

Class 6: The grammar of graphics

API-201

Quantitative Analysis and Empirical Methods I

Profs. Goel and Taylor
Harvard Kennedy School

Agenda

1. Introduce the final exercise
2. Review basics of ggplot
3. Practice making plots

Final exercise

Goal. To apply the empirical tools you have acquired in the course to a policy question of interest to you.

Final work products.

- memo [5 single-spaced pages]
- presentation [5 minutes]
- technical appendix [no limit but keep it succinct]

Teams. Work in groups of 3-4 of your choosing, and can include students across C/D sections. If you have trouble finding a group, let us know and we'll help.

Proposal. Due on Friday, Sep. 29. We encourage you to meet with the teaching team to discuss your ideas.

Grammar of graphics

The grammar of graphics is a sophisticated system for describing and building graphs.

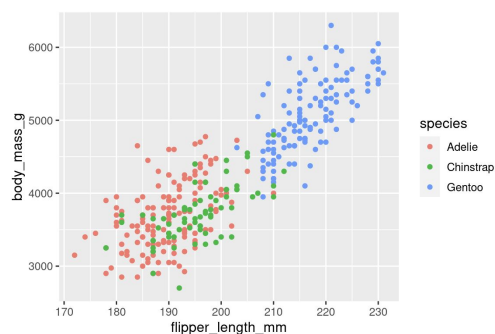
Three key features of this system are:

- There is an explicit mapping between variables and visual properties [with the `aes` function]
- This grammar lets us describe plots in general terms, leaving the details to the visualization engine
- Plots are built up additively, in layers

An illustrative example

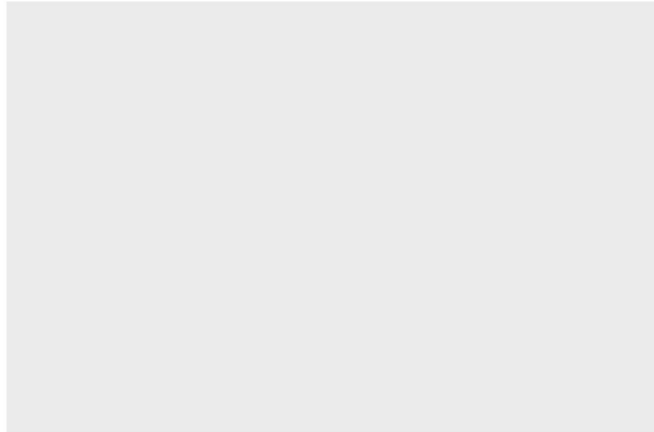
We'll walk through an illustrative example that highlights the key principles of this grammar: a comparison of flipper length and body mass among penguins.

[Taken from R4DS — Chapter 2]



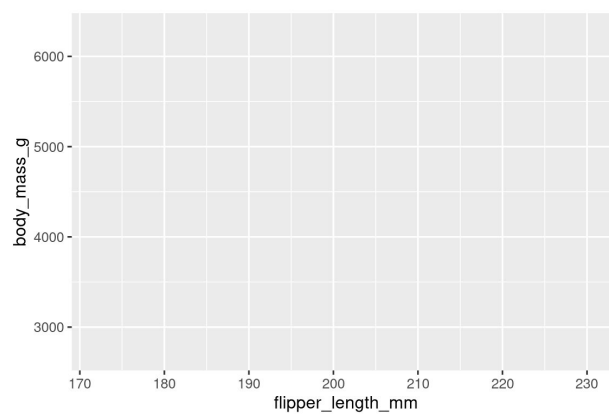
An illustrative example

```
ggplot(data = penguins)
```



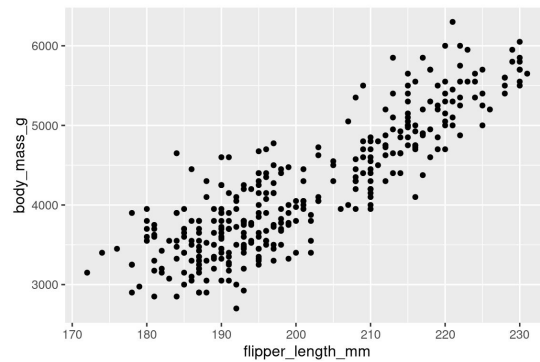
An illustrative example

```
ggplot(  
  data = penguins,  
  mapping = aes(x = flipper_length_mm, y = body_mass_g)  
)
```



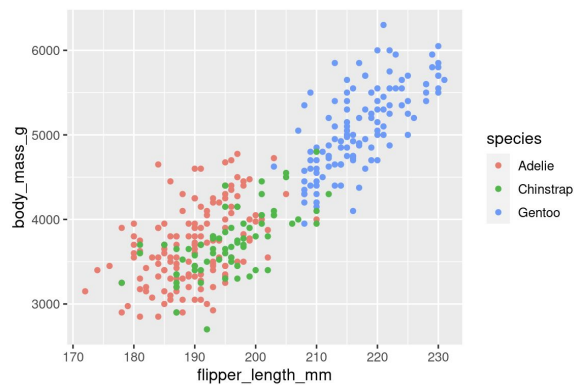
An illustrative example

```
ggplot(  
  data = penguins,  
  mapping = aes(x = flipper_length_mm, y = body_mass_g)  
) +  
  geom_point()  
#> Warning: Removed 2 rows containing missing values (`geom_point()`).
```



An illustrative example

```
ggplot(  
  data = penguins,  
  mapping = aes(x = flipper_length_mm, y = body_mass_g, color = species)  
) +  
  geom_point()
```



Key takeaways

Grammar of graphics is a rich system for describing and building plots.

With this grammar, we:

- Explicitly **map variables** to visual properties
- Describe plots in **general terms**, leaving the details to the visualization engine [e.g., we don't need to specify the colors]
- Build plots up additively, in **layers**

Plotting in practice

Now we'll use what we've learned to make our own plots!

<https://bit.ly/API201-ggplot>