Summary report for penguins

統計114 張肇麟

2024-09-18

Table of contents

Summary Staistic

```r
library(Hmisc)
```

```
Attaching package: 'Hmisc'

The following objects are masked from 'package:base':

    format.pval, units
```

```r
library("palmerpenguins")
latex(describe(penguins), file = "", caption.placement = "top")
```

1

penguins
8 Variables   344 Observations

---

species

n	missing	distinct
344	0	3

Value	Adelie	Chinstrap	Gentoo
Frequency	152	68	124
Proportion	0.442	0.198	0.360

---

island

n	missing	distinct
344	0	3

Value	Biscoe	Dream	Torgersen
Frequency	168	124	52
Proportion	0.488	0.360	0.151

---

bill_length_mm

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	164	1	43.92	6.274	35.70	36.60	39.23	44.45	48.50	50.80	51.99

lowest : 32.1 33.1 33.5 34   34.1, highest: 55.1 55.8 55.9 58   59.6

---

bill_depth_mm

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	80	1	17.15	2.267	13.9	14.3	15.6	17.3	18.7	19.5	20.0

lowest : 13.1 13.2 13.3 13.4 13.5, highest: 20.7 20.8 21.1 21.2 21.5

---

flipper_length_mm

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	55	0.999	200.9	16.03	181.0	185.0	190.0	197.0	213.0	220.9	225.0

lowest : 172 174 176 178 179, highest: 226 228 229 230 231

---

body_mass_g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	94	1	4202	911.8	3150	3300	3550	4050	4750	5400	5650

lowest : 2700 2850 2900 2925 2975, highest: 5850 5950 6000 6050 6300

2

sex

```
        n   missing   distinct
      333        11          2
```

```
Value       female     male
Frequency      165      168
Proportion   0.495    0.505
```

year

```
        n   missing   distinct   Info    Mean      Gmd
      344         0          3   0.888    2008    0.8919
```

```
Value         2007   2008   2009
Frequency      110    114    120
Proportion   0.320  0.331  0.349
```

For the frequency table, variable is rounded to the nearest 0

Descriptive statistics

```
library(table1)
library(dplyr)

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" "Ad
 $ 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 ...
```

3

```
$ 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")=List of 3
 ..$ cols   :List of 17
 .. ..$ studyName          : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
 .. ..$ Sample Number      : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
 .. ..$ Species            : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
 .. ..$ Region             : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
 .. ..$ Island             : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
 .. ..$ Stage              : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
 .. ..$ Individual ID      : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
 .. ..$ Clutch Completion  : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
 .. ..$ Date Egg           :List of 1
 .. .. ..$ format: chr ""
 .. .. ..- attr(*, "class")= chr [1:2] "collector_date" "collector"
 .. ..$ Culmen Length (mm) : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
 .. ..$ Culmen Depth (mm)  : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
 .. ..$ Flipper Length (mm): list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
 .. ..$ Body Mass (g)      : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
 .. ..$ Sex                : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
 .. ..$ Delta 15 N (o/oo)  : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
 .. ..$ Delta 13 C (o/oo)  : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
 .. ..$ Comments           : list()
 .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
```

4

```
  ..$ default: list()
  .. ..- attr(*, "class")= chr [1:2] "collector_guess" "collector"
  ..$ skip   : num 1
  ..- attr(*, "class")= chr "col_spec"
```

```r
Mypenguins <- penguins_raw %>%
  mutate(Species = case_when(
    Species == "Adelie Penguin (Pygoscelis adeliae)" ~ "Adelie",
    Species == "Chinstrap penguin (Pygoscelis antarctica)" ~ "Chinstrap",
    Species == "Gentoo penguin (Pygoscelis papua)" ~ "Gentoo",
    TRUE ~ Species  # Keeps other values unchanged
  ))
Mypenguins[,c(3:6,8,14)] <- lapply(Mypenguins[,c(3:6,8,14)],as.factor)
str(Mypenguins)
```

```
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            : Factor w/ 3 levels "Adelie","Chinstrap",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Region             : Factor w/ 1 level "Anvers": 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 ...
 $ Stage              : Factor w/ 1 level "Adult, 1 Egg Stage": 1 1 1 1 1 1 1 1 1 1 ...
 $ Individual ID      : chr [1:344] "N1A1" "N1A2" "N2A1" "N2A2" ...
 $ Clutch Completion  : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 1 1 2 2 ...
 $ 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                : Factor w/ 2 levels "FEMALE","MALE": 2 1 1 NA 1 2 1 2 NA 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")=List of 3
  ..$ cols   :List of 17
  .. ..$ studyName          : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
  .. ..$ Sample Number      : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ Species            : list()
```

5

```
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Region           : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Island           : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Stage            : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Individual ID    : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Clutch Completion : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Date Egg         :List of 1
.. .. ..$ format: chr ""
.. .. ..- attr(*, "class")= chr [1:2] "collector_date" "collector"
.. ..$ Culmen Length (mm) : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Culmen Depth (mm)  : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Flipper Length (mm): list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Body Mass (g)    : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Sex              : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Delta 15 N (o/oo) : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Delta 13 C (o/oo) : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Comments         : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
..$ default: list()
.. ..- attr(*, "class")= chr [1:2] "collector_guess" "collector"
..$ skip   : num 1
..- attr(*, "class")= chr "col_spec"
```

```
table1(~ Region + Island + `Culmen Length (mm)` +
        `Culmen Depth (mm)` + `Flipper Length (mm)` +
        `Body Mass (g)` + Sex| Species, data = Mypenguins)
```

6

|	Adelie	Chinstrap	Gentoo	Overall
	(N=152)	(N=68)	(N=124)	(N=344)
Region				
Anvers	152 (100%)	68 (100%)	124 (100%)	344 (100%)
Island				
Biscoe	44 (28.9%)	0 (0%)	124 (100%)	168 (48.8%)
Dream	56 (36.8%)	68 (100%)	0 (0%)	124 (36.0%)
Torgersen	52 (34.2%)	0 (0%)	0 (0%)	52 (15.1%)
Culmen Length (mm)				
Mean (SD)	38.8 (2.66)	48.8 (3.34)	47.5 (3.08)	43.9 (5.46)
Median [Min, Max]	38.8 [32.1, 46.0]	49.6 [40.9, 58.0]	47.3 [40.9, 59.6]	44.5 [32.1, 59.6]
Missing	1 (0.7%)	0 (0%)	1 (0.8%)	2 (0.6%)
Culmen Depth (mm)				
Mean (SD)	18.3 (1.22)	18.4 (1.14)	15.0 (0.981)	17.2 (1.97)
Median [Min, Max]	18.4 [15.5, 21.5]	18.5 [16.4, 20.8]	15.0 [13.1, 17.3]	17.3 [13.1, 21.5]
Missing	1 (0.7%)	0 (0%)	1 (0.8%)	2 (0.6%)
Flipper Length (mm)				
Mean (SD)	190 (6.54)	196 (7.13)	217 (6.48)	201 (14.1)
Median [Min, Max]	190 [172, 210]	196 [178, 212]	216 [203, 231]	197 [172, 231]
Missing	1 (0.7%)	0 (0%)	1 (0.8%)	2 (0.6%)
Body Mass (g)				
Mean (SD)	3700 (459)	3730 (384)	5080 (504)	4200 (802)
Median [Min, Max]	3700 [2850, 4780]	3700 [2700, 4800]	5000 [3950, 6300]	4050 [2700, 6300]
Missing	1 (0.7%)	0 (0%)	1 (0.8%)	2 (0.6%)
Sex				
FEMALE	73 (48.0%)	34 (50.0%)	58 (46.8%)	165 (48.0%)
MALE	73 (48.0%)	34 (50.0%)	61 (49.2%)	168 (48.8%)
Missing	6 (3.9%)	0 (0%)	5 (4.0%)	11 (3.2%)

Visualization

Missing Values

```
library(Hmisc)
library(DataExplorer)
plot_missing(Mypenguins)
```

7
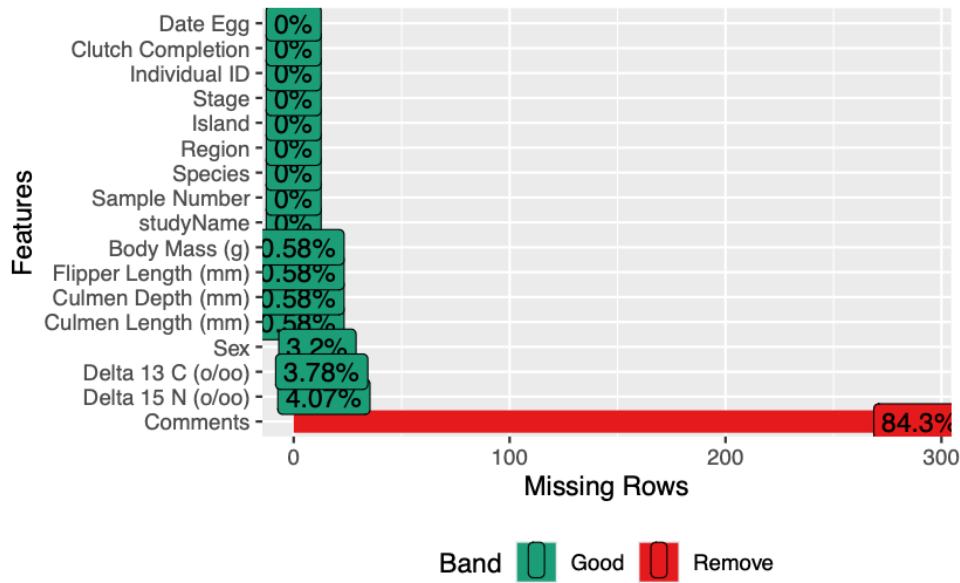
Plots

Scatterplots

```
library(ggplot2)
ggplot(data = Mypenguins, aes(x = `Flipper Length (mm)`, y = `Body Mass (g)`)) +
  geom_point(na.rm = TRUE, aes(color = Species), size = 2, alpha = 0.9) +
  scale_color_manual(values = c("#7BD5F5", "#787FF6", "#FB7B8E")) +
  labs(title = "Flipper and Body Mass by Species",
       x = "Flipper Length (mm)",
       y = "Body Mass (g)") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```
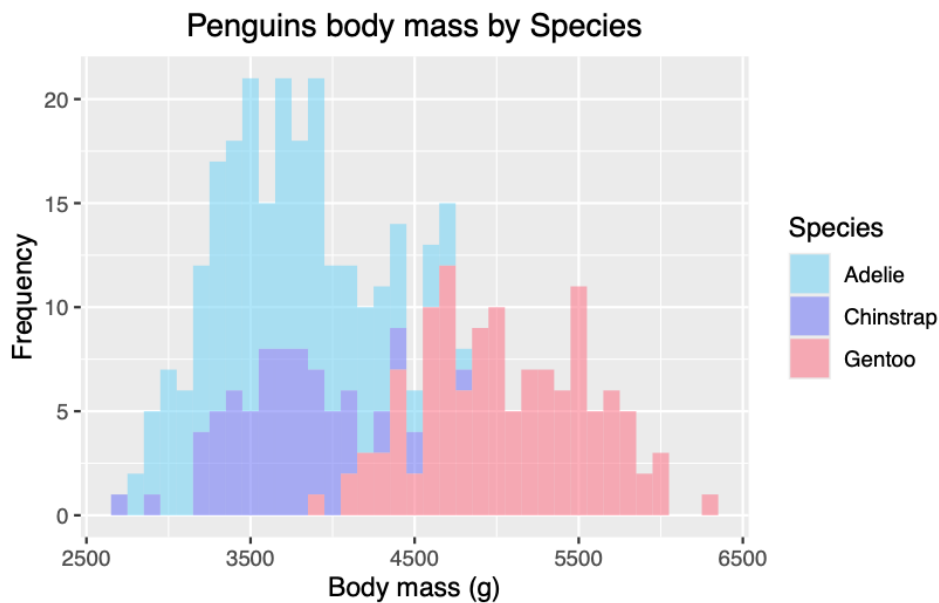
8

Flipper and Body Mass by Species
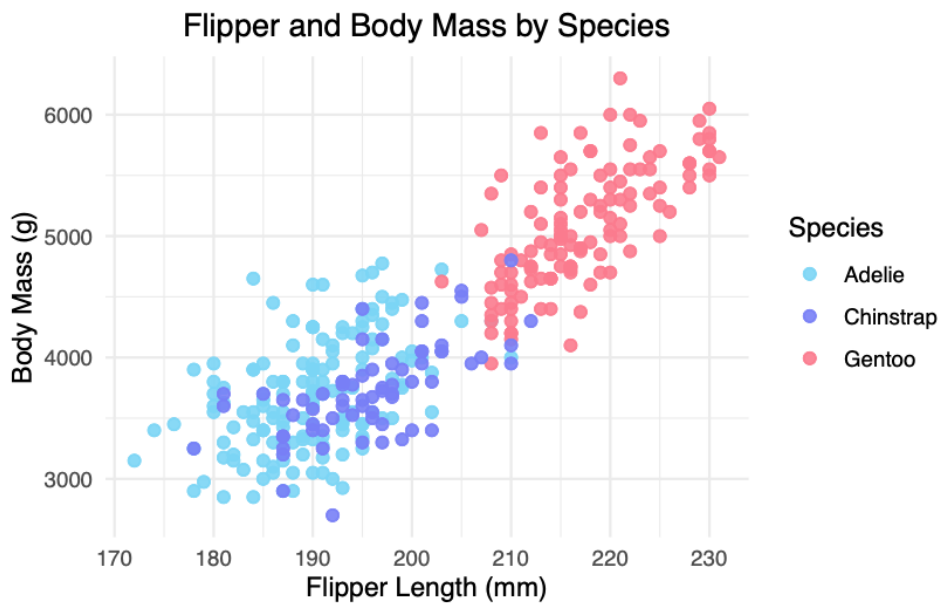

```
ggplot(data = Mypenguins, aes(x = `Flipper Length (mm)`, y = `Body Mass (g)`)) +
  geom_point(na.rm = TRUE, aes(color = Island), size = 2, alpha = 0.9) +
  scale_color_manual(values = c("#7BD5F5", "#787FF6", "#FB7B8E")) +
  labs(title = "Flipper and Body Mass by Island",
       x = "Flipper Length (mm)",
       y = "Body Mass (g)") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```

9

Flipper and Body Mass by Island
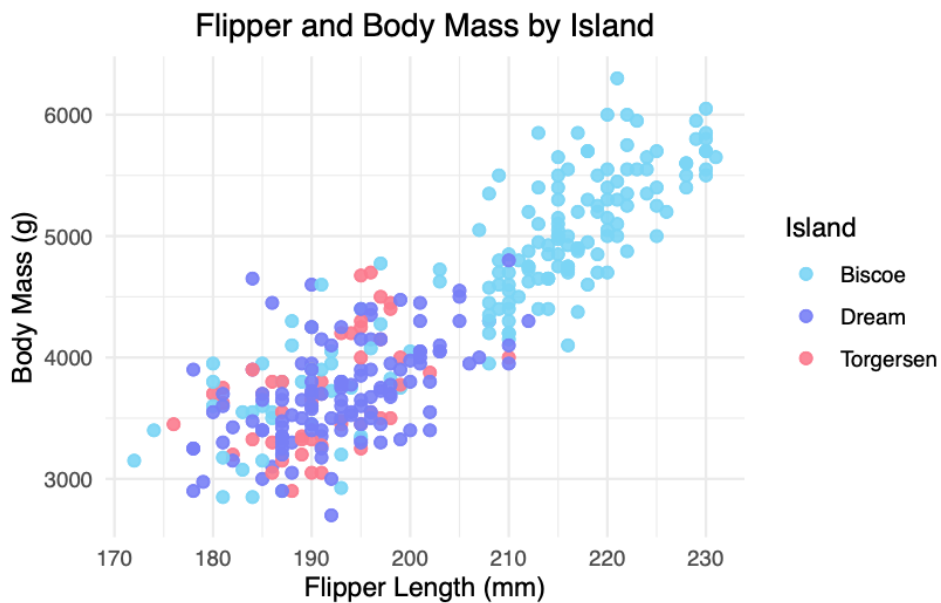
Histograms

```
ggplot(data = Mypenguins, aes(x = `Body Mass (g)`)) +
  geom_histogram(aes(fill = Species),
                 alpha = 0.6,
                 na.rm = TRUE,
                 binwidth = 100) +
  scale_fill_manual(values = c("#7BD5F5", "#787FF6", "#FB7B8E")) +
  labs(x = "Body mass (g)",
       y = "Frequency",
       title = "Penguins body mass by Species") +
  theme(plot.title = element_text(hjust = 0.5))
```

10

Penguins body mass by Species


```
ggplot(data = Mypenguins, aes(x = `Body Mass (g)`)) +
  geom_histogram(aes(fill = Island),
                 alpha = 0.6,
                 na.rm = TRUE,
                 binwidth = 100) +
  scale_fill_manual(values = c("#7BD5F5", "#787FF6", "#FB7B8E")) +
  labs(x = "Body mass (g)",
       y = "Frequency",
       title = "Penguins body mass by Island") +
  theme(plot.title = element_text(hjust = 0.5))
```

11

Penguins body mass by Island
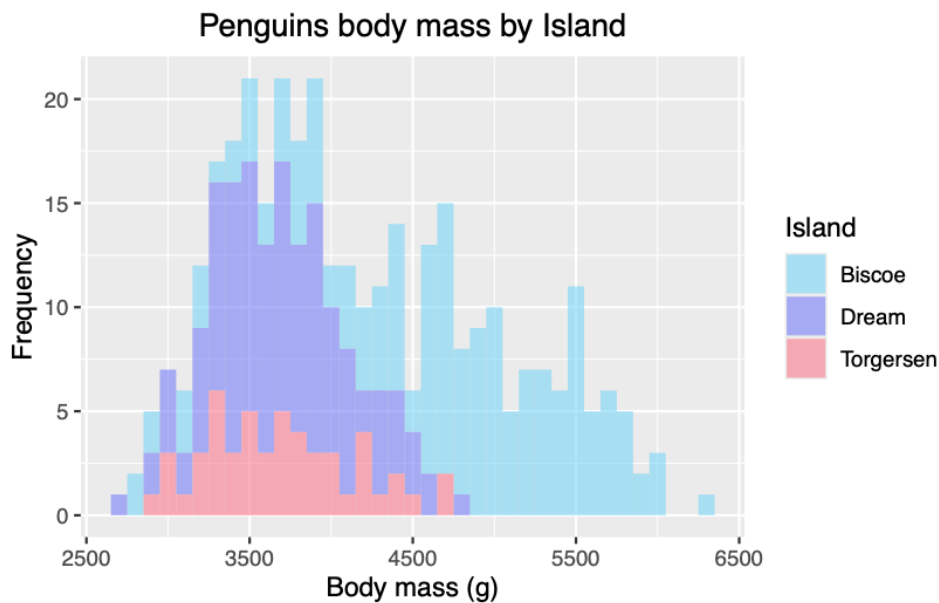

```
ggplot(Mypenguins, aes(x = `Flipper Length (mm)`,
                       y = `Body Mass (g)`)) +
  geom_point(aes(color = Sex)) +
  scale_color_manual(values = c("#FB7B8E", "#7BD5F5"),
                     na.translate = FALSE) +
  labs(title = "Penguin Flipper Length and Body Mass",
       x = "Flipper Length (mm)",
       y = "Body Mass (g)",
       color = "Penguin sex") +
  theme(legend.position = "bottom",
        plot.title.position = "plot",
        plot.title = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.5),
        plot.caption.position = "plot") +
  facet_wrap(~Species)
```

Warning: Removed 11 rows containing missing values or values outside the scale range
(`geom_point()`).

12

Penguin Flipper Length and Body Mass
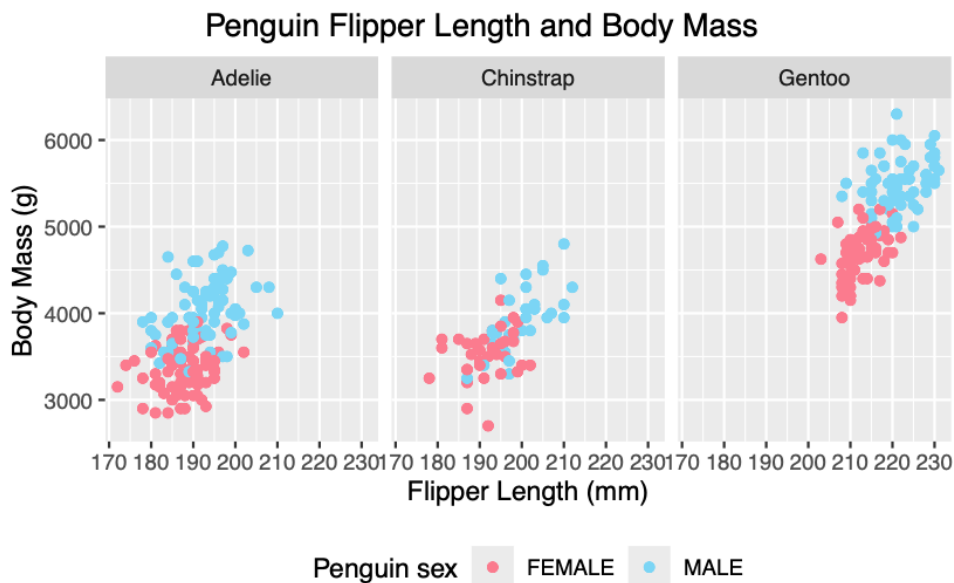

```
ggplot(Mypenguins, aes(x = `Culmen Length (mm)`,
                       y = `Culmen Depth (mm)`)) +
  geom_point(aes(color = Sex)) +
  scale_color_manual(values = c("#FB7B8E", "#7BD5F5"),
                     na.translate = FALSE) +
  labs(title = "Penguin Culmen Length and Culmen Depth",
       x = "Flipper Length (mm)",
       y = "Body Mass (g)",
       color = "Penguin sex") +
  theme(legend.position = "bottom",
        plot.title.position = "plot",
        plot.title = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.5),
        plot.caption.position = "plot") +
  facet_wrap(~Species)
```

```
Warning: Removed 11 rows containing missing values or values outside the scale range
(`geom_point()`).
```

13

Penguin Culmen Length and Culmen Depth
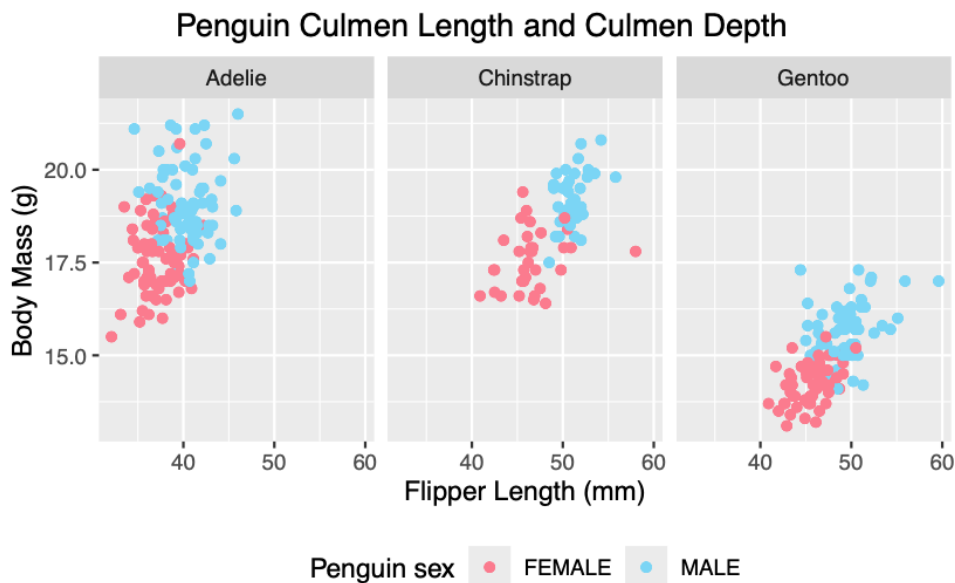

```
ggplot(data = Mypenguins, aes(x = Species, y = `Flipper Length (mm)`)) +
  geom_boxplot(aes(color = Species),
               width = 0.3, show.legend = FALSE) +
  geom_jitter(aes(color = Species), alpha = 0.8,
              show.legend = FALSE, position = position_jitter(width = 0.225)) +
  scale_color_manual(values = c("#7BD5F5", "#787FF6", "#FB7B8E")) +
  labs(title = "Flipper length between species",
    x = "Species",
    y = "Flipper length (mm)") +
  theme(plot.title = element_text(hjust = 0.5))
```

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

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

14

Flipper length between species