ggplot2 package – qplot

Steven Chiou

This document gives a brief introduction to one of R's most powerful package for visualization – the ggplot2 package with an emphaze in the qplot function.

The two main functions in ggplot2 are ggplot and qplot. Of those, qplot is made to mimic the format of plot from the base R, requires less syntax for many common tasks, and is more beginner-friendly.

**Motivating example We will use the mpg dataset to demanstrate some key features of 'qplot.

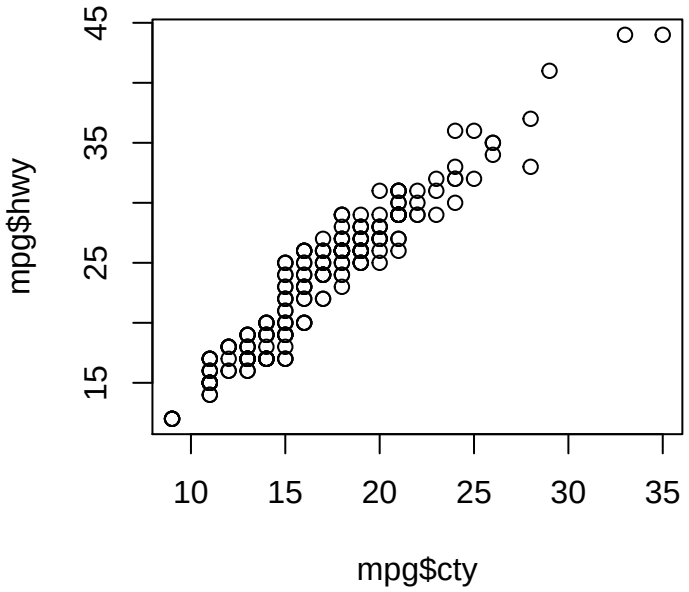
```
> library(ggplot2)
> mpg
# A tibble: 234 x 11
   manufacturer model displ  year   cyl trans drv     cty   hwy fl    cla~
   <chr>        <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <ch>
 1 audi         a4      1.8  1999     4 auto~ f        18    29 p     com~
 2 audi         a4      1.8  1999     4 manu~ f        21    29 p     com~
 3 audi         a4      2    2008     4 manu~ f        20    31 p     com~
 4 audi         a4      2    2008     4 auto~ f        21    30 p     com~
 5 audi         a4      2.8  1999     6 auto~ f        16    26 p     com~
 6 audi         a4      2.8  1999     6 manu~ f        18    26 p     com~
 7 audi         a4      3.1  2008     6 auto~ f        18    27 p     com~
 8 audi         a4 q~   1.8  1999     4 manu~ 4        18    26 p     com~
 9 audi         a4 q~   1.8  1999     4 auto~ 4        16    25 p     com~
10 audi         a4 q~   2    2008     4 manu~ 4        20    28 p     com~
# ... with 224 more rows
```

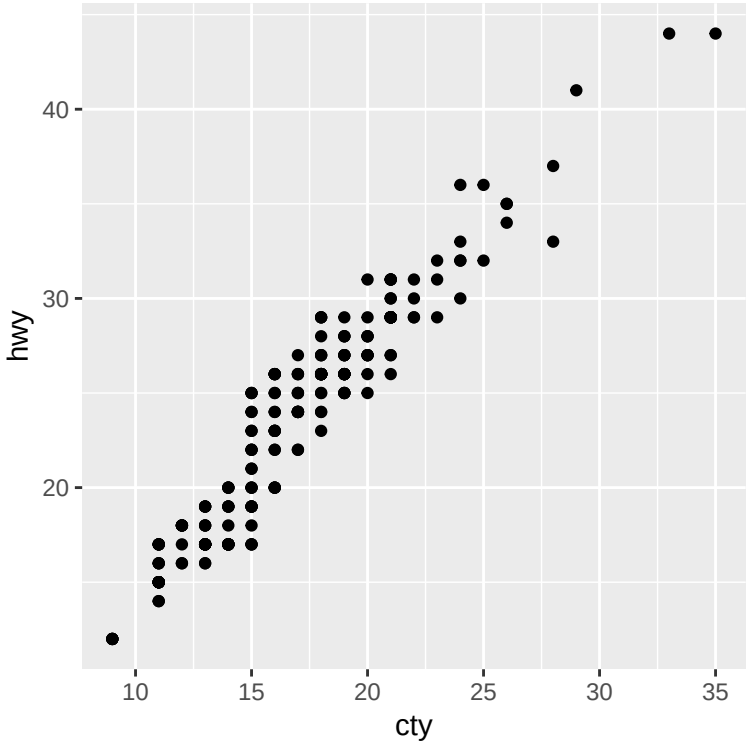
As with plot, the first two arguments in qplot are the "x"" and "y". On the other hand, ggplot relies on aesthetic mapping.

```
> args(qplot)
function (x, y, ..., data, facets = NULL, margins = FALSE, geom = "auto",
    xlim = c(NA, NA), ylim = c(NA, NA), log = "", main = NULL,
    xlab = NULL, ylab = NULL, asp = NA, stat = NULL, position = NULL)
NULL
> args(ggplot)
function (data = NULL, mapping = aes(), ..., environment = parent.frame())
NULL
```

Here is a simple histogram comparison:

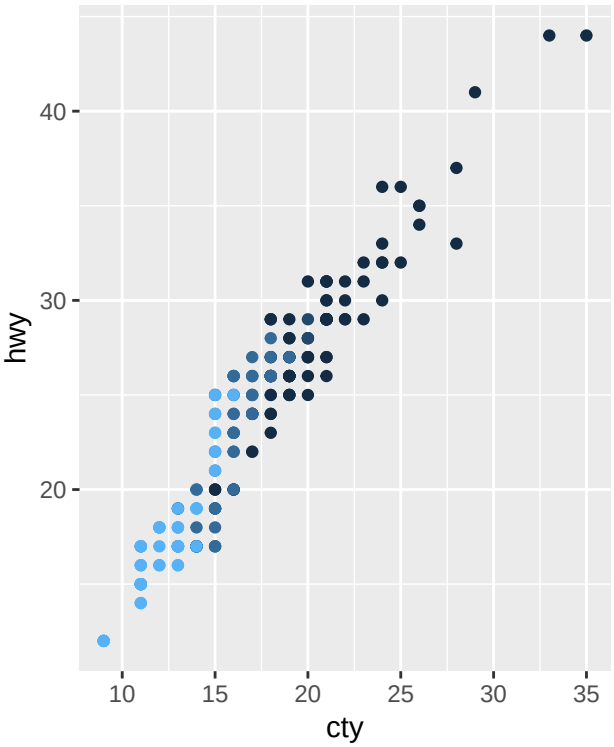
```
> plot(mpg$cty, mpg$hwy)
> qplot(cty, hwy, data = mpg)
```

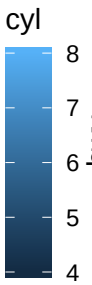


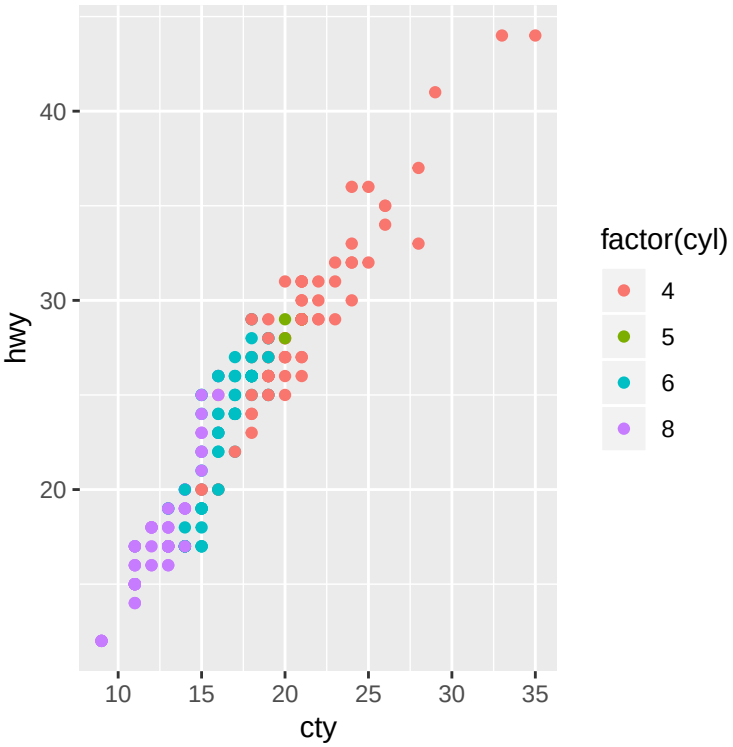


You can add a color layer based on information about `cyl`.

```
> qplot(cty, hwy, data = mpg, color = cyl)
> qplot(cty, hwy, data = mpg, color = factor(cyl))
```

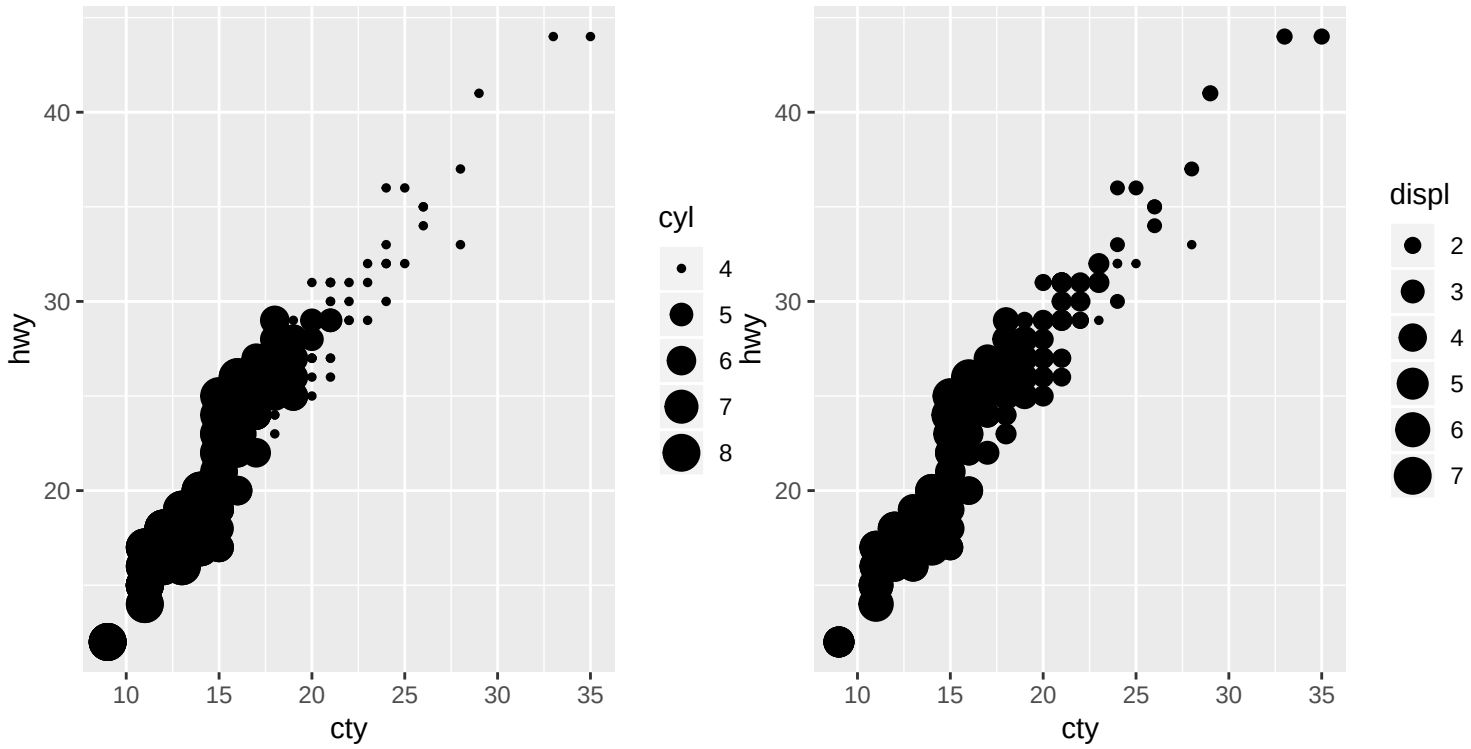






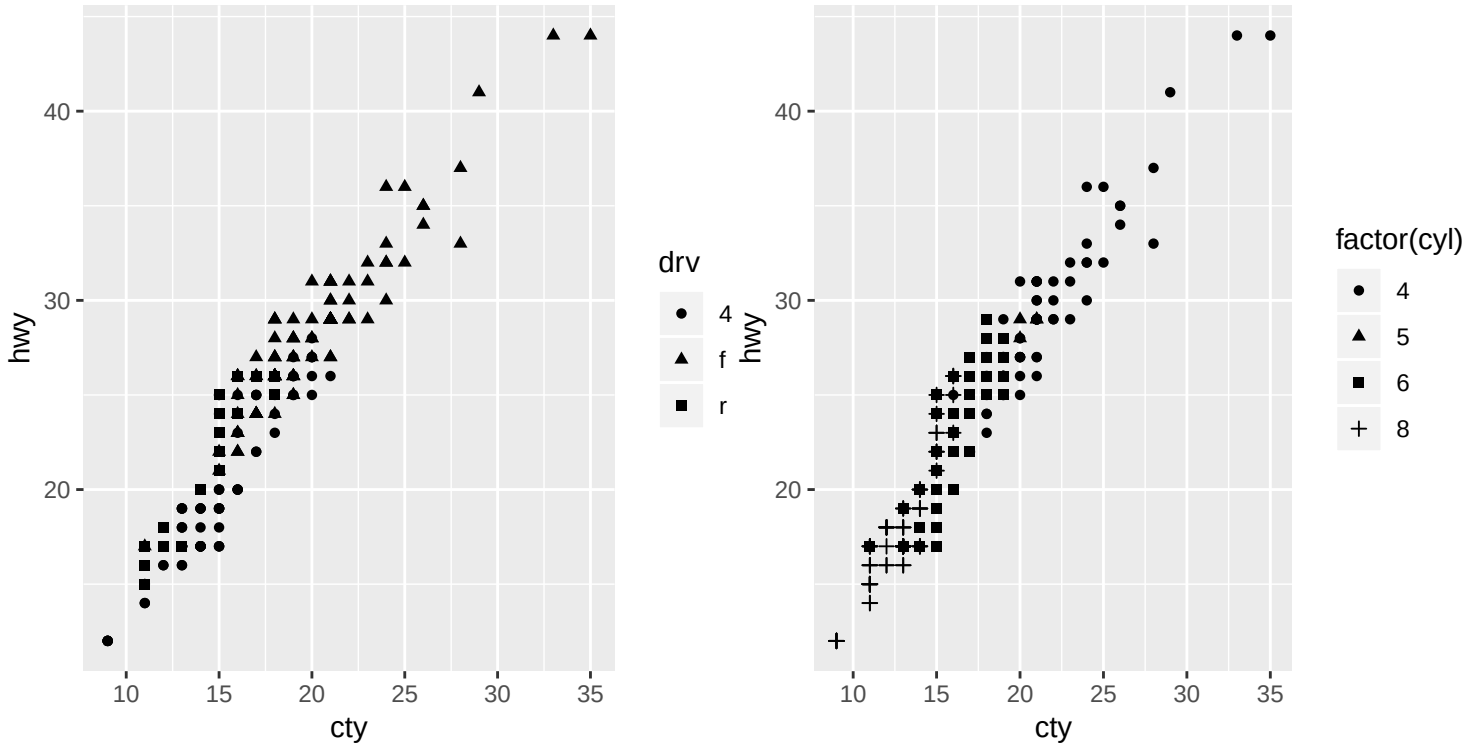
Or a size layer.

```
> qplot(cty, hwy, data = mpg, size = cyl)
> qplot(cty, hwy, data = mpg, size = displ)
```



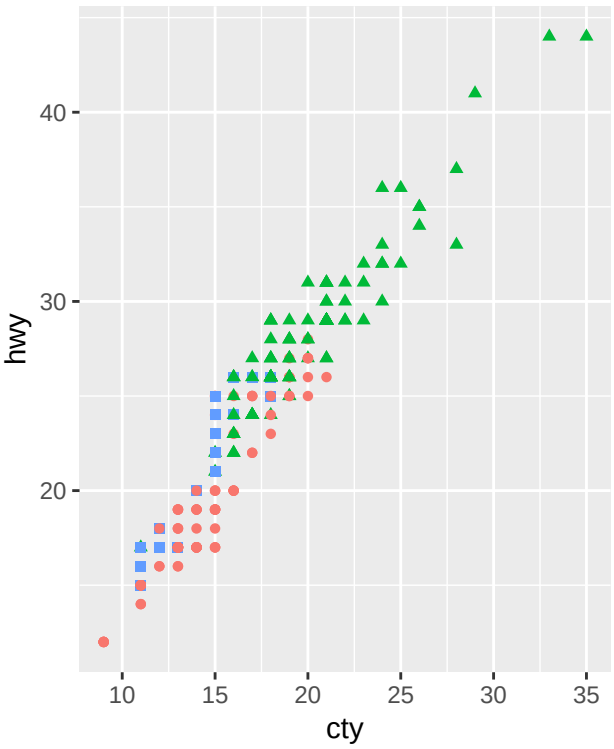
Even shapes.

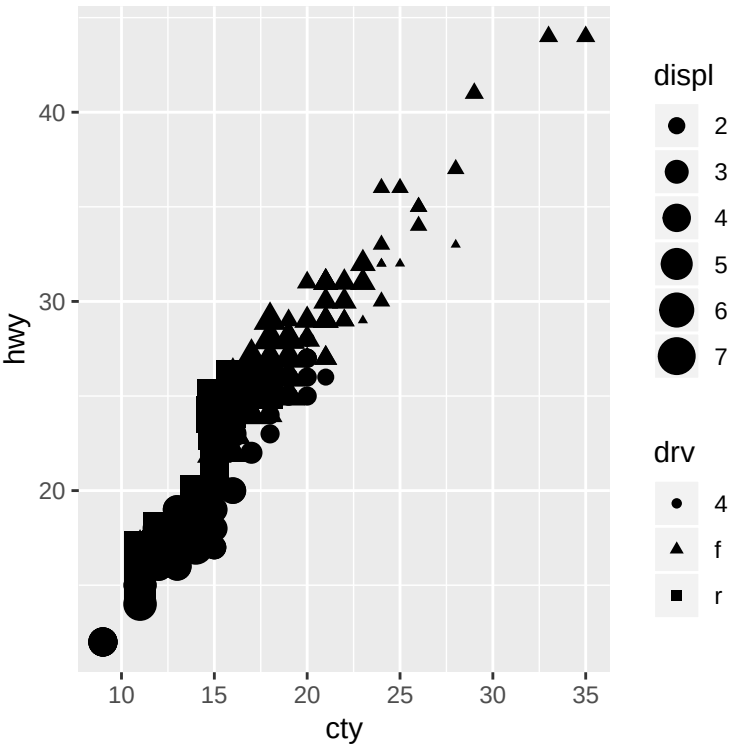
```
> qplot(cty, hwy, data = mpg, shape = drv)
> qplot(cty, hwy, data = mpg, shape = factor(cyl))
```



With combine mapping.

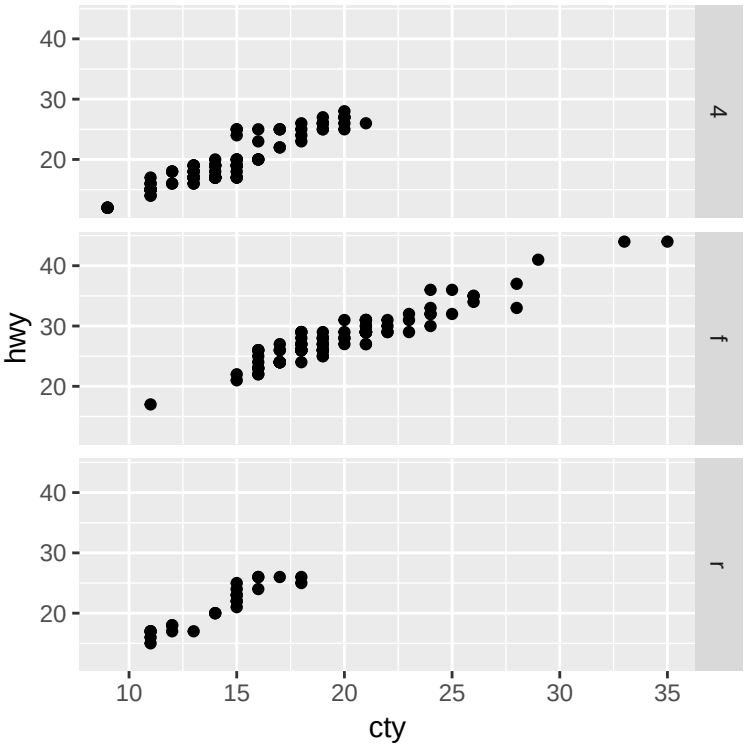
```
> qplot(cty, hwy, data = mpg, shape = drv, color = drv)
> qplot(cty, hwy, data = mpg, shape = drv, size = displ)
```

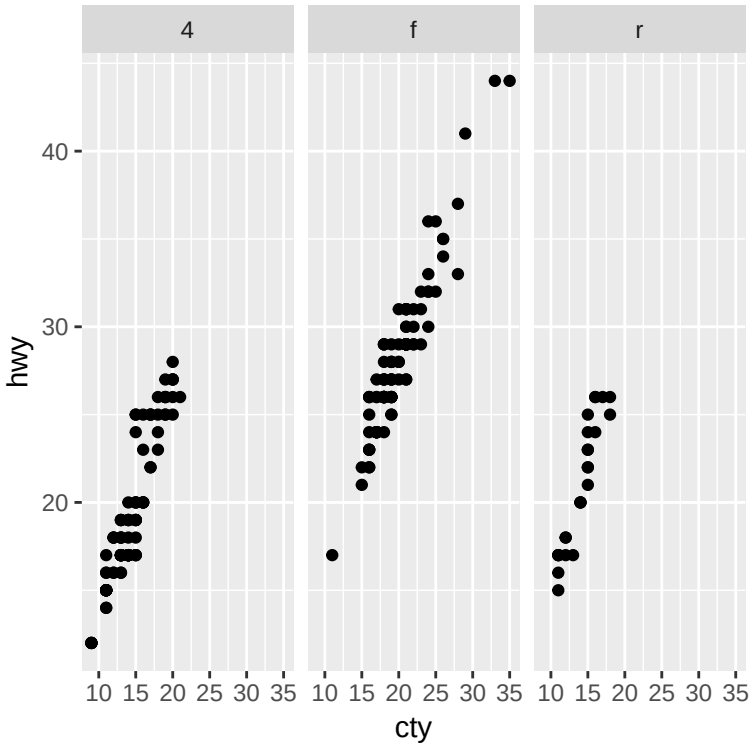




Different types of aesthetic attributes work better with different types of variables. An alternative is to use "faceting", a mechanism for automatically laying out multiple plots on a page.

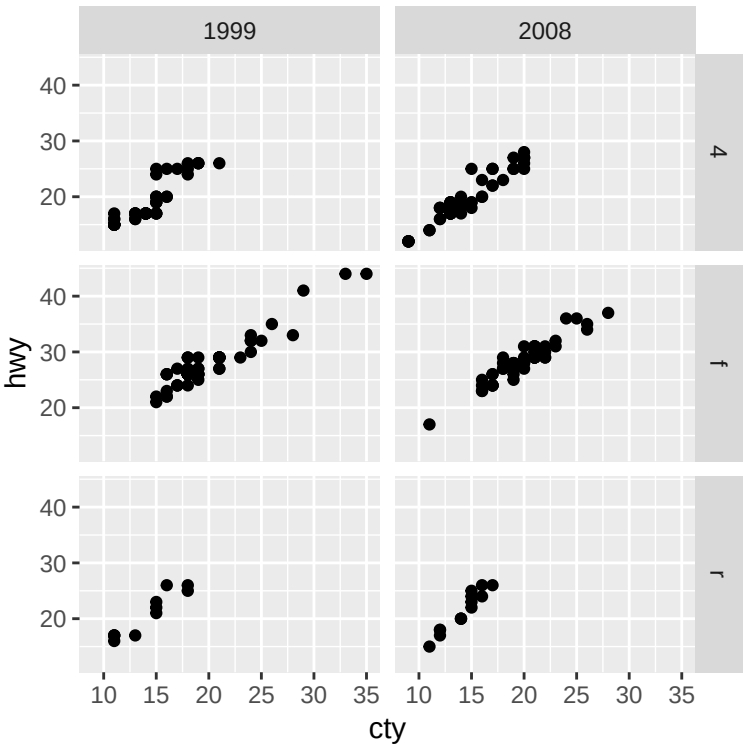
```
> qplot(cty, hwy, data = mpg, facets = drv ~ .)
> qplot(cty, hwy, data = mpg, facets = . ~ drv)
```

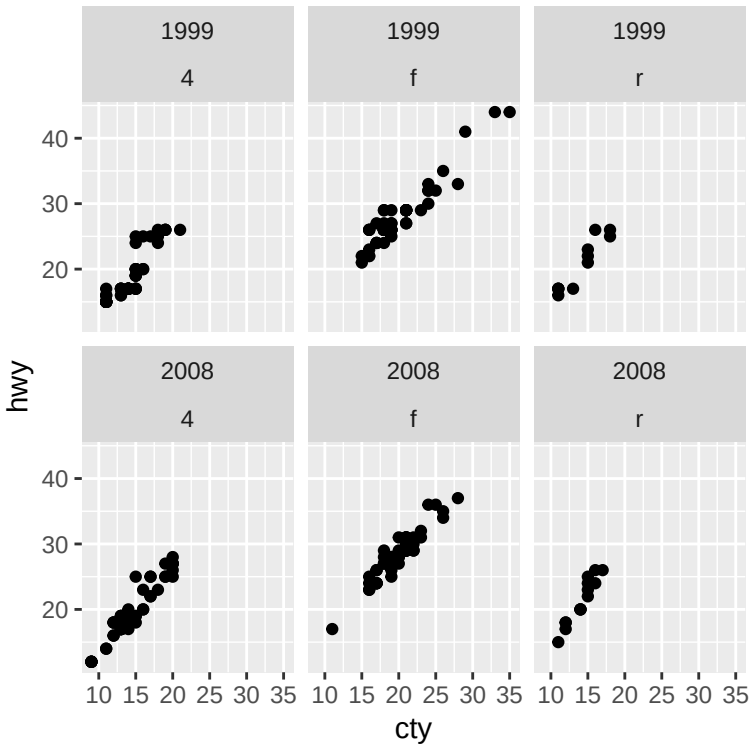




More examples

```
> qplot(cty, hwy, data = mpg, facets = drv ~ year)
> qplot(cty, hwy, data = mpg, facets =  ~ year + drv)
```


1999
2008
4
f
r
hwy
cty
40
30
20
10
15
20
25
30
35


1999
4
1999
f
1999
r
2008
4
2008
f
2008
r
hwy
cty
40
30
20
10
15
20
25
30
35