

Code-a-long 2.2

Learning Outcomes

- Students will be able to use ggplot to generate box plots.
- Students will be able to apply error bars to box plots and bar plots.

```
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
```

```
library(palmerpenguins)
```

```
penguins<-palmerpenguins::penguins
```

```
penguins
```

```
# A tibble: 344 x 8
  species island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
  <fct>   <fct>         <dbl>         <dbl>         <int>         <int>
1 Adelie Torgersen     39.1          18.7          181          3750
2 Adelie Torgersen     39.5          17.4          186          3800
3 Adelie Torgersen     40.3           18           195          3250
```

```

4 Adelie Torgersen      NA      NA      NA      NA
5 Adelie Torgersen    36.7    19.3    193    3450
6 Adelie Torgersen    39.3    20.6    190    3650
7 Adelie Torgersen    38.9    17.8    181    3625
8 Adelie Torgersen    39.2    19.6    195    4675
9 Adelie Torgersen    34.1    18.1    193    3475
10 Adelie Torgersen    42      20.2    190    4250
# i 334 more rows
# i 2 more variables: sex <fct>, year <int>

```

Box plots with error bars

Strategy: Use `geom_boxplot()`, and in the `aes` function, set `x` equal to a categorical column, which will automatically group them. For the horizontal lines (error bars), add `stat_boxplot(geom = "errorbar")`.

```

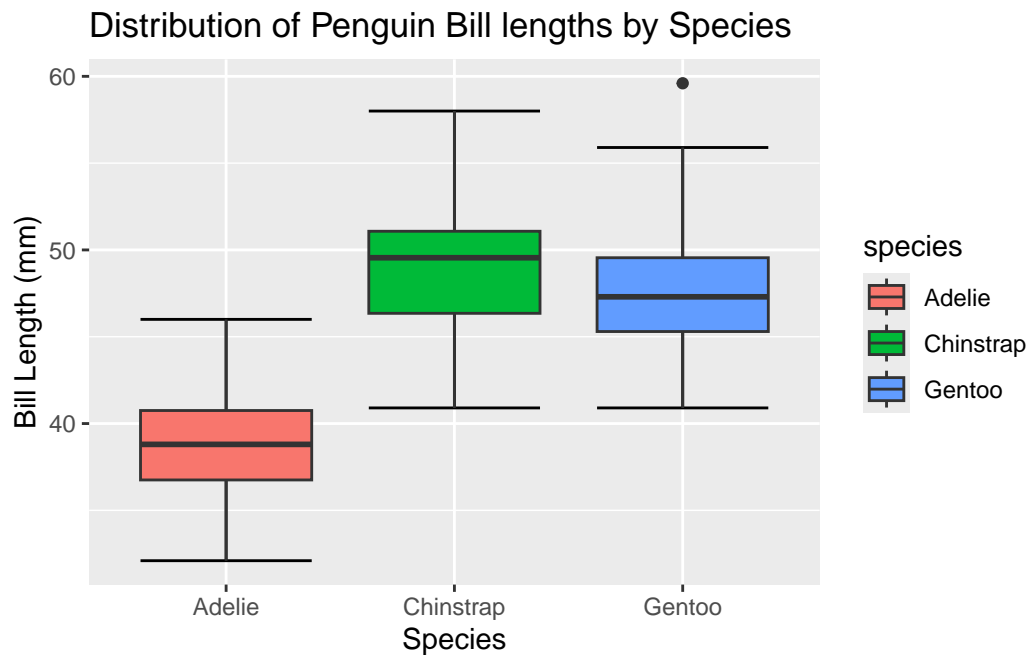
ggplot(data=penguins, mapping=aes(x=species, y=bill_length_mm, fill=species))+
  stat_boxplot(geom="errorbar")+
  geom_boxplot()+
  labs(title="Distribution of Penguin Bill lengths by Species",
       x="Species",
       y="Bill Length (mm)")

```

```

Warning: Removed 2 rows containing non-finite outside the scale range
(`stat_boxplot()`).
Removed 2 rows containing non-finite outside the scale range
(`stat_boxplot()`).

```



Bar plots with error bars

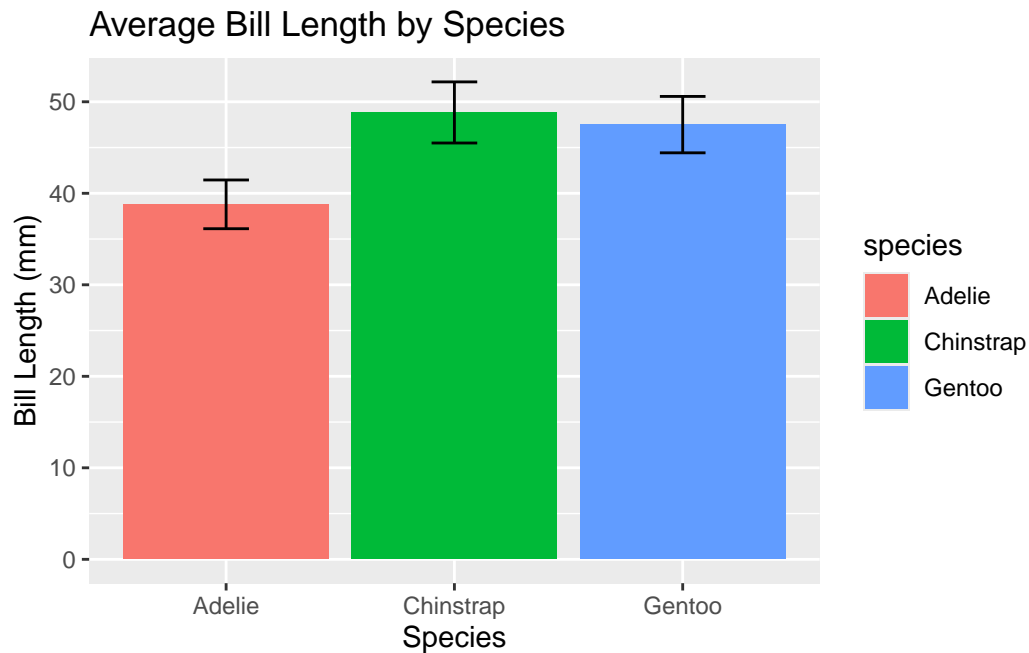
Strategy: use `group_by/summarize` to calculate group mean and standard deviation. Add `geom_errorbar`, in which the top bar is the mean + standard deviation, and the lower bar is the mean - standard deviation:

```
groupedBillLength<- penguins %>%
  group_by(species) %>%
  summarize(avgBillLength=mean(bill_length_mm, na.rm=TRUE),
            sdBillLength=sd(bill_length_mm, na.rm=TRUE))

groupedBillLength
```

```
# A tibble: 3 x 3
  species avgBillLength sdBillLength
  <fct>      <dbl>         <dbl>
1 Adelie      38.8           2.66
2 Chinstrap   48.8           3.34
3 Gentoo     47.5           3.08
```

```
ggplot(data=groupedBillLength, mapping=aes(x=species, y=avgBillLength, fill=species))+
  geom_bar(stat="identity")+
  geom_errorbar(mapping=aes(ymin=avgBillLength-sdBillLength,
                           ymax=avgBillLength+sdBillLength),
               width=0.2)+
  labs(title="Average Bill Length by Species",
       x="Species",
       y="Bill Length (mm)")
```



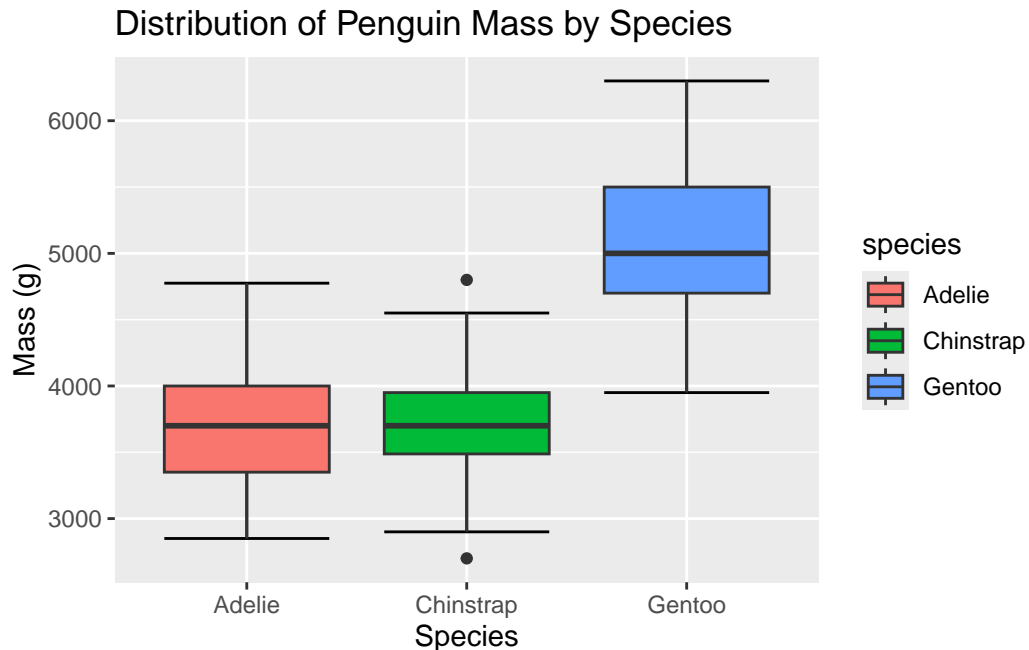
Practice

Create a box plot comparing penguin mass among species. Include error bars, a legend, and labels.

```
# create box plot below

ggplot(data=penguins, mapping=aes(x=species, y=body_mass_g, fill=species))+
  stat_boxplot(geom="errorbar")+
  geom_boxplot()+
  labs(title="Distribution of Penguin Mass by Species",
       x="Species", y="Mass (g)")
```

Warning: Removed 2 rows containing non-finite outside the scale range
 (`stat_boxplot()`).
 Removed 2 rows containing non-finite outside the scale range
 (`stat_boxplot()`).



Create a bar plot comparing penguin mass among species. Include error bars (± 1 sd), a legend, and labels.

```
# create bar plot below
```

```
groupedMass<- penguins %>%
  group_by(species) %>%
  summarize(avgMass=mean(body_mass_g, na.rm=TRUE),
            sdMass=sd(body_mass_g, na.rm=TRUE))
```

```
groupedMass
```

```
# A tibble: 3 x 3
  species   avgMass sdMass
  <fct>     <dbl>  <dbl>
1 Adelie   3701.   459.
2 Chinstrap 3733.   384.
3 Gentoo   5076.   504.
```

```
ggplot(data=groupedMass, mapping=aes(x=species, y=avgMass, fill=species))+
  geom_bar(stat="identity")+
  geom_errorbar(mapping=aes(ymin=avgMass-sdMass,
                           ymax=avgMass+sdMass),
               width=0.2)+
  labs(title="Average Penguin Mass by Species",
       x="Species",
       y="Mass (g)")
```

