ggplot2

Cedric Arisdakessian

2022-09-12

Basics

- Data in `ggplot()` call or specific function
- Aesthetics: `aes()` or `aes_string()`
- Add layers with +

```
ggplot(data, aes(...)) +
  layer1 +
  layer2 + ...
# OR
ggplot() +
  layer1(data1, aes()) +
  layer2(data2, aes()) + ...
```

Aesthetics

Depending on the type of layer you use:

- color
- fill
- shape
- alpha
- lwd

Note: For `geom_point()`, depending on the selected shape, fill is valid or not. https://ggplot2.tidyverse.org/reference/scale_shape-6.png

Geometry

- `geom_point()`
- `geom_boxplot()`
- `geom_histogram()` / `geom_density()`
- `...` and many more

How to choose the right plot for your data?

- How many dimensions to plot?
- Types of variables: categorical/continuous/compositional/...?
- Desired level of detail (e.g. barplot vs boxplot)

https://r-graph-gallery.com/

Subplots

- Just another layer, defined by a factor column in your data
  - you might have to convert your table from wide to long format
  - if not possible, consider using `cowplot::plot_grid()`
- 2 types of layers
  - `facet_grid(rows ~ cols)`: 1 grid for every combinations of row/cols (some plots might be empty)
  - `facet_wrap(rows ~ cols)`: Only non-empty plots.
- important option `scales`, that can be set to `free`, `free_x` or `free_y`

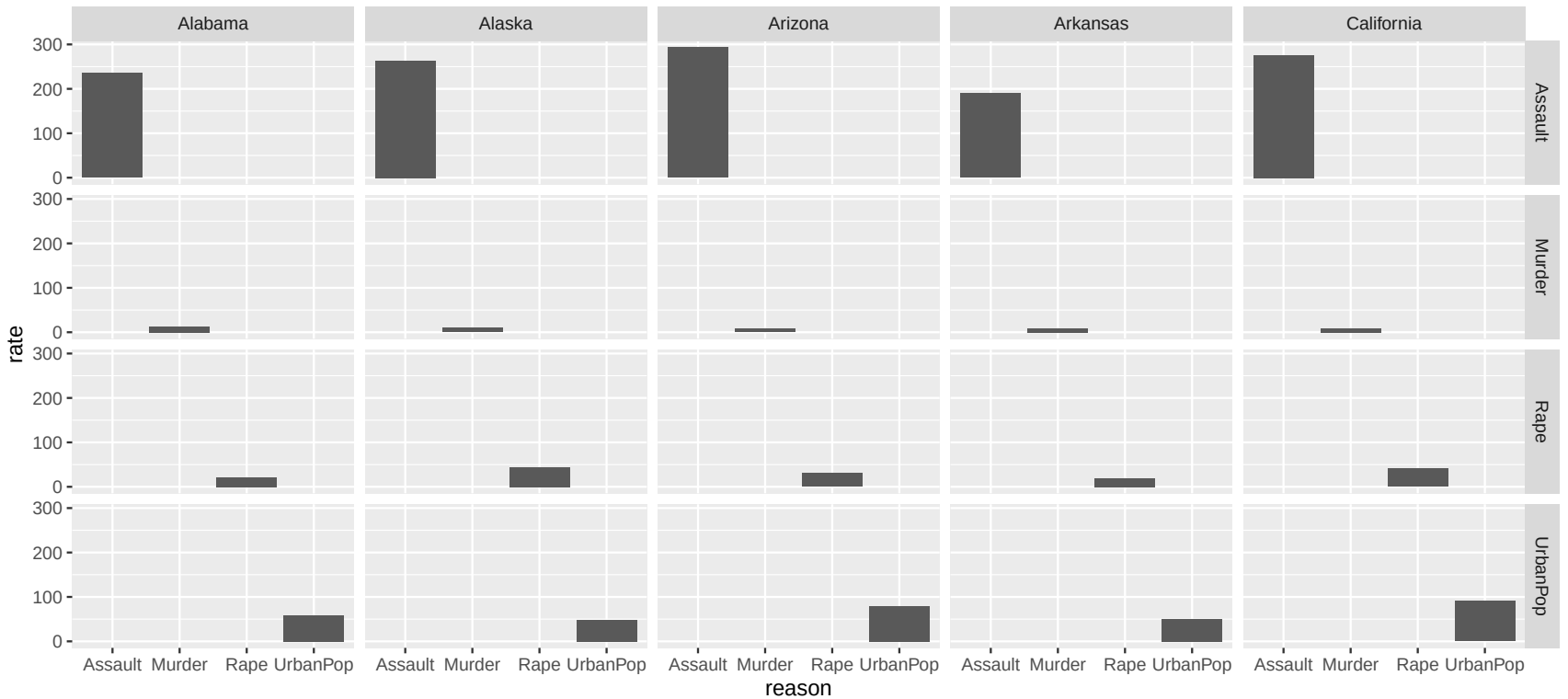
```
library(ggplot2)
library(tidyr)
library(tibble)

data("USArrests")
cols <- colnames(USArrests)

head(USArrests, 5)
```

```
##            Murder Assault UrbanPop Rape
## Alabama      13.2     236       58 21.2
## Alaska       10.0     263       48 44.5
## Arizona       8.1     294       80 31.0
## Arkansas      8.8     190       50 19.5
## California    9.0     276       91 40.6
```

```
USArrests %>% head(5) %>% rownames_to_column("State") %>%
  pivot_longer(all_of(cols), names_to="reason", values_to="rate"
  ggplot(aes(x=reason, y=rate)) + geom_col() +
    facet_grid(reason~State)
```



Customize plot

- Axes labels: `xlab()`, `ylab()`, `ggtitle()`
  or at the same time: `labs(x=, y=, title=, fill=, color=)`
- Axes limits: `xlim(0, 10)`, `ylim()`
  or at the same time: `coord_cartesian(xlim=, ylim=)`
- Axes scale: `scale_x_continuous(trans="log10")` or `scale_x_log10()`
- Color/fill scales (palette= int or string):
  - `scale_color_continuous(palette=...)`
  - `scale_color_brewer(palette=...)`
  - `scale_color_manual(values=...)`
- Themes: https://ggplot2.tidyverse.org/reference/ggtheme.html
- Custom theme: `theme(axis.title.x = element_text(family, face, colour, size))`

Readable code

- Use natural line breaks after "+"
- Avoid copy-pasting, instead try to use the %+% operator
- Avoid creating new variables with the 'piping' operator: %>% (cf: USarrests example)

```
library(cowplot)
library(ggplot2)

p <- ggplot(data=points, aes(x=PCA1, y=PCA2)) + geom_points()

p1 <- p %+% aes(fill=Seasonal)
p2 <- p %+% aes(fill=Aquifer)
p3 <- p %+% aes(fill=Lifestyle)

plot_grid(p1, p2, p3, nrow=2)
```

```
base_plot <- ggplot(data, aes(x=x, y=y)) + geom_point()
p1 <- base_plot %+% data[enriched,]
p2 <- base_plot %+% data[!enriched,]
plot_grid(p1, p2)
```

Bonus: ggplot2 extensions

https://exts.ggplot2.tidyverse.org/gallery/

- ggsci (more color palettes)
- ggthemes (more themes)
- gganimate (dynamic plots)
- ggpubr (add stats to boxplots)
- ggrepel (text labels positioning)
- etc.