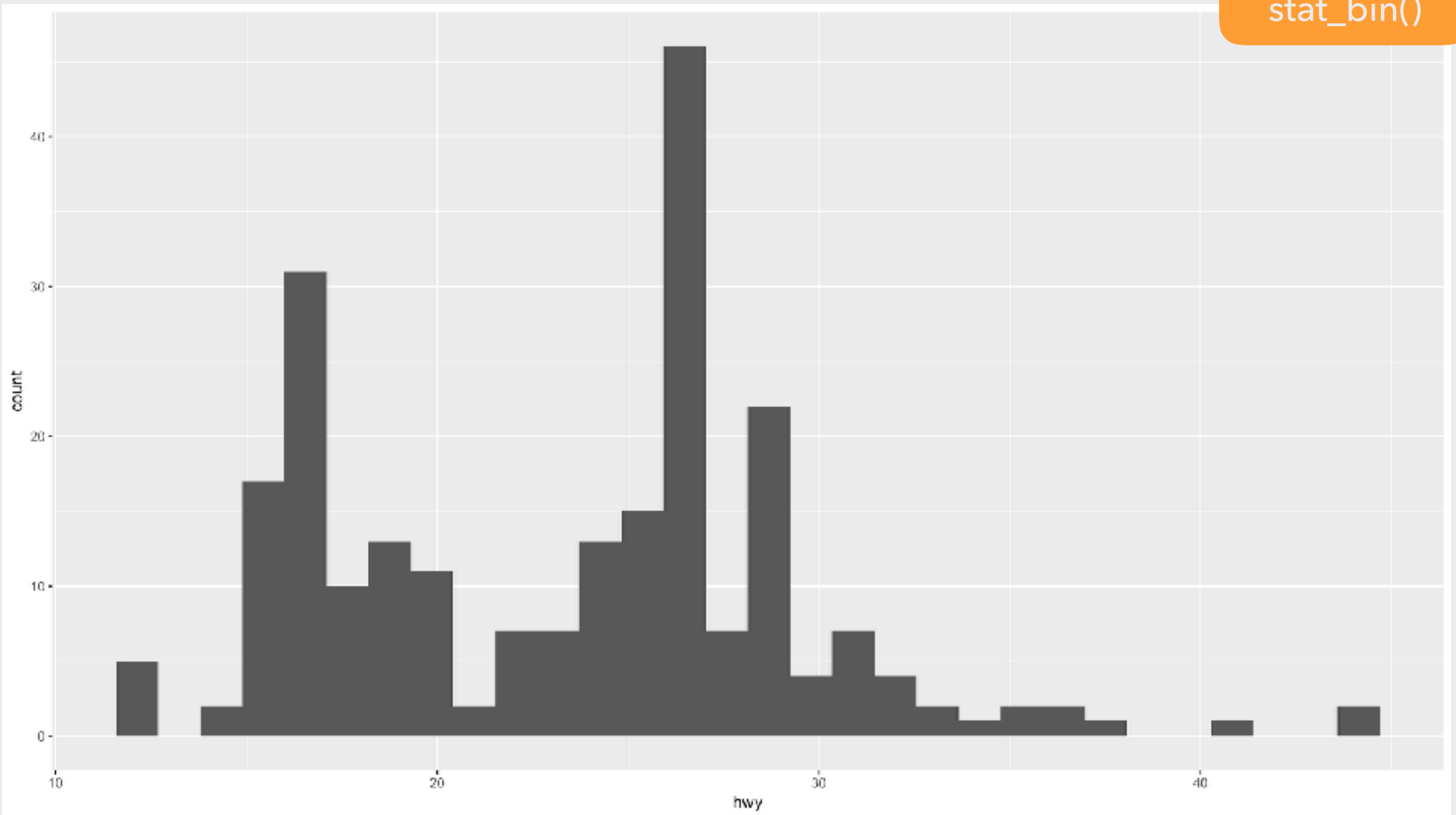
```
ggplot2::ggplot(data = dataFrame) +
   geom_bar(mapping = aes(aesthetic), stat = “identity”)
```

```
Example – the mpg data from ggplot2:
   ggplot(data = mpg) +
      geom_bar(mapping = aes(class), stat = “identity”)
```

The stat “identity” is implied by default when you use the bar geom

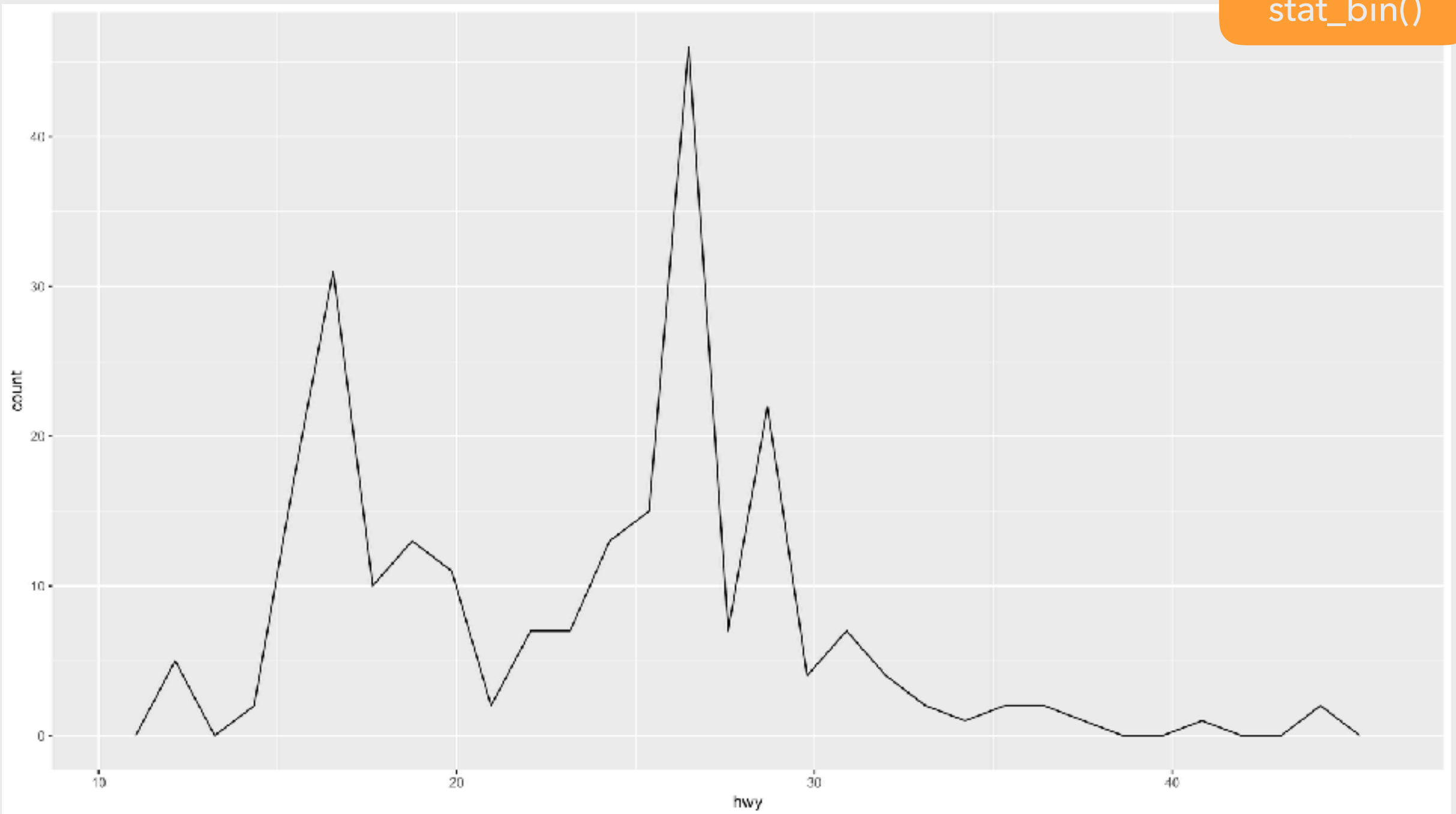
CONTINUOUS VARIABLES

stat_bin()



LINE PLOTS

stat_bin()



AREA PLOTS

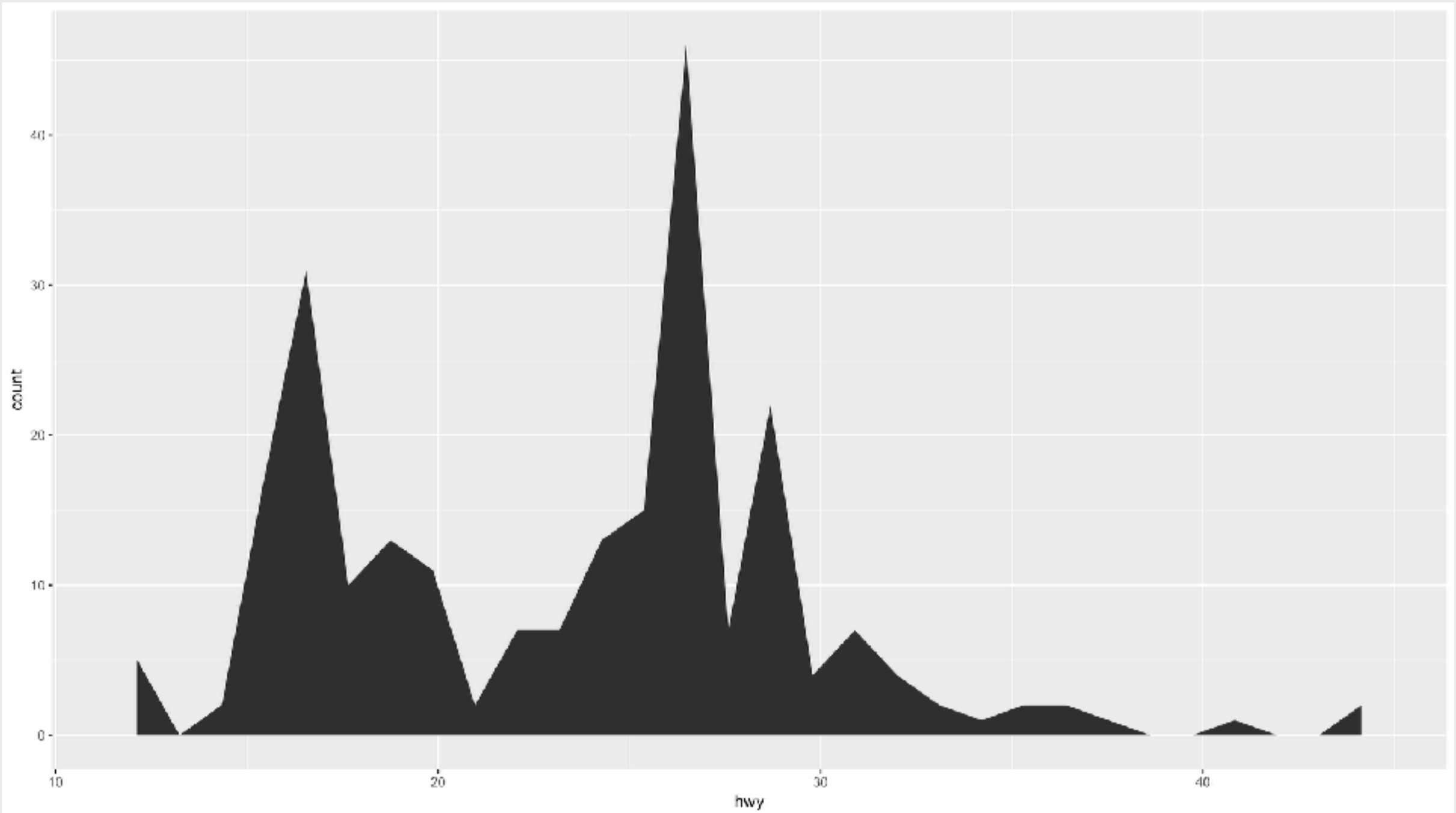
```
ggplot2::ggplot(data = dataFrame) +

   geom_area(mapping = aes(aesthetic), stat = “bin”)
```

Example - the `mpg` data from `ggplot2`:

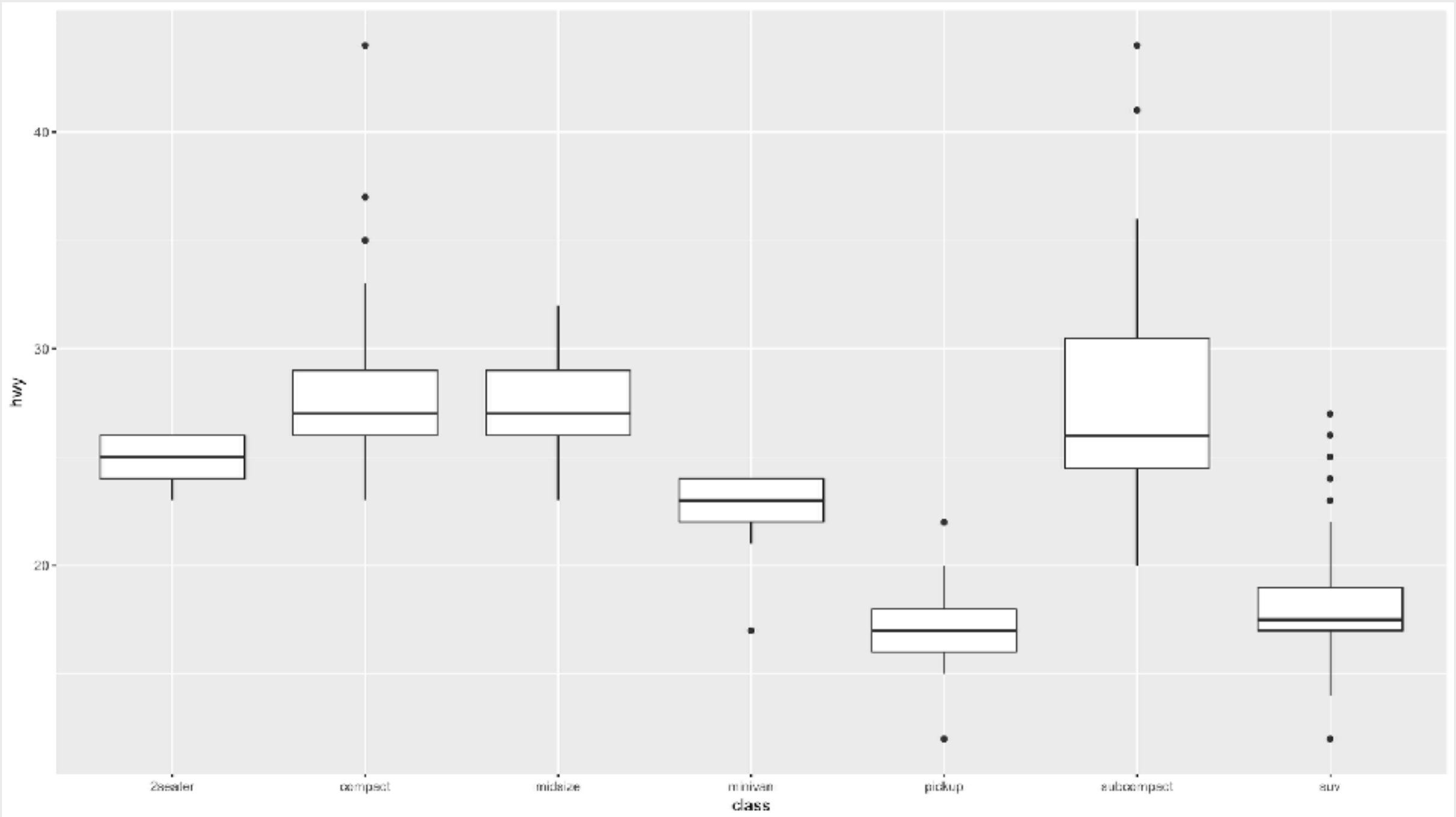
```
   ggplot(data = mpg) +

      geom_area(mapping = aes(hwy), stat = “bin”)
```

This geom is for use with one continuous variable, stat *must* be included

AREA PLOTS



BOX PLOTS

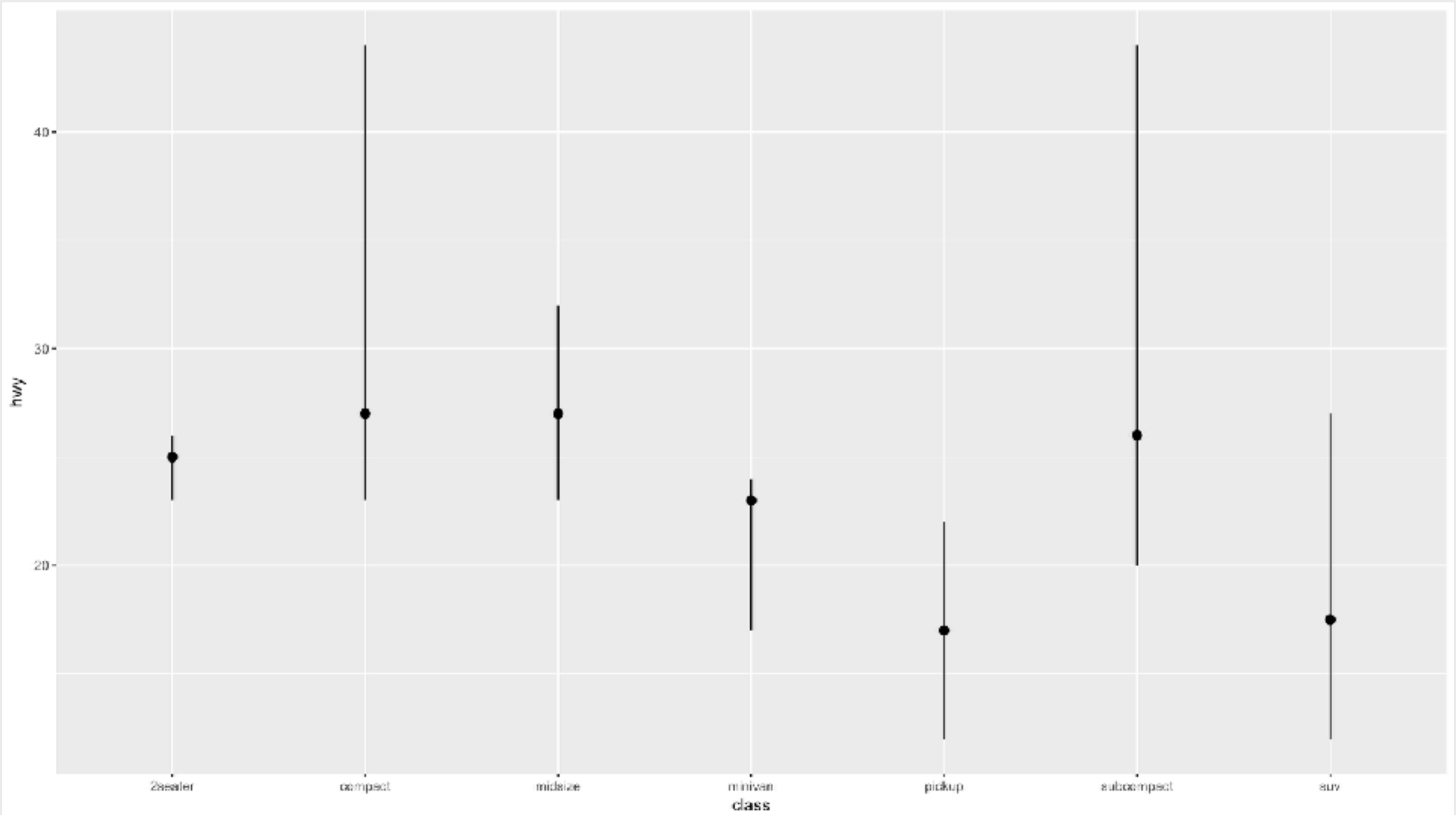

40
30
20
hwy
2seater
compact
midsize
minivan
pickup
subcompact
suv
class

SUMMARIZING VALUES

Example – the mpg data from ggplot2:

```
  ggplot(data = mpg) +
    stat_summary(
      mapping = aes(x = class, y = hwy),
      fun.ymin = min,
      fun.ymax = max,
      fun.y = median
    )
```

SUMMARIZING VALUES



2 AESTHETIC ADJUSTMENTS

ADDING COLOR ARBITRARILY

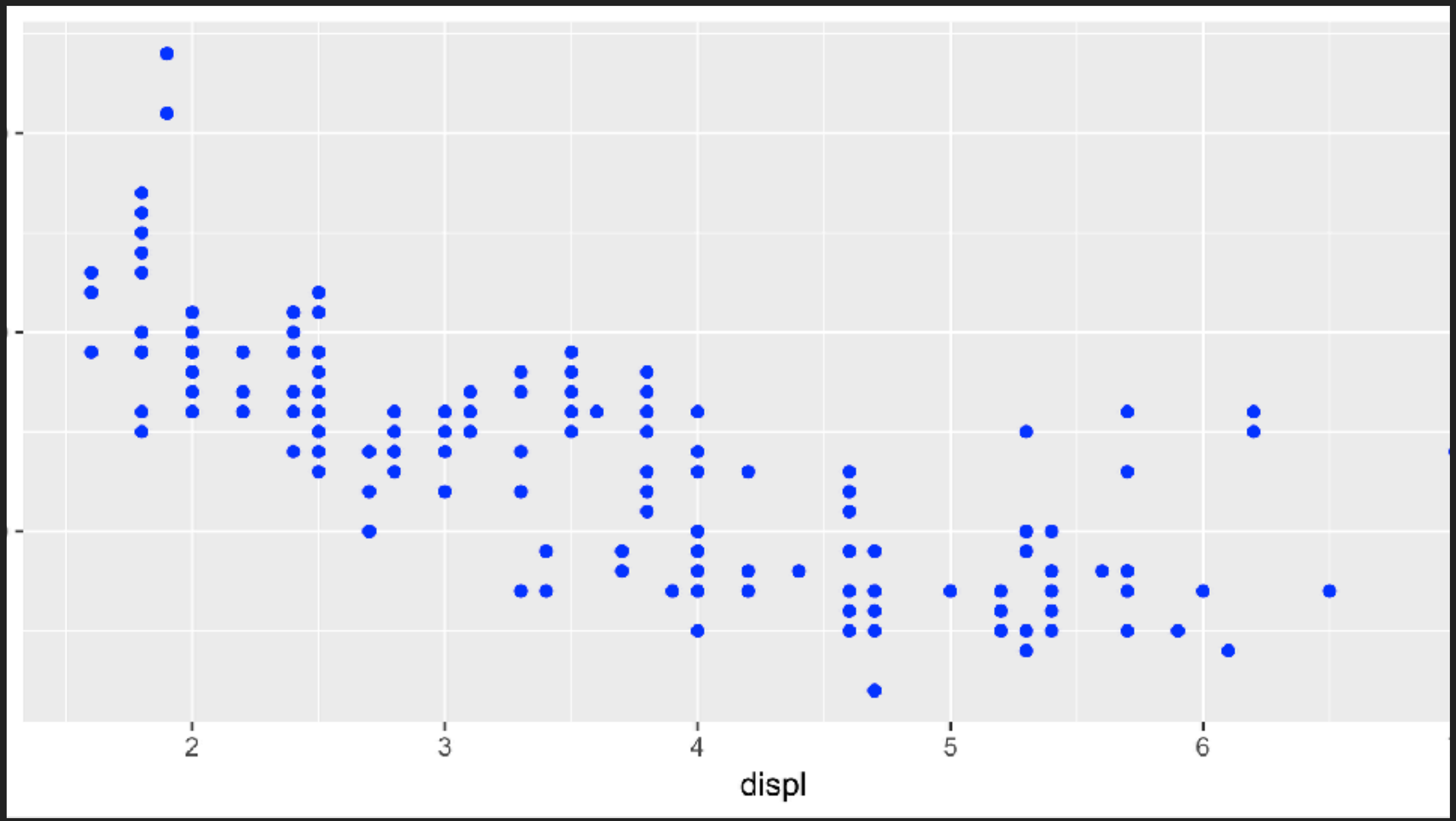
```
ggplot2::ggplot(data = dataFrame) +

   geom_point(mapping = aes(x = var1, y = var2), color = “color”)
```

Example – the mpg data from ggplot2:

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

You cannot use the fill option instead of color with this geom

ADDING COLOR ARBITRARILY



ADDING COLOR ARBITRARILY

```
ggplot2::ggplot(data = dataFrame) +
   geom_point(mapping = aes(x = var1, y = var2), color = “color”)
```

Example – the mpg data from ggplot2:

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

You cannot use the fill option instead of color with this geom

For now, avoid setting colors arbitrarily and let `ggplot2` do the work instead!

ADDING COLOR TO THE AESTHETIC

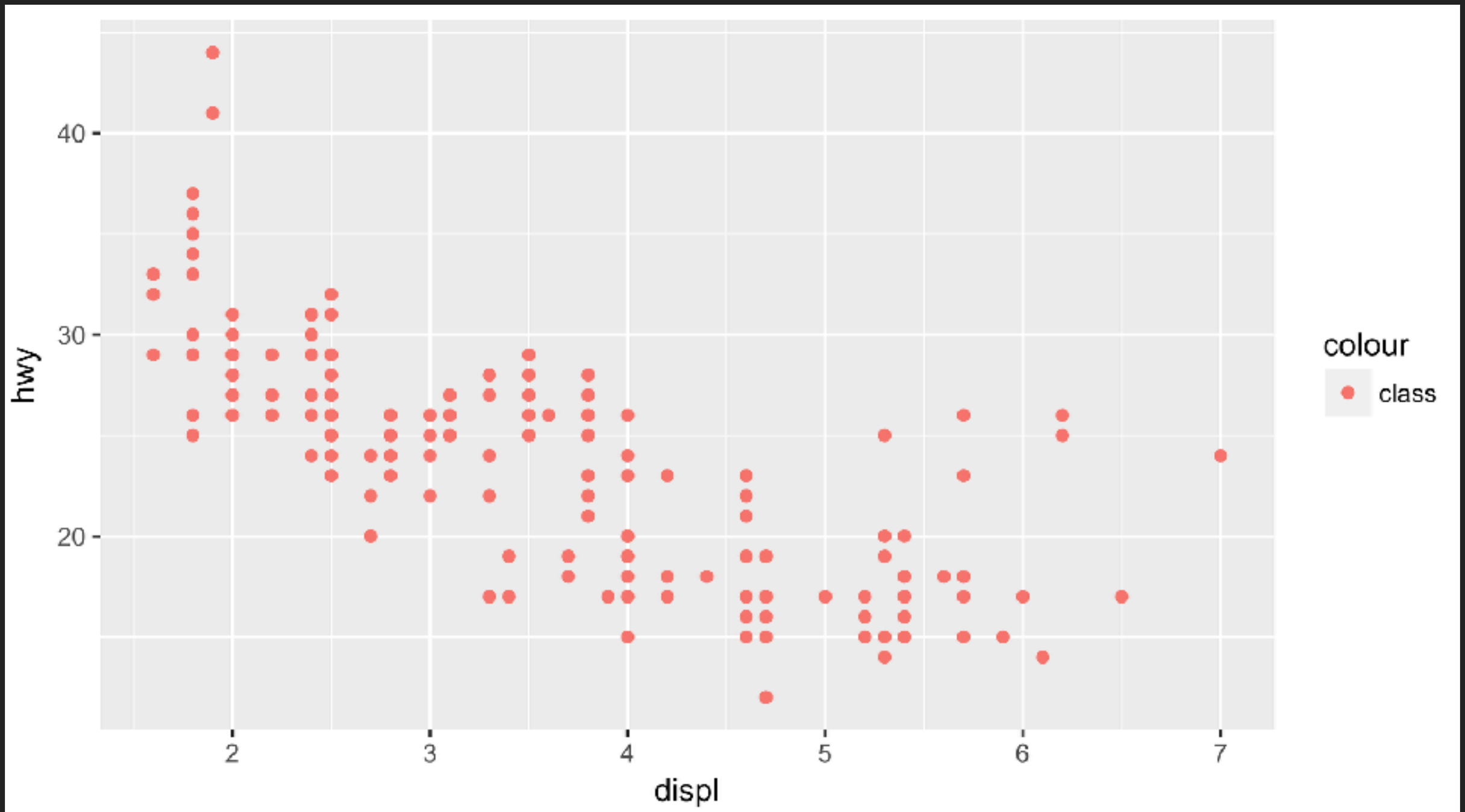
```
ggplot2::ggplot(data = dataFrame) +
   geom_point(mapping = aes(x = var1, y = var2, color = “color”))
```

Example – the mpg data from ggplot2:

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

You cannot use the fill option instead of color with this geom

ADDING COLOR TO THE AESTHETIC



ADDING COLOR BASED ON OTHER VALUES

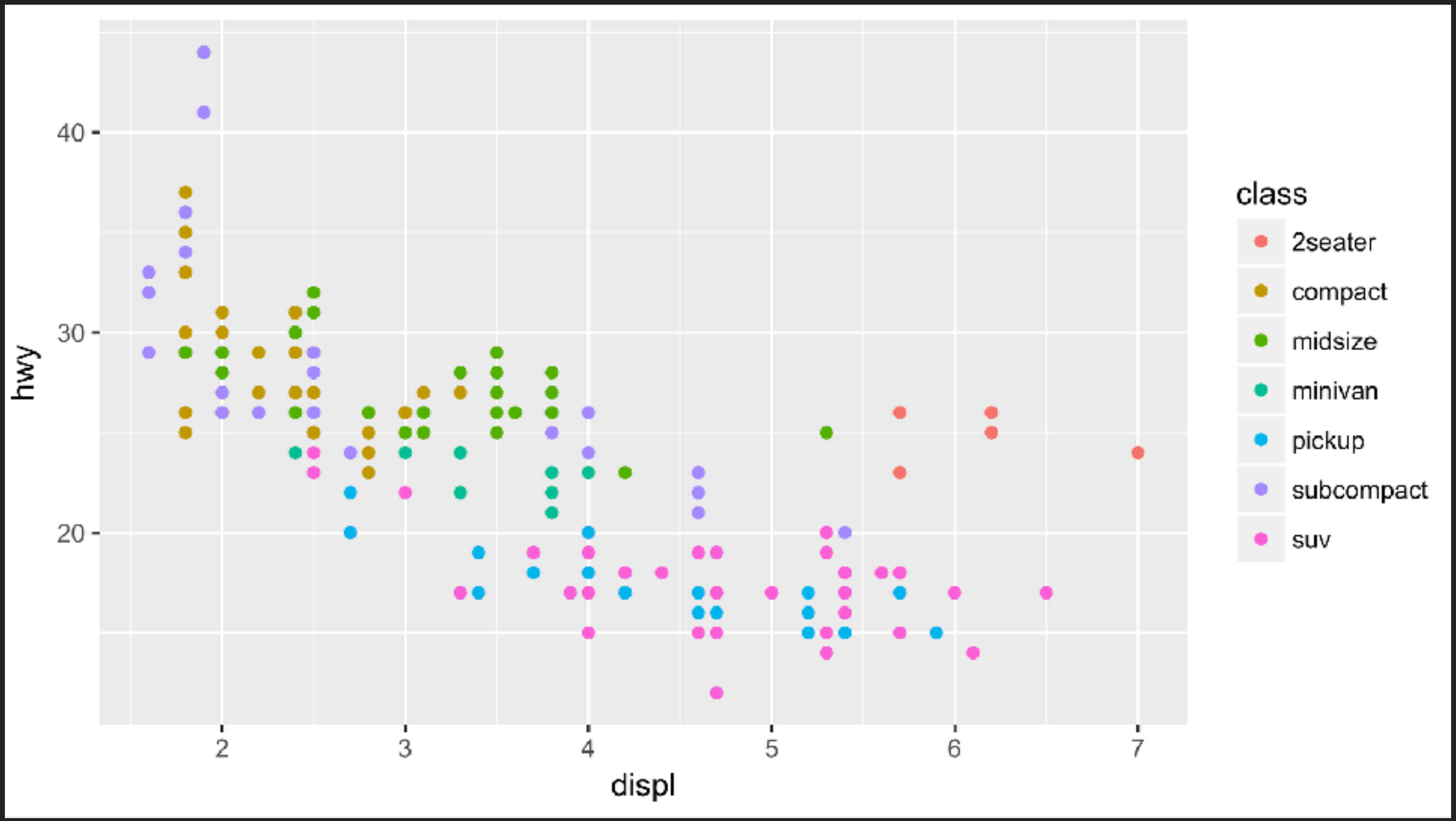
```
ggplot2::ggplot(data = dataFrame) +
   geom_point(mapping = aes(x = var1, y = var2, color = var3))
```

Example – the mpg data from ggplot2:

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

You cannot use the fill option instead of color here

ADDING COLOR BASED ON OTHER VALUES



ADDING COLOR BASED ON OTHER VALUES

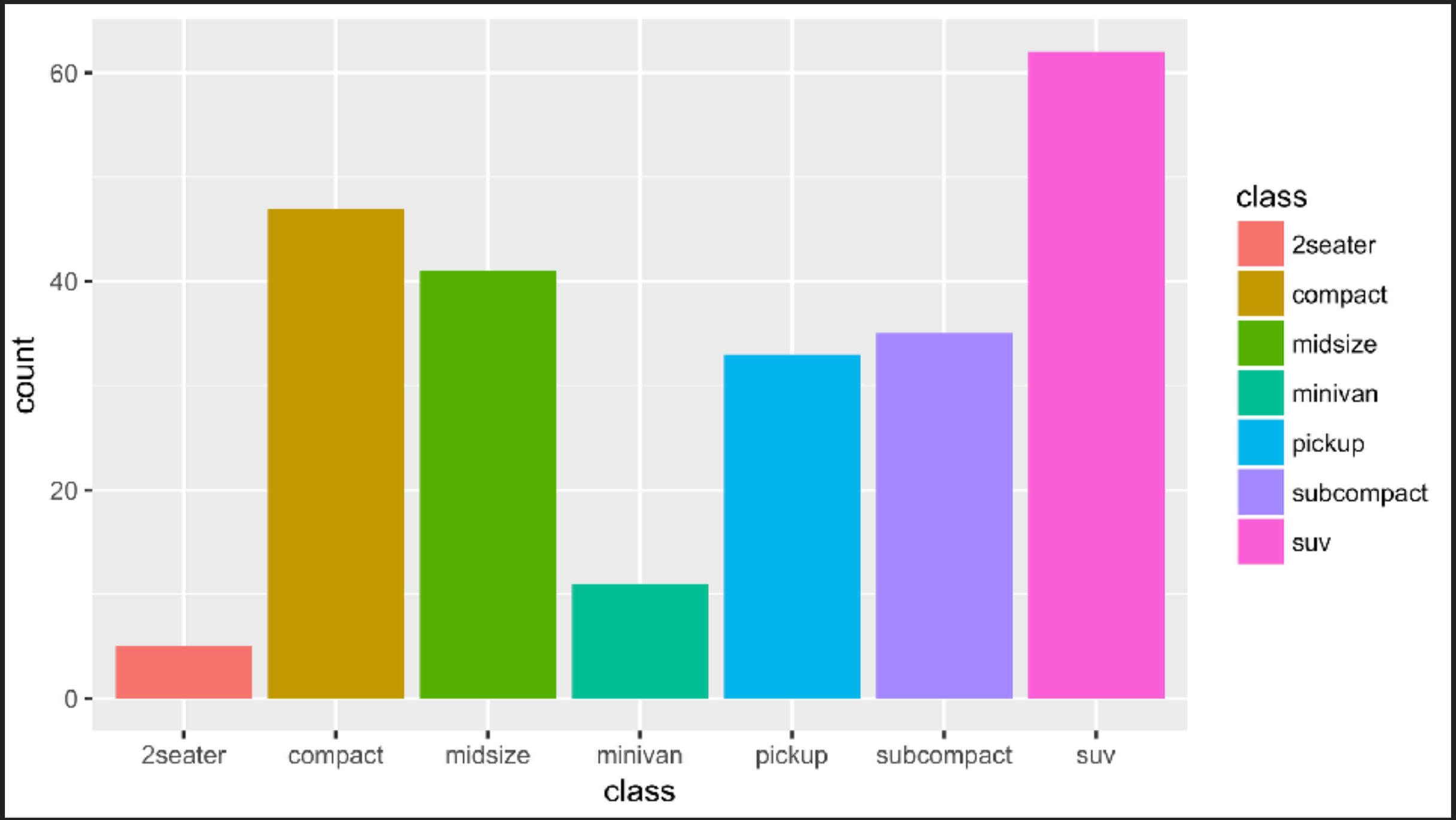
```
ggplot2::ggplot(data = dataFrame) +
   geom_bar(mapping = aes(x = var, fill = var))
```

Example – the mpg data from ggplot2:

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

This can also be done with the color option, but fill is preferable

ADDING COLOR BASED ON OTHER VALUES



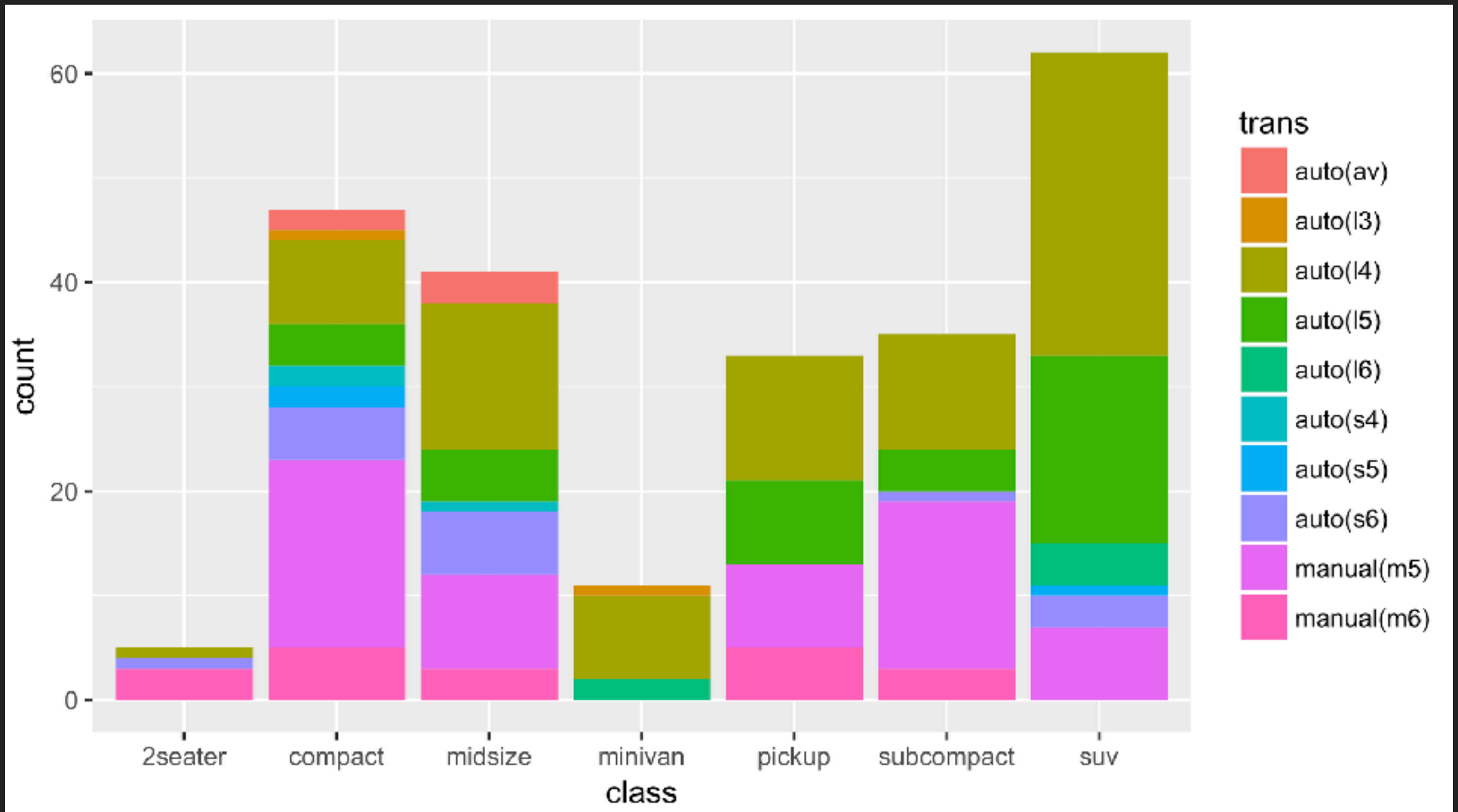
ADDING COLOR BASED ON OTHER VALUES

```
ggplot2::ggplot(data = dataFrame) +
   geom_bar(mapping = aes(x = var, fill = var))
```

Example – the mpg data from ggplot2:

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

ADDING COLOR BASED ON OTHER VALUES



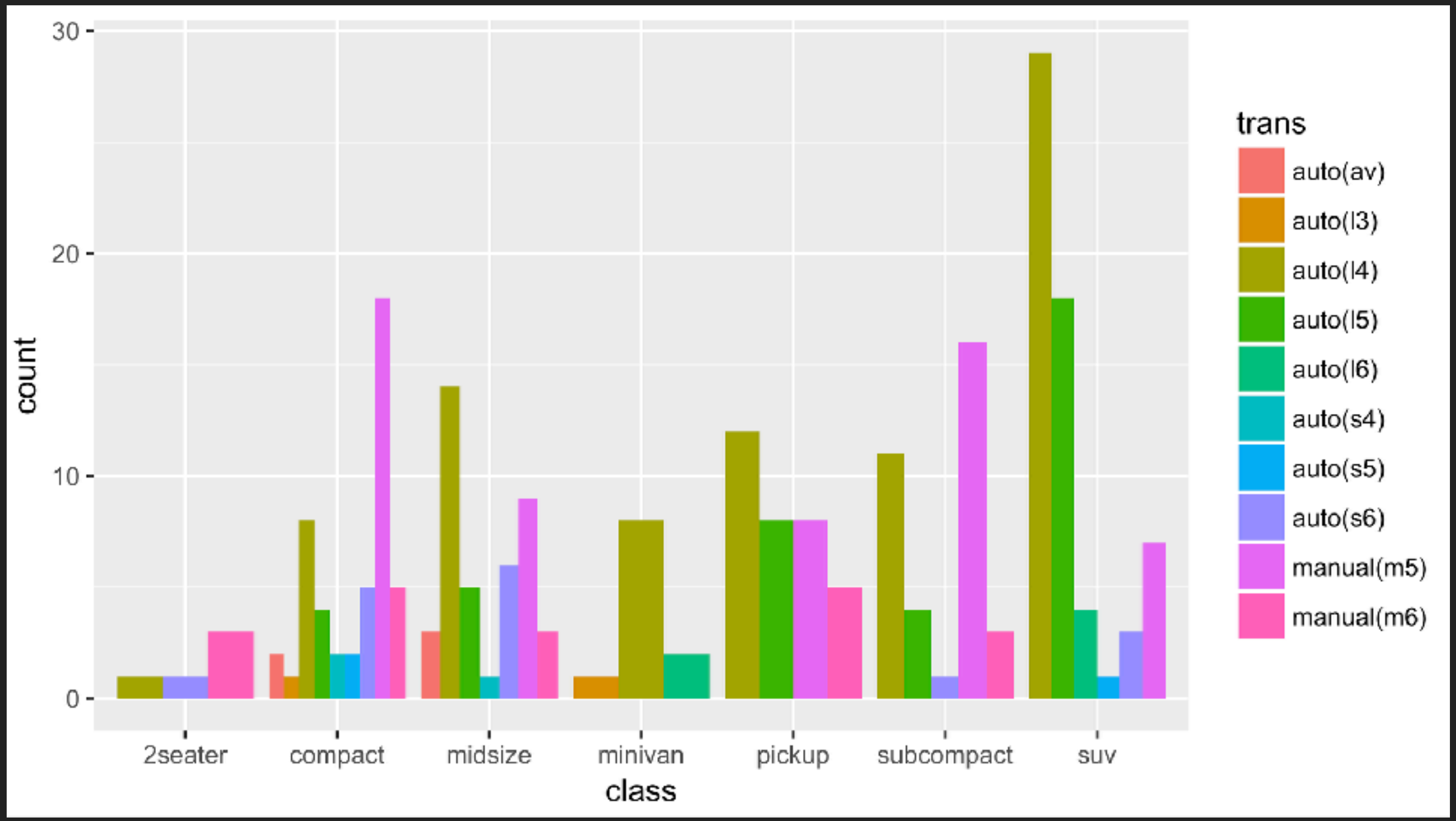
POSITION ADJUSTMENTS – DODGE

```
ggplot2::ggplot(data = dataFrame) +
   geom_bar(mapping = aes(x = var, fill = var),
   position = “dodge”)
```

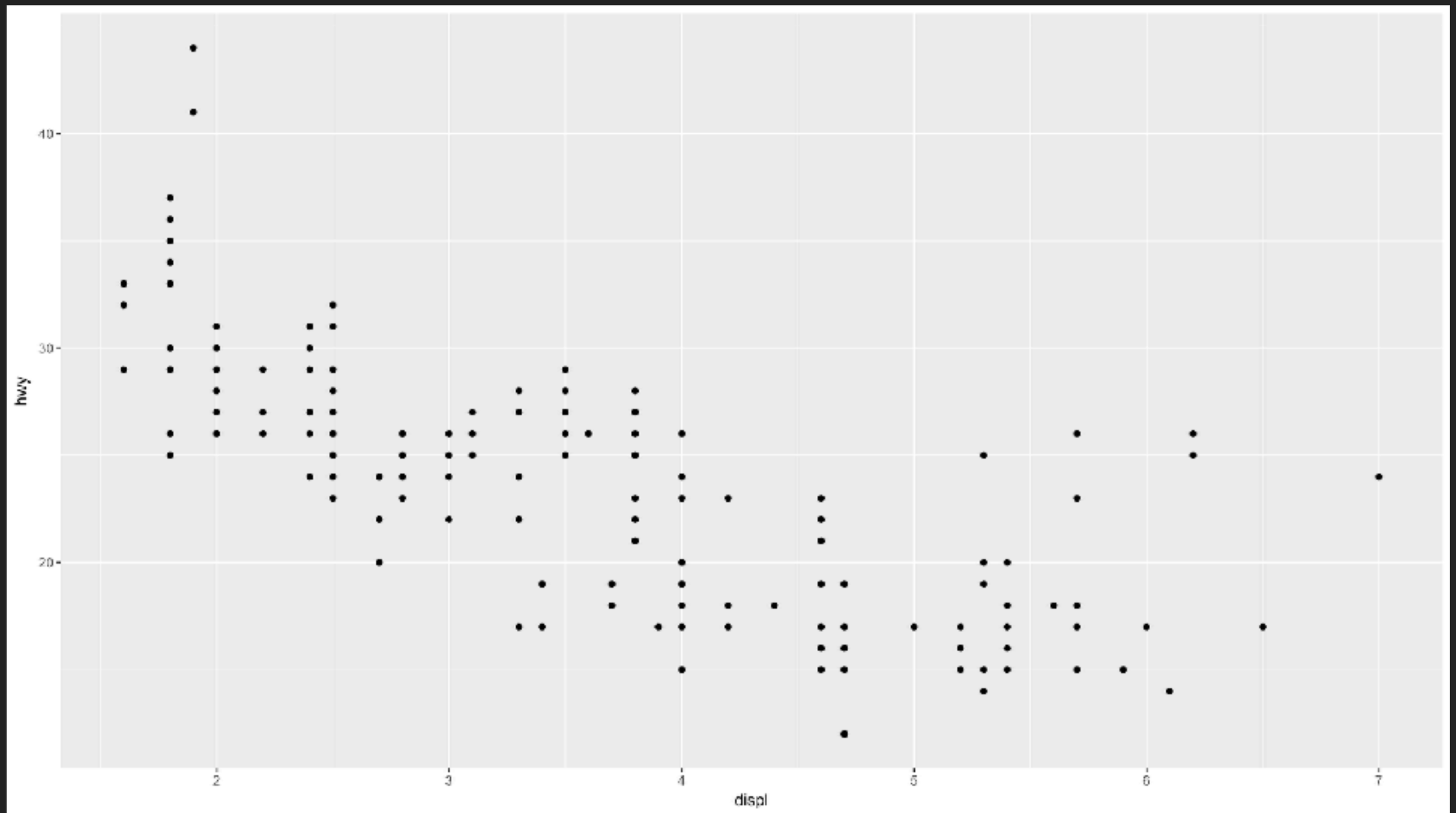
Example – the mpg data from ggplot2:

```
   ggplot(data = mpg) +
      geom_bar(mapping = aes(x = class, fill = trans),
      position = “dodge”)
```

POSITION ADJUSTMENTS – DODGE



"MISSING" POINTS



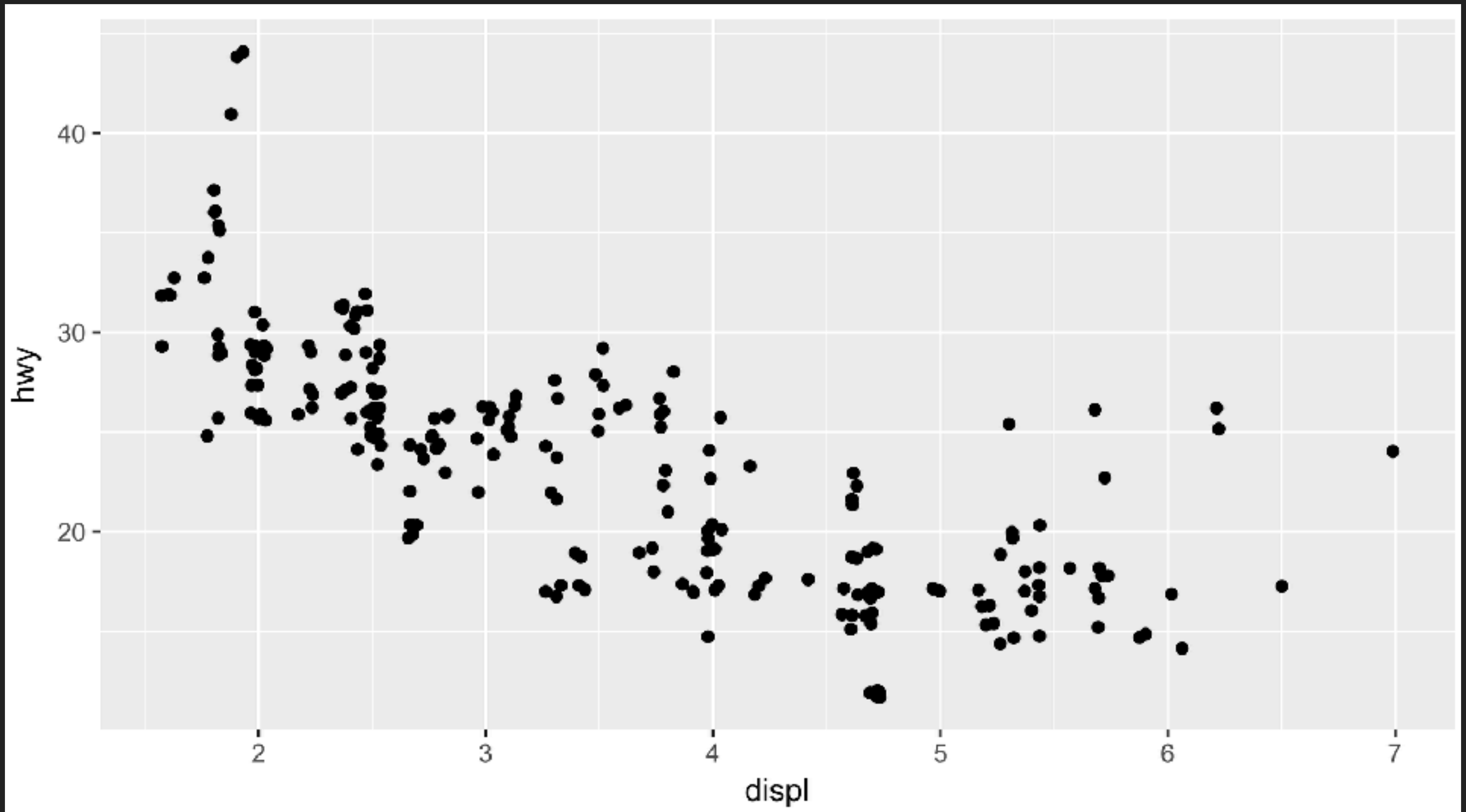
POSITION ADJUSTMENTS – JITTER

```
ggplot2::ggplot(data = dataFrame) +
   geom_point(mapping = aes(x = var1, y = var2),
   position = “jitter”)
```

Example – the mpg data from ggplot2:

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

POSITION ADJUSTMENTS – JITTER

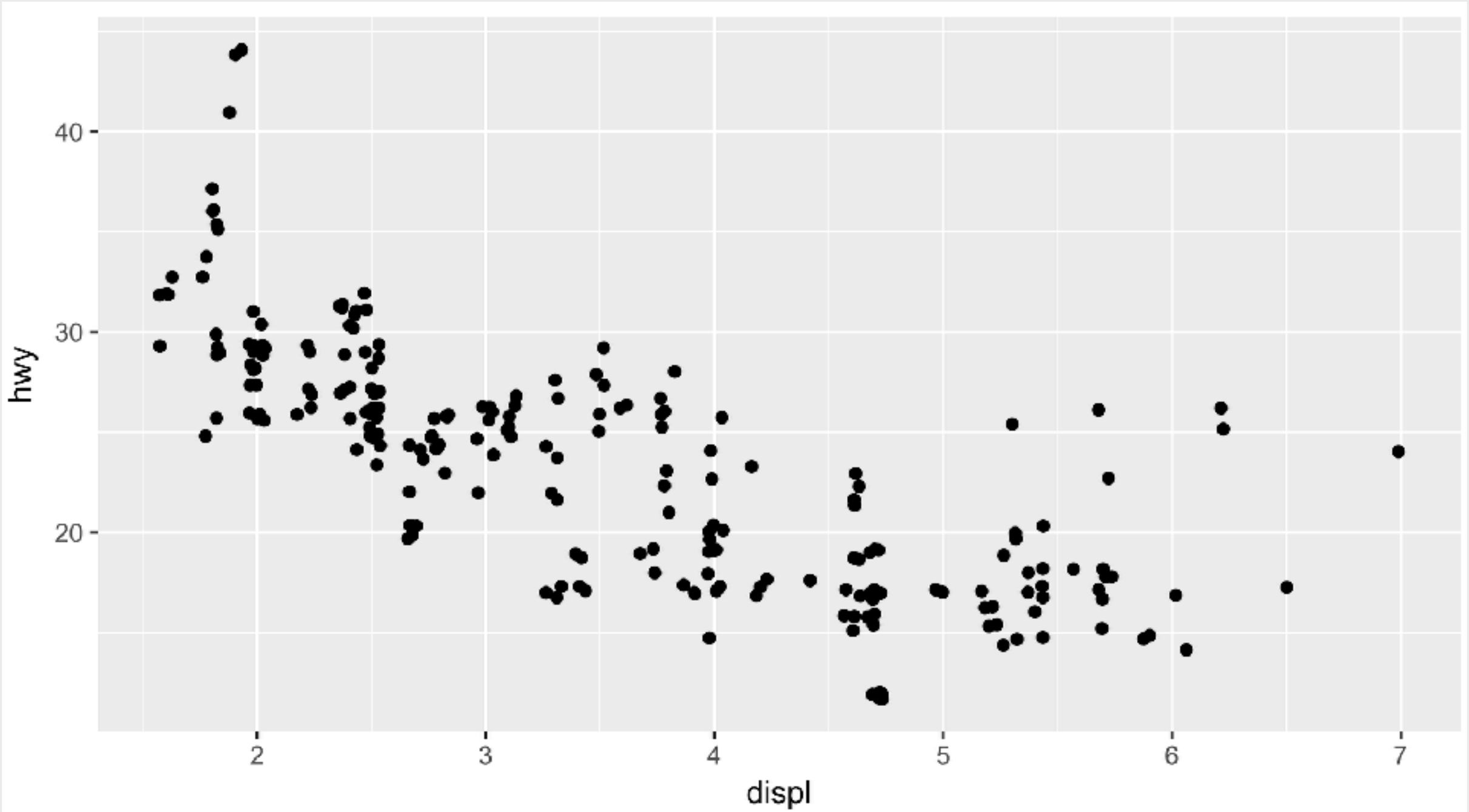


3 COORDINATE SYSTEMS

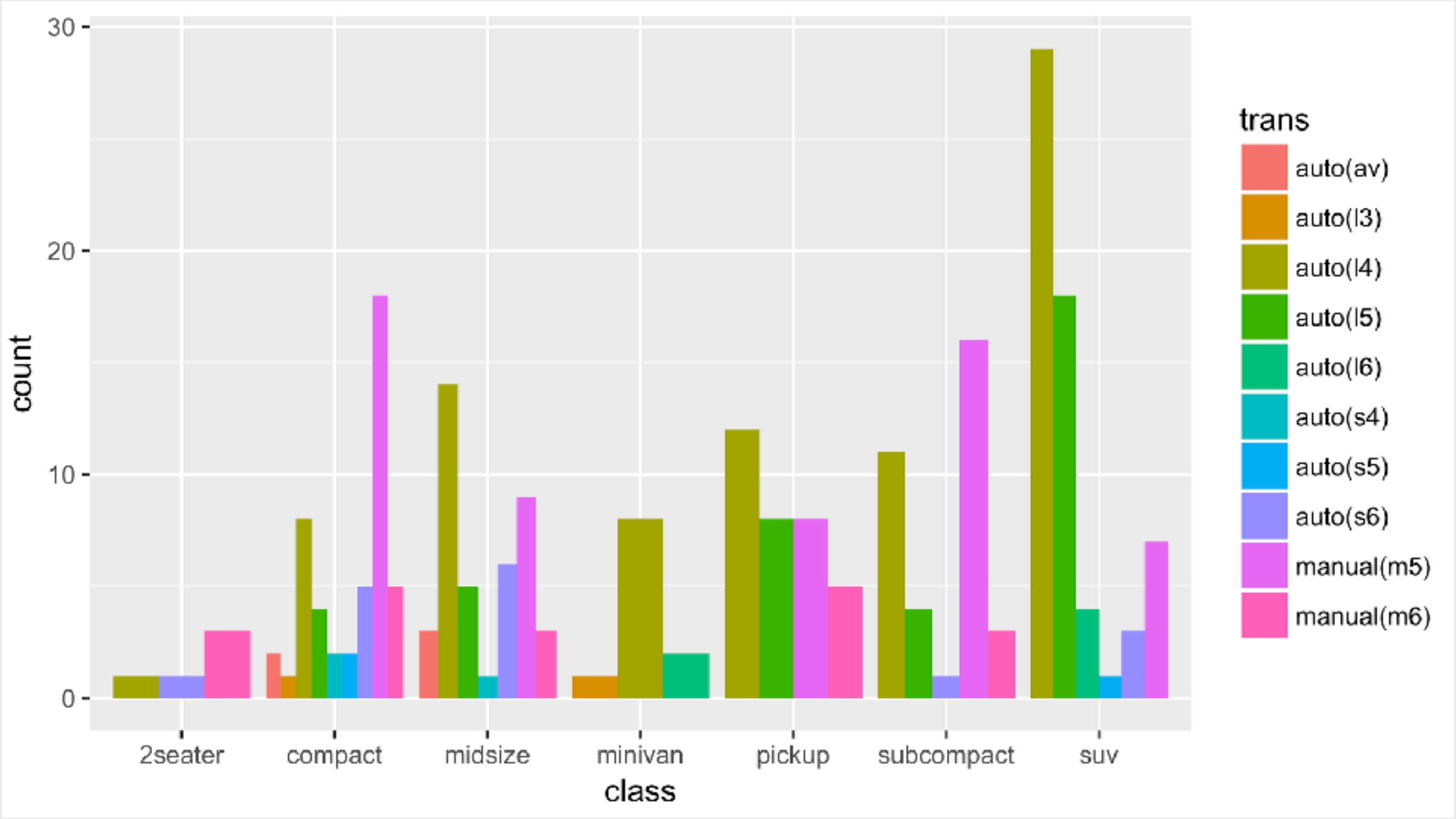
RENÉ DESCARTES

- French/Dutch philosopher and mathematician
- Published the idea of data arrayed along an axes in 1647
- Has been transformed into what we now understand to be Cartesian coordinate systems
- A number of other thinkers wrote about similar ideas independently in the same period

CARTESIAN COORDINATE SYSTEMS



CARTESIAN COORDINATE SYSTEMS



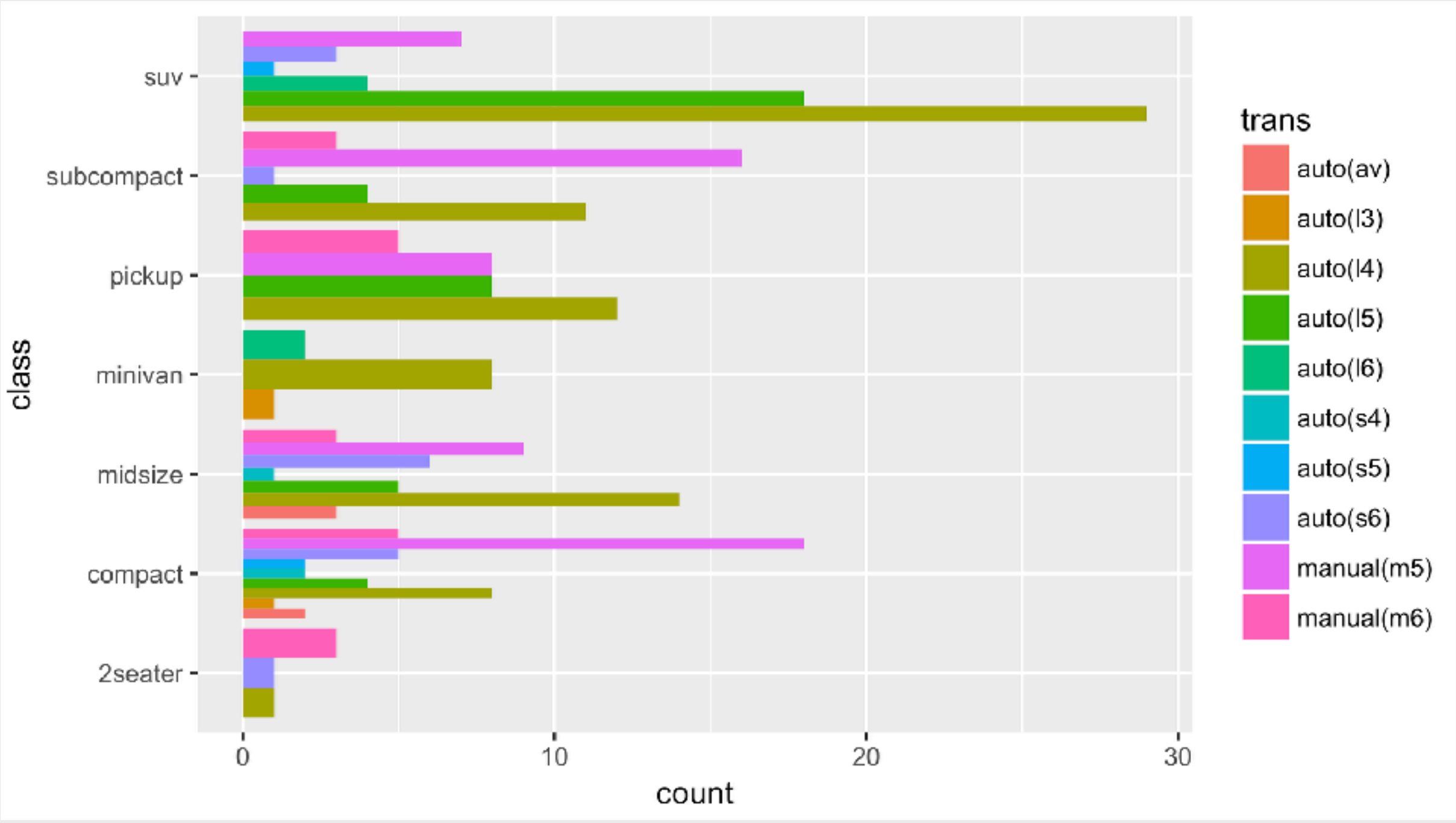
FLIPPING COORDINATES

```
ggplot2::ggplot(data = dataFrame) +
   geom_bar(mapping = aes(x = var, fill = var) +
   coord_flip()
```

Example – the mpg data from ggplot2:

```
   ggplot(data = mpg) +
      geom_bar(mapping = aes(x = class, fill = trans),
        position = “dodge”) +
      coord_flip()
```

FLIPPING COORDINATES



4 THE GRAMMAR OF GRAPHICS

A BASIC TEMPLATE

```
ggplot2::ggplot(data = dataFrame) +
   geom(mapping = aes(aesthetics),
      stat = statistics,
      position = position,
) +
coordinateFunction +
facetFunction
```