STAT2430: Assignment 1

Dylan, Dempster and B00902362

2025-01-26

Look at the data

Here is a condensed overview of the `penguins` data

```
str(penguins)
```

```
## tibble [344 x 8] (S3: tbl_df/tbl/data.frame)
##  $ species          : Factor w/ 3 levels "Adelie","Chinstrap",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ island           : Factor w/ 3 levels "Biscoe","Dream",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ bill_length_mm   : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
##  $ bill_depth_mm    : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
##  $ flipper_length_mm: int [1:344] 181 186 195 NA 193 190 181 195 193 190 ...
##  $ body_mass_g      : int [1:344] 3750 3800 3250 NA 3450 3650 3625 4675 3475 4250 ...
##  $ sex              : Factor w/ 2 levels "female","male": 2 1 1 NA 1 2 1 2 NA NA ...
##  $ year             : int [1:344] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
```

and an overview of the `penguins_raw` supplementary data.

```
str(penguins_raw)
```

```
## tibble [344 x 17] (S3: tbl_df/tbl/data.frame)
##  $ studyName          : chr [1:344] "PAL0708" "PAL0708" "PAL0708" "PAL0708" ...
##  $ Sample Number      : num [1:344] 1 2 3 4 5 6 7 8 9 10 ...
##  $ Species            : chr [1:344] "Adelie Penguin (Pygoscelis adeliae)" "Adelie Penguin (Pygoscelis
##  $ Region             : chr [1:344] "Anvers" "Anvers" "Anvers" "Anvers" ...
##  $ Island             : chr [1:344] "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...
##  $ Stage              : chr [1:344] "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" "
##  $ Individual ID      : chr [1:344] "N1A1" "N1A2" "N2A1" "N2A2" ...
##  $ Clutch Completion  : chr [1:344] "Yes" "Yes" "Yes" "Yes" ...
##  $ Date Egg           : Date[1:344], format: "2007-11-11" "2007-11-11" ...
##  $ Culmen Length (mm) : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
##  $ Culmen Depth (mm)  : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
##  $ Flipper Length (mm): num [1:344] 181 186 195 NA 193 190 181 195 193 190 ...
##  $ Body Mass (g)      : num [1:344] 3750 3800 3250 NA 3450 ...
##  $ Sex                : chr [1:344] "MALE" "FEMALE" "FEMALE" NA ...
##  $ Delta 15 N (o/oo)  : num [1:344] NA 8.95 8.37 NA 8.77 ...
##  $ Delta 13 C (o/oo)  : num [1:344] NA -24.7 -25.3 NA -25.3 ...
##  $ Comments           : chr [1:344] "Not enough blood for isotopes." NA NA "Adult not sampled." ...
##  - attr(*, "spec")=
##   .. cols(
```

```
##   ..    studyName = col_character(),
##   ..    'Sample Number' = col_double(),
##   ..    Species = col_character(),
##   ..    Region = col_character(),
##   ..    Island = col_character(),
##   ..    Stage = col_character(),
##   ..    'Individual ID' = col_character(),
##   ..    'Clutch Completion' = col_character(),
##   ..    'Date Egg' = col_date(format = ""),
##   ..    'Culmen Length (mm)' = col_double(),
##   ..    'Culmen Depth (mm)' = col_double(),
##   ..    'Flipper Length (mm)' = col_double(),
##   ..    'Body Mass (g)' = col_double(),
##   ..    Sex = col_character(),
##   ..    'Delta 15 N (o/oo)' = col_double(),
##   ..    'Delta 13 C (o/oo)' = col_double(),
##   ..    Comments = col_character()
##   .. )
```

We will use the penguins dataset.

New visualization
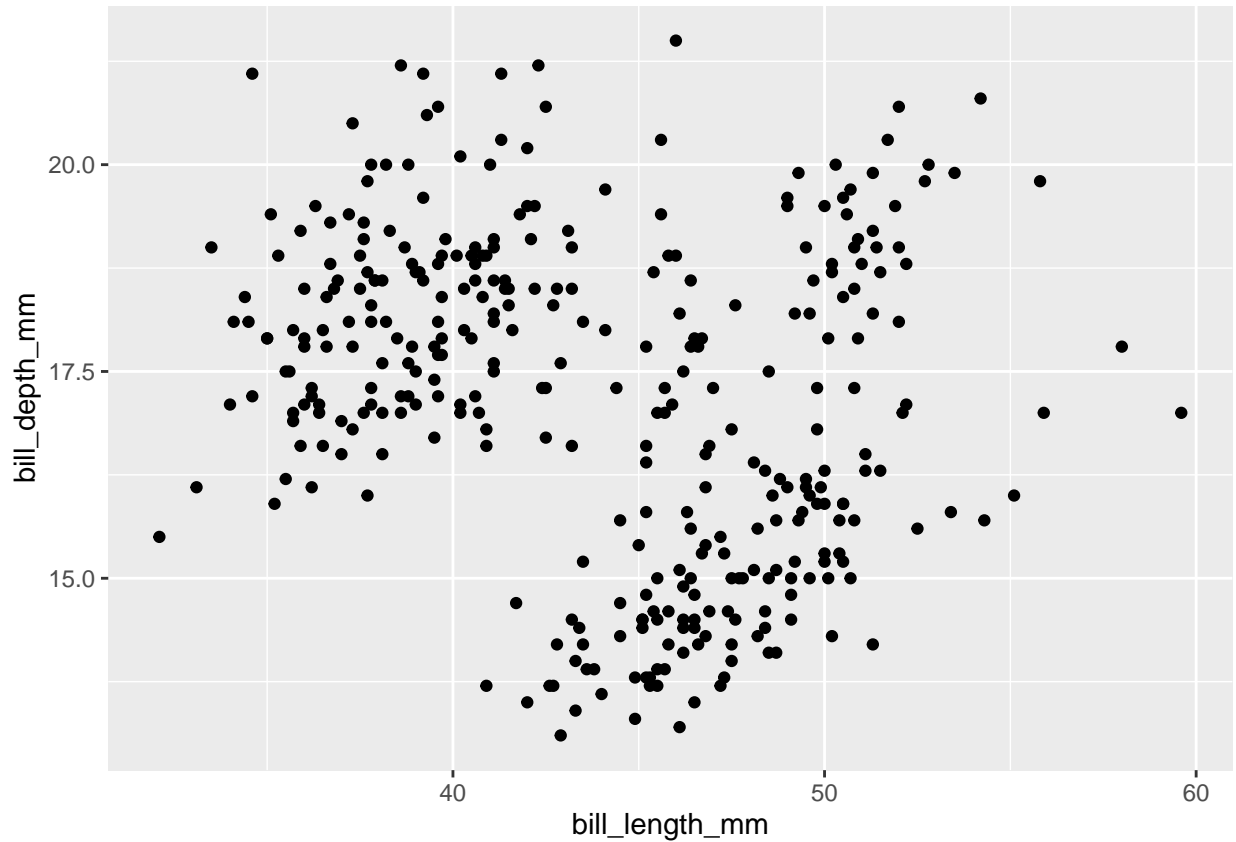
Step 1 (3 pts)

```
colnames(penguins)
```

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

```
ggplot(penguins, aes(x=bill_length_mm,y=bill_depth_mm))+geom_point()
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## ('geom_point()').
```
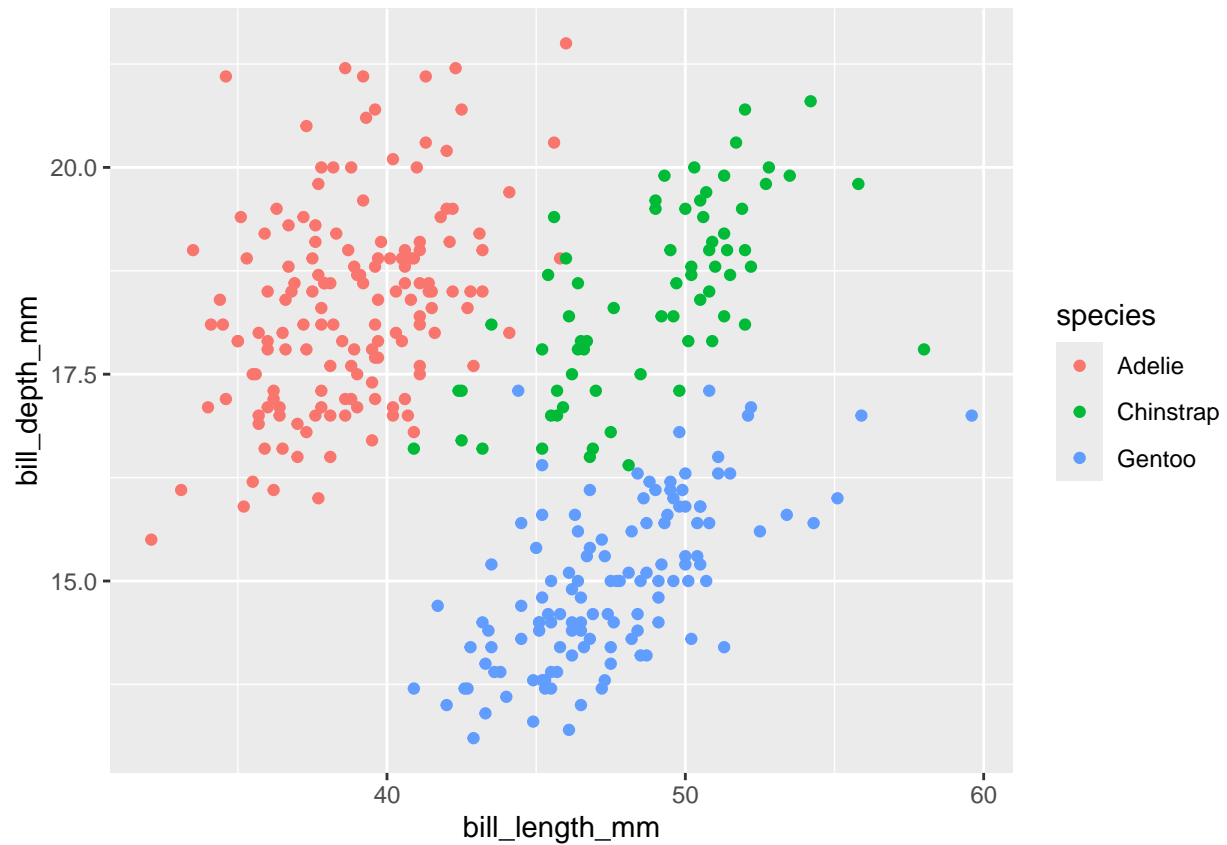

This scatter plot shows points scattered around, but with no clear differences between bill lenth and bill depth.

Step 2 (2 pts)

```
ggplot(penguins, aes(x=bill_length_mm,y=bill_depth_mm,colour=species))+geom_point()
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## ('geom_point()').
```

The addition of colour shows that each species has different cluster patterns, highlighting how bill length and bill depth differs across species of penguin.

Step 3 (2 pts)

```
ggplot(penguins, aes(x=bill_length_mm,y=bill_depth_mm,colour=species))+geom_point()+labs(
  title= "Bill Length and Bill Depth in Different Species of Penguins",
  x= "Bill Length (mm)",
  y= "Bill Depth (mm)"
)
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## ('geom_point()').
```

Bill Length and Bill Depth in Different Species of Penguins
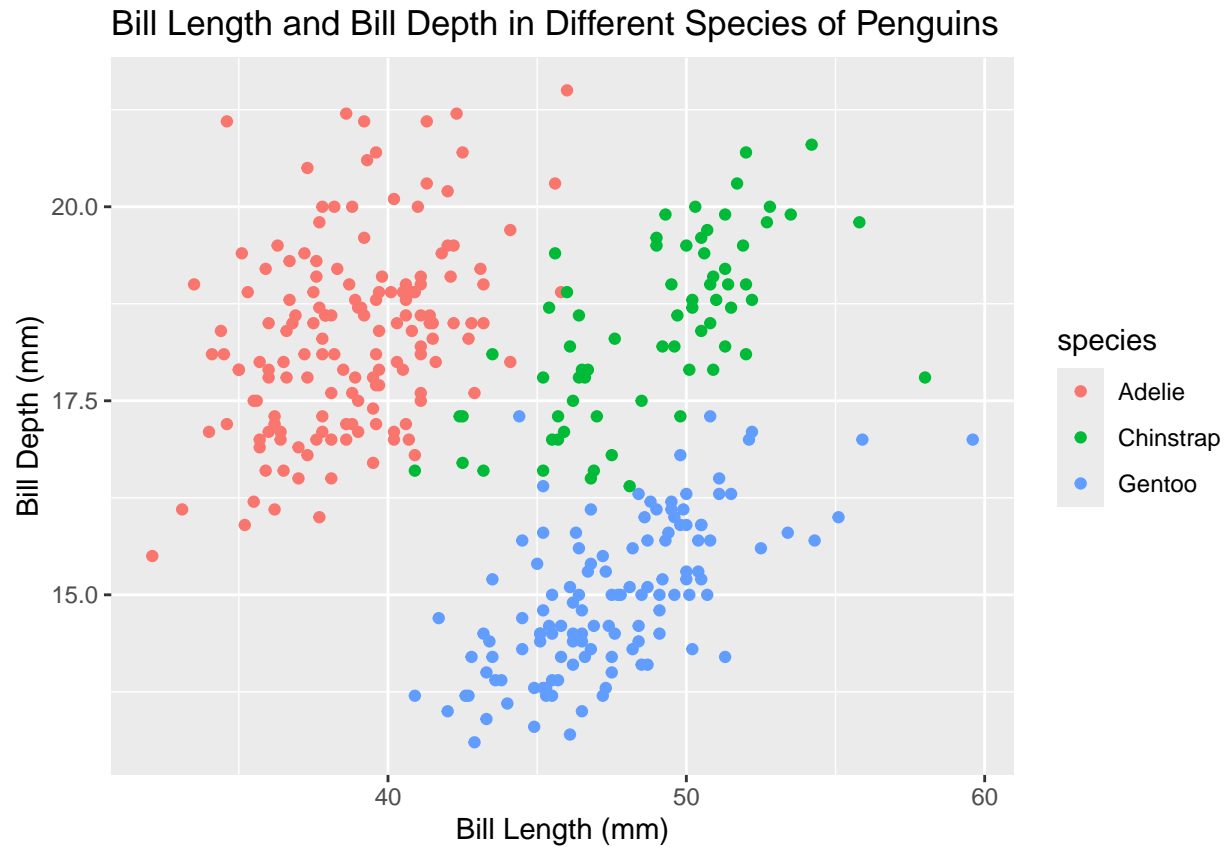
Another comparison

```
ggplot(penguins, aes(x=species, y=body_mass_g, fill=species))+geom_boxplot()+
  labs(
    title= "The Differences in Penguins Body Mass across Species",
    x= "Penguin Species",
    y= "Body Mass (g)"
  )
```

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

The Differences in Penguins Body Mass across Species