Lab Exercises Week 03

Alison Lawyer

2024-09-30

```
### keep this chunk in all your RMarkdown scripts
knitr::opts_chunk$set(echo = TRUE)
knitr::opts_chunk$set(tidy.opts = list(width.cutoff = 60), tidy = TRUE)
```

```
# List required packages
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(palmerpenguins)
```

For this lab, you will create various data visualizations of the palmerpenguins dataset to practice your understanding of ggplot and the grammar of graphics and apply visualization best practices.

```
# load the palmerpenguins data this time, you can use the
# cleaned dataset called 'penguins'
data("penguins")
```

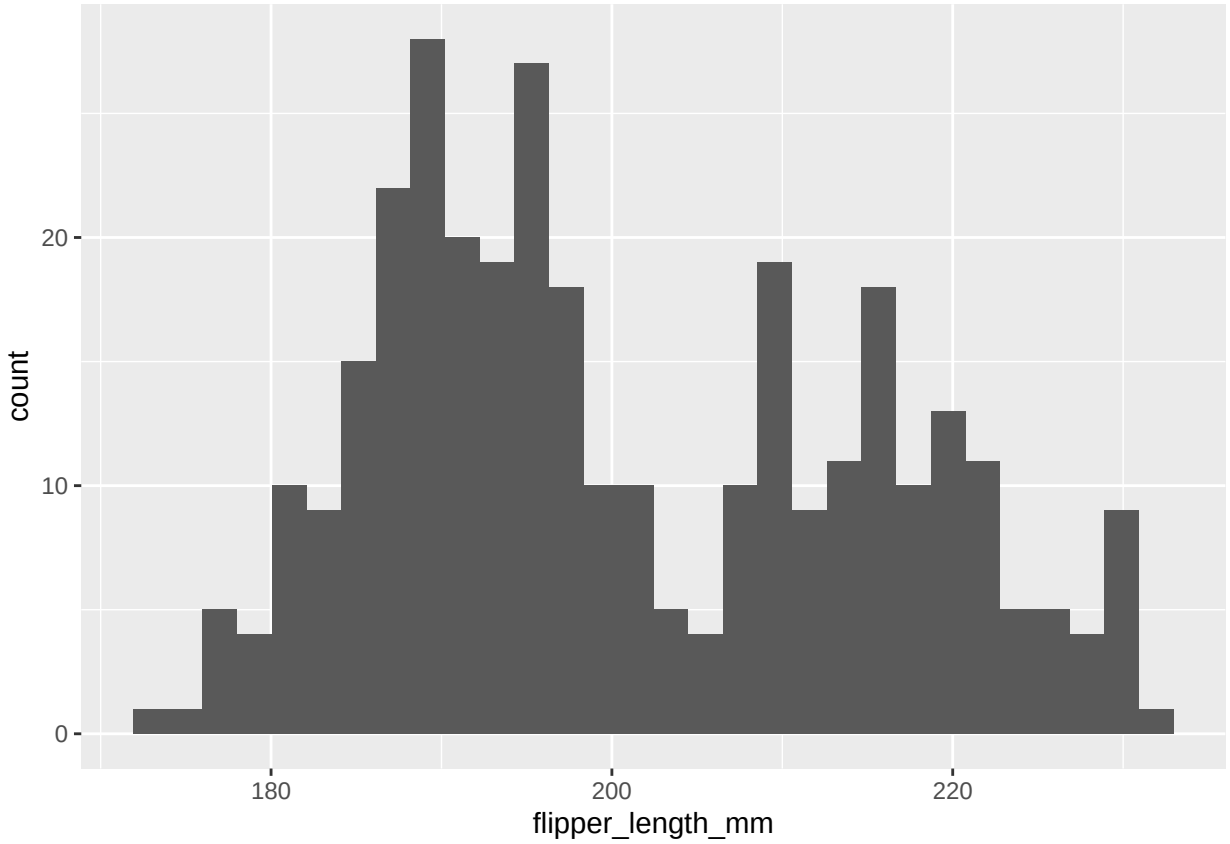
Part 1

Exercise 1

You have already created histograms previously using the hist() function from base R. Now you will learn how to do so with ggplot()

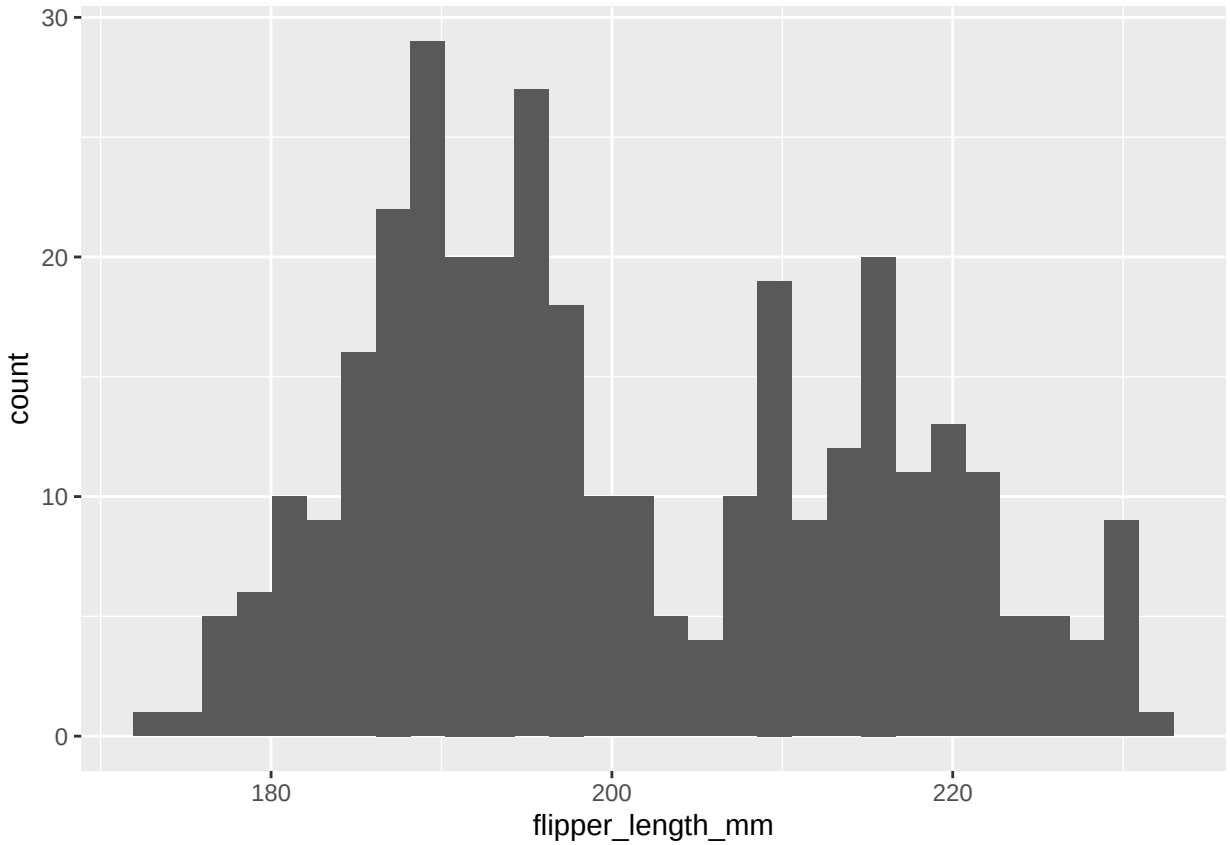
```
# Create a histogram of flipper length using
# geom_histogram() Remember to start with the basic ggplot
# layer and add aesthetics first. What do you put inside
# aes()?
ggplot(na.omit(penguins), aes(x = flipper_length_mm)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

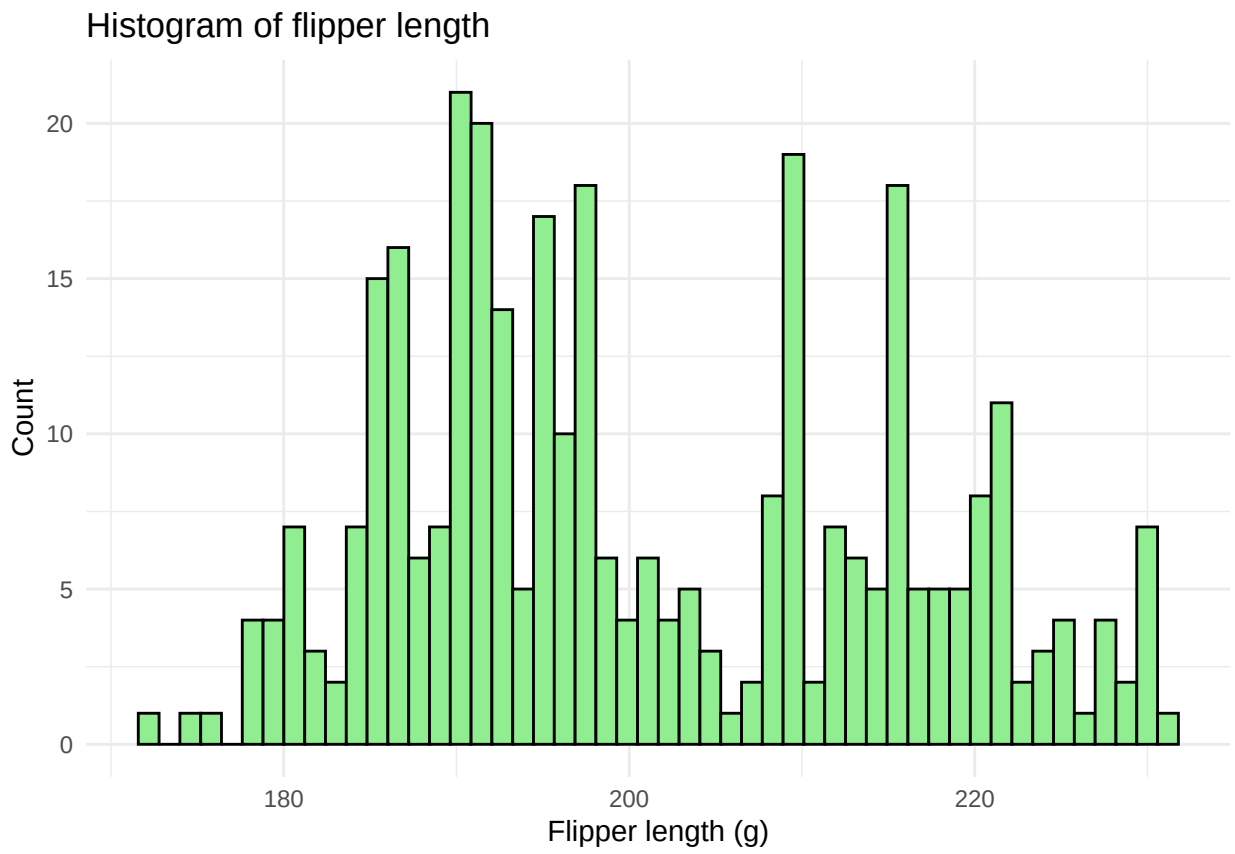



```
## OR
ggplot(penguins, aes(x = flipper_length_mm)) + geom_histogram(na.rm = TRUE)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(na.omit(penguins), aes(x = flipper_length_mm)) + geom_histogram(bins = 50,
    color = "black", fill = "lightgreen") + labs(title = "Histogram of flipper length",
    x = "Flipper length (g)", y = "Count") + theme_minimal()
```

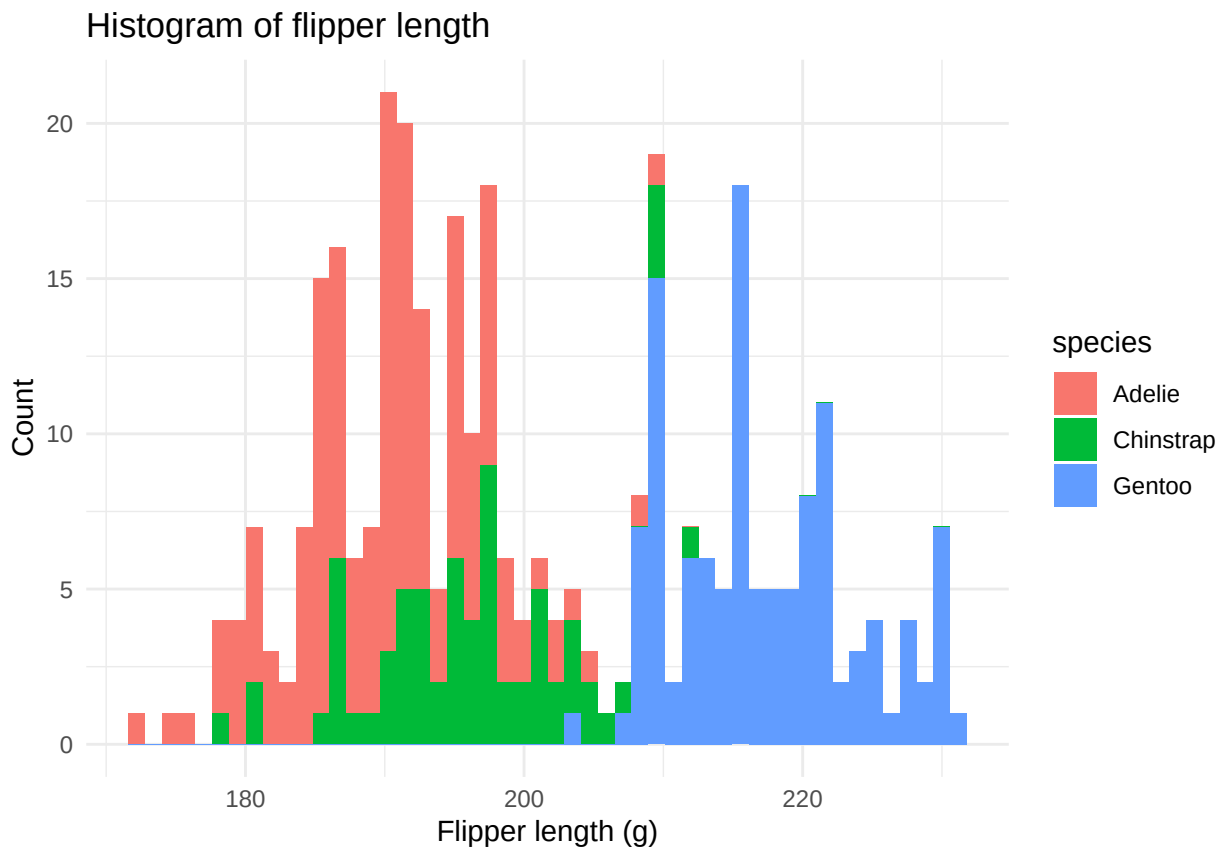


Exercise 2

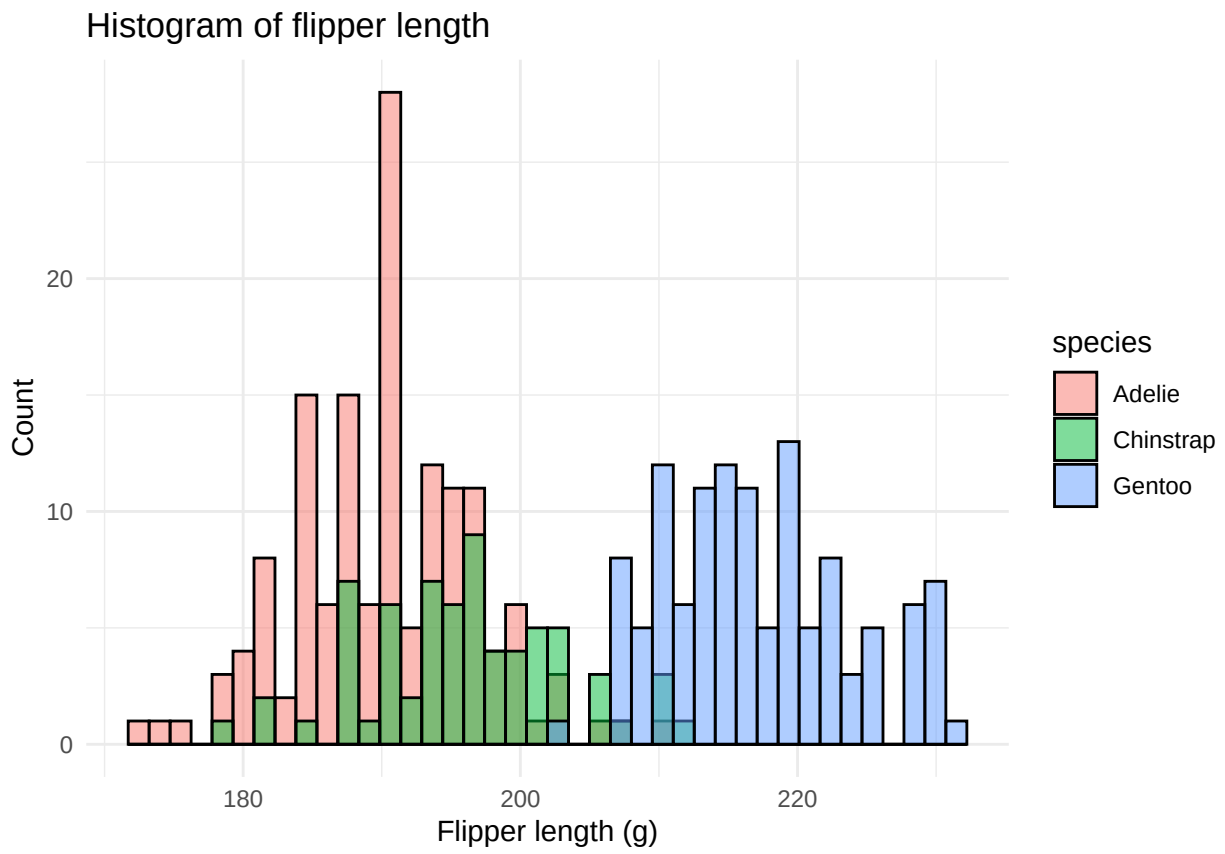
In this exercise you will practice modifying the look of your graph using a variable from the data.

```
# Modify your histogram of flipper length by coloring the
# bars by species Remember that to achieve this you will
# want to add a variable signifying species into the aes()
# function of ggplot

ggplot(na.omit(penguins), aes(x = flipper_length_mm, fill = species)) +
    geom_histogram(bins = 50) + labs(title = "Histogram of flipper length",
    x = "Flipper length (g)", y = "Count") + theme_minimal()
```



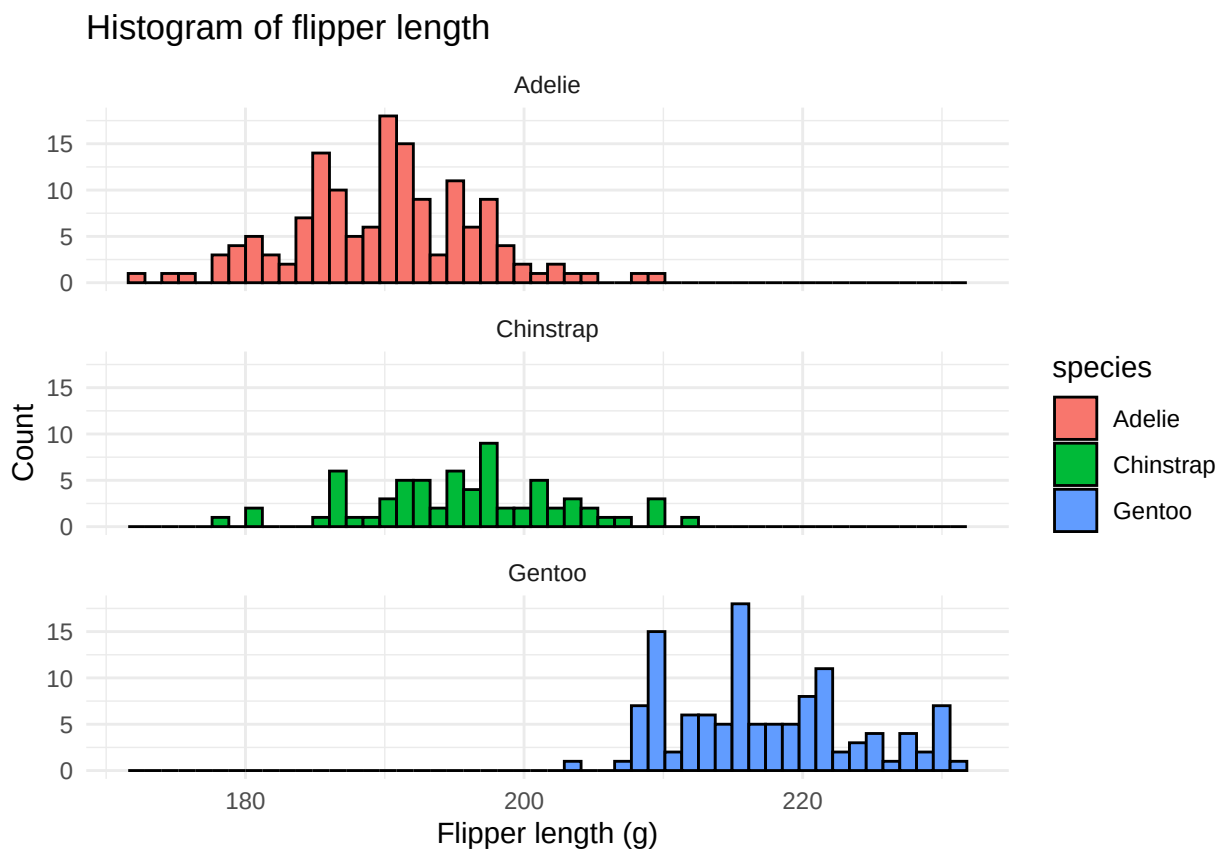

```
# You will see some of the bars for different species
# overlap To deal with this, within the geom_histogram(),
# which draws your bars, use transparency (alpha) on the
# fill of the bars so that overlapping bars are visible
ggplot(na.omit(penguins), aes(x = flipper_length_mm, fill = species)) +
    geom_histogram(position = "identity", alpha = 0.5, bins = 40,
        color = "black") + labs(title = "Histogram of flipper length",
    x = "Flipper length (g)", y = "Count") + theme_minimal()
```



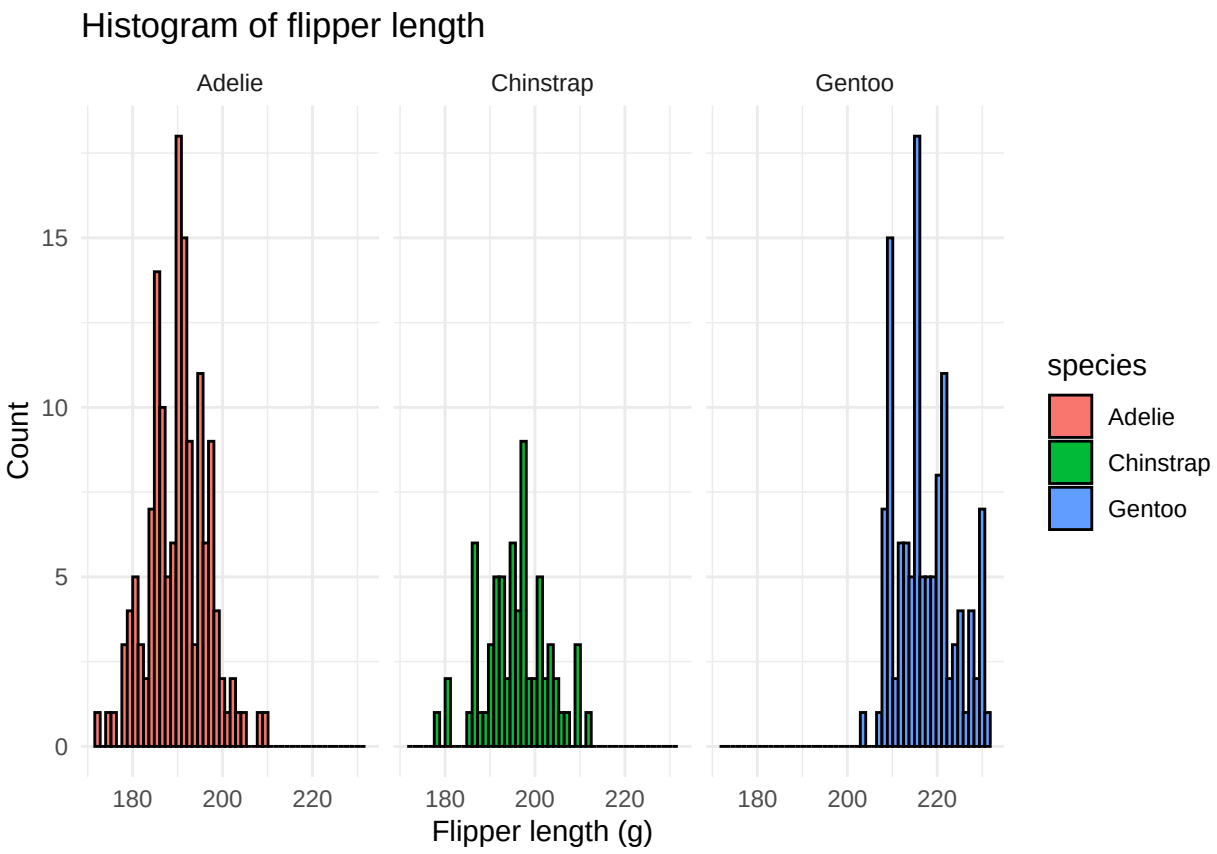
Exercise 3

Now experiment with the facet_wrap and facet_grid functions to create a histogram of flipper lengths for all three penguin species, but instead of one figure with overlapping bars you will now separate out the histogram for each species.

```
# Instead of coloring bars by species, it is cleaner to
# create separate graphs for each species Use the
# facet_wrap() function in ggplot to create multiples of a
# figure for a grouping variable.
ggplot(na.omit(penguins), aes(x = flipper_length_mm, fill = species)) +
    theme_minimal() + geom_histogram(bins = 50, color = "black") +
    labs(title = "Histogram of flipper length", x = "Flipper length (g)",
        y = "Count") + facet_wrap(~species, nrow = 3)
```




```
# Now explore the use of facet_grid instead and notice the
# difference in behavior of the resulting plots.
ggplot(na.omit(penguins), aes(x = flipper_length_mm, fill = species)) +
    theme_minimal() + geom_histogram(bins = 50, color = "black") +
    labs(title = "Histogram of flipper length", x = "Flipper length (g)",
        y = "Count") + facet_grid(cols = vars(species))
```

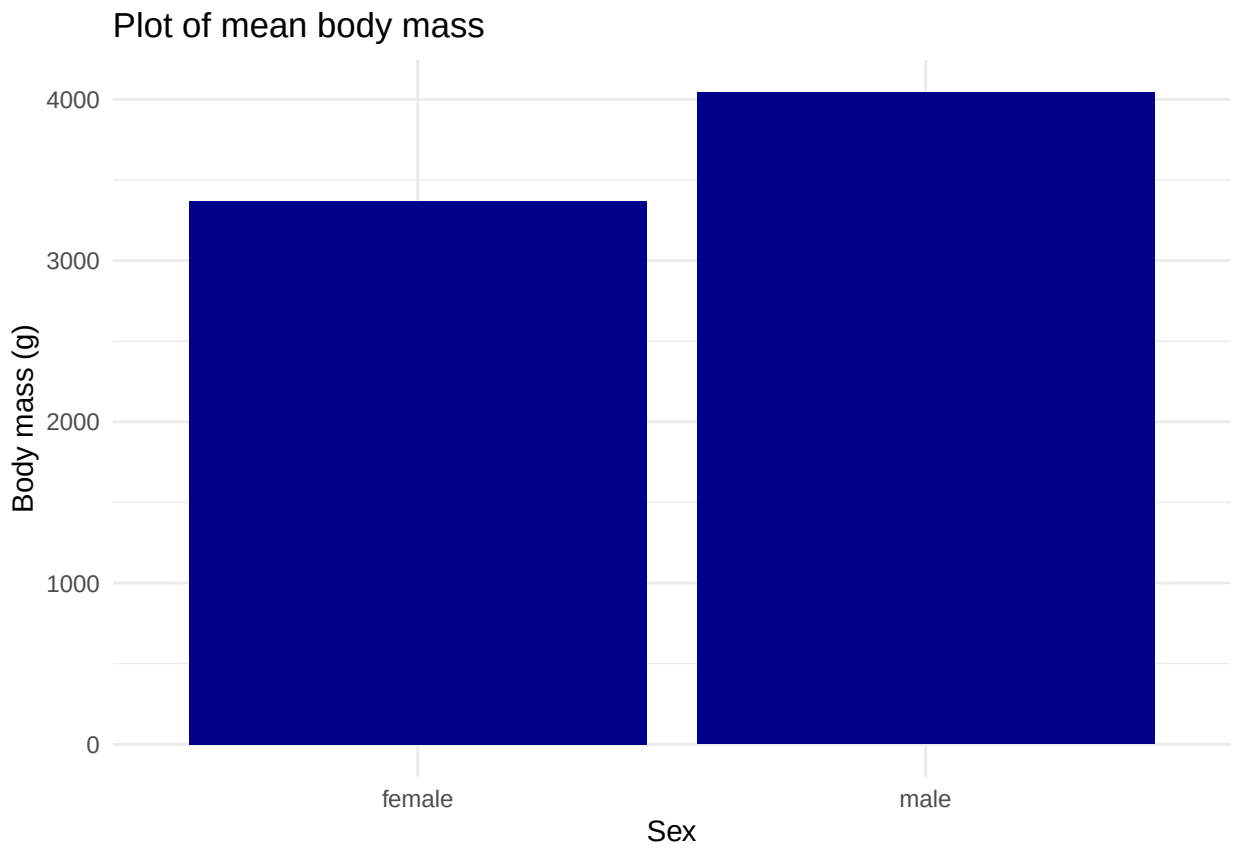


Exercise 4

Now let's explore the sexual dimorphism among Adelie penguins. Create a graph of the mean body mass by sex.

```
# Create a data summary by sex and species that calculates
# the mean and standard deviation of body mass
adelie_mass_summary <- na.omit(penguins) %>%
    filter(species == "Adelie") %>%
    group_by(sex) %>%
    summarize(mean_body_mass = mean(body_mass_g), sd_body_mass = sd(body_mass_g))

# Plot the mean body mass for males and females for adelie
# penguins.  Add appropriate labels.
adelie_mass_summary %>%
    ggplot(aes(x = sex, y = mean_body_mass)) + theme_minimal() +
    geom_col(fill = "darkblue") + labs(title = "Plot of mean body mass",
    x = "Sex", y = "Body mass (g)")
```




```
# Apply a theme of your choice to the graph. You can
# explore different themes by typing ?theme and going to
# complete themes or by adding a theme to your plot
# starting with theme_ and trying out the different options
# that pop up. There are even more themes available when
# you install and load the package 'ggthemes'

## already been using theme in functions above.
```

Exercise 5

A common element of bar charts that plot mean values is to add a measure of variation. In this exercise you will practice adding errorbars to a graph using the standard deviation from your data summary.

```
# Add error bars to the graph. Error bars should represent
# one standard deviation.  Note: Use the geometry
# geom_errorbar() for creating your error bars
adelie_mass_summary %>%
    ggplot(aes(x = sex, y = mean_body_mass, fill = sex)) + theme_minimal() +
    geom_col() + labs(title = "Plot of mean body mass", x = "Sex",
    y = "Body mass (g)") + geom_errorbar(aes(ymin = mean_body_mass -
    sd_body_mass, ymax = mean_body_mass + sd_body_mass), width = 0.3)
```


Plot of mean body mass
Body mass (g)
4000
3000
2000
1000
0
female
male
Sex
sex
female
male