http://courses.had.co.nz

Visualisation

Hadley Wickham

Assistant Professor / Dobelman Family Junior Chair

Department of Statistics / Rice University

July 2012

1. Scatterplots & aesthetics
2. Facetting
3. Geoms

Diving in

Scatterplot basics

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

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

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

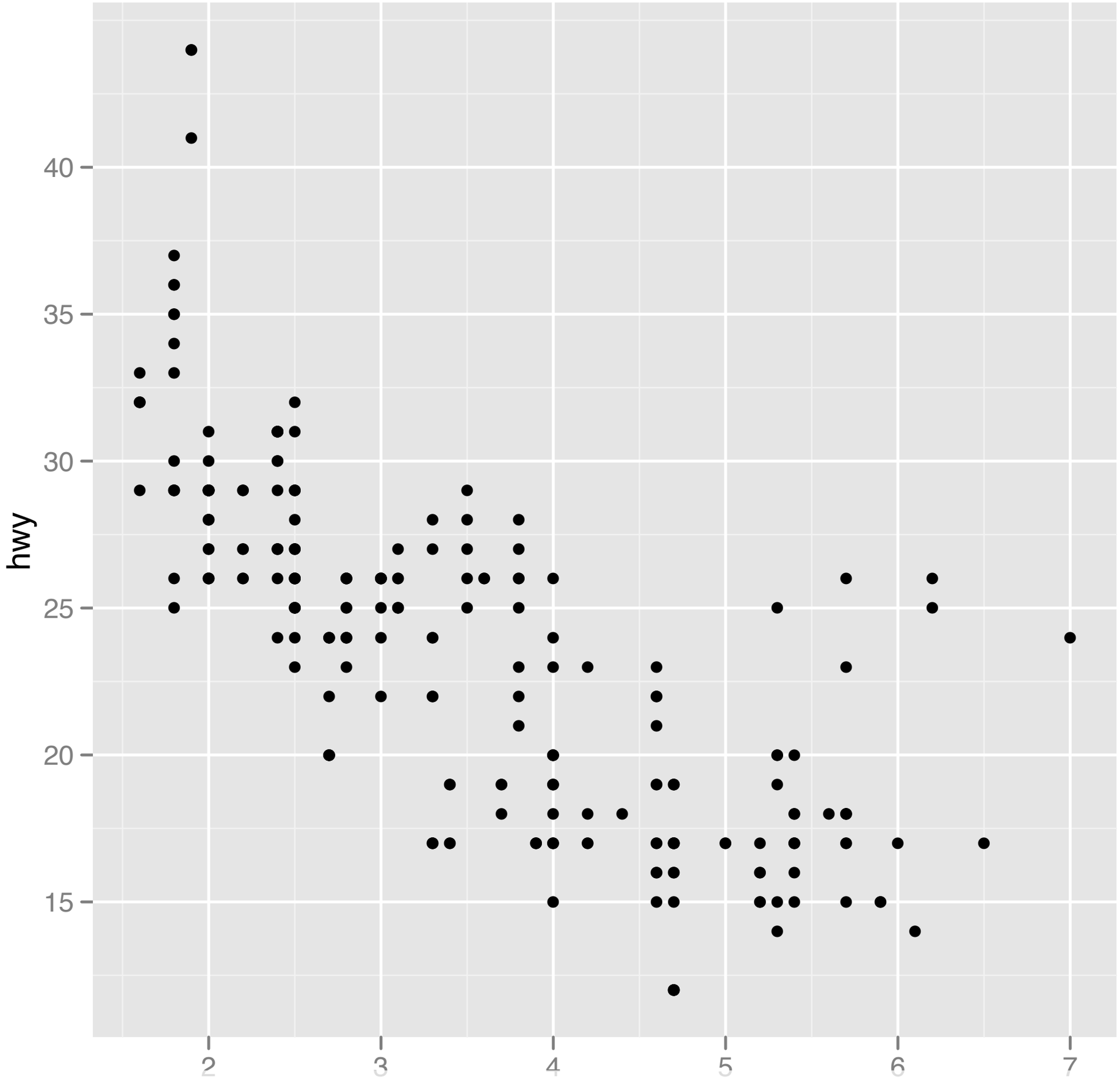
Scatterplot basics

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

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

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

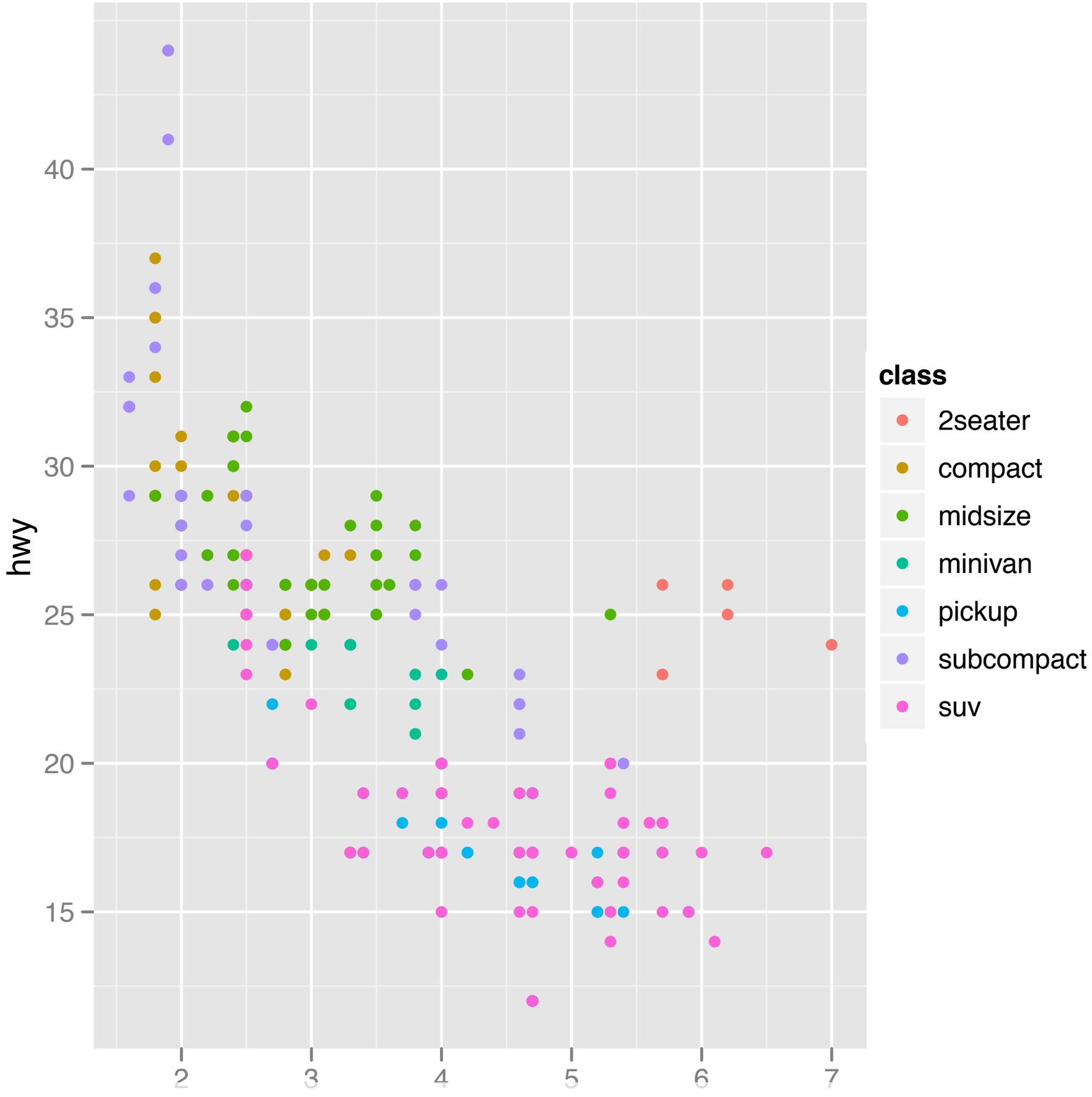
Always explicitly specify the data



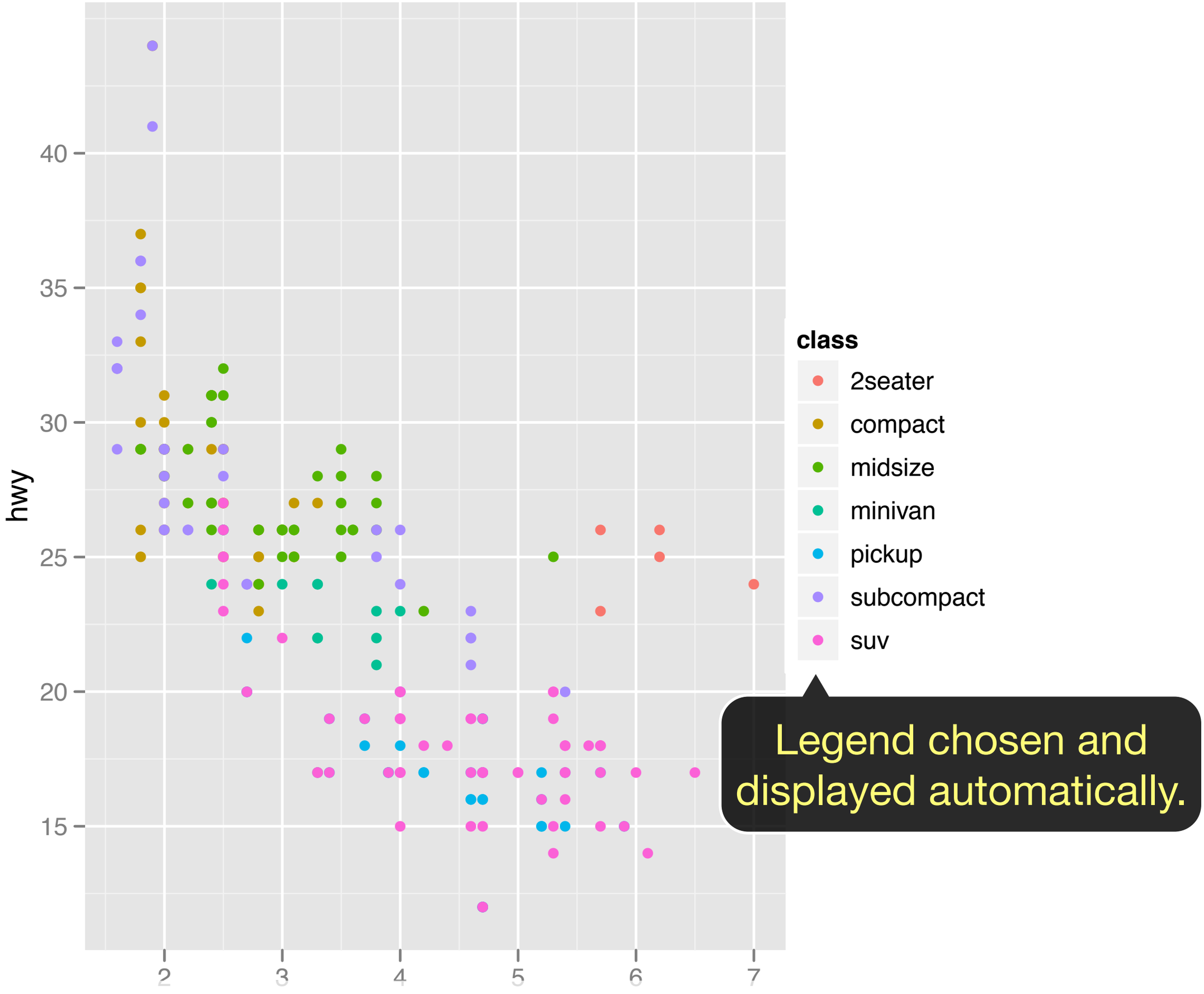

```
qplot(displ, hwy, data = mpg)
```

Additional variables

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




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




```
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	Colour gradient
Size	Discrete size steps	Linear mapping between radius and value
Shape	Different shape for each	Doesn’t work

Facetting

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

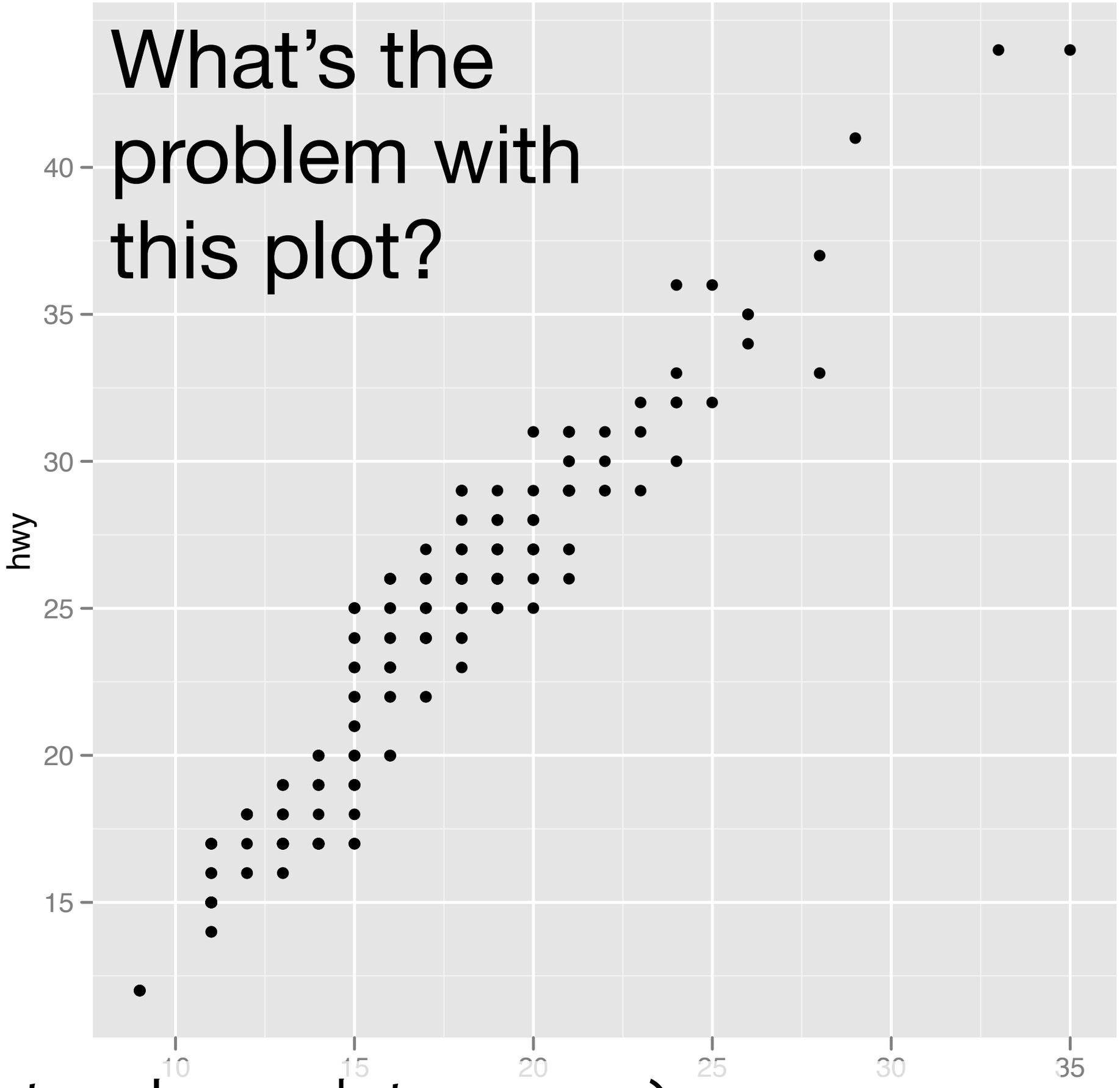
Aside: workflow

Keep a copy of the slides open so that you can copy and paste the code.

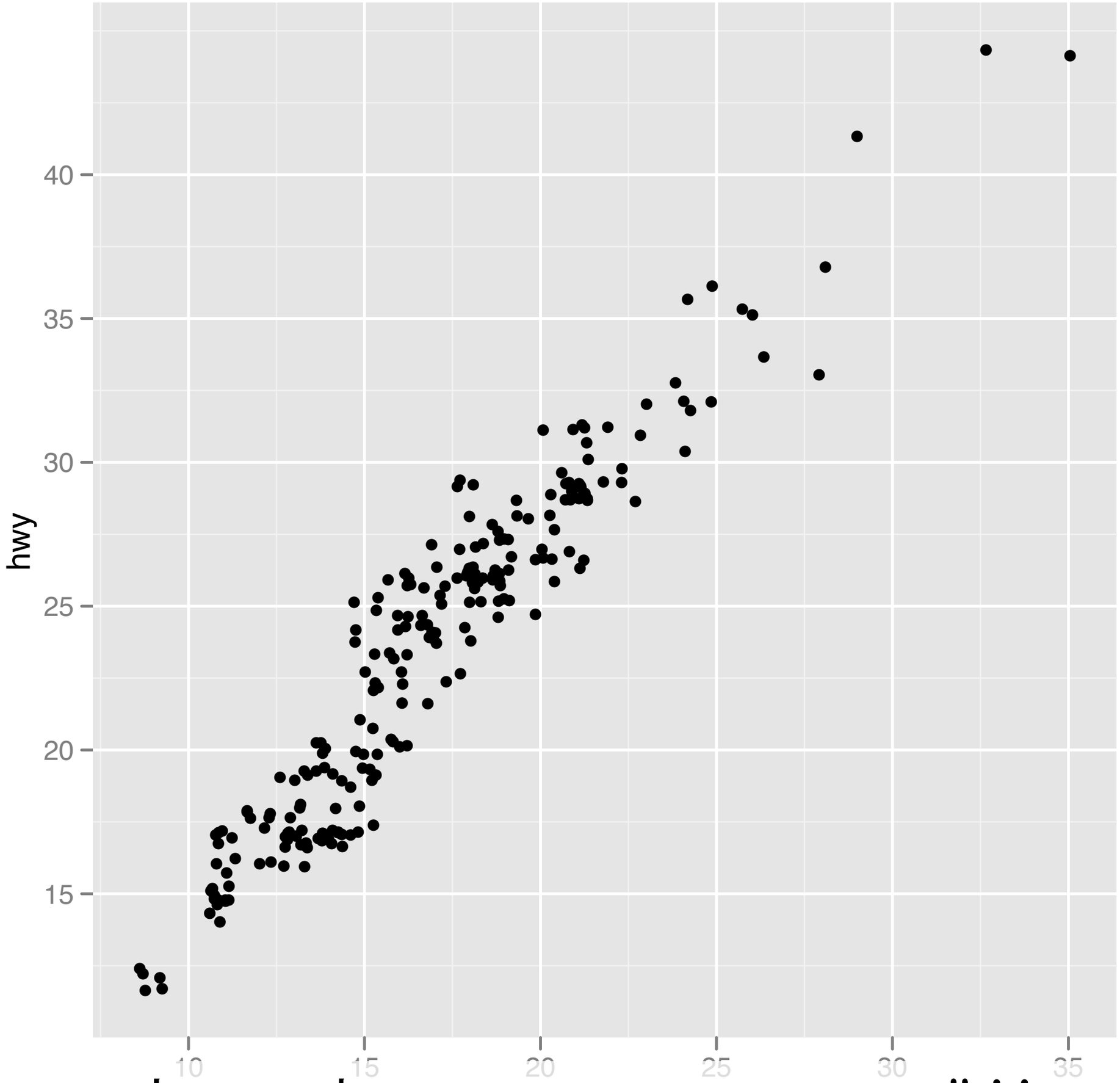
For complicated commands, write them in the editing area and then run.

Geoms

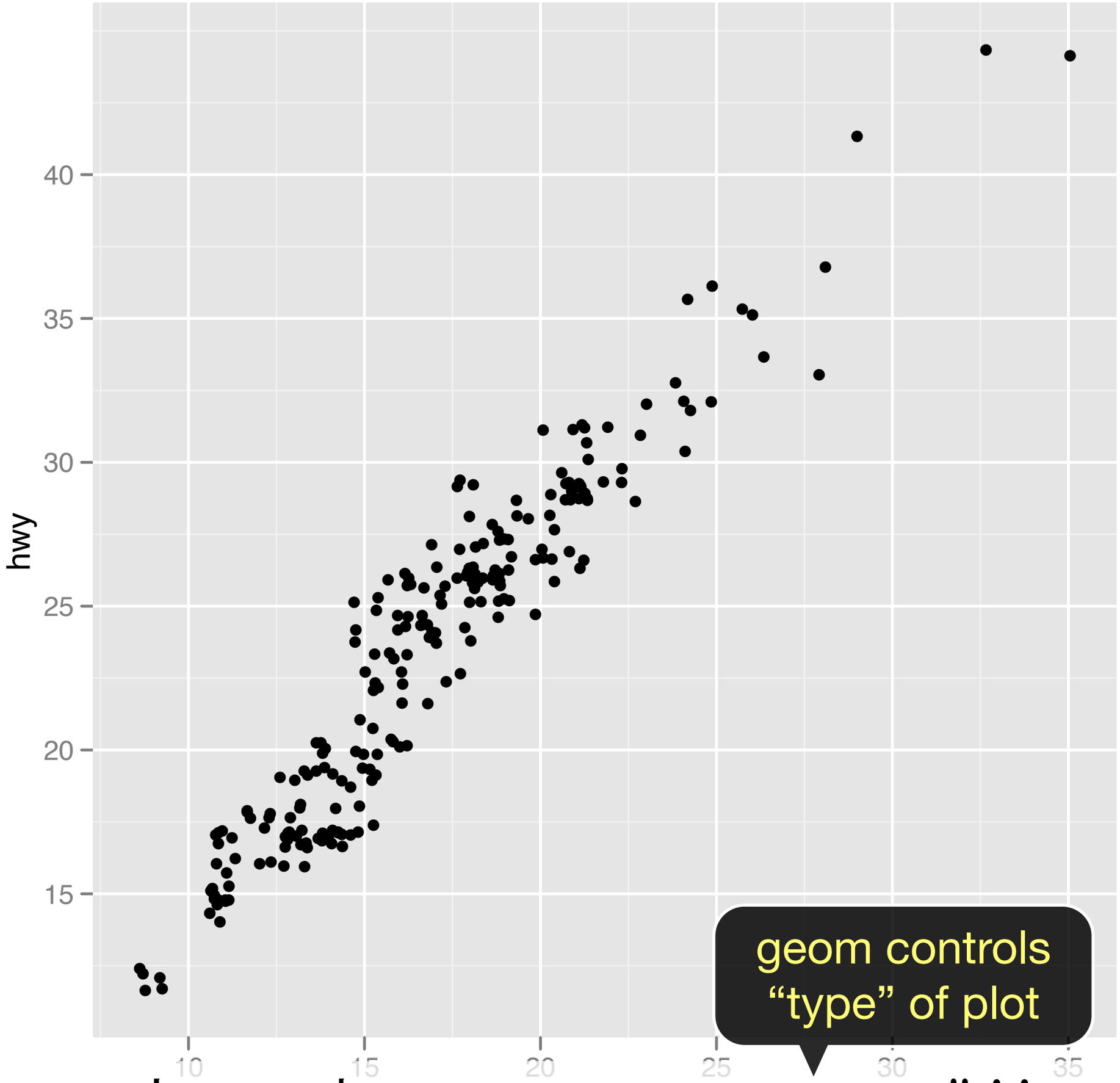
What's the problem with this plot?



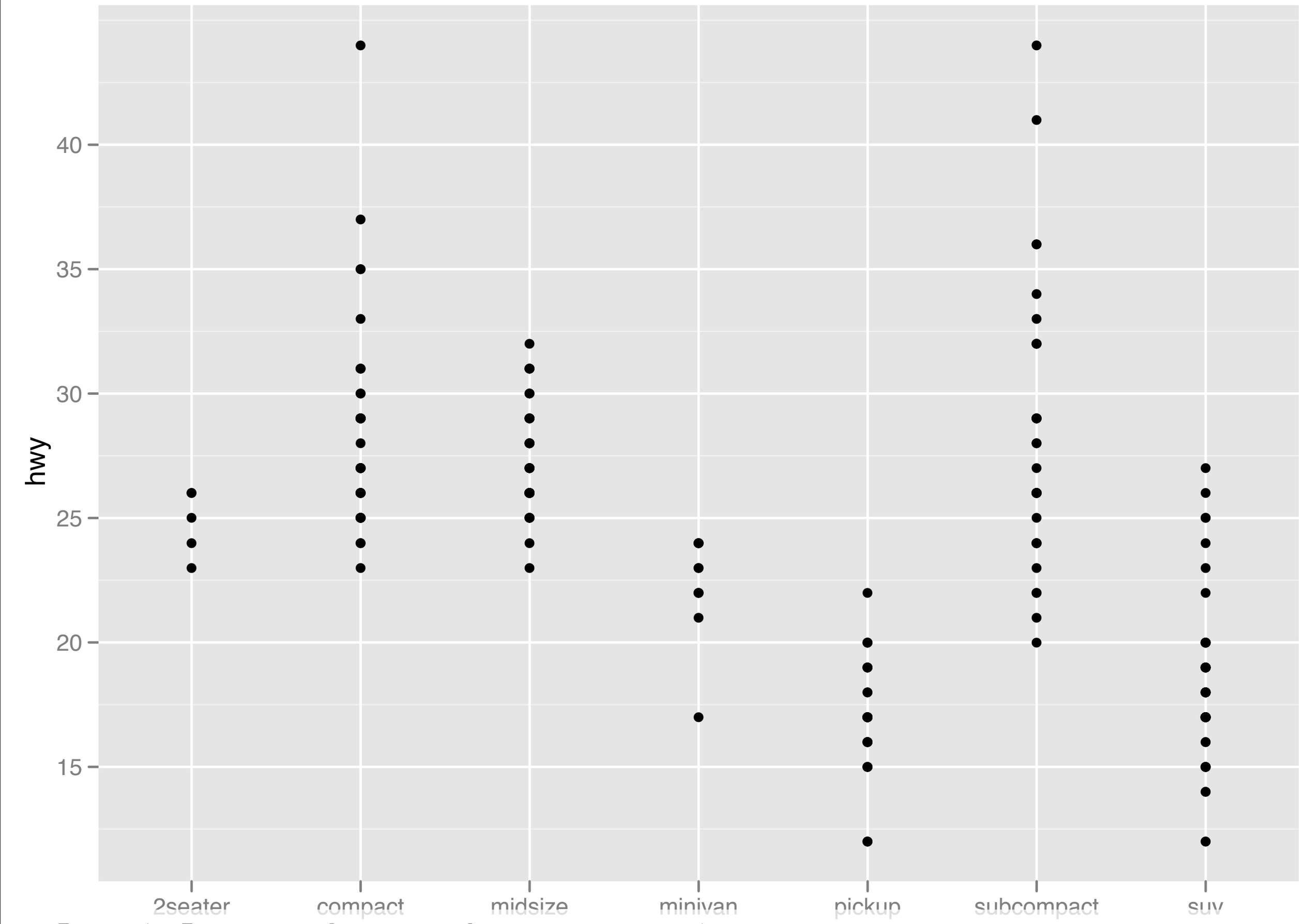

```
qplot(cty, hwy, data = mpg)
```




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



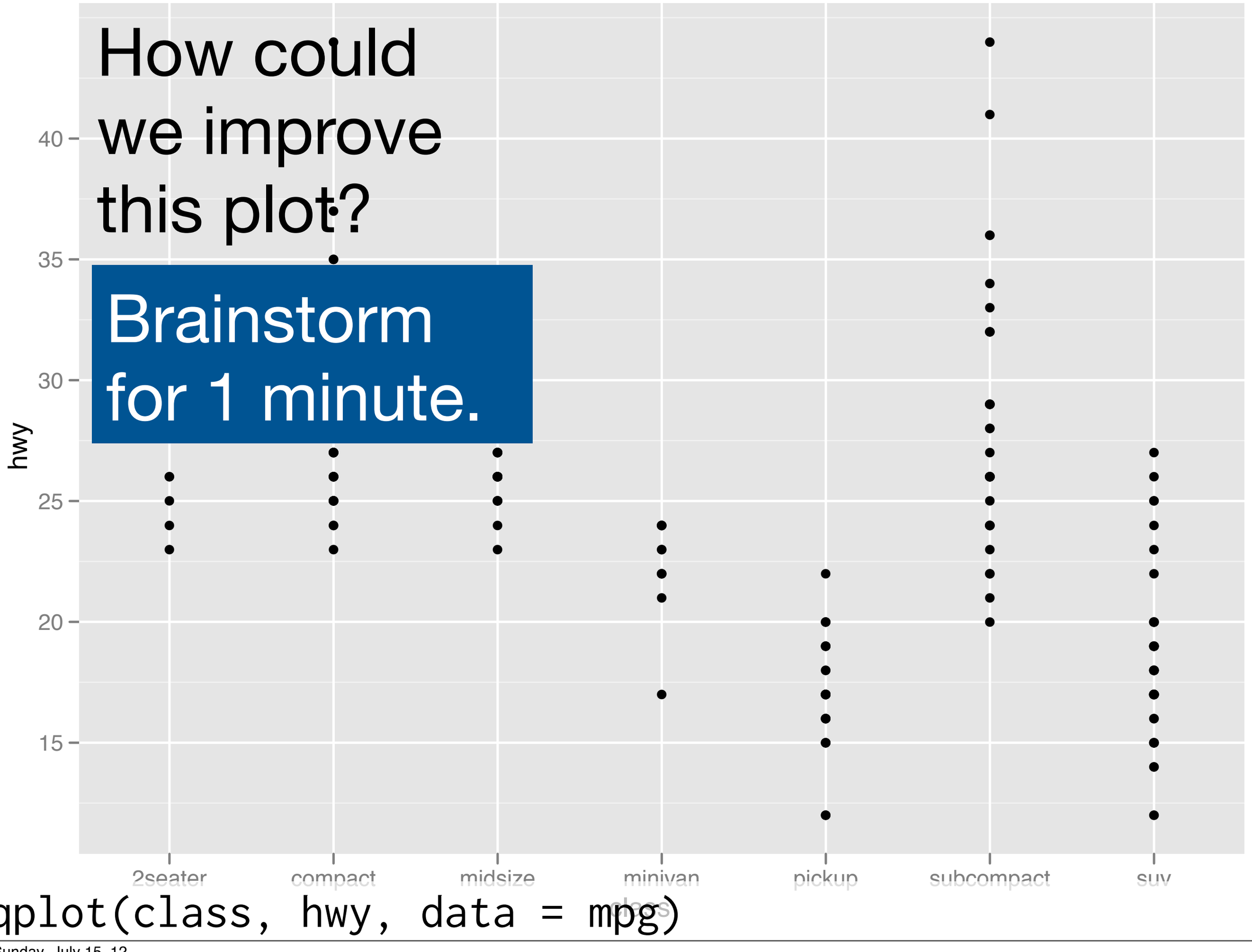

```
qplot(cty, hwy, data = mpg, geom = "jitter")
```



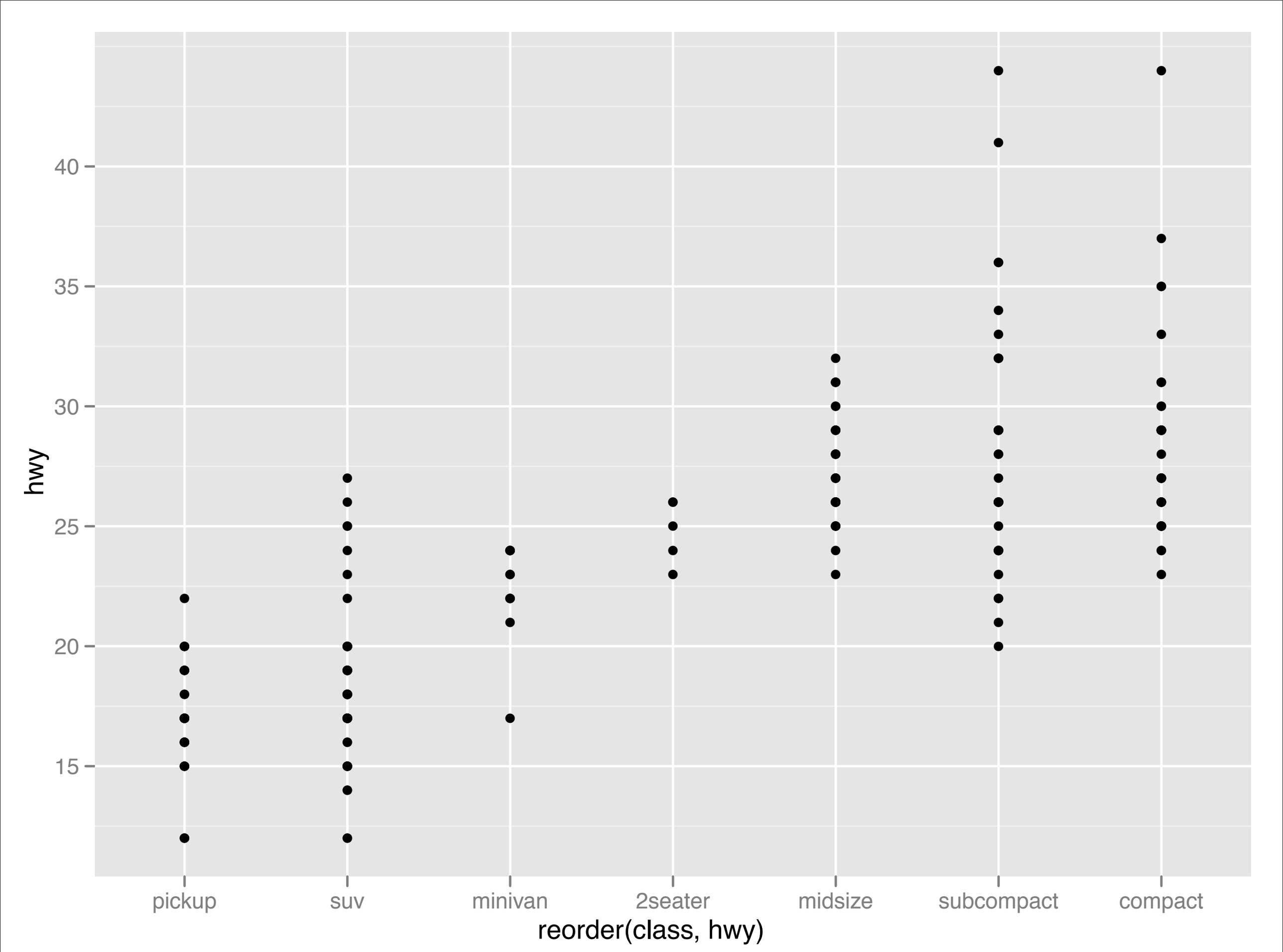

```
qplot(class, hwy, data = mpg)
```

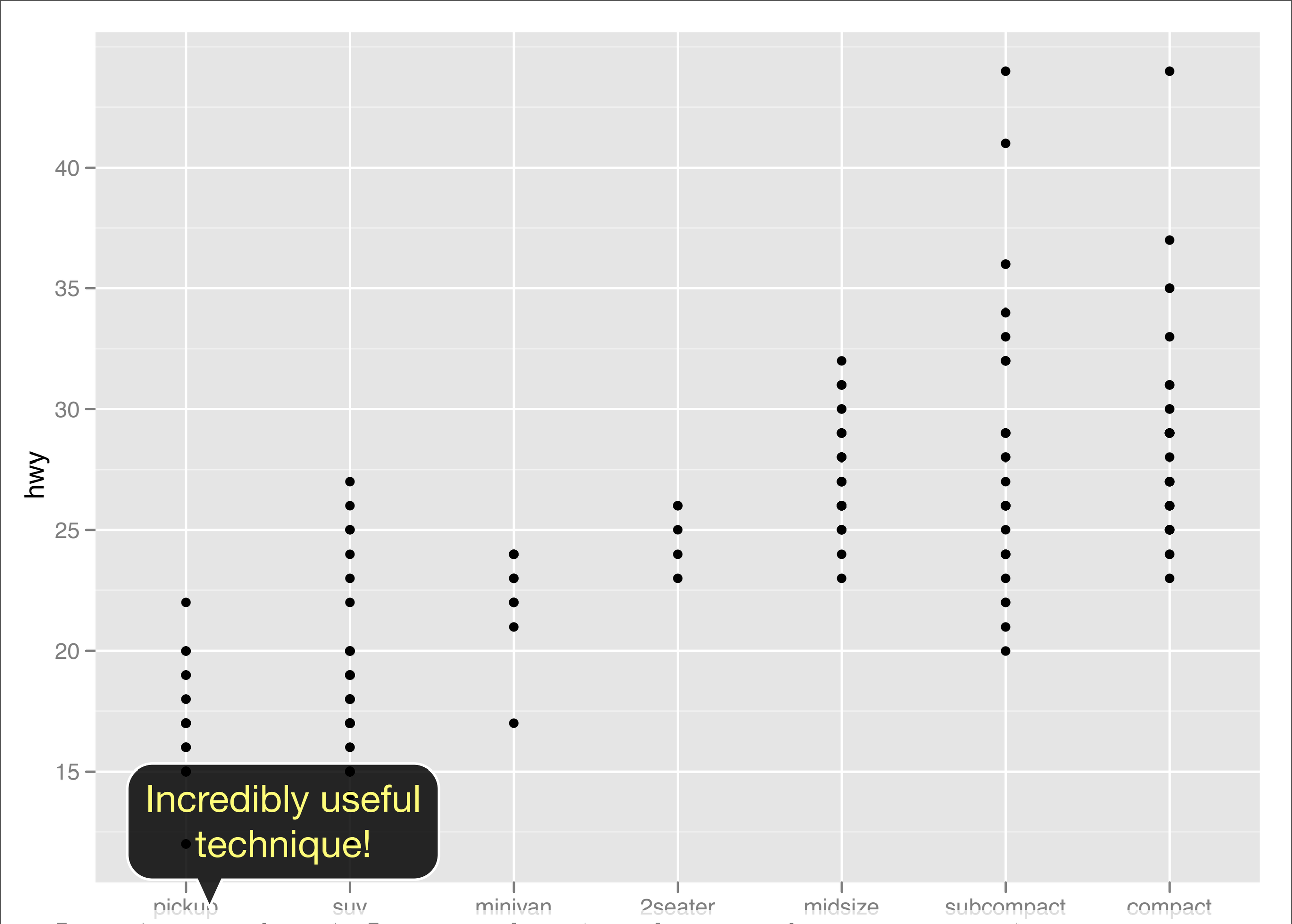
How could we improve this plot?

Brainstorm for 1 minute.

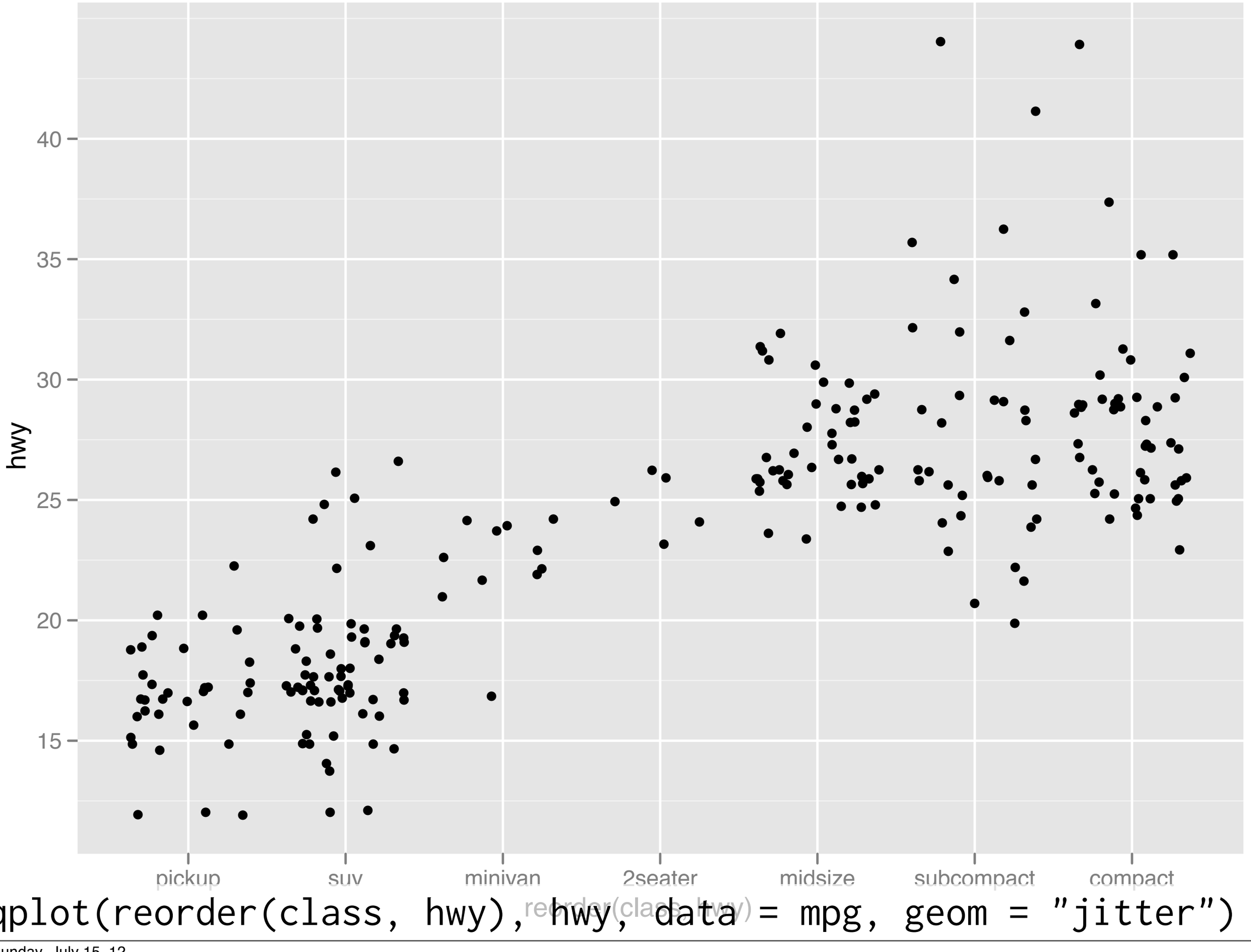



```
qplot(class, hwy, data = mpg)
```

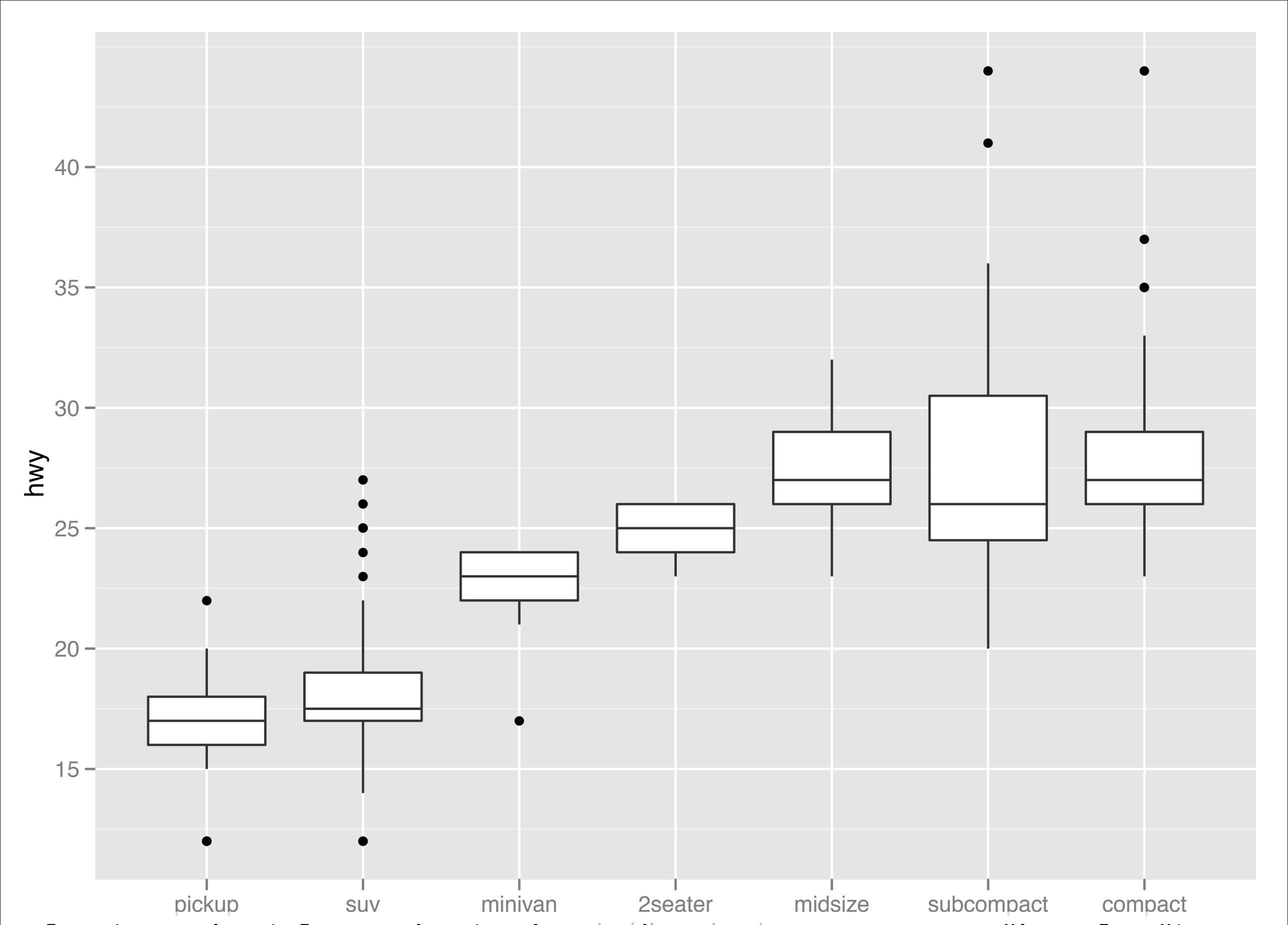

hwy
40
35
30
25
20
15
pickup
suv
minivan
2seater
midsize
subcompact
compact
reorder(class, hwy)



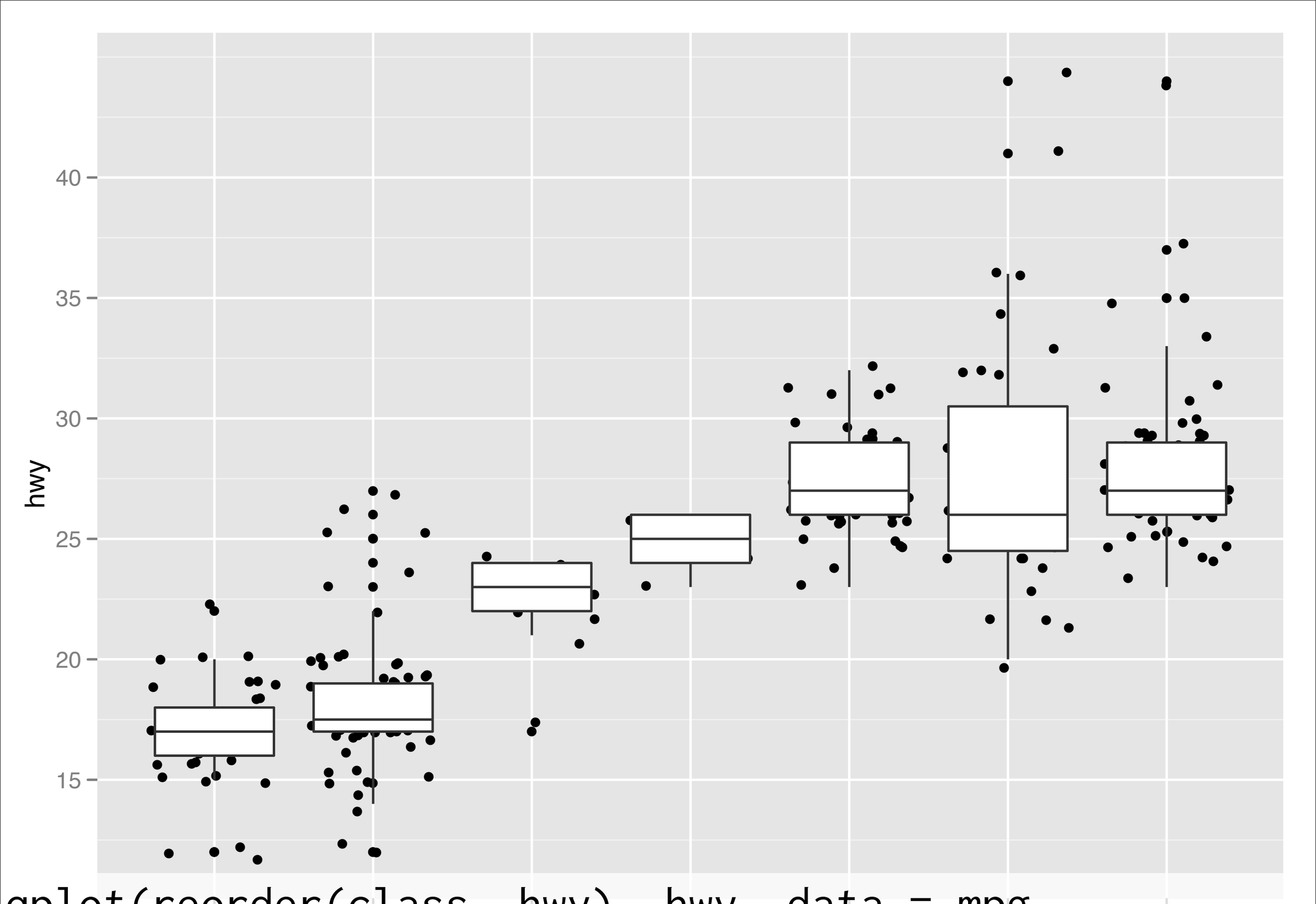

```
qplot(reorder(class, hwy), hwy, data = mpg)
```




```
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 previous 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. Copy from the history panel.
2. Build up the important bits as you go. (recommended)

Learning ggplot2

Book
http://amzn.com/0387981403

Blog
http://blog.ggplot2.org/

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

stackoverflow
http://stackoverflow.com/tags/ggplot2

Cookbook for common graphics
http://wiki.stdout.org/rcookbook/Graphs/