Plotting Map Data

Eric Hare, Andee Kaplan, Sam Tyner

May 17, 2016

States Data

To make a map, let's load up the states data and take a look:

```
states <- map_data("state")
```

```
##
##  # maps v3.1: updated 'world': all lakes moved to separate new #
##  # 'lakes' database. Type '?world' or 'news(package="maps")'.  #
```

```
head(states)
```

```
##        long      lat group order  region subregion
## 1 -87.46201 30.38968     1     1 alabama      <NA>
## 2 -87.48493 30.37249     1     2 alabama      <NA>
## 3 -87.52503 30.37249     1     3 alabama      <NA>
## 4 -87.53076 30.33239     1     4 alabama      <NA>
## 5 -87.57087 30.32665     1     5 alabama      <NA>
## 6 -87.58806 30.32665     1     6 alabama      <NA>
```

Basic Map Data

What needs to be in the data set in order to plot a basic map?

- Need latitude/longitude points for all map boundaries
- Need to know which boundary group all lat/long points belong
- Need to know the order to connect points within each group

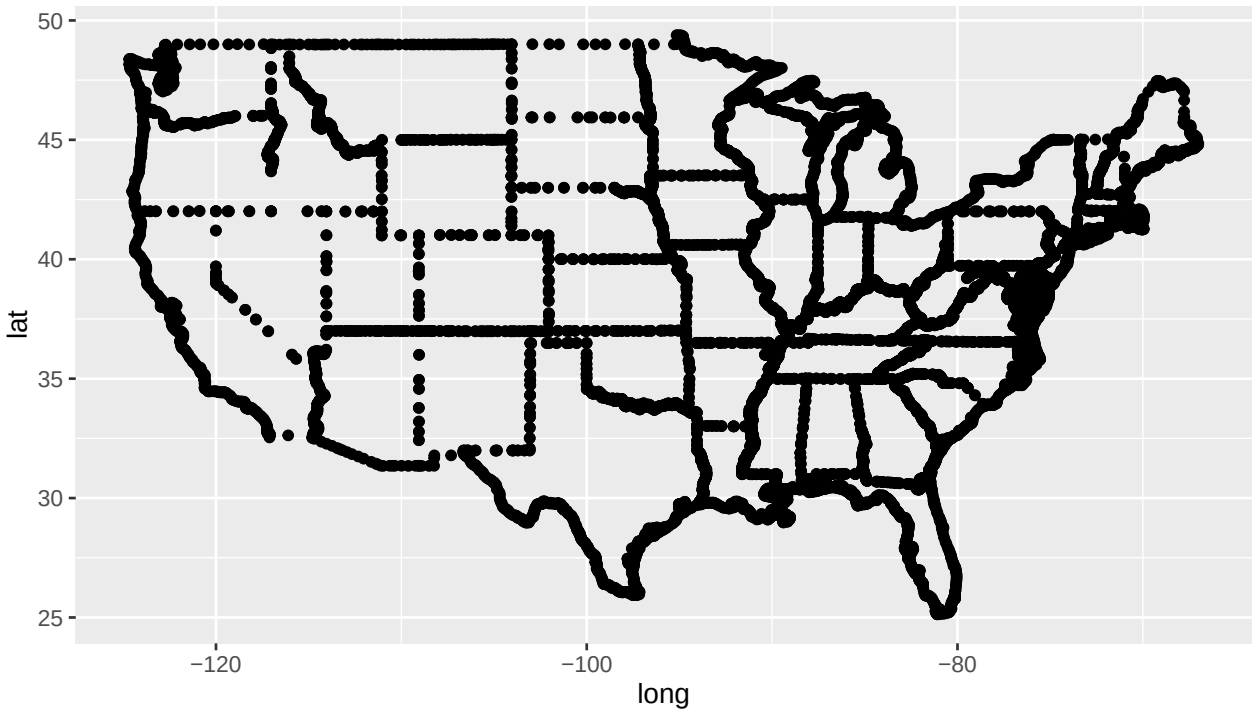
Data for Building Basic State Map

Our *states* data has all necessary information

A Basic (Rather Hideous) Map

A bunch of latitude longitude points...

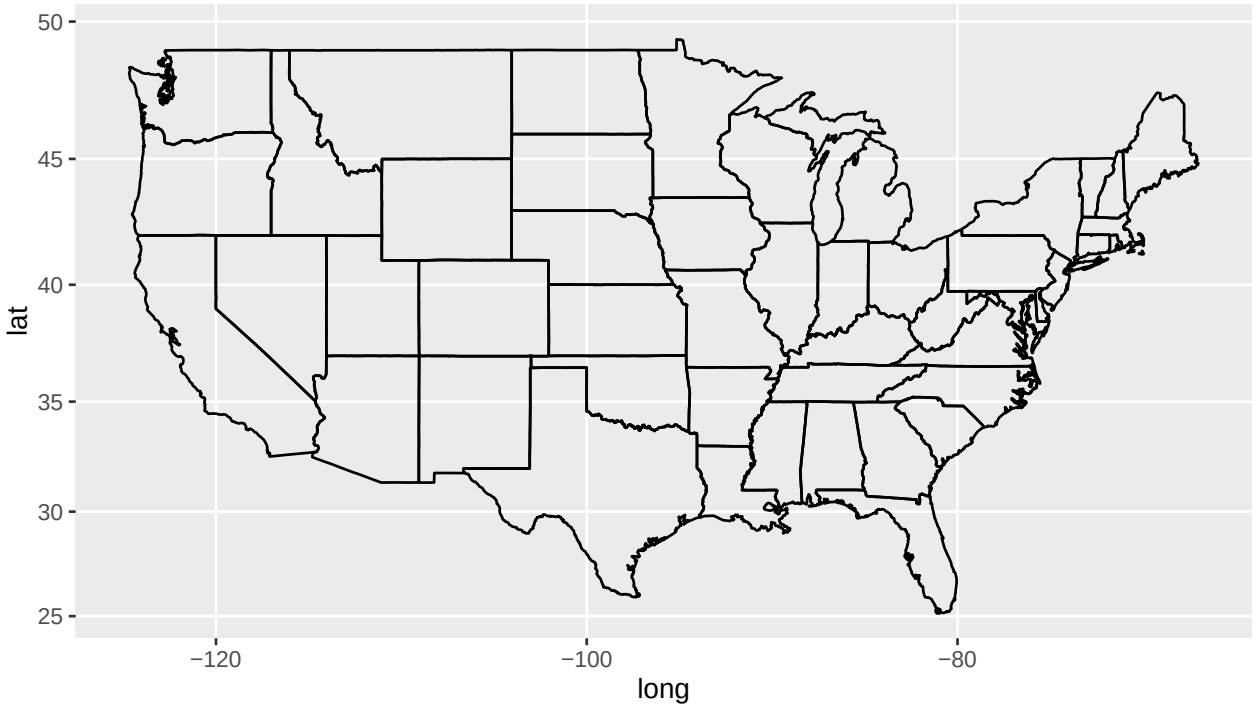
```
qplot(long, lat, geom = "point", data = states)
```



A Bit Better of a Map

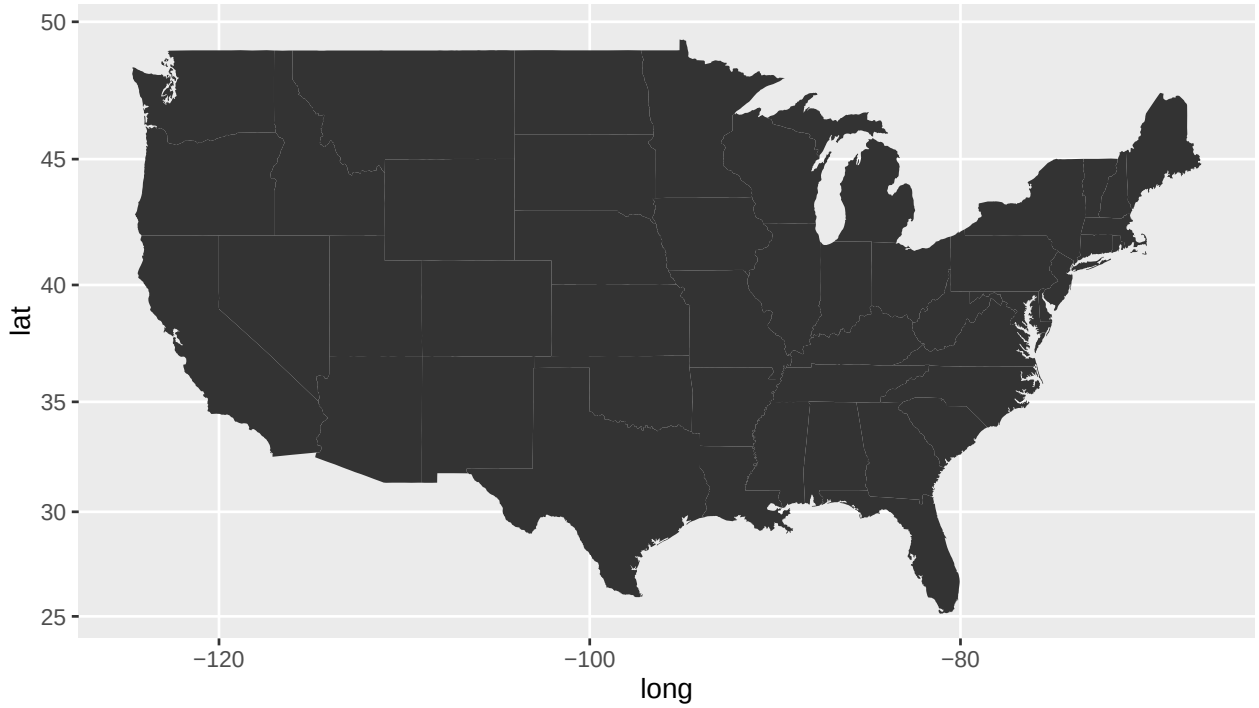
. . . that are connected with lines in a very specific order.

```
qplot(long, lat, geom = "path", data = states, group = group) +
    coord_map()
```



Polygon instead of Path

```
qplot(long, lat, geom = "polygon", data = states, group = group) +
    coord_map()
```



Incorporating Information About States

We want to incorporate additional information into the plot:

- Add other geographic information by adding geometric layers to the plot
- Add non-geopgraphic information by altering the fill color for each state
  - Use `geom = "polygon"` to treat states as solid shapes to add color
  - Incorporate numeric information using color shade or intensity
  - Incorporate categorical informaion using color hue

Categorical Information Using Hue

If a categorical variable is assigned as the fill color then `qplot` will assign different hues for each category. Let's load in a state regions dataset:

```
statereg <- read.csv("http://heike.github.io/rwrks/02-r-graphics/data/statereg.csv")

head(statereg)
```

```
##        State StateGroups
## 1 california        West
## 2     nevada        West
## 3     oregon        West
## 4 washington        West
## 5      idaho        West
```

```
## 6    montana        West
```

Joining Data

We need to `join` or `merge` our original states data with this new information on the regions. We can use the `left_join` function to do so (more about this later):

```
states.class.map <- left_join(states, statereg, by = c("region" = "State"))
```

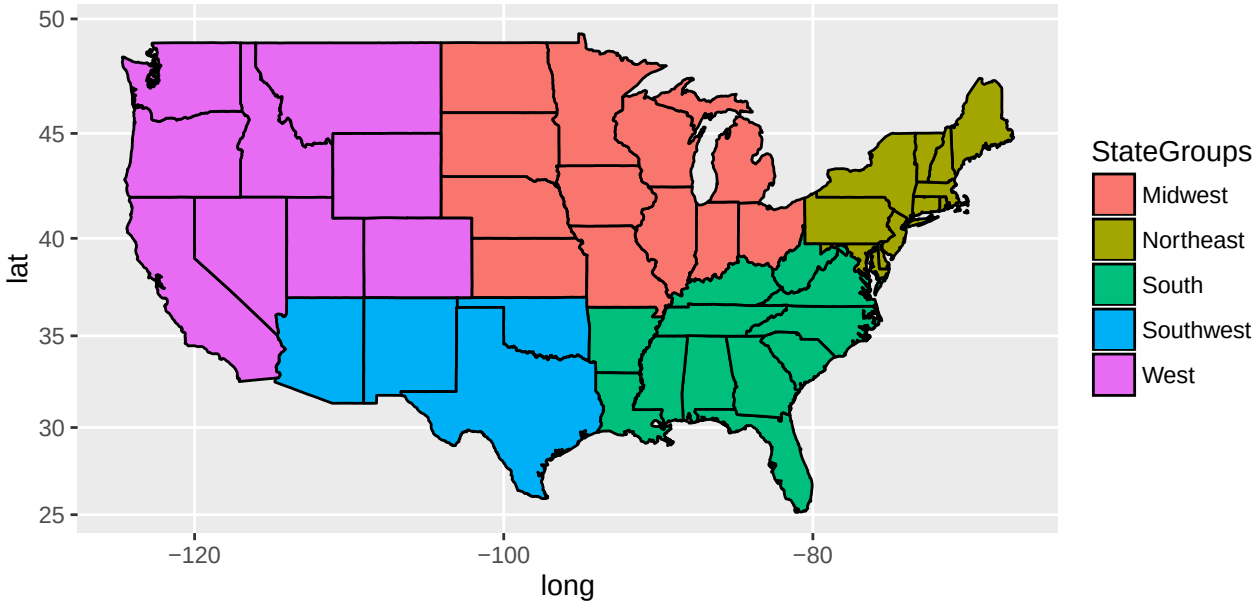
```
## Warning in left_join_impl(x, y, by$x, by$y): joining factor and character
## vector, coercing into character vector
```

```
head(states.class.map)
```

```
##        long      lat group order  region subregion StateGroups
## 1 -87.46201 30.38968     1     1 alabama      <NA>       South
## 2 -87.48493 30.37249     1     2 alabama      <NA>       South
## 3 -87.52503 30.37249     1     3 alabama      <NA>       South
## 4 -87.53076 30.33239     1     4 alabama      <NA>       South
## 5 -87.57087 30.32665     1     5 alabama      <NA>       South
## 6 -87.58806 30.32665     1     6 alabama      <NA>       South
```

Plotting the Result

```
qplot(long, lat, geom = "polygon", data = states.class.map,
      group = group, fill = StateGroups, colour = I("black")) +
    coord_map()
```



Numerical Information Using Shade and Intensity

To show how was can add numerical information to map plots we will use the BRFSS data

- Behavioral Risk Factor Surveillance System

- 2008 telephone survey run by the Center for Disease Control (CDC)
- Ask a variety of questions related to health and wellness
- Cleaned data with state aggregated values posted on website

BRFSS Data Aggregated by State

```
states.stats <- read.csv("http://heike.github.io/rwrks/02-r-graphics/data/states.stats.csv")
head(states.stats)
```

```
##   state.name   avg.wt avg.qlrest2   avg.ht  avg.bmi avg.drnk
## 1    alabama 180.7247    9.051282 168.0310 29.00222 2.333333
## 2     alaska 189.2756    8.380952 172.0992 28.90572 2.323529
## 3    arizona 169.6867    5.770492 168.2616 27.04900 2.406897
## 4   arkansas 177.3663    8.226619 168.7958 28.02310 2.312500
## 5 california 170.0464    6.847751 168.1314 27.23330 2.170000
## 6   colorado 167.1702    8.134715 169.6110 26.16552 1.970501
```

We must join this data again

```
states.map <- left_join(states, states.stats, by = c("region" = "state.name"))
```

```
## Warning in left_join_impl(x, y, by$x, by$y): joining factor and character
## vector, coercing into character vector
```

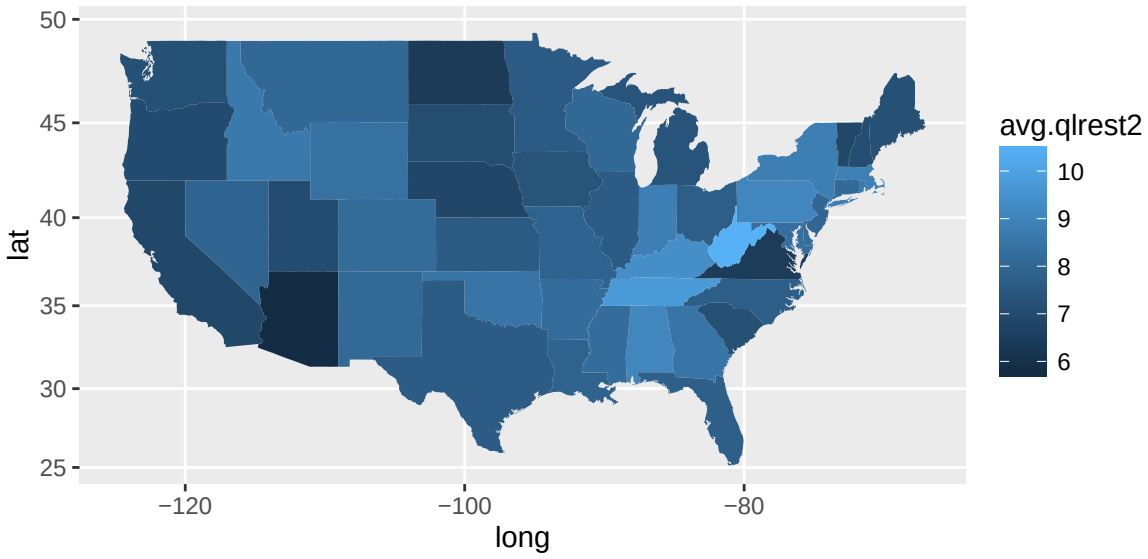
```
head(states.map)
```

```
##        long      lat group order  region subregion   avg.wt avg.qlrest2
## 1 -87.46201 30.38968     1     1 alabama      <NA> 180.7247    9.051282
## 2 -87.48493 30.37249     1     2 alabama      <NA> 180.7247    9.051282
## 3 -87.52503 30.37249     1     3 alabama      <NA> 180.7247    9.051282
## 4 -87.53076 30.33239     1     4 alabama      <NA> 180.7247    9.051282
## 5 -87.57087 30.32665     1     5 alabama      <NA> 180.7247    9.051282
## 6 -87.58806 30.32665     1     6 alabama      <NA> 180.7247    9.051282
##    avg.ht  avg.bmi avg.drnk
## 1 168.031 29.00222 2.333333
## 2 168.031 29.00222 2.333333
## 3 168.031 29.00222 2.333333
## 4 168.031 29.00222 2.333333
## 5 168.031 29.00222 2.333333
## 6 168.031 29.00222 2.333333
```

Numerical Information Using Shade and Intensity

Average number of days in the last 30 days of insufficient sleep by state

```
qplot(long, lat, geom = "polygon", data = states.map,
      group = group, fill = avg.qlrest2) + coord_map()
```



BRFSS Data Aggregated by State

```
states.sex.stats <- read.csv("http://heike.github.io/rwrks/02-r-graphics/data/states.sex.stats.csv")
head(states.sex.stats)
```

```
##   state.name SEX   avg.wt avg.qlrest2   avg.ht  avg.bmi avg.drnk    sex
## 1    alabama   1 198.8936    8.648936 177.5729 28.50714 3.033333   Male
## 2    alabama   2 173.0315    9.224771 163.9956 29.21280 2.041667 Female
## 3     alaska   1 203.3919    7.236111 178.3896 28.91494 2.487179   Male
## 4     alaska   2 169.5660    9.907407 163.1296 28.89286 2.103448 Female
## 5    arizona   1 191.3739    5.163793 177.1724 27.63152 2.814286   Male
## 6    arizona   2 156.2054    6.142857 162.7043 26.67683 2.026667 Female
```

One More Join

```
states.sex.map <- left_join(states, states.sex.stats, by = c("region" = "state.name"))
```

```
## Warning in left_join_impl(x, y, by$x, by$y): joining factor and character
## vector, coercing into character vector
```

```
head(states.sex.map)
```

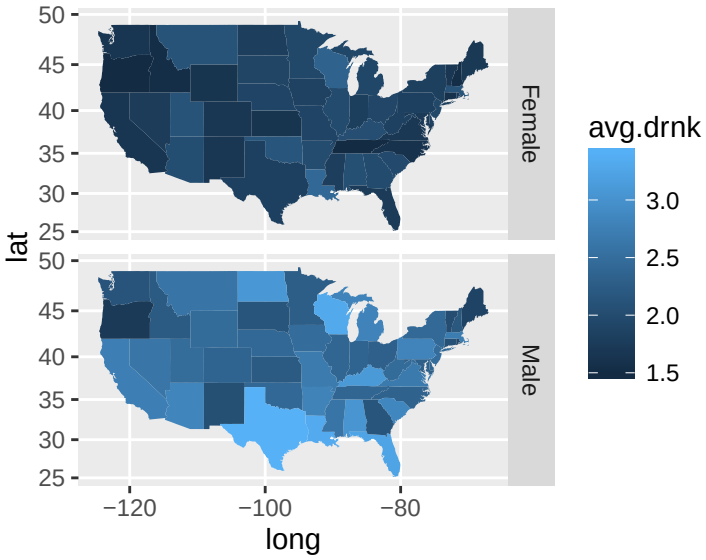
```
##        long      lat group order  region subregion SEX   avg.wt
## 1 -87.46201 30.38968     1     1 alabama      <NA>   1 198.8936
## 2 -87.46201 30.38968     1     1 alabama      <NA>   2 173.0315
## 3 -87.48493 30.37249     1     2 alabama      <NA>   1 198.8936
## 4 -87.48493 30.37249     1     2 alabama      <NA>   2 173.0315
## 5 -87.52503 30.37249     1     3 alabama      <NA>   1 198.8936
## 6 -87.52503 30.37249     1     3 alabama      <NA>   2 173.0315
##   avg.qlrest2   avg.ht  avg.bmi avg.drnk    sex
## 1    8.648936 177.5729 28.50714 3.033333   Male
## 2    9.224771 163.9956 29.21280 2.041667 Female
## 3    8.648936 177.5729 28.50714 3.033333   Male
## 4    9.224771 163.9956 29.21280 2.041667 Female
```

```
## 5    8.648936 177.5729 28.50714 3.033333   Male
## 6    9.224771 163.9956 29.21280 2.041667 Female
```

Adding Numerical Information

Average number of alcoholic drinks per day by state and gender

```
qplot(long, lat, geom = "polygon", data = states.sex.map,
      group = group, fill = avg.drnk) + coord_map() +
    facet_grid(sex ~ .)
```



Your Turn

- Use `left_join` to combine child healthcare data with maps information. You can load in the child healthcare data with:

```
states.health.stats <- read.csv("http://heike.github.io/rwrks/02-r-graphics/data/states.health.stats.cs
```

- Use `qplot` to create a map of child healthcare undercoverage rate by state

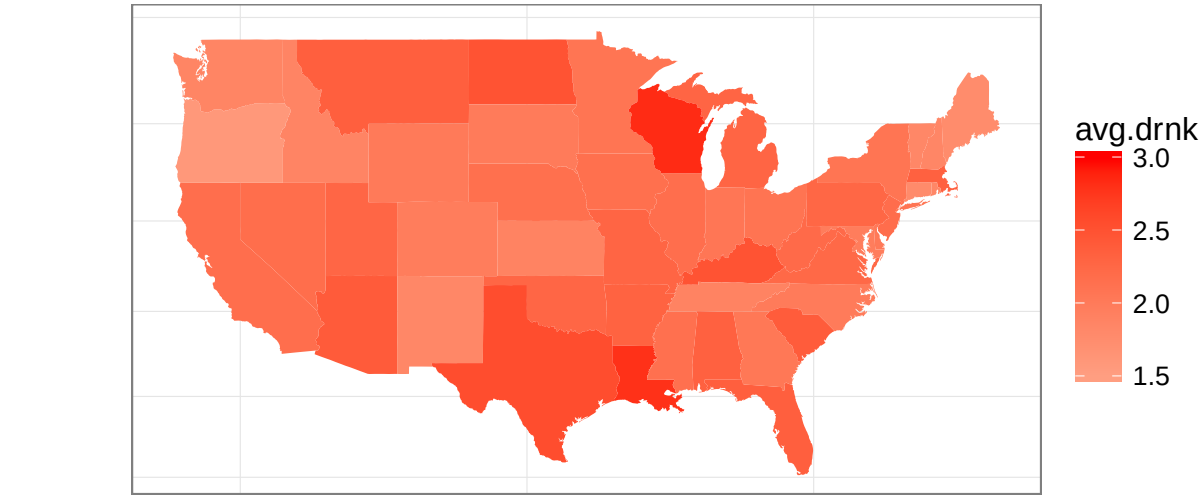
Cleaning Up Your Maps

Use ggplot2 options to clean up your map!

- Adding Titles `+ ggtitle(...)`
- Might want a plain white background `+ theme_bw()`
- Extremely familiar geography may eliminate need for latitude and longitude axes `+ theme(...)`
- Want to customize color gradient `+ scale_fill_gradient2(...)`
- Keep aspect ratios correct `+ coord_map()`

Cleaned Up Map

```
qplot(long, lat, geom="polygon", data = states.map, group = group, fill = avg.drnk) +
  coord_map() +  theme_bw() +
  scale_fill_gradient2(limits = c(1.5, 3),low = "lightgray", high = "red") +
  theme(axis.ticks = element_blank(),
        axis.text.x = element_blank(),
        axis.title.x = element_blank(),
        axis.text.y = element_blank(),
        axis.title.y = element_blank()) +
  ggtitle("Map of Average Number of Alcoholic Beverages Consumed Per Day by State")
```



Your Turn

Use options to polish the look of your map of child healthcare undercoverage rate by state!