Introduction

Hadley Wickham

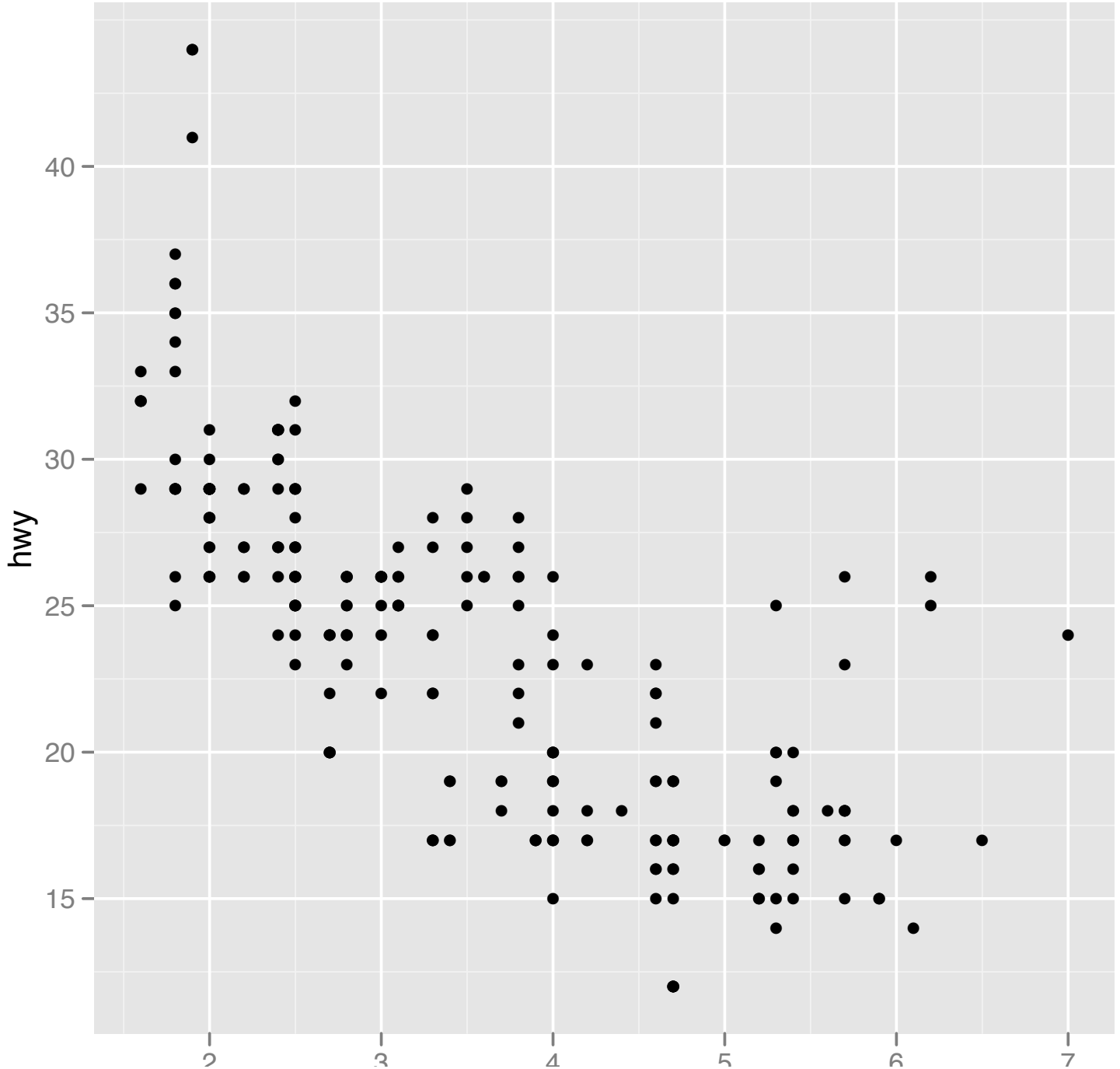
October 2009

http://lookinatdata.com/
ismi-2009

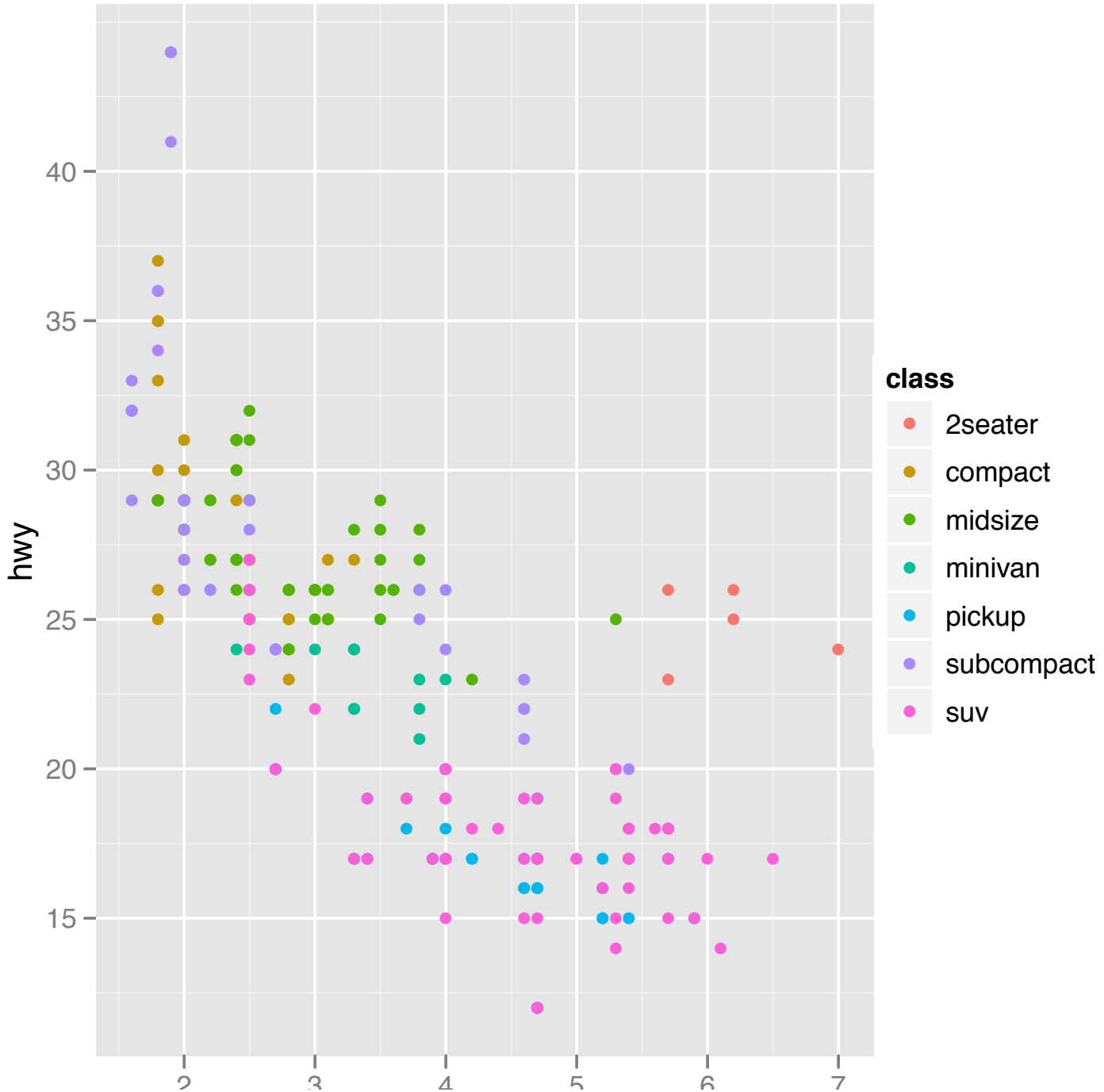
1. Preview of today
2. About ggplot2
3. More resources
4. Diving in

Fuel economy

Basic graphics



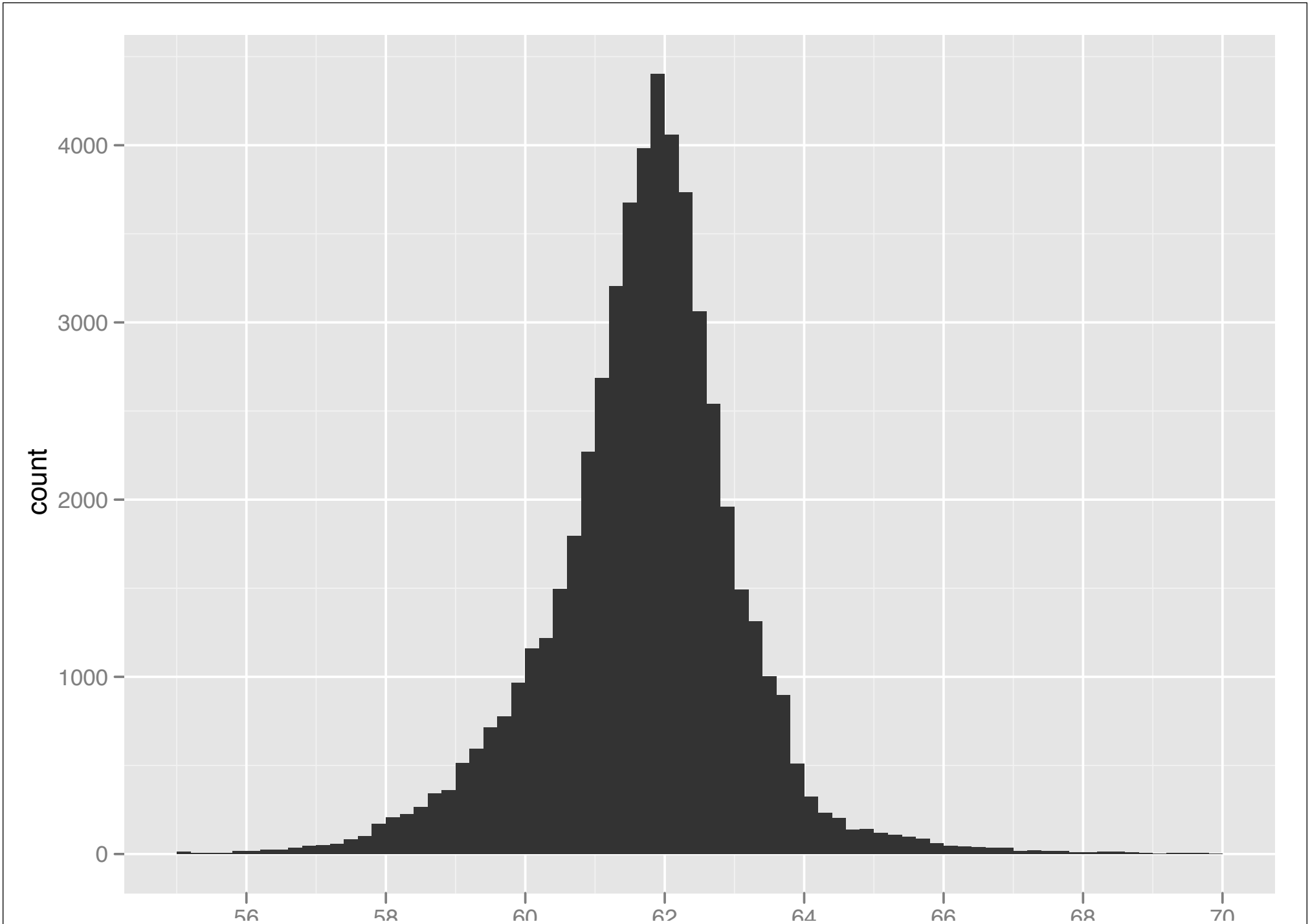

```
qplot(displ, hwy, data = mpg)
```



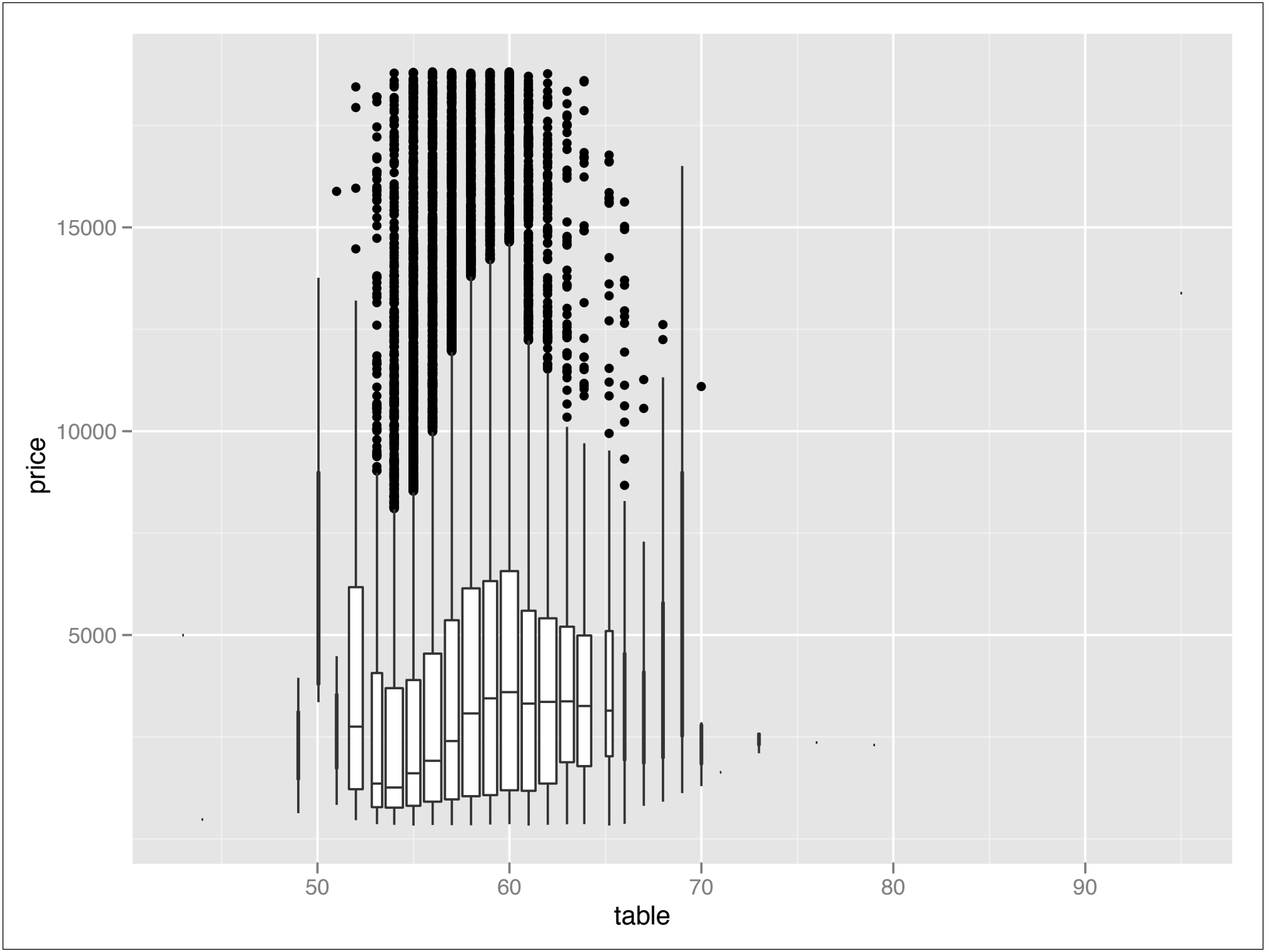

```
qplot(displ, hwy, colour = class, data = mpg)
```

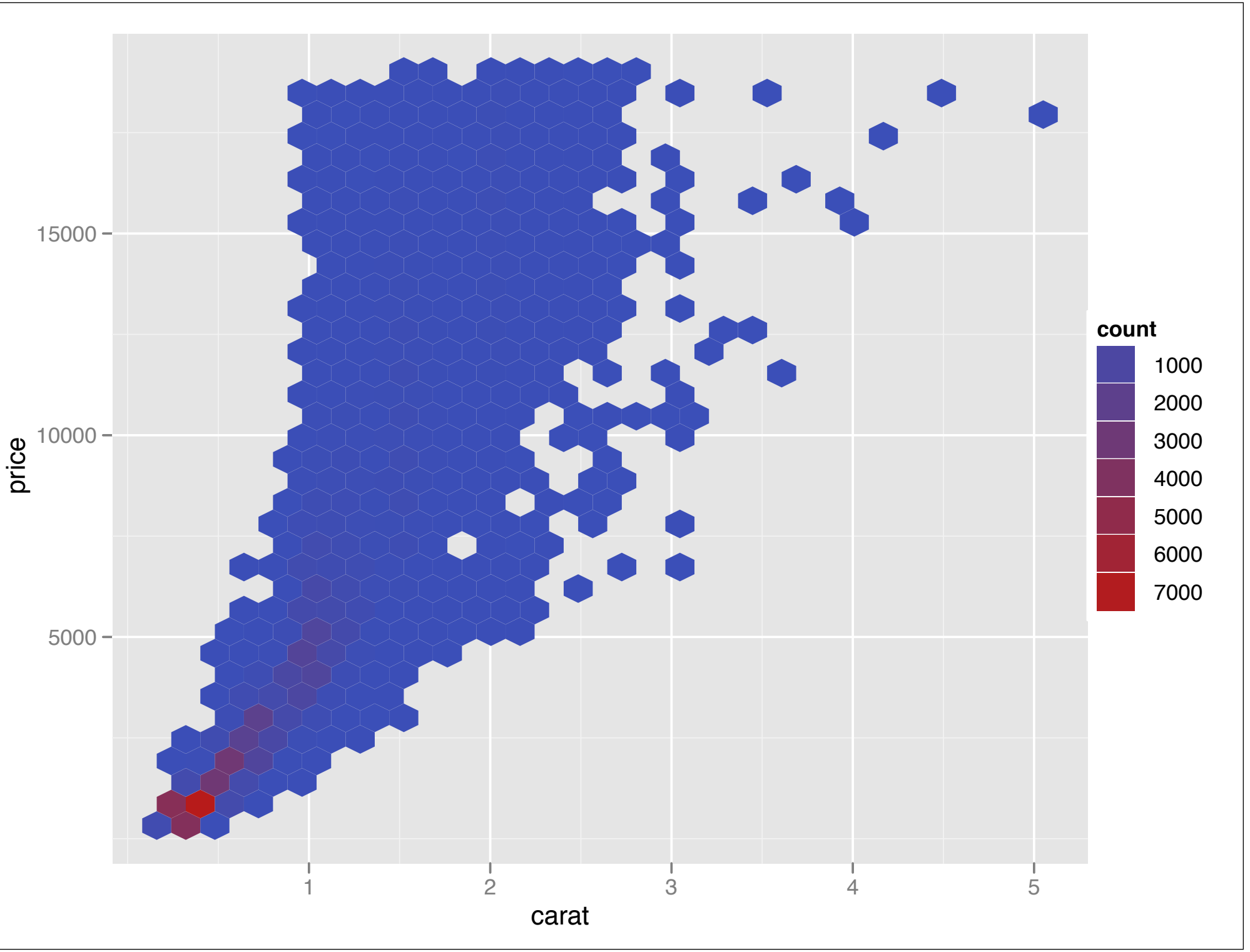
Diamond prices

Displaying large data




```
qplot(depth, data = diamonds, binwidth = 0.2)
```

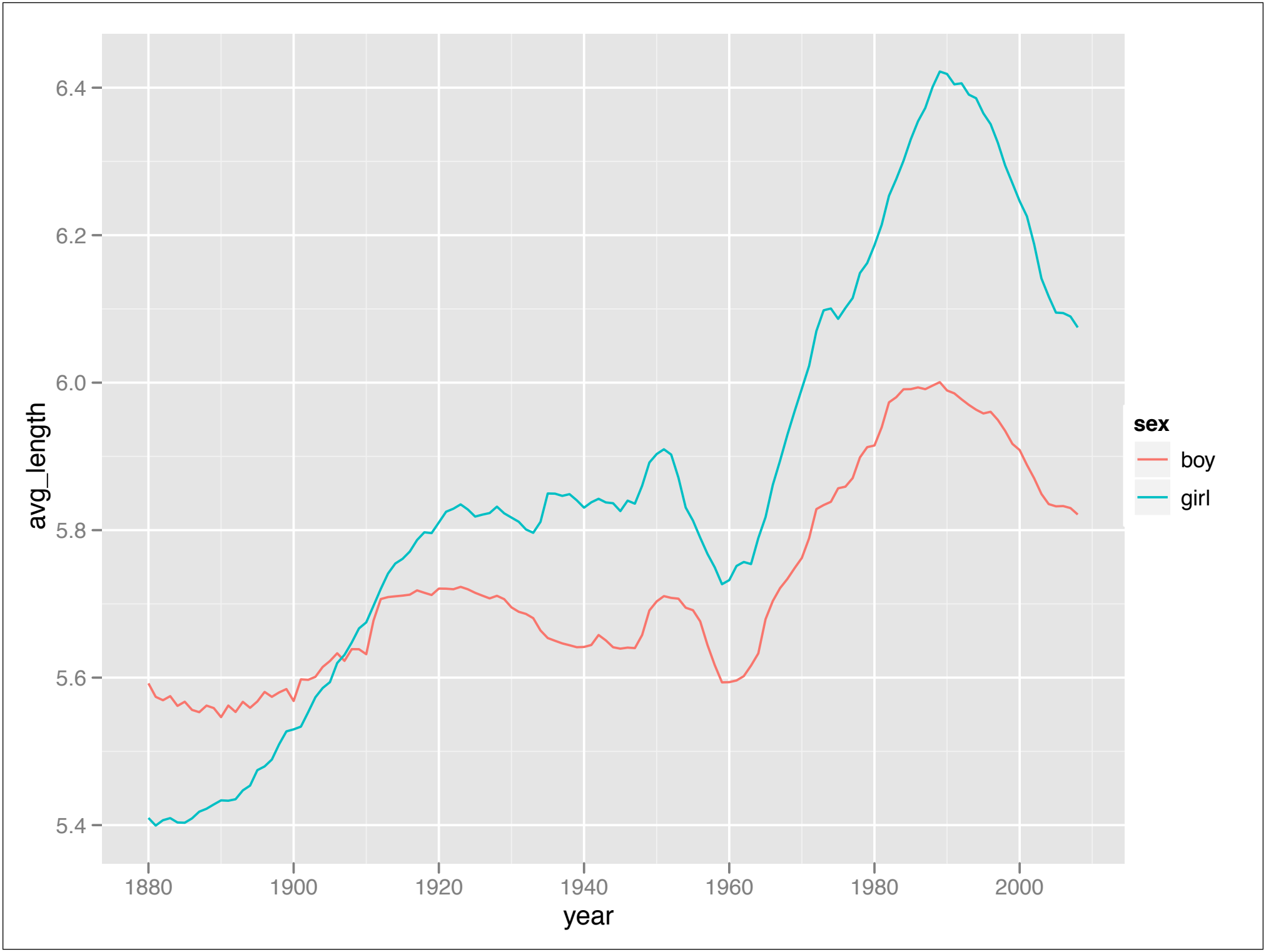

15000
10000
5000
price
50
60
70
80
90
table

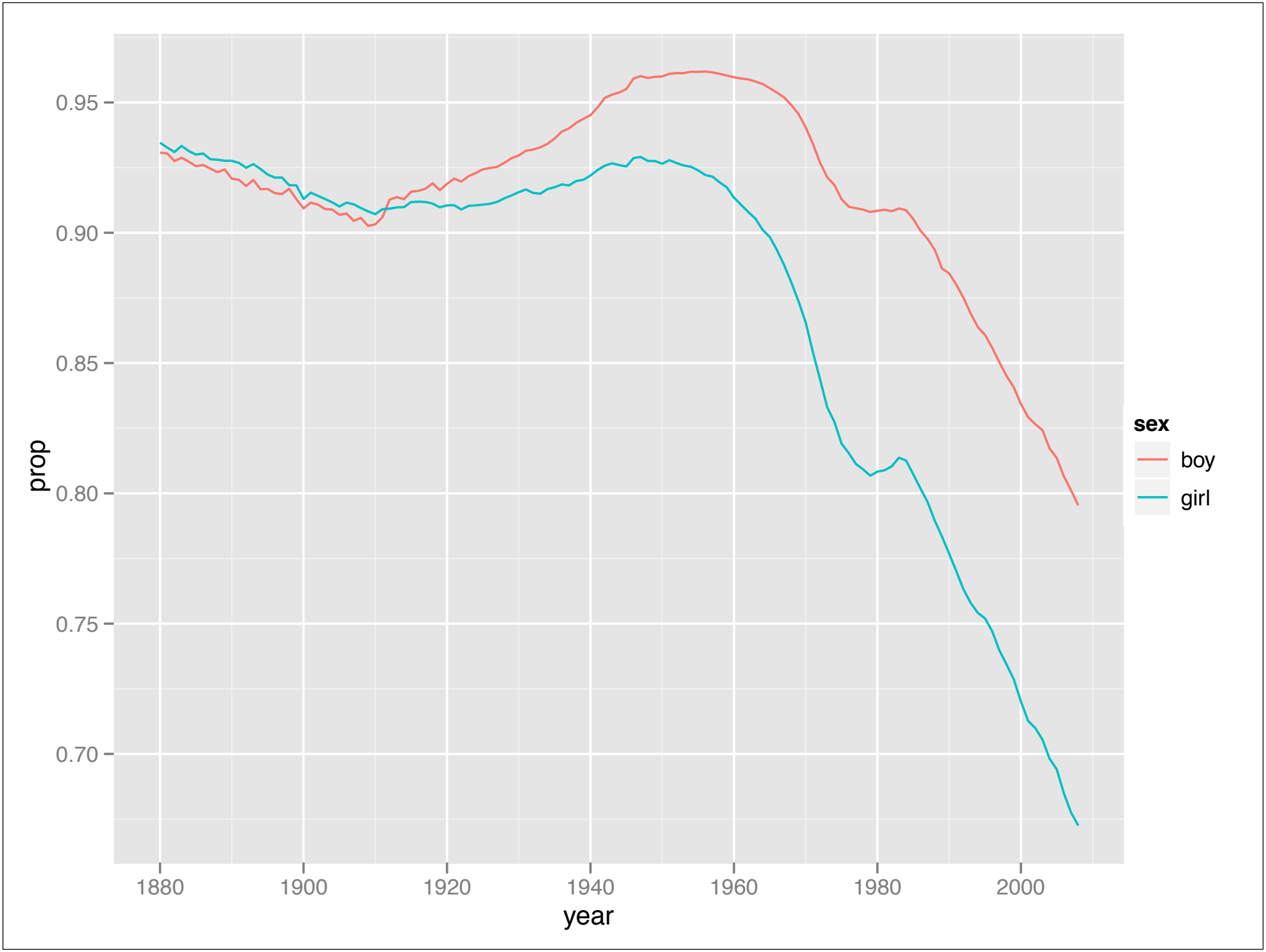

price
15000
10000
5000
1
2
3
4
5
carat
count
1000
2000
3000
4000
5000
6000
7000

US baby names

US baby names

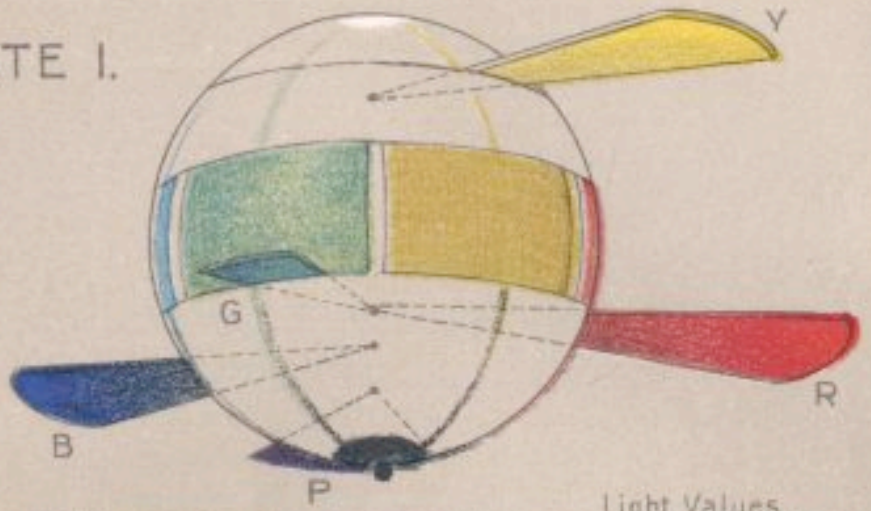
Transforming data


6.4
6.2
6.0
5.8
5.6
5.4
avg_length
1880
1900
1920
1940
1960
1980
2000
year
sex
boy
girl


0.95
0.90
0.85
0.80
0.75
0.70
prop
1880
1900
1920
1940
1960
1980
2000
year
sex
boy
girl

Polishing your plots

PLATE I.



Light Values

Middle Values

Dark Values

P B G Y R

Copyright 1907 by A. H. Munsell

1. Saving your work
2. Labels & ticks
3. Themes

ggplot2

About ggplot2

Graphical grammar (domain specific language), based on “The Grammar of Graphics” by Leland Wilkinson.

Specify what you want, not how to create it. Many fiddly details taken care of.

“Instead of spending time making your graph look pretty, you can focus on creating a graph that bests reveals the messages in your data.”

Useful resources

http://had.co.nz/ggplot2

http://had.co.nz/ggplot2/book

http://groups.google.com/group/ggplot2

http://learnr.wordpress.com

http://ggplot2.wik.is

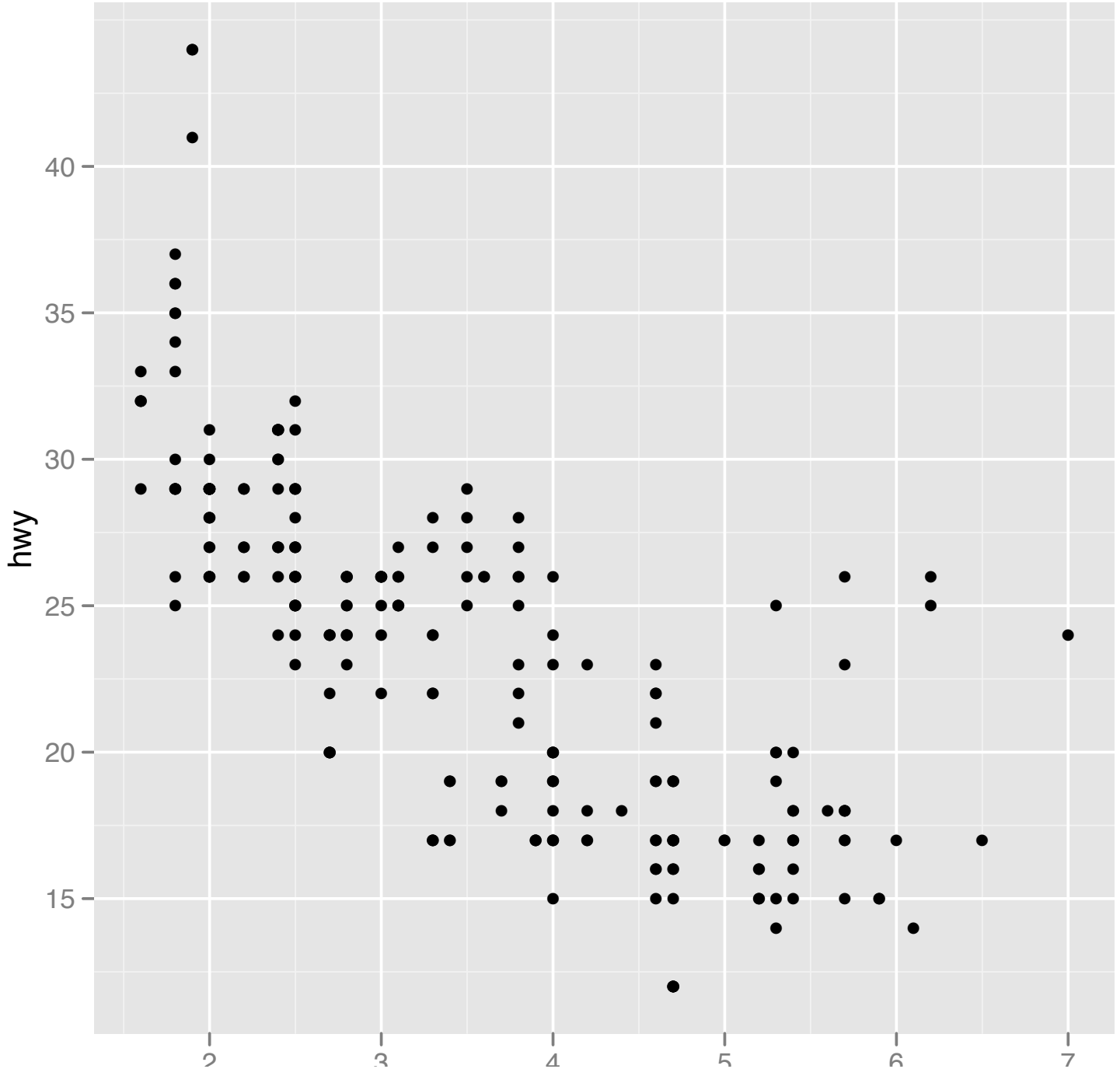
Scatterplot basics

```
install.packages("ggplot2")
library(ggplot2)

?mpg
head(mpg)
str(mpg)
summary(mpg)

qplot(displ, hwy, data = mpg)
```

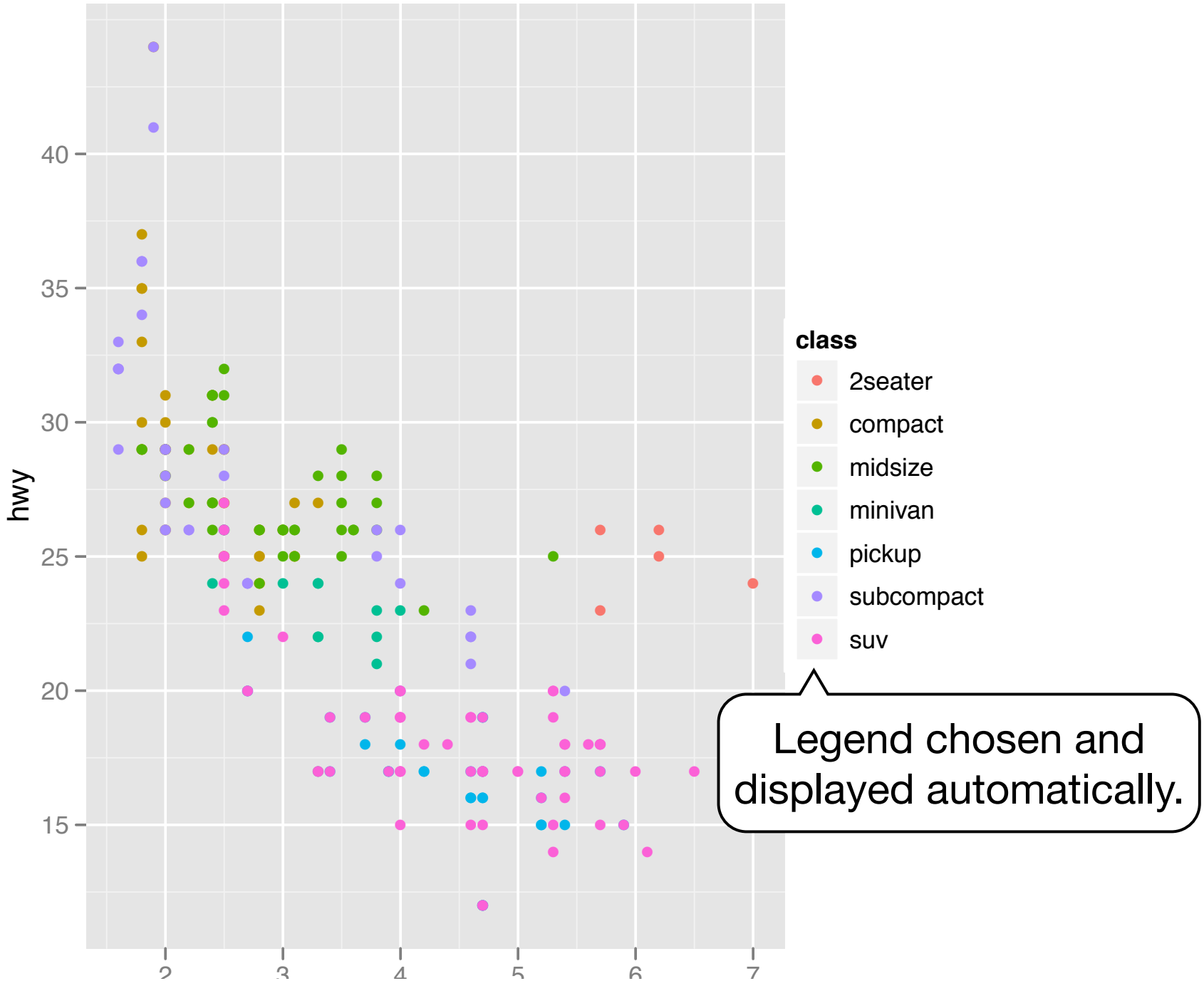
In ggplot2, we always explicitly specify the data




```
qplot(displ, hwy, data = mpg)
```

Additional variables

Can display additional variables with **aesthetics** (like shape, colour, size) or **facetting** (small multiples displaying different subsets)




```
qplot(displ, hwy, colour = class, data = mpg)
```

Your turn

Experiment with colour, size, and shape aesthetics.

What's the difference between discrete or continuous variables?

What happens when you combine multiple aesthetics?

	Discrete	Continuous
Colour	Rainbow of colours	Gradient from red to blue
Size	Discrete size steps	Linear mapping between radius and value
Shape	Different shape for each	Doesn’t work

Faceting

Small multiples displaying different subsets of the data.

Useful for exploring conditional relationships. Useful for large data.

Your turn

```
qplot(displ, hwy, data = mpg) +
facet_grid(. ~ cyl)
```

```
qplot(displ, hwy, data = mpg) +
facet_grid(drv ~ .)
```

```
qplot(displ, hwy, data = mpg) +
facet_grid(drv ~ cyl)
```

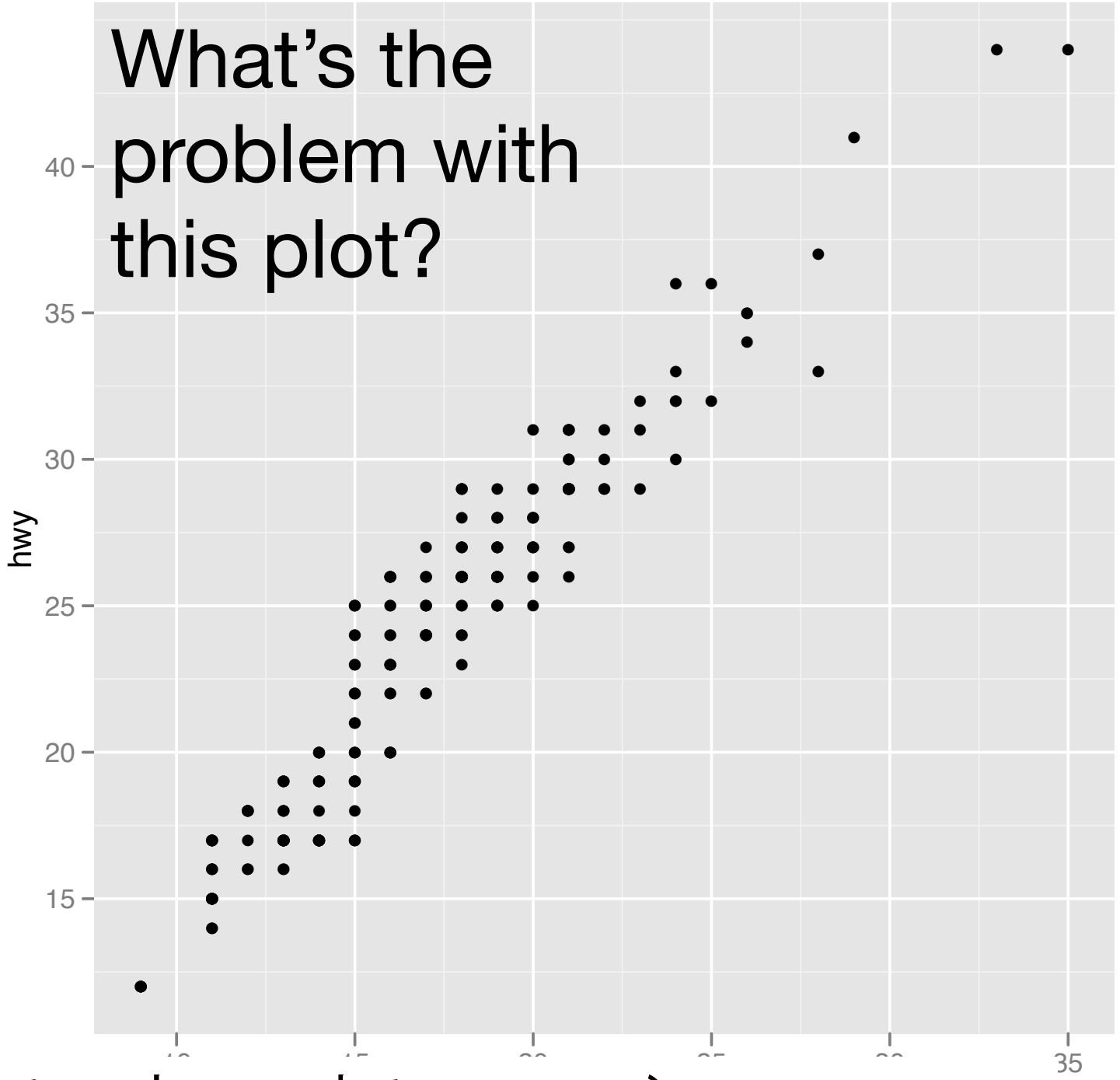
```
qplot(displ, hwy, data = mpg) +
facet_wrap(~ class)
```

Summary

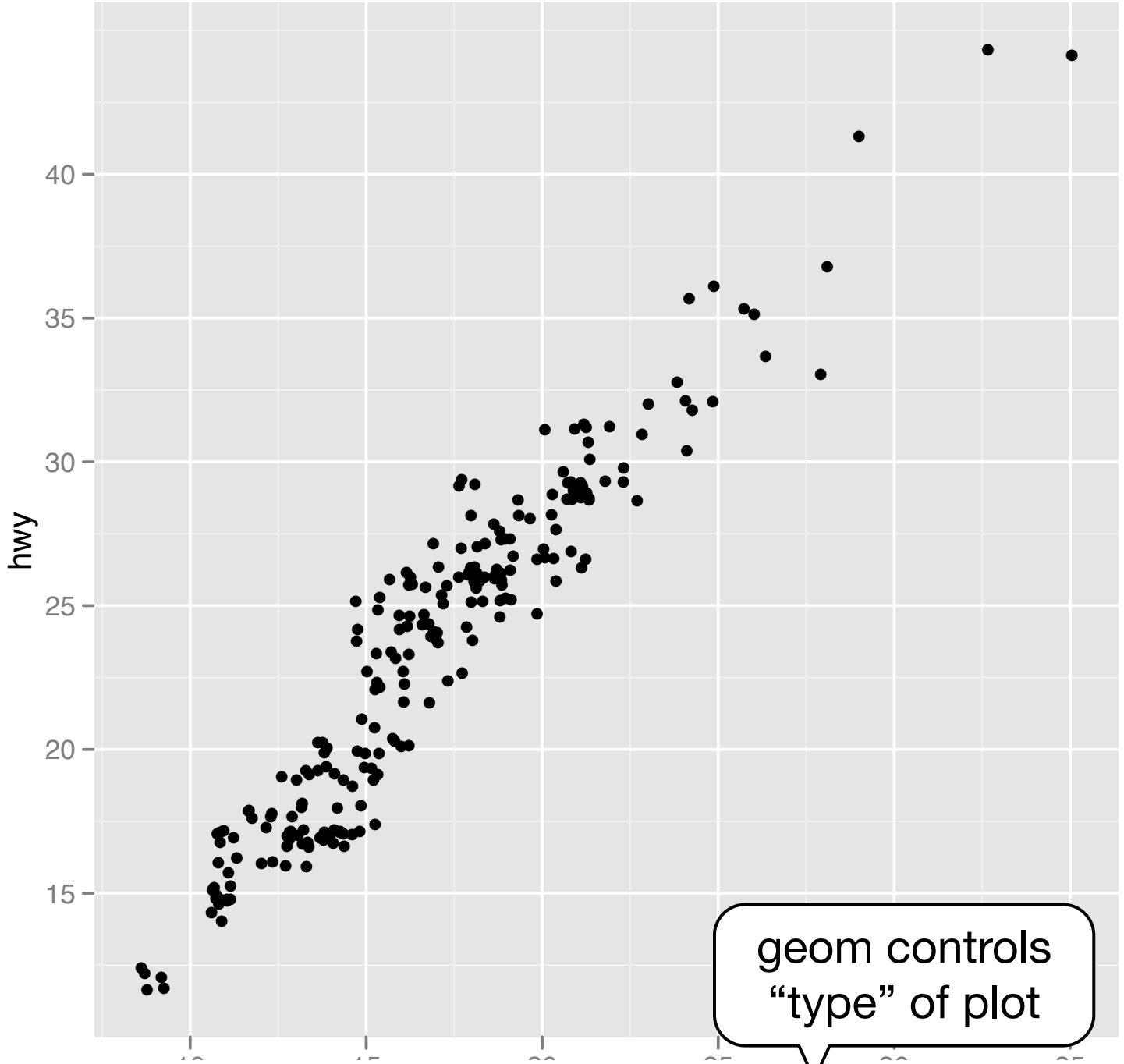
`facet_grid()`: 2d grid, rows ~ cols, . for no split

`facet_wrap()`: 1d ribbon wrapped into 2d

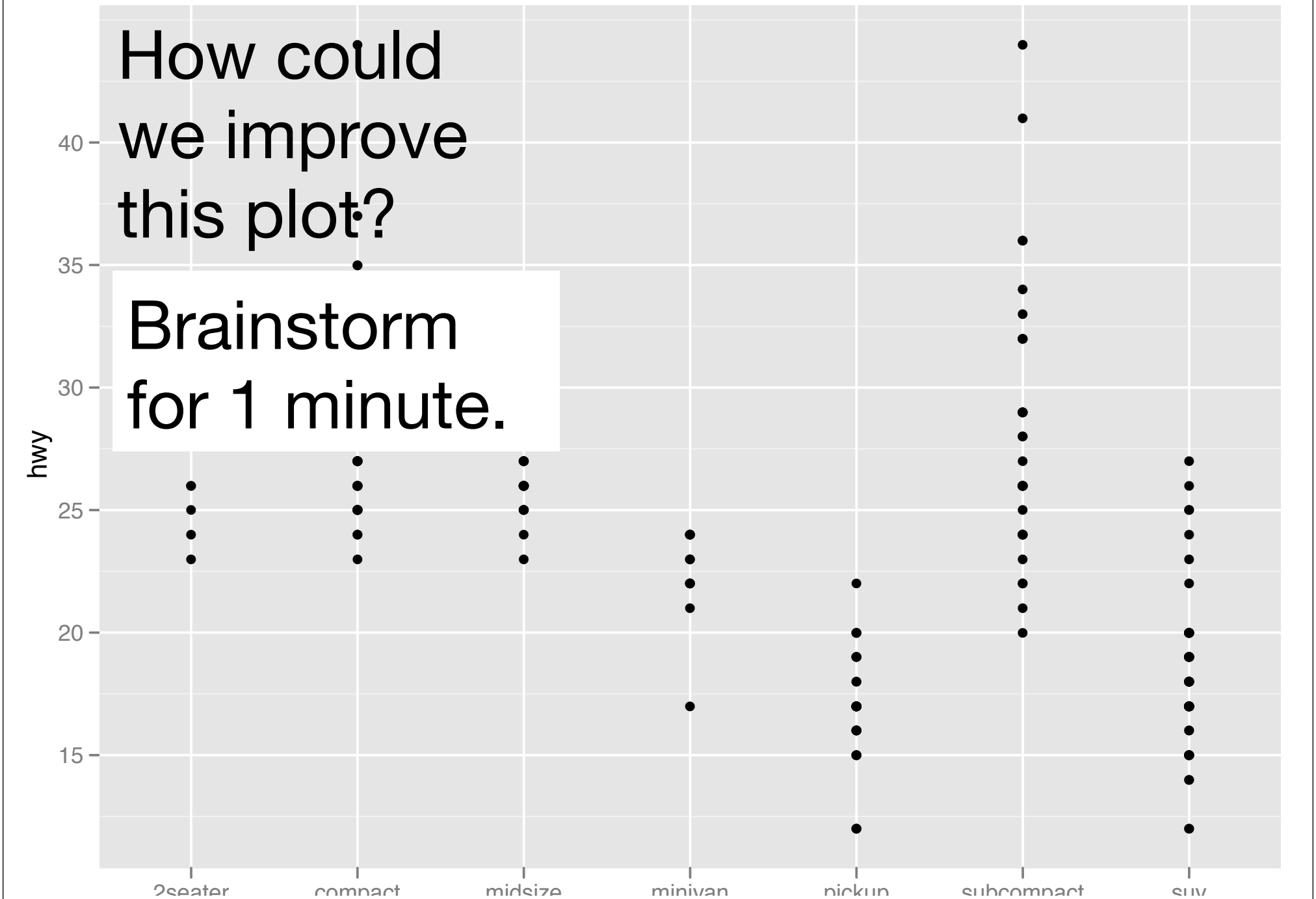
Scales argument controls whether position scales are fixed or free.



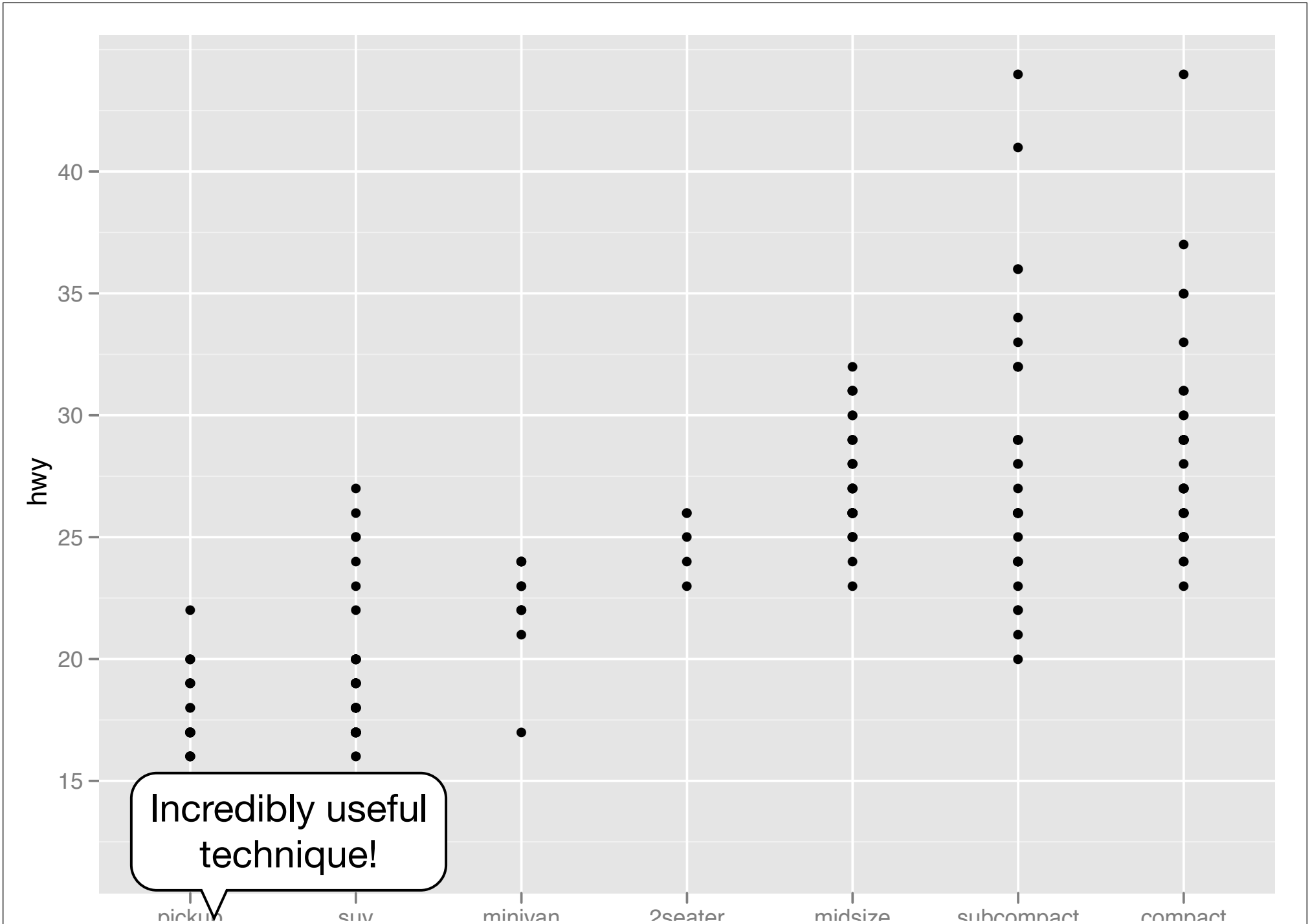

```
qplot(cty, hwy, data = mpg)
```



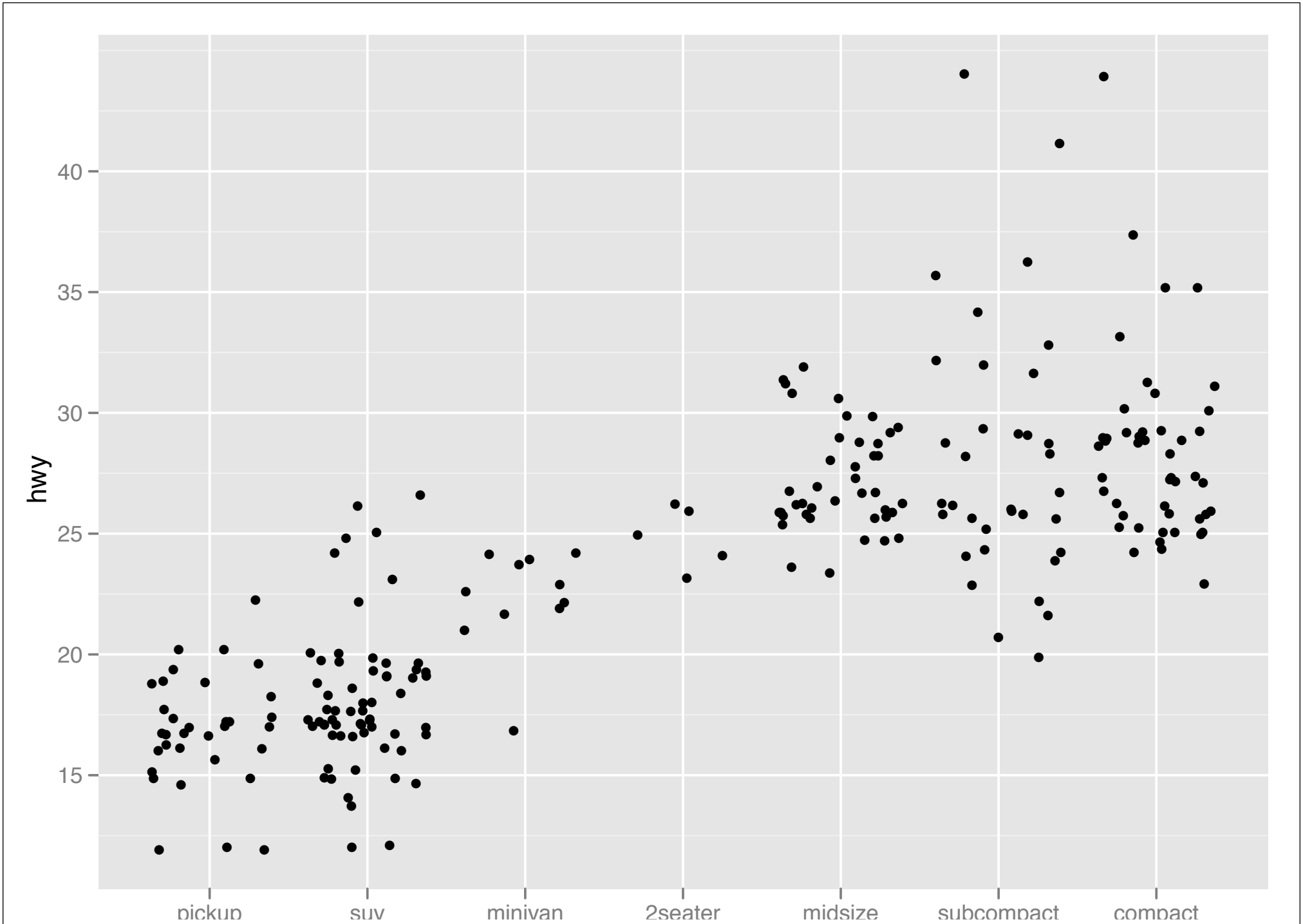

```
qplot(cty, hwy, data = mpg, geom = "jitter")
```



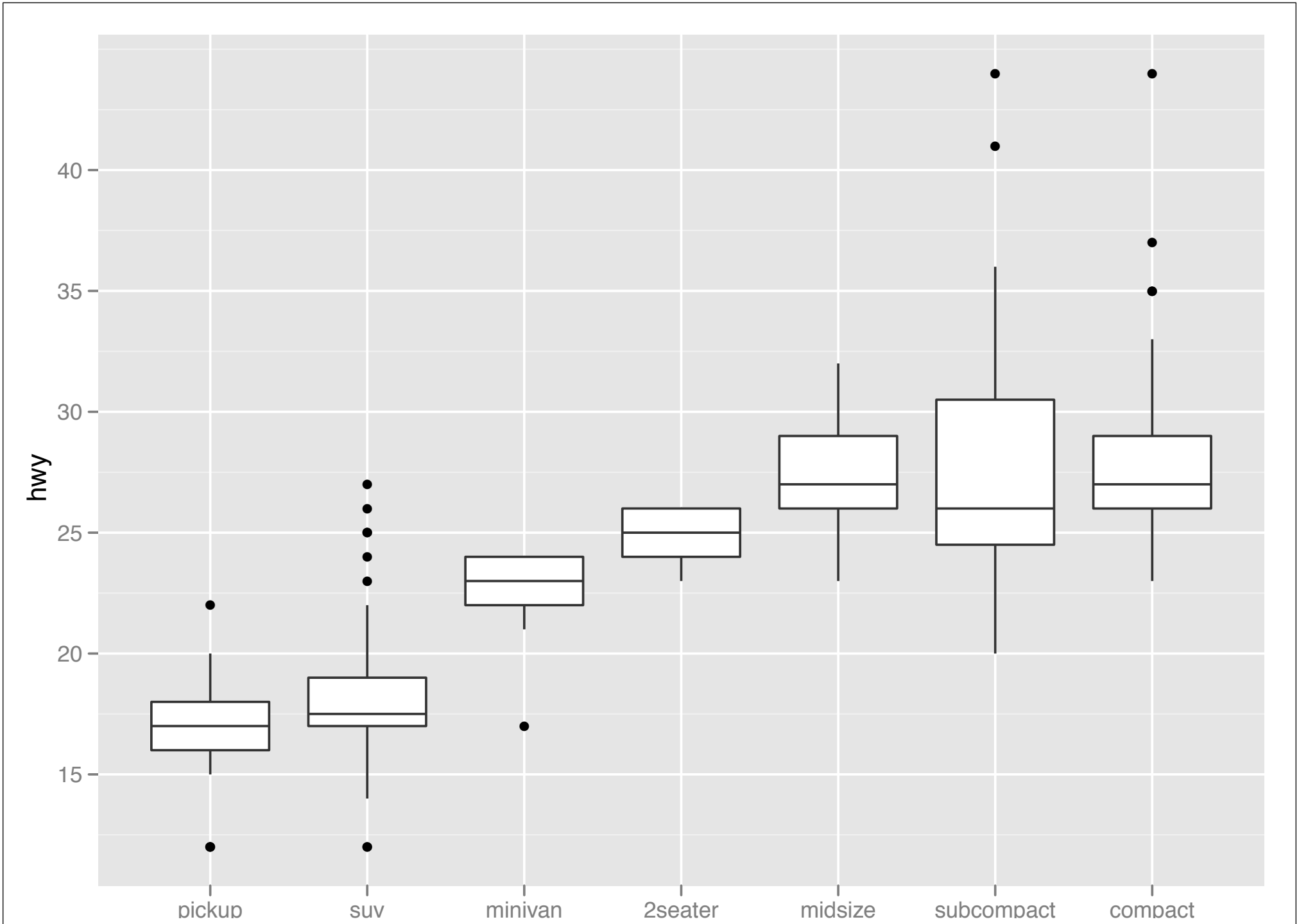

```
qplot(class, hwy, data = mpg)
```



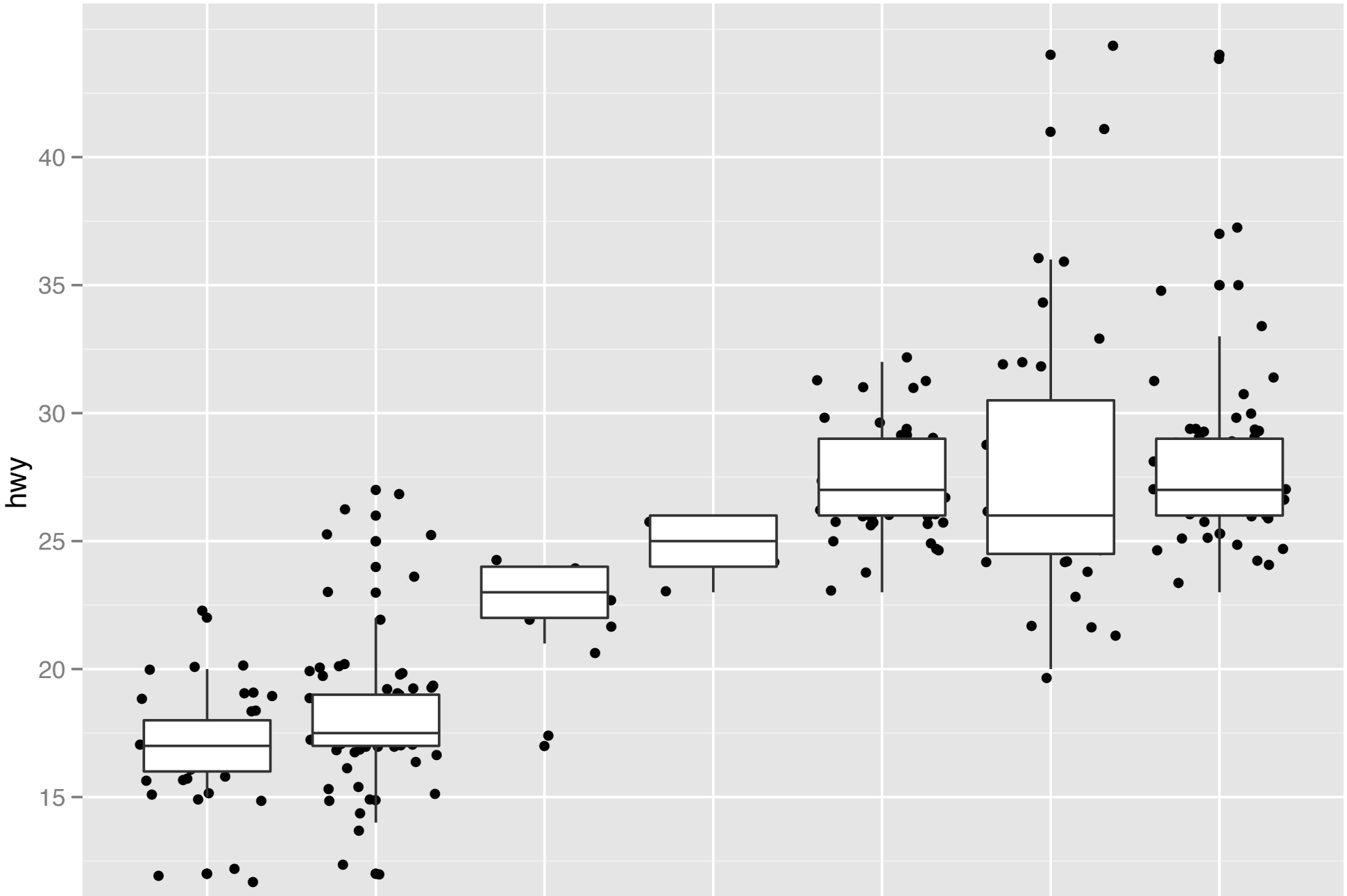

```
qplot(reorder(class, hwy), hwy, data = mpg)
```




```
qplot(reorder(class, hwy), hwy, data = mpg, geom = "jitter")
```




```
qplot(reorder(class, hwy), hwy, data = mpg, geom = "boxplot")
```




```
qplot(reorder(class, hwy), hwy, data = mpg,
  geom = c("jitter", "boxplot"))
```

Your turn

Read the help for reorder. Redraw the previously plots with class ordered by median hwy.

How would you put the jittered points on top of the boxplots?

Aside: coding strategy

At the end of each interactive session, you want a summary of everything you did. Two options:

1. Save everything you did with `savehistory()` then remove the unimportant bits.

2. Build up the important bits as you go. (this is how I work)

This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 United States License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/us/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.