Visualising space

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

July 2010

1. Choropleth (thematic) maps
2. **Texas mortality**
3. **TB notifications**
4. Bubble maps

Getting started

```
options(stringsAsFactors = FALSE)
library(ggplot2)

pop <- read.csv("tx-pop.csv")
deaths <- read.csv("tx-deaths.csv")
borders <- read.csv("tx-borders.csv")
```

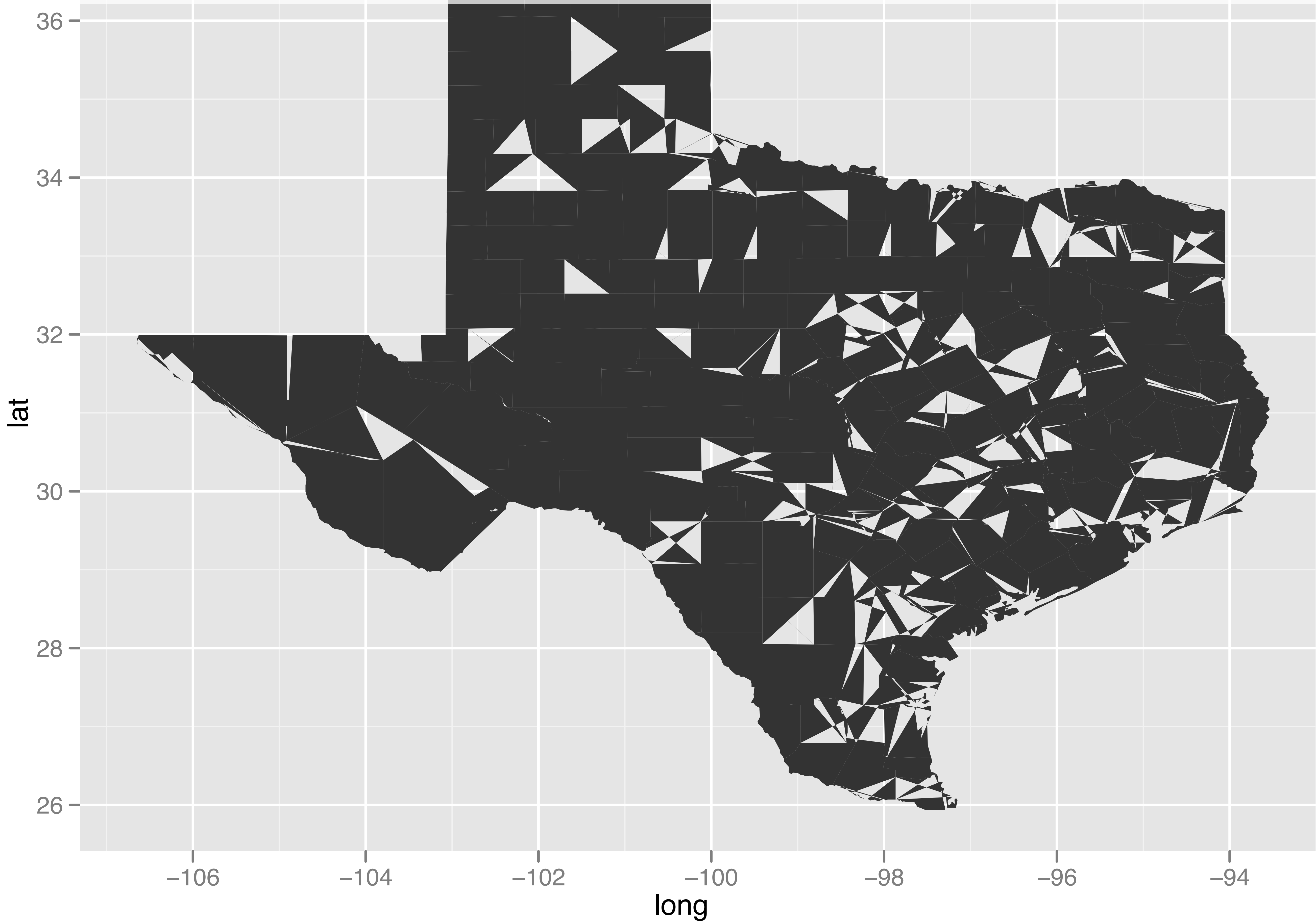
Choropleth maps

Make fill colour of map areas proportional to some value (e.g. a rate)

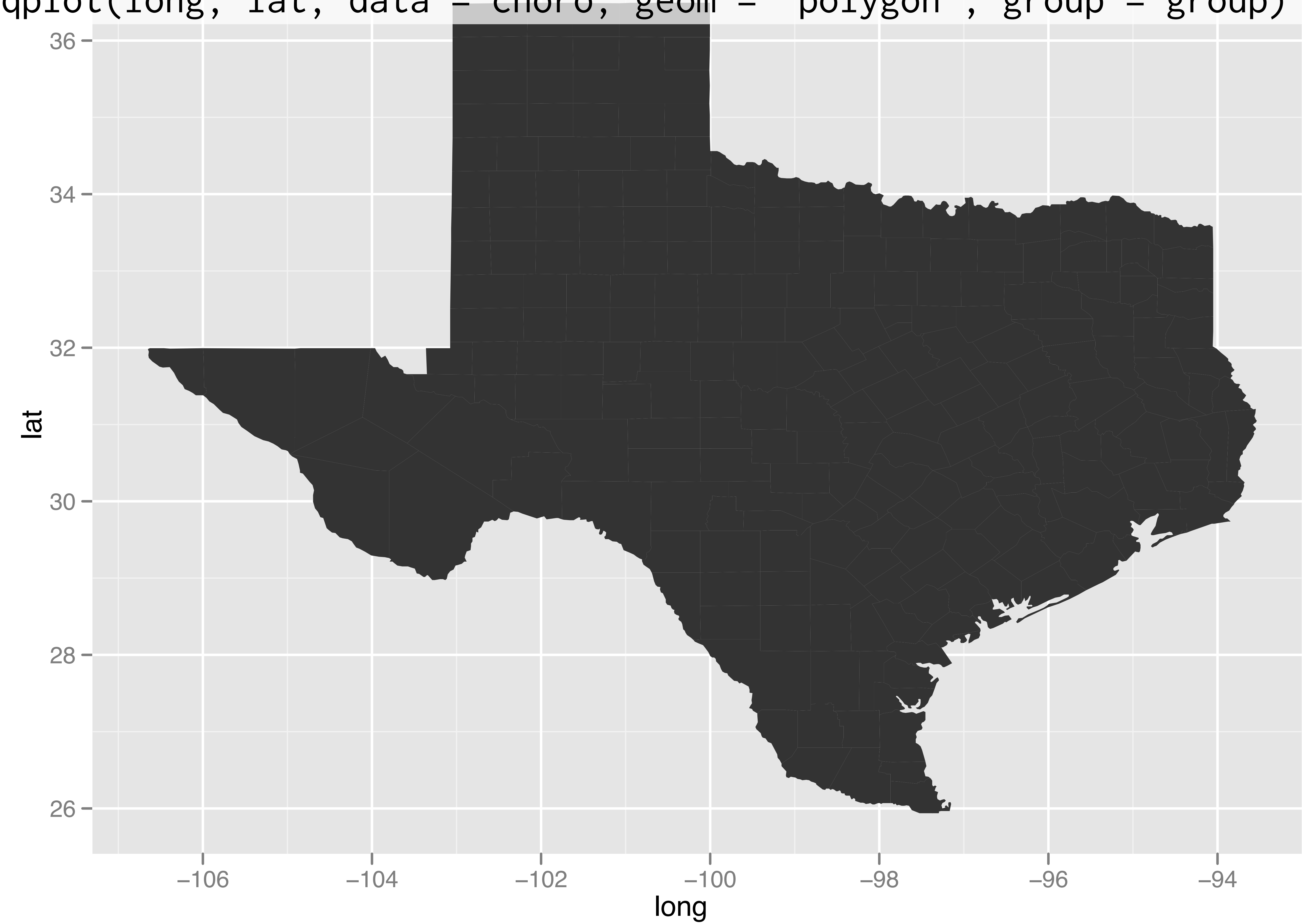
We'll demonstrate this with deaths for each county (254) in Texas, broken down by ICD-9 chapter.

First step is to **merge** borders with area level information.

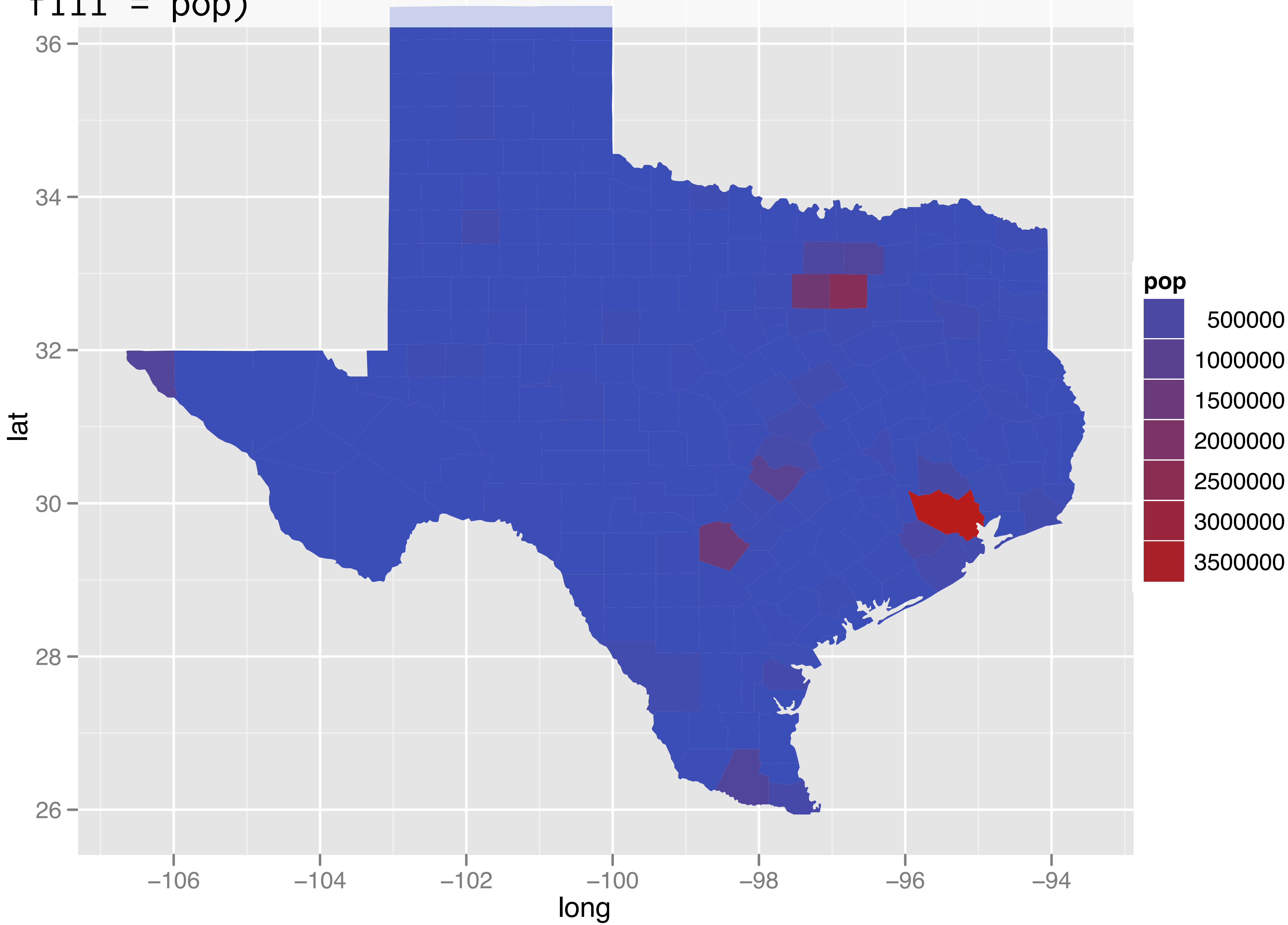
```
choro <- merge(borders, pop)
qplot(long, lat, data = choro, geom = "polygon", group = group)
```



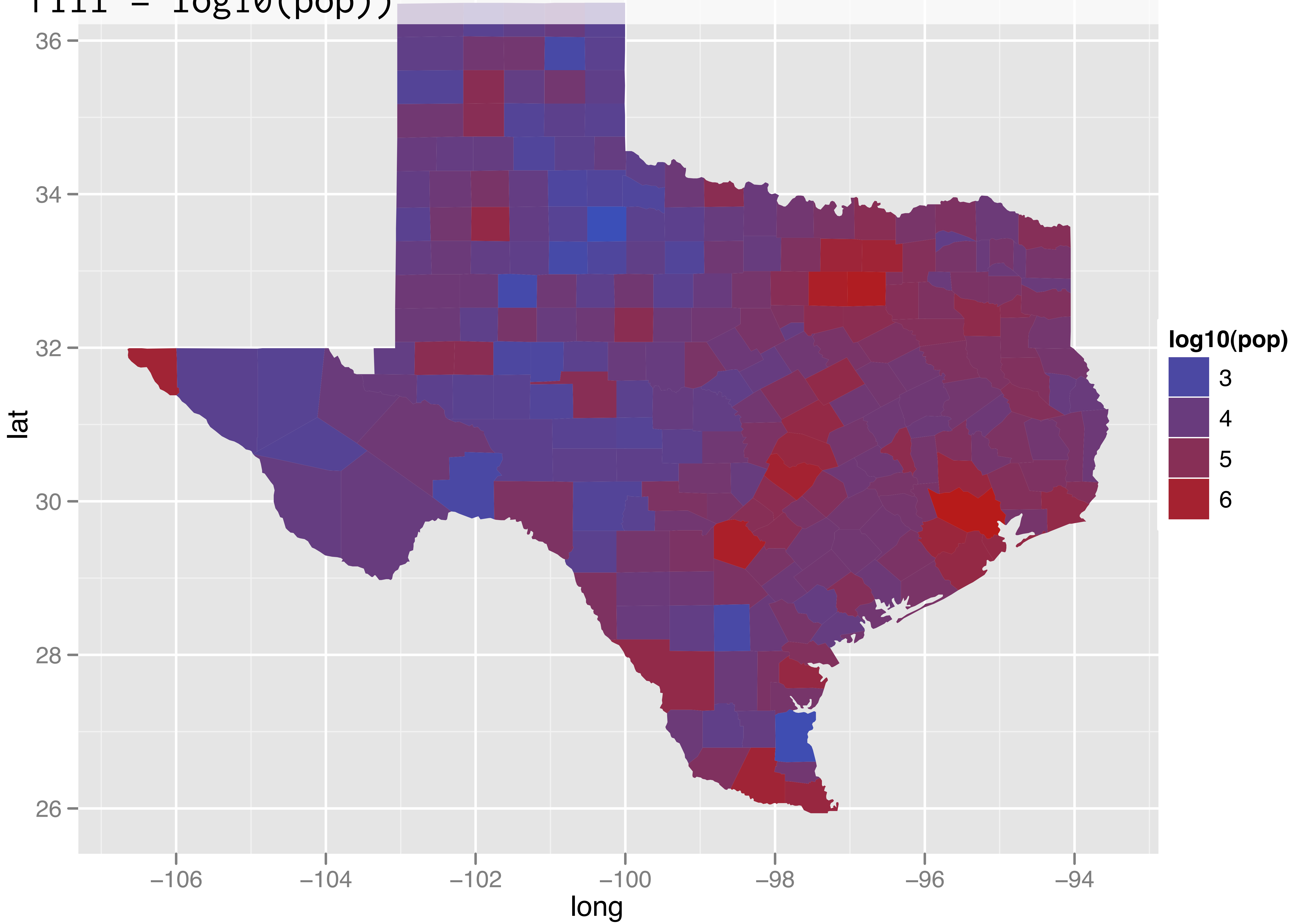
```
choro <- choro[order(choro$order), ]
qplot(long, lat, data = choro, geom = "polygon", group = group)
```



```
qplot(long, lat, data = choro, geom = "polygon", group = group,
  fill = pop)
```



```
qplot(long, lat, data = choro, geom = "polygon", group = group,
  fill = log10(pop))
```



Binning

Presentation more reliable if we bin into a small number of categories (4-9). Perception not harmed.

Usually best to bin so we have the same number of cases in each bin (quantiles).

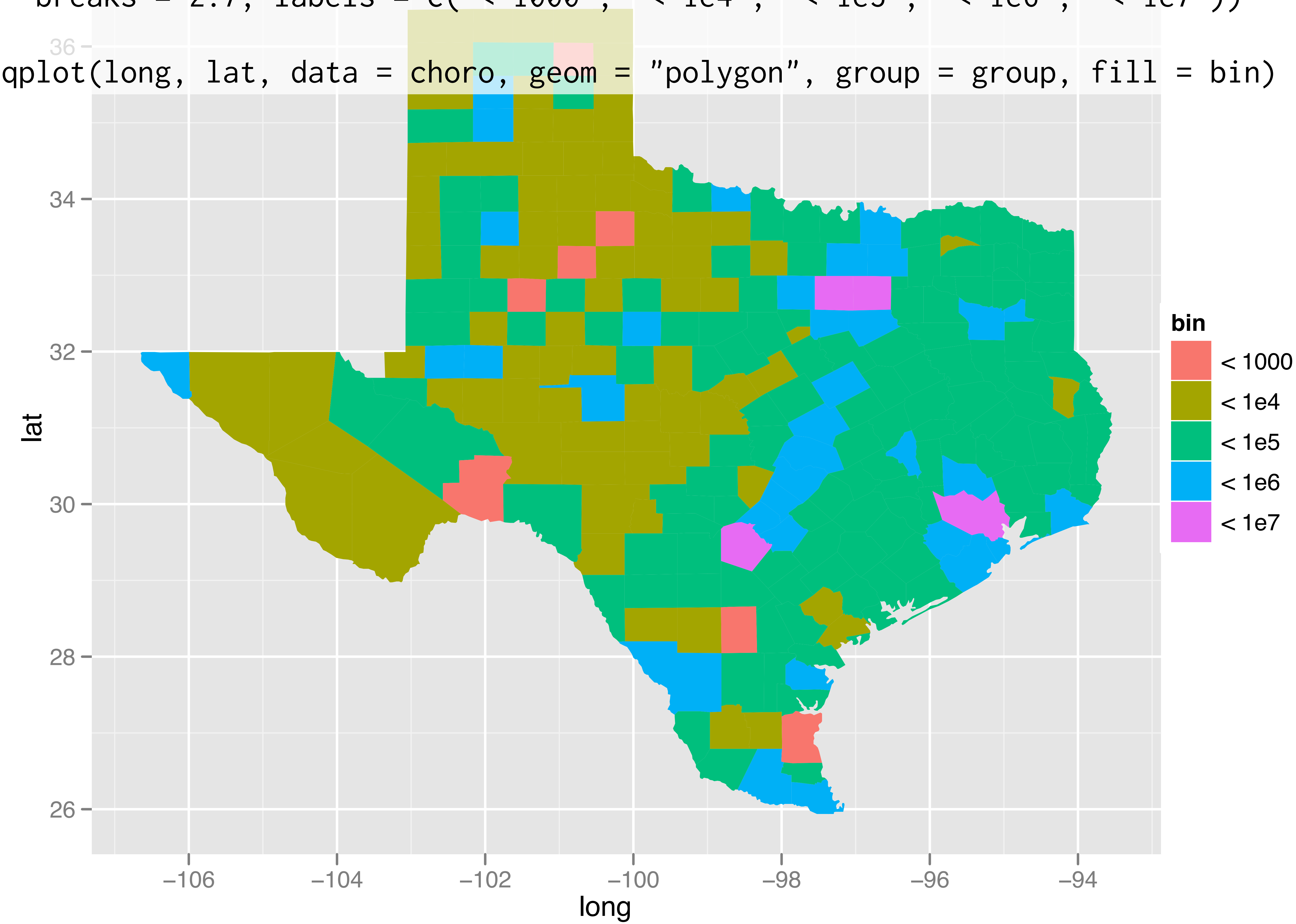
Cut

`cut`: If you know exactly what intervals you want

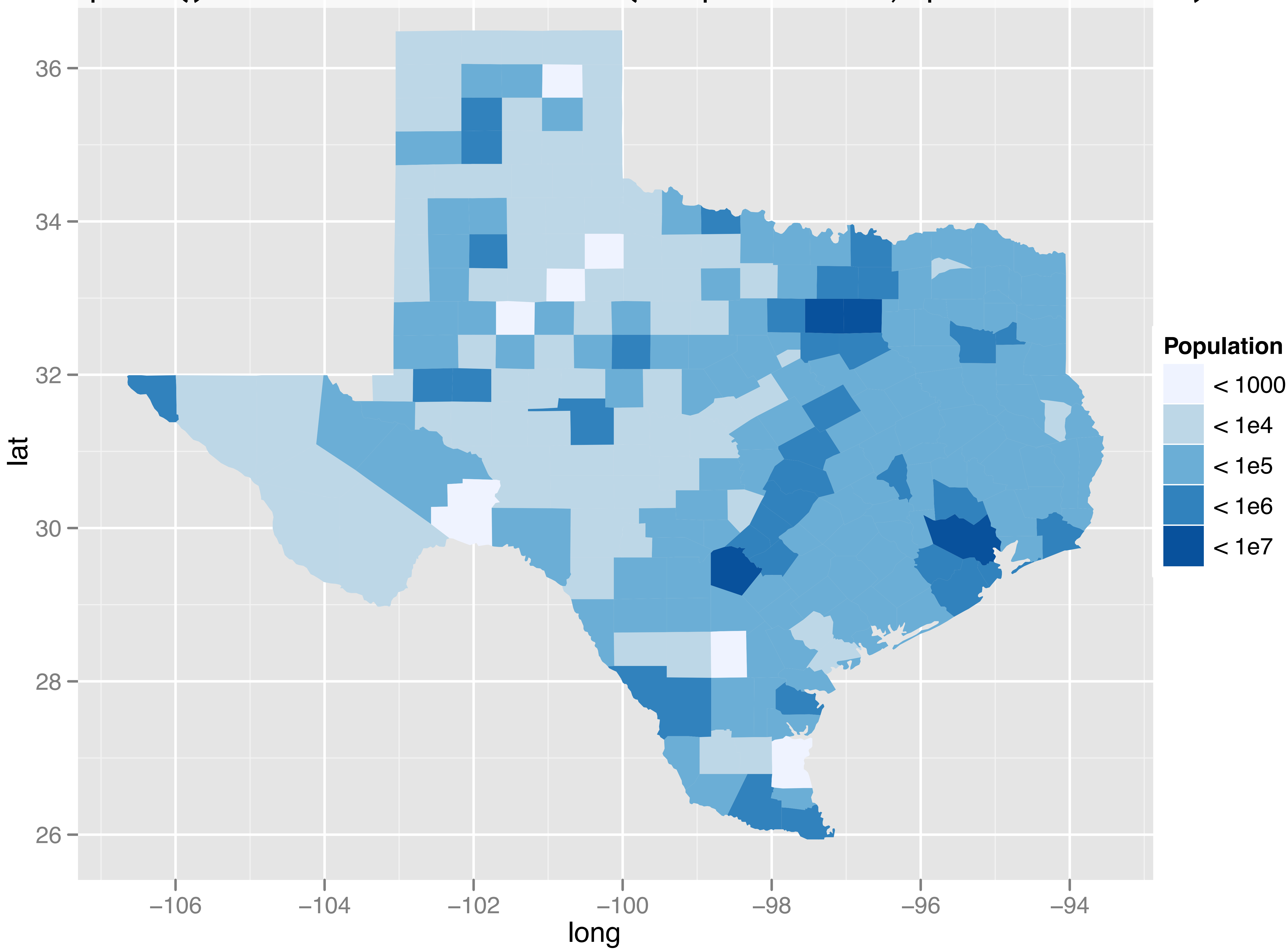
`cut_number`: to cut into n intervals each containing an equal **number of cases**.

`cut_interval`: to cut into n intervals of equal **length** (quantiles).

```
choro$bin <- cut(log10(choro$pop),
  breaks = 2:7, labels = c("< 1000", "< 1e4", "< 1e5", "< 1e6", "< 1e7"))

qplot(long, lat, data = choro, geom = "polygon", group = group, fill = bin)
```



```
last_plot() + scale_fill_brewer("Population", pal = "Blues")
```



ColorBrewer

http://colorbrewer2.org/

Helps to pick good colour scales for maps.

Your turn

Experiment with using `cut_interval` and `cut_number` to break up the population in different ways (on both original and log scales). Read the help if you get stuck.

Try out different colour brewer scales.

```
cancer <- subset(deaths, disease == "Neoplasms")
choro <- merge(borders, cancer)
choro <- choro[order(choro$order), ]

qplot(long, lat, data = choro, geom = "polygon",
  group = group, fill = cut_number(rate, 5)) +
  scale_fill_brewer(pal = "Blues")

# Can you find a disease with a strong spatial
# component?
```

TB Notifications

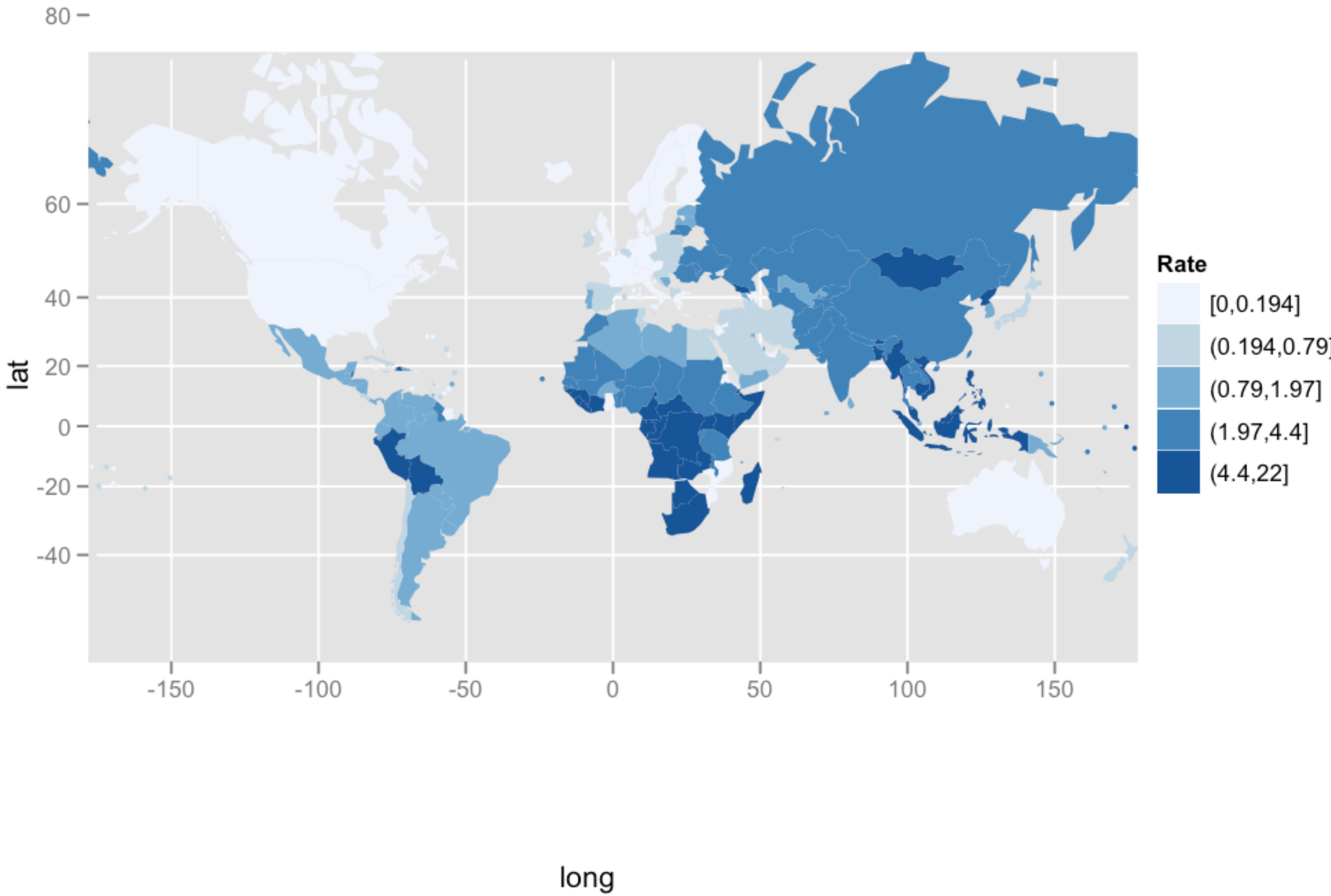
Number and rate of TB cases broken down by country and year (for all countries with at least 10 years of data).

Use with official WHO shape file.

Getting started

```
library(ggplot2)

tb <- read.csv("tb.csv")
tb2008 <- subset(tb, year == 2008)

boundaries <- read.csv("world-boundaries.csv")

choro <- merge(tb2008, worlddf, by = "iso2")
choro <- choro[order(choro$order), ]
```

```
qplot(long, lat, data = choro, fill = cut_number(rate, 5),
 geom = "polygon", group = group) + scale_fill_brewer("Rate", pal = "Blues")
```



Problems?

What is a big problem with this plot?
What is the problem with choropleth plots?

Take one minute to brainstorm some possible issues.

Problems

Big areas most striking. But in the US (as with most countries) big areas tend to least populated. Most populated areas tend to be small and dense - e.g. the East coast.

(Another computational problem: need to push around a lot of data to create these plots)

Alternative

Bubble (proportional symbol) maps.

Draw a point in the centre of each country, and map colour or size to rate.

To compute centres, figure out centroid of biggest piece of country. See `3-maps.r` for complete details

Aside

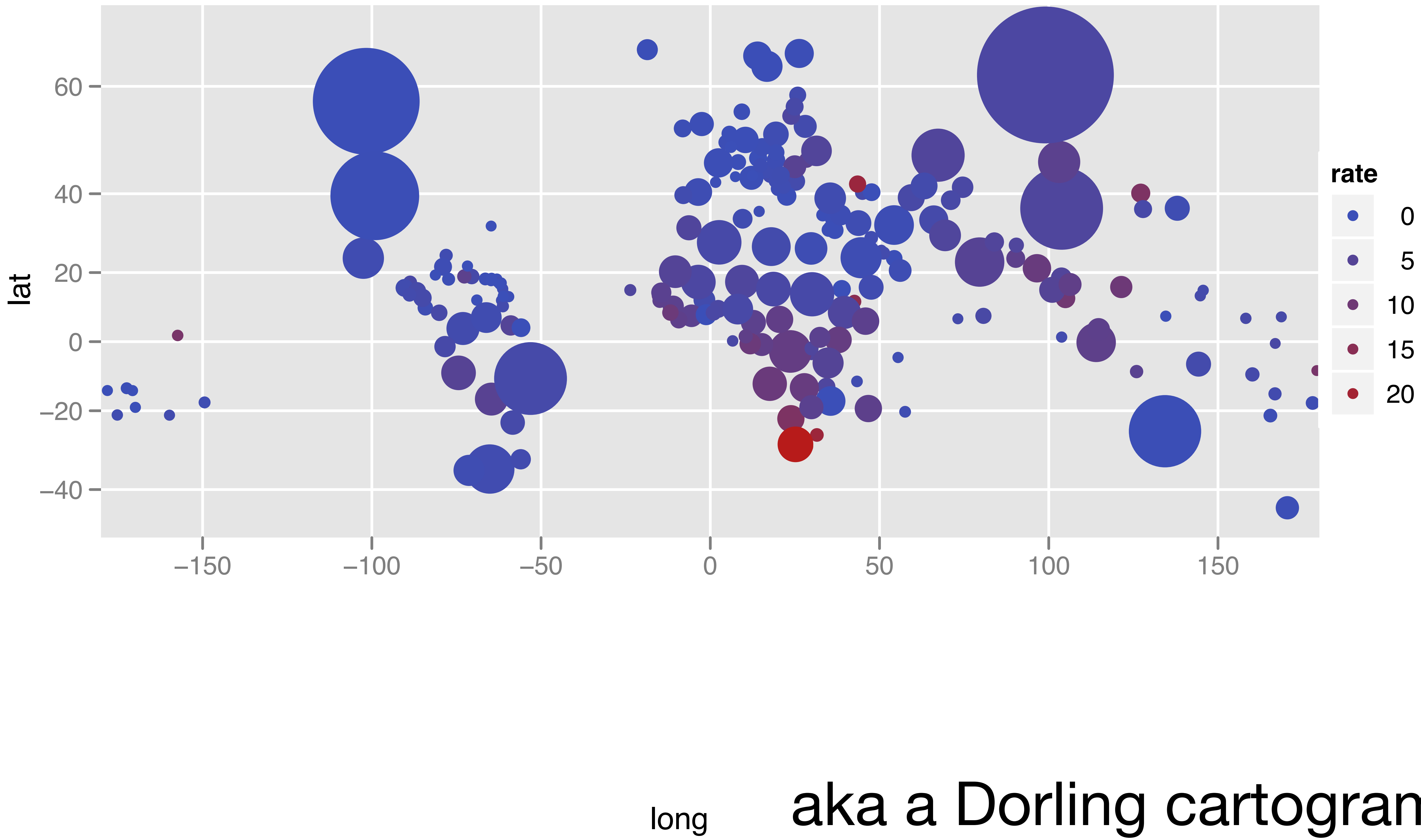
Notice anything missing?

You already know one easy solution to this (draw a background layer)

Getting started

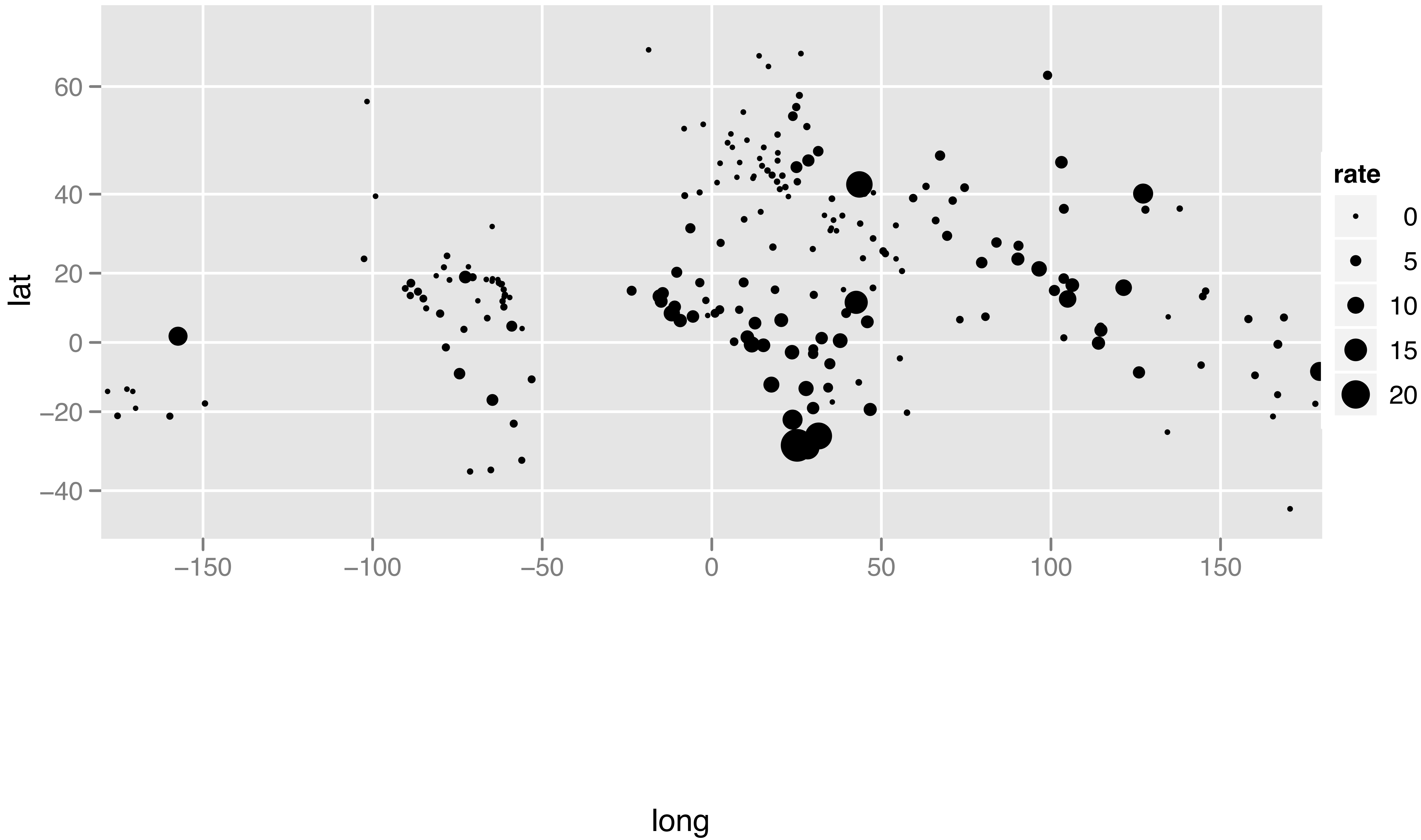
```
centres <- read.csv("world-centres.csv")
bubble <- merge(centres, tb2008m, by = "iso2")

world_coord <- coord_map(
  xlim = c(-180, 180),
  ylim = c(-50, 70))
```

```
qplot(long, lat, data = bubble, size = area, colour = rate) +
  scale_area(to = c(0.1, 25), legend = FALSE) +
  world_coord
```

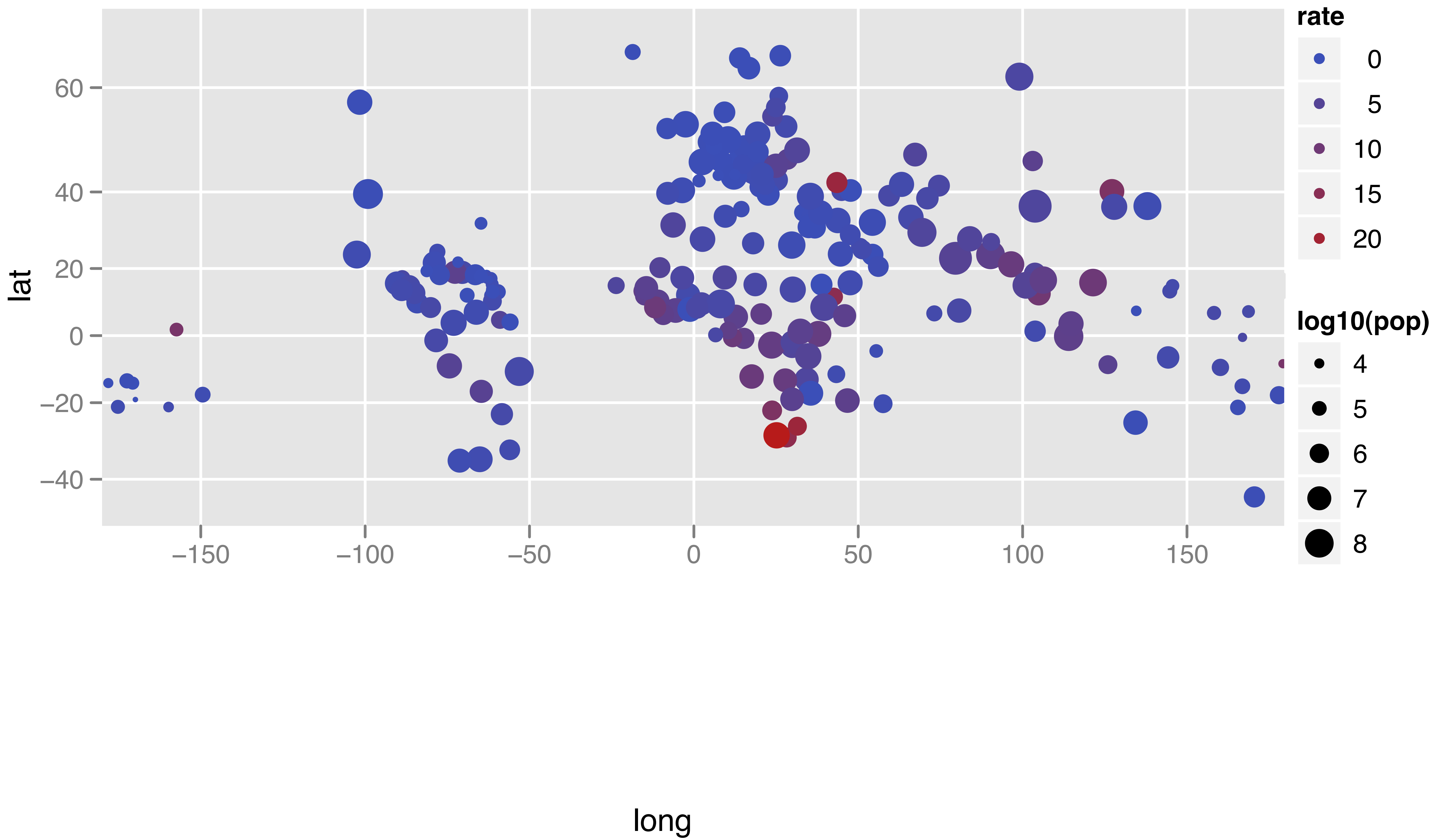


aka a Dorling cartogram

```
qplot(long, lat, data = bubble, size = rate) + world_coord
```



```
qplot(long, lat, data = bubble, size = log10(pop), colour = rate) +
  world_coord
```



Your turn

Practice recreating a bubble map. What other point aesthetics might you be able to use?

Think about the disadvantages of a bubble map.