

ggplot2 basics

Hadley Wickham

Assistant Professor / Dobelman Family Junior Chair
Department of Statistics / Rice University

June 2012



Wednesday, June 13, 12

1. Rstudio

2. Diving in: scatterplots & aesthetics

3. Facetting

4. Geoms

Rstudio

RStudio

Workspace History

Files Plots Packages Help

Zoom Export Clear All

```
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)  
>
```

hwy

displ

The image shows the RStudio interface. The top-left pane is the Source editor, containing the code `1 qplot(displ, hwy, data = mpg)`. The top-right pane is the Plots pane, displaying a scatter plot of highway mileage (hwy) versus engine displacement (displ) for the mpg dataset. The plot shows a negative correlation, with a grid background. The y-axis (hwy) ranges from 15 to 40, and the x-axis (displ) ranges from 2 to 7. The bottom-left pane is the Console, showing the execution of the code and the loading of the ggplot2 package along with its dependencies (reshape, plyr, grid, proto). The console output includes: `> 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)`, and `>`.

Console – run code here

RStudio

Workspace History

Untitled1* x

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)  
>
```

Files Plots Packages Help

Zoom Export Clear All

The plot displays a scatter of data points from the 'mpg' dataset. The x-axis is labeled 'displ' and ranges from approximately 1.8 to 7.0. The y-axis is labeled 'hwy' and ranges from 15 to 45. The data points generally show a downward trend, indicating that as engine displacement increases, highway mileage tends to decrease. There are several notable outliers with low displacement (around 1.8-2.0) and high highway mileage (between 35 and 45).

Output – plots and help

The image shows the RStudio interface. The top-left pane is the R Script editor, titled 'Untitled1*'. It contains the following code:

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

The bottom-left pane is the Console, showing the following output:

```
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)  
>
```

The right pane displays a scatter plot of highway mileage (hwy) versus engine displacement (displ). The x-axis is labeled 'displ' and ranges from approximately 1.8 to 7.0. The y-axis is labeled 'hwy' and ranges from approximately 12 to 45. The plot shows a negative correlation between engine displacement and highway mileage.

Editor – save code here

Hint: learn the keyboard shortcuts

The image shows the RStudio interface. The top-left pane is the R Script editor, containing the following code:

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

The bottom-left pane is the Console, showing the following output:

```
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)  
>
```

The right pane displays a scatter plot of highway mileage (hwy) on the y-axis versus engine displacement (displ) on the x-axis. The plot shows a negative correlation between the two variables, with a grid background. The x-axis ranges from approximately 1.8 to 7.0, and the y-axis ranges from 15 to 40.

Editor – save code here

Short cuts

In editor:

Command + 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

Divining in



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
```

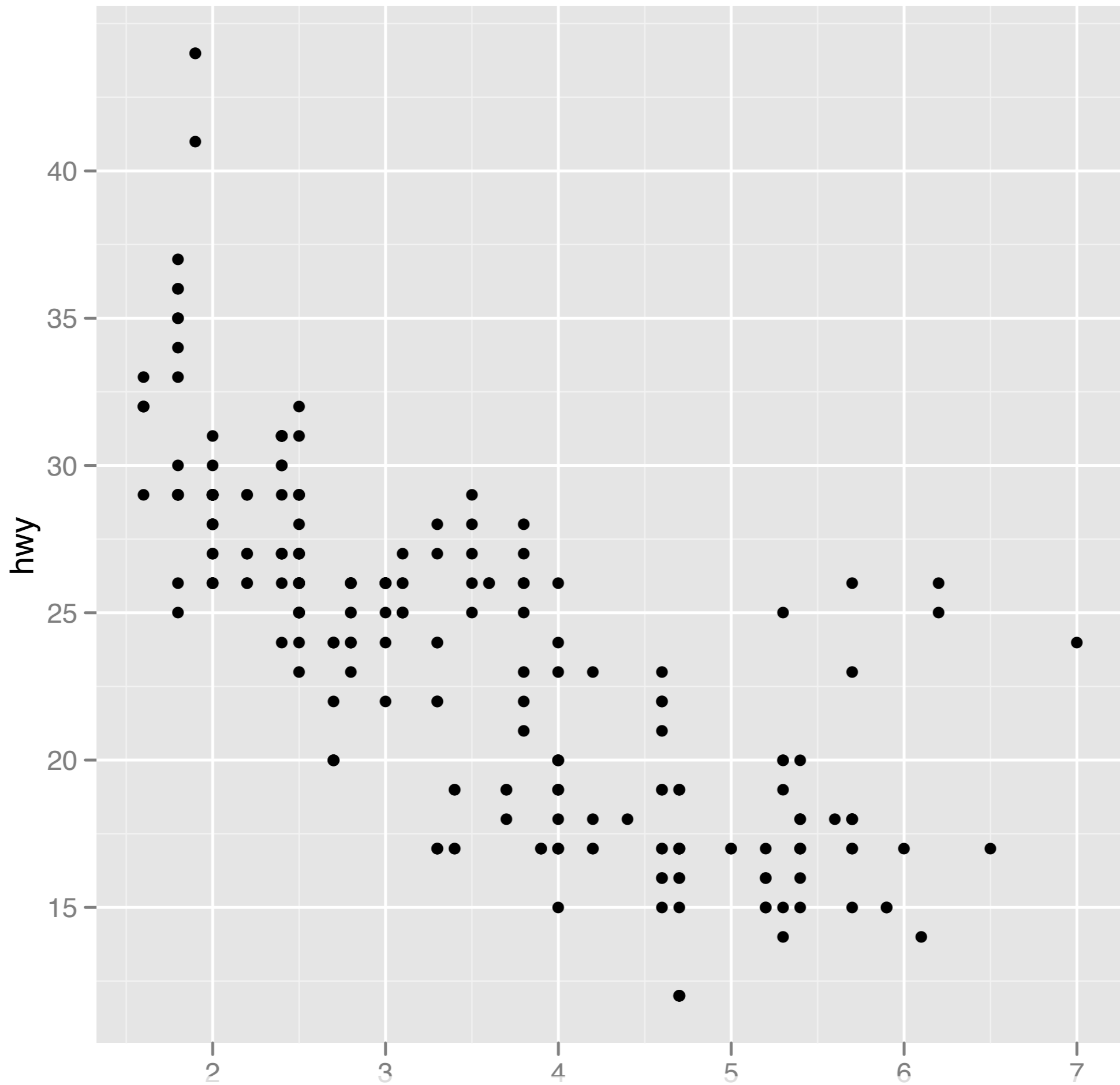
```
head(mpg)
```

```
str(mpg)
```

```
summary(mpg)
```

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

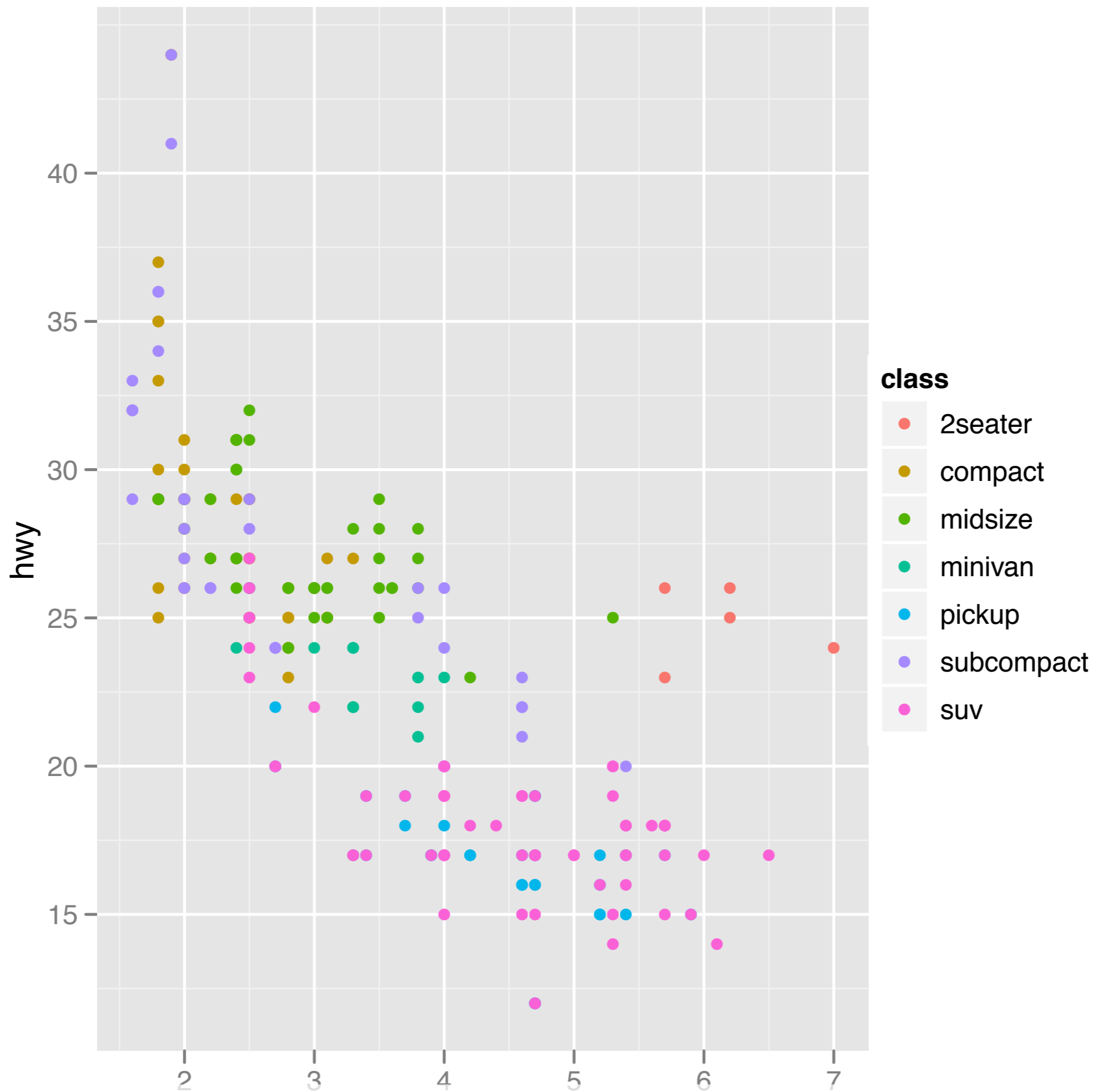
Always explicitly
specify the data



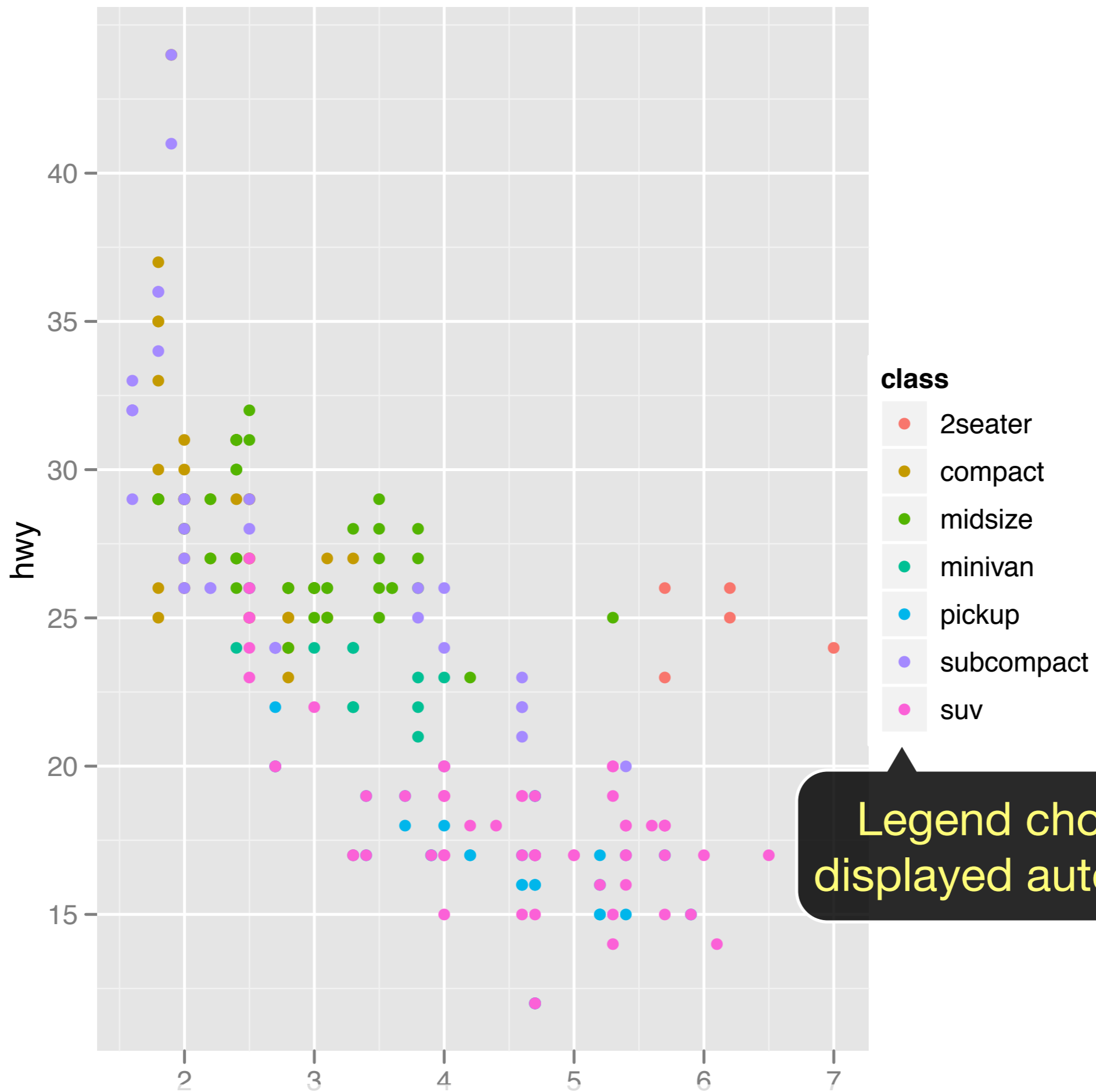
`qplot(displ, hwy, data = mpg)`

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	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

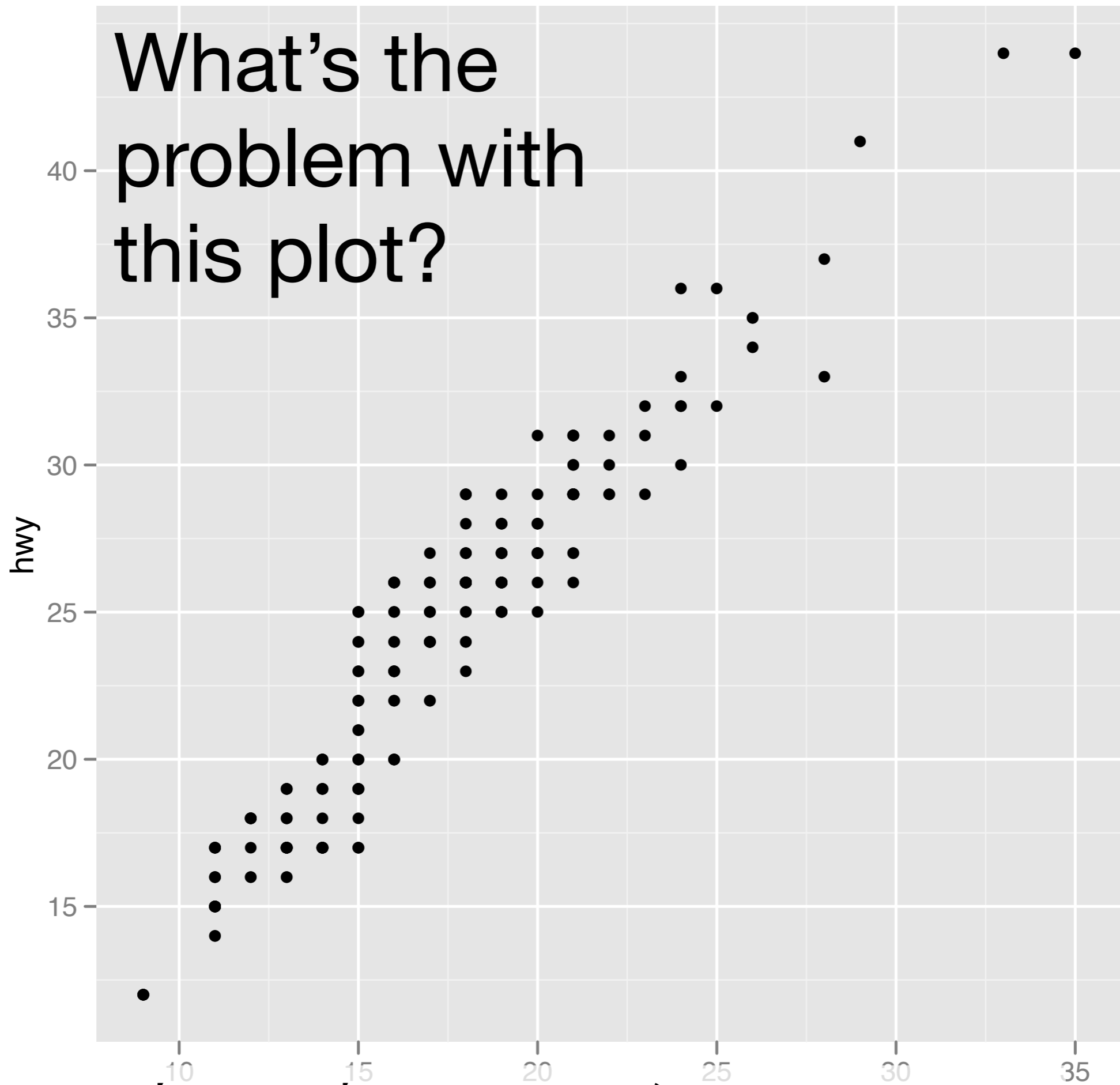
Aside: workflow

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

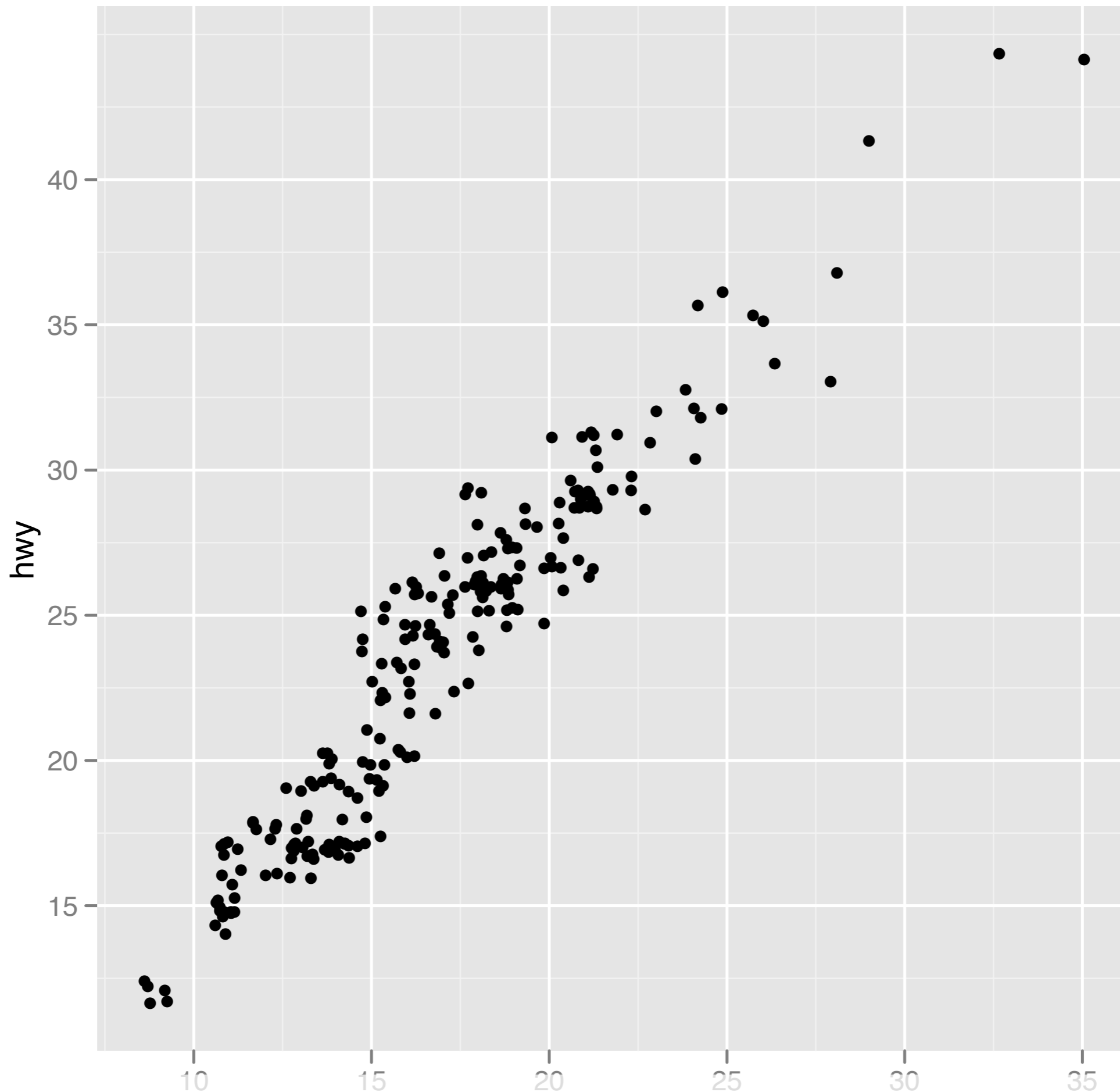
For complicated commands, write them in the editing area and then copy and paste.

Geoms

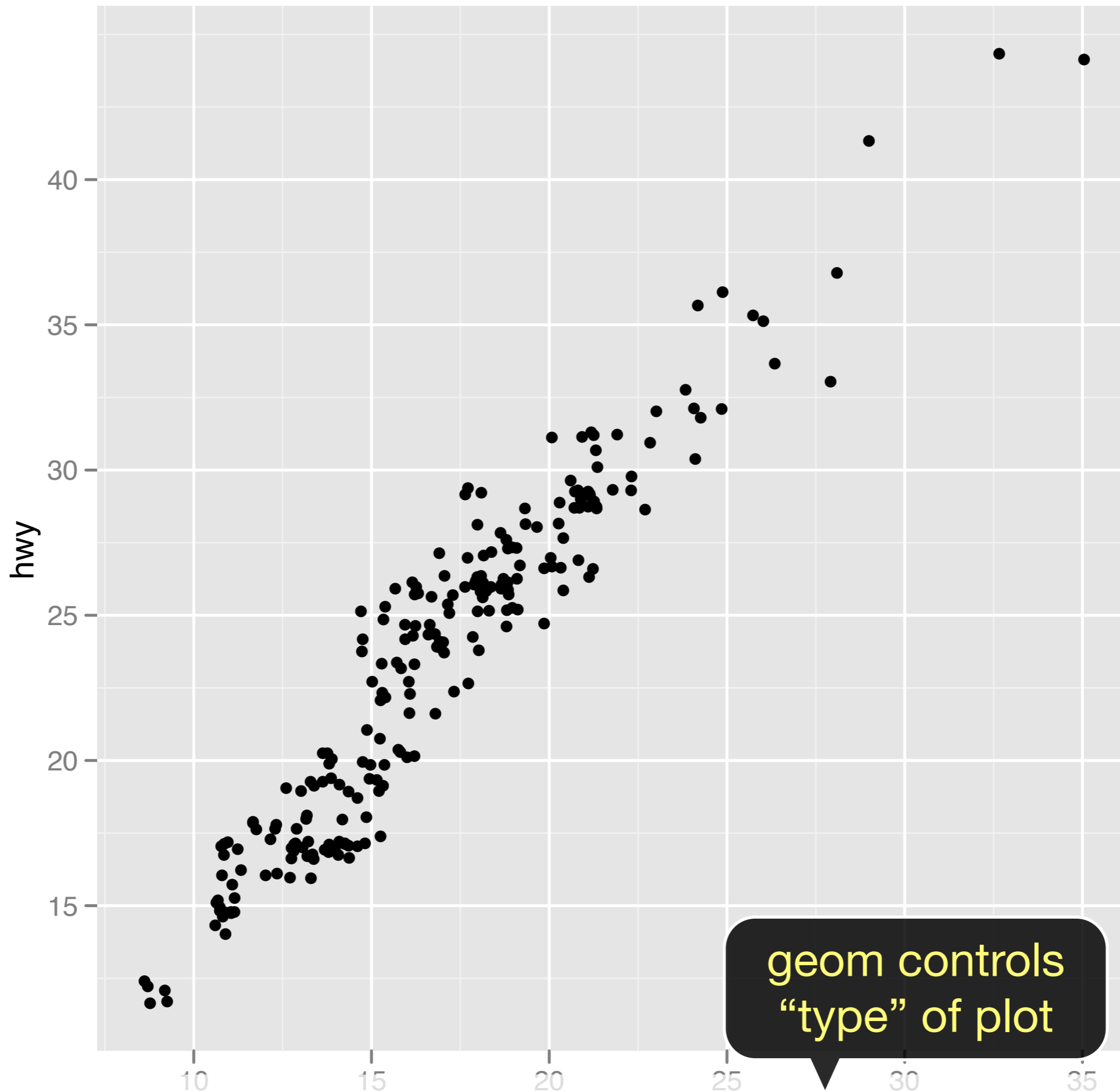
What's the
problem with
this plot?



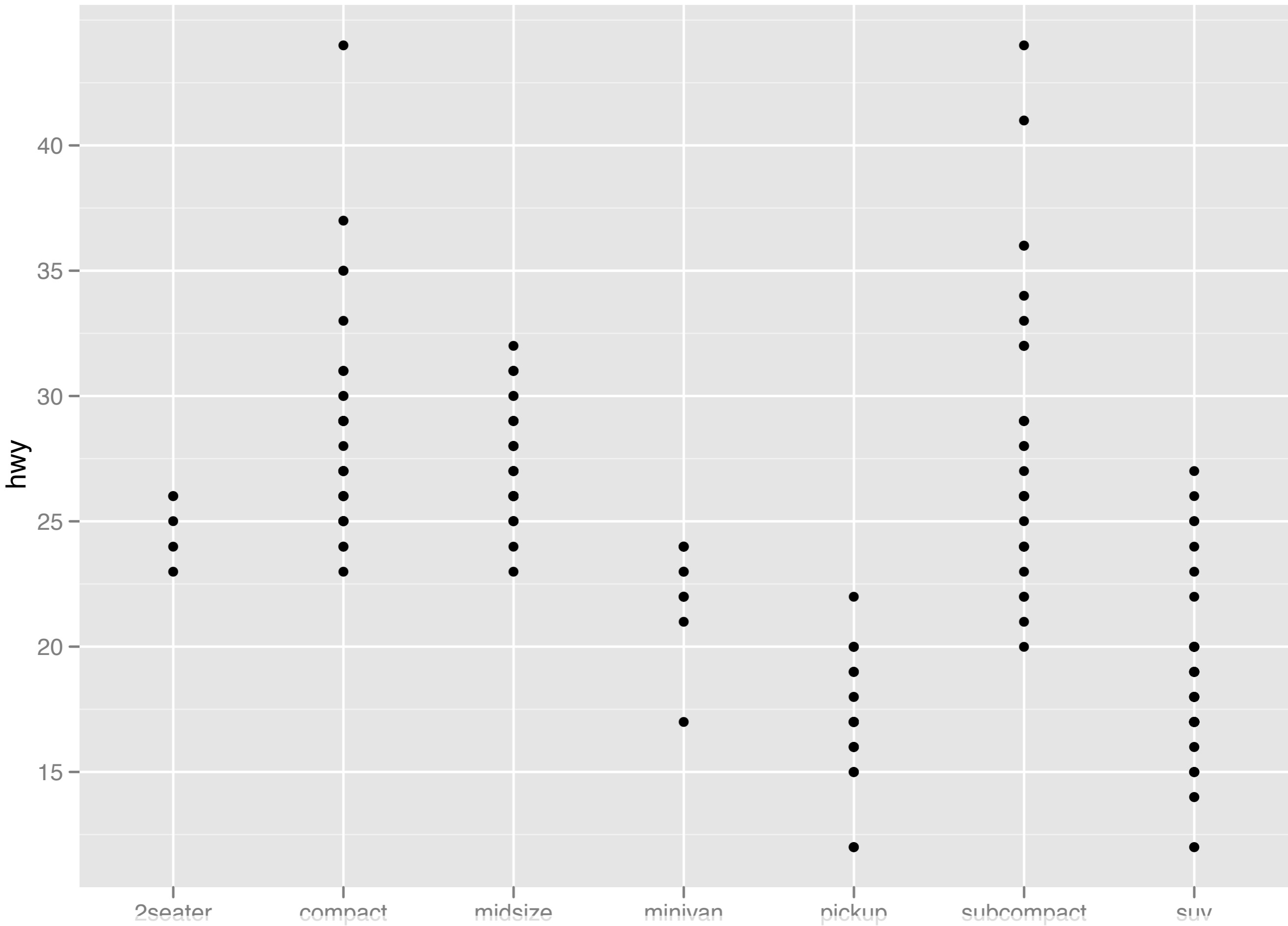
```
qplot(cty, hwy, data = mpg)
```



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



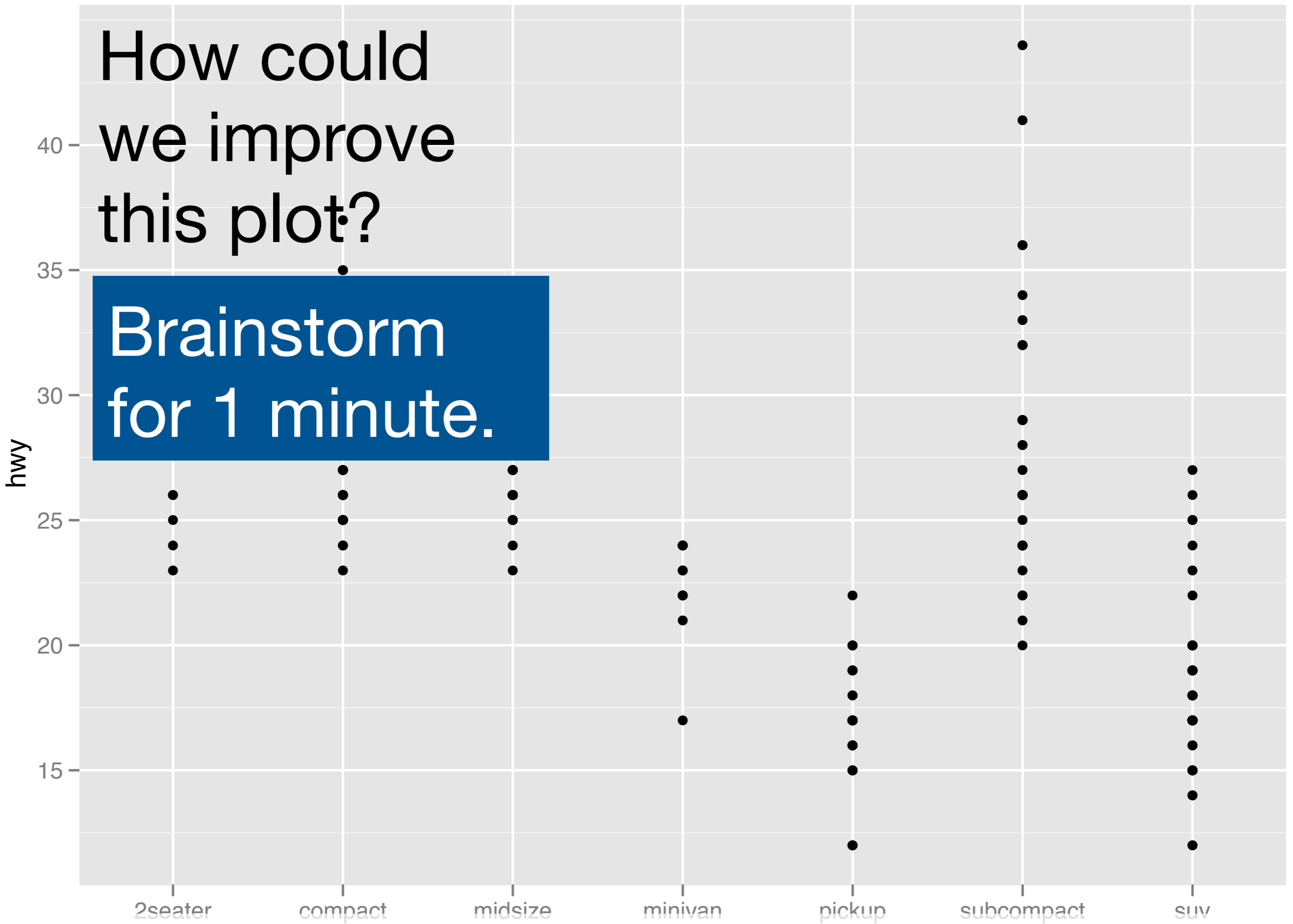
```
qplot(cty, hwy, data = mpg, geom = "jitter")
```



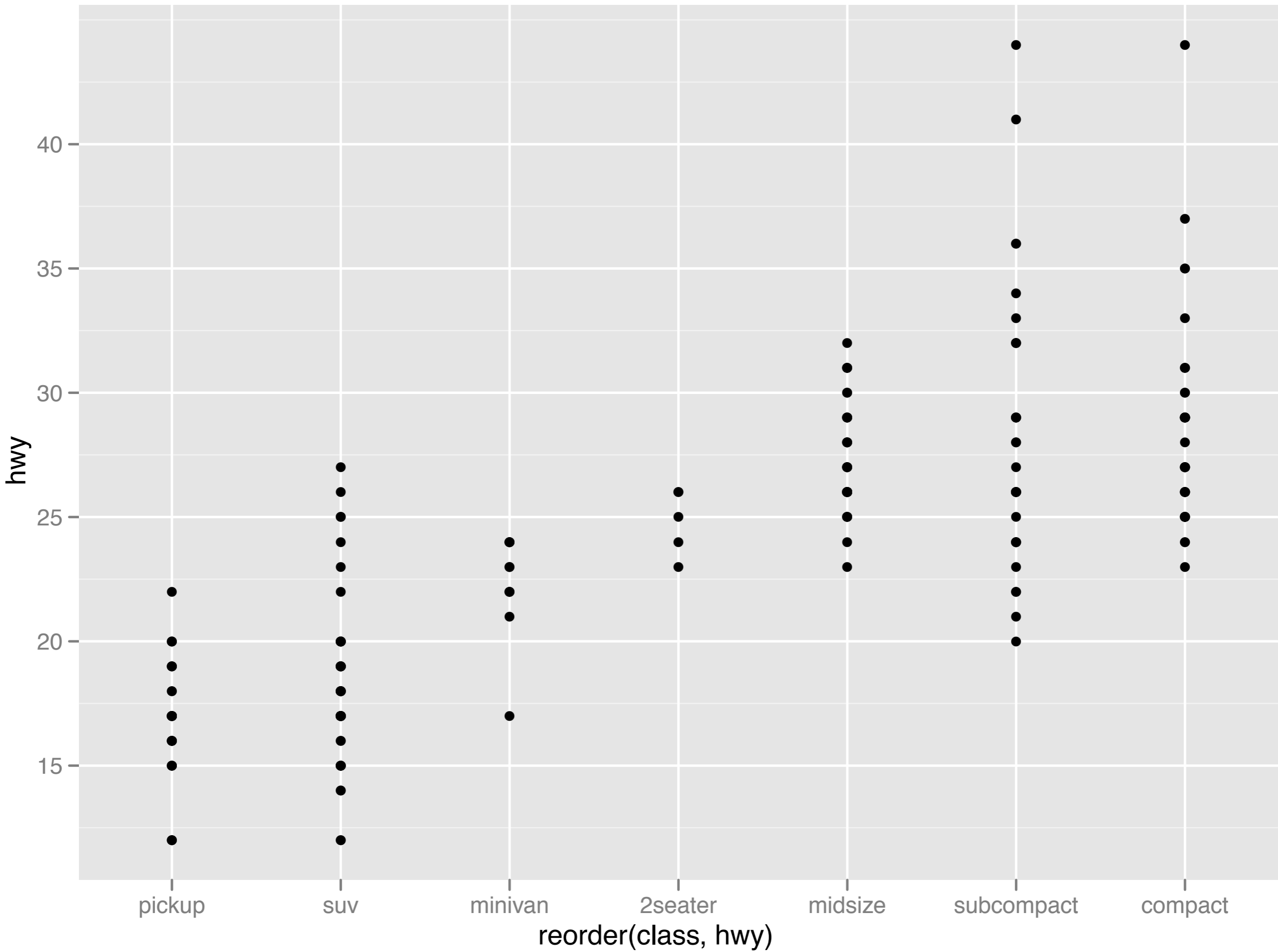
```
qplot(class, hwy, data = mpg)
```

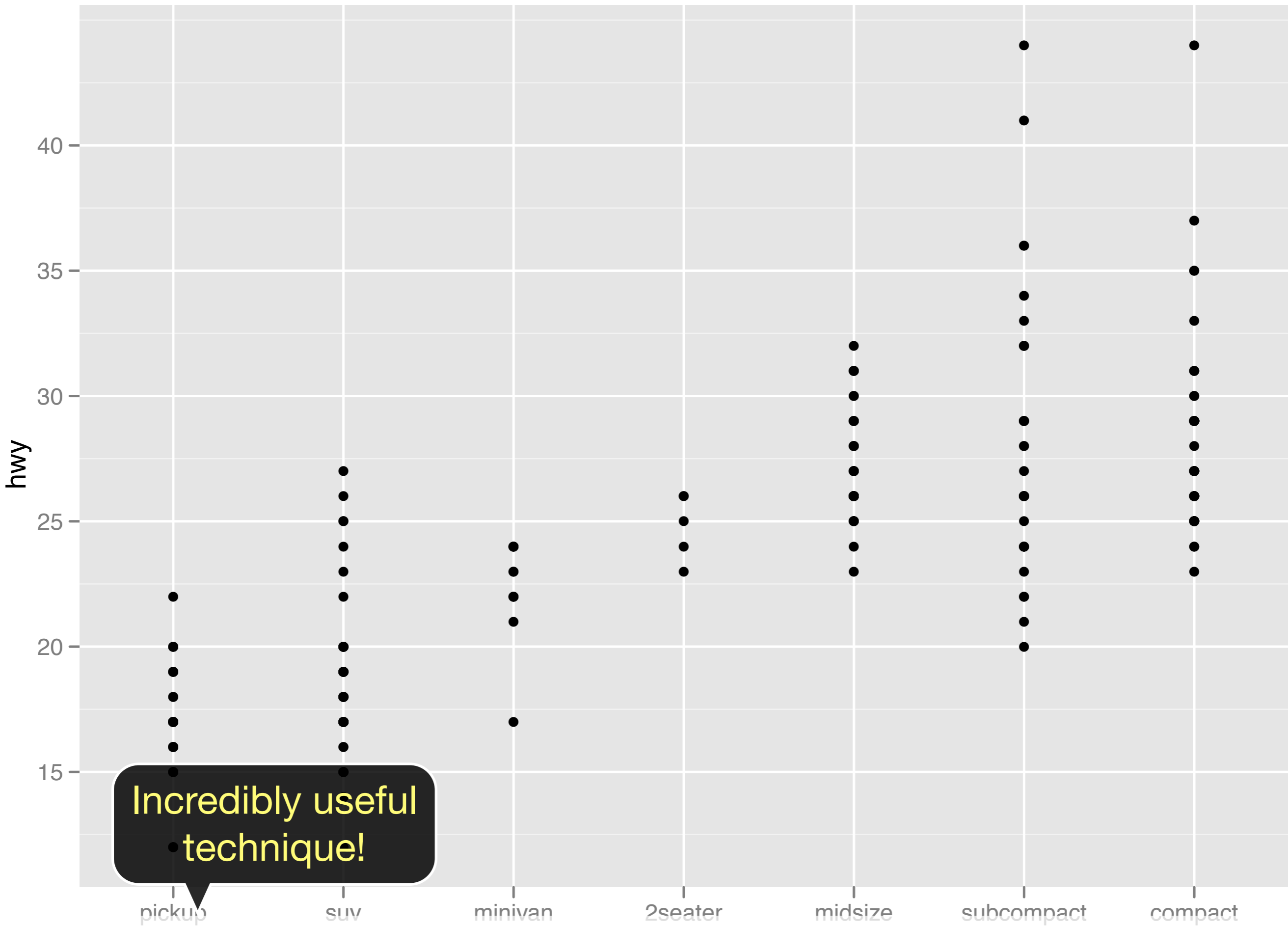
How could we improve this plot?

Brainstorm for 1 minute.



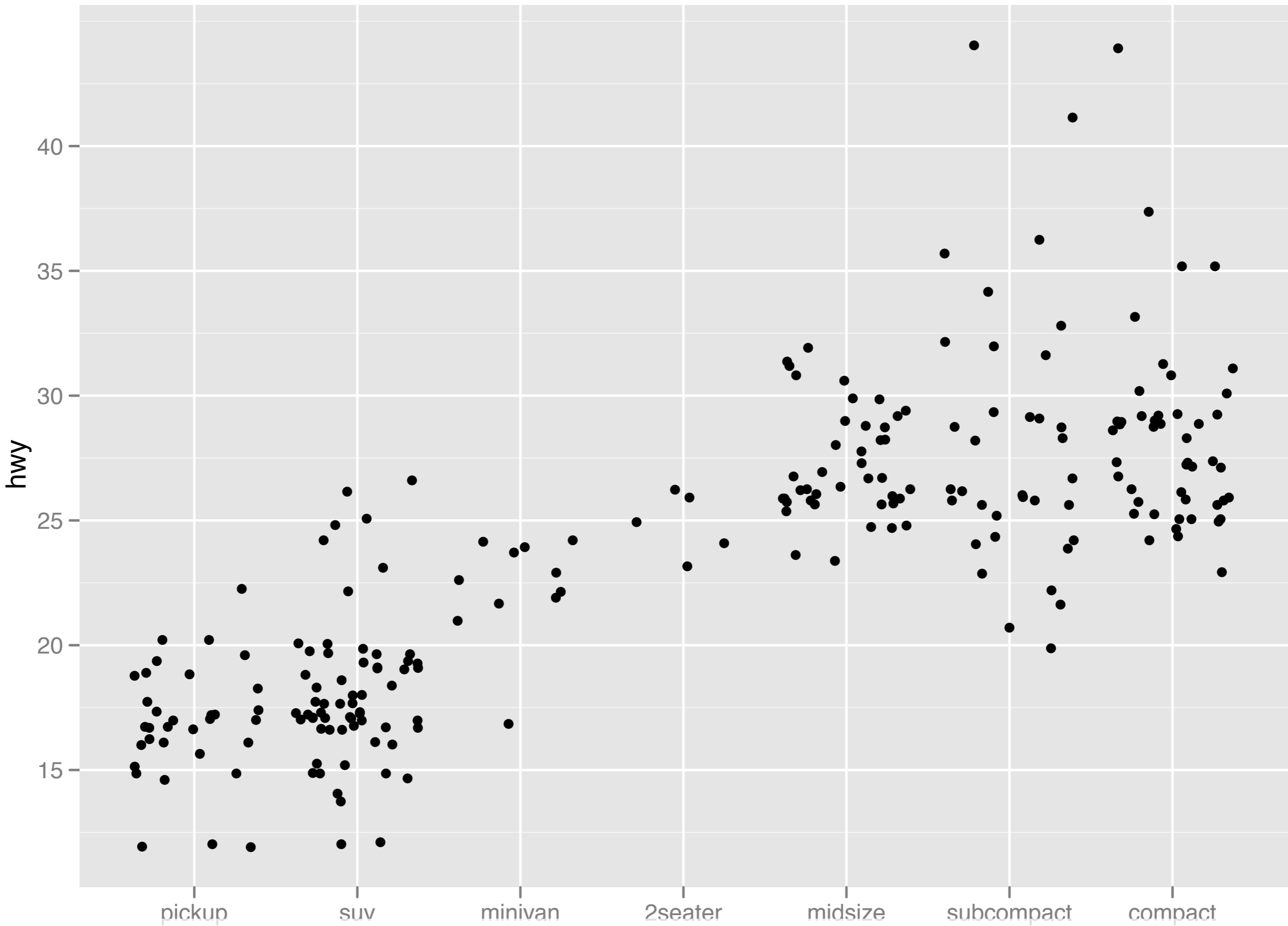
```
qplot(class, hwy, data = mpg)
```



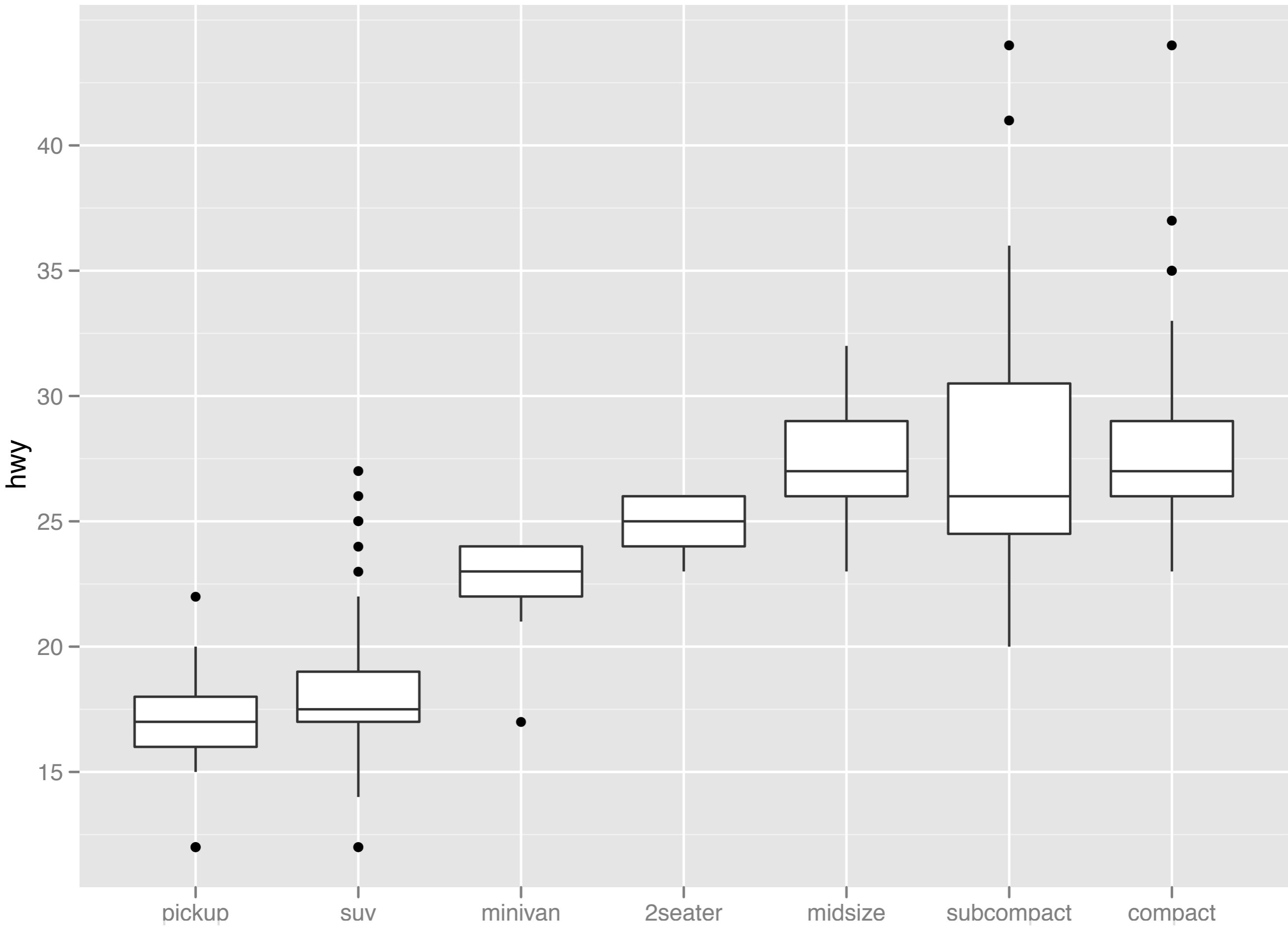


Incredibly useful
• technique!

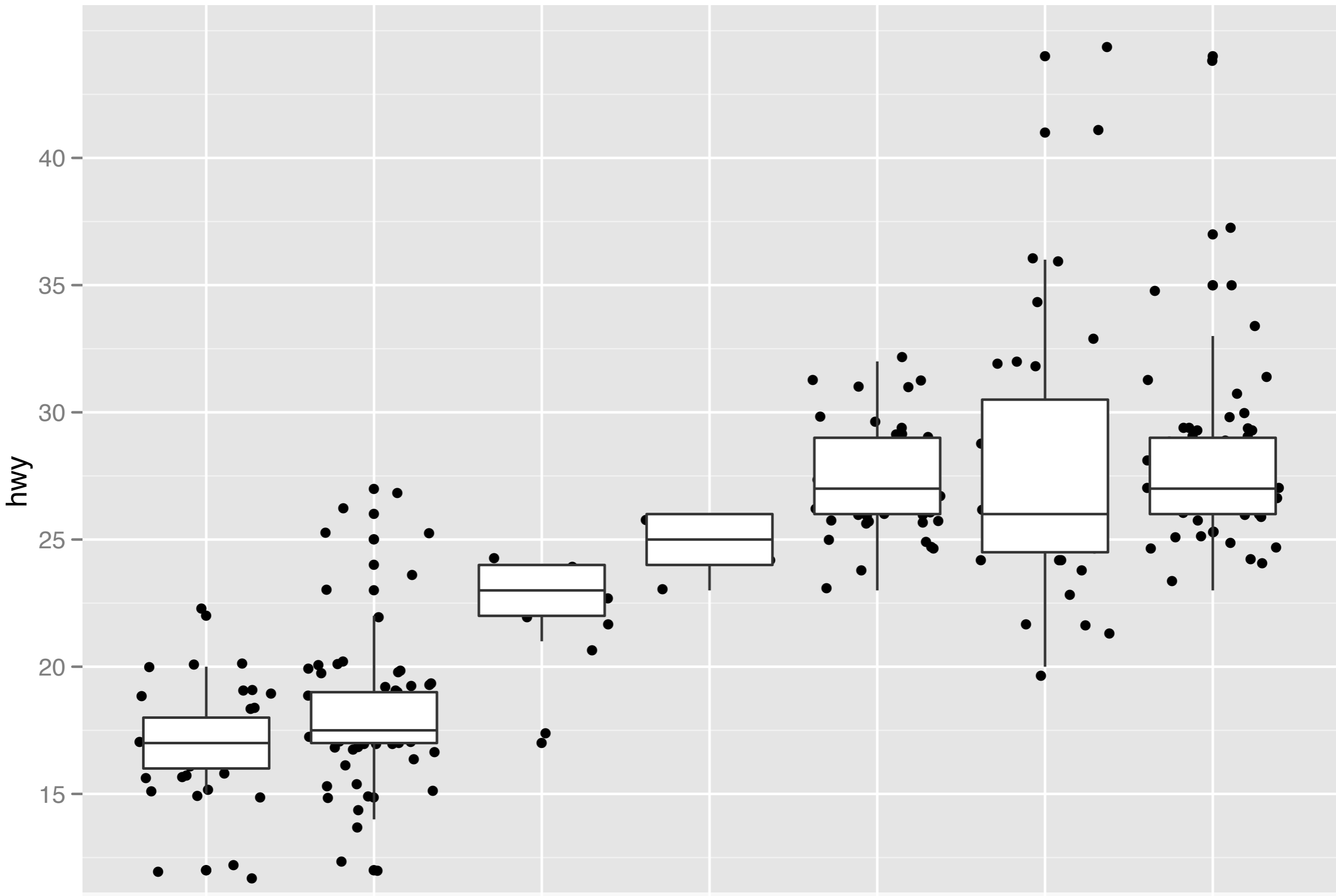
```
qplot(reorder(class, hwy), hwy, data = mpg)
```

```
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(reorder(class, hwy))"))
```

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)

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.