Introduction to Choropleths

Data Wrangling and Husbandry

03/30/2020

Choropleth Maps

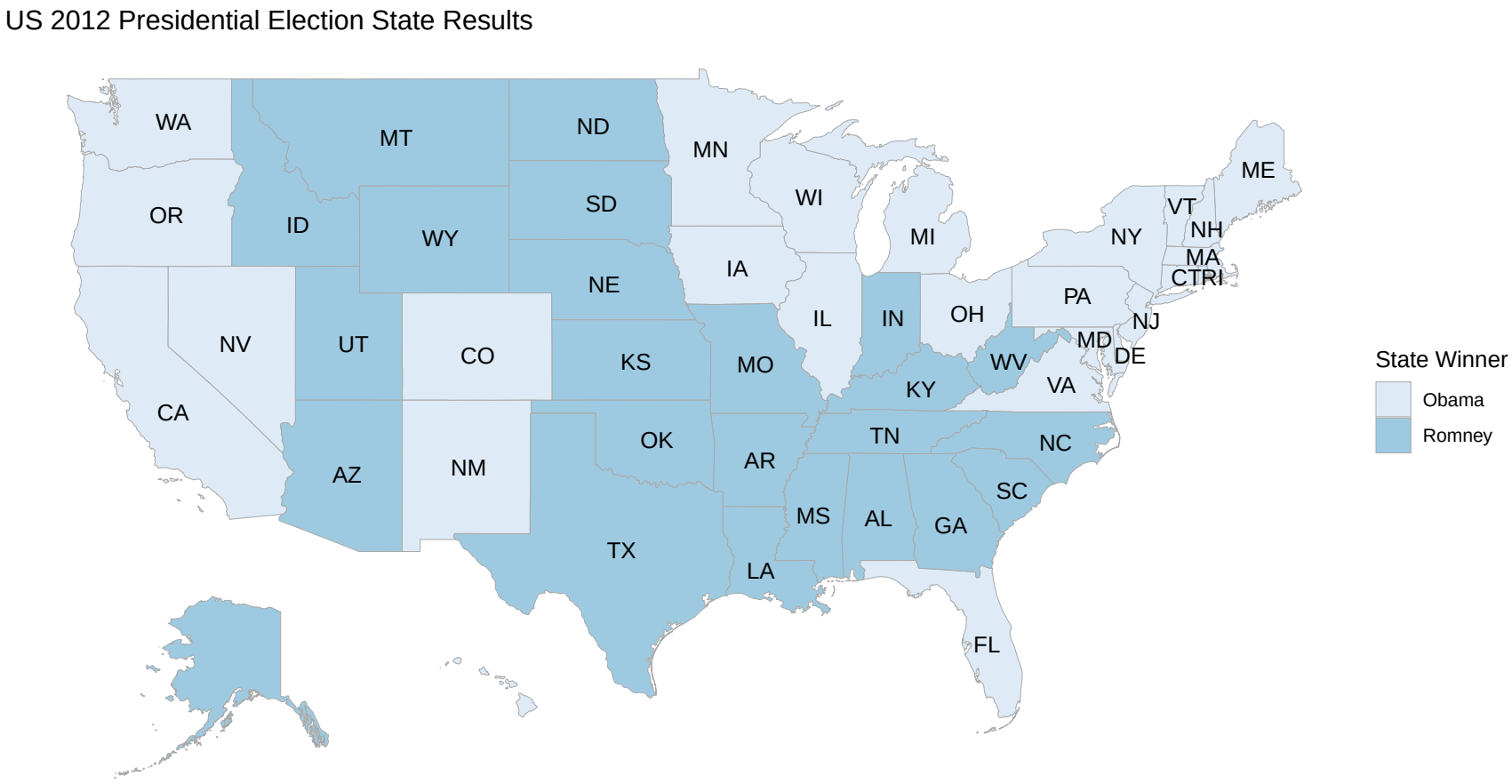
A choropleth map is a map in which areas are shaded relative to a measurement such as population density. They are widely used, and quite useful, but one must be careful because

- regionally aggregated data can be misleading
- the use of counts or sums to represent magnitude rather than normalized values such as rates leads the eye giving incorrect weights to larger regions

Because of this, dasymetric maps are increasing popular. Nonetheless, choropleths are quite useful and have a nice implementation in R in the `choroplethr` package.

Here's a very simple example, from the help file

```
library(choroplethr)
data(df_president)
state_choropleth(df_president,
                 title  = "US 2012 Presidential Election St
                 legend = "State Winner")
```



- The only requirement is that the dataframe (`df_president` in this case) have two columns, `region` and `value`. The `region` value must exactly match the names of the states in the `state.map` dataframe (full lower case name).
- This requirement of the `choroplethr` functions (that the input is a data frame with specified names) makes the use of pipes an obvious choice

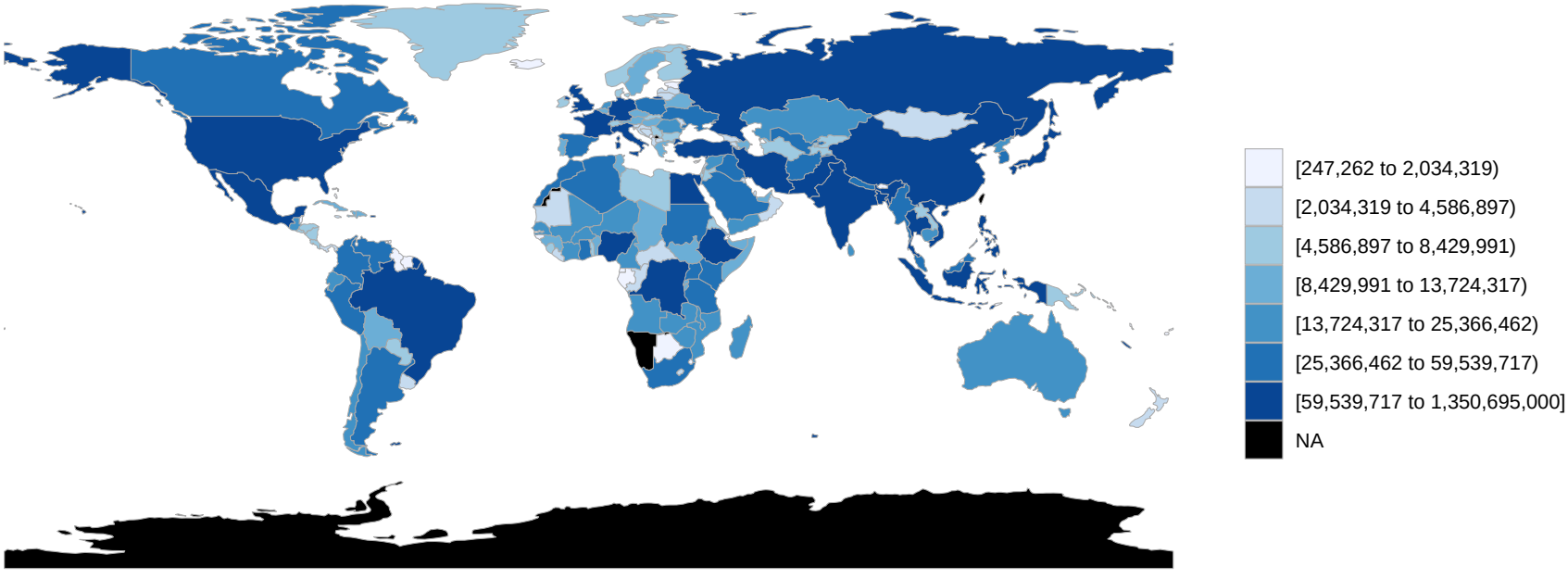
```
head(df_president)
```

```
##       region  value
## 1    alabama Romney
## 2     alaska Romney
## 3    arizona Romney
## 4   arkansas Romney
## 5 california  Obama
## 6   colorado  Obama
```

There are also `country_choropleth()`, `admin1_choropleth()` (states or provinces), and `county_choropleth()`.

```
data(df_pop_country)
country_choropleth(df_pop_country)
```

```
## Warning in self$bind(): The following regions were missi
## NA: namibia, western sahara, taiwan, antarctica, kosovo
```



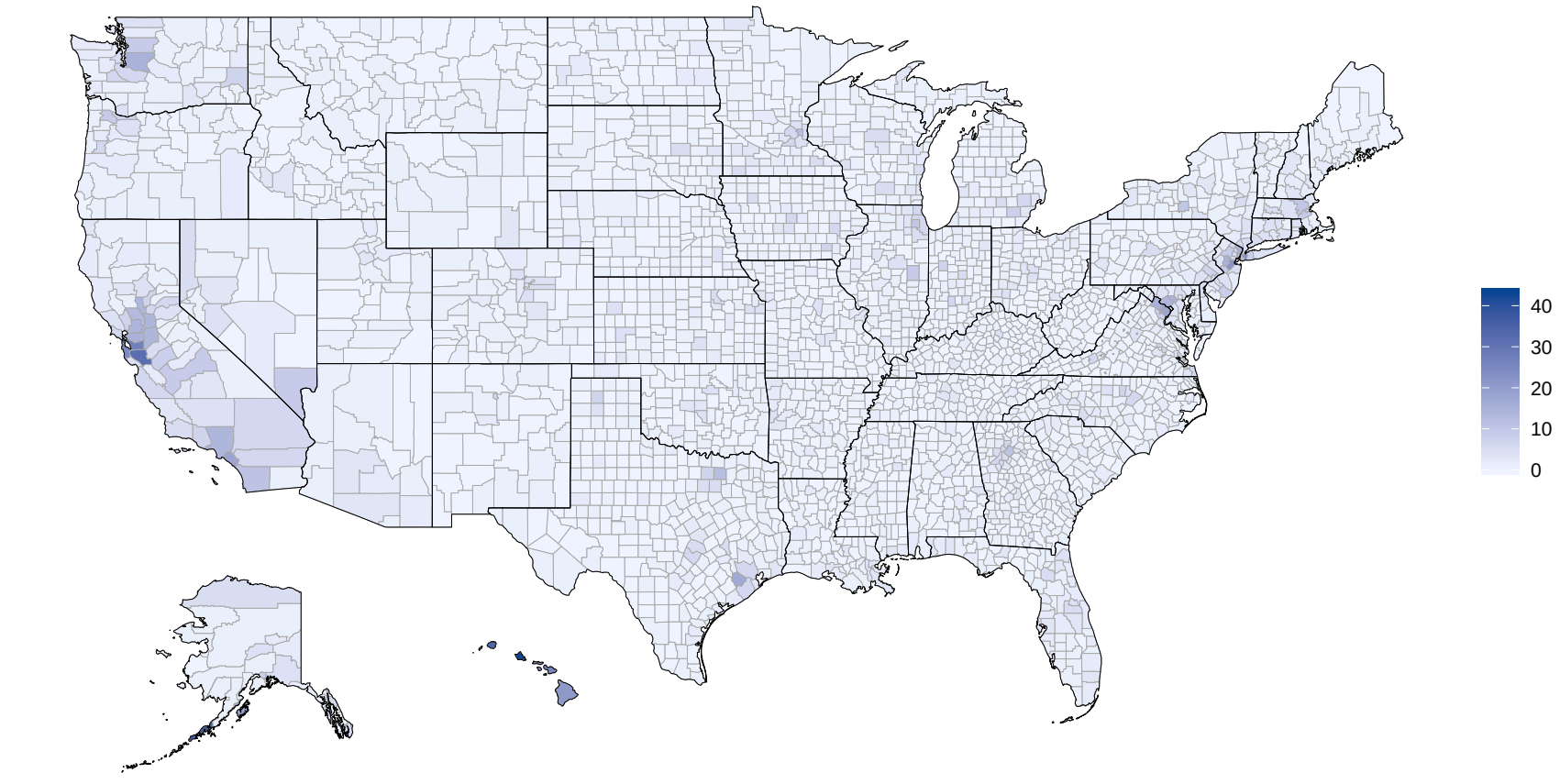
We'll work through an example using `county_choropleth()`.

```
data(df_county_demographics)
head(df_county_demographics, n = 2)
```

```
##   region total_population percent_white percent_black pe
## 1   1001             54907            76            18
## 2   1003            187114            83             9
##   percent_hispanic per_capita_income median_rent median_
## 1                2              24571         668       3
## 2                4              26766         693       4
```

Notice that counties are given by codes, not names. They are known as county FIPS codes

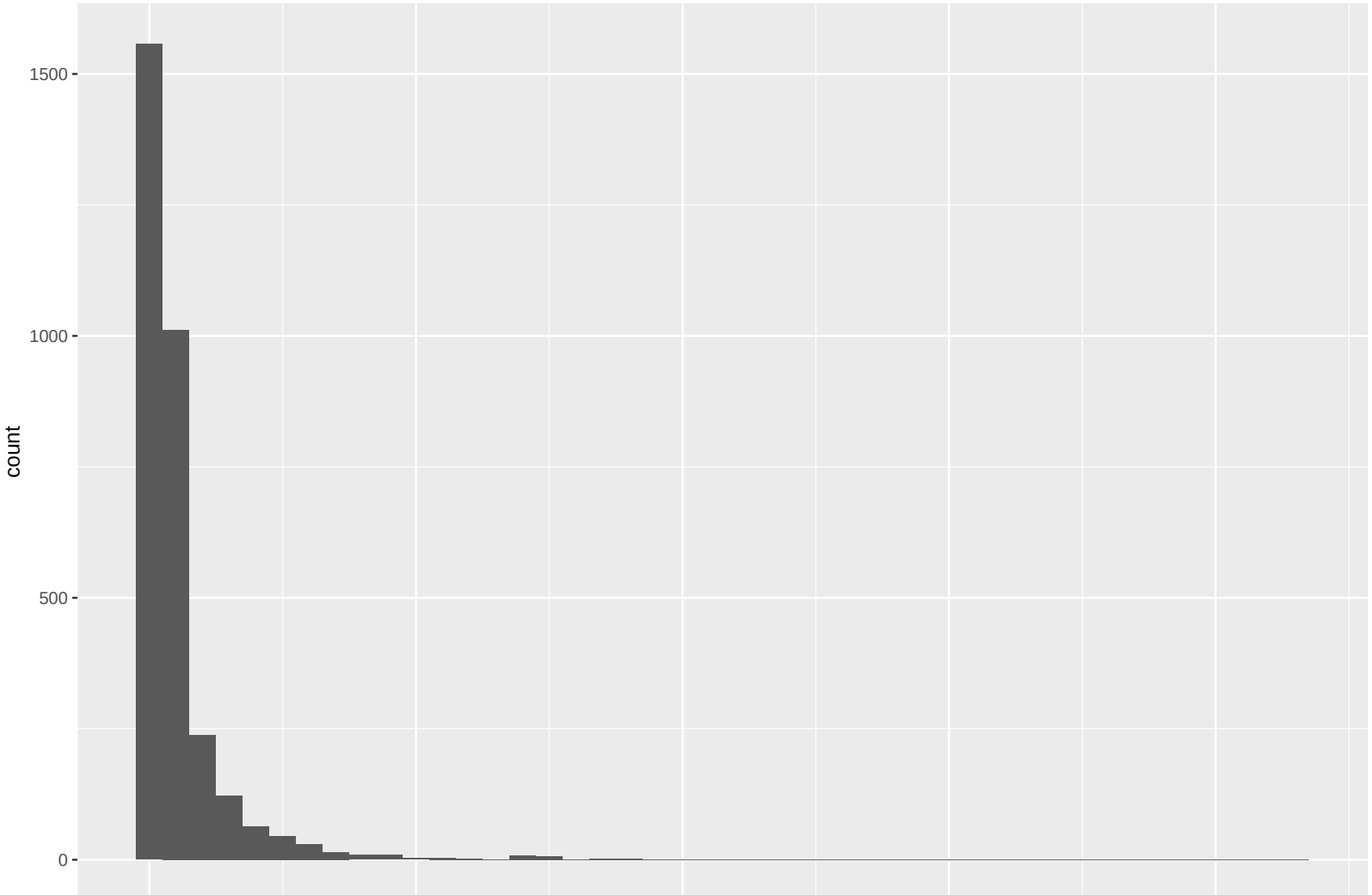
```
df_county_demographics %>%
  dplyr::rename(value = percent_asian) %>%
  county_choropleth(title = "Asian Percentage of Populatio
```




```
# the num_colors = 1 option uses a continuous color palett
```

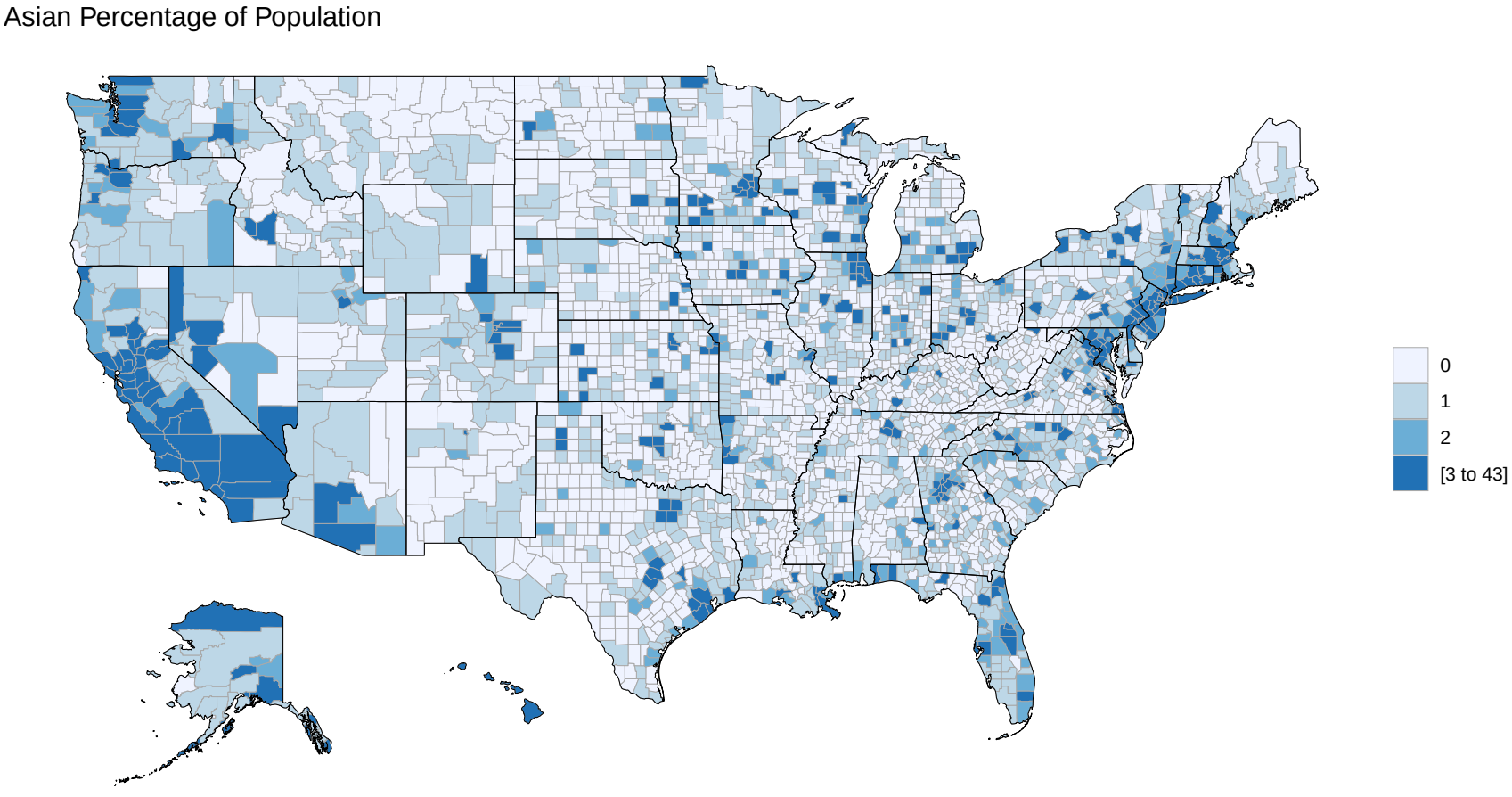
The color scheme in the map is so-so. There are a few counties with high proportions, leading the scale to be evenly distributed from 0 to 43, but almost all of the counties are in a very tight range.

```
ggplot(df_county_demographics, aes(percent_asian)) + geom_h
```



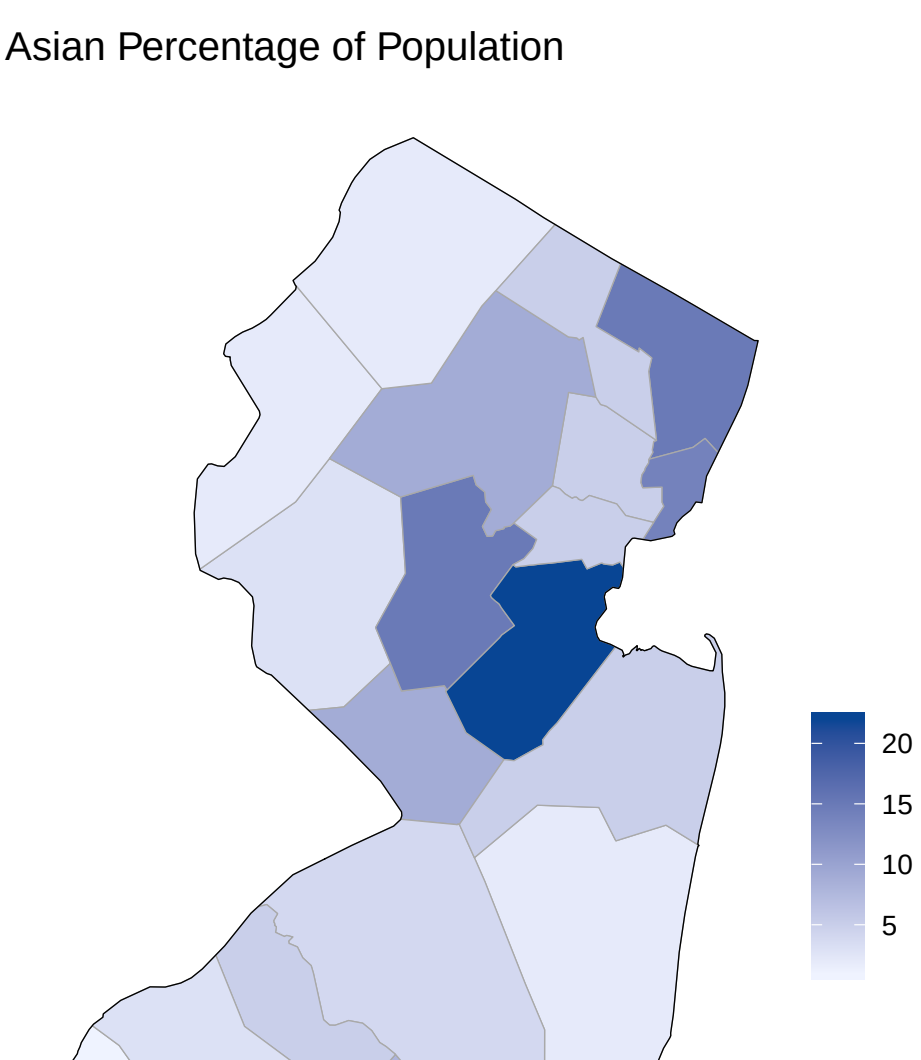
In this situation, it can be helpful to use a fixed number of colors which forces the choropleth function to bin the values.

```
df_county_demographics %>%
  dplyr::rename(value = percent_asian) %>%
  county_choropleth(title = "Asian Percentage of Populatio
```



Let's look at just NJ. Note the `state_zoom` option.

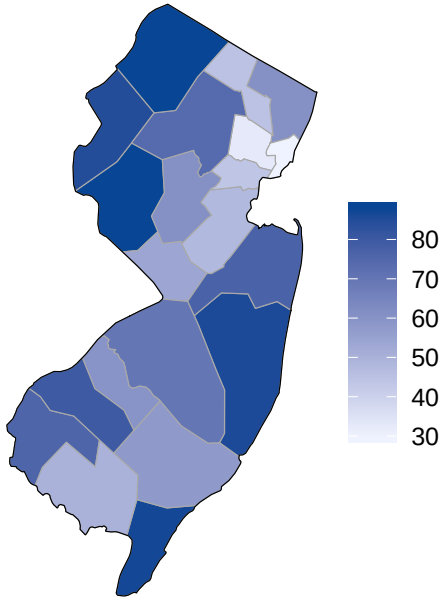
```
df_county_demographics %>%
  dplyr::rename(value = percent_asian) %>%
  county_choropleth(title = "Asian Percentage of Populatio
                    num_colors = 1, state_zoom = "new jers
```



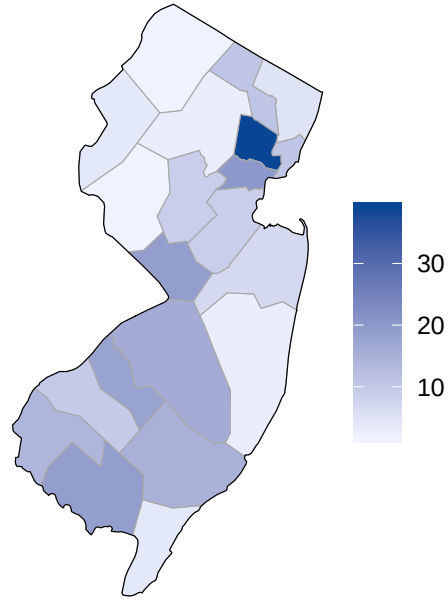
```r
library(gridExtra)
grid.arrange(
  df_county_demographics %>% dplyr::rename(value = percent_
  df_county_demographics %>% dplyr::rename(value = percent_
  df_county_demographics %>% dplyr::rename(value = percent_
  df_county_demographics %>% dplyr::rename(value = percent_
  ncol=2, nrow=2
)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:acs':
##
##     combine

## The following object is masked from 'package:dplyr':
##
##     combine
```

White Percentage of Population



African American Percentage of Population



Asian Percentage of Population

Hispanic Percentage of Population

It's possible to get finer control, but it requires some custom coding. (We'll see some other approaches shortly.)

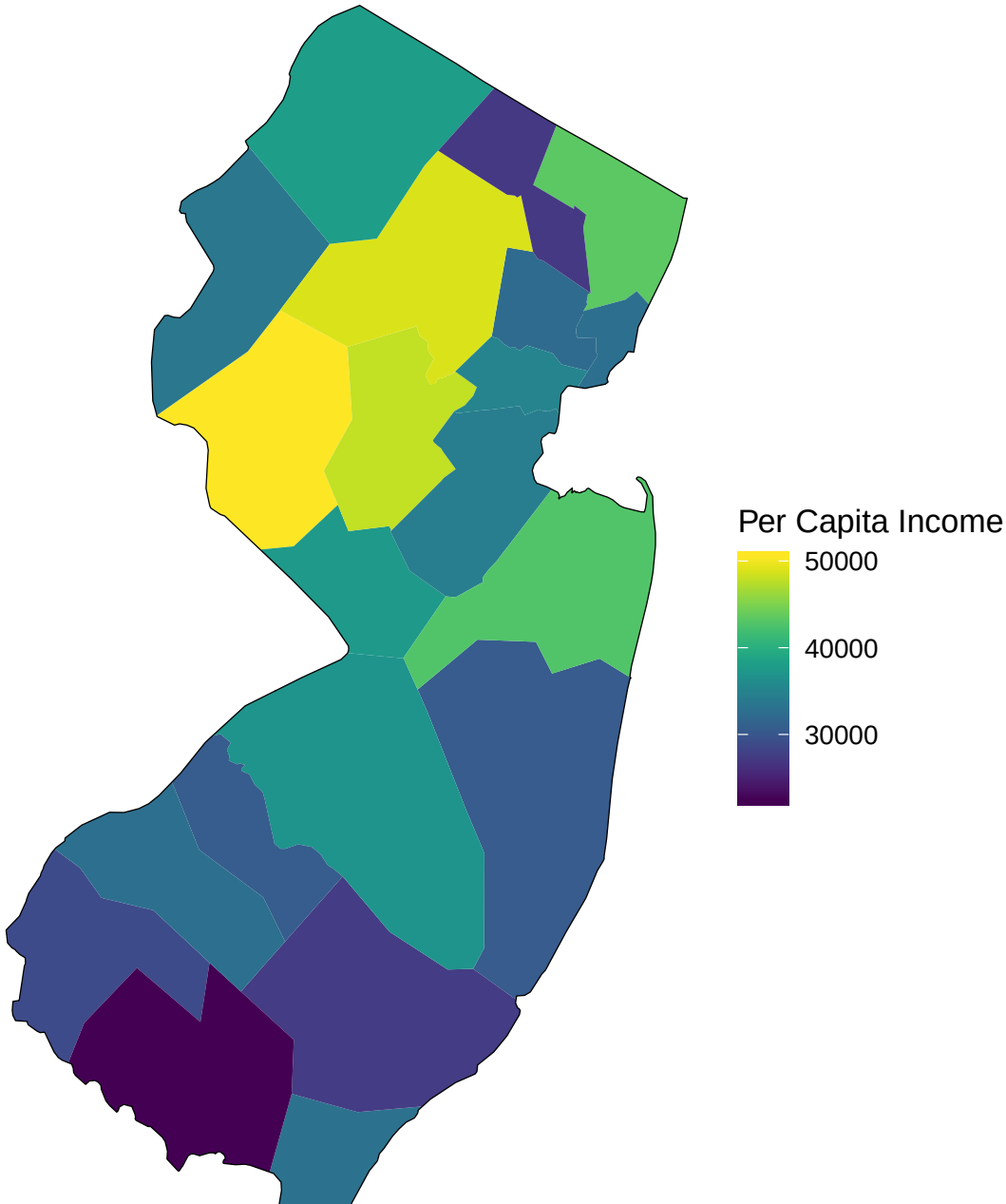
```
library(viridis)
```

```
## Loading required package: viridisLite
```

```
temp_demo <- df_county_demographics %>%
  mutate(value = per_capita_income)
custom_choro <- CountyChoropleth$new(temp_demo)
custom_choro$title <- "Per Capita Income By County"
custom_choro$set_num_colors(1)
custom_choro$ggplot_polygon <- geom_polygon(aes(fill = valu
custom_choro$ggplot_scale <- scale_fill_gradientn(name = "I
custom_choro$set_zoom("new jersey")
```

```
custom_choro$render()
```

Per Capita Income By County



Note that for county-level data, the values for region are codes, not names. The are FIPS county codes. The first two digits correspond to the state, and are in fact the FIPS state codes.

```
tail(df_county_demographics)
```

```
##      region total_population percent_white percent_black
## 3138  56035             10178            86             0
## 3139  56037             44437            80             0
## 3140  56039             21575            82             0
## 3141  56041             21040            88             0
## 3142  56043              8468            83             0
## 3143  56045              7153            94             0
##      percent_hispanic per_capita_income median_rent medi
## 3138                7              35944         979
## 3139               15              30517         789
## 3140               15              43444        1001
## 3141                9              25513         532
## 3142               14              28308         394
## 3143                3              28764         551
```

In class exercise

1. Download the SPSS file "U.S. Religion Census Religious Congregations and Membership Study, 2010 (County File).SAV" from the class dropbox folder and read it into your R session. (There is a codebook available.)

2. The census contains information on adherents to 236 religious groups but also includes a few aggregate values. The county variable is called FIPS. Try making national and NJ choropleth maps of TOTRATE, the total rate of adherence (religious attendence per 1000 residents) and one of the aggregate rates:

- Evangelical Protestant–Rates of adherence per 1,000 population (2010) (EVANRATE)
- Black Protestant–Rates of adherence per 1,000 population (2010) (BPRTRATE)
- Mainline Protestant–Rates of adherence per 1,000 population (2010) (MPRTRATE)
- Catholic–Rates of adherence per 1,000 population (2010) (CATHRATE)