

ggplot2 is a grammar for graphics

Wickham, Hadley (2016) ggplot2: Elegant Graphics for Data Analysis. Second Edition. Springer Use R! Series.

Plots are composed of:

- **Data** + set of aesthetic **mappings** describing how variables are mapped to aesthetic attributes
- **Layers** of geometric elements and “transformations”; geometric objects or **geom** correspond to what you see (points, polygons, points, etc.) Statistical transformations or **stat**, summarize the data (e.g., binning and counting observations to create a histogram, summarizing a relationship with a linear model)
- The **scales** map values in the data to values in aesthetic space (color, size, or shape). Draws legends or axes.
- A coordinate system, **coord**, describes how data coordinates are mapped to the graph (cartesian, polar or map projections).

- A faceting specification, or **facet**, describes how to break up data into subsets and how to display those subsets as small multiples.
- A **theme** which controls the finer points of display (e.g., font size or background color).

What the ggplot2 grammar doesn't do:

- Doesn't suggest what graphics you should use to answer the questions you are interested in.
- It does not describe interactivity, only static graphics (other options available such as ggvis);

**drivetrain: front wheel (f), rear wheel (r),
4 wheel (4)**

```
> library(ggplot2)
```

```
> head(mpg)
```

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

```
class
1 compact
2 compact
3 compact
4 compact
5 compact
6 compact
```

engine size

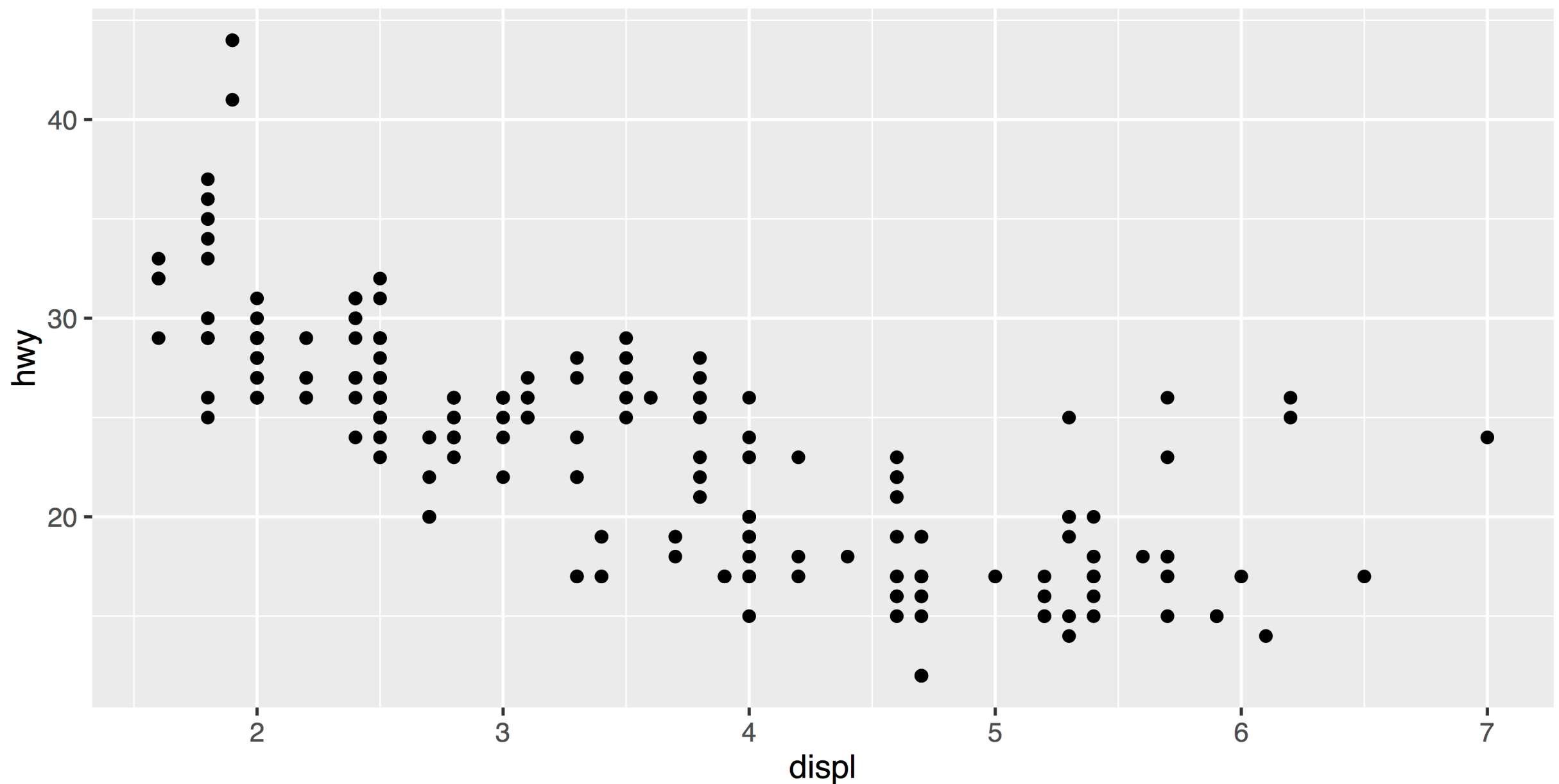
miles per gallon (mpg) for city and highway

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

data →

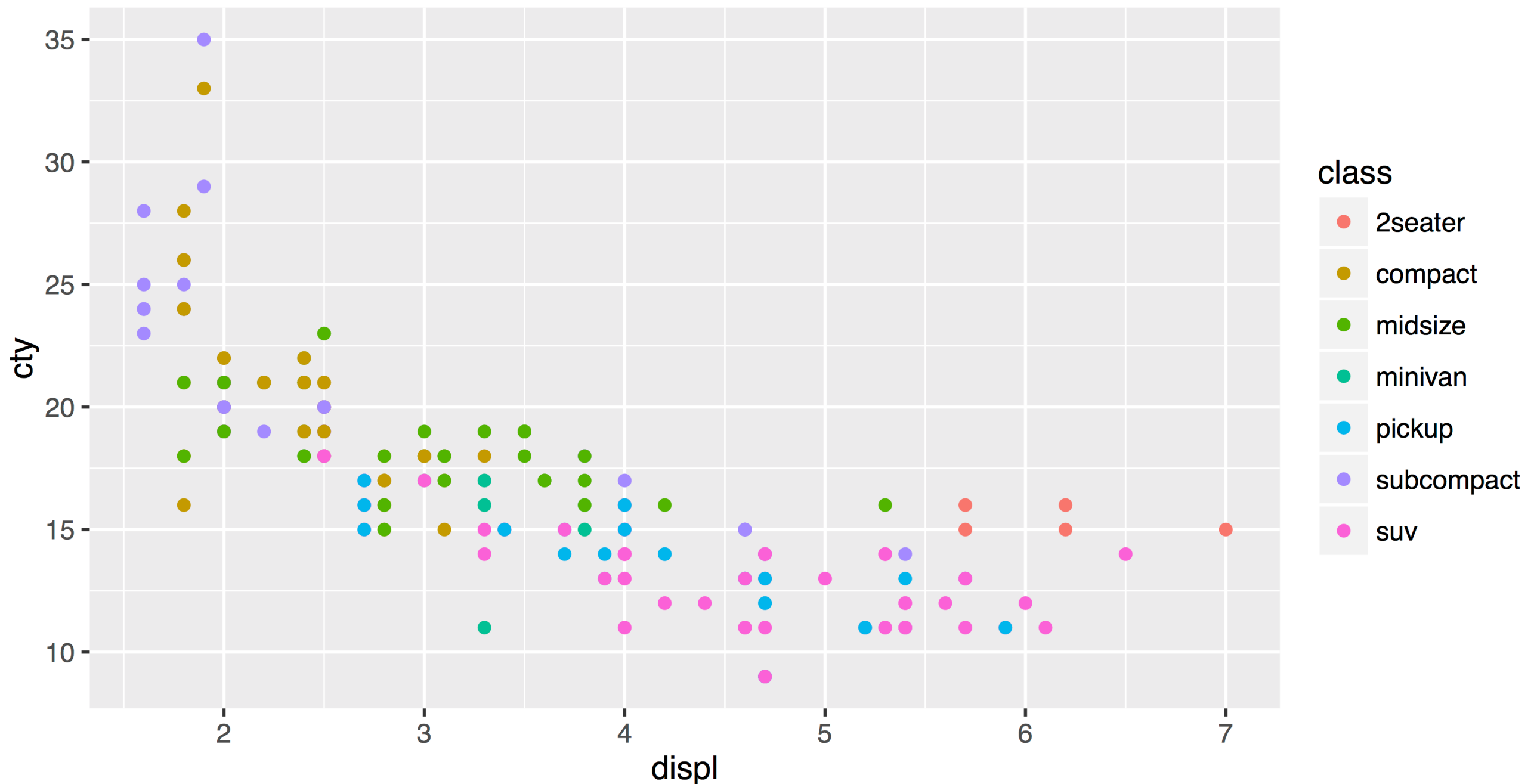
→ **aesthetic mapping: engine size to x,
fuel economy to y**

→ **layer: points**



- Aesthetic attributes: color, size, shape

```
> ggplot(mpg, aes(displ, cty, color=class)) +  
+   geom_point()
```



- If you want a single color for all the points:

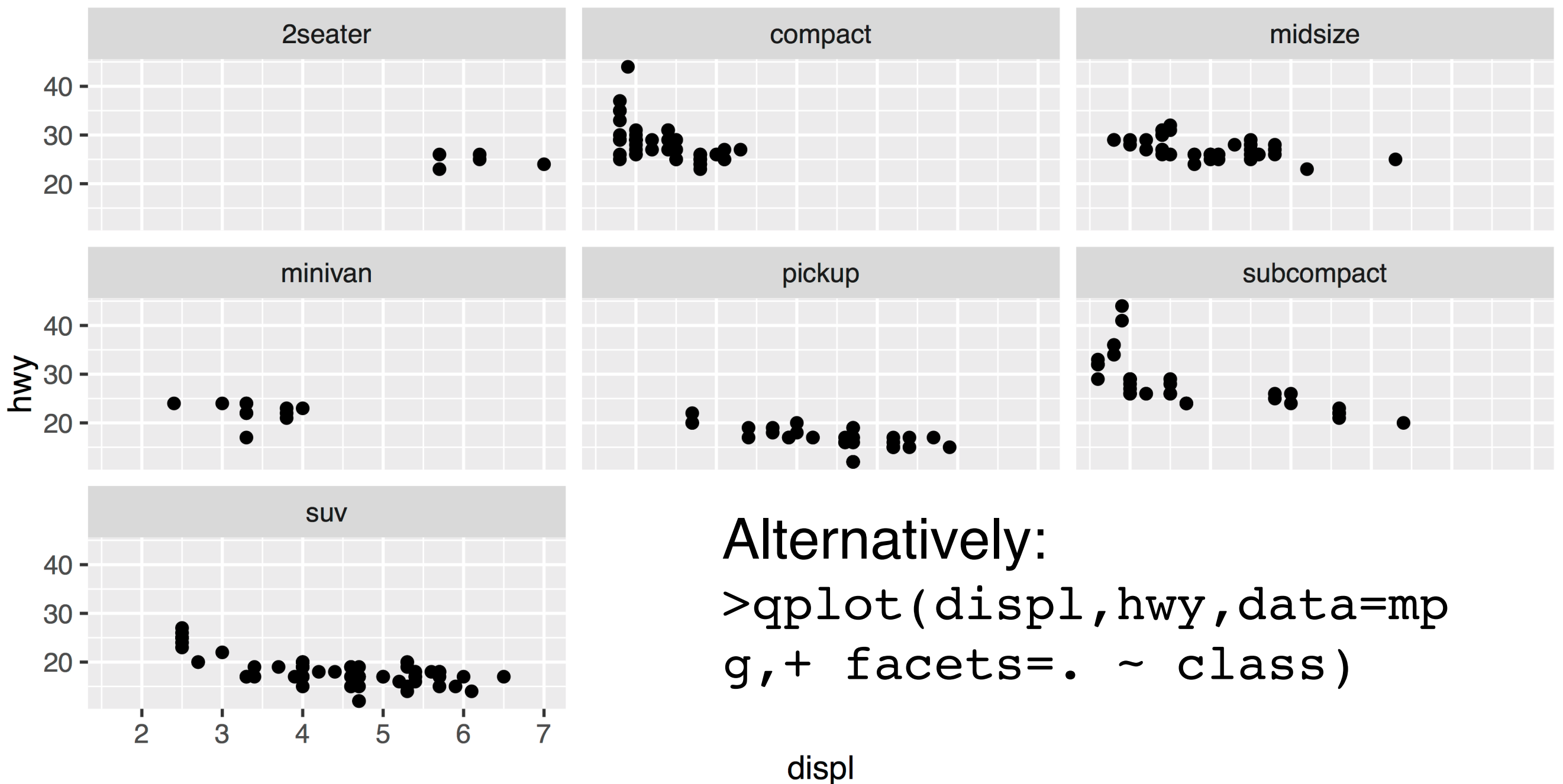
```
> ggplot(mpg, aes(displ, cty)) +  
+   geom_point(color="blue")
```

instead of:

```
> ggplot(mpg, aes(displ, cty)) +  
+   geom_point(aes(color="blue"))
```

Faceting: technique for displaying additional categorical variables. It creates tables of graphics by splitting the data into subsets and displaying the same graph for each subset. Two types: grid and wrapped. Wrapped is the most useful.

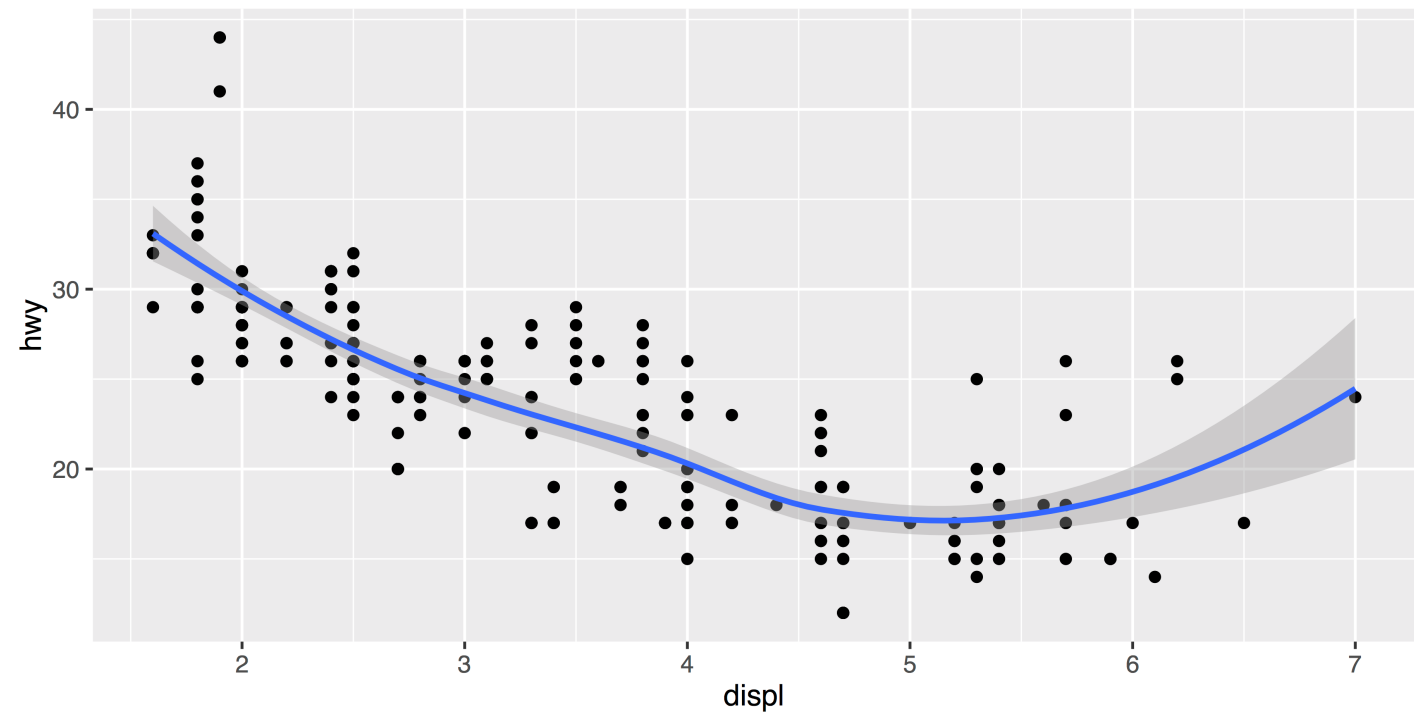
```
> ggplot(mpg, aes(displ, hwy)) +  
  + geom_point() +  
  + facet_wrap(~class)
```



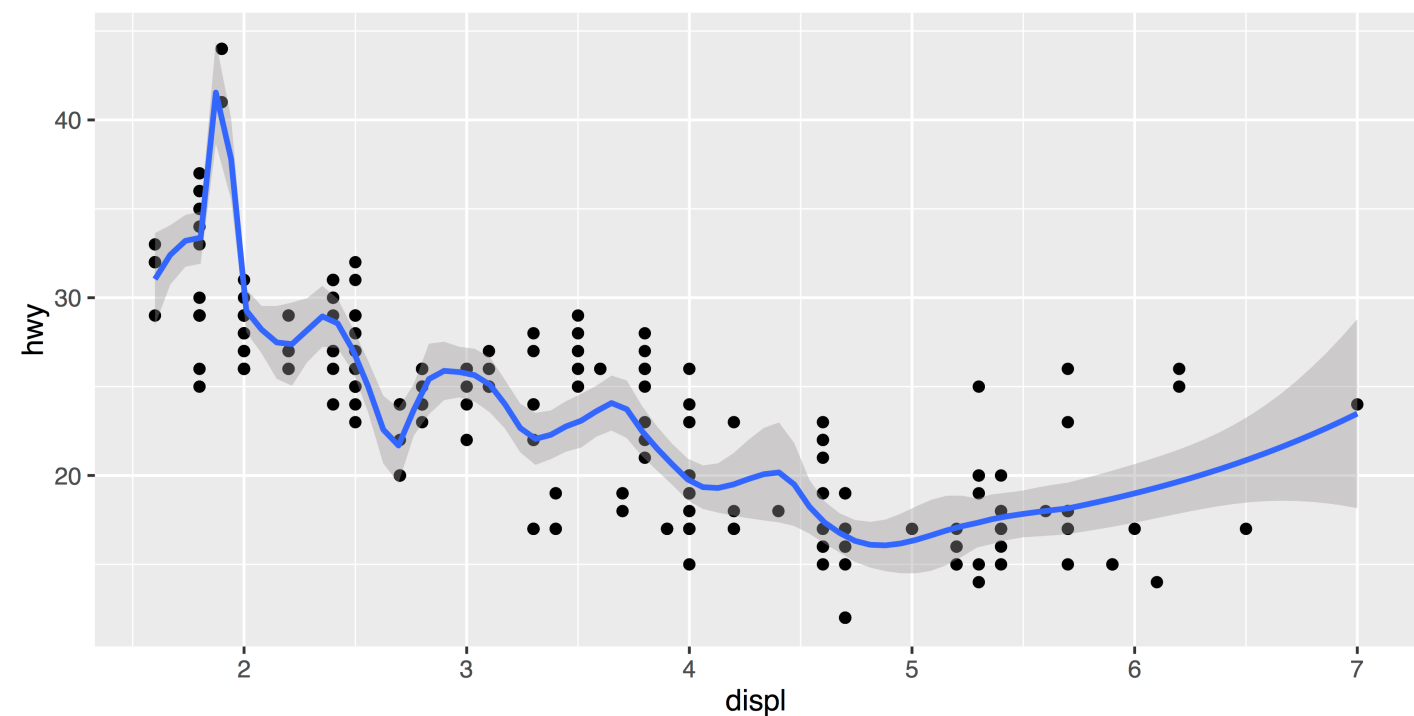
We also have different types of geom functions (these are not the only ones):

- `geom_smooth()` : fits a smoother to the data, displays smooth function + standard error. Default option is `lowess`
- `geom_boxplot()` : summarizes the distribution of a set of points using boxplots
- `geom_histogram()` and `geom_freqpoly()` : show distributions of continuous variables
- `geom_bar()` : barplots for categorical variables
- `geom_path()` and `geom_line()` : in line plots lines travel left to right, while paths can go in any direction

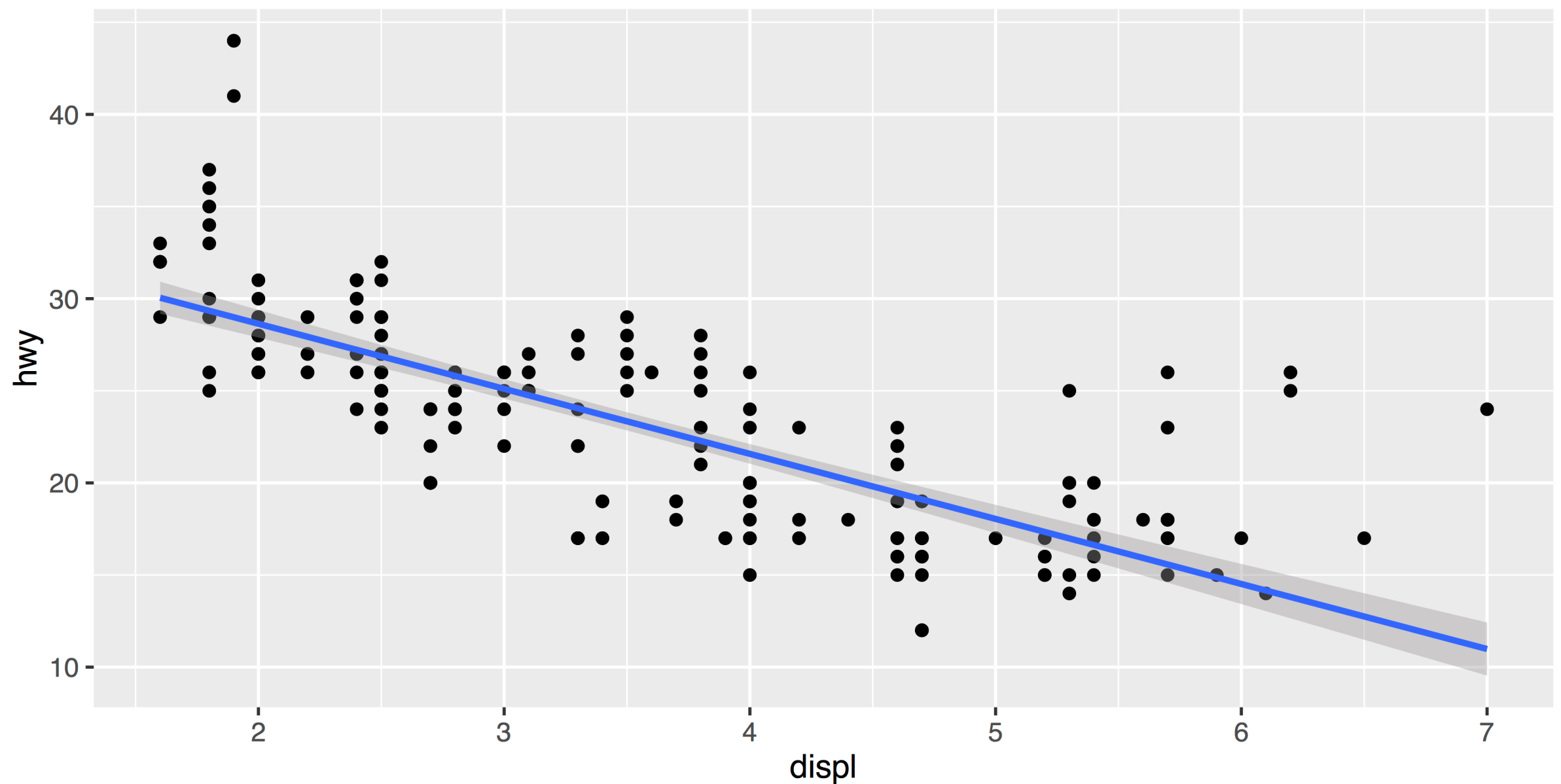
```
> ggplot(mpg, aes(displ, hwy)) +  
+   geom_point() +  
+   geom_smooth()
```



```
> ggplot(mpg,  
aes(displ, hwy)) +  
+   geom_point() +  
+   geom_smooth(span=0.2)
```

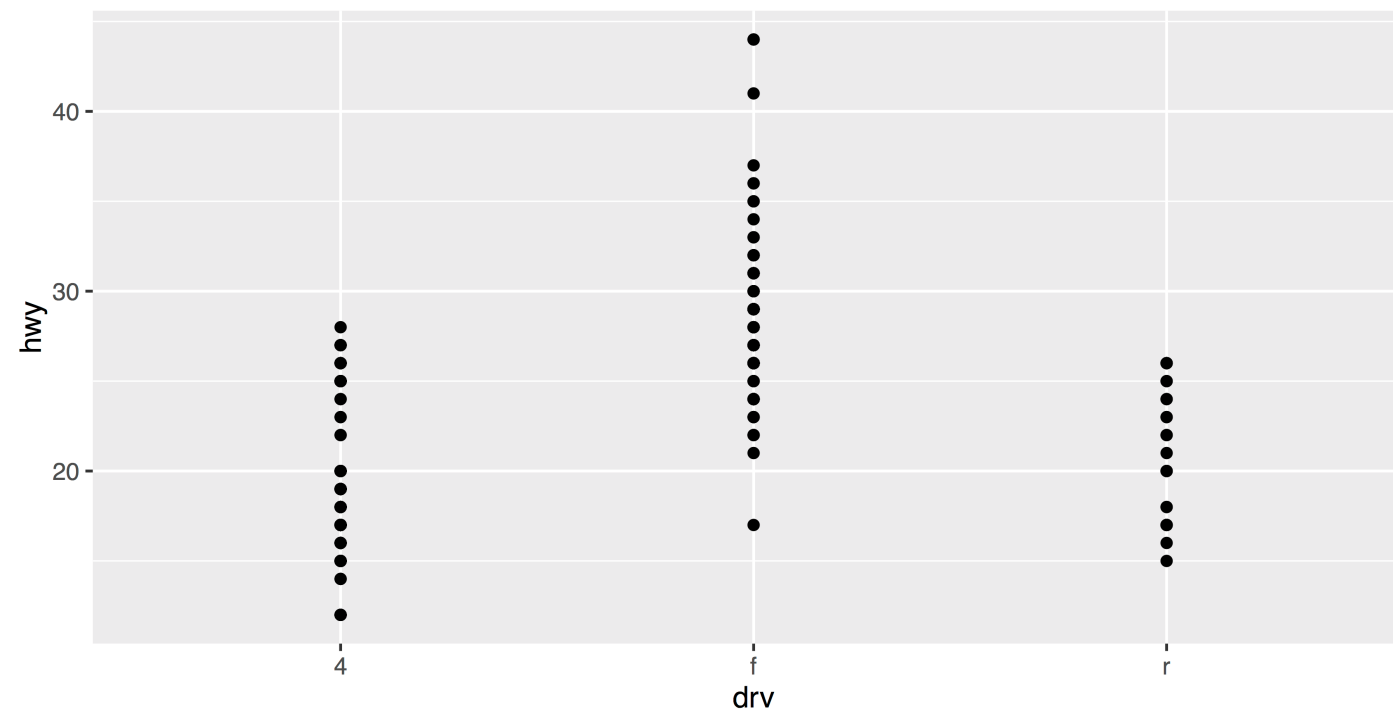


```
> ggplot(mpg, aes(displ, hwy)) + geom_point() +  
+   geom_smooth(method="lm")
```

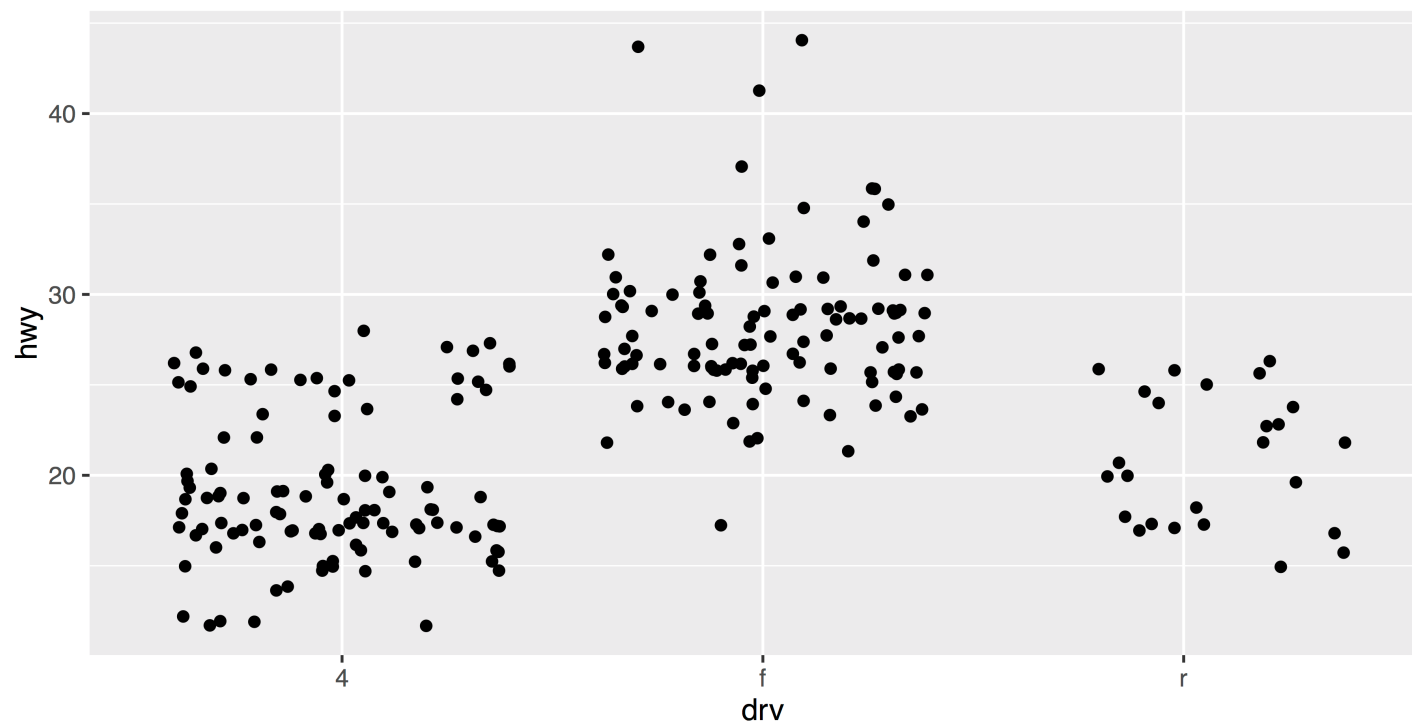


Including categorical variables:

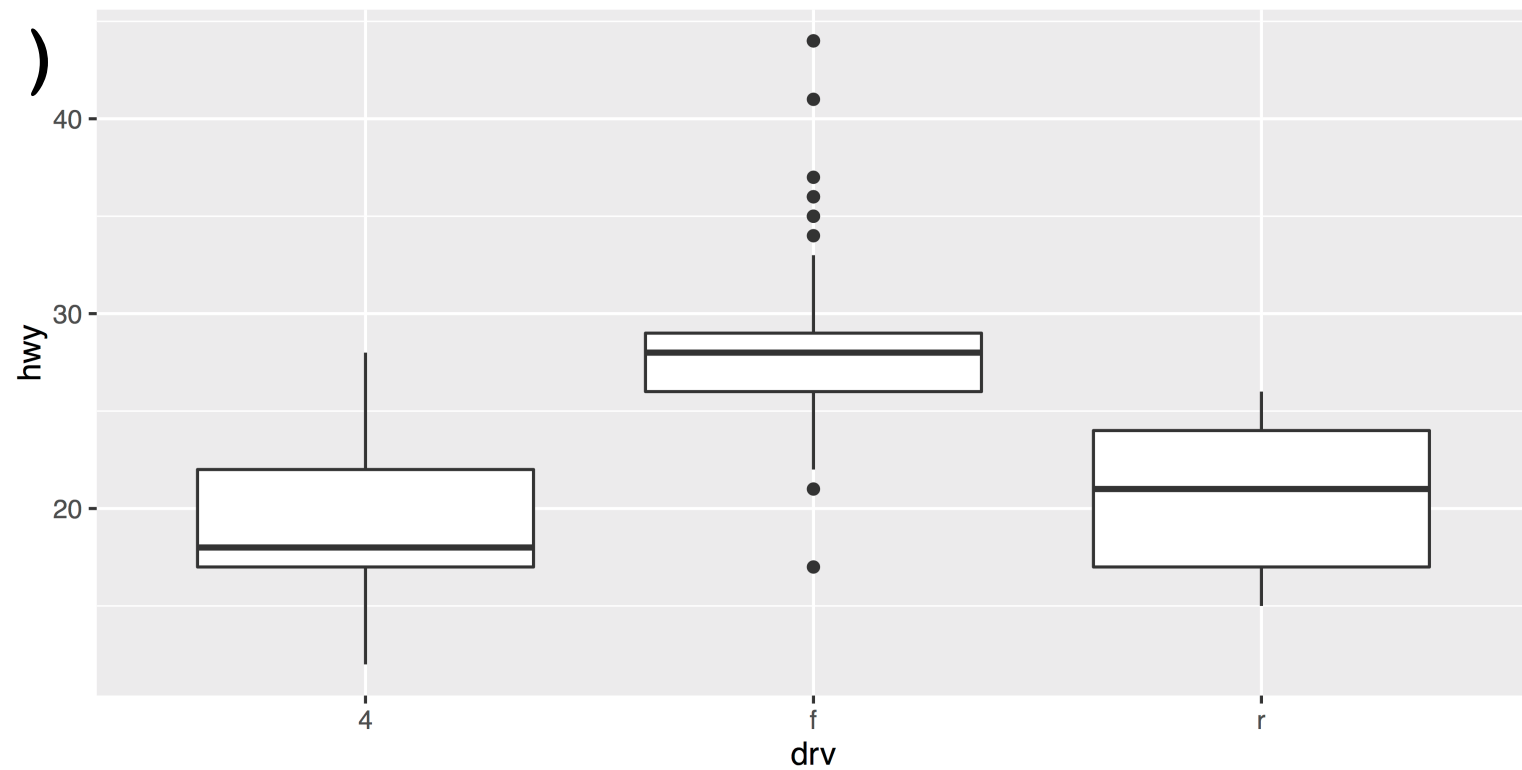
```
> ggplot(mpg, aes(drv, hwy)) + geom_point()
```



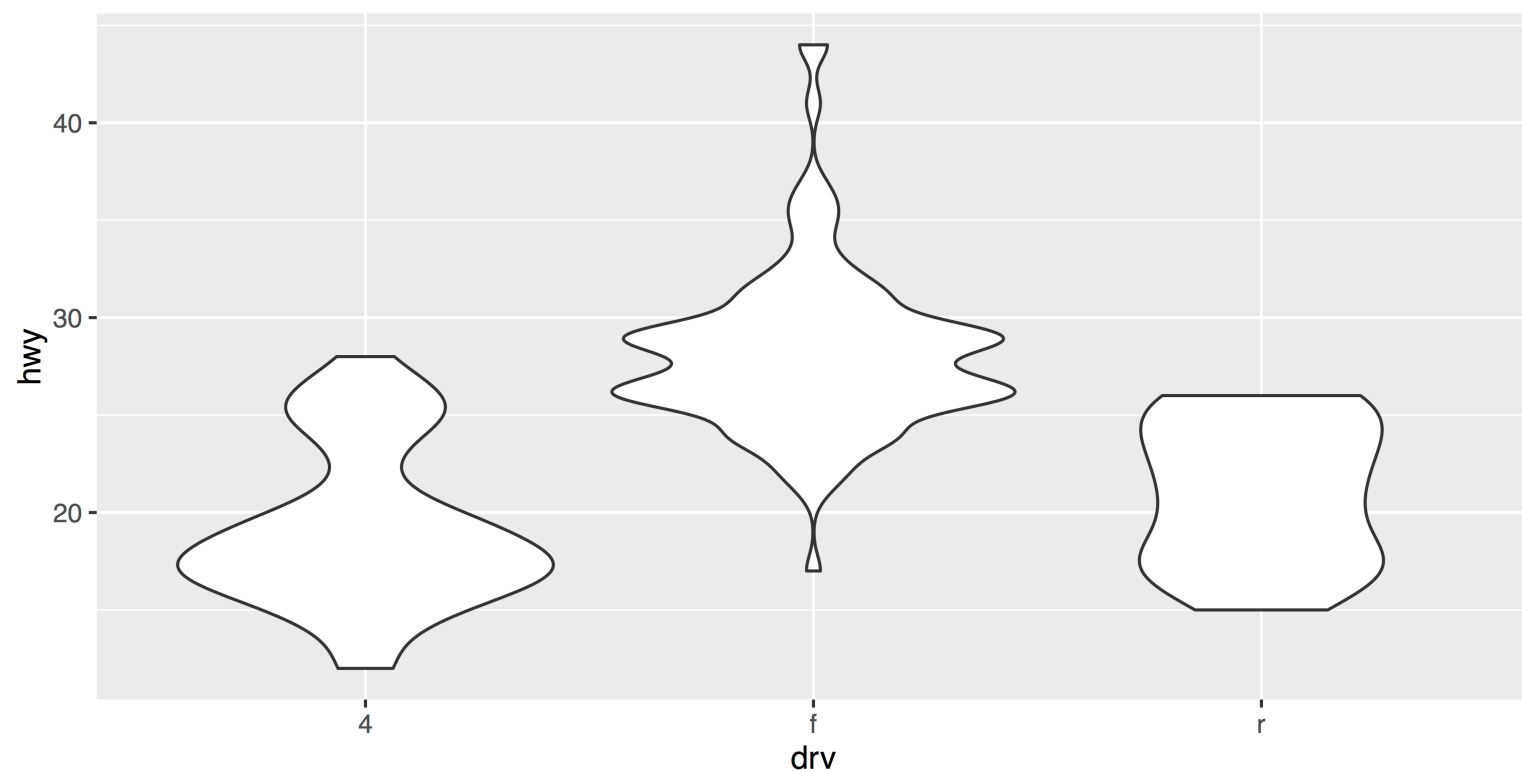
```
> ggplot(mpg, aes(drv, hwy)) + geom_jitter()
```



```
> ggplot(mpg, aes(drv, hwy)) +  
+   geom_boxplot()
```

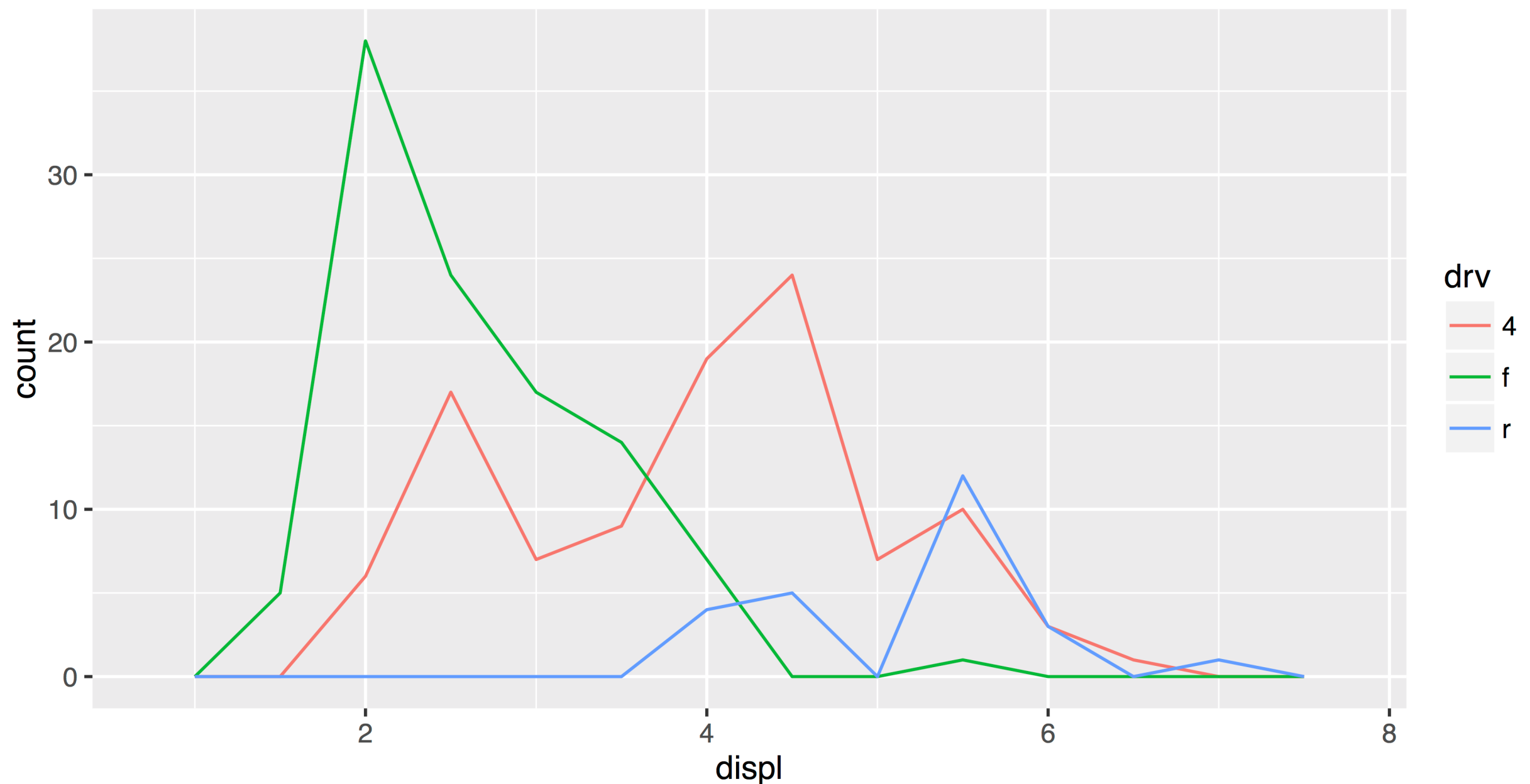


```
> ggplot(mpg, aes(drv, hwy)) +  
+   geom_violin()
```

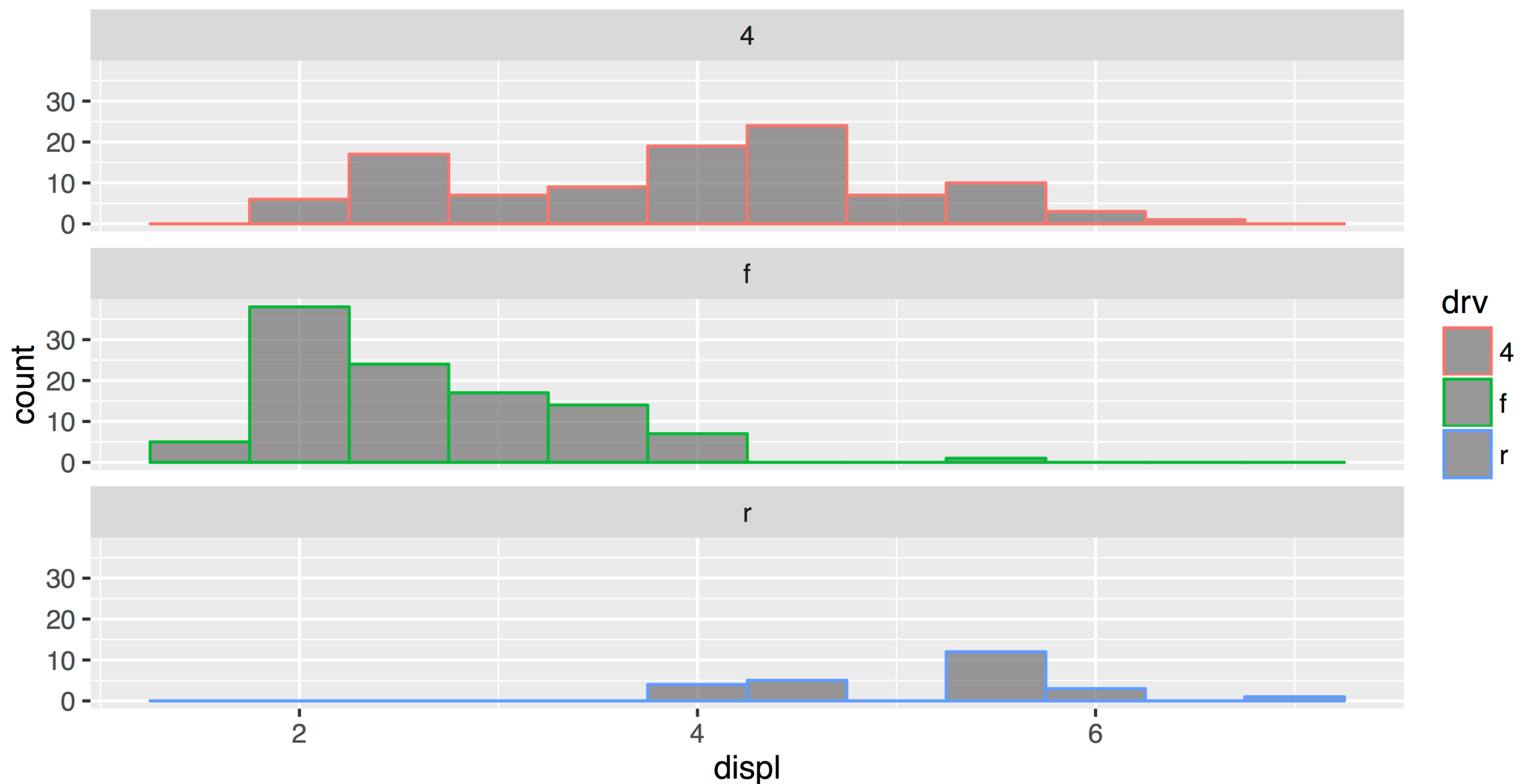


Histograms & frequency polygons:

```
> ggplot(mpg, aes(displ, color=drv)) +  
  + geom_freqpoly(binwidth=0.5)
```



```
> ggplot(mpg, aes(x=displ, color=drv)) +  
+   geom_histogram(binwidth=0.5, alpha=0.6) +  
+   facet_wrap(~drv, ncol=1)
```



- Maps: you can use `library(maps)`
- You can also use `ggmap`