COMM101: Data Visualization with ggplot

Welcome to the grammar of graphics

MARINCS 100B | Intro to Marine Data Science | Winter 2025

Key concepts

3 components: Data, Geometries, Mapping

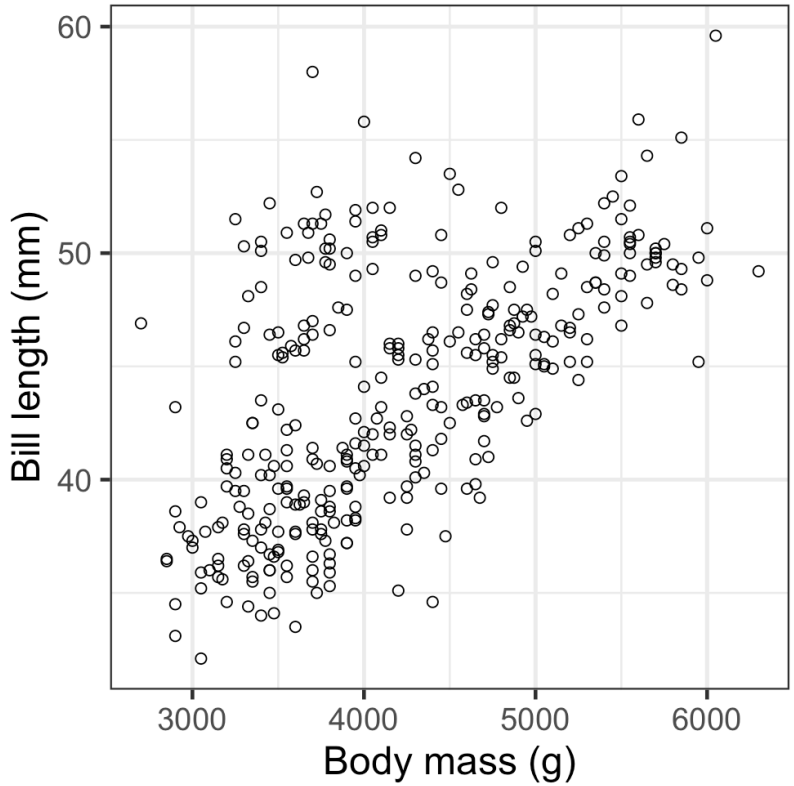
Refinement: Scales and Themes

Best practices: Labeling, visual interpretation, negative space

Data, mapping, and geometries

Geometries!

species	body_mass_g	bill_length_mm
Adelie	3750	39.1
Gentoo	5400	49.9
Chinstrap	3500	46.5



Data frame

Row = observation, Column = var

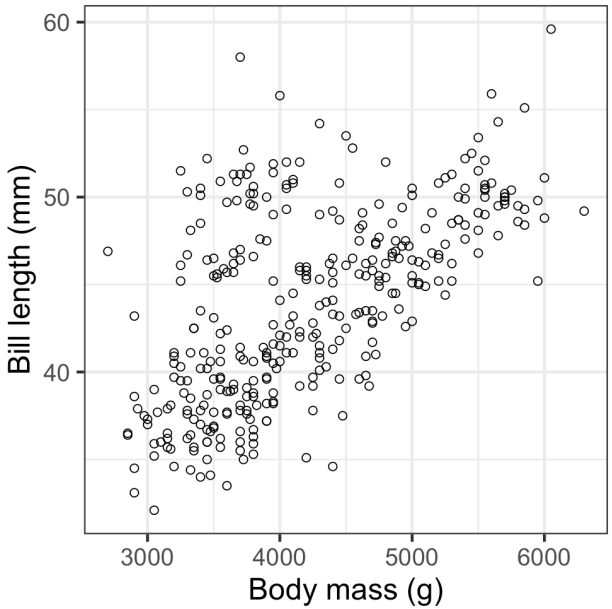
Mapping = vars to axis

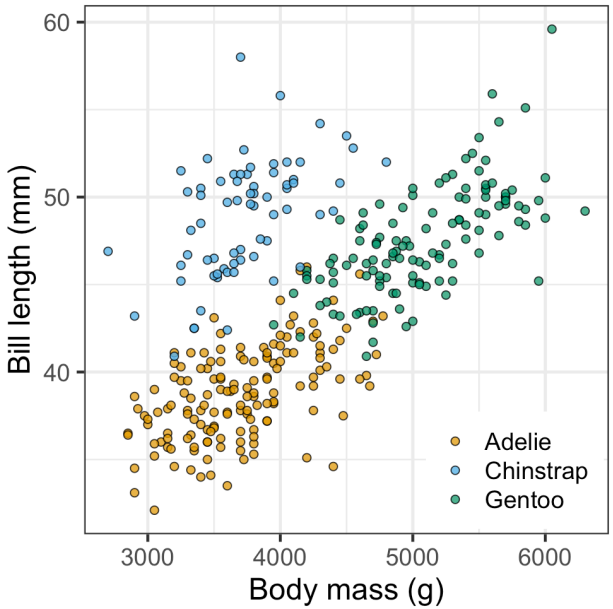
Data, mapping, and geometries

Data = df
geometries = actual form in figure(points)
mapping = connections between vars and axis

Scales and themes

species	body_mass_g	bill_length_mm
Adelie	3750	39.1
Gentoo	5400	49.9
Chinstrap	3500	46.5





Same data, same geom.
Same mapping

Added mapping color and color scale
Theme positions legend

Scales and themes

Scales customize how mapping interacts with geometry

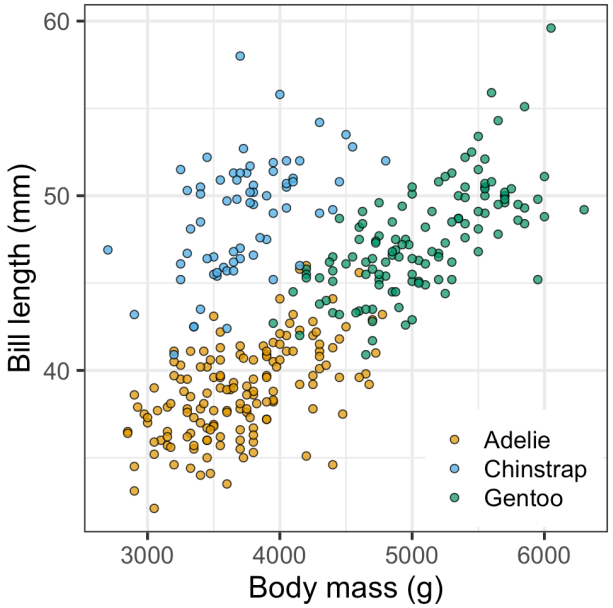
Theme customizes overall

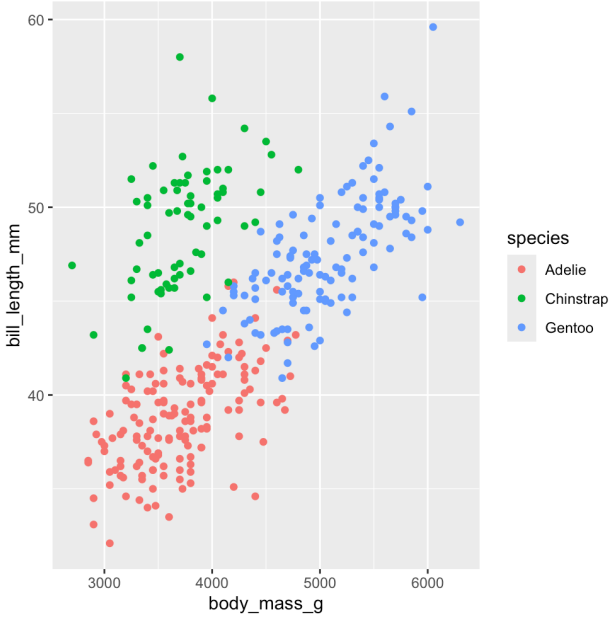
Visualization best practices

Fix labeling

Easy to visually interpret

Better negative space





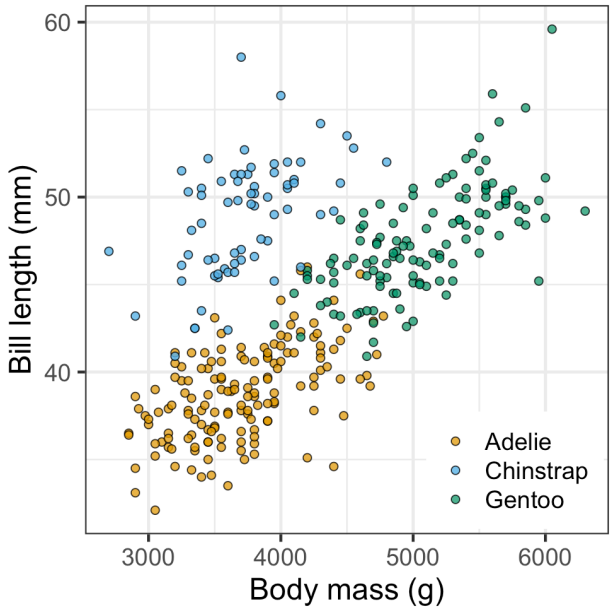
Recap

Figures always include Data, Geometries, and Mappings

We can refine visual via scales and themes

Label, negative space, etc

species	body_mass_g	bill_length_mm
Adelie	3750	39.1
Gentoo	5400	49.9
Chinstrap	3500	46.5



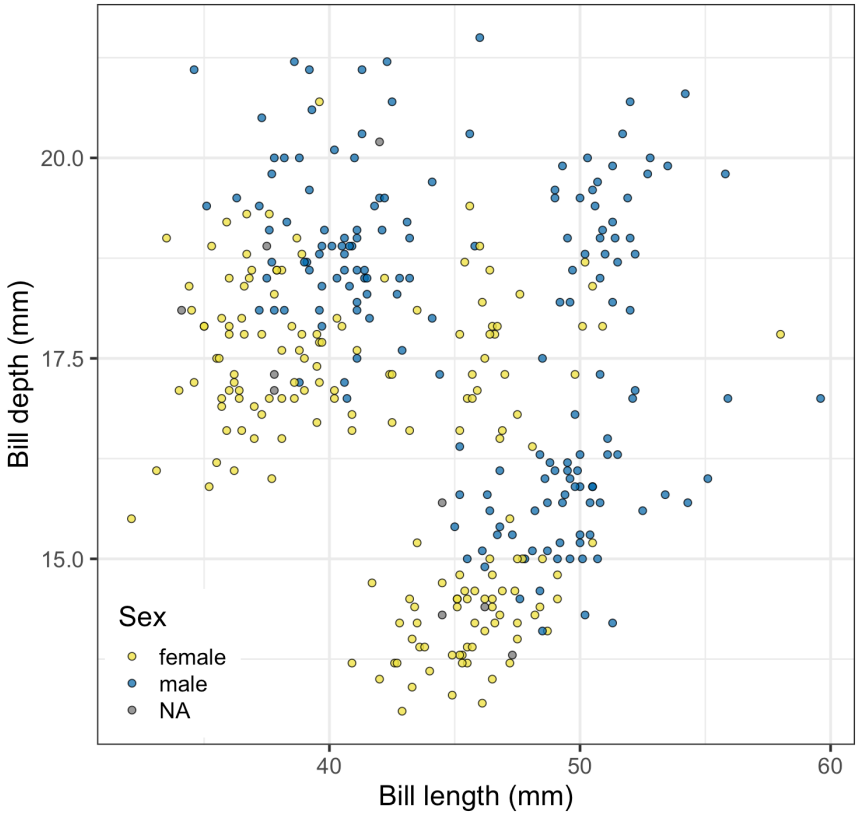
New vocabulary and lingering questions

New vocabulary

Lingering questions

Exercises

Describe the grammar of graphics components (data, geometries, mapping, scales, theme) in the figure below.



The data is being represented by the points(geometry). The bill length and depth have been mapped to the x and y axis(mapping). The scale turned the points to be blue and yellow(scales), and the theme changed the background color and the legend position.

COMM101: Data Visualization with ggplot

Introducing ggplot

MARINCS 100B | Intro to Marine Data Science | Winter 2025

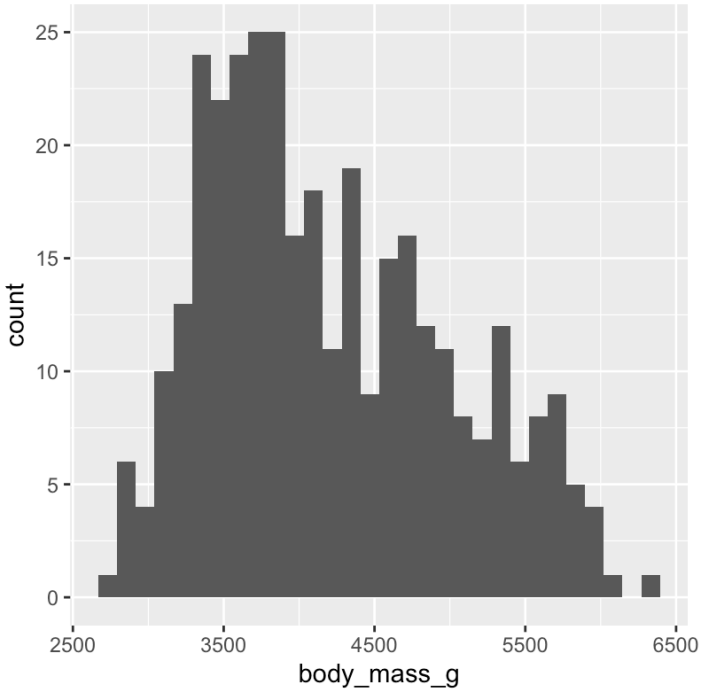
Key concepts

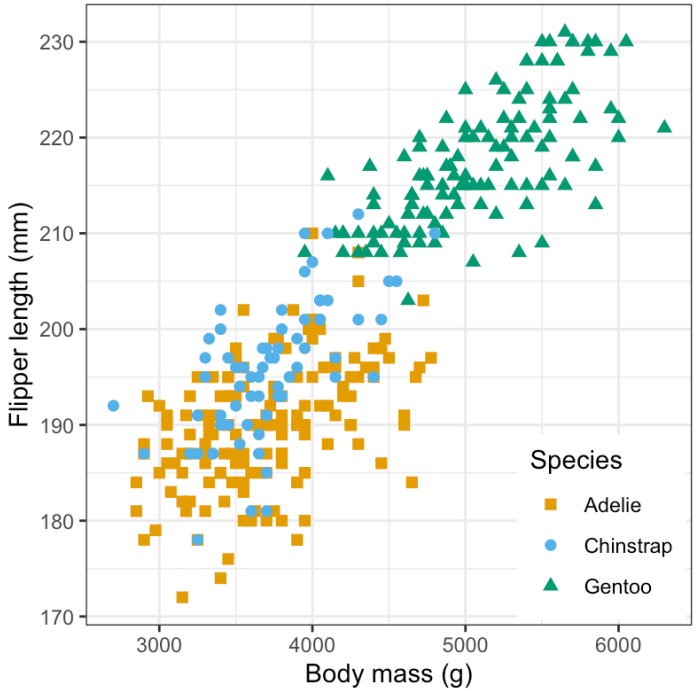
GGplot is R-based implementation of grammar of graphics

Adds layers

Choose geometries wisely to represent vars

Demo in R





Insert cool code stuff here

Recap

Built plots in ggplot
How to choose geometries
Add layers to plots

New vocabulary and lingering questions

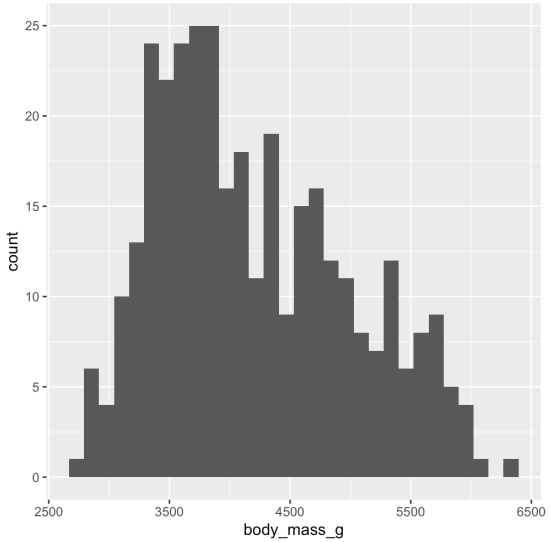
New vocabulary

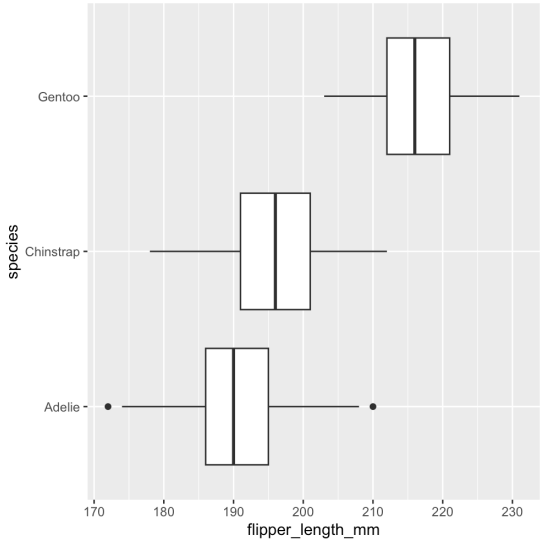
Lingering questions

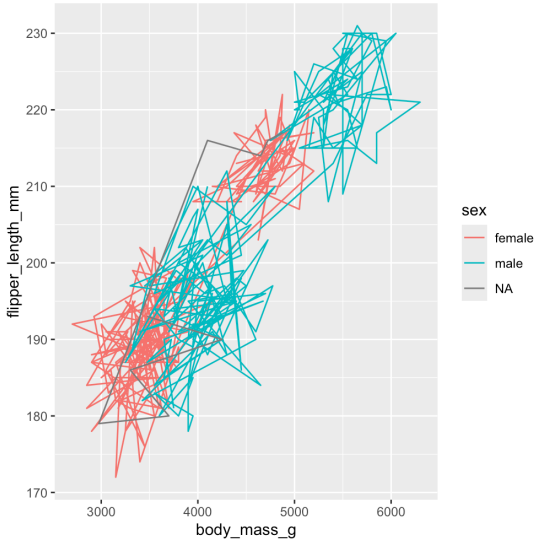
Why use geom_path()?
*for the exercise

Exercises

Here are three figures. Edit the code in comm101b.R so the outputs match the figures below.







COMM101: Data Visualization with ggplot

Customization with scales and themes

MARINCS 100B | Intro to Marine Data Science | Winter 2025

Key concepts

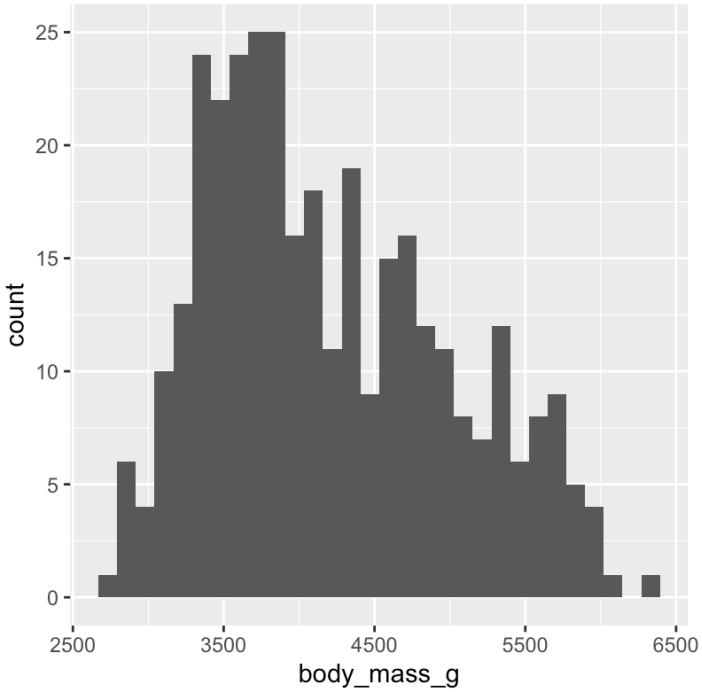
Visual presentation matters

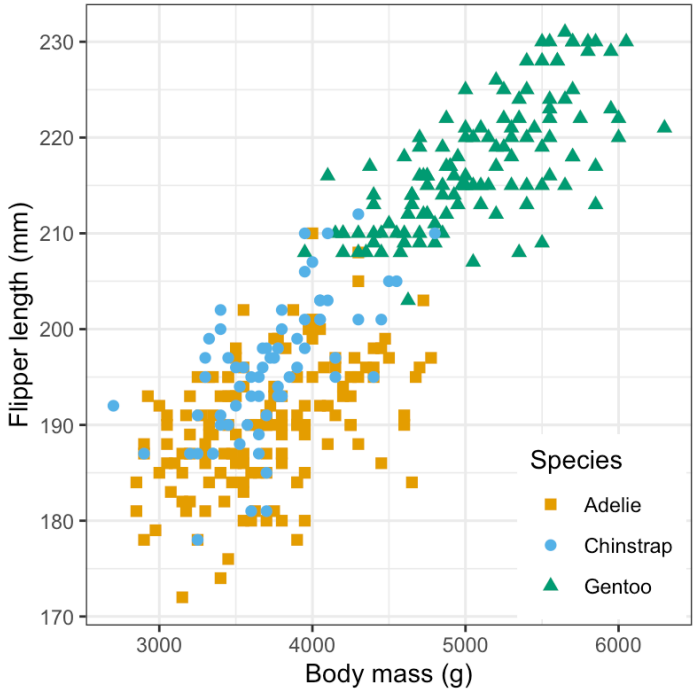
Scales customize mapping between data and geometries

Scales let us change color

Themes customize overall presentation, like background

Demo in R


count
25
20
15
10
5
0
2500
3500
4500
5500
6500
body_mass_g


Flipper length (mm)
230
220
210
200
190
180
170
3000
4000
5000
6000
Body mass (g)
Species
Adelie
Chinstrap
Gentoo

Recap

Visual presentation is important
Scales is data --> mappings
Themes is overall stuff

New vocabulary and lingering questions

New vocabulary

Lingering questions

Exercises

comm101c.R contains the code to make the figure below. Edit the code to use scales and themes to improve the visual presentation.