Penguins Student Lab

Insert Student Name

Insert Date

Objectives

- Connect to Github to retrieve this script
- Pull an update to the script
- The object orientation of R
- Functions & Arguments
- Different types of values: numeric vs. character
- Simple data manipulation
- Graphing
- Statistical tests
- There is always more than one way to do things
- Knit final result into a PDF

```
library(tidyverse)
library(palmerpenguins)
```

Loading Libraries

Looking at the data. View your dataset.

For the purposes of Knitting this document afterwards, please use the View command, both in the console and in a chunk. Once you have used it in a chunk, try commenting it out. Remember your "#."

```
#View()
```

Find your column names.

```
colnames(penguins)
```

```
## [1] "species"          "island"            "bill_length_mm"
## [4] "bill_depth_mm"    "flipper_length_mm" "body_mass_g"
## [7] "sex"              "year"
```

Find the dimensions of the dataset.

```
dim(penguins)
```

```
## [1] 344   8
```

What are the classes of variables? Use the class function and don't forget your $ unless using the pipe operator!

What kind of variable is species?

```
class(penguins$species)
```

```
## [1] "factor"
```

```
penguins$species %>%
  class()
```

```
## [1] "factor"
```

What kind of variable is bill length?

```
class(penguins$bill_length_mm)
```

```
## [1] "numeric"
```

```
penguins$bill_length_mm %>%
  class()
```

```
## [1] "numeric"
```

Insert explanation here

```
two_species <- filter(penguins, species == "Adelie" | species == "Gentoo")

#OR

two_species <- penguins %>%
  filter(species == "Adelie" | species == "Gentoo")
```

Filter the gentoo and adelie species, and put it in a new dataset called `two_species`.

Making Boxplots   Make a box plot using your new two_species set looking at the differences between bill length of the two species of penguin. Add color based on the two species! Explain and interpret this plot.
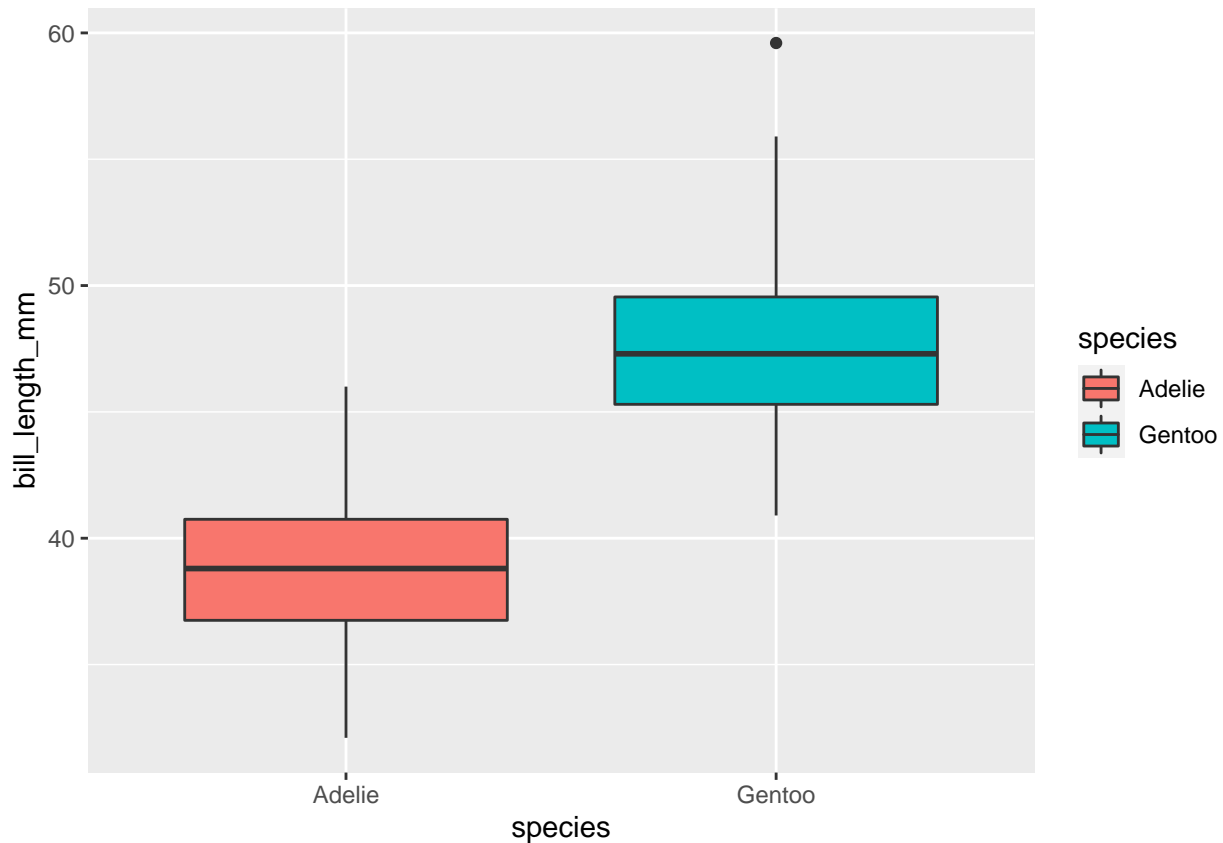
Insert explanation here

```
ggplot(two_species,
       aes(x = species, y = bill_length_mm, fill = species)) +
       geom_boxplot()
```
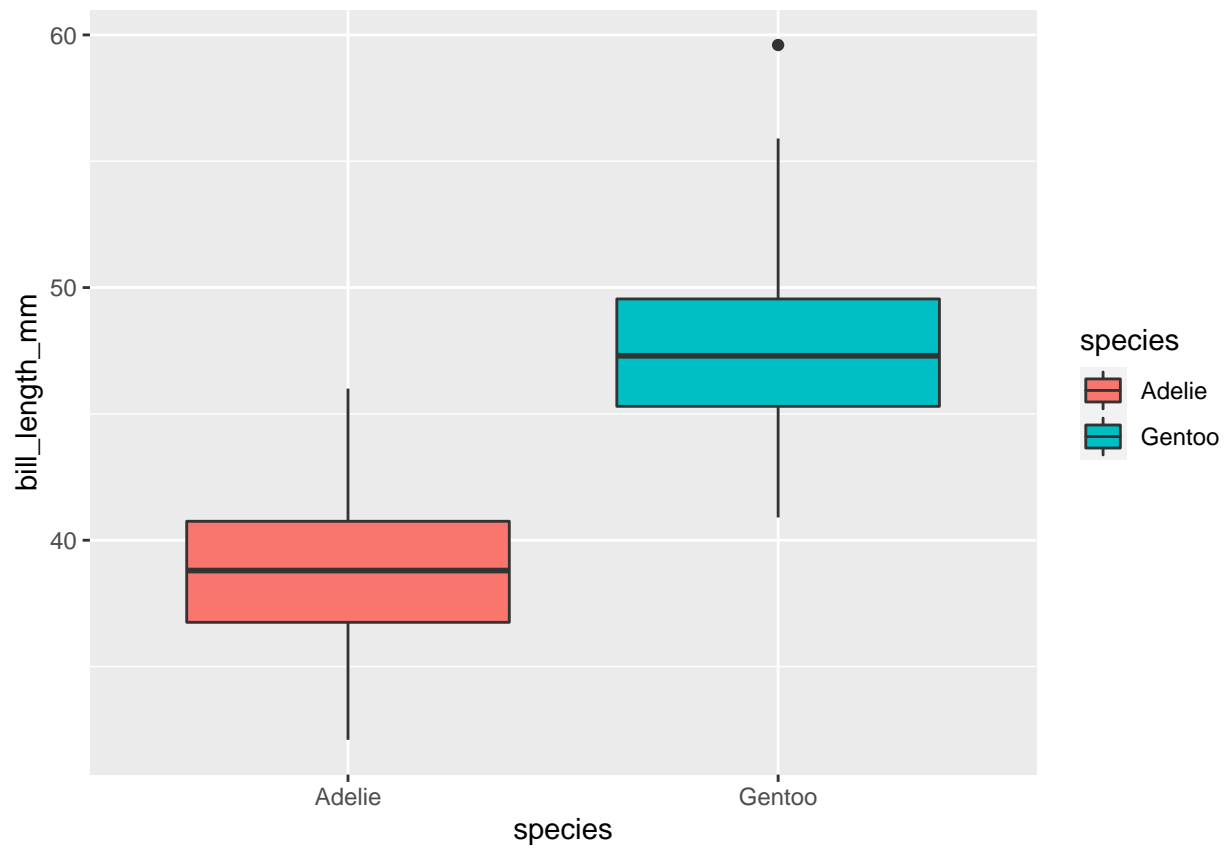
```
#OR

two_species %>%
  ggplot(
    aes(x = species, y = bill_length_mm, fill = species)) +
    geom_boxplot()
```

Run a t-test and a linear regression to see if there is a significant difference between the two species bill length averages. Hint use bill_length_mm for your dependent variable and species as your independent variable. *Insert explanation of t-test and model here*

```
t.test(bill_length_mm ~ species, data = two_species)
```

```
##
##  Welch Two Sample t-test
##
## data:  bill_length_mm by species
## t = -24.725, df = 242.58, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.407672 -8.019303
## sample estimates:
## mean in group Adelie mean in group Gentoo
##             38.79139             47.50488
```

```
mod_0 <- lm(bill_length_mm ~ species, data = two_species)

summary(mod_0)
```

```
##
## Call:
## lm(formula = bill_length_mm ~ species, data = two_species)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

4

```
## -6.6914 -2.1015 -0.0981  2.0052 12.0951
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   38.7914     0.2326  166.75   <2e-16 ***
## speciesGentoo  8.7135     0.3472   25.09   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.859 on 272 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.6984, Adjusted R-squared:  0.6973
## F-statistic: 629.8 on 1 and 272 DF,  p-value: < 2.2e-16
```

**Knit your final result for your own reference and to learn how to create an assignment that
could be turned in for future homework.**