MassMutual DSDP 2018:

CRASH COURSE IN `ggplot2`

June 15, 2018

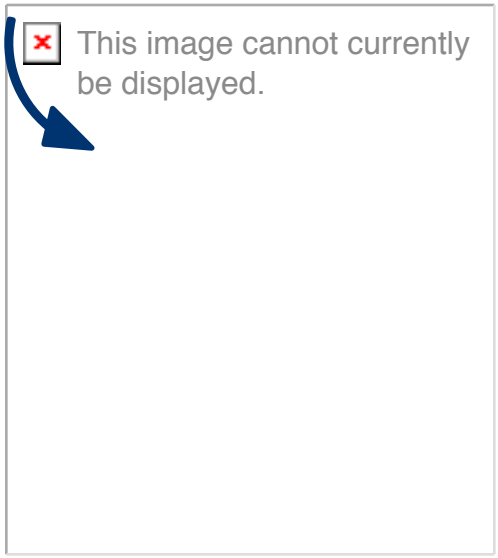
R. Jordan Crouser
Assistant Professor of Computer Science
Smith College

Outline

- ✓Introductions
- ✓Visualization overview
  - ✓Flashback to early experiences in data wrangling
  - ✓Visualization (def.)
  - ✓Data (def.)
- ✓Graphical primitives
- ✓Visual dimensions
- ✓Pre-lunch activity: mapping visual to data dimensions
- After lunch: ggplot2 crash course

10-minute crash course

“(g)rammar of (g)raphics”



What is the “Grammar of Graphics”?

- **Big idea**: independently specify plot *building blocks* and combine them to create graphical displays
- Building blocks include:
  - data (obvi.)
  - geometric objects (the literal stuff we draw)
  - aesthetic mappings (how we draw that stuff)
  - statistical transformations (underlying model)
  - scales (range of values, colors, etc.)
  - faceting (small multiples)

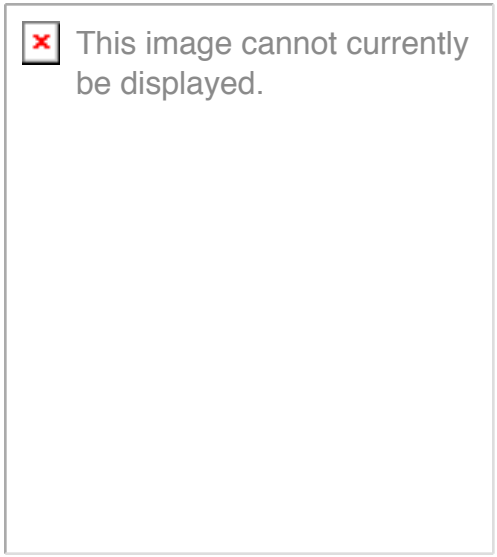
Geometric objects (`geom`)

- In `ggplot2`, the actual marks we put on a plot are called *geometric objects* or `geoms`
- Examples:
  - points (`geom_point`, for scatter plots, dot plots, etc.)
  - lines (`geom_line`, for time series, trend lines, etc.)
  - boxplot (`geom_boxplot`, for, well, boxplots!)
  - … and many more!

Aesthetic mapping (`aes`)

- In `ggplot2`, an *aesthetic* means "something you can see"
- Aesthetic mappings are set with the `aes()` function
- Each type of `geom` accepts only a subset of all aesthetics.
- Examples include:
  - position (i.e., on the x and y axes)
  - color ("outside" color)
  - fill ("inside" color)
  - shape (of points)
  - line type
  - size
  - … and many more!

All you really need to know…




```
geoms + aes()
```

Lab: `ggplot2` refresher