http://courses.had.co.nz

Visualisation with R and ggplot2

Hadley Wickham

Assistant Professor / Dobelman Family Junior Chair

Department of Statistics / Rice University

August 2011

cc BY NC

http://courses.had.co.nz

1. Introductions

2. Course outline

HELLO

my name is

Hadley

Your turn

Throughout the course, there will be many opportunities for you to practice what you've learned and receive one-on-one help.

Please get things started by introducing yourself: what's your name and what sort of data do you work with.

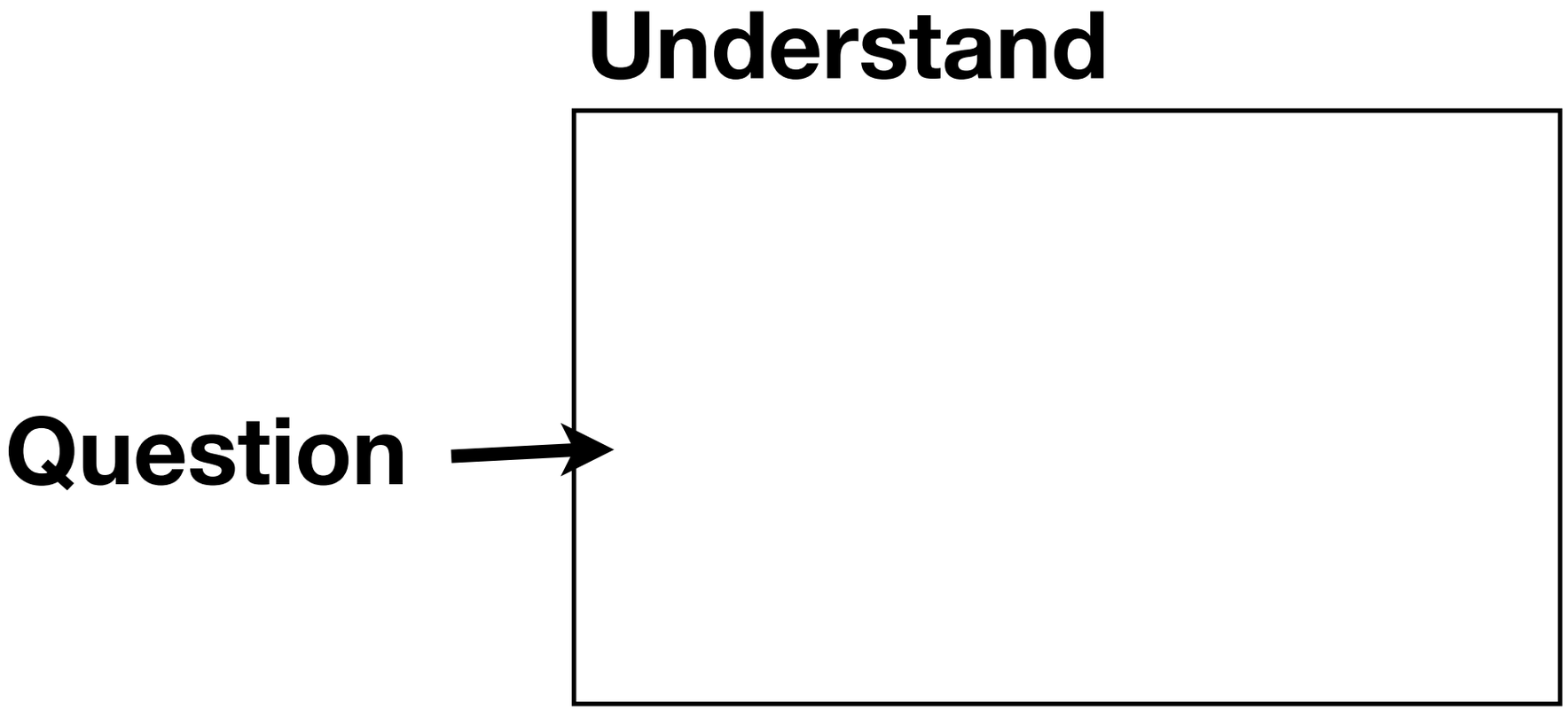
Day 1

Why? Data analysis

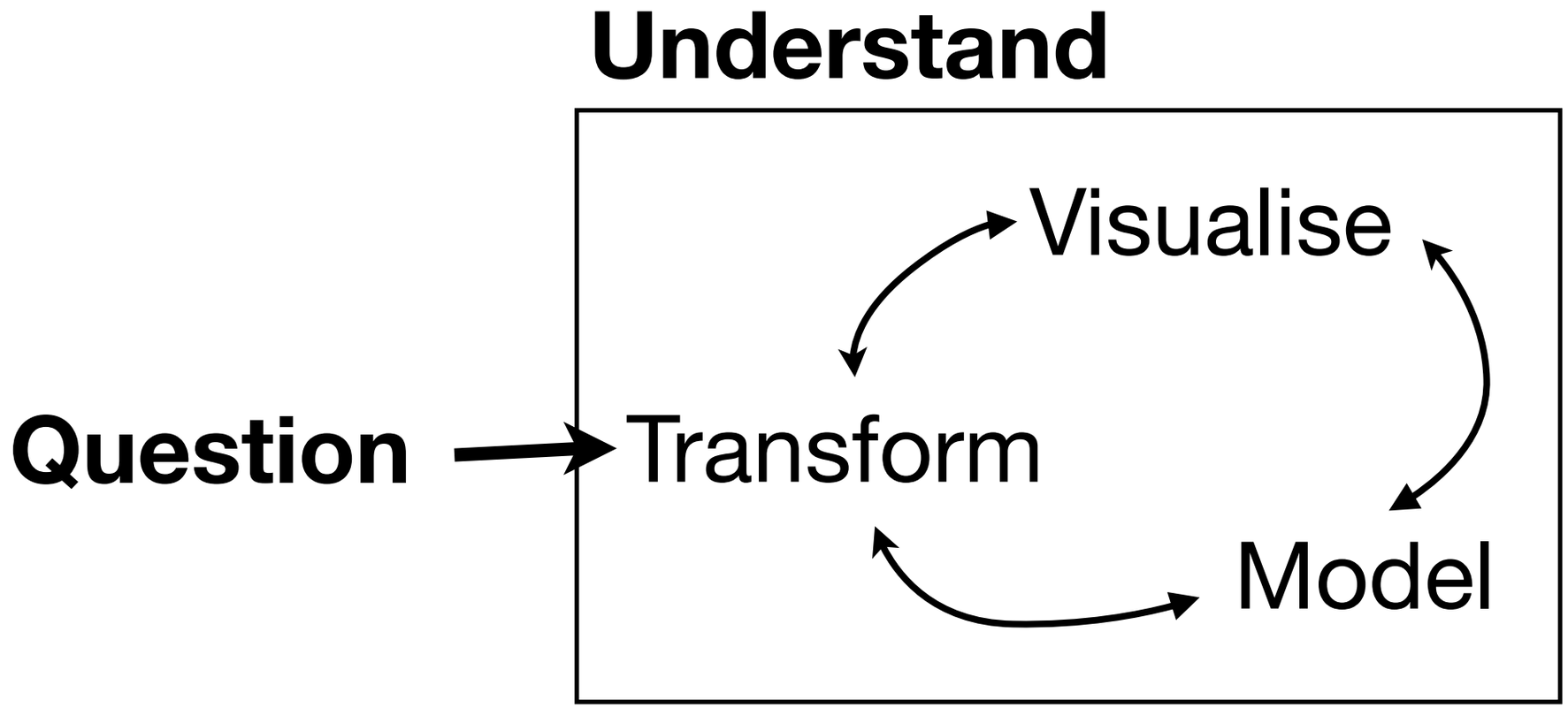
Why? Data analysis

Question

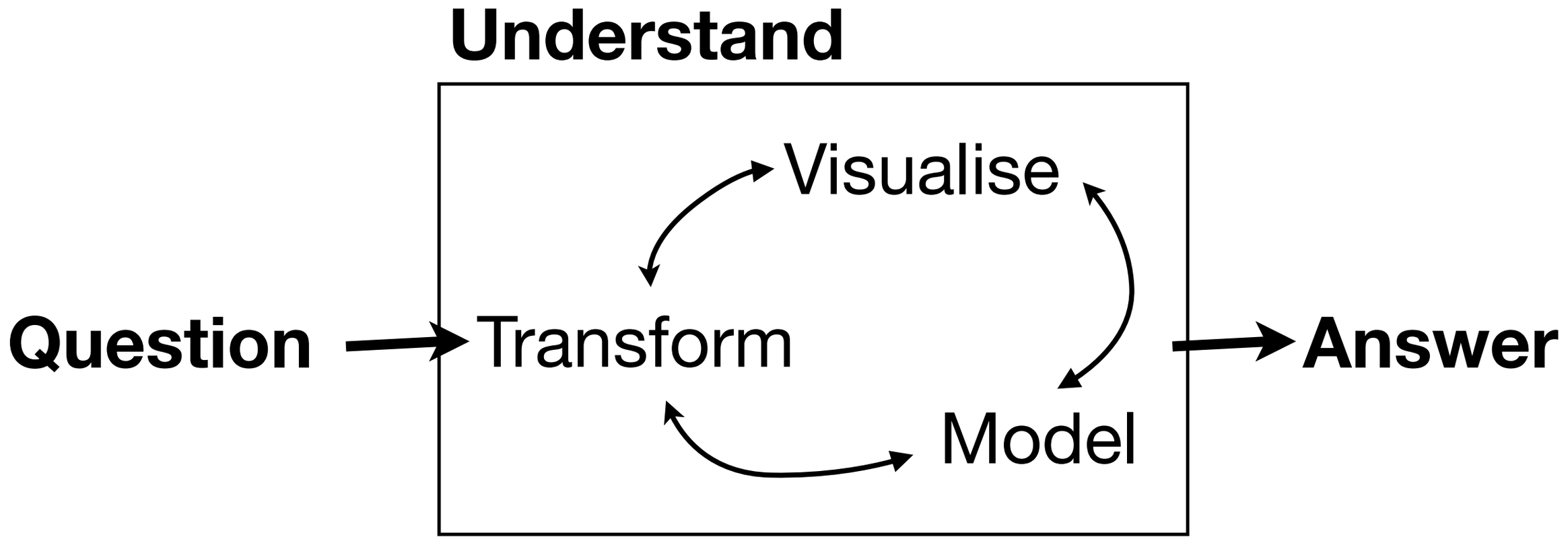
Why? Data analysis



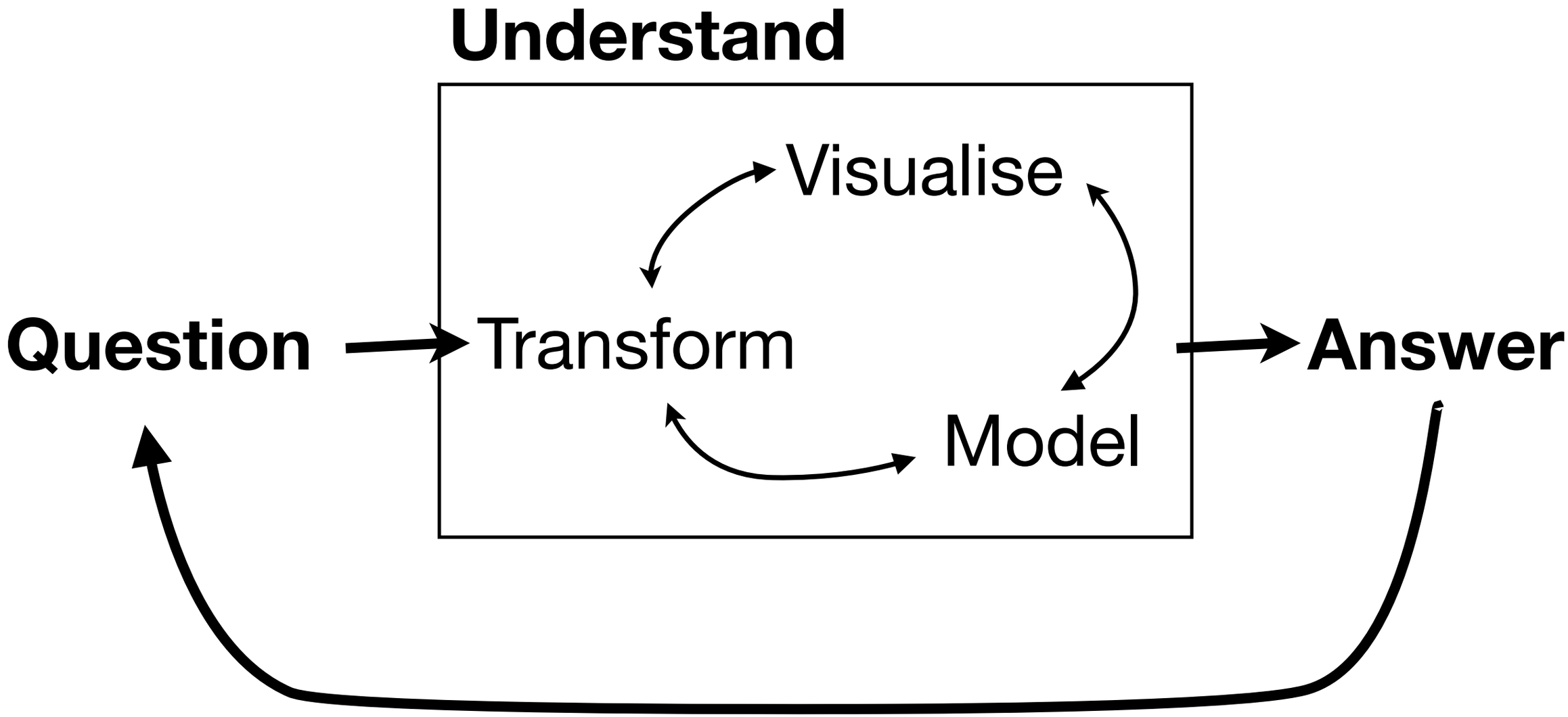
Why? Data analysis

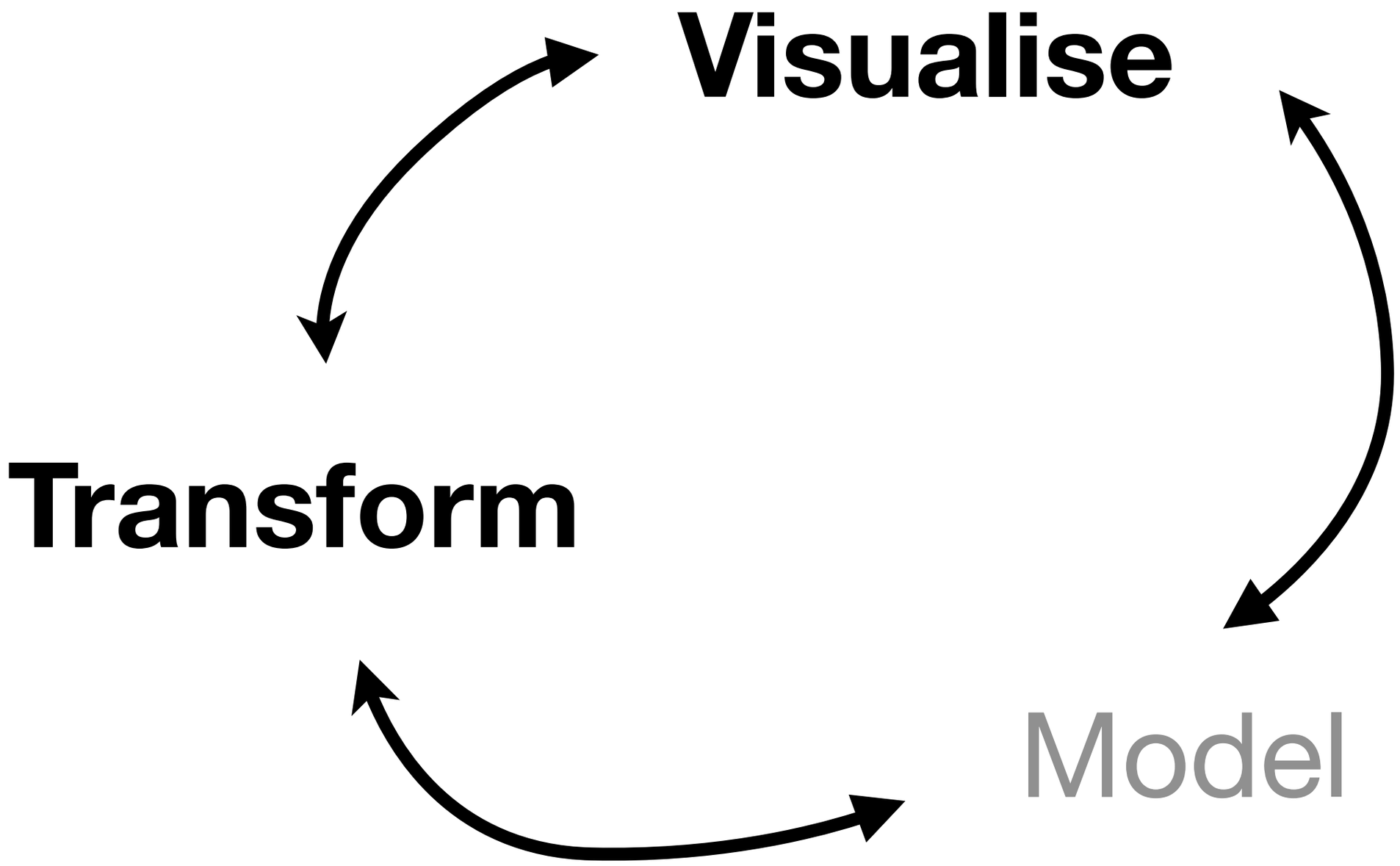


Why? Data analysis

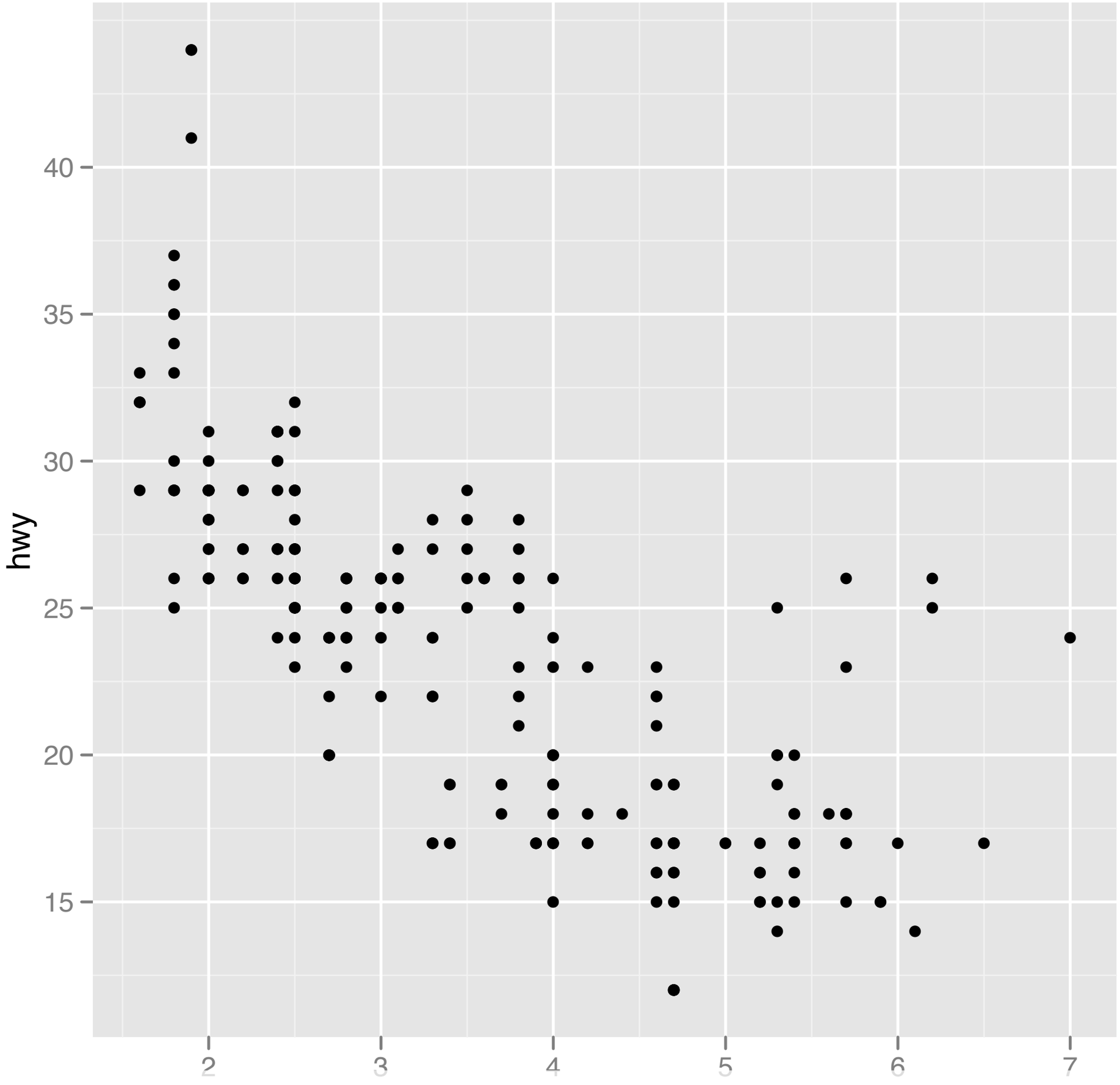


Why? Data analysis

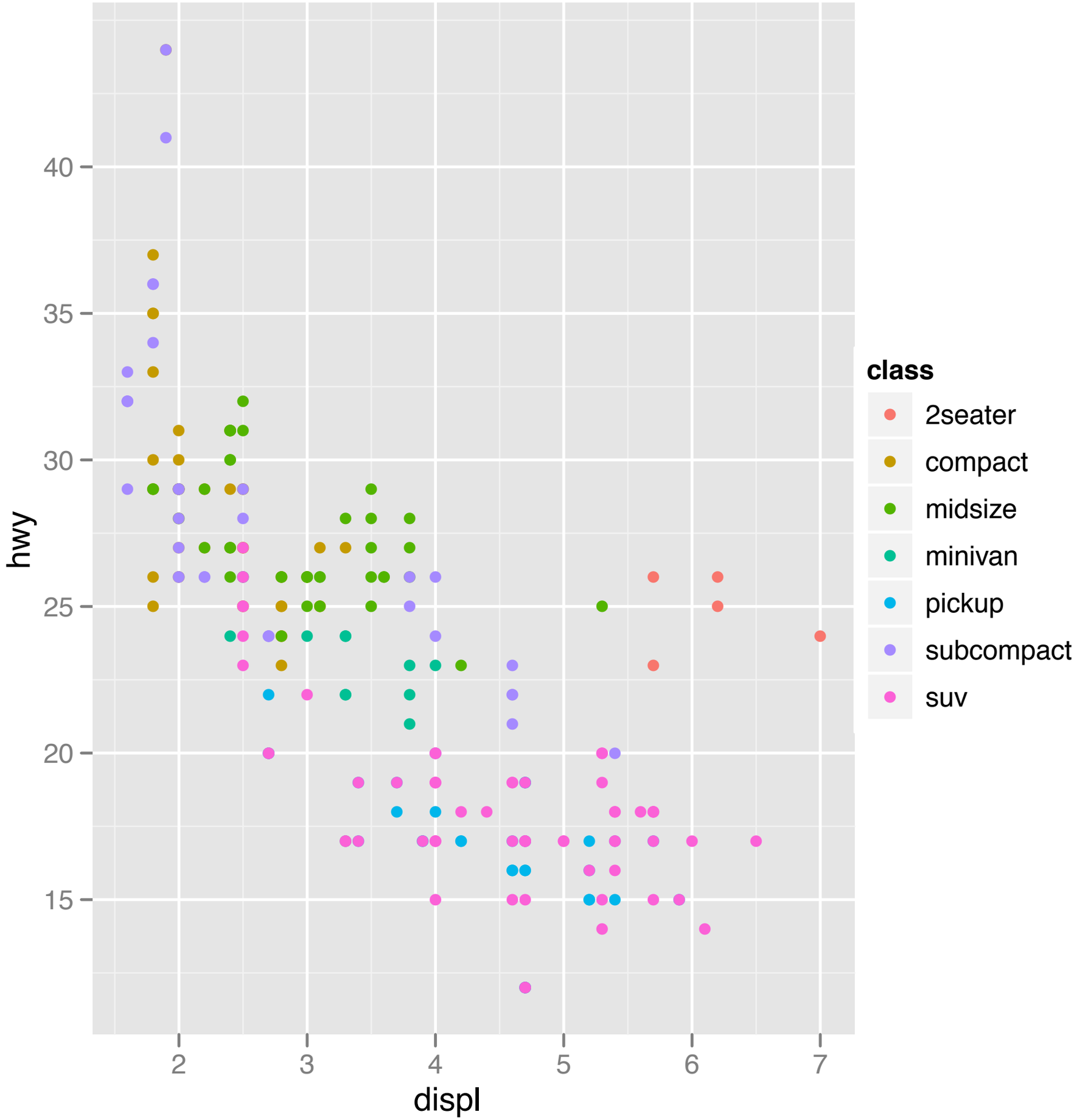


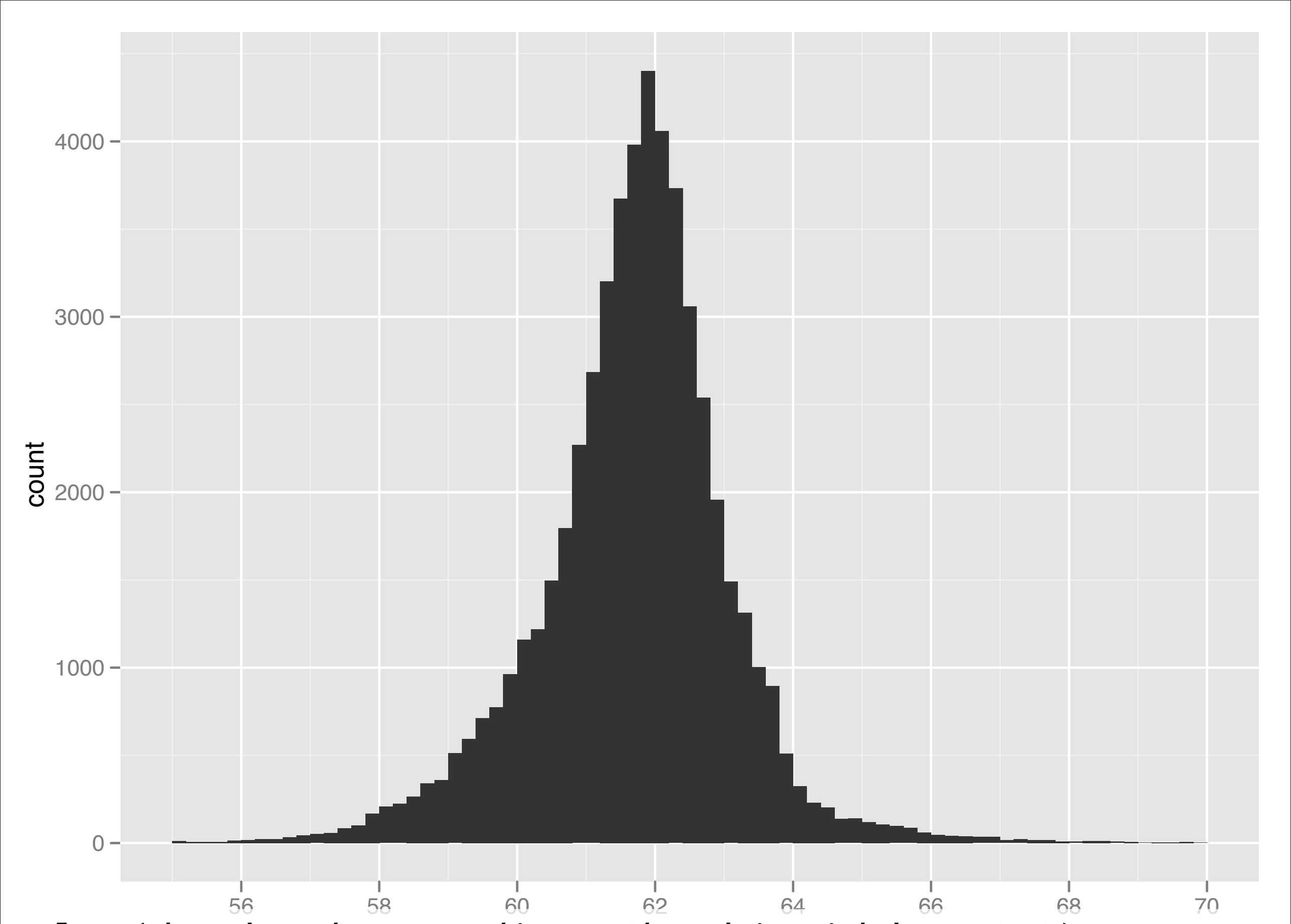

Visualise
Transform
Model

ggplot2 basics

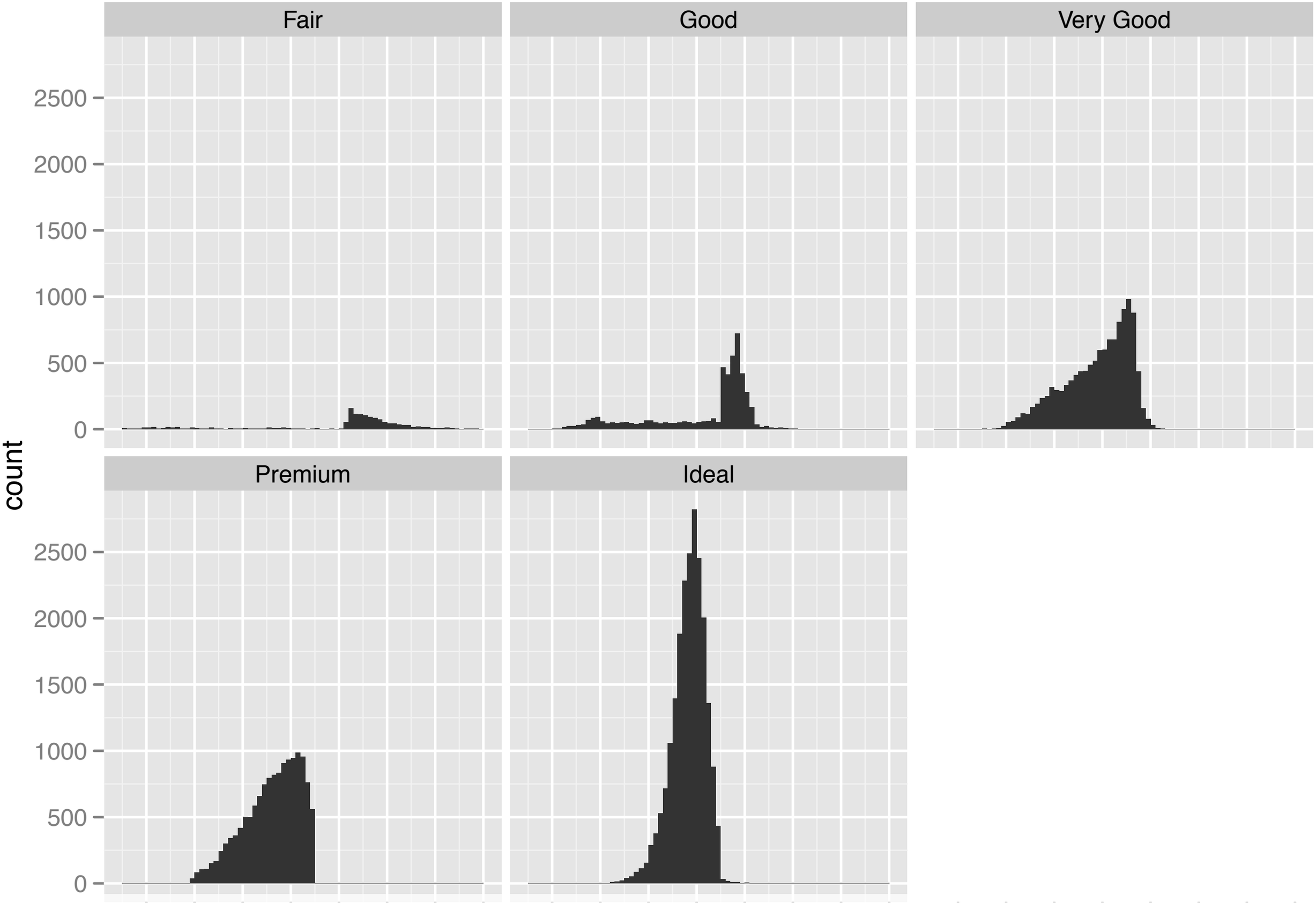



```
qplot(displ, hwy, data = mpg)
```


hwy
40
35
30
25
20
15
2
3
4
5
6
7
displ
class
2seater
compact
midsize
minivan
pickup
subcompact
suv




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

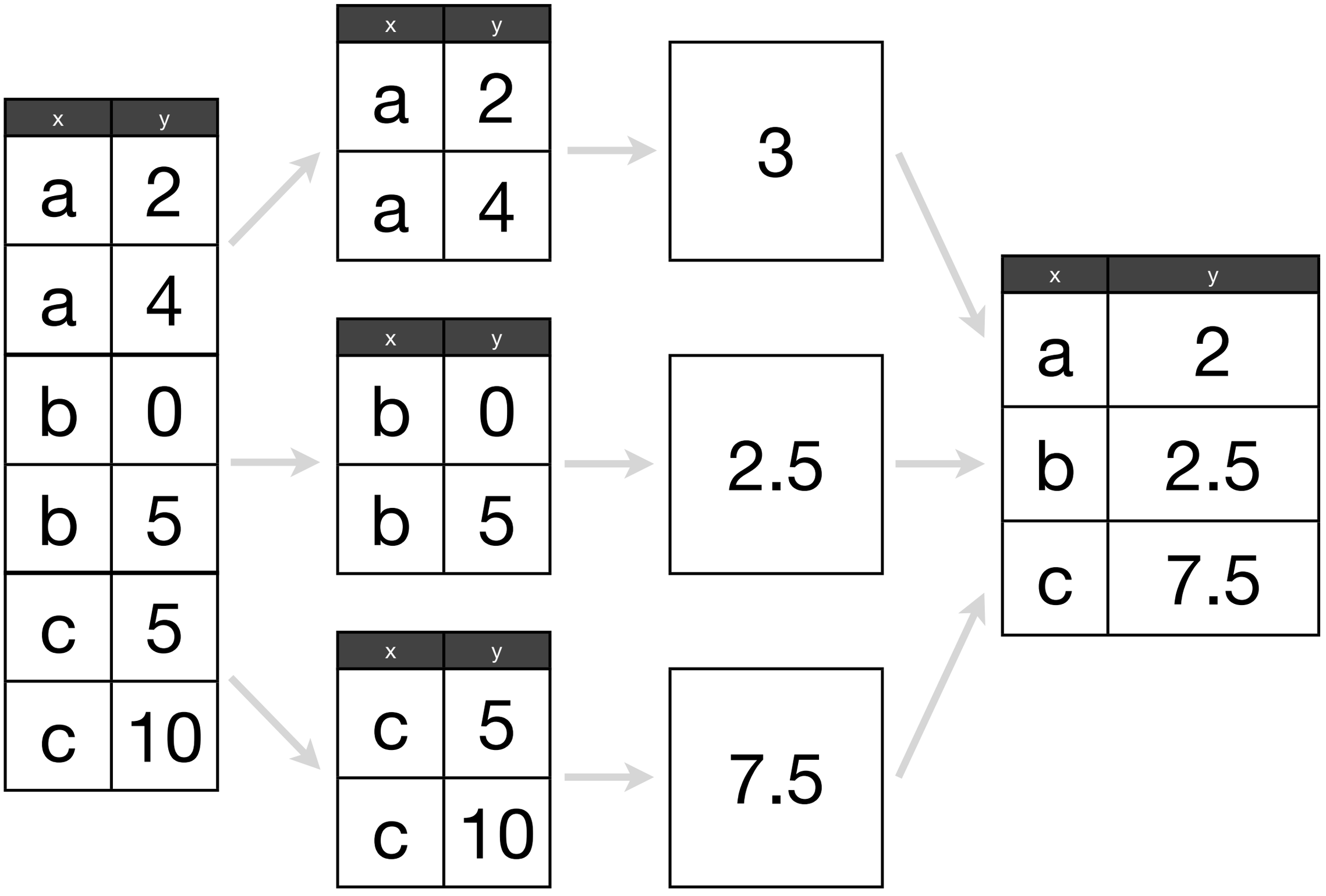


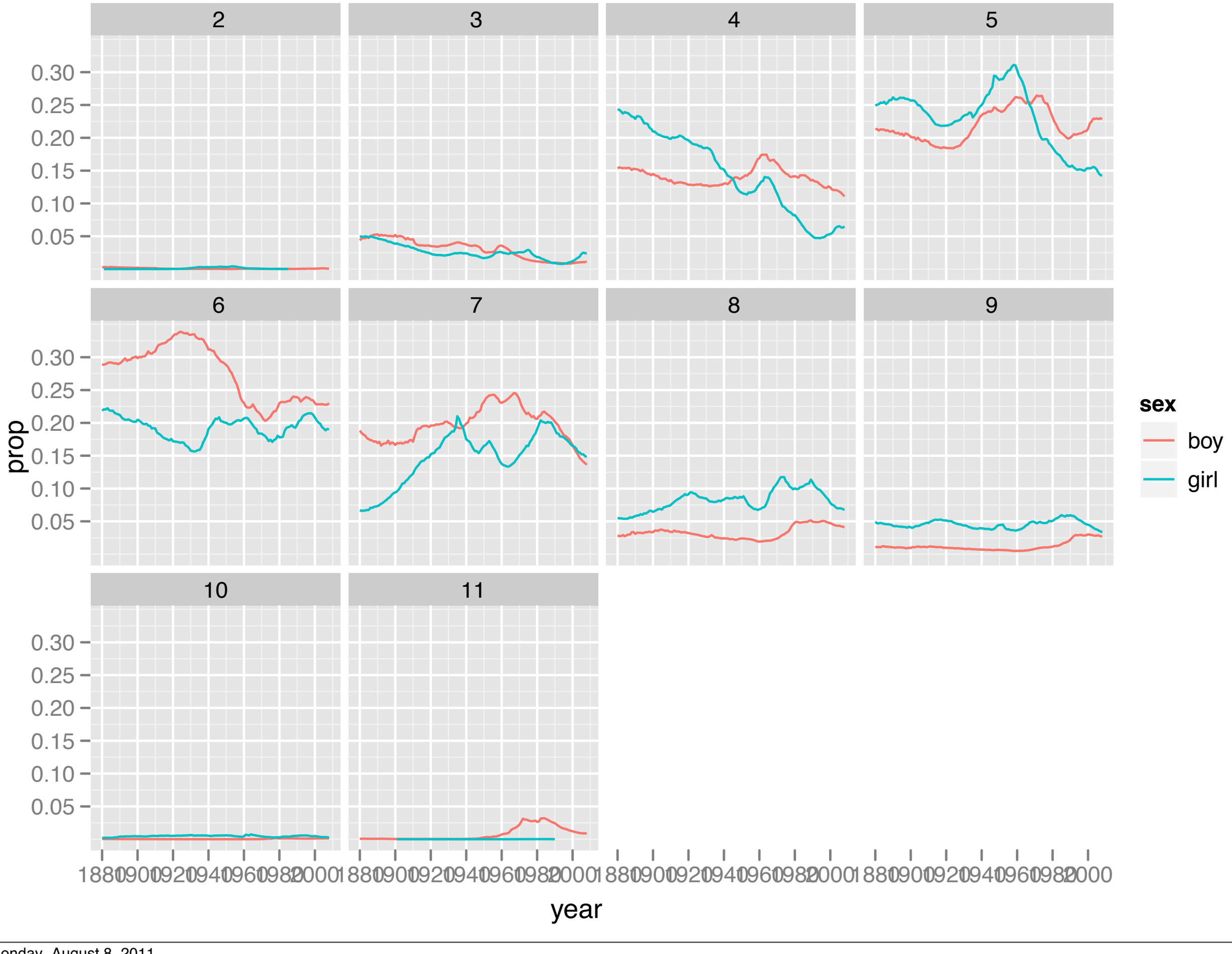

```
qplot(depth, data = diamonds, binwidth = 0.2) +
  xlim(55, 70) + facet_wrap(~cut)
```

Data manipulation

Function	Package
`subset`	`base`
`summarise`	`plyr`
`mutate`	`plyr`
`arrange`	`plyr`

Split Apply Combine

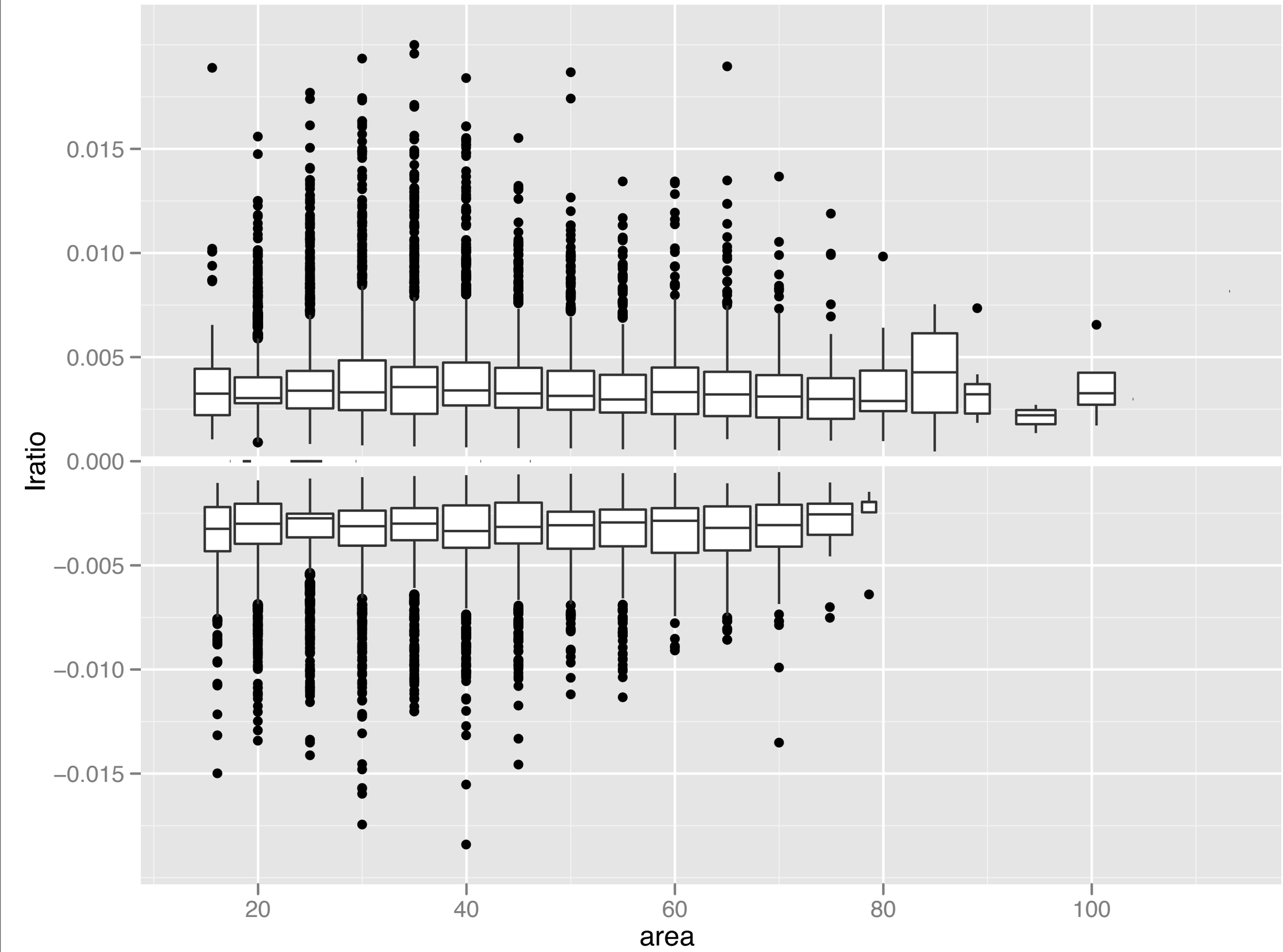



2
3
4
5
6
7
8
9
10
11
0.30
0.25
0.20
0.15
0.10
0.05
prop
1880
1900
1920
1940
1960
1980
2000
year
sex
boy
girl

Graphics: critique and creation

What should I plot?

How can I plot it?


0.015
0.010
0.005
0.000
−0.005
−0.010
−0.015
lratio
20
40
60
80
100
area

Polishing plots for presentation

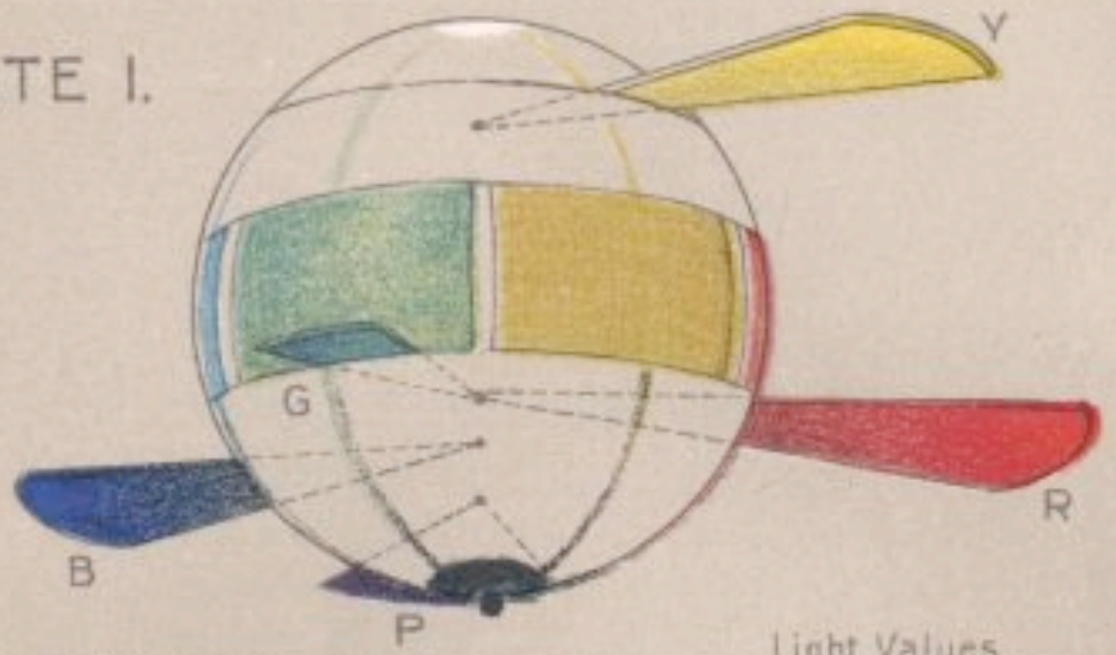
1. **Scales**: used to override default perceptual mappings, and tune parameters of axes and legends.

2. **Coordinate systems**: override default Cartesian coordinate system

3. **Themes**: control presentation of non-data elements.

4. **Saving your work**: to include in reports, presentations, etc.

PLATE I.



Light Values

Middle Values

Dark Values

P B G Y R

Copyright 1907 by A. H. Munsell

Learning a new
language is hard!