Data Visualization with ggplot2

Anthony Chau

UCI Center for Statistical Consulting

7/24/2021 (updated: 2021-07-24)

What is ggplot2?

- `ggplot2` is a R package for creating statistical and data graphics
- `ggplot2`'s approach to graphics is based on The Grammar of Graphics
- Mature (14 years old) and popular package
- Powerful and extensible

Grammar of Graphics

- Big idea: a visualization is constructed from many independent components
- We put together different components to create our desired visualization
- Components of a plot:
  - Data
  - Aesthetic mappings
  - Geometric objects
  - Scales
  - Facet specification
  - Statistical Transformation
  - Coordinate System

Example

- Let's use ggplot2 on an example dataset to illustrate how it works
- Dataset contains demographics information for the US Midwest from 2000 Census

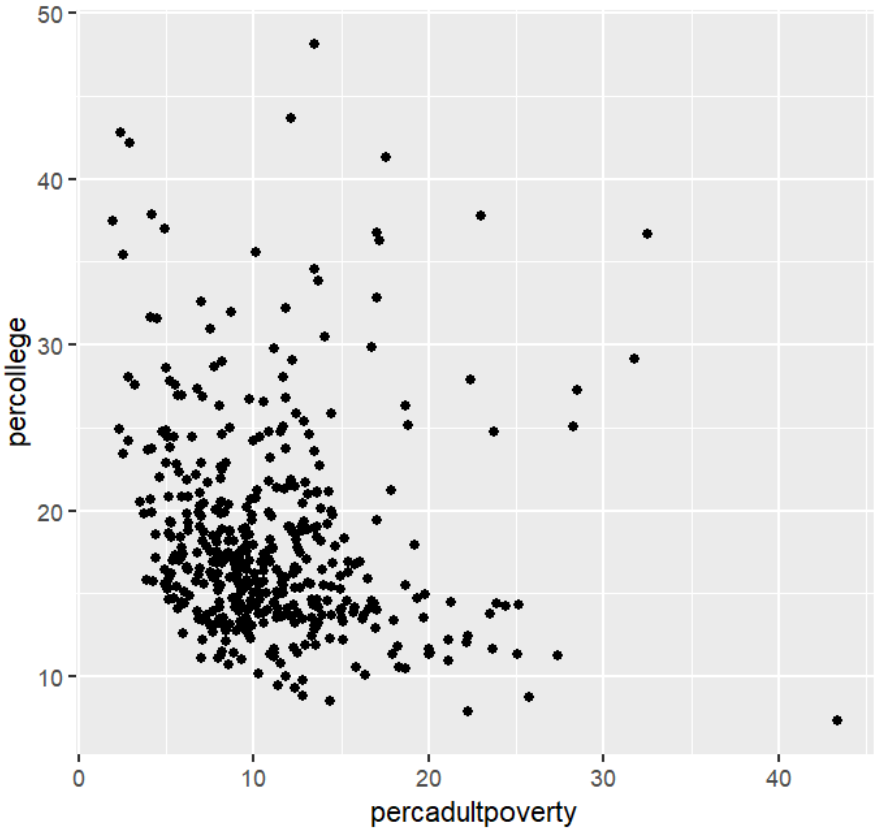
```
midwest ← ggplot2::midwest
head(midwest, 3)
#> # A tibble: 3 x 28
#>     PID county    state  area poptotal popdensity popwhite popblack popame
#>   <int> <chr>     <chr> <dbl>    <int>      <dbl>    <int>    <int>
#> 1   561 ADAMS     IL    0.052    66090      1271.    63917     1702
#> 2   562 ALEXANDER IL    0.014    10626       759      7054     3496
#> 3   563 BOND      IL    0.022    14991       681.    14477      429
#> # ... with 19 more variables: popasian <int>, popother <int>, percwhite <d
#> #   percblack <dbl>, percamerindan <dbl>, percasian <dbl>, percother <dbl>
#> #   popadults <int>, perchsd <dbl>, percollege <dbl>, percprof <dbl>,
#> #   poppovertyknown <int>, percpovertyknown <dbl>, percbelowpoverty <dbl>,
#> #   percchildbelowpovert <dbl>, percadultpoverty <dbl>, percelderlypoverty
#> #   inmetro <int>, category <chr>
dim(midwest)
#> [1] 437  28
colnames(midwest)
```

Midwest Demographics

- Suppose we wanted to know the relationship between the percent of people below poverty line and the percent of people college educated
- What would you expect the graph to look like?

Midwest Demographics

```
ggplot(midwest,
       aes(x = percadultpover
           y = percollege)) +
  geom_point()
```



Recap

- Intialize a plot with the `ggplot` function
- Specify our data source
- *Aesthetic* properties: choose which variables to use for x and y position
- *Geometric* object (geom): Specify the type of plot
- We used the *point geom* `geom_point()` which produces a scatterplot

Improving the plot

- Intialize a plot with the `ggplot` function
- Specify our data source
- *Aesthetic* properties: choose which variables to use for x and y position
- *Geometric* object (geom): Specify the type of plot
- We used the *point geom* `geom_point()` which produces a scatterplot