Visualisation with R

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

July 2010

1. Course essentials
2. Preview of today
3. Learning R
4. Diving in

HELLO

my name is

Hadley

http://had.co.nz/who-vis

Resources

Zip file contains all the data we'll use.

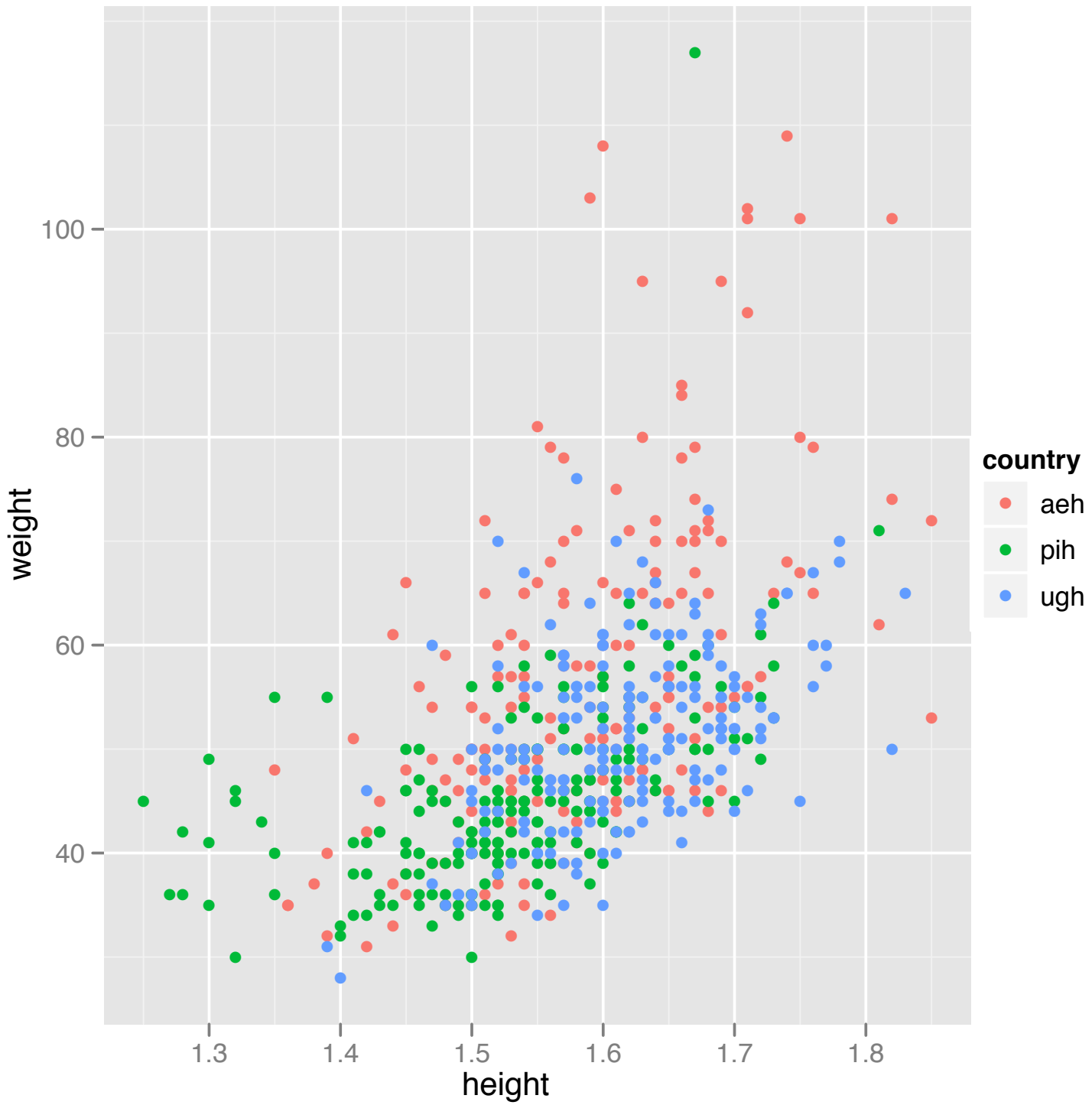
Slides contain all the code you should need to get up and running.

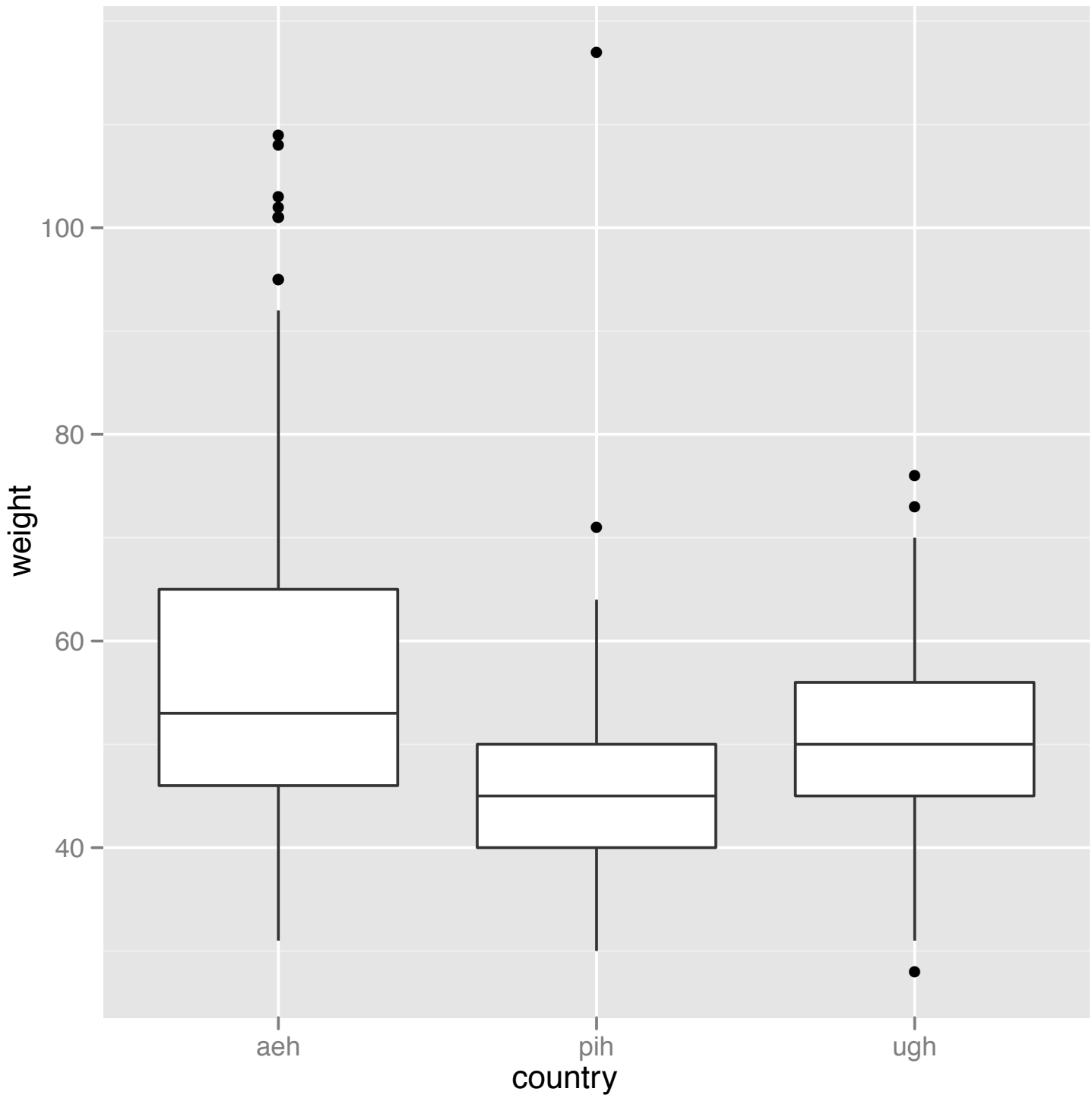
Code files contain pretty much everything that I'll show you (plus some extras).

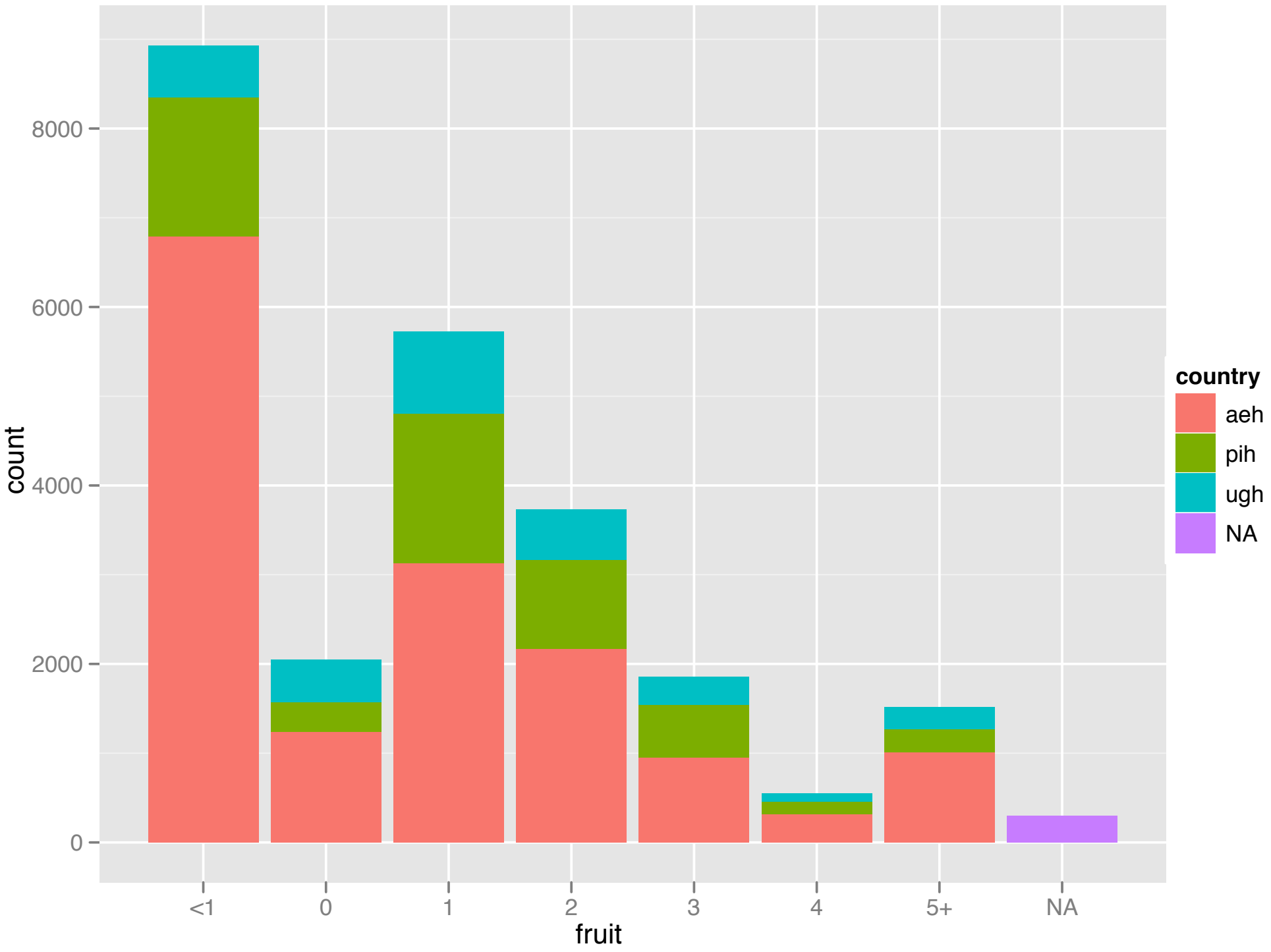
Overview

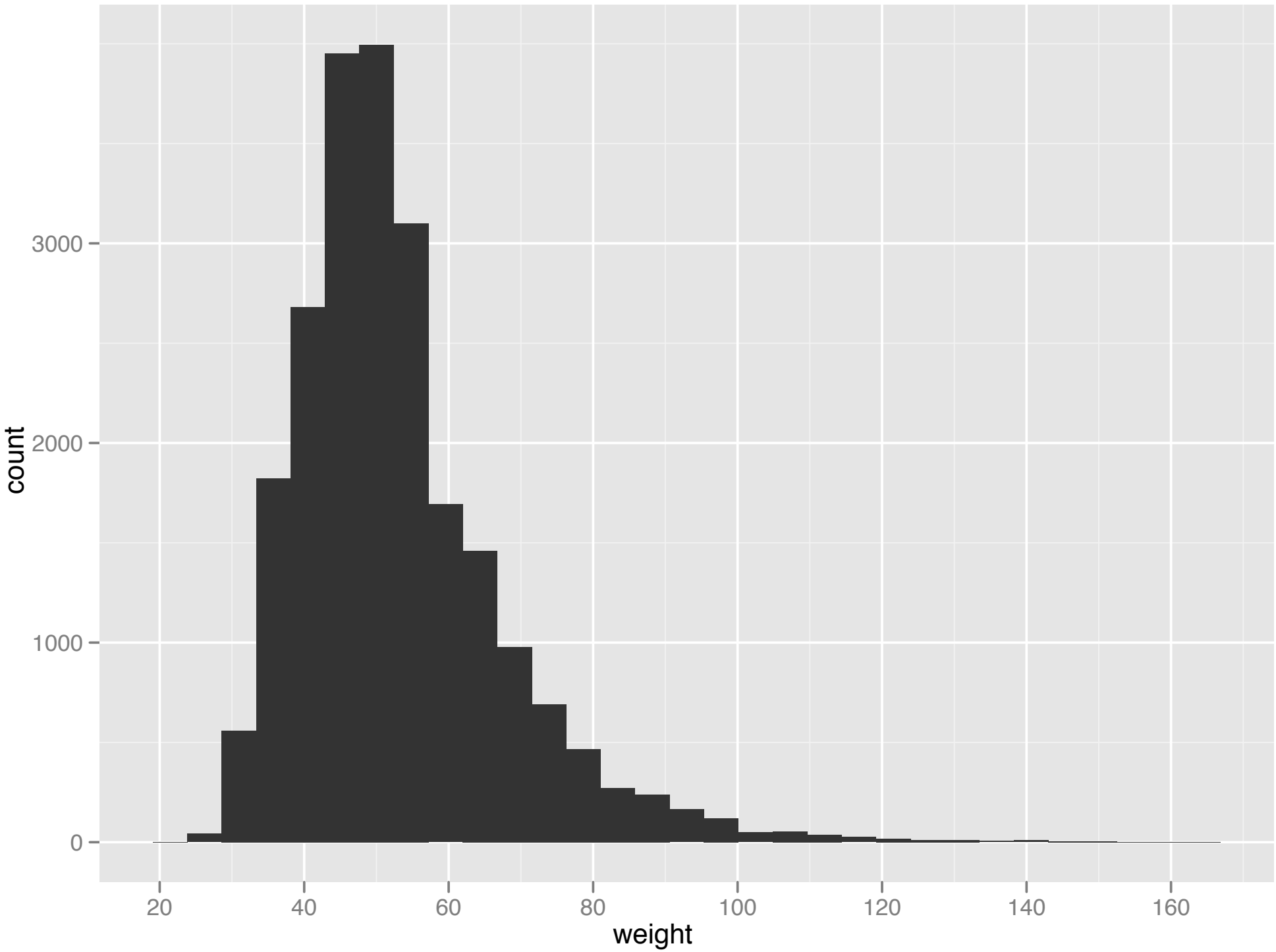
Global school health survey

Basic graphics


weight
100
80
60
40
1.3
1.4
1.5
1.6
1.7
1.8
height
country
aeh
pih
ugh


100
80
60
40
weight
aeh
pih
ugh
country

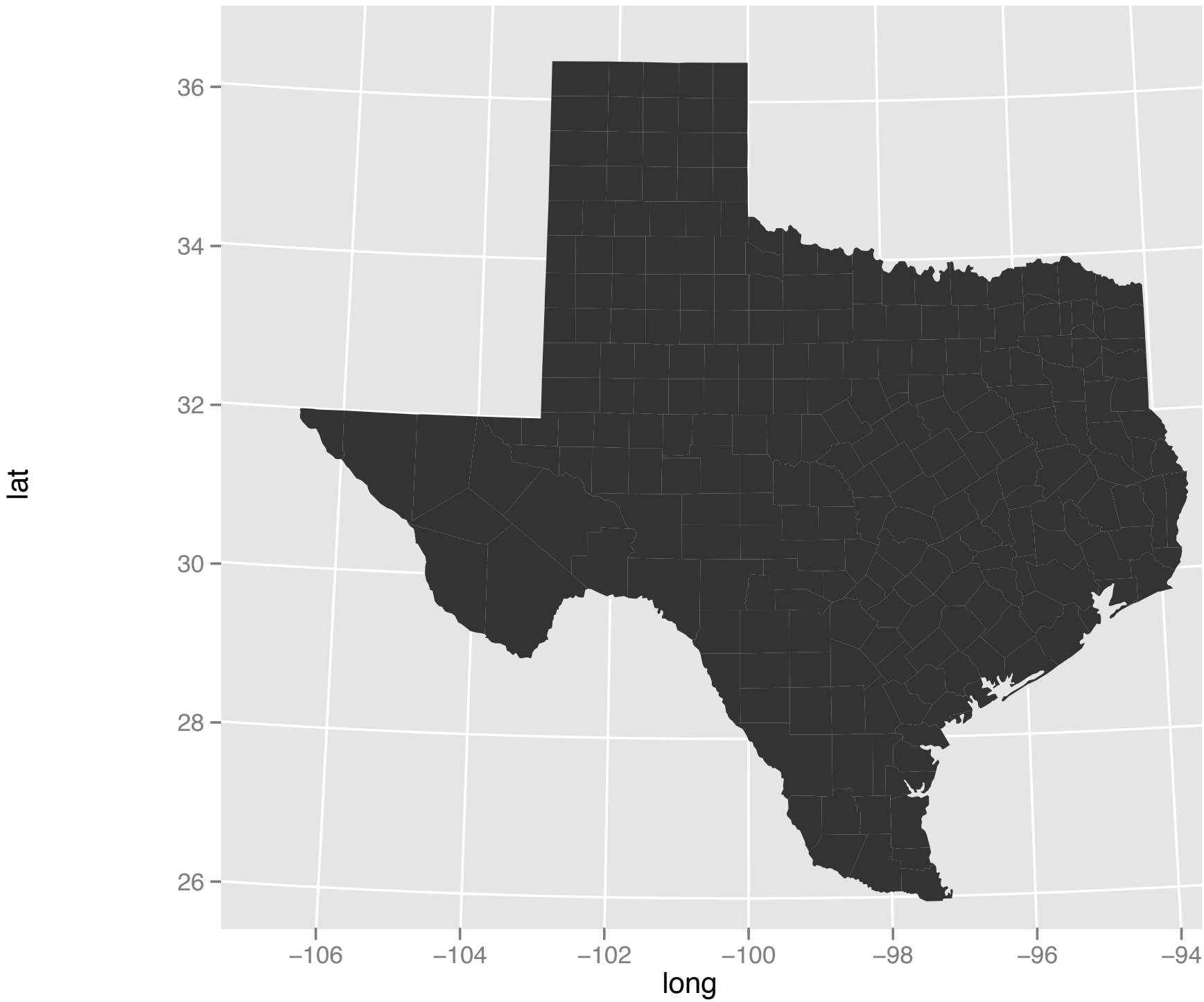

count
8000
6000
4000
2000
0
<1
0
1
2
3
4
5+
NA
fruit
country
aeh
pih
ugh
NA

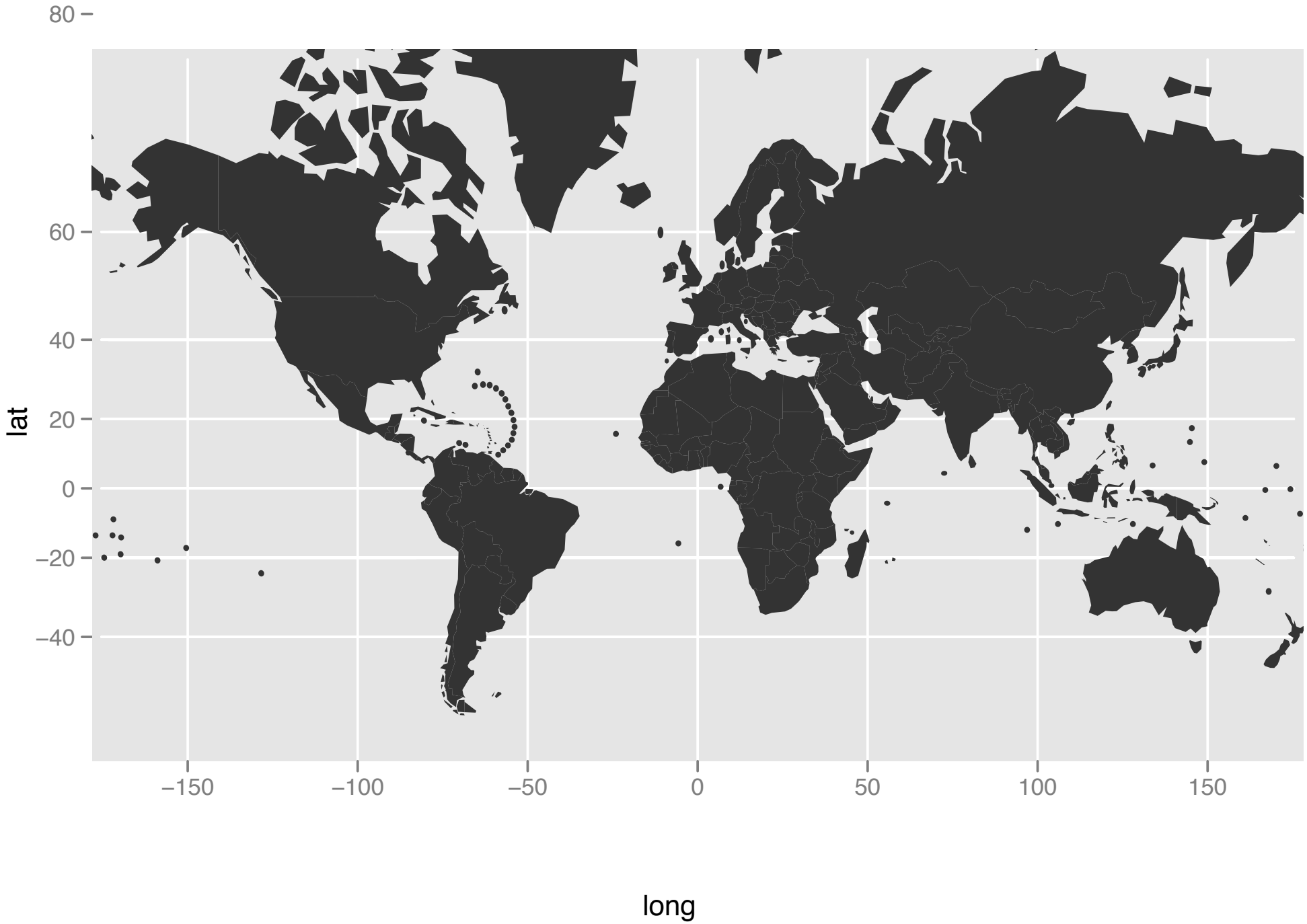

count
3000
2000
1000
0
20
40
60
80
100
120
140
160
weight

Texas & world maps

Introduction to spatial data

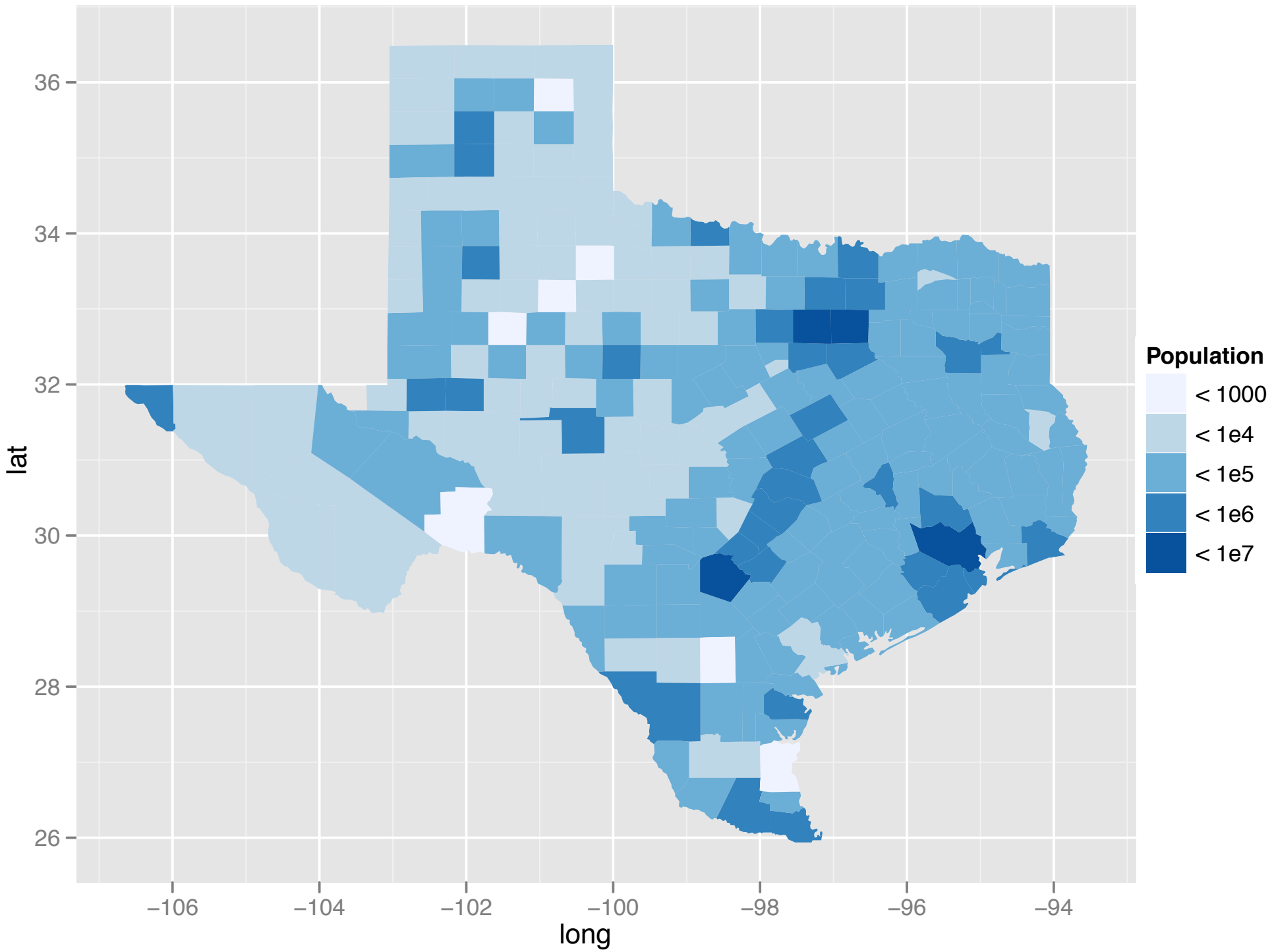
```
> head(borders, 20)
        long      lat group order state   county
1  -95.75271 31.53560     1     1 texas anderson
2  -95.76989 31.55852     1     2 texas anderson
3  -95.76416 31.58143     1     3 texas anderson
4  -95.72979 31.58143     1     4 texas anderson
5  -95.74698 31.61008     1     5 texas anderson
6  -95.72405 31.63873     1     6 texas anderson
7  -95.75271 31.67311     1     7 texas anderson
8  -95.76989 31.66738     1     8 texas anderson
9  -95.77563 31.63300     1     9 texas anderson
10 -95.79855 31.63873     1    10 texas anderson
11 -95.81000 31.67311     1    11 texas anderson
12 -95.79282 31.71321     1    12 texas anderson
13 -95.82146 31.70748     1    13 texas anderson
14 -95.87876 31.71321     1    14 texas anderson
15 -95.87876 31.75332     1    15 texas anderson
16 -95.91887 31.78197     1    16 texas anderson
17 -95.94751 31.78197     1    17 texas anderson
18 -95.98762 31.80489     1    18 texas anderson
19 -95.98189 31.83927     1    19 texas anderson
20 -95.99908 31.86218     1    20 texas anderson
```

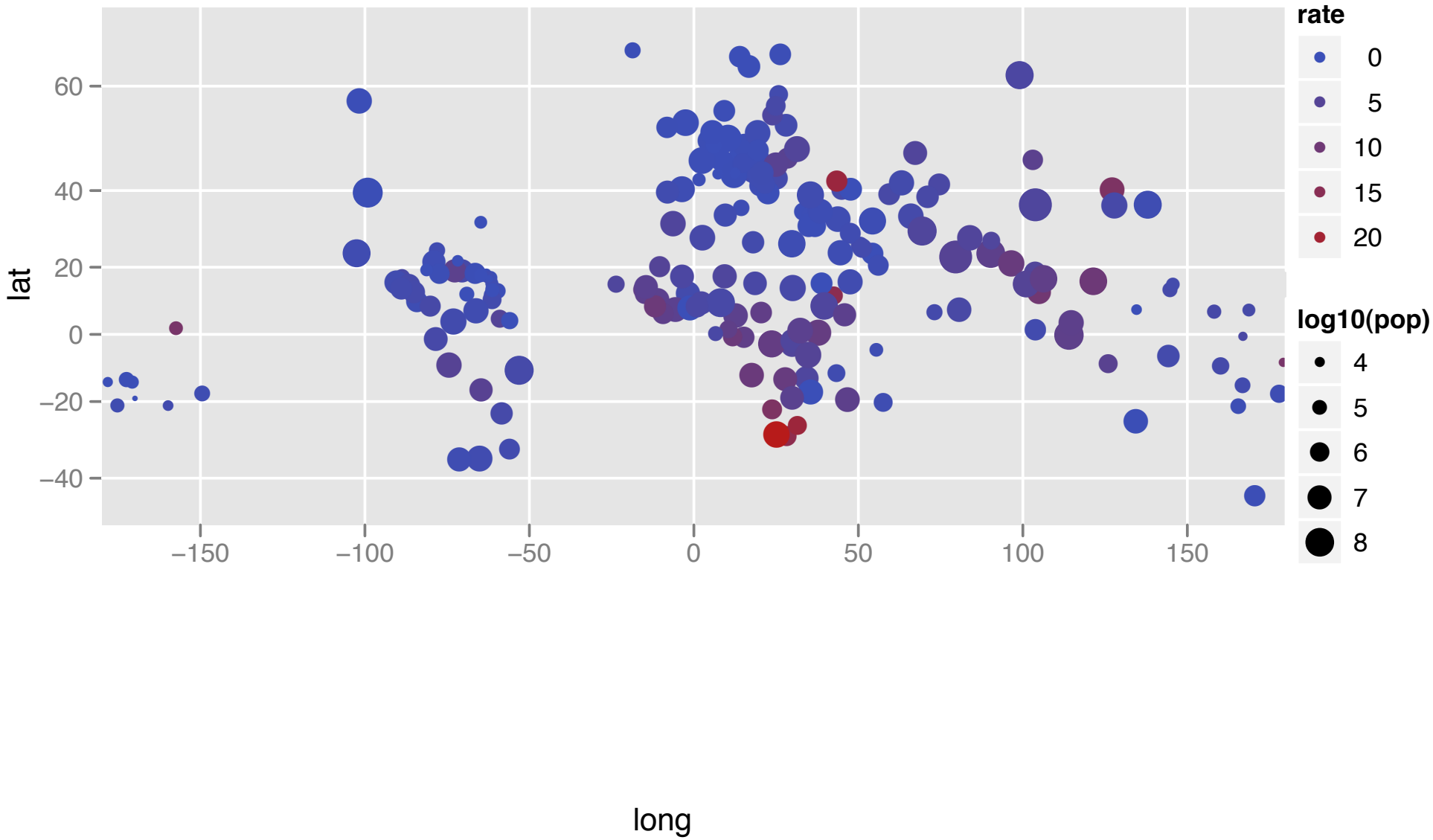

36
34
32
30
28
26
lat
−106
−104
−102
−100
−98
−96
−94
long


80
60
40
20
0
−20
−40
lat
−150
−100
−50
0
50
100
150
long

Texas deaths & TB notifications

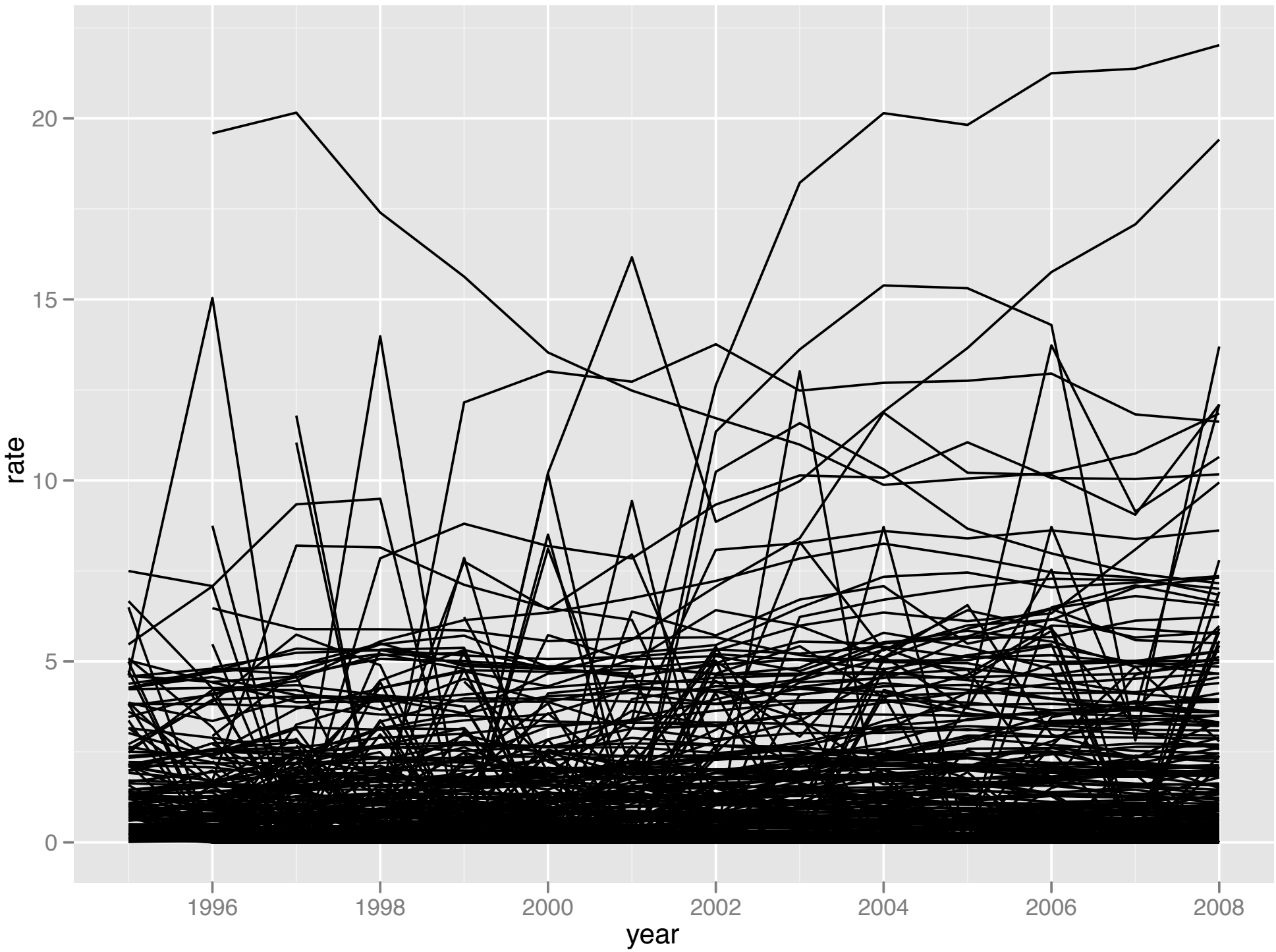
Visualising space (and time)

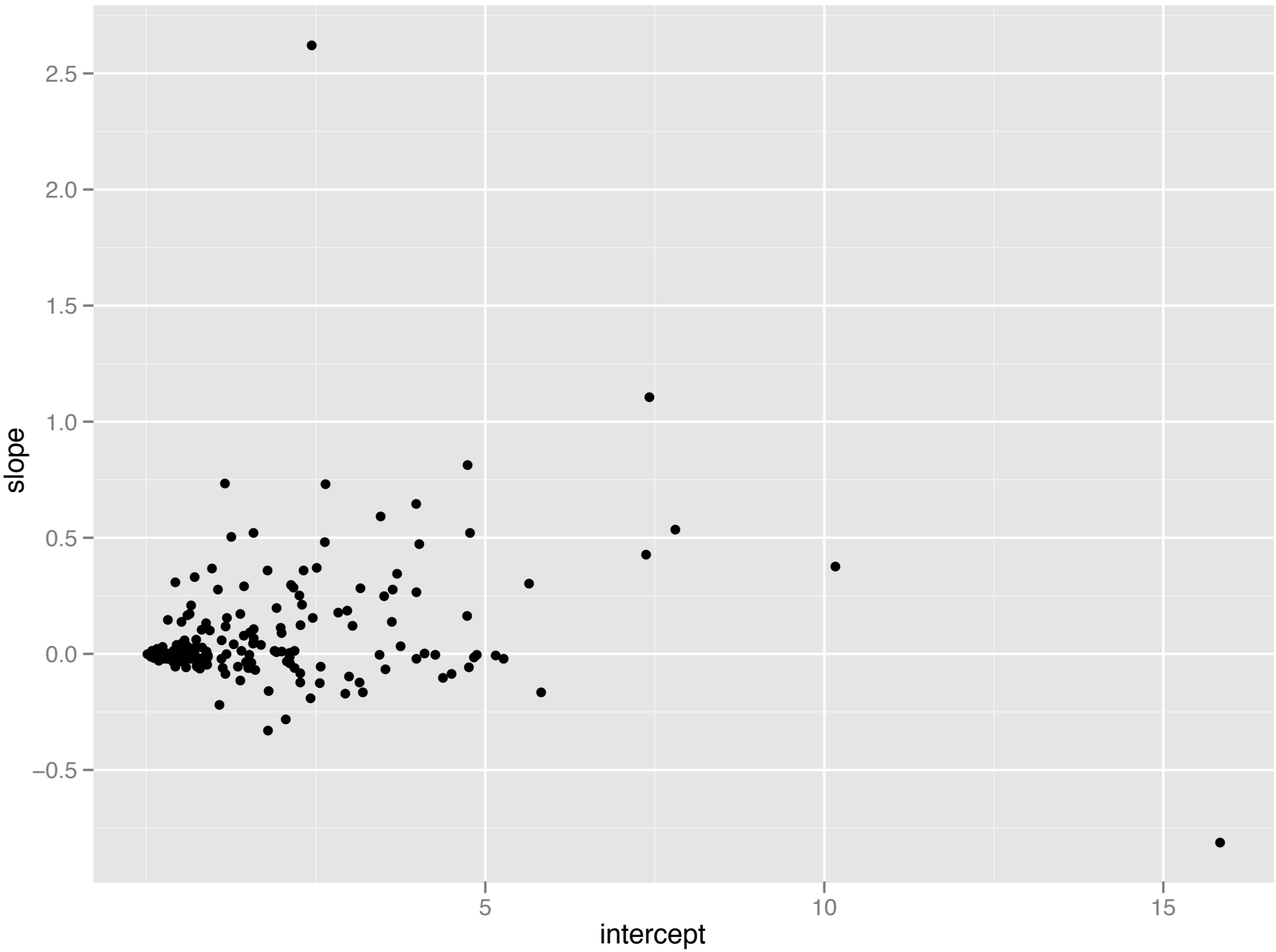

36
34
32
30
28
26
lat
−106
−104
−102
−100
−98
−96
−94
long
Population
< 1000
< 1e4
< 1e5
< 1e6
< 1e7


rate
0
5
10
15
20
log10(pop)
4
5
6
7
8
lat
60
40
20
0
−20
−40
−150
−100
−50
0
50
100
150
long

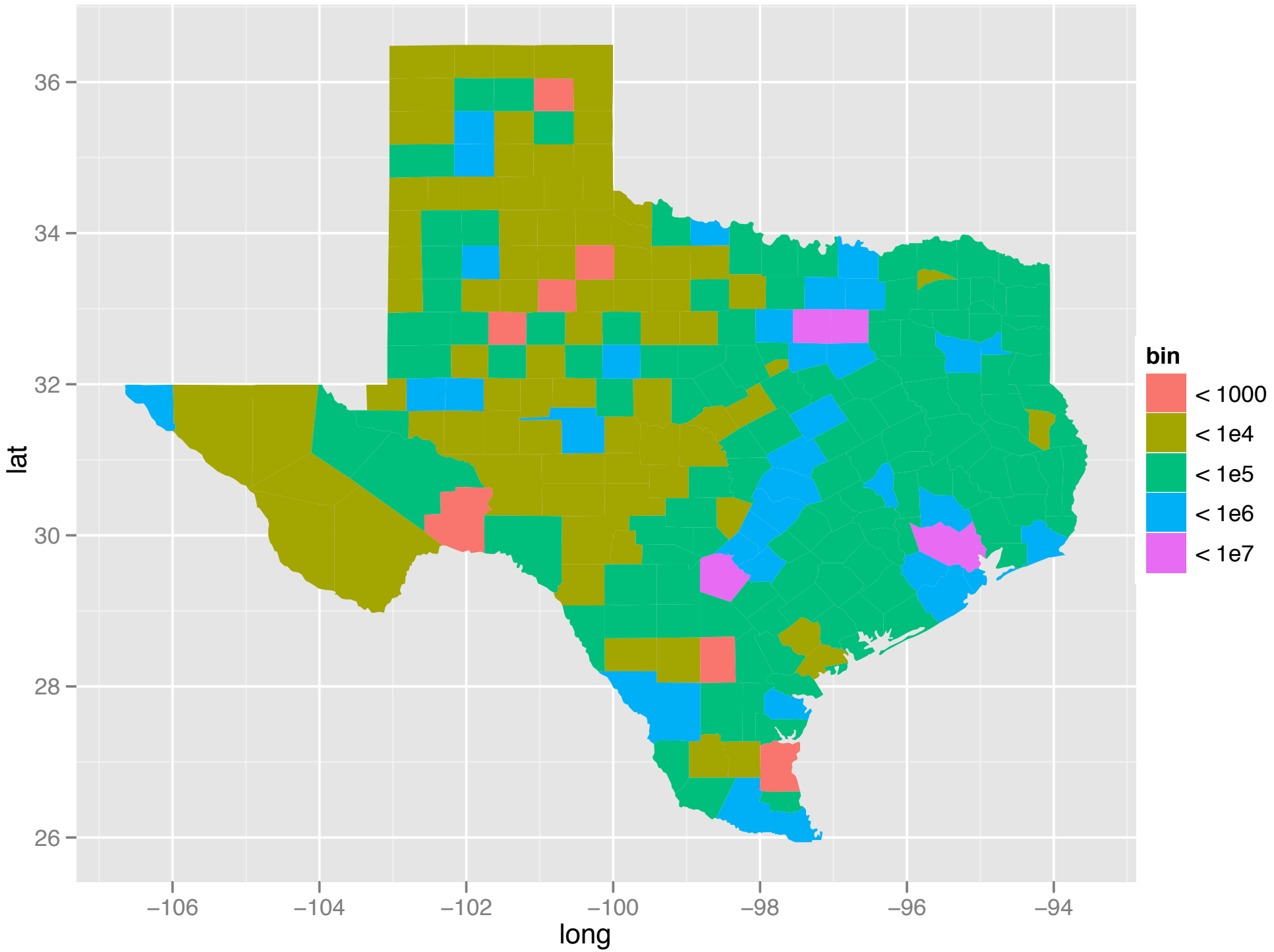
TB notifications

Exploring trends


rate
20
15
10
5
0
1996
1998
2000
2002
2004
2006
2008
year


2.5
2.0
1.5
1.0
0.5
0.0
−0.5
slope
5
10
15
intercept

Polishing plots for presentation


36
34
32
30
28
26
lat
−106
−104
−102
−100
−98
−96
−94
long
bin
< 1000
< 1e4
< 1e5
< 1e6
< 1e7

Where to next

1. Learning R
2. Learning ggplot2
3. Learning visualisation

Learning a new
language is hard!

Let’s get started!

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.