Stat405

Statistical computing & graphics

Hadley Wickham

1. Introductions
2. Syllabus
3. Introduction to linux
4. Introduction to R
5. Diving in

HELLO

my name is

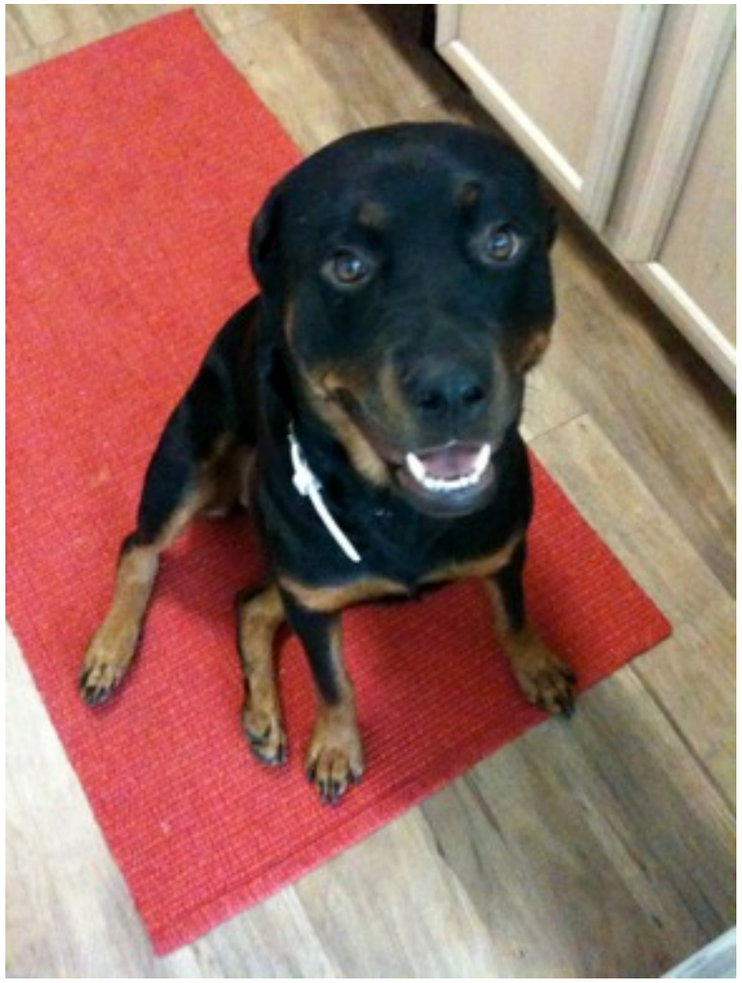
Hadley

At Rice, I'm a

- McMurtry divisional advisor,
- major advisor for Statistics

Away from Rice, I

- love to cook
- have two dogs
- travel too much



http://stat405.had.co.nz

hadley@rice.edu

Syllabus

Tools

Computer: mac/windows/linux

Software: R, text editor, latex (= Rstudio)

Brain: scepticism, curiosity, organisation

Homework

Lowest grade dropped.

Honour code: you can discuss ideas with other class members, but you must present your own work. All code should be your own.

Late policy: 20% penalty if turned in by 9am Monday. Homeworks not accepted after that time.

All homeworks must be submitted in physical form. Electronic versions will only be accepted in exceptional circumstances.

Team projects

3 bigger team projects, culminating in poster presentation at the end of year.

Teams of 4-5 people, assigned by Hadley.

Will teach team work skills. Option to disband after first project. Firing and quitting.

Rstudio

Setup

Install R and Rstudio on your computer, following the instructions on the class website.

You can also use Rstudio online:
https://www.clear.rice.edu/rstudio
(note that the files are saved in your Rice computing account)

RStudio

Untitled1*

Source on Save | Run | Source

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

1:30 (Top Level) R Script

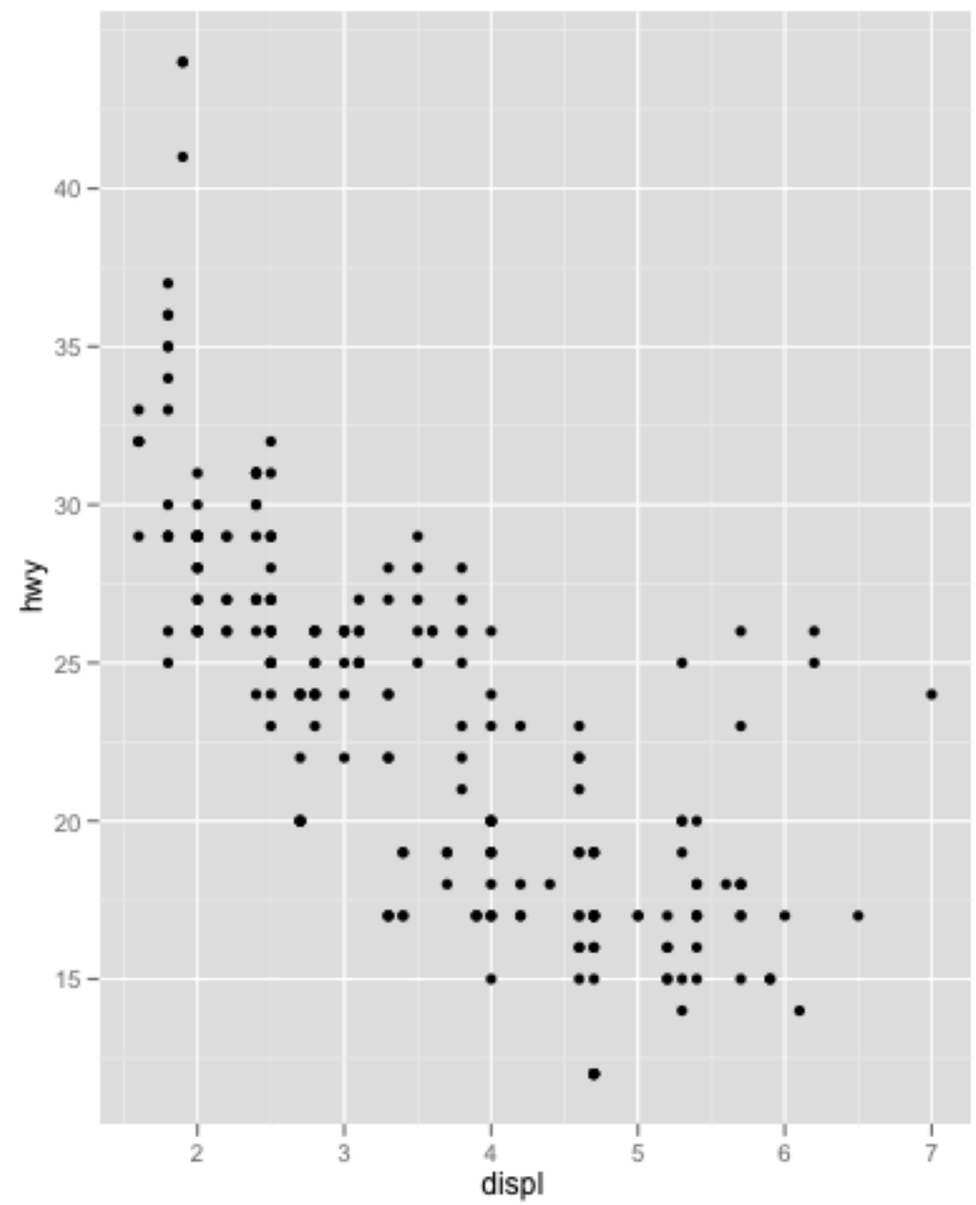
Console ~/

```
or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(ggplot2)
Loading required package: reshape
Loading required package: plyr

Attaching package: 'reshape'

The following object(s) are masked from 'package:plyr':

    rename, round_any

Loading required package: grid
Loading required package: proto
> qplot(displ, cyl, data = mpg)
> qplot(displ, hwy, data = mpg)
>
```

Workspace | History

Files | Plots | Packages | Help

Zoom | Export | Clear All



RStudio

Untitled1*

Source on Save | Run | Source

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

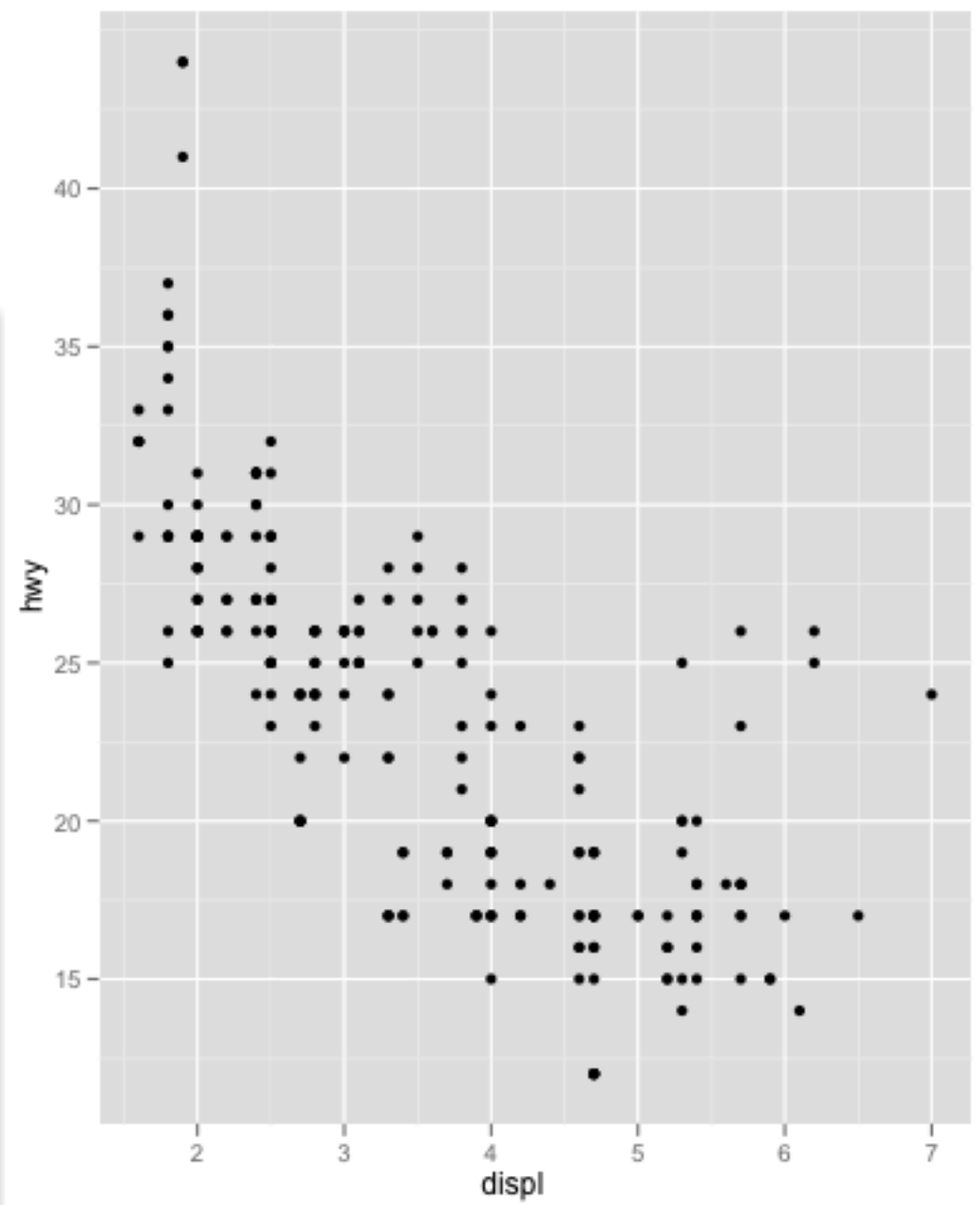
Console ~/

```
or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(ggplot2)
Loading required package: reshape
Loading required package: plyr

Attaching package: 'reshape'

The following object(s) are masked from 'package:plyr':

    rename, round_any

Loading required package: grid
Loading required package: proto
> qplot(displ, cyl, data = mpg)
> qplot(displ, hwy, data = mpg)
>
```

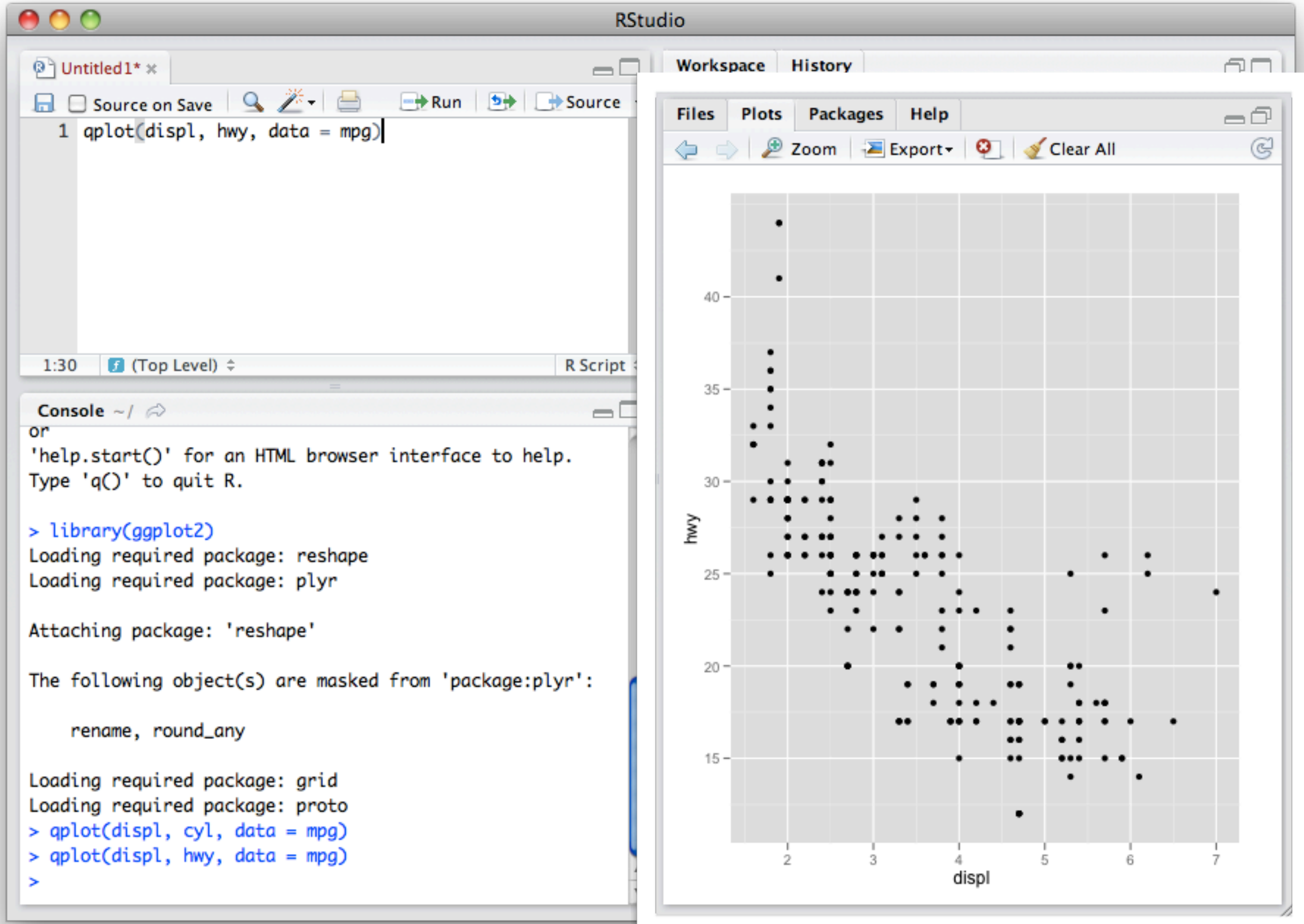
Workspace | History

Files | Plots | Packages | Help

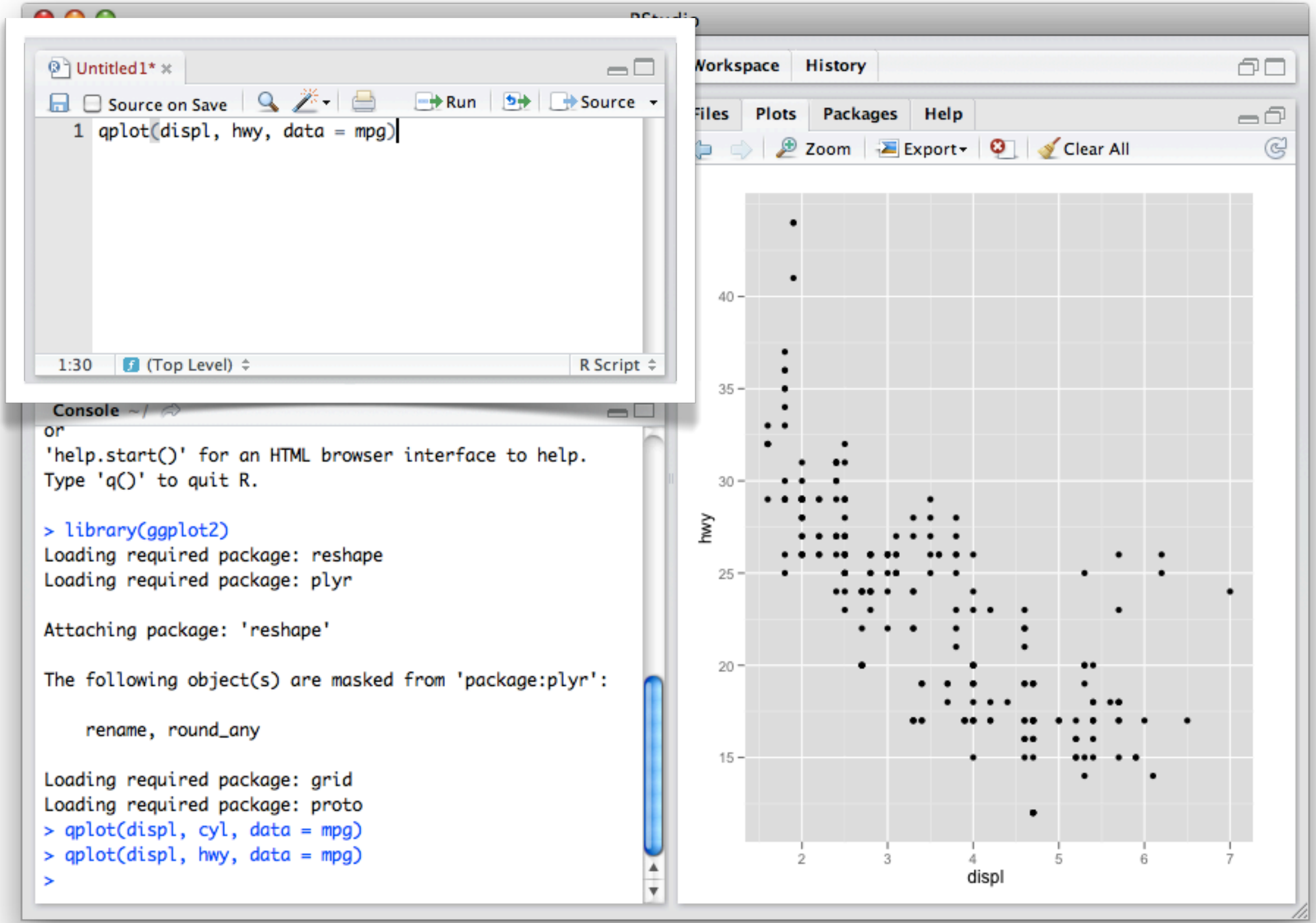
Zoom | Export | Clear All



Console – run code here



Output – plots and help



Editor – save code here

Hint: learn the keyboard shortcuts

Untitled1*

Source on Save | Run | Source

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

1:30 (Top Level) R Script

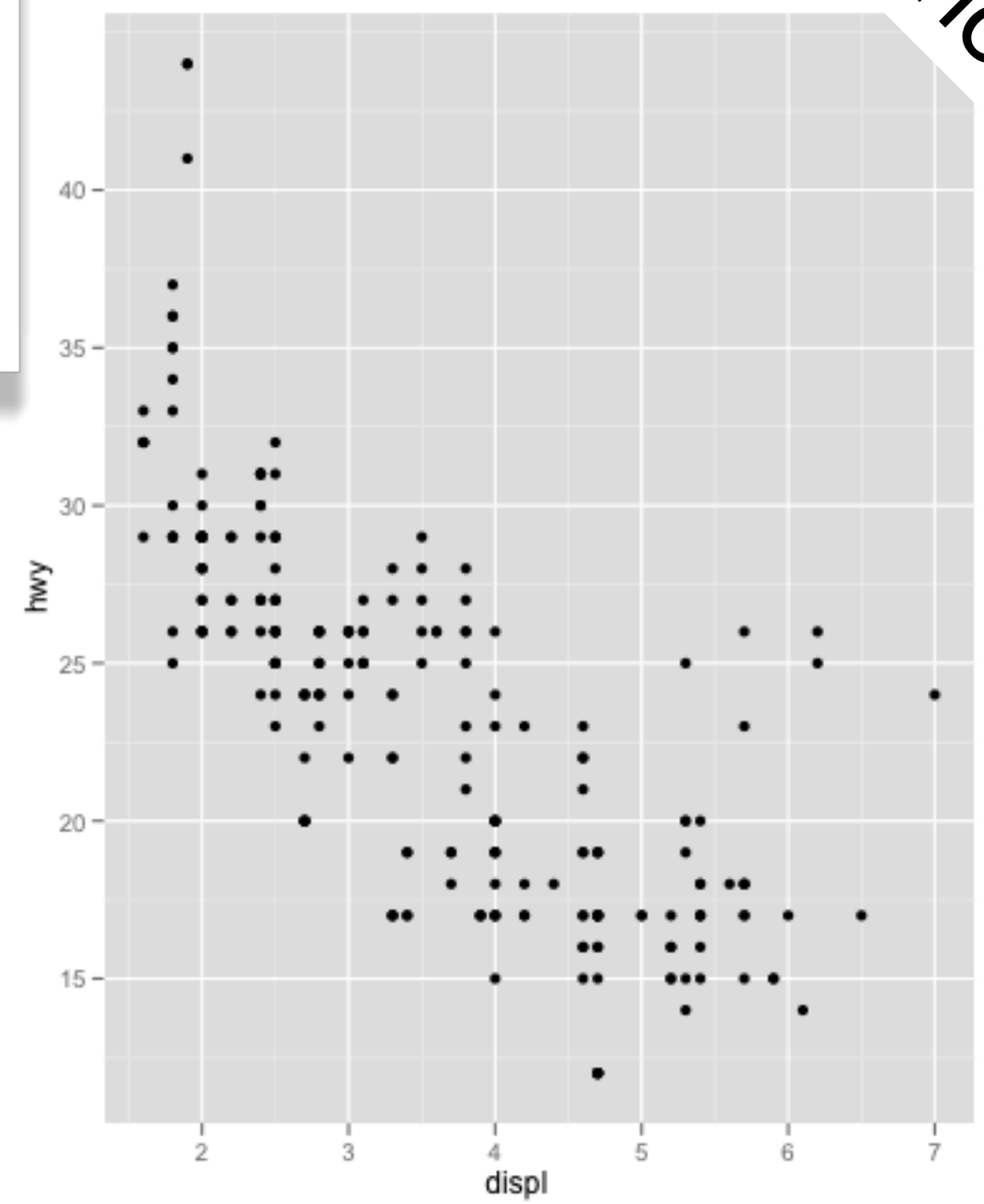
Console ~/

```
or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(ggplot2)
Loading required package: reshape
Loading required package: plyr

Attaching package: 'reshape'

The following object(s) are masked from 'package:plyr':

    rename, round_any

Loading required package: grid
Loading required package: proto
> qplot(displ, cyl, data = mpg)
> qplot(displ, hwy, data = mpg)
>
```

Workspace | History

Files | Plots | Packages | Help

Zoom | Export | Clear All



Editor – save code here

Short cuts

In editor:

Command/ctrl + enter: send code to console

Ctrl + 2: move cursor to console

In console:

Up arrow: retrieve previous command

Ctrl + up arrow: search commands

Ctrl + 1: move cursor to editor

Introduction to R


Learning a new
language is hard!

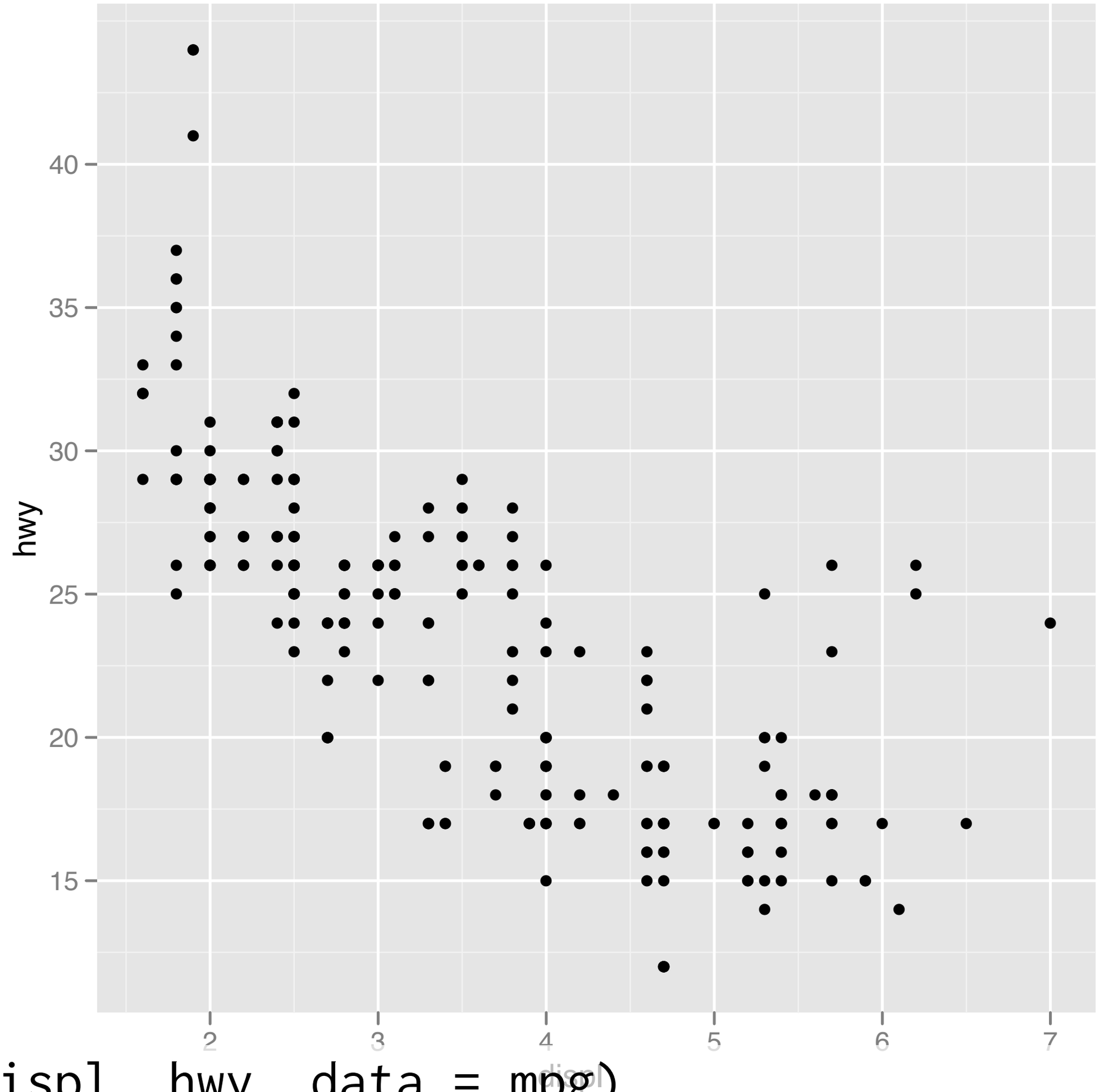
Scatterplot basics

```
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)
```

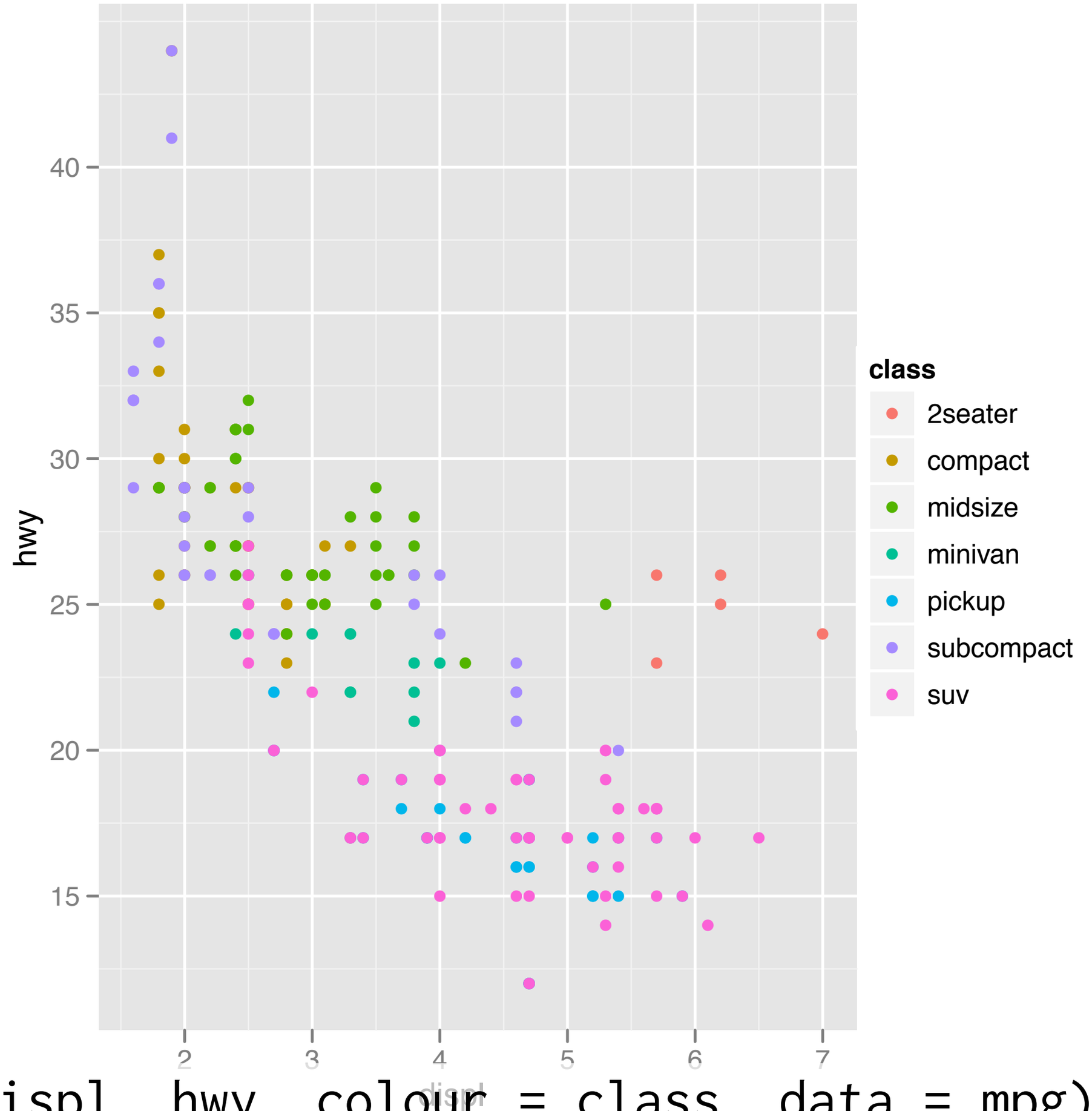
Always explicitly specify the data



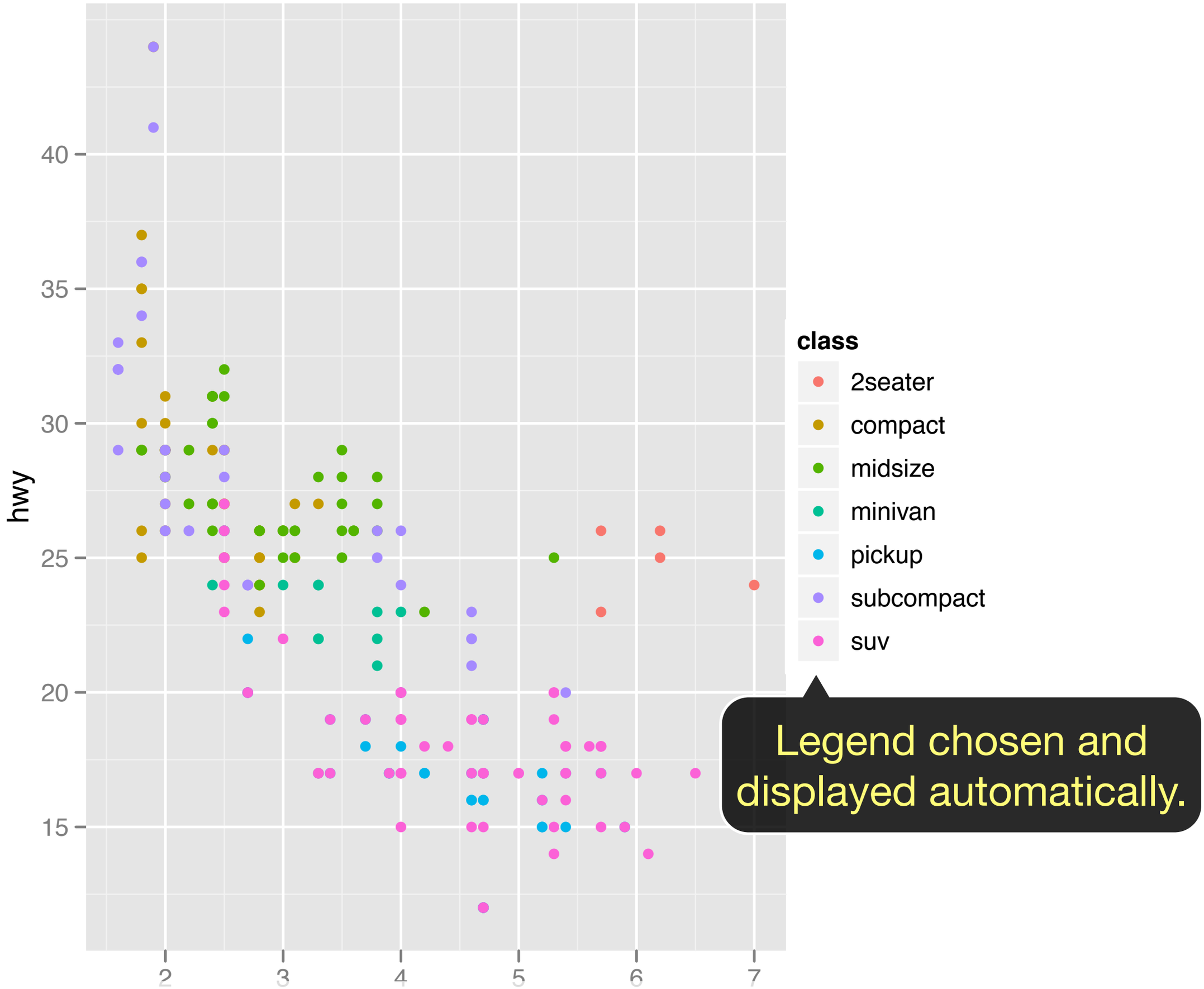

```
qplot(displ, hwy, data = mpg)
```

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	Gradient from red to blue
Size	Discrete size steps	Linear mapping between radius and value
Shape	Different shape for each	Shouldn’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

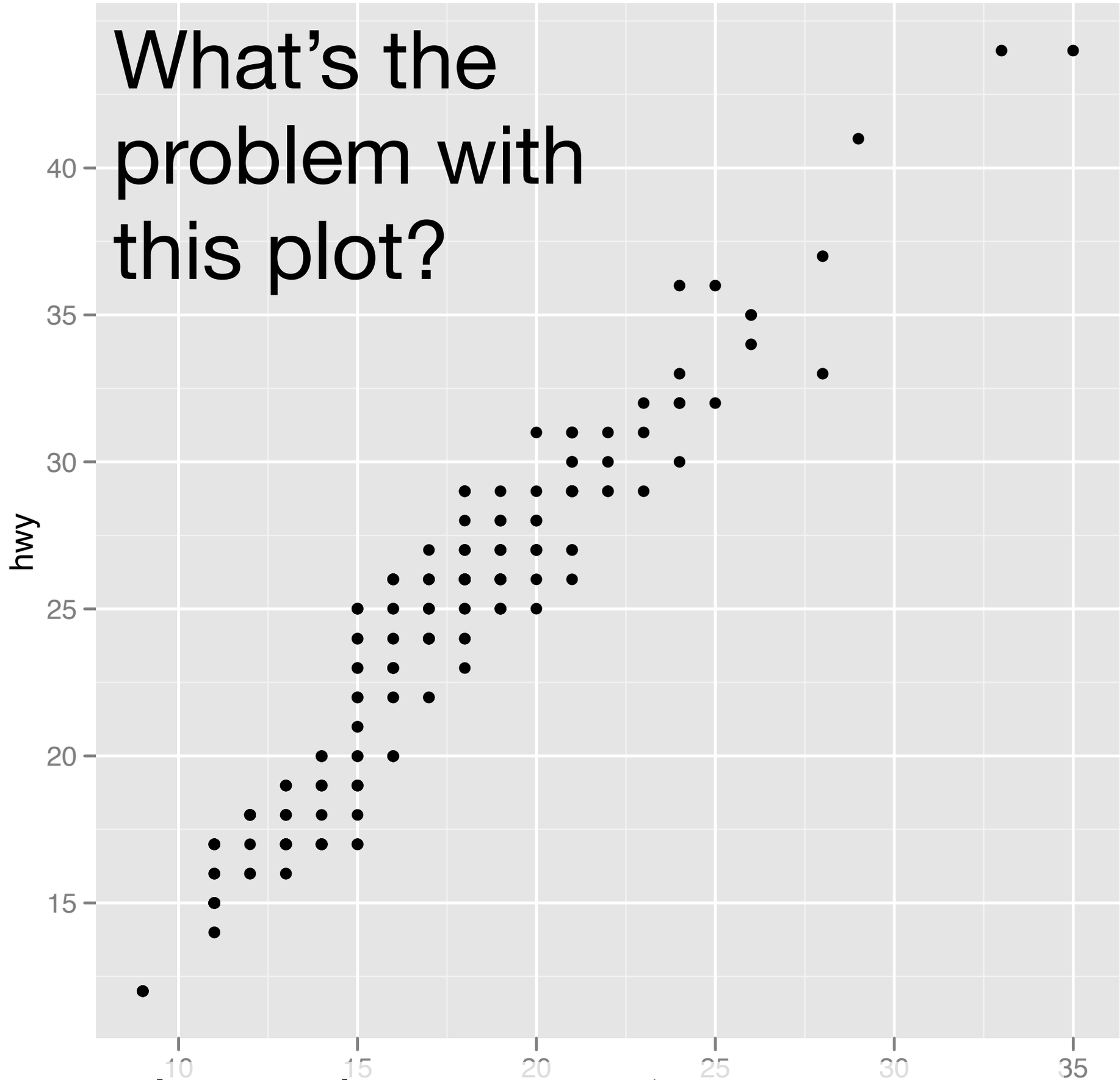
`facet_wrap()`: 1d ribbon wrapped into 2d

Aside: workflow

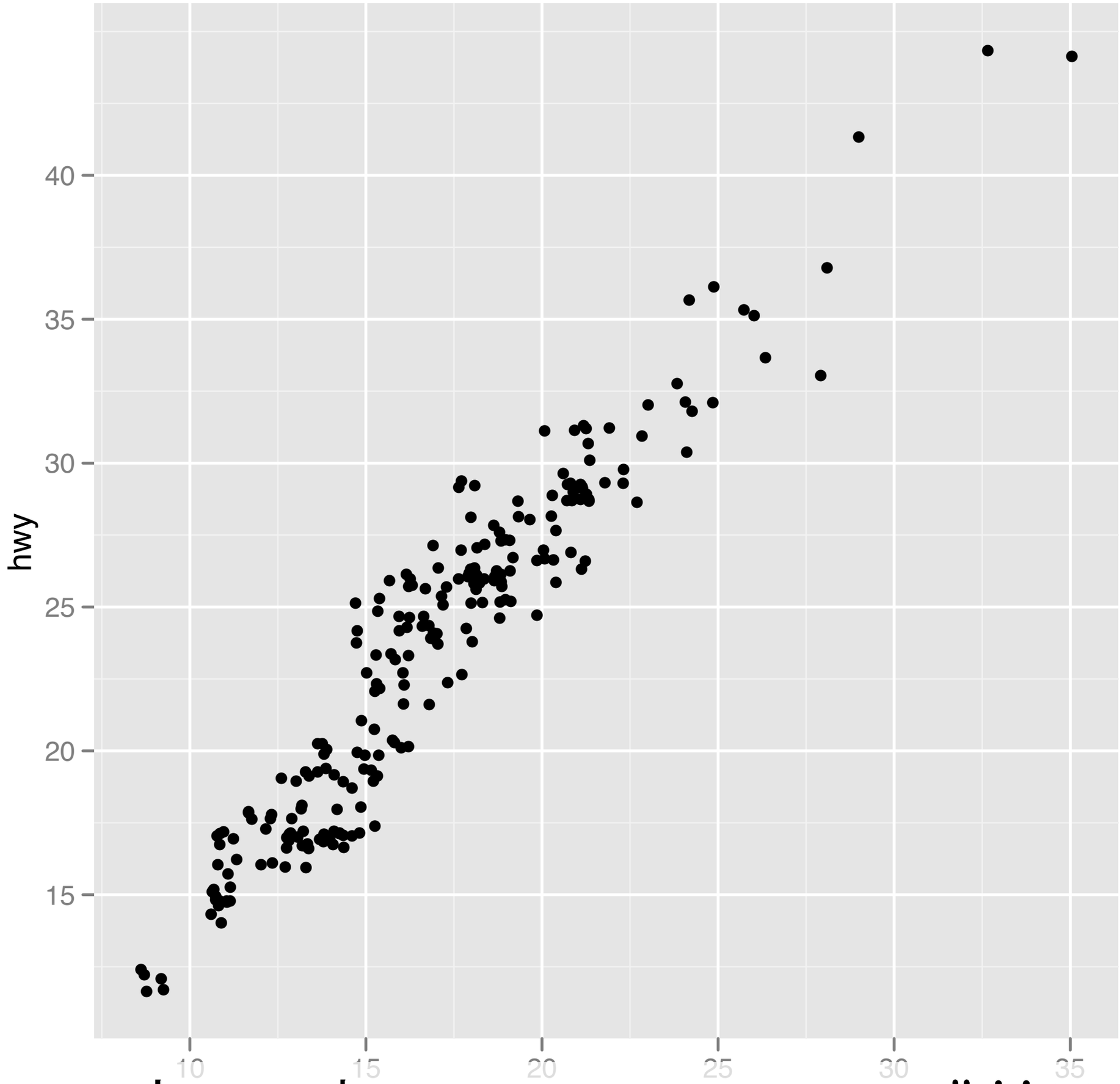
Keep a copy of the slides open so that you can copy and paste the code.

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

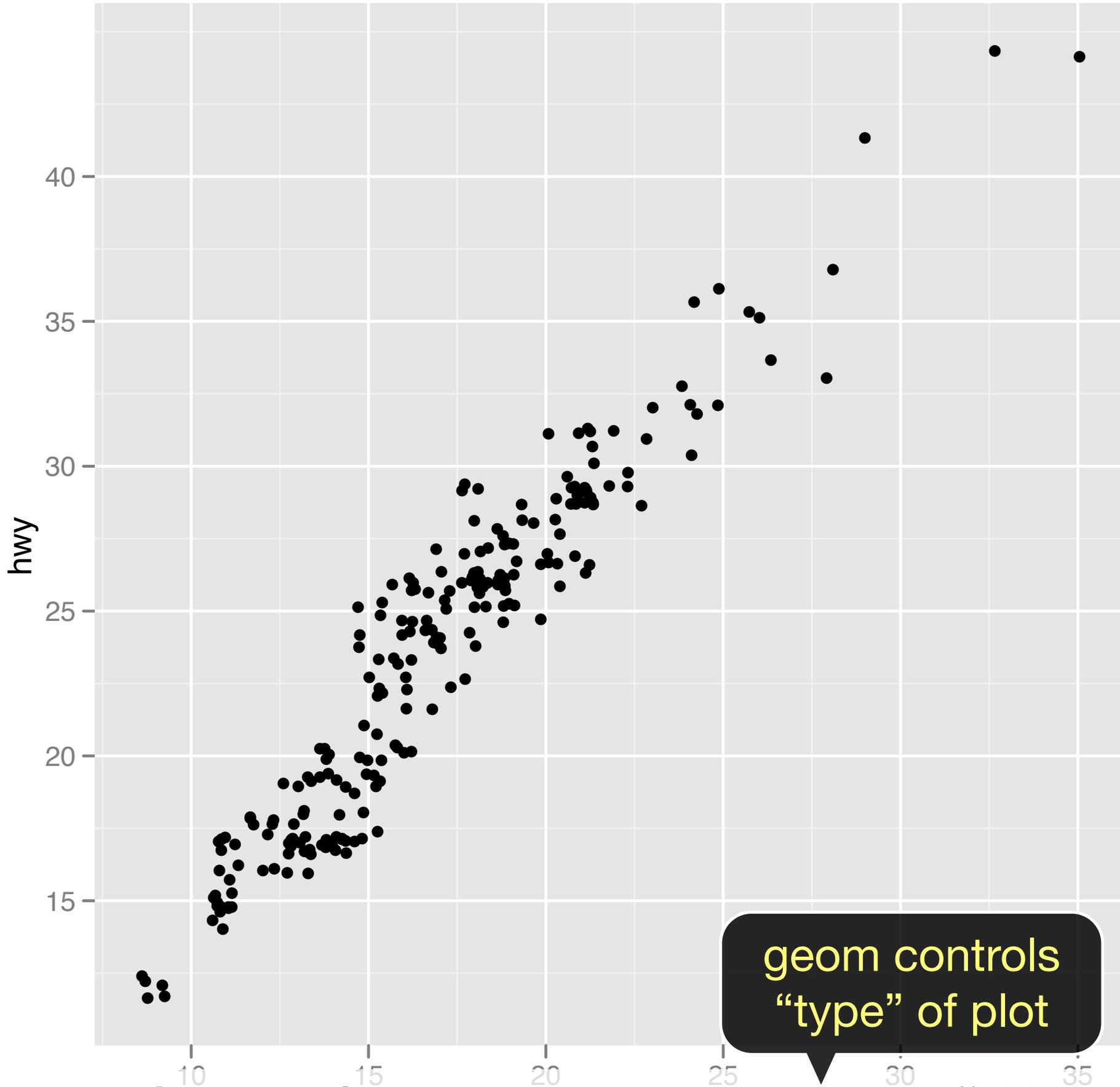
What's the problem with this plot?




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

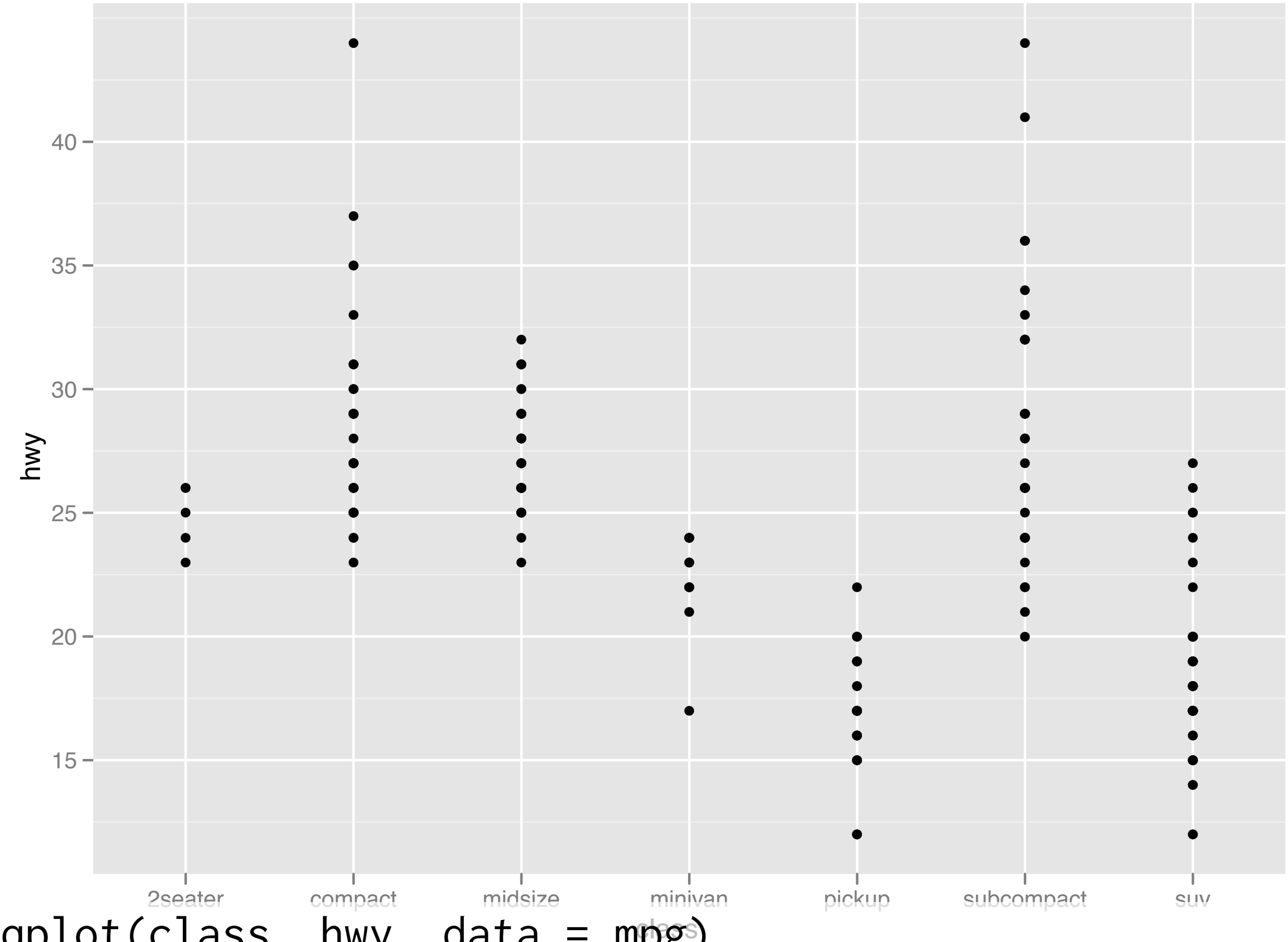



```
qplot(cty, hwy, data = mpg, geom = "jitter")
```



geom controls
"type" of plot

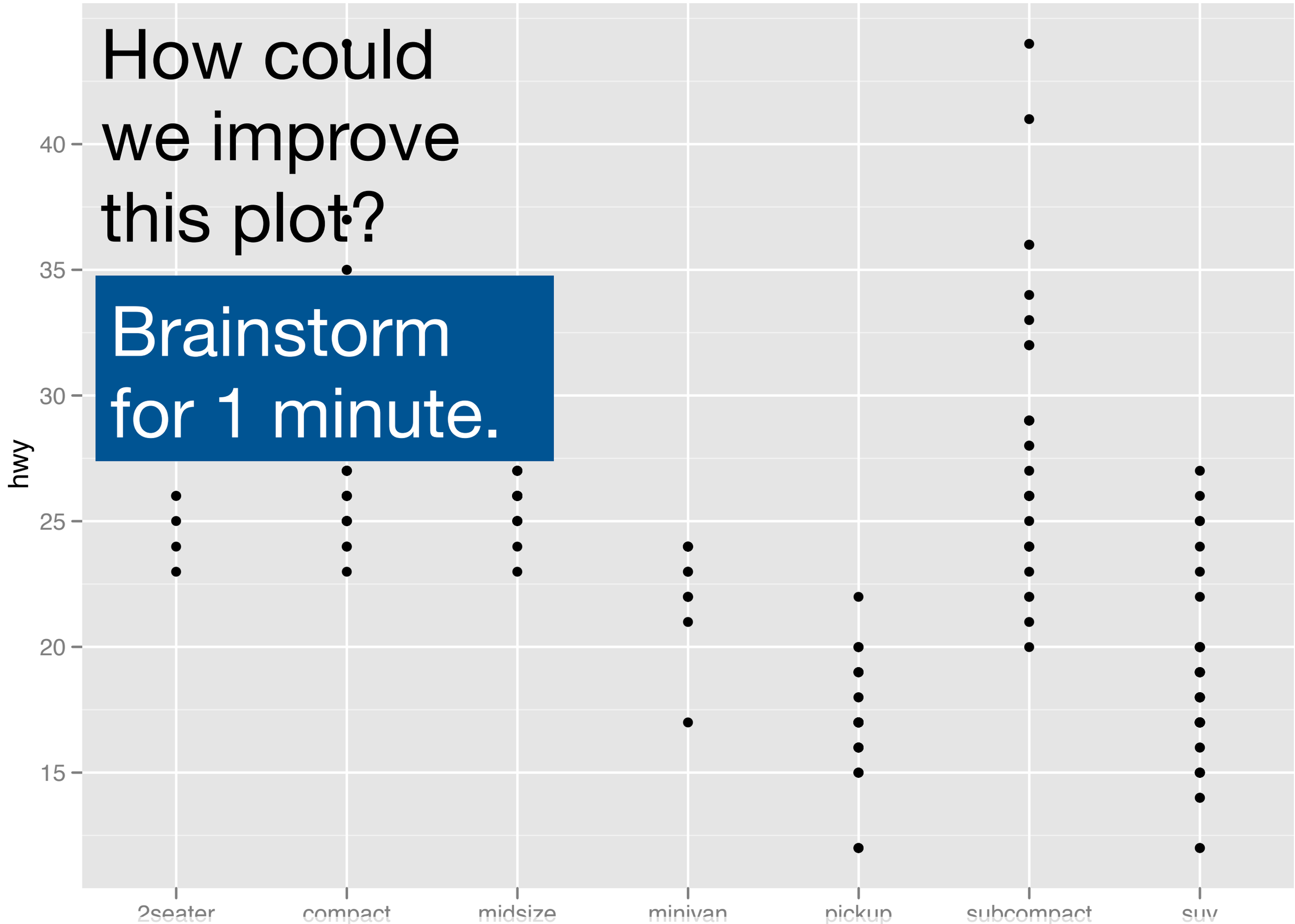
```
qplot(cty, hwy, data = mpg, geom = "jitter")
```



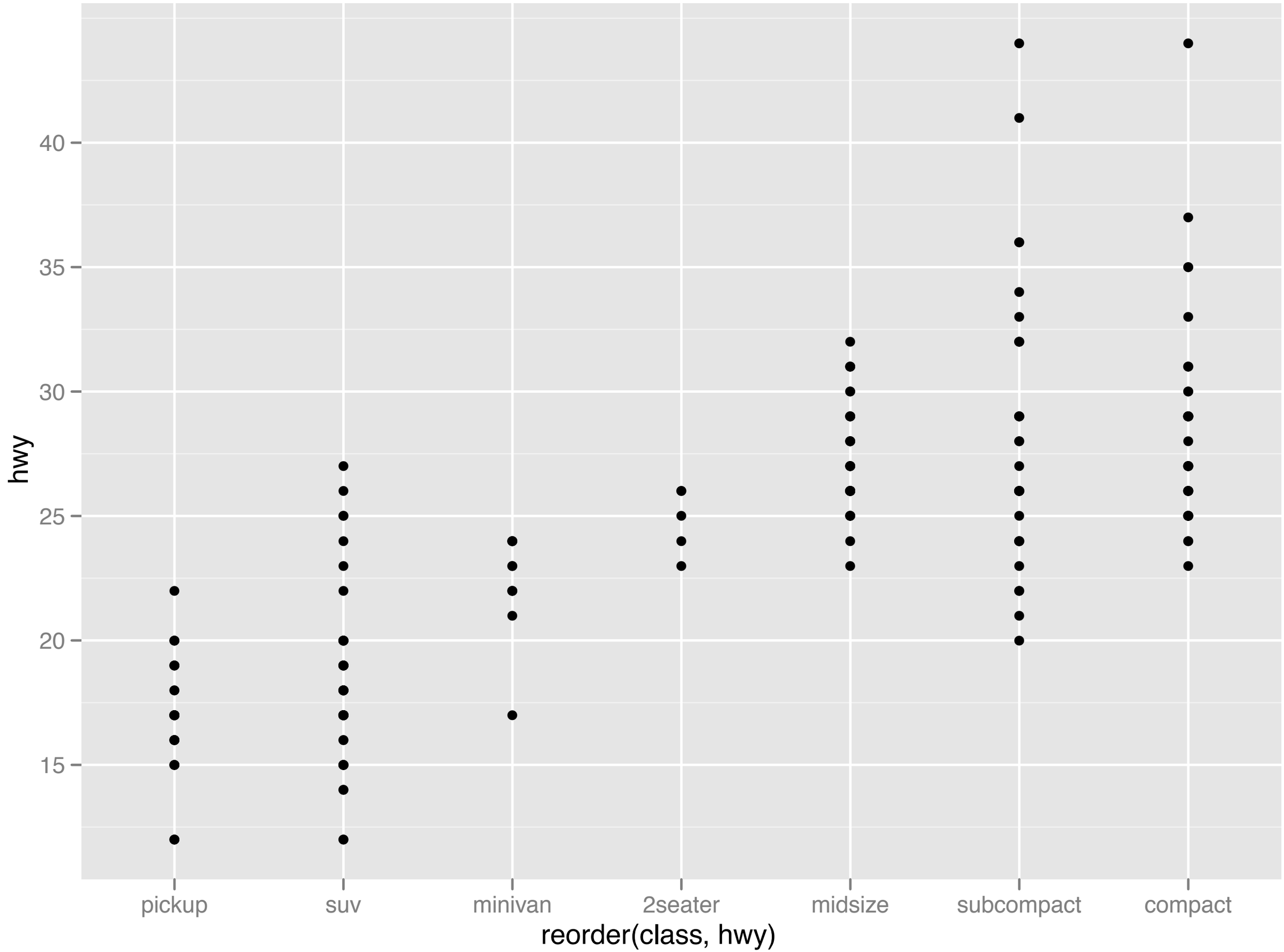

```
qplot(class, hwy, data = mpg)
```

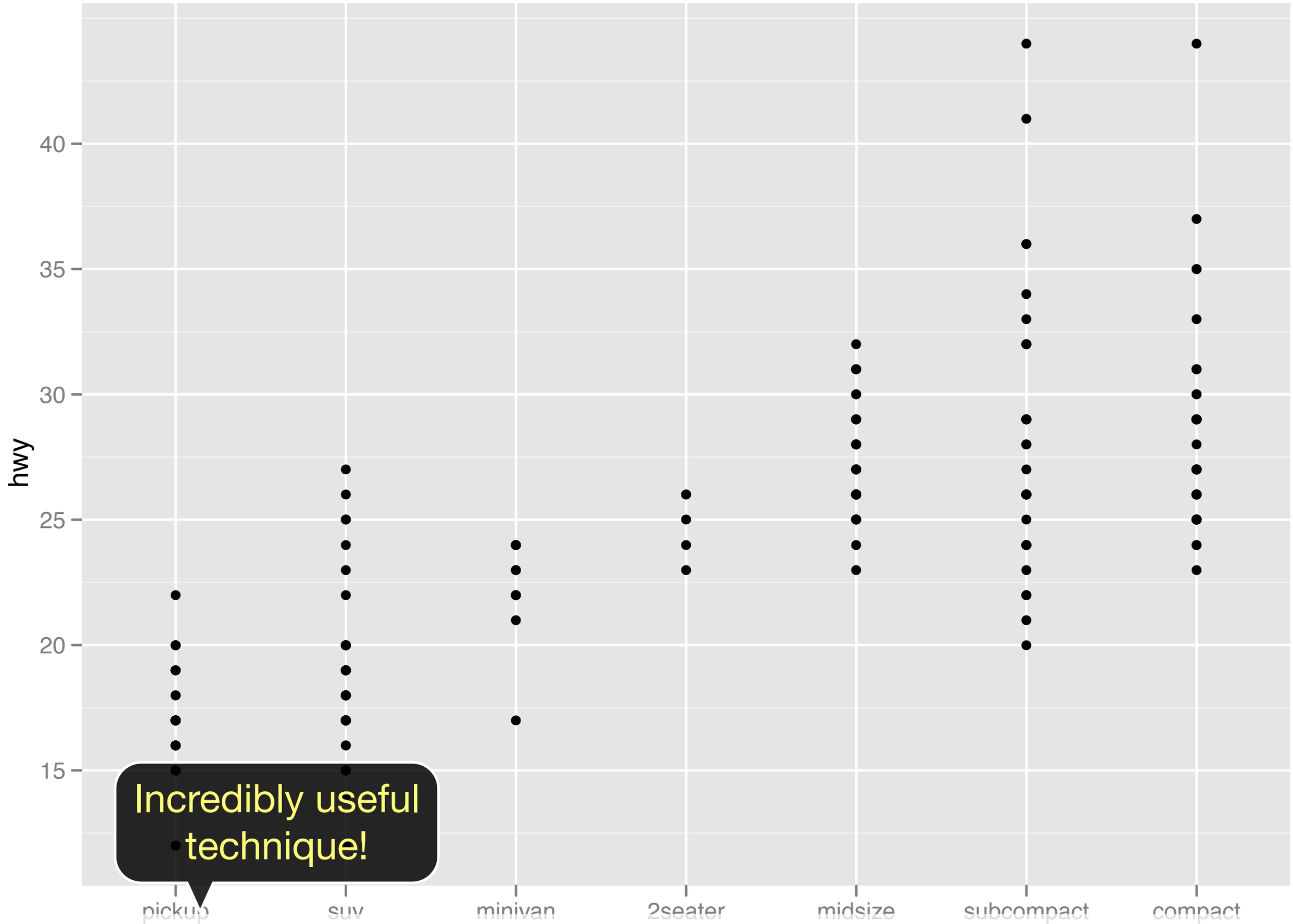
How could we improve this plot?

Brainstorm for 1 minute.

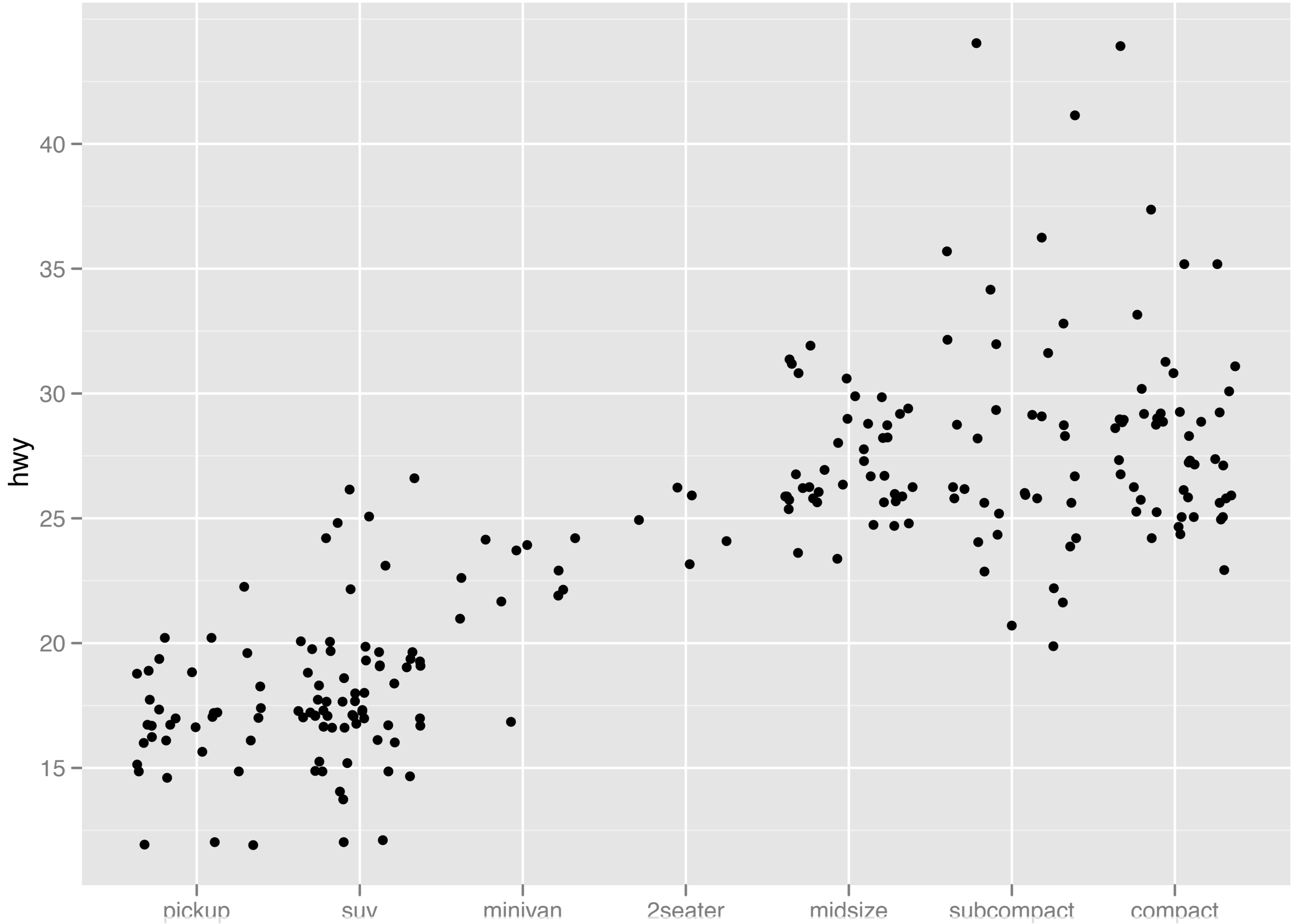



```
qplot(class, hwy, data = mpg)
```

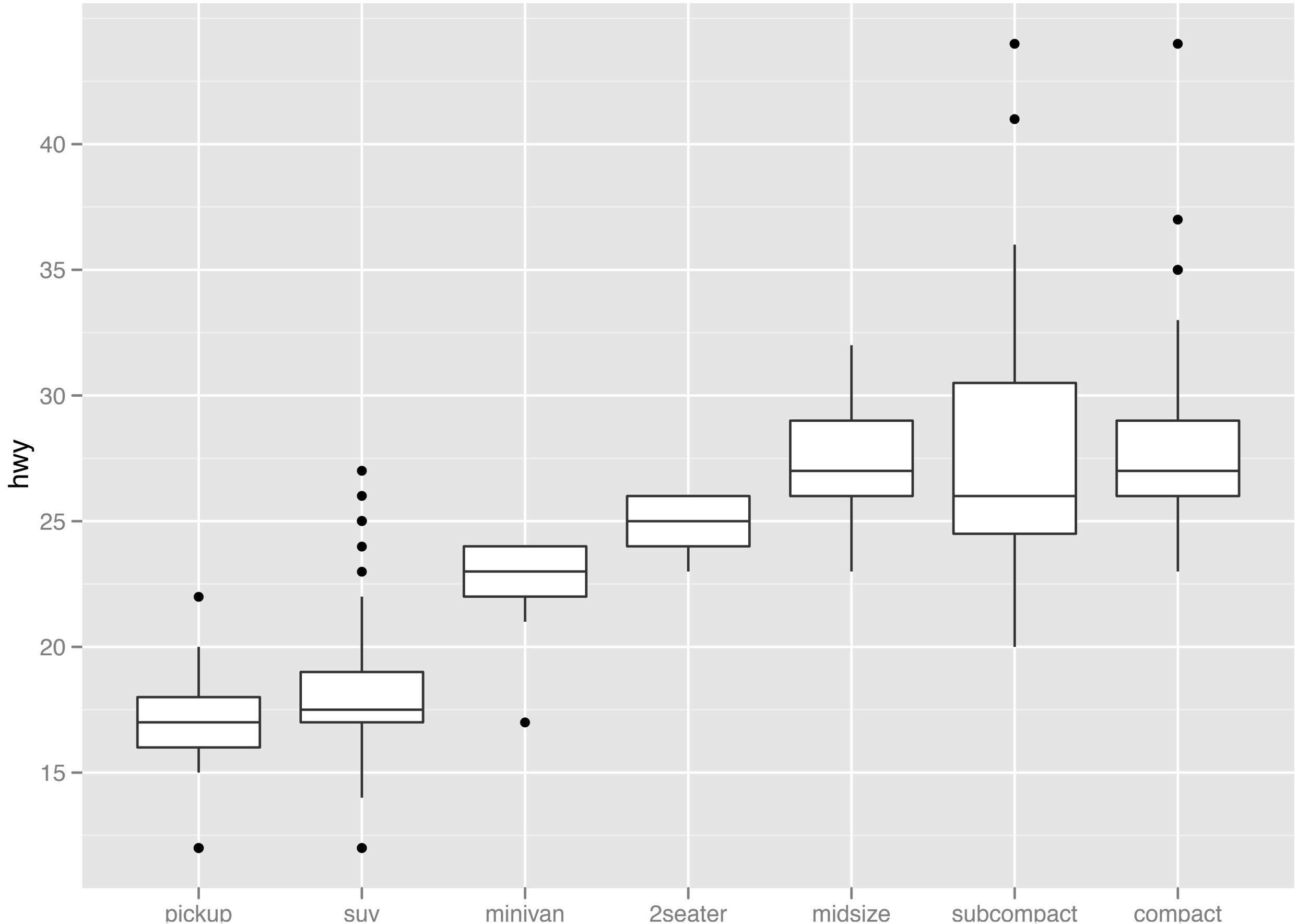

40
35
30
25
20
15
hwy
pickup
suv
minivan
2seater
midsize
subcompact
compact
reorder(class, hwy)



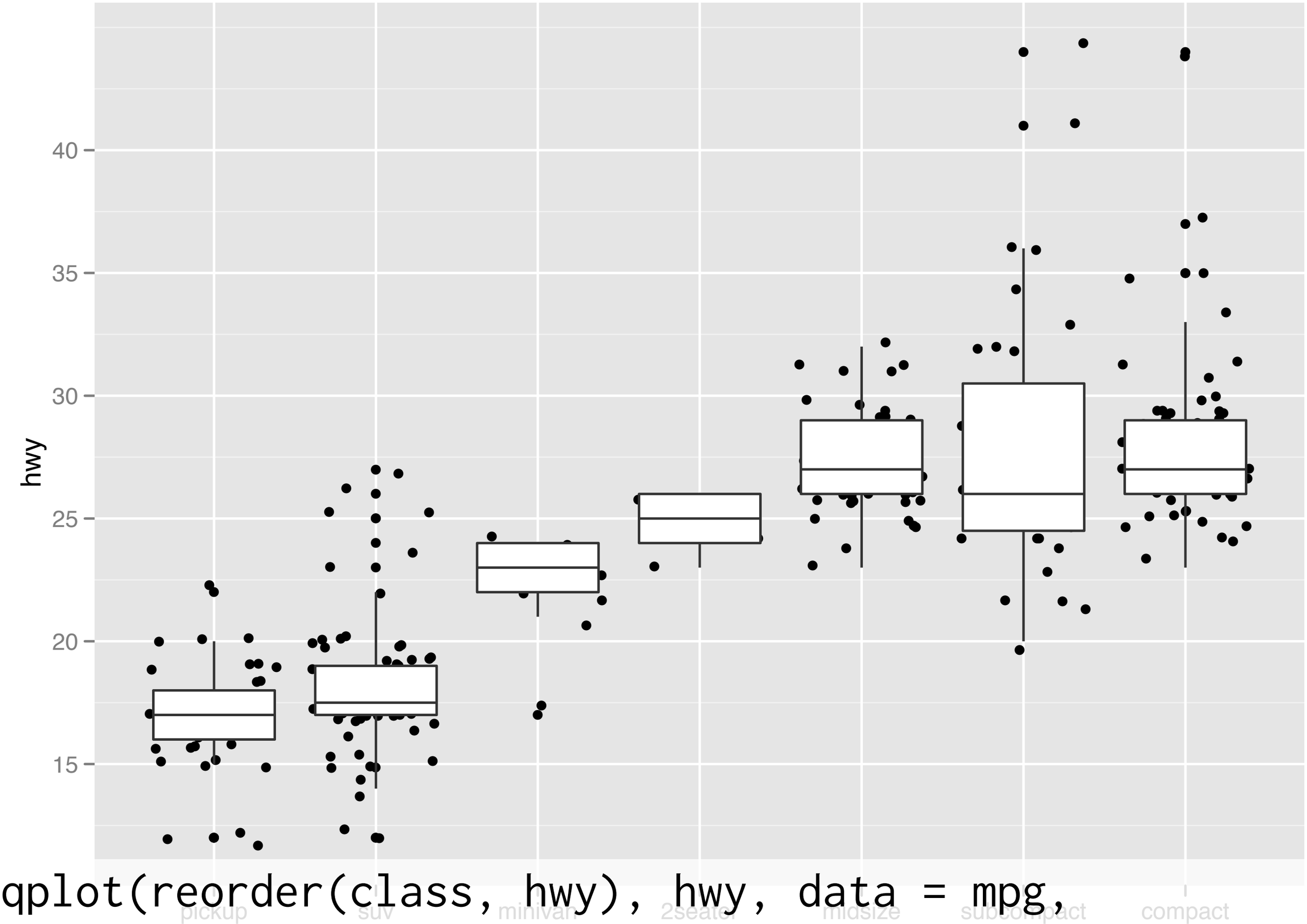

```
qplot(reorder(class, hwy), hwy, data = mpg)
```




```
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)