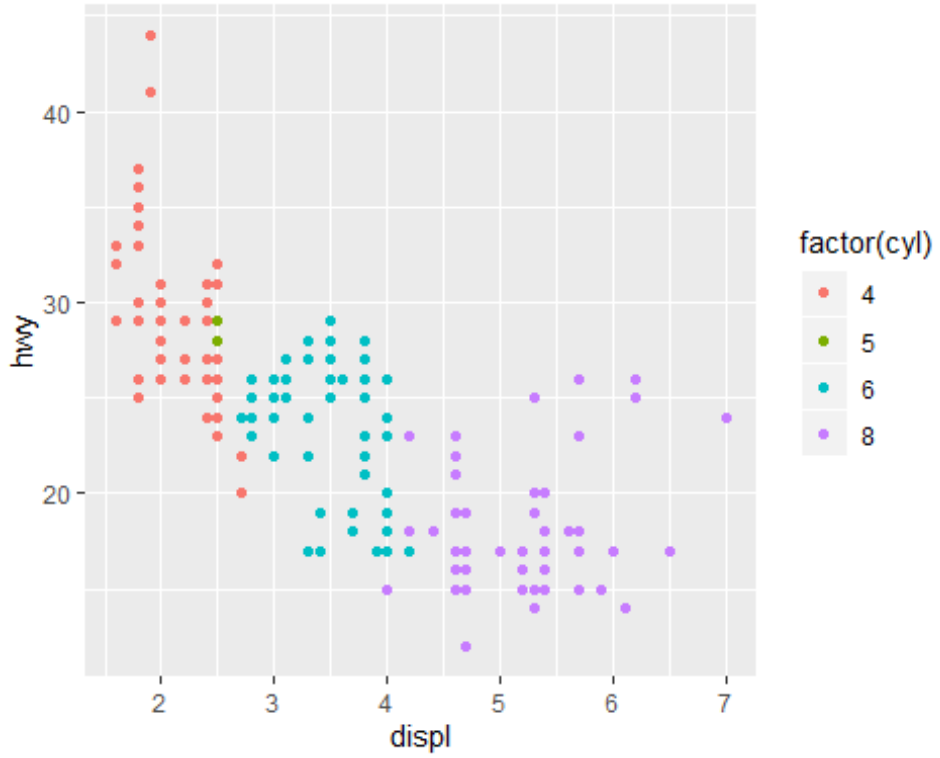
chapter 4 : Mastering the Grammar

This chapter describes the theoretical basis of ggplot2: the layered grammar of graphics.

```
library(ggplot2)
```

4.2 Building a Scatterplot

```
ggplot(mpg, aes(displ, hwy, color = factor(cyl))) +
  geom_point()
```



How does ggplot2 draw this plot?

4.2.1 Mapping Aesthetics to Data

scatterplot : each points has horizontal and vertical position, size, color, shape - these arrtibutes called **aesthetics**

each aesthetic can be mapped to a variable, or set to a constant value. We can create many different types of plots using this data. **Geoms** determine the “type” of the plot.

4.2.2 Scaling

We need to convert them from data units to graphical units that computer can display. - this conversion process is called **scaling**

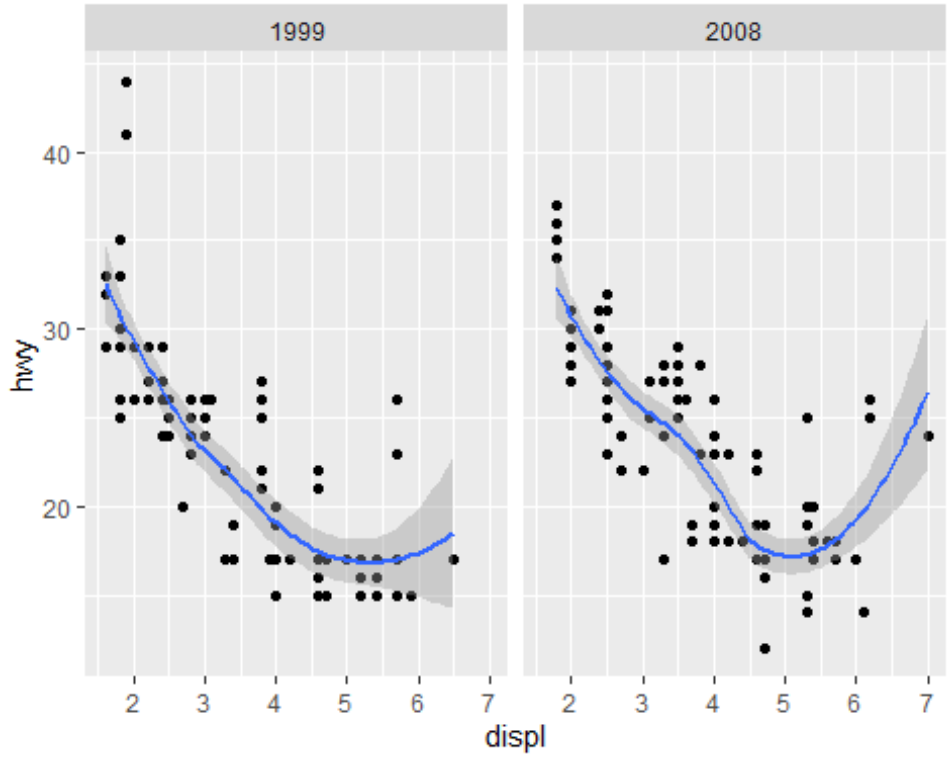
We don't use exact pixels (for position) because the drawing system that ggplot2 uses **grid**, take care of that final conversion for us. Final step determines how the two positions (x and y) are combined to form the final location on the plot. Done by the coordinate system **coord**.

Finally, need to render this data to create the graphical objects - combine graphical objects form three sources : *data*, *scales and coordinate system*, *plot annotations*

4.3 Adding Complexity

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

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



- New components : facets, multiple layers, statistics

The smooth layer is different to the point layer. It displays a statistical transformation of the data. -> requires an additional step in the process : after mapping the data to aesthetics, the data is passed to a statistical transformation, **stat**.

Scaling occurs in three parts : transforming, training, mapping - Scale transformation occurs before statistical transformation. - after statistics are computed, each scale is trained

on every dataset from all the layers and facets. - finally the scales map the data values into aesthetic values. This is a local operation.

4.4 Components of the Layered Grammar

data, mappings, stat, geom, position adjustment form a **layer**.

4.4.1 Layers

- A layer is composed of five parts :

1. Data
2. Aesthetic mappings
3. A statistical transformation (stat)
4. A geometric object (geom)
5. A position adjustment

4.4.2 Scales

- scale controls the mapping from data to aesthetic attributes
- we need a scale for every aesthetic used on a plot

A scale is a function and its inverse, along with a set of parameters. The inverse function is used to draw a guide so that you can read values from the graph. Guides are either axes (for position scales) or legends (for everything else).

4.4.3 Coordinate System

coord(coordinate system) maps the position of objects onto the plane of the plot.

- scaling is performed before statistical transformation, while coordinate transformations occur afterward

4.4.4 Facetting

This is a powerful tool when investigating whether patterns hold across all conditions.