

Data Viz with ggplot2

(CMP595 PPGC/INF/UFRGS)



Data Viz with ggplot2 (CMP595 PPGC/INF/UFRGS)

Lucas Mello Schnorr, Jean-Marc Vincent

INF/UFRGS

Porto Alegre, Brazil – October 20th, 2017

{

}

Make sure your data is tidy

The ggplot2 package (part of tidyverse)

Implements the grammar of graphics

- ▶ Deep features behind all statistical graphics
- ▶ Layered to enable interactive construction
- ▶ Set of **verbs** (when combined, produce a plot)

The ggplot2 package (part of tidyverse)

Implements the grammar of graphics

- ▶ Deep features behind all statistical graphics
- ▶ Layered to enable interactive construction
- ▶ Set of **verbs** (when combined, produce a plot)

How it works

The grammar enables a mapping

Rows (observations) →

Columns (variables) →

The ggplot2 package (part of tidyverse)

Implements the grammar of graphics

- ▶ Deep features behind all statistical graphics
- ▶ Layered to enable interactive construction
- ▶ Set of **verbs** (when combined, produce a plot)

How it works

The grammar enables a mapping

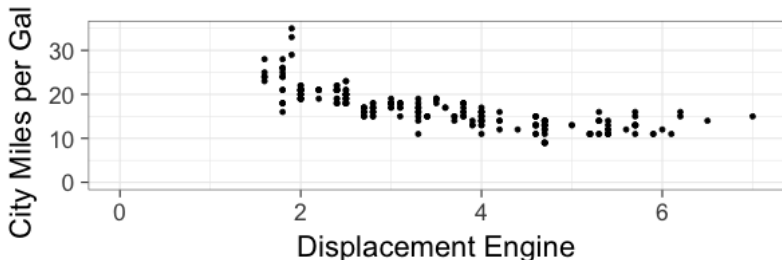
Rows (observations) →	Geometric objects (points, lines) and theirs
Columns (variables) →	Asthetic attributes (color, shape, size)

Plots

- ▶ (your tidy) Data
- ▶ Layers (each type of geometric object is a layer)
- ▶ Scales (from the values in the data space to the aesthetic space)
- ▶ Coord
- ▶ Faceting
- ▶ Theme (see Tufte)

Demonstration #1 (mpg)

```
suppressMessages(library(tidyverse));  
mpg %>%  
  ggplot(aes(x=displ, y=cty)) +  
    geom_point() +  
    theme_bw(base_size=22) +  
    xlim(0,NA) +  
    ylim(0,NA) +  
    xlab("Displacement Engine") +  
    ylab("City Miles per Galon")
```

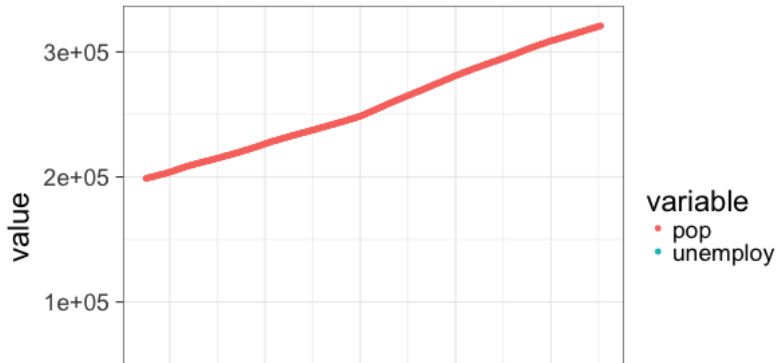


Demonstration #2 (presidential)

```
presidential %>%  
  arrange(start) %>%  
  mutate(Order = 1:nrow(.)) %>%  
  ggplot(aes(x=start, y=Order, fill=party)) +  
  geom_rect(aes(xmin=start, xmax=end, ymin=Order, ymax=Order)) +  
  theme_bw(base_size=22) +  
  scale_y_continuous(breaks=c(seq(1:11)+0.5), labels=(presidential$party[1:11])) +  
  scale_fill_manual(values = c("blue", "red")) +  
  coord_cartesian(xlim=c(as.Date("1953-01-20"), as.Date("2017-01-20"))) +  
  theme (  
    plot.margin = unit(c(0,0,0,0), "cm"),  
    legend.margin = unit(.1, "line"),  
    axis.title.x = element_blank(),  
    axis.title.y = element_blank(),  
    legend.position = "top",  
    legend.justification = "left",  
    legend.title = element_blank()  
  )
```

Demonstration #3 (economics)

```
economics %>%  
  gather(variable, value, -date) %>%  
  filter(variable %in% c("pop", "unemploy")) %>%  
  ggplot(aes(x=date, y=value, color=variable)) +  
  theme_bw(base_size=22) +  
  geom_point() +  
  ylim(0,NA)
```



References

The Grammar of Graphics by Leland Wilkinson

- ▶ Long book: 20 chapters

ggplot2: Elegant Graphics for Data Analysis by Hadley Wickham

```
git clone https://github.com/hadley/ggplot2-book.git
cd ggplot2-book
make clean
make
```