COMM101: Data Visualization with ggplot

Welcome to the grammar of graphics

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

Key concepts

Components: data, geometrics, mapping

Refine: scales and themes

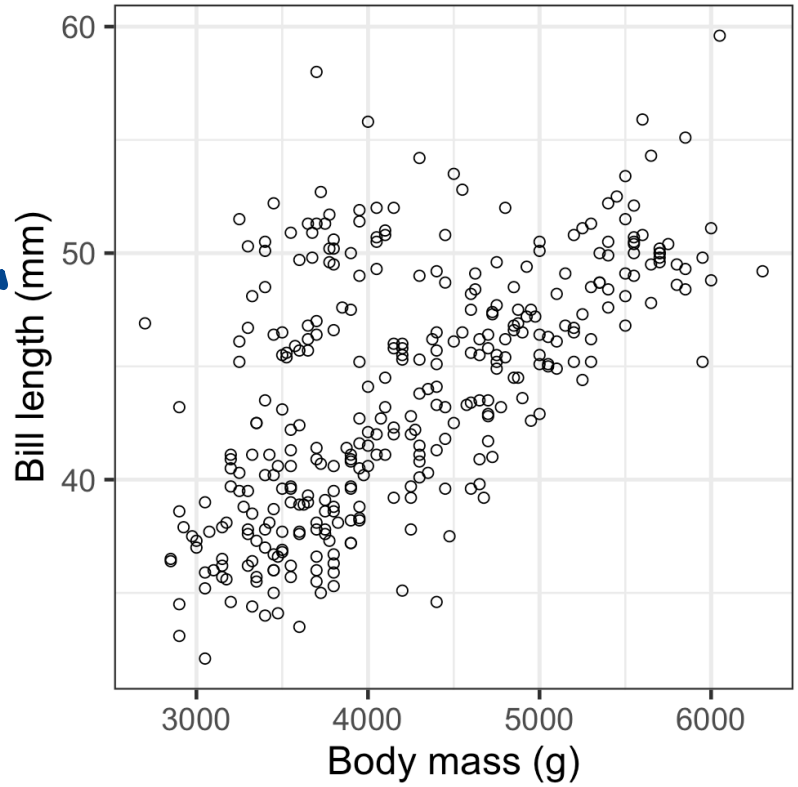
Best practices: labeling, visual interpretation, negative space

Data, mapping, and geometries

Data Frame

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

mapping



individual points:

geometries

Data, mapping, and geometries

data: data frame itself

geometrics: visual form of data in the figure

mapping: connections between data columns and aspects of geometries

Scales and themes

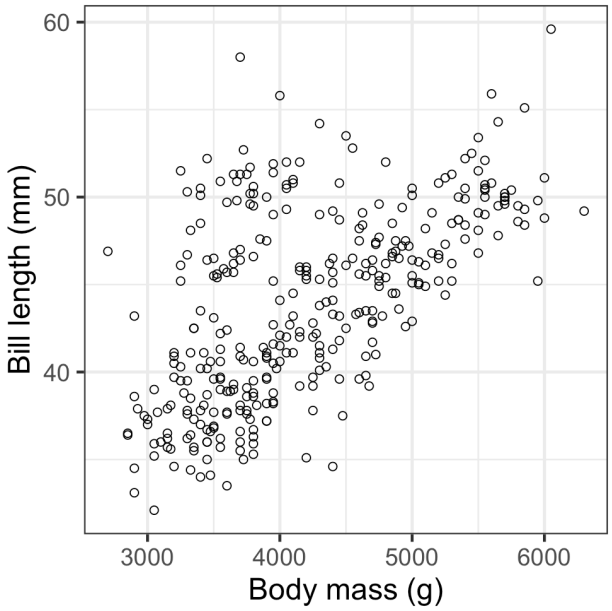
species	body_mass_g	bill_length_mm
Adelie	3750	39.1
Gentoo	5400	49.9
Chinstrap	3500	46.5

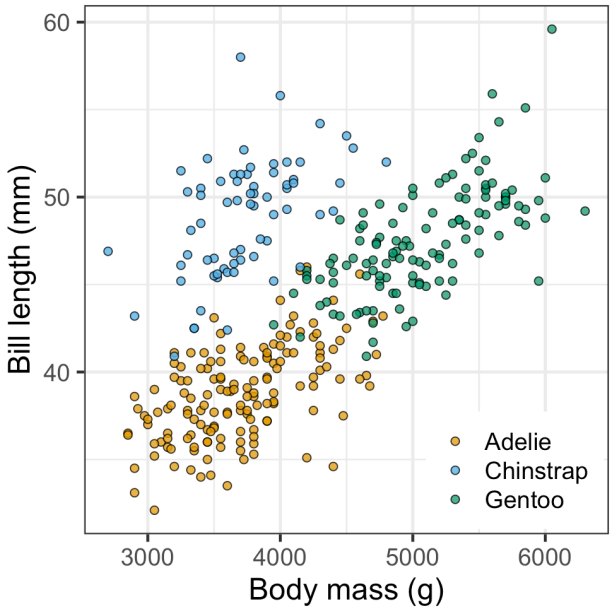
Same:

- data
- geometry
- x, y mappings

Different:

- added a mapping (color)
- color scale
- theme to position the legend





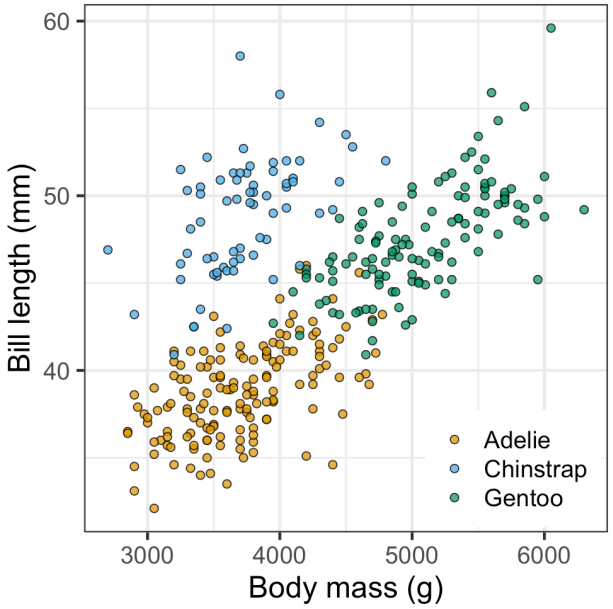
Scales and themes

scales: customize how mappings interact w/ geometries

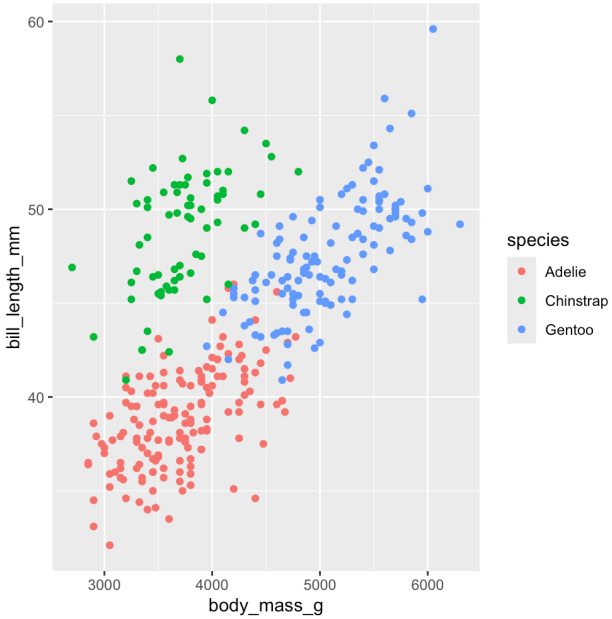
themes: customize the overall visual appearance

Visualization best practices

A Good Plot:



Original:



Fix the Labeling

visual interpretation

Fix negative space

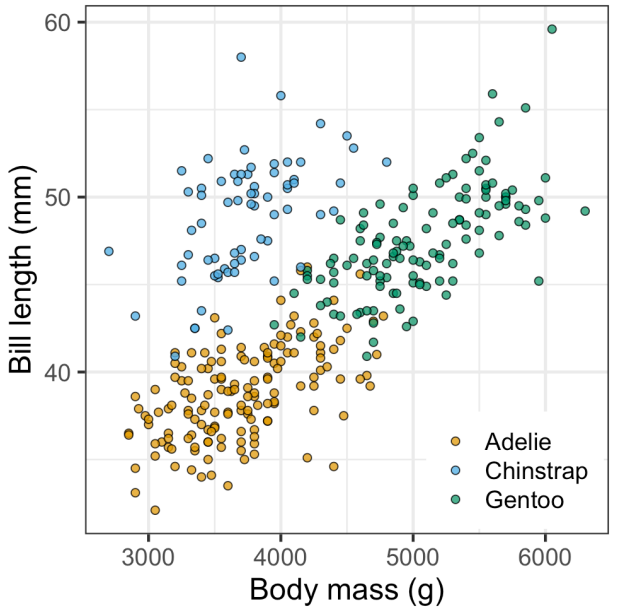
Recap

Figures always include:
data, geometrics, mappings

Refine the visual appearance
using scales and themes

Best practices: labeling and negative space

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

mapping: the connection between data and geometries

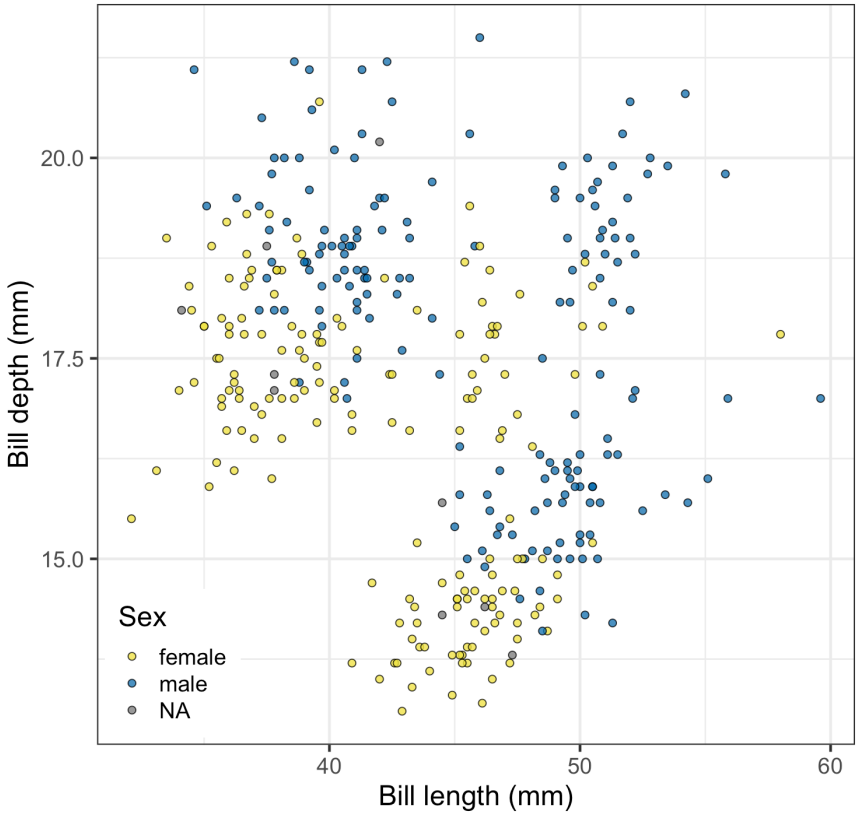
geometries: data in a visual form

scales: customization of mappings in relation to a geometry

Lingering questions

Exercises

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



mapping: axis labels and colored data

theme: color of background, moved legend

scale: the organization of colors

geometrics: each individual plotted point

data: the table used to create the plot

COMM101: Data Visualization with ggplot

Introducing ggplot

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

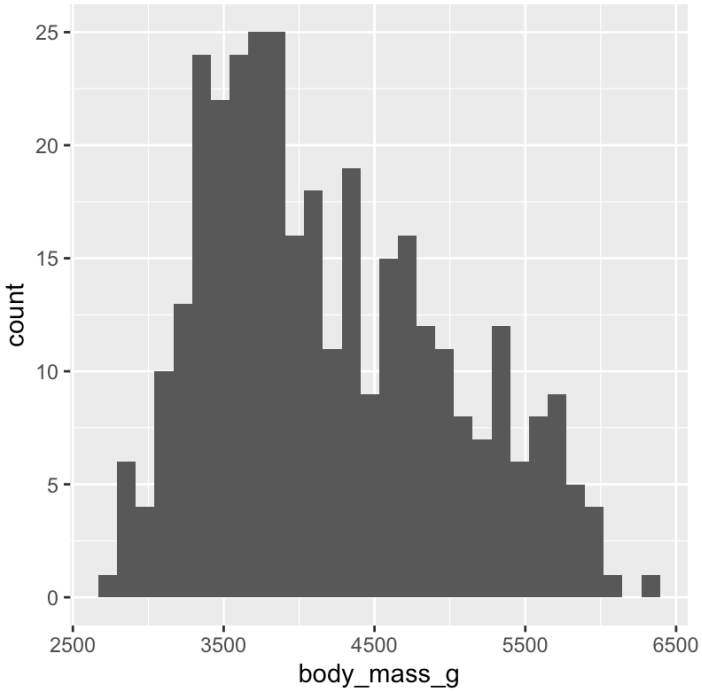
Key concepts

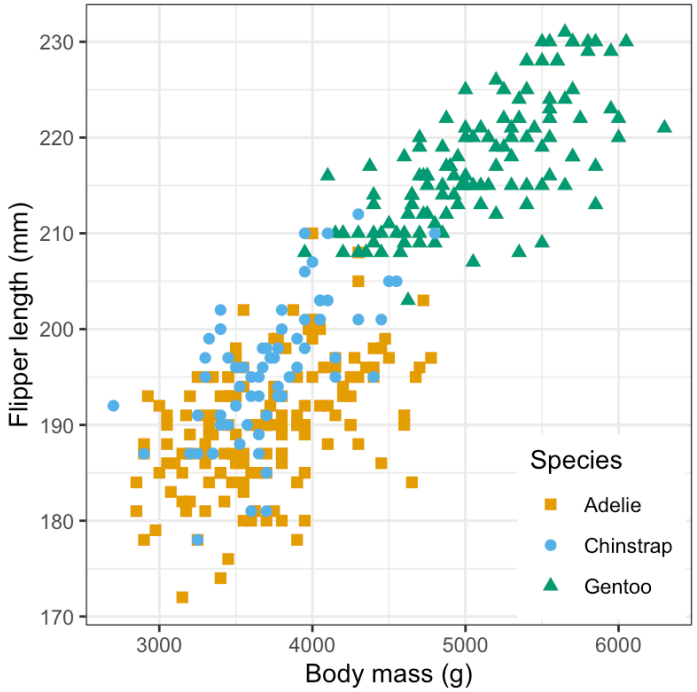
ggplot is an R-based implementation of the grammar of graphics

Adding layers (e.g. geometries, scales etc.)

Choosing geometries to represent variables (and combinations)

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

```
library(ggplot2)

library(palmerpenguins)

# geometries

# visualizing one continuous variable: histogram

ggplot(penguins, aes(x= body_mass_g)) + geom_histogram()

# visualizing one continuous variable and categorical valie: boxplots

ggplot(penguins, aes(x= body_mass_g, y= species)) + geom_boxplot()

# visualizing two continuous variables: scatter plot

ggplot(penguins, aes(x= body_mass_g, y= flipper_length_mm)) + geom_point()

# Aesthetics

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

Recap

ggplot is R's implementation of the grammar of graphics

build up plots by adding layers

how to choose geometries to best represent our variables

New vocabulary and lingering questions

New vocabulary

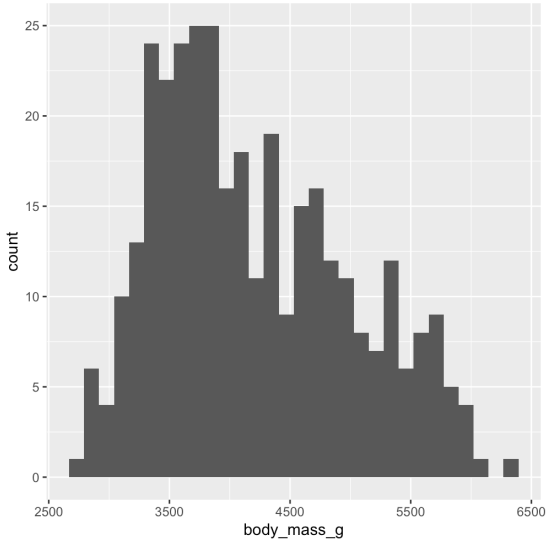
ggplot: grammer of graphics

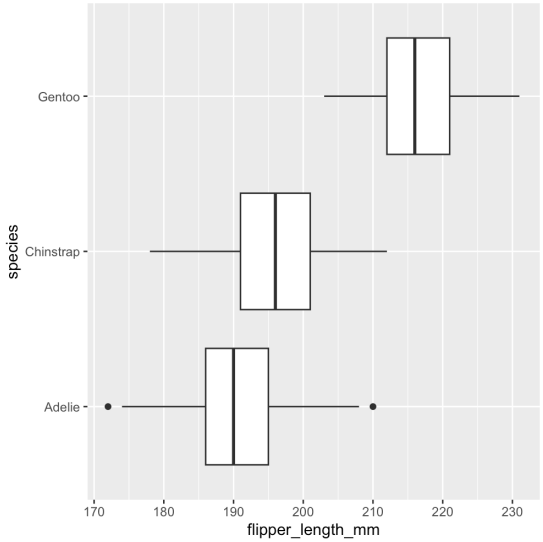
in relation to a table/plot

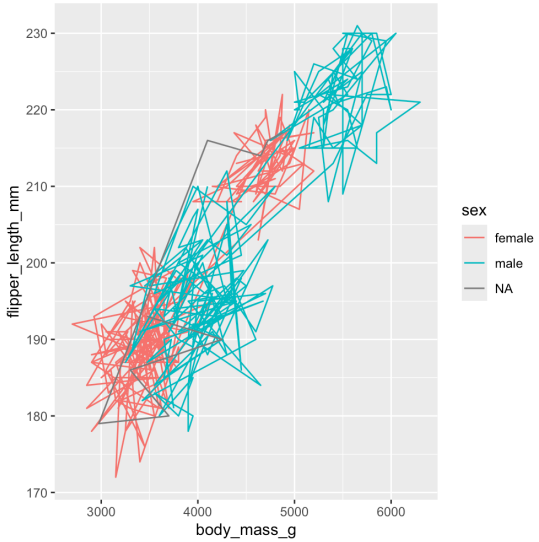
Lingering questions

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

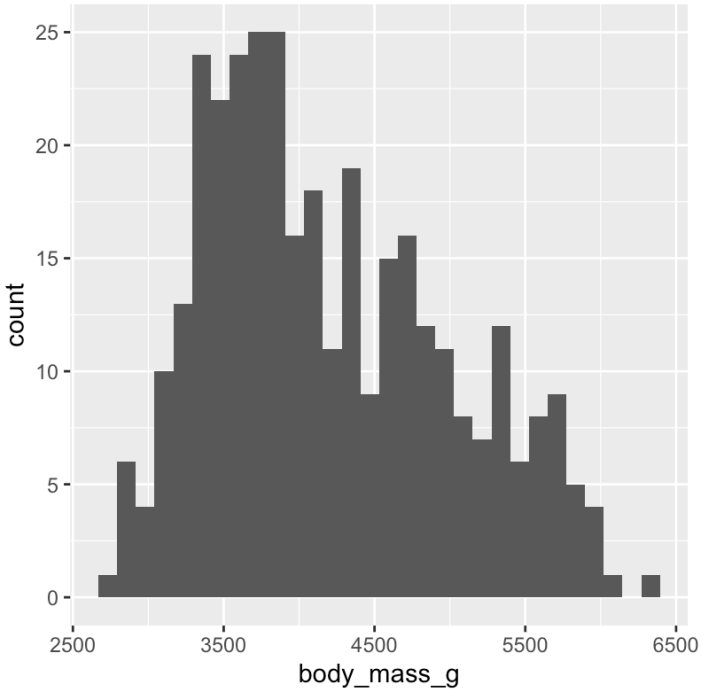
Key concepts

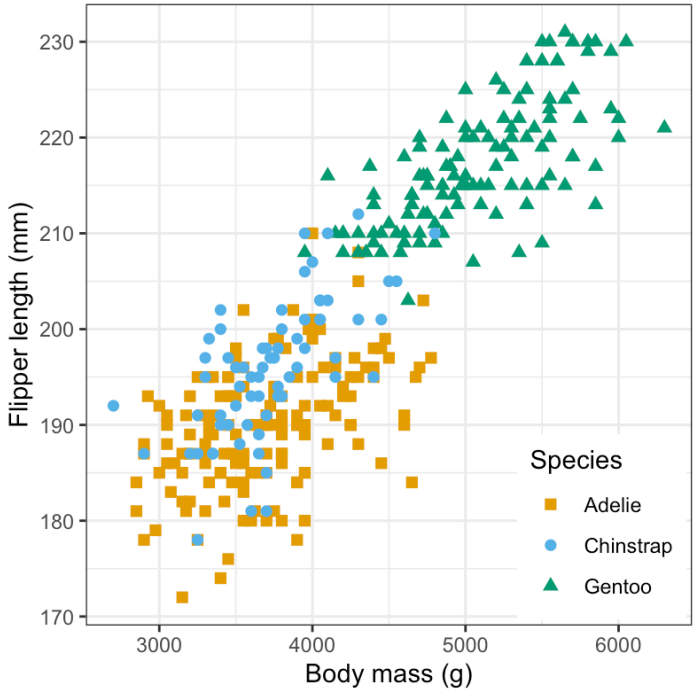
Visual presentation matters

scales customize the mapping between data and geometries

themes customize the overall appearance

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

```
library ggplot

library palmerpenguins

# scales customize the expression of mappings


# manually override default values

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

# use colorblind-friendly pallete instead

Okabe_Ito <- c("#E69F00", "#56B4E9","#009E73","#F0E442", "#0072B2")

ggplot (penguins, aes (x= body_mass_g, y= flipper_length_mm, color = species)) + geom_point () +
  scale_color_manual (values = Okabe_Ito)

# can do the same with shapes

ggplot (penguins, aes (x= body_mass_g, y= flipper_length_mm, color = species, shape = species)) +
  geom_point () + scale_color_manual (values = Okabe_Ito)

ggplot (penguins, aes (x= body_mass_g, y= flipper_length_mm, color = species, shape = species)) +
  geom_point () + scale_color_manual (values = Okabe_Ito) + scale_shape_manual (values = 15:19)
```

Continued...

```
# Built in and custom themes
ggplot(penguins, aes(x= body_mass_g, y= flipper_length_mm, color = species, shape = species)) +
  geom_point() + scale_color_manual(values = okabe_ito) + scale_shape_manual(values = 15:19)
  + theme_bw()

# move legend inside figure panel
ggplot(penguins, aes(x= body_mass_g, y= flipper_length_mm, color = species, shape = species)) +
  geom_point() + scale_color_manual(values = okabe_ito) + scale_shape_manual(values = 15:19)
  + theme_bw() + theme(legend.position = "inside", legend.position.inside = c(0.99,0.01),
  legend.justification = c(1,0))
```

Recap

visual presentation is important for interpretation

scales ← relationship between mappings and geometries

themes change the overall appearances

New vocabulary and lingering questions

New vocabulary

theme: change how a graph looks as a whole rather than a single individual part

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.