

# Introduction

Hadley Wickham

**October 2009**



<http://had.co.nz/umsl>

1. Preview of today & tomorrow

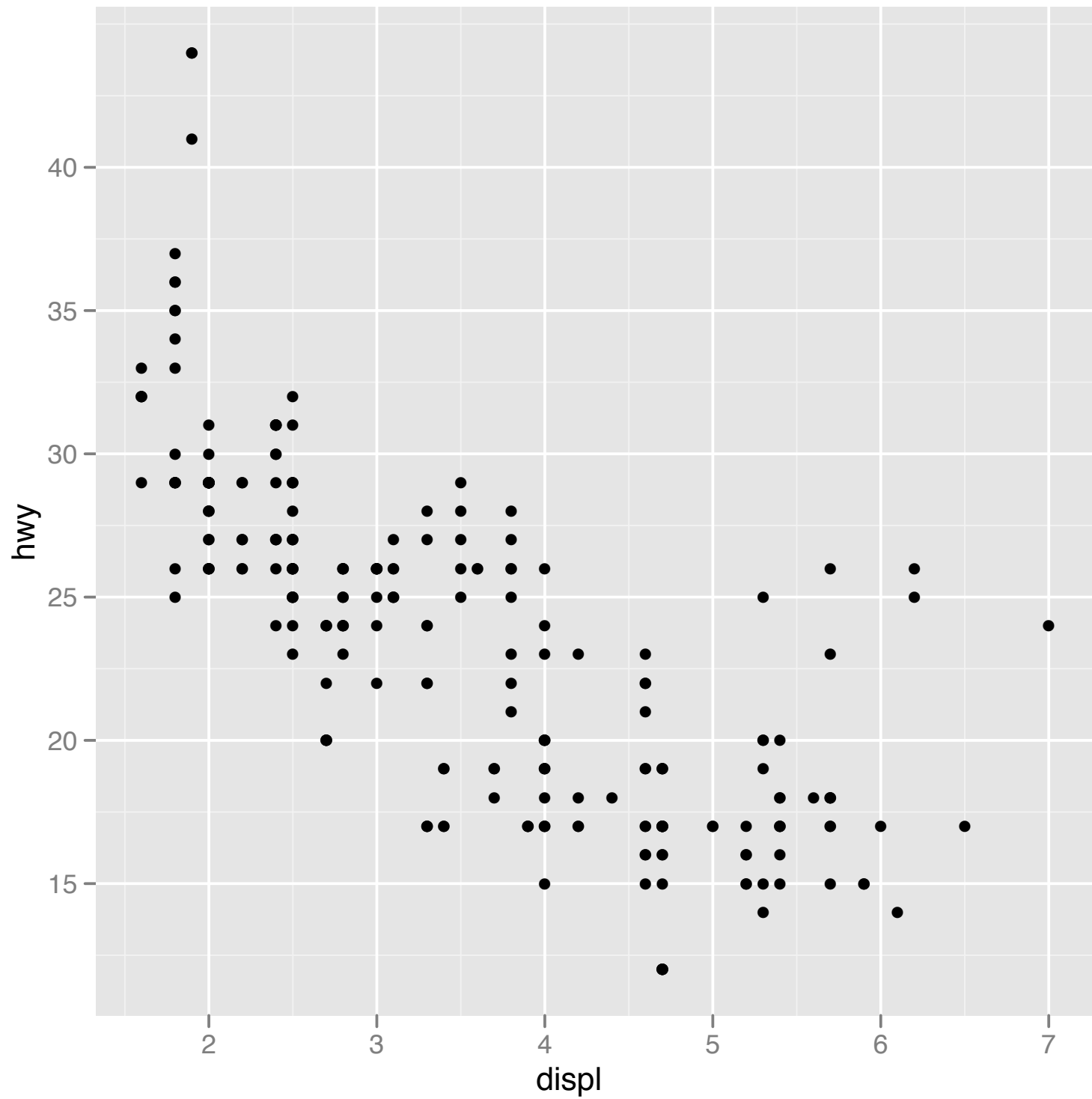
2. About ggplot2

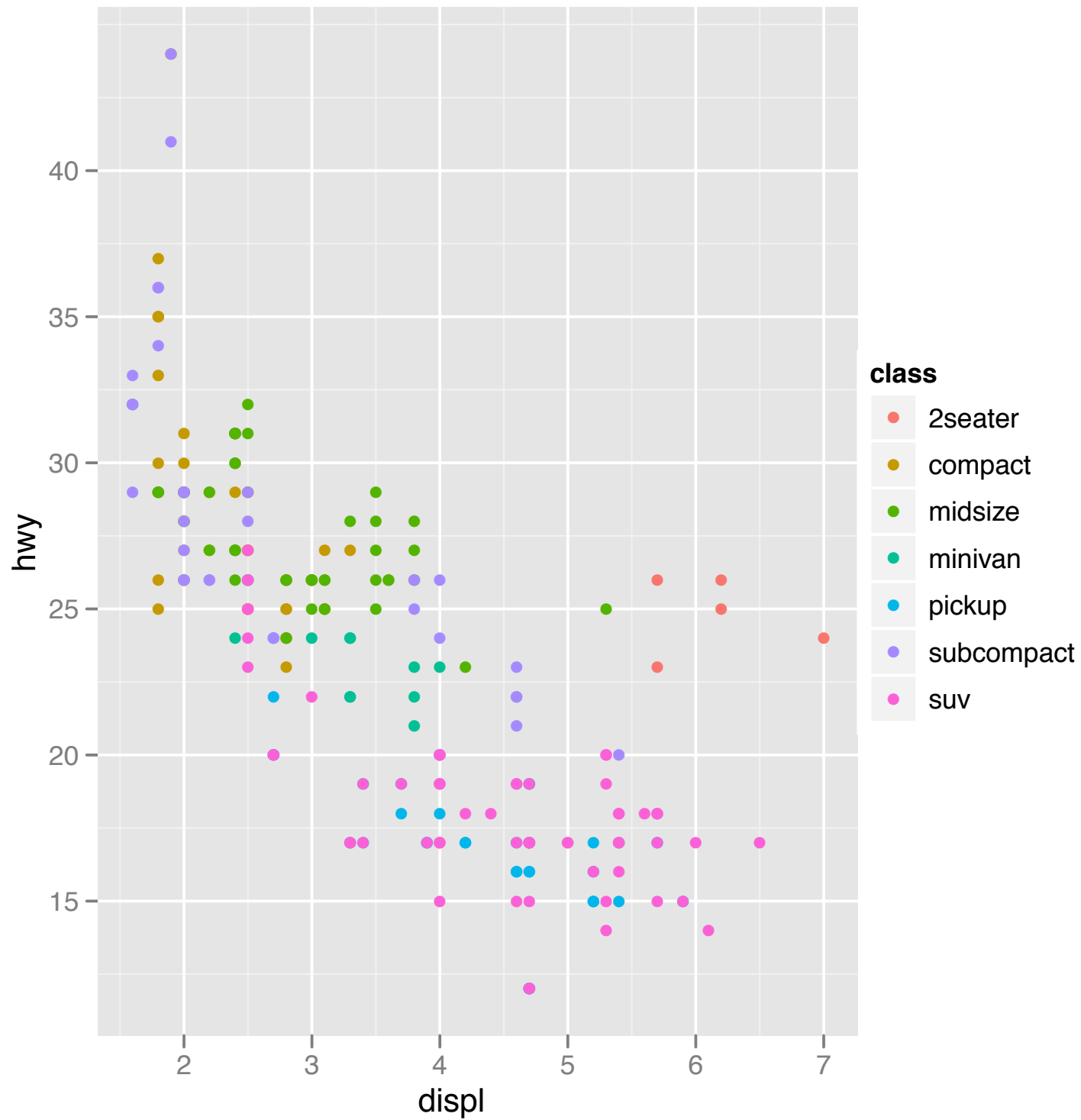
3. More resources

4. Diving in

# **Fuel economy**

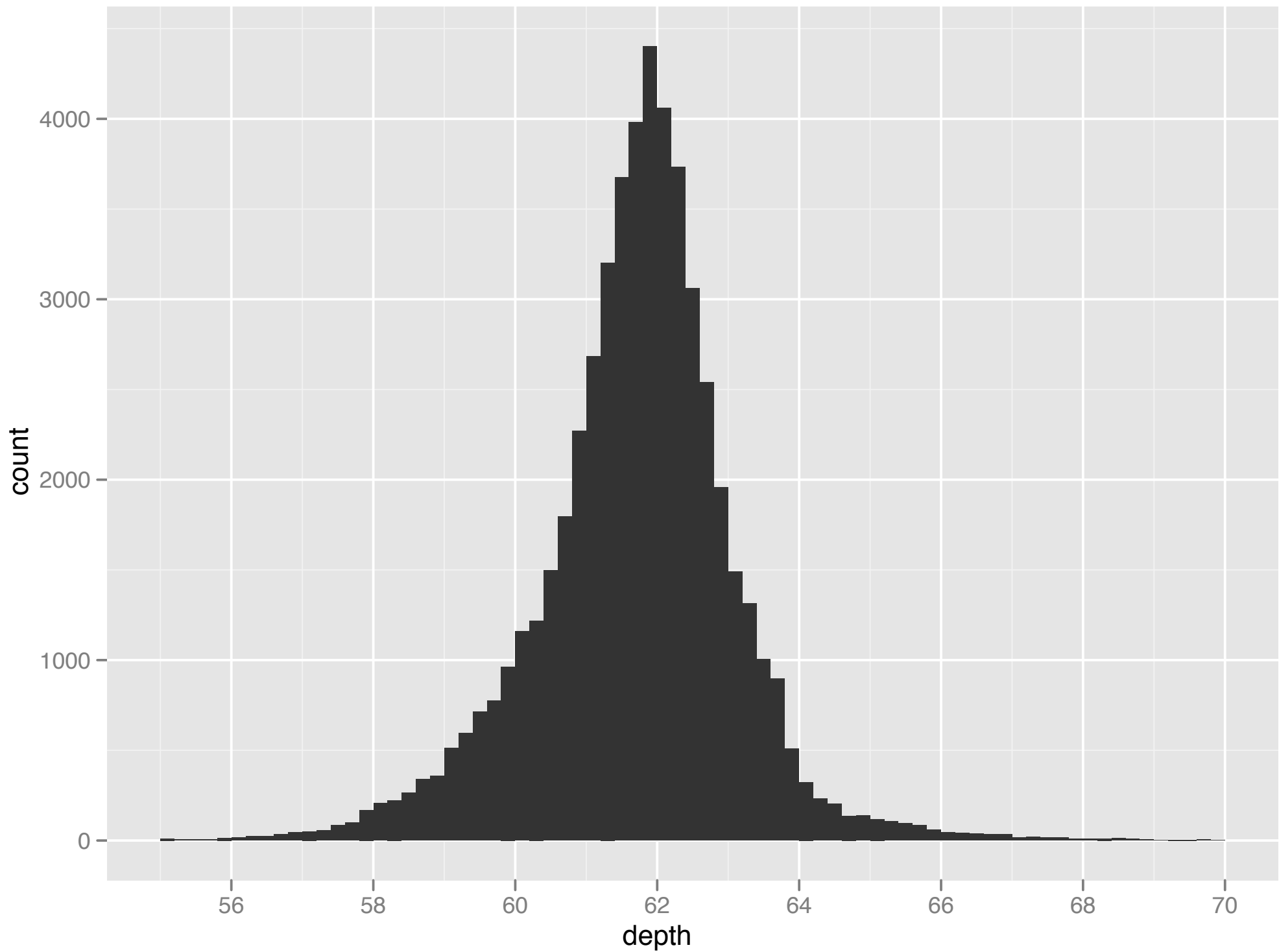
## **Basic graphics**



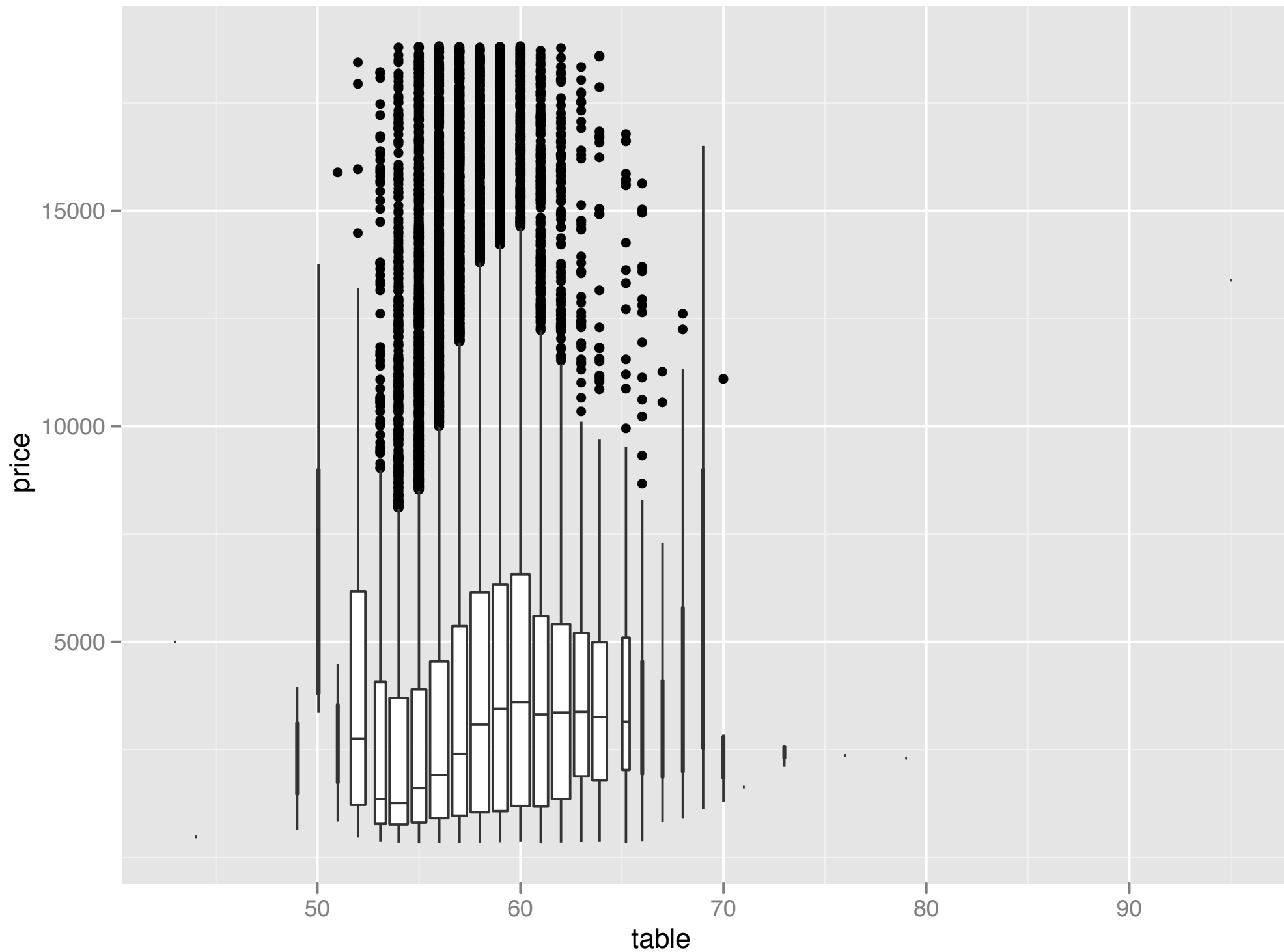


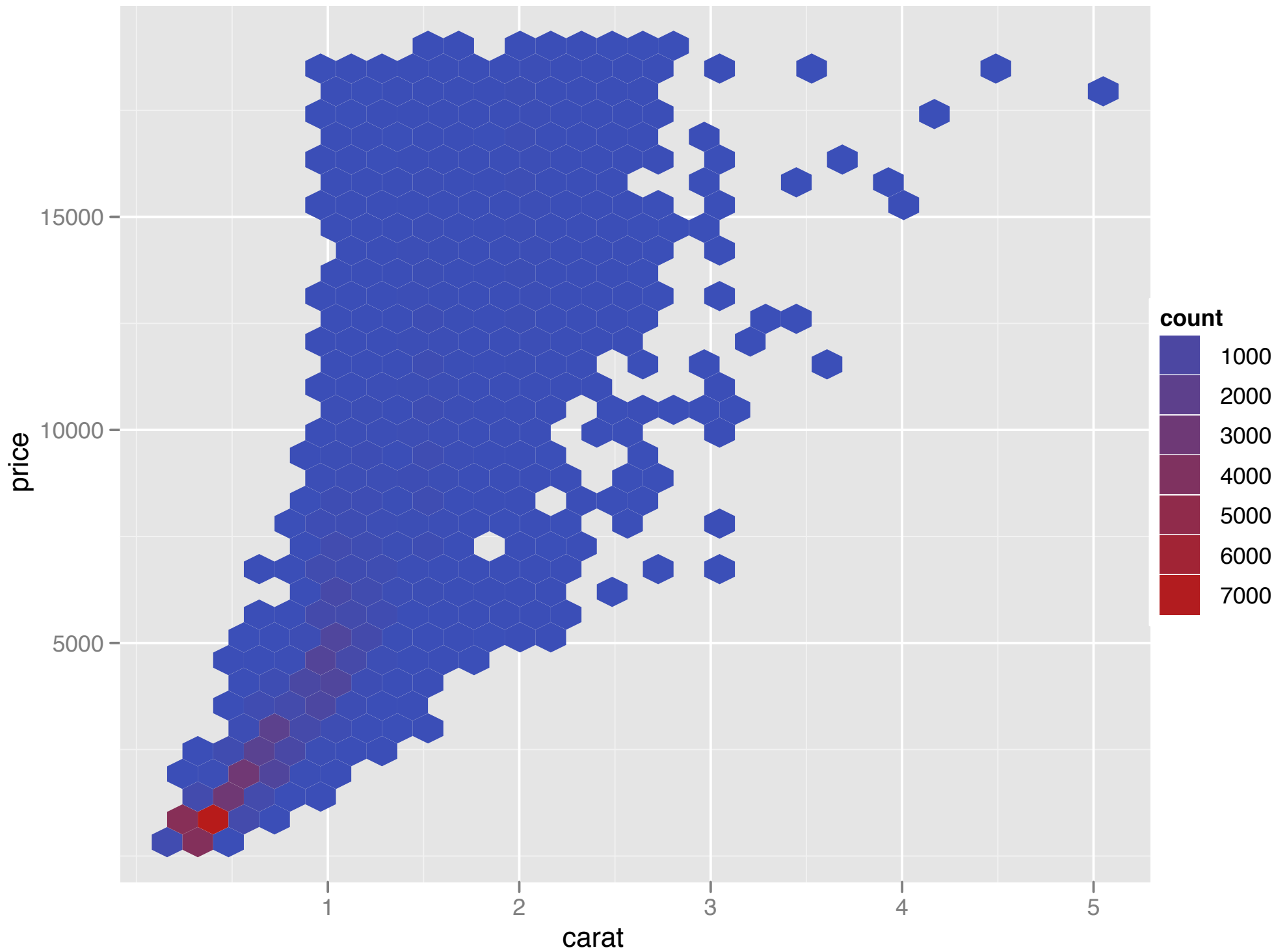
# **Diamond prices**

## Displaying large data





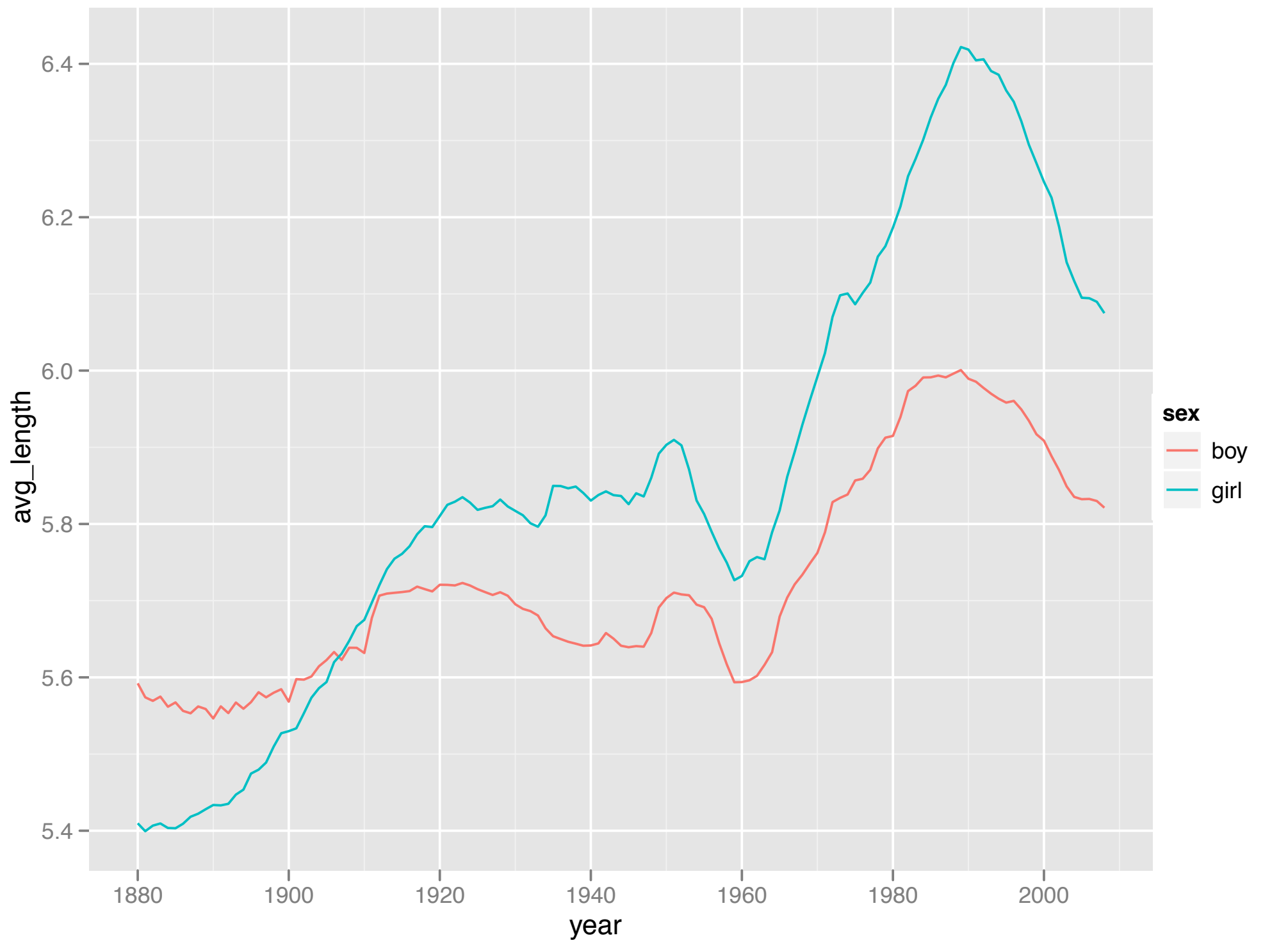


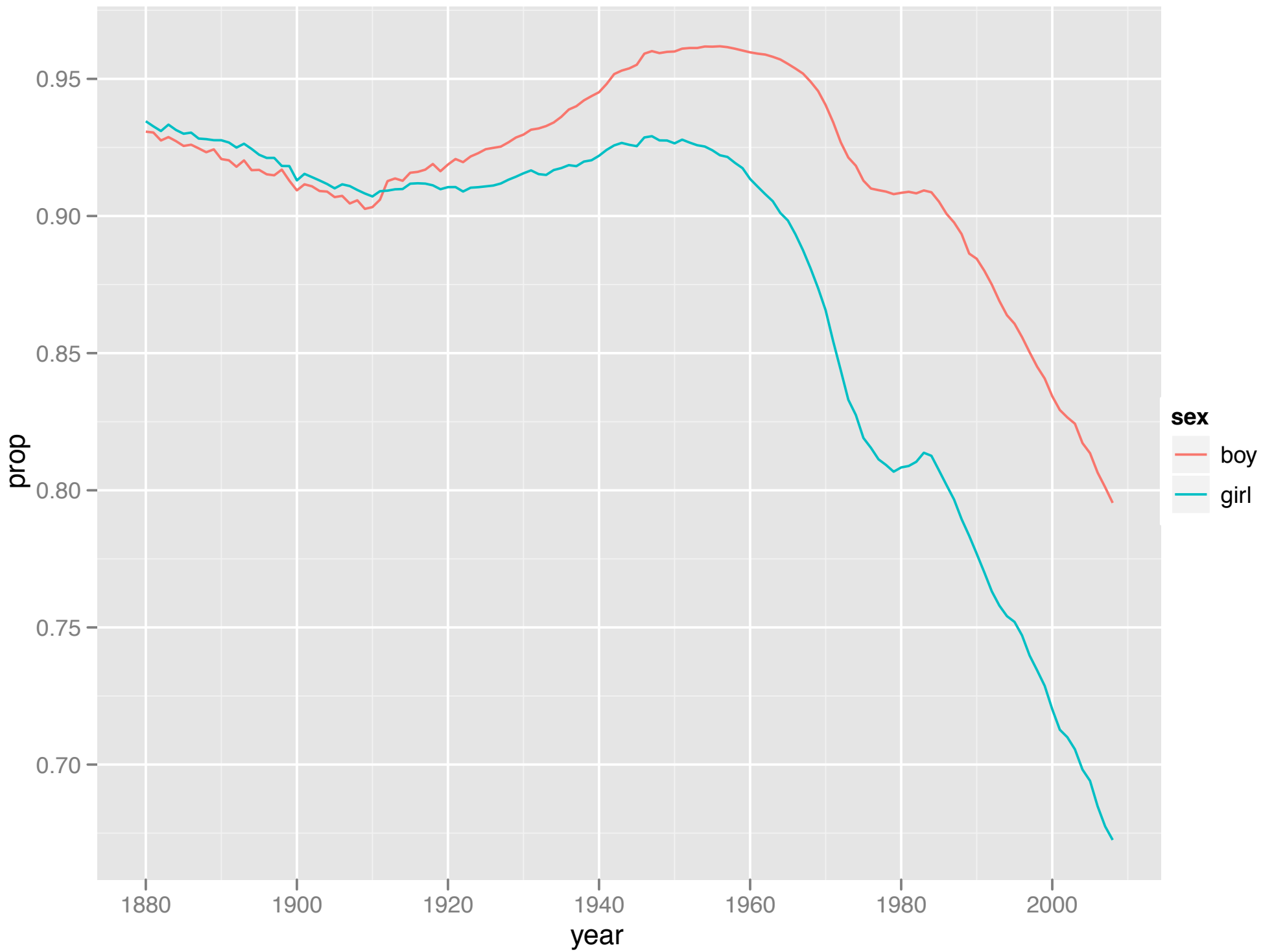


US baby names

**US baby names**

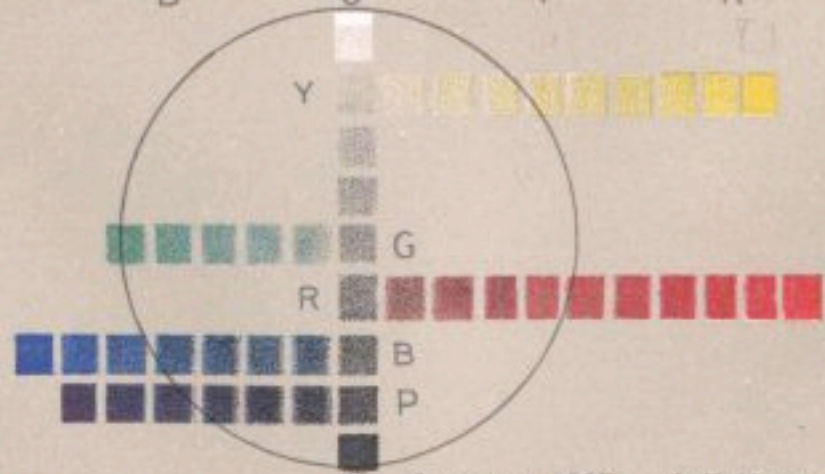
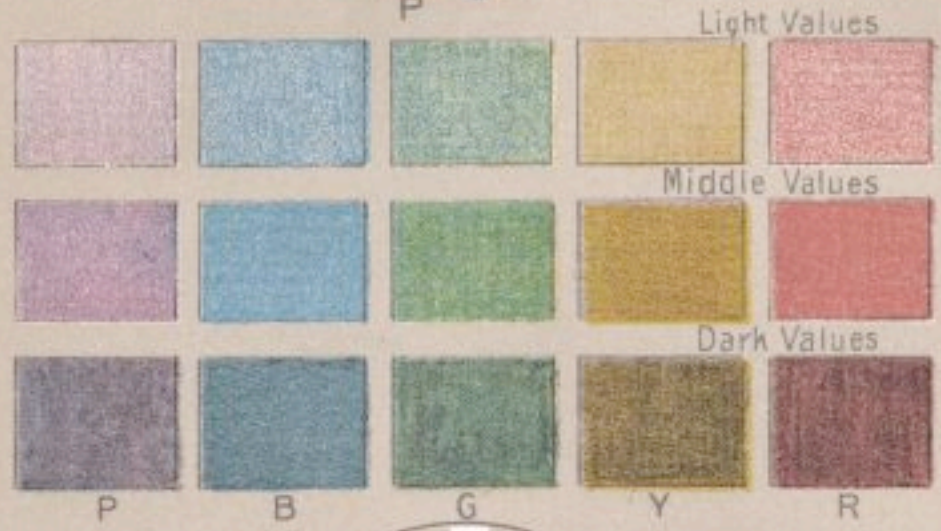
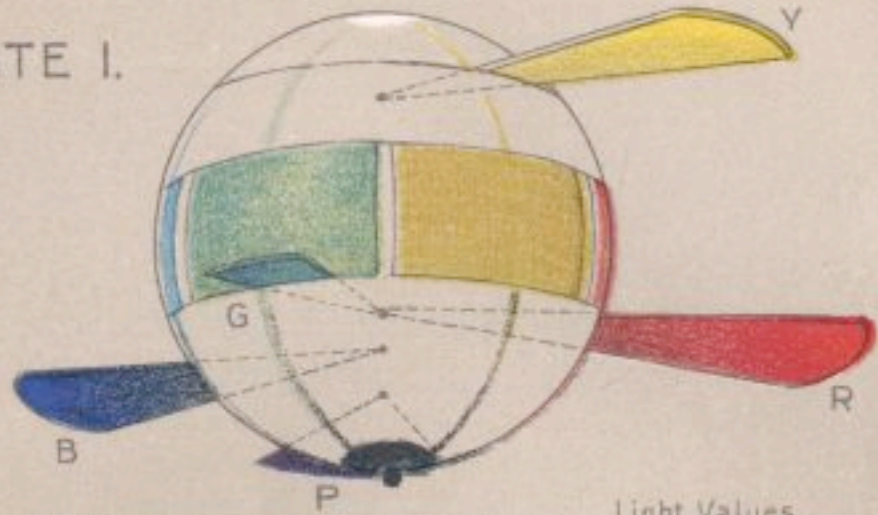
Transforming data





# Polishing your plots

PLATE I.



1. Saving your work

2. Labels & ticks

3. Themes



**Tomorrow**

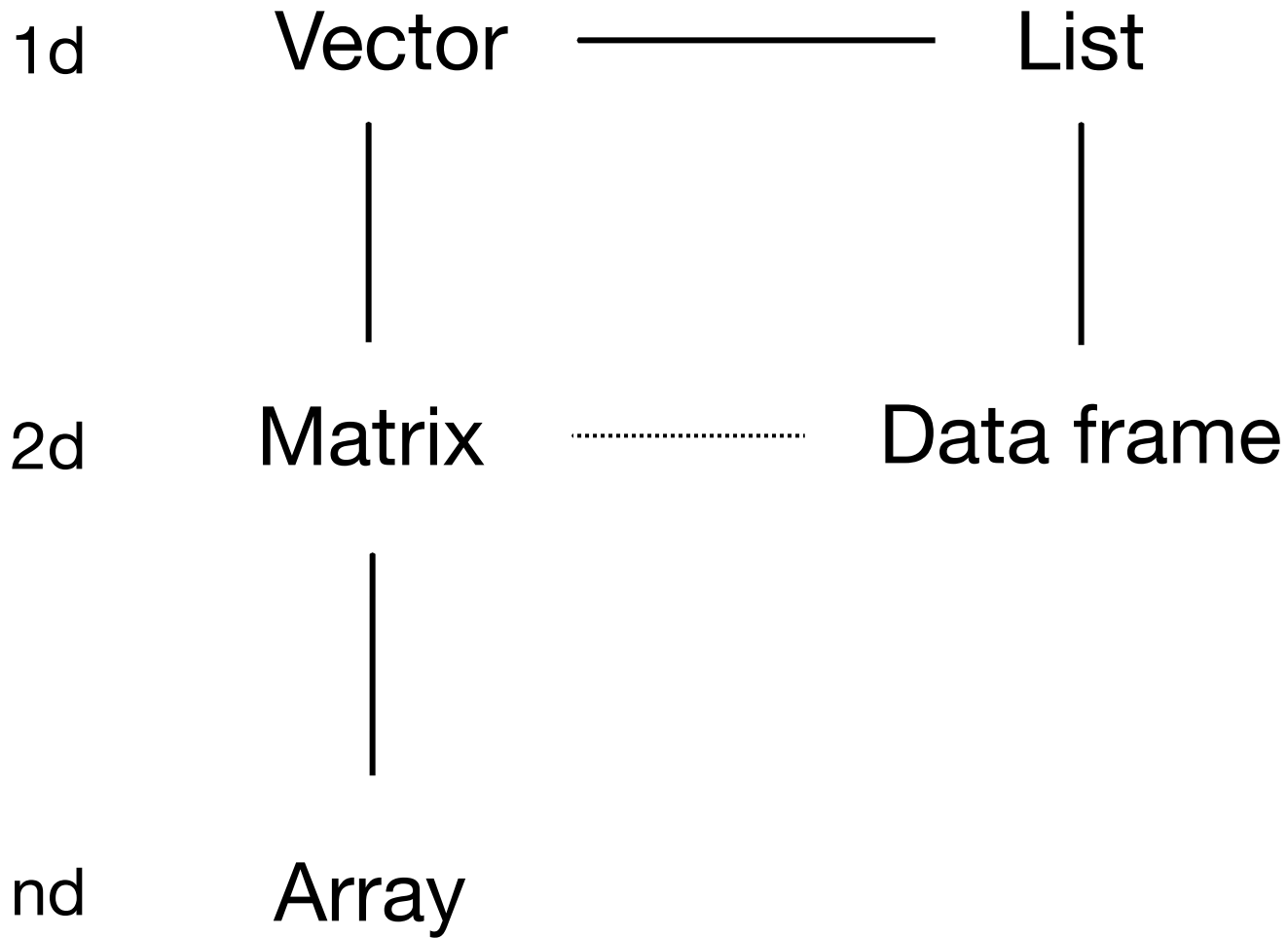
# **Subsetting & data structures**

**blank** include all

**integer** **+ve:** include  
**-ve:** exclude

**logical** include TRUEs

**character** lookup by name



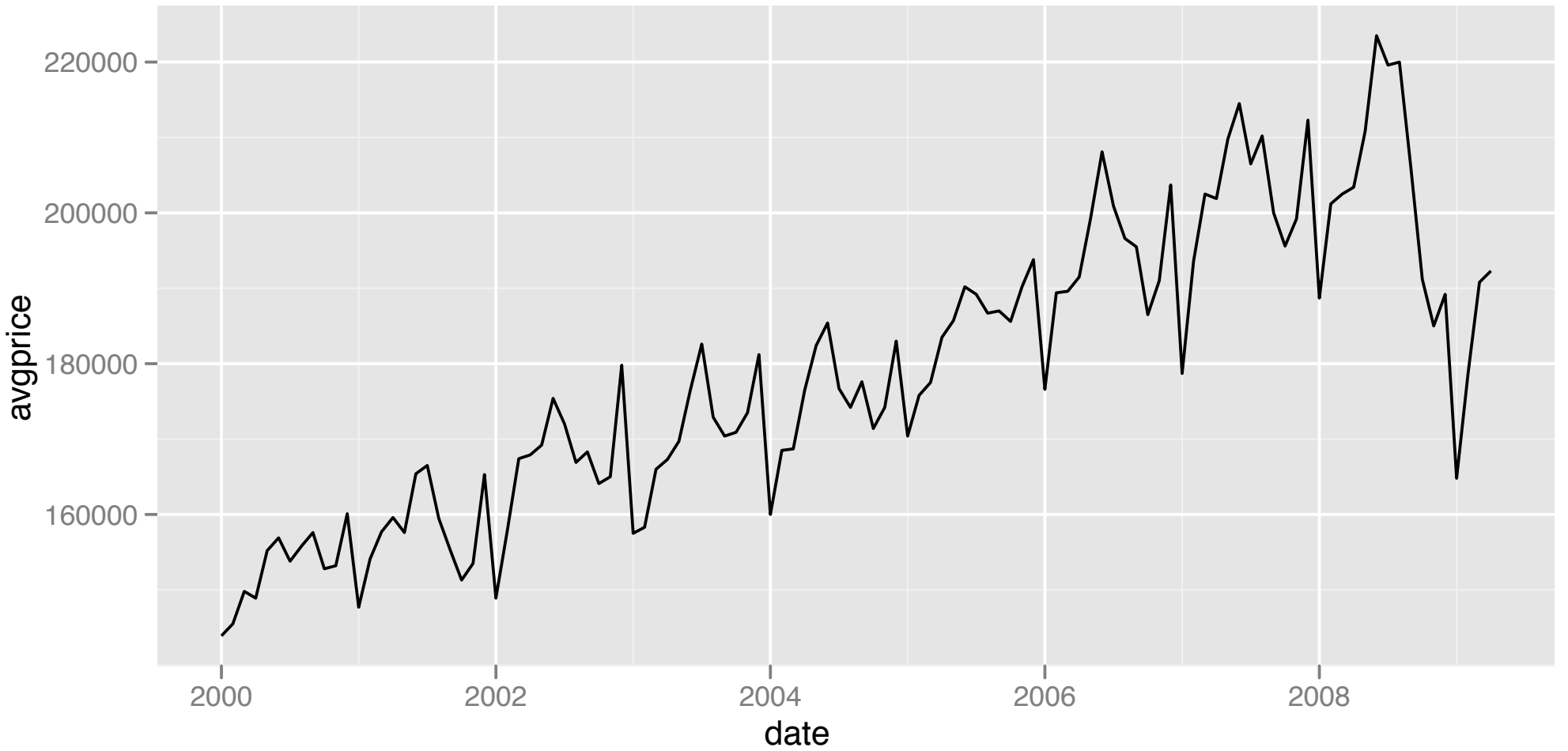
Same types

Different types

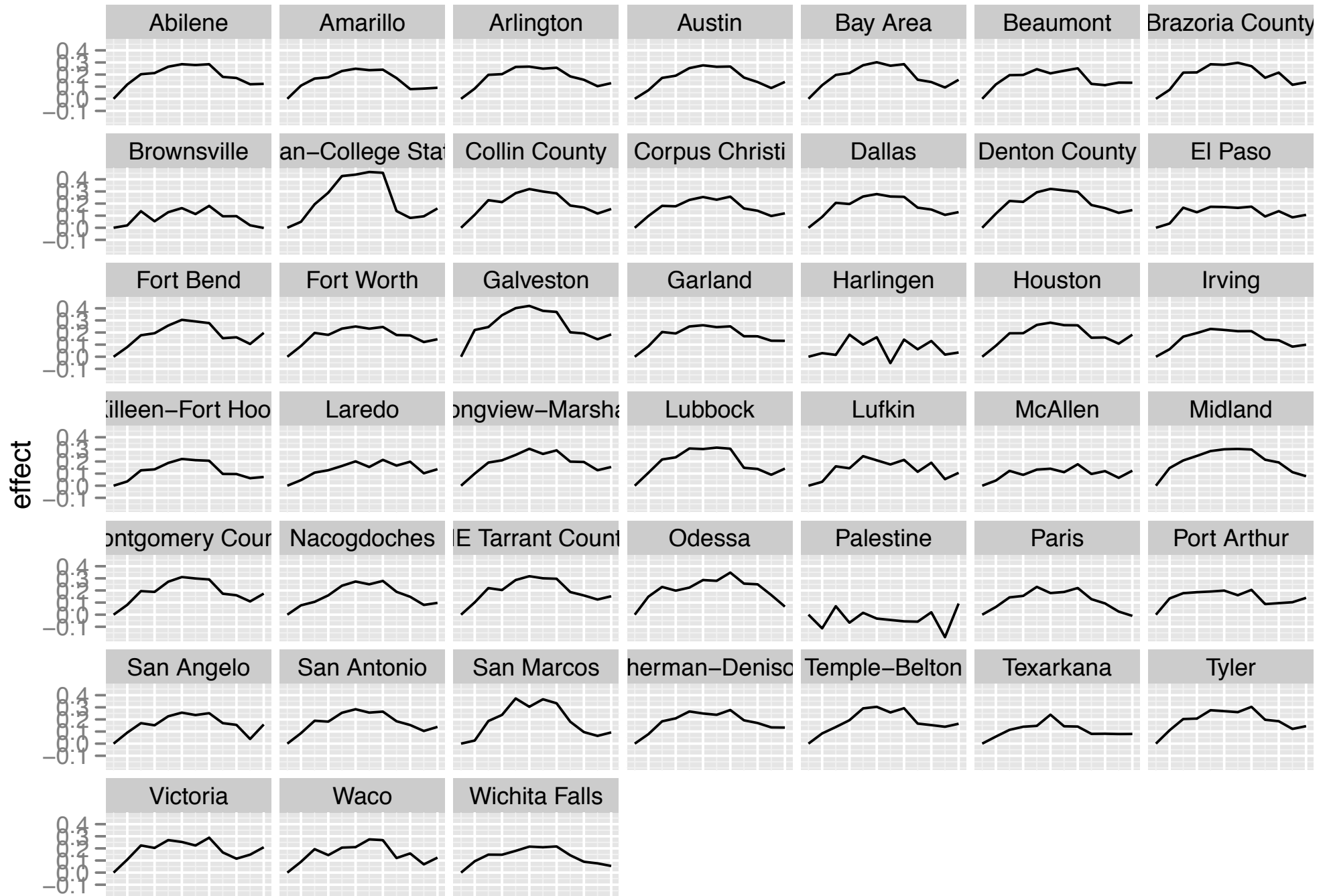
US body names

# **Butterfly communities & Texas house sales Model visualisation**





```
qplot(date, avgprice, data = houston, geom = "line")
```



```

qplot(month, effect, data = coef2, geom = "line") +
  facet_wrap(~ city)

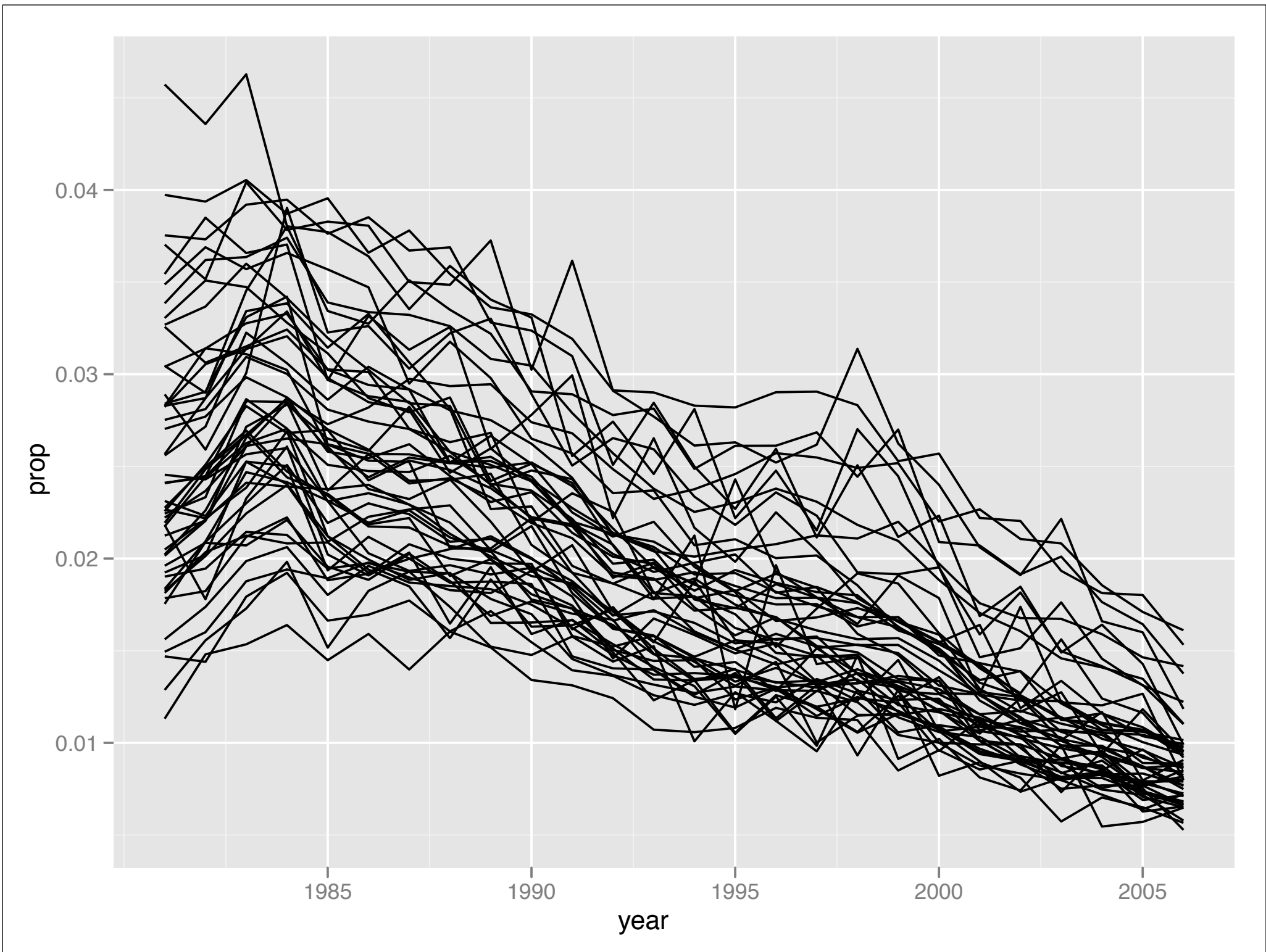
```

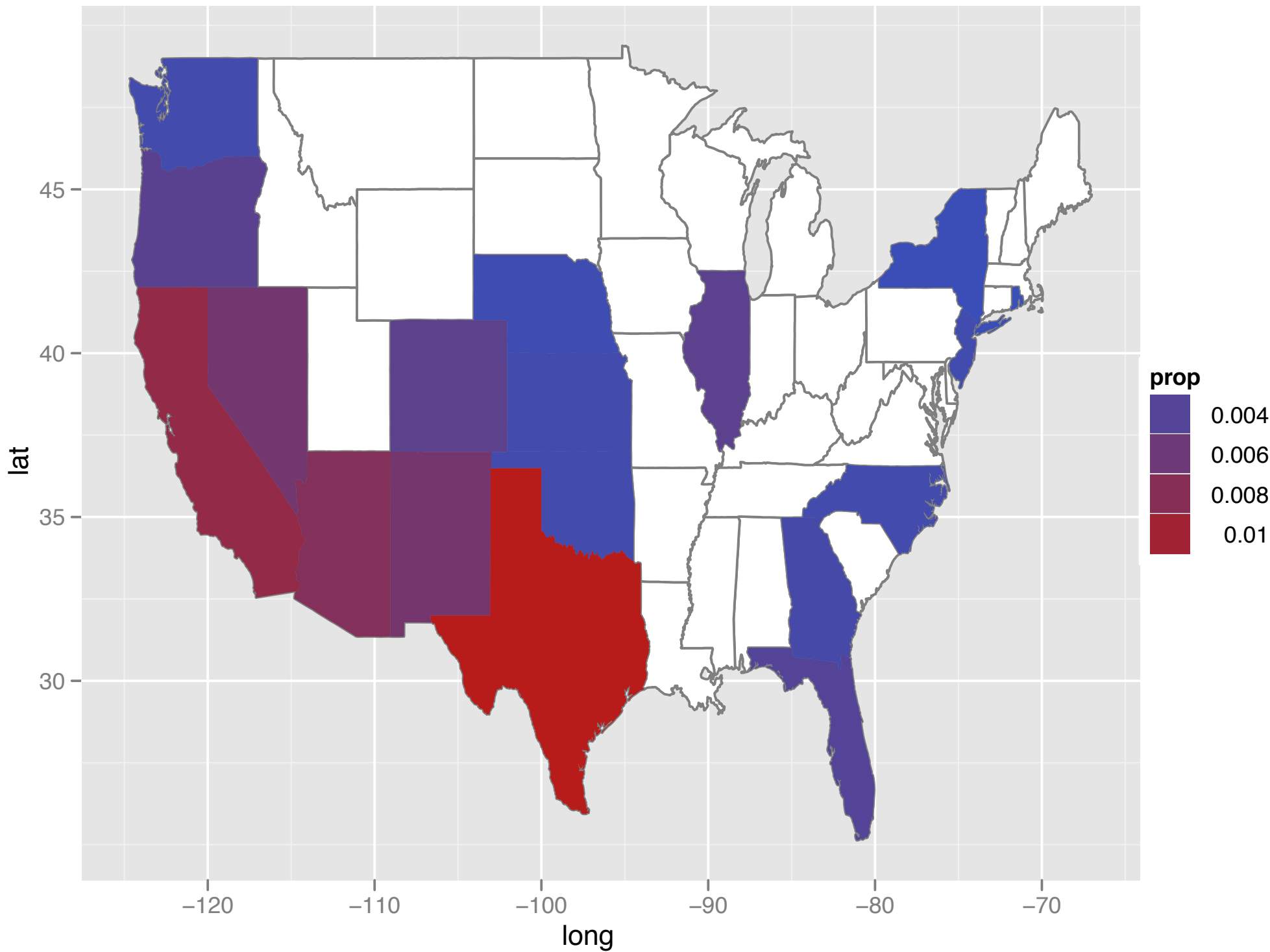


US baby names

# **Baby names by state**

Visualising time & space







US baby names

# ggplot2 theory & graphical critique

# Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Mimard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Léger, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk et Mohilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

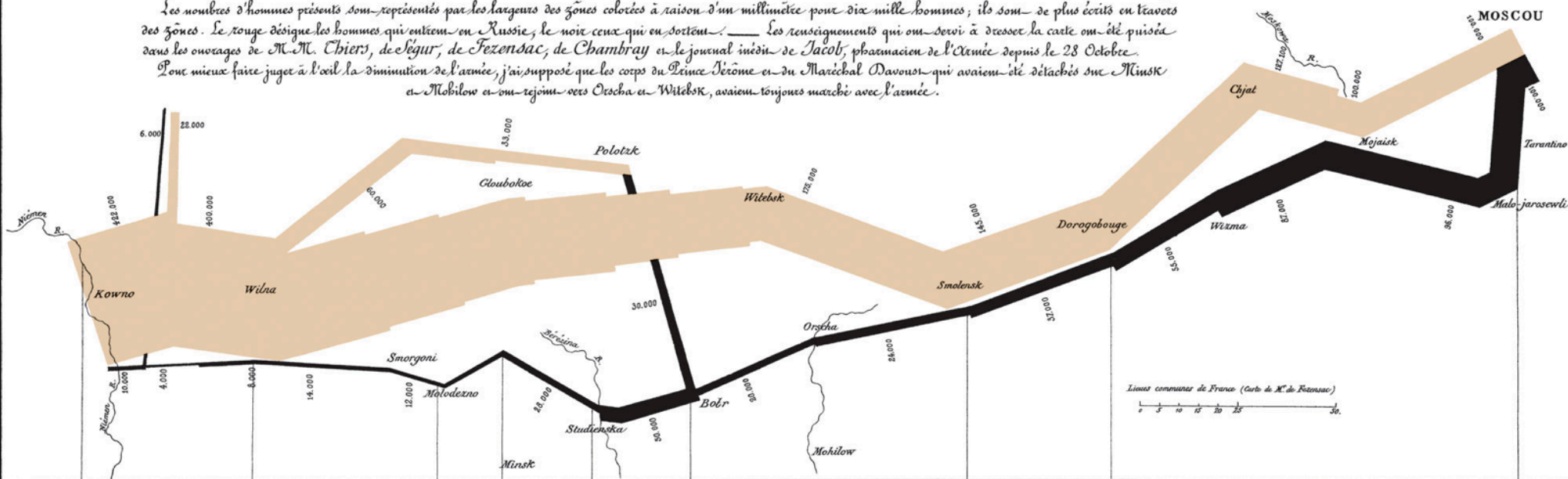
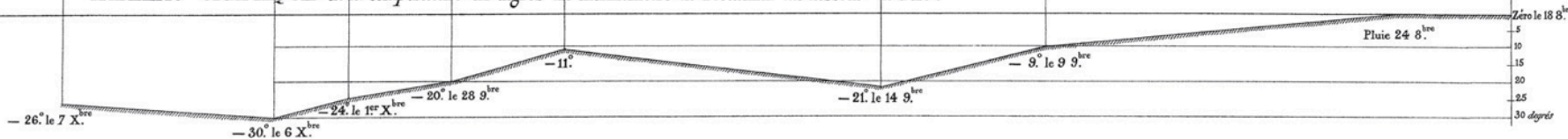


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop le Niémen gelé.



# Content

# Construction







**Context**



**Consumption**

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
```

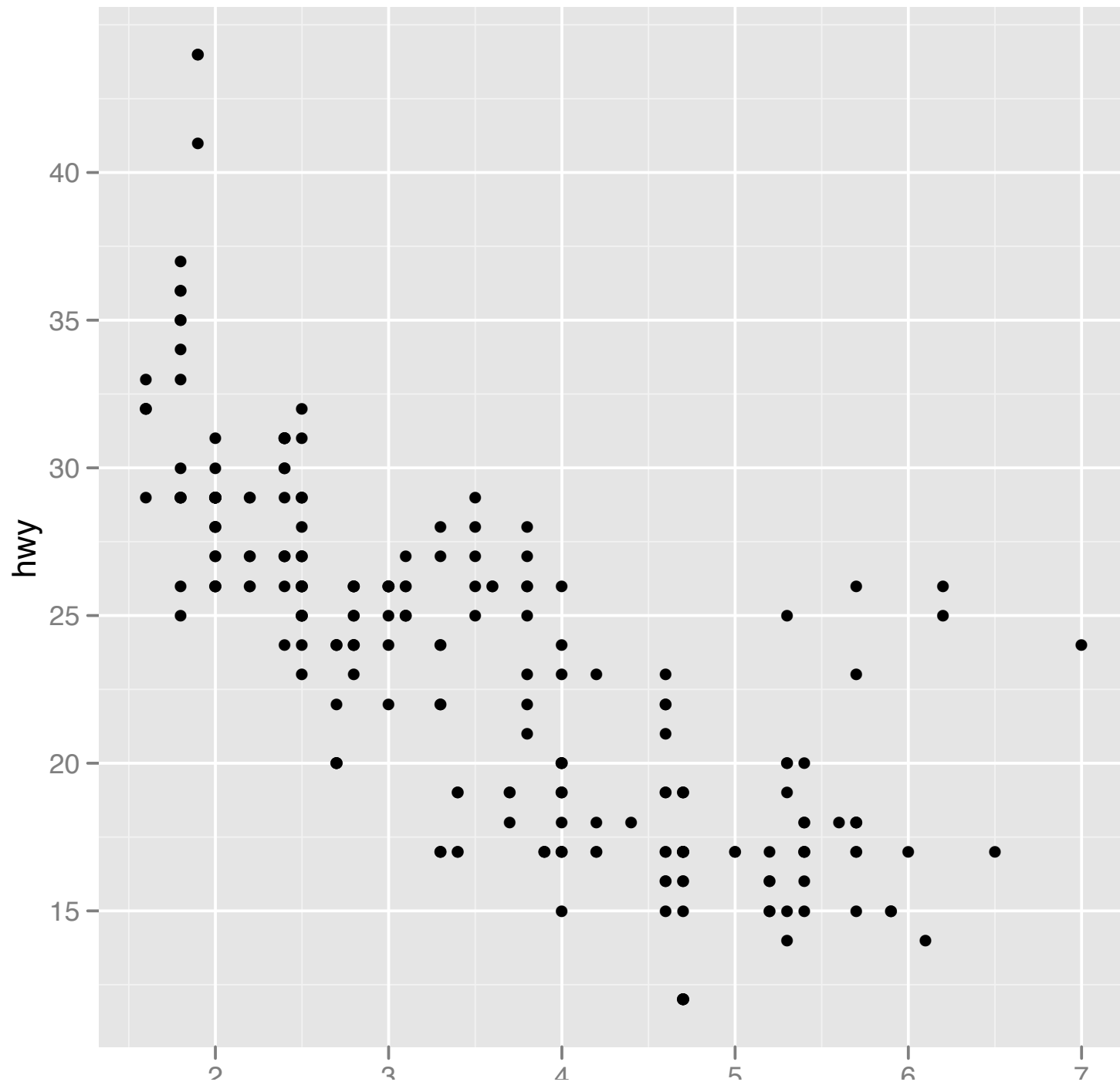
```
head(mpg)
```

```
str(mpg)
```

```
summary(mpg)
```

In ggplot2, we  
always explicitly  
specify the data

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

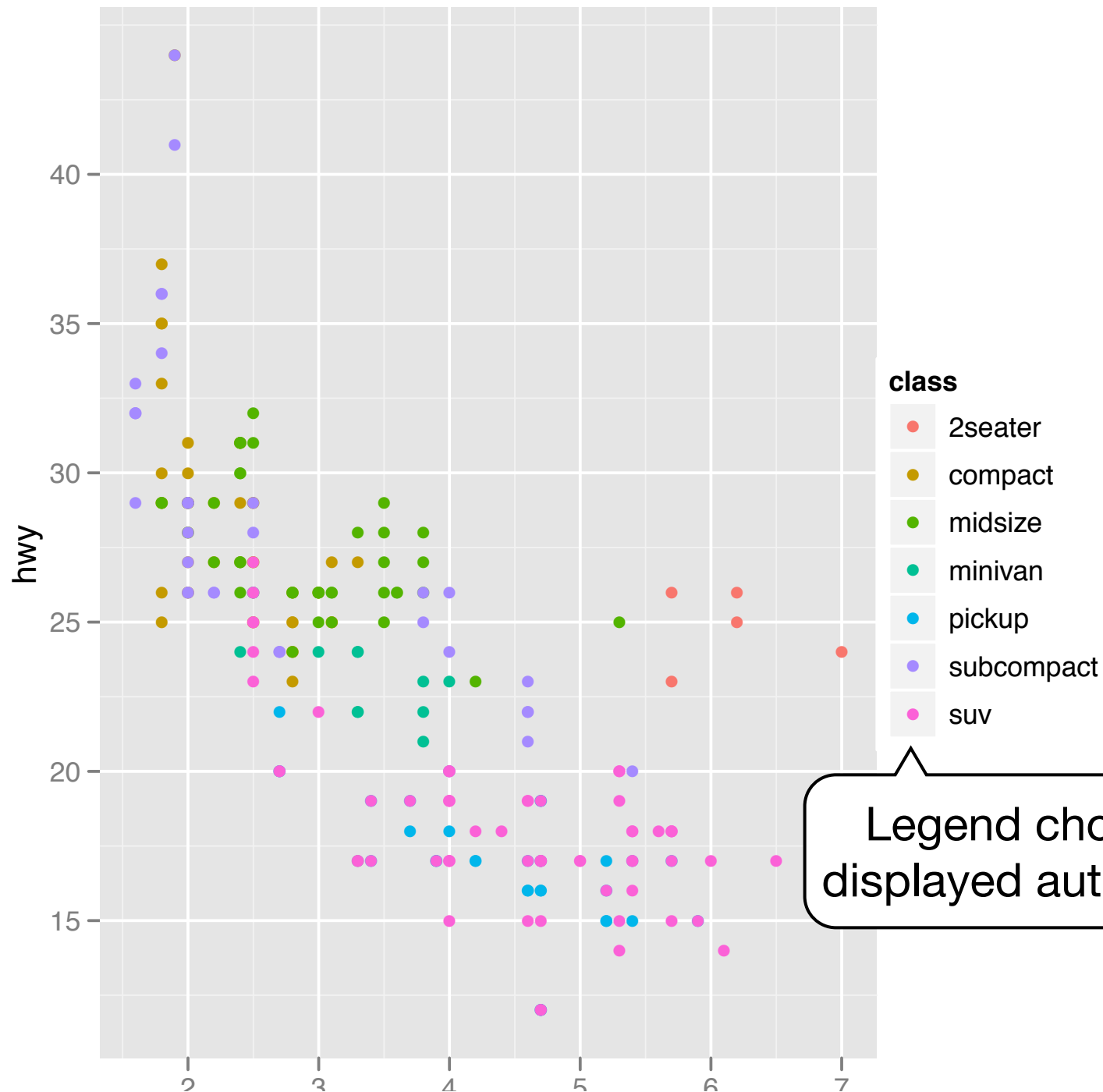


```
qplot(displ, hwy, data = mpg)
```

# 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)
```

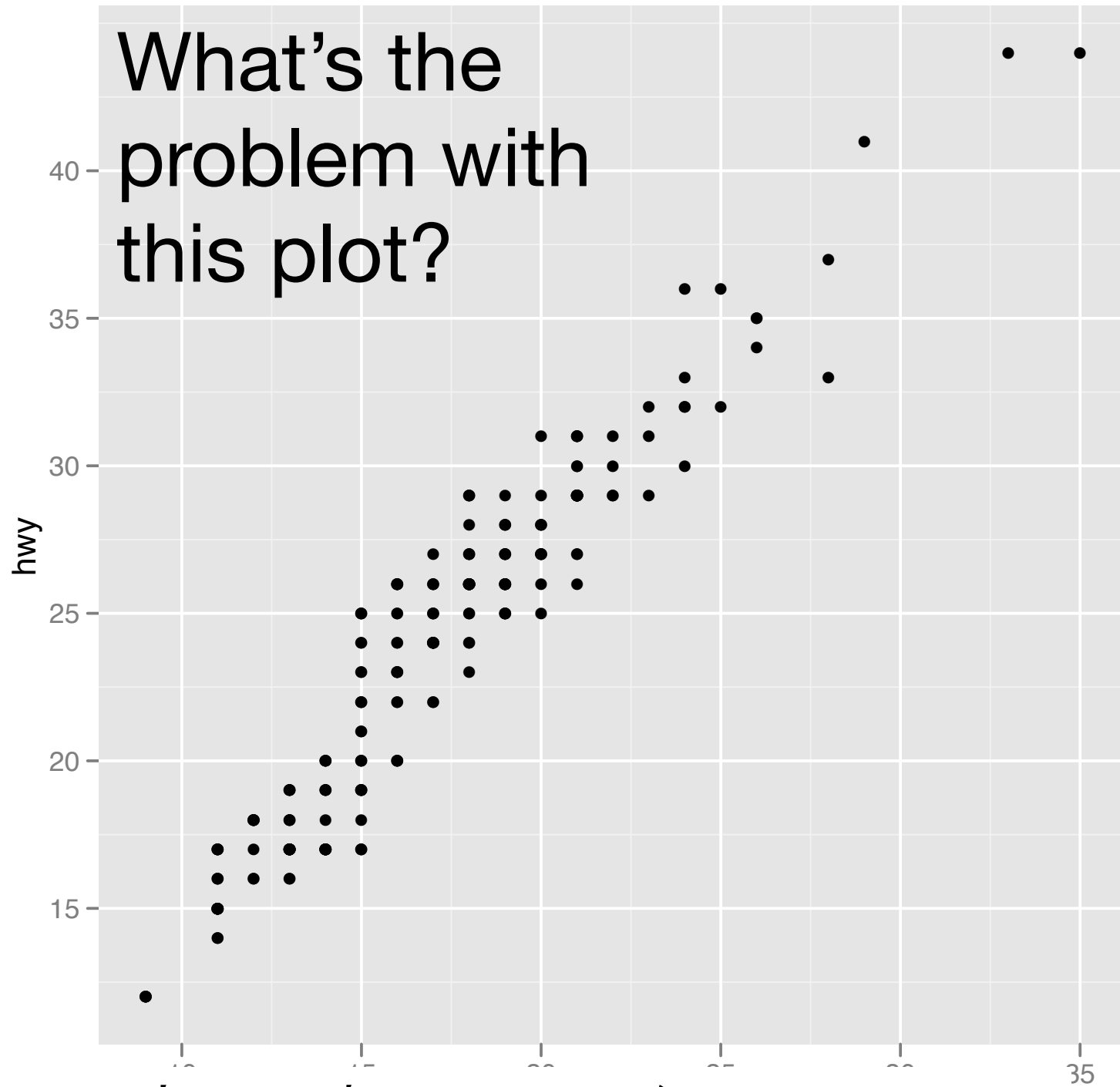
# Summary

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

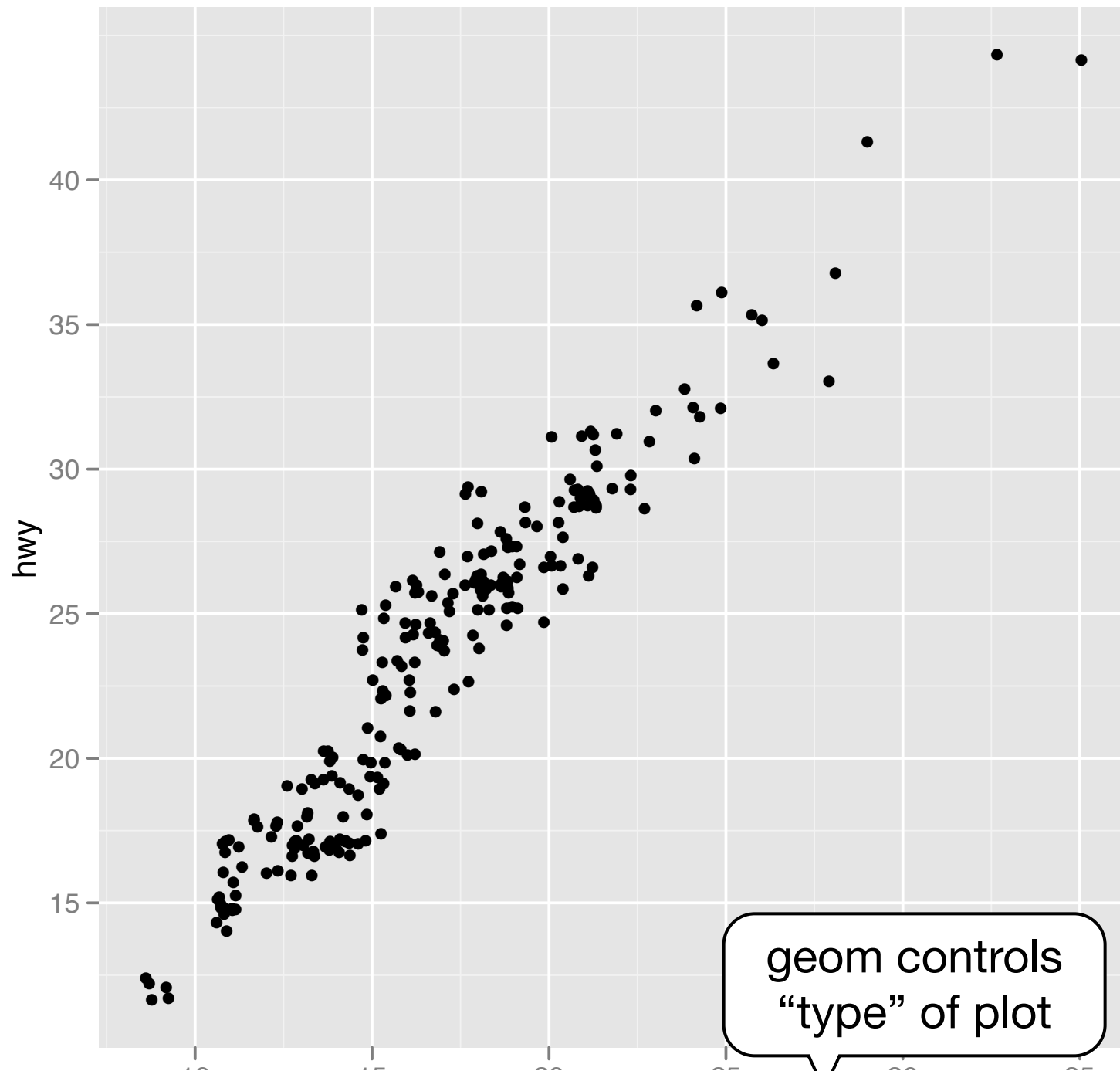
`facet_wrap()`: 1d ribbon wrapped into 2d

Scales argument controls whether position scales are fixed or free.

What's the  
problem with  
this plot?



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



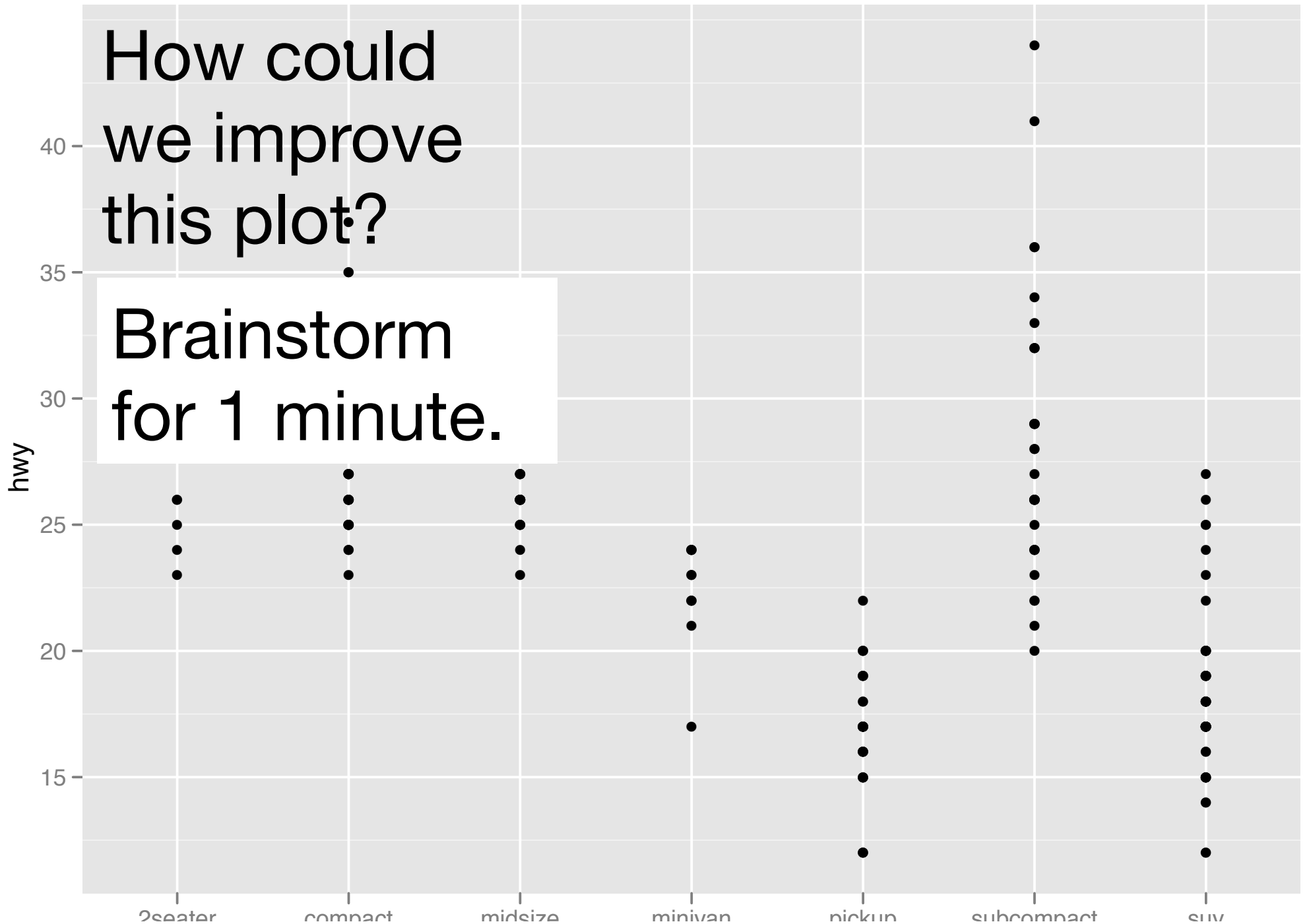
geom controls  
"type" of plot

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

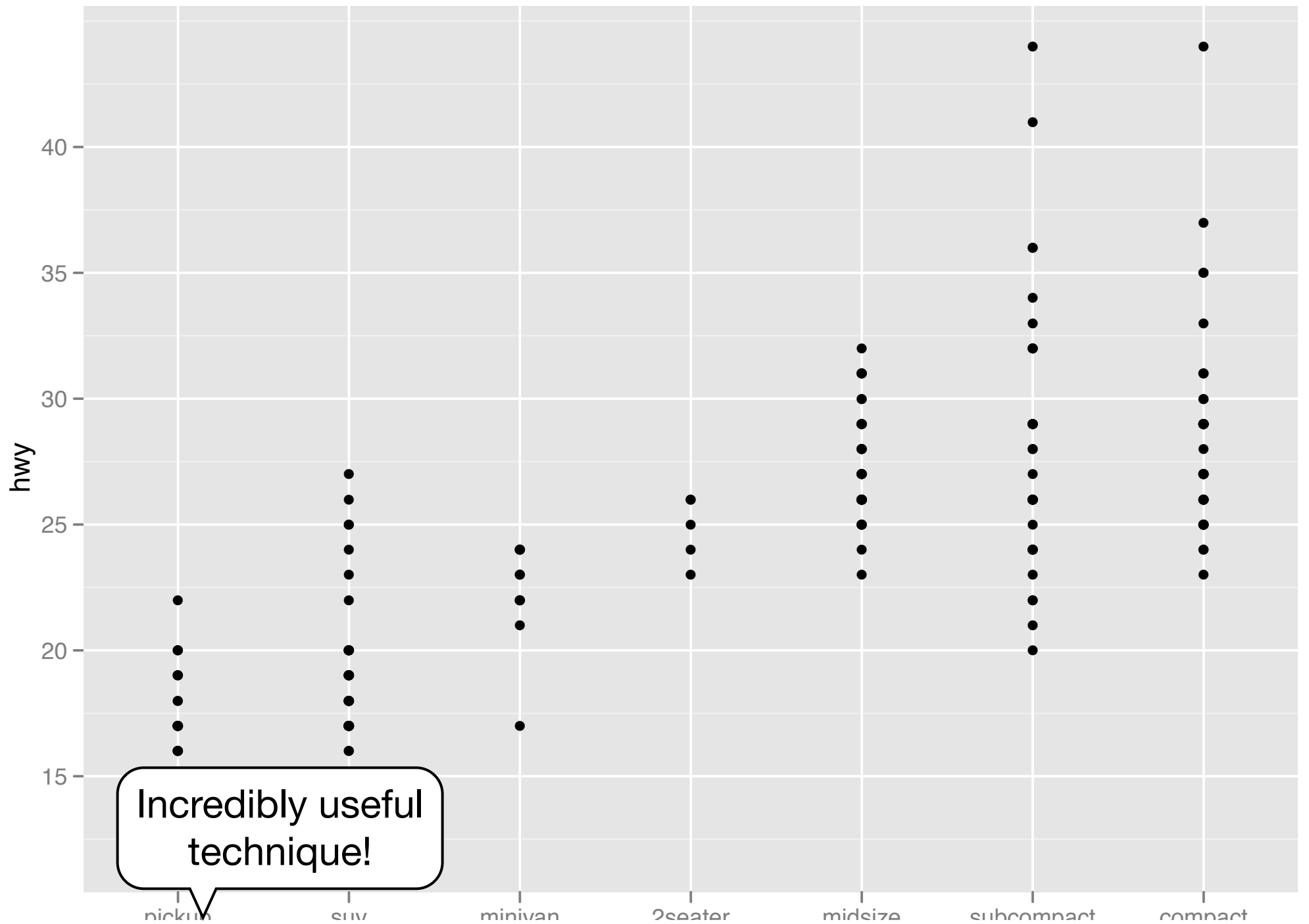


How could  
we improve  
this plot?

Brainstorm  
for 1 minute.

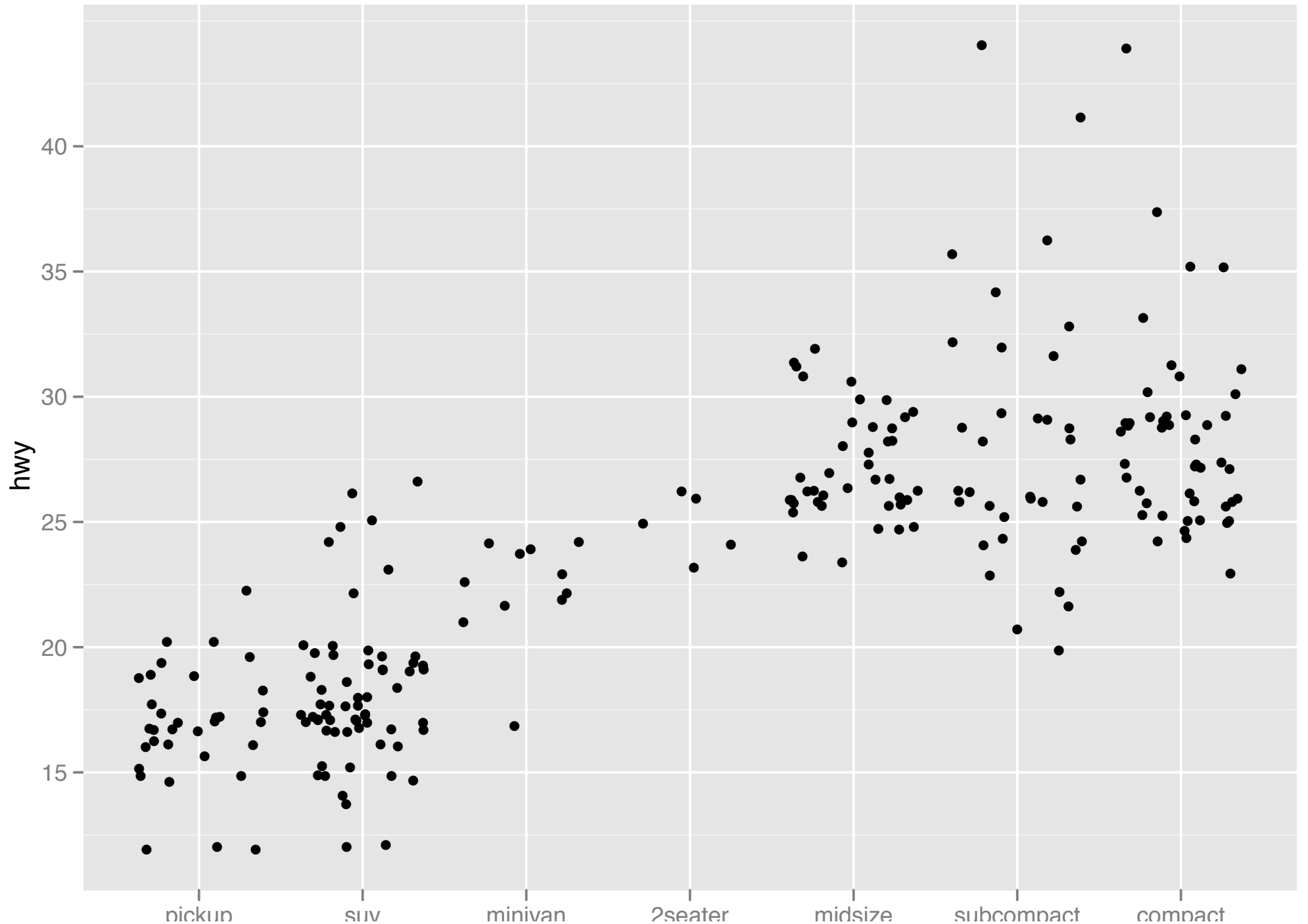


```
qplot(class, hwy, data = mpg)
```

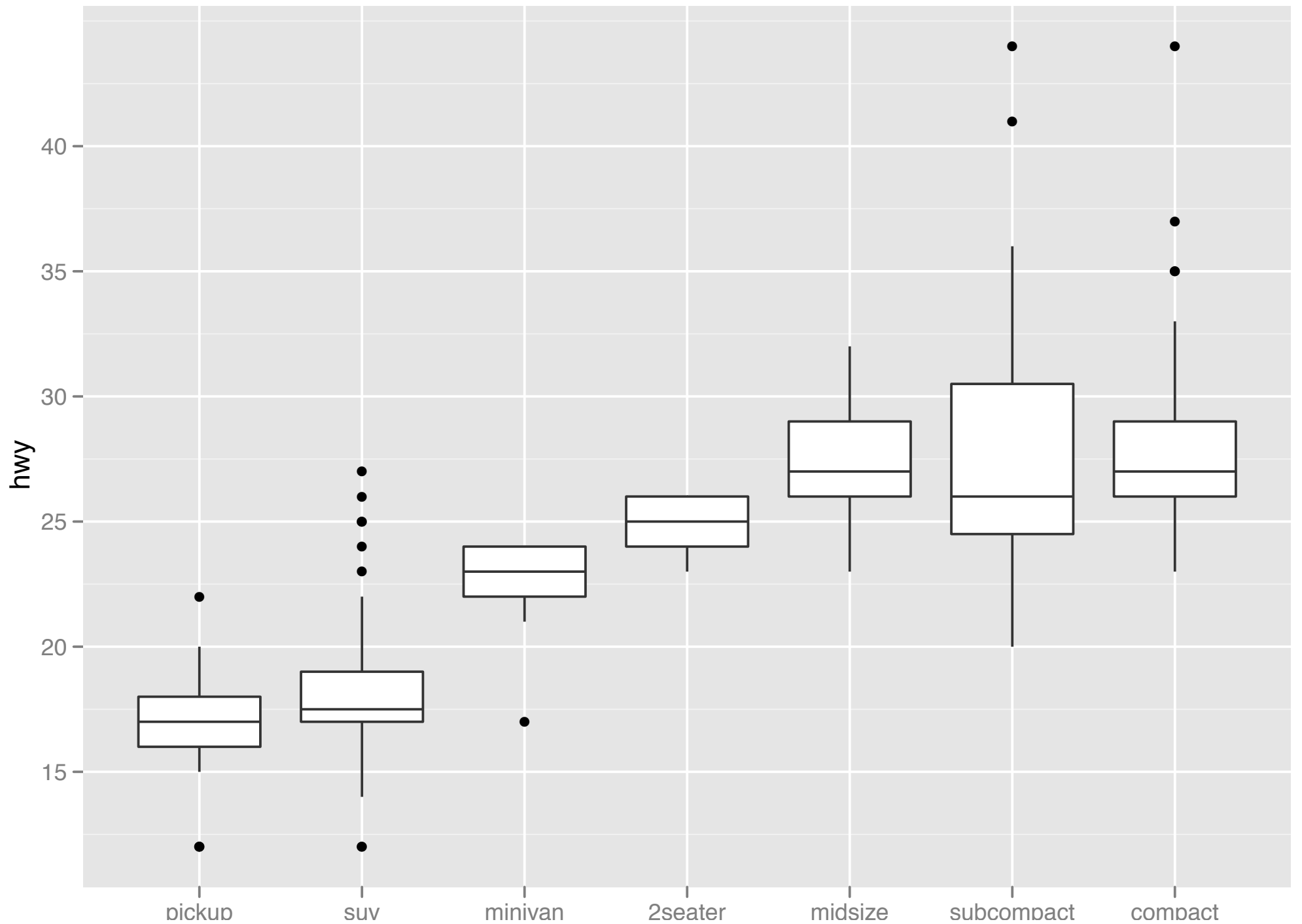


Incredibly useful  
technique!

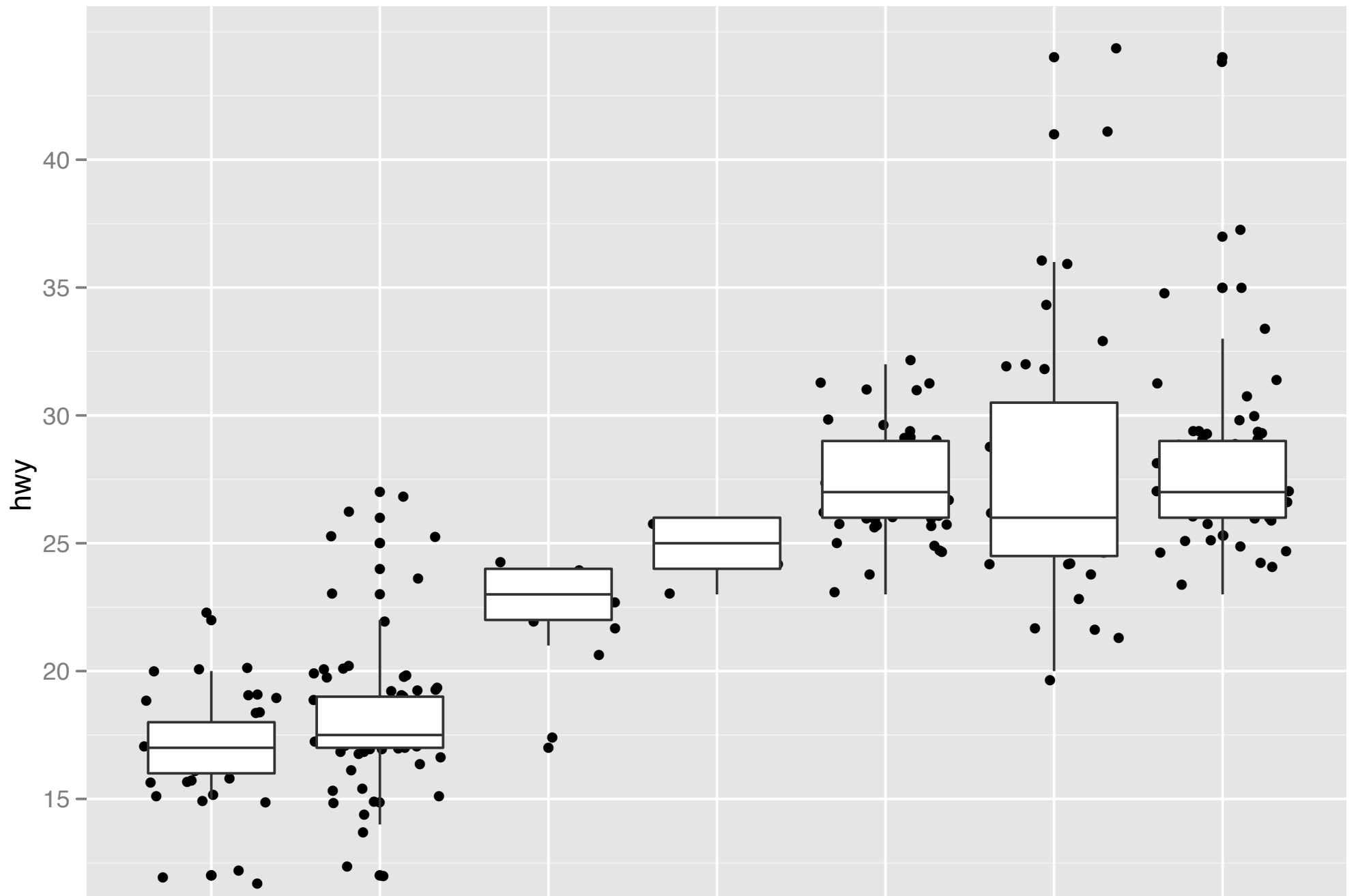
```
qplot(reorder(class, hwy), hwy, data = mpg)
```



```
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"))
```

# Your turn

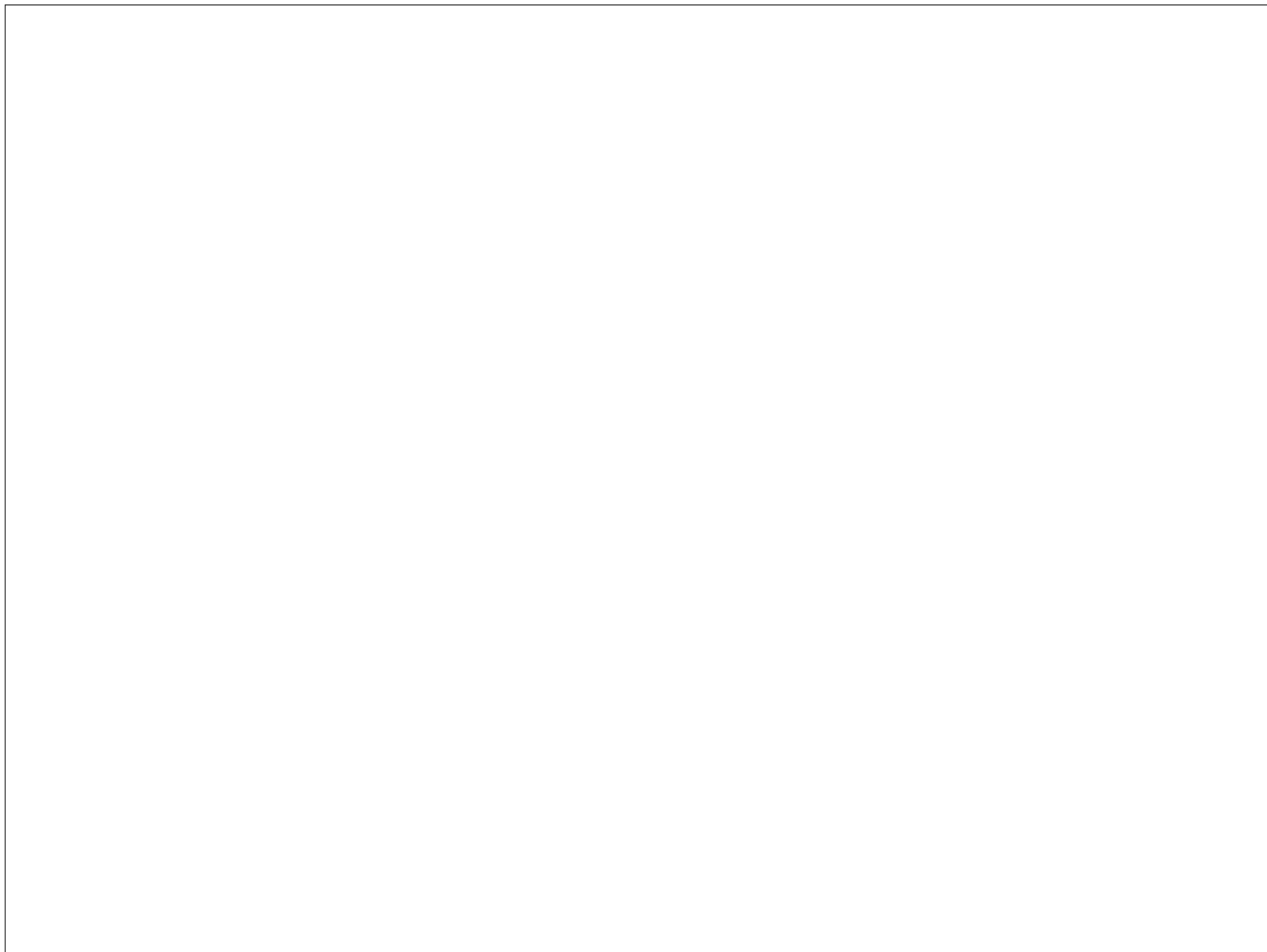
Read the help for reorder. Redraw the previously plots with class ordered by median hwy.

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.