Example Markdown

Andrew Moore

2015-11-10

Plotting the `midwest` data-set

`midwest` is a data-set contained in the `ggplot2` library. It contains demographics on counties in the midwest, such as *college-achievement*, *income*, and *population density*. I'll be making some visualizations at the county-level eventually, and I'd like to get used to working with this kind of data.

The data can be retrieved after the `ggplot2` package is loaded. The code-chunk below sets up my workspace. We'll start by taking a quick look at the data.

```r
# load the R packages I want to use
library(ggplot2)
library(dplyr)
library(knitr)
library(printr)

# access the dataset stored in R
data(midwest)

# pull out only Michigan's data
midwest <- filter(midwest, state == "MI")

# open up the data, see how it's formatted and look at the first 10 columns
glimpse(midwest[, 1:10])
```

```
## Observations: 83
## Variables: 10
## $ PID          (int) 1197, 1198, 1199, 1200, 1201, 1202, 1203, 1204, ...
## $ county       (chr) "ALCONA", "ALGER", "ALLEGAN", "ALPENA", "ANTRIM"...
## $ state        (chr) "MI", "MI", "MI", "MI", "MI", "MI", "MI", "MI", ...
## $ area         (dbl) 0.041, 0.051, 0.049, 0.034, 0.031, 0.021, 0.054,...
## $ poptotal     (int) 10145, 8972, 90509, 30605, 18185, 14931, 7954, 5...
## $ popdensity   (dbl) 247.4390, 175.9216, 1847.1225, 900.1471, 586.612...
## $ popwhite     (int) 10026, 8422, 86760, 30372, 17895, 14695, 6971, 4...
## $ popblack     (int) 27, 213, 1448, 35, 23, 10, 49, 104, 1242, 30, 24...
## $ popamerindian (int) 56, 304, 543, 93, 211, 139, 918, 188, 726, 237, ...
## $ popasian     (int) 26, 24, 411, 85, 24, 38, 10, 144, 428, 35, 1487,...
```

The data look tidy, everything seems formatted well; I can probably proceed with some visualizations. But maybe we should start with a summary table.

```r
# create a new data-frame, summarizing data by metro/non-metro counties
my_summary <- midwest %>%
  group_by(inmetro) %>%
  summarize(
    avg_density    = mean(popdensity),
    med_density    = median(popdensity),
    avg_perc_white = mean(percwhite)
  )

my_summary %>%
  kable(col.names = c("Metro/Non-Metro", "Average density", "Median density",
                      "Mean % of white individuals"),
        caption = "Density and % Caucasian, by metropolitan vs non-metropolitan counties",
        digits = 2)
```

Table 1: Density and % Caucasian, by metropolitan vs non-metropolitan counties

Metro/Non-Metro	Average density	Median density	Mean % of white individuals
0	743.79	627.21	96.53
1	8270.29	4297.04	89.61

Metro counties have a higher level of non-white residents (89.61%).