

A Summary Report for the palmerpenguins

統計114 簡玲萱

2024-09-15

Table of contents

Summary Statistics	1
Descriptive statistics	3
Visualization	5
Missing value	5
Scatter plot	6
Bar plot	8
Boxplot	10

Summary Statistics

```
library(palmerpenguins)
data(package='palmerpenguins')
penguins_data <- penguins_raw
penguins_data[,c(3:5,8,14)] <- lapply(penguins_data[,c(3:5,8,14)], as.factor)

library(Hmisc)
```

Attaching package: 'Hmisc'

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

format.pval, units

```
latex(describe(penguins_data), file = "", caption.placement = "top")
```

penguins_data
17 Variables 344 Observations

studyName						
n	missing	distinct				
344	0	3				
Value	PAL0708	PAL0809	PAL0910			
Frequency	110	114	120			
Proportion	0.320	0.331	0.349			

Sample Number



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
344	0	152	1	63.15	46.35	6.15	12.00	29.00	58.00	95.25	121.00	134.85

lowest : 1 2 3 4 5, highest: 148 149 150 151 152

Species



n	missing	distinct
344	0	3

Value	Adelie Penguin (Pygoscelis adeliae)	Chinstrap penguin (Pygoscelis antarctica)
Frequency	152	68
Proportion	0.442	0.198
Value	Gentoo penguin (Pygoscelis papua)	
Frequency	124	
Proportion	0.360	

Region

n	missing	distinct	value
344	0	1	Anvers

Value	Anvers
Frequency	344
Proportion	1

Island



n	missing	distinct
344	0	3

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

Stage

n	missing	distinct	value
344	0	1	Adult, 1 Egg Stage

Value	Adult, 1 Egg Stage
Frequency	344
Proportion	1

Individual ID

n	missing	distinct
344	0	190

lowest : N100A1 N100A2 N10A1 N10A2 N11A1 , highest: N98A2 N99A1 N99A2 N9A1 N9A2

Clutch Completion

n	missing	distinct
344	0	2

Value	No	Yes
Frequency	36	308
Proportion	0.105	0.895

Date Egg



n	missing	distinct	Info	Mean	Gmd	.05	.10
344	0	50	0.999	2008-11-27	328	2007-11-12	2007-11-16
.25	.50	.75	.90	.95			
2007-11-28	2008-11-09	2009-11-16	2009-11-22	2009-11-26			

lowest : 2007-11-09 2007-11-10 2007-11-11 2007-11-12 2007-11-13
highest: 2009-11-22 2009-11-23 2009-11-25 2009-11-27 2009-12-01

Culmen 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

Culmen 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

Sex

n	missing	distinct
333	11	2

Value	FEMALE	MALE
Frequency	165	168
Proportion	0.495	0.505

Δ 15 N (o/oo):

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
330	14	330	1	8.733	0.6323	7.897	8.047	8.300	8.652	9.172	9.491	9.689

lowest : 7.6322 7.63452 7.63884 7.68528 7.6887 , highest: 9.93727 9.98044 10.0202 10.0237 10.0254

Δ 13 C (o/oo):

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
331	13	331	1	-25.69	0.9093	-26.79	-26.69	-26.32	-25.83	-25.06	-24.53	-24.36

lowest : -27.0185 -26.9547 -26.8964 -26.8648 -26.8635, highest: -24.1657 -24.1026 -23.9031 -23.8902 -23.7877

Comments

n	missing	distinct
54	290	10

lowest : Adult not sampled.

highest: No blood sample obtained.

Adult not sampled. Nest never observed with full clutch.
No delta15N data received from lab.

Descriptive statistics

```
library(table1)
str(penguins_data)
```

```

tibble [344 x 17] (S3: tbl_df/tbl/data.frame)
 $ studyName      : chr [1:344] "PAL0708" "PAL0708" "PAL0708" ...
 $ Sample Number  : num [1:344] 1 2 3 4 5 6 7 8 9 10 ...
 $ Species        : Factor w/ 3 levels "Adelie Penguin (Pygoscelis adeliae)",...: 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          : 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 : 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()
 .. .. ..- 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(~ `Culmen Length (mm)`+`Culmen Depth (mm)`+`Flipper Length (mm)`+
`Body Mass (g)`+`Delta 15 N (o/oo)`+`Delta 13 C (o/oo)`| Species,
data=penguins_data)

```

	Adelie Penguin (Pygoscelis adeliae)	Chinstrap penguin (Pygoscelis antarctica)	Gentoo penguin (Pygoscelis papua)	Overall
	(N=152)	(N=68)	(N=124)	(N=344)
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%)
Delta 15 N (o/oo)				
Mean (SD)	8.86 (0.426)	9.36 (0.369)	8.25 (0.264)	8.73 (0.552)
Median [Min, Max]	8.88 [7.70, 9.80]	9.37 [8.47, 10.0]	8.25 [7.63, 8.83]	8.65 [7.63, 10.0]
Missing	11 (7.2%)	1 (1.5%)	2 (1.6%)	14 (4.1%)
Delta 13 C (o/oo)				
Mean (SD)	-25.8 (0.588)	-24.5 (0.239)	-26.2 (0.539)	-25.7 (0.794)
Median [Min, Max]	-26.0 [-26.8, -23.9]	-24.6 [-25.1, -23.8]	-26.2 [-27.0, -25.0]	-25.8 [-27.0, -23.8]
Missing	11 (7.2%)	0 (0%)	2 (1.6%)	13 (3.8%)

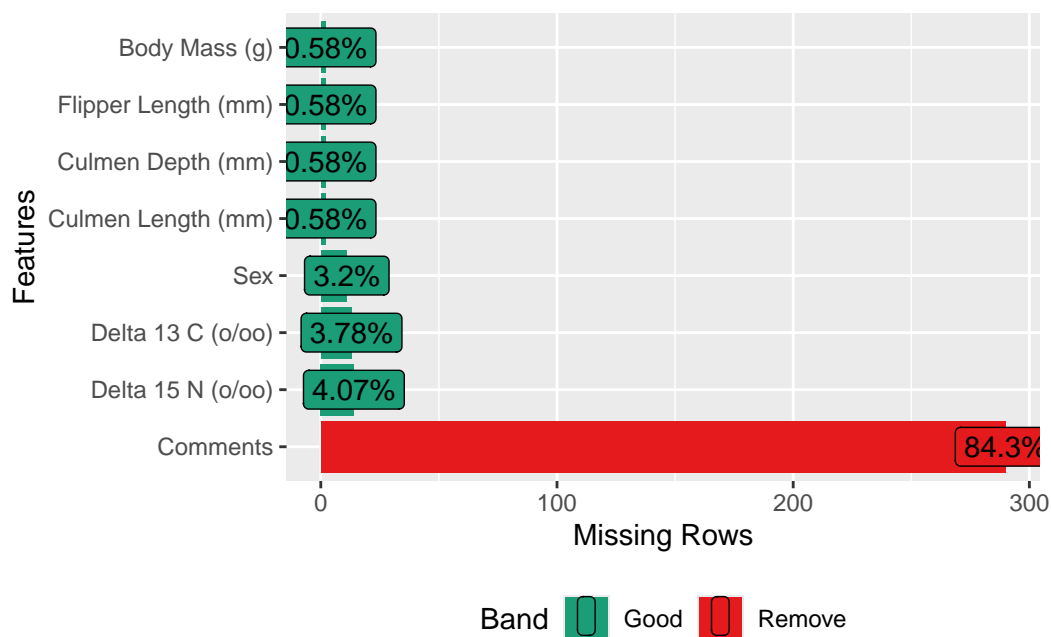
```
table1(~ Region + Island +`Clutch Completion`+ Sex | Species, data=penguins_data)
```

	Adelie Penguin (Pygoscelis adeliae)	Chinstrap penguin (Pygoscelis antarctica)	Gentoo penguin (Pygoscelis papua)	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%)
Clutch Completion				
No	14 (9.2%)	14 (20.6%)	8 (6.5%)	36 (10.5%)
Yes	138 (90.8%)	54 (79.4%)	116 (93.5%)	308 (89.5%)
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 value

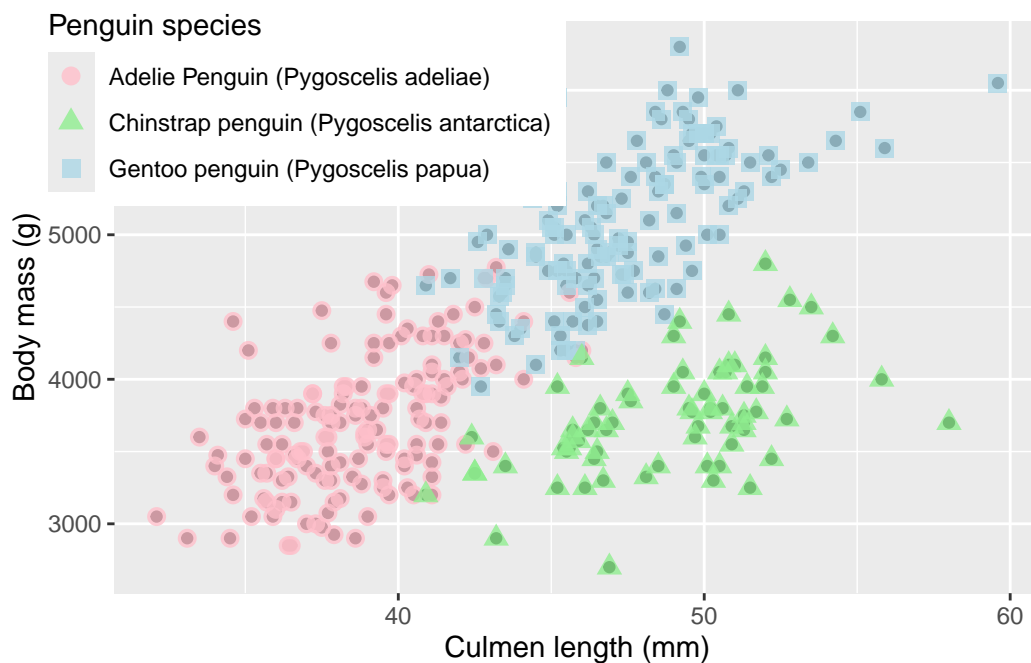
```
library(DataExplorer)
plot_missing(penguins_data, missing_only=TRUE)
```



Scatter plot

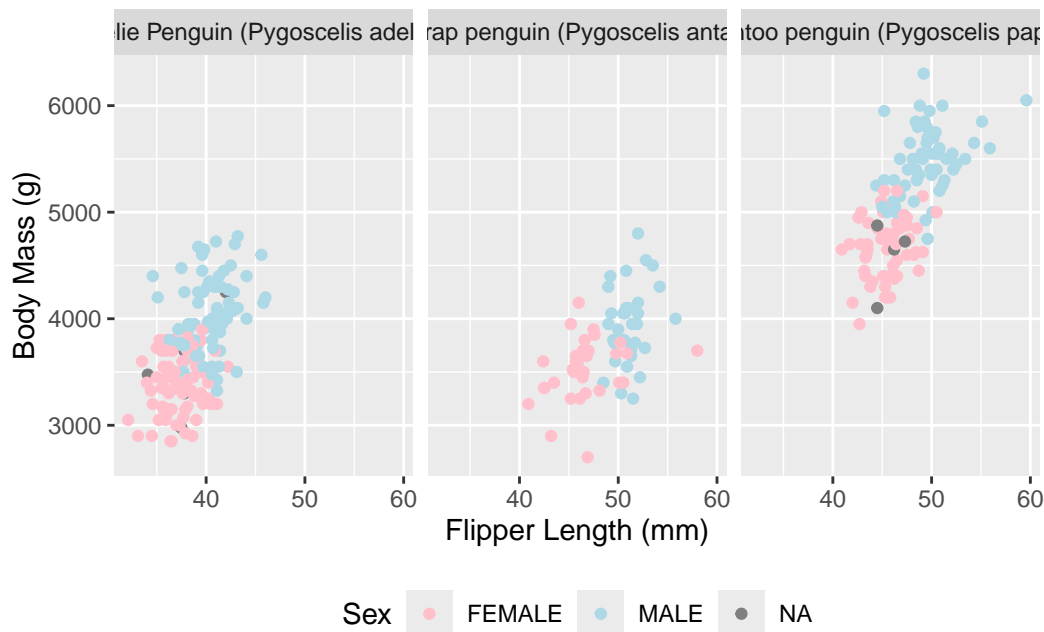
Culmen length and body mass for species

```
library(ggplot2)
ggplot(data = penguins_data, aes(x = `Culmen Length (mm)`, y = `Body Mass (g)`) +
  geom_point(na.rm = TRUE) +
  geom_point(aes(color = Species, shape = Species), size = 3, alpha = 0.8) +
  scale_color_manual(values = c("pink", "lightgreen", "lightblue")) +
  labs(x = "Culmen length (mm)", y = "Body mass (g)", color = "Penguin species", shape = "Penguin species") +
  theme(legend.position = c(0.2, 0.86))
```



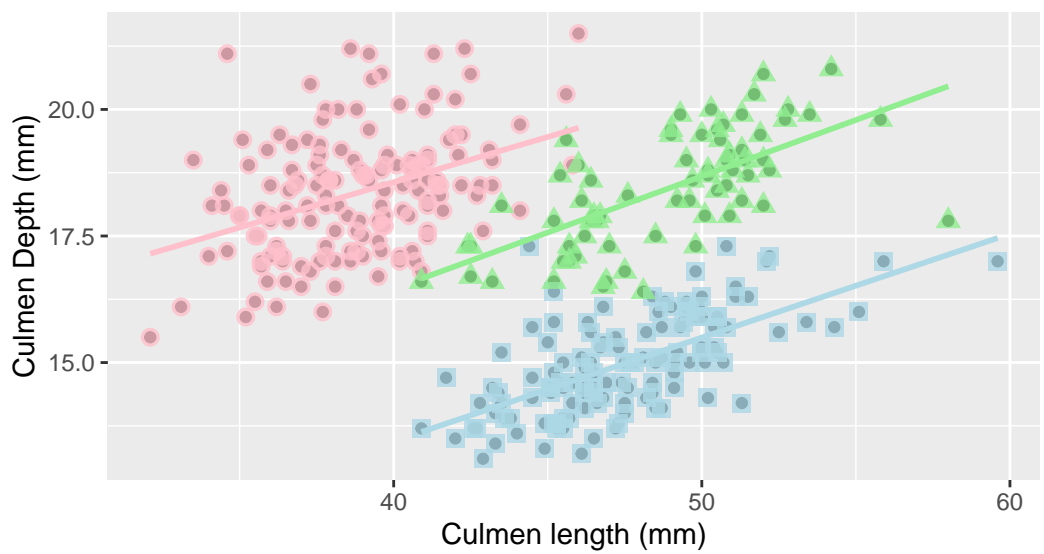
```
ggplot(data = penguins_data, aes(x = `Culmen Length (mm)`, y = `Body Mass (g)`) +
  geom_point(aes(color = Sex)) +
  scale_color_manual(values = c("pink", "lightblue")) +
```

```
labs(x = "Flipper Length (mm)", y = "Body Mass (g)") +
theme(legend.position = 'bottom') +
facet_wrap(~Species)
```



Culmen length and depth for species

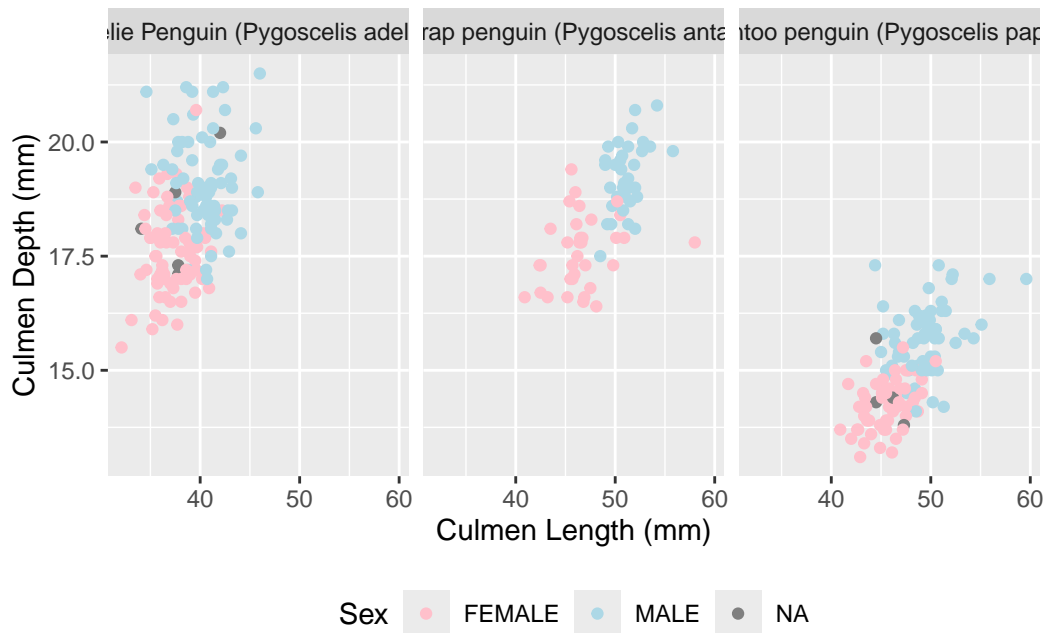
```
ggplot(data = penguins_data, aes(x = `Culmen Length (mm)`, y = `Culmen Depth (mm)`) +
  geom_point(na.rm = TRUE) +
  geom_point(aes(color = Species, shape = Species), size = 3, alpha = 0.8) +
  geom_smooth(method = "lm", se = FALSE, aes(color = Species)) +
  scale_color_manual(values = c("pink", "lightgreen", "lightblue")) +
  labs(x = "Culmen length (mm)", y = "Culmen Depth (mm)",
       color = "Penguin species", shape = "Penguin species") +
  theme(legend.position = 'bottom')
```



Adelia Penguin (*Pygoscelis adeliae*) Chinstrap penguin (*Pygoscelis antarctica*) King penguin (*Pygoscelis patagonicus*)

Culmen Length and depth for sex and species

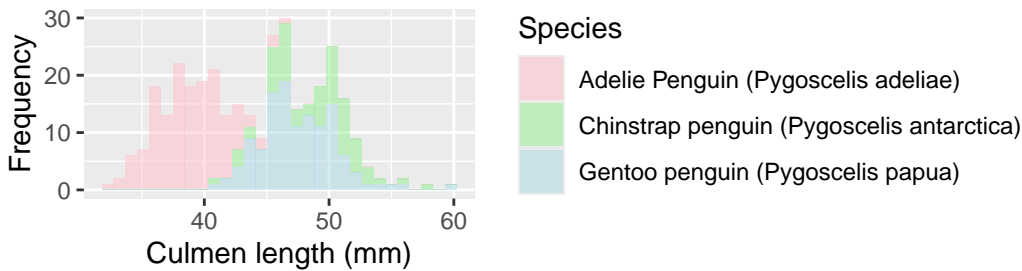
```
ggplot(data = penguins_data, aes(x = `Culmen Length (mm)`, y = `Culmen Depth (mm)`) +
  geom_point(aes(color = Sex)) +
  scale_color_manual(values = c("pink", "lightblue")) +
  labs(x = "Culmen Length (mm)", y = "Culmen Depth (mm)") +
  theme(legend.position = 'bottom') +
  facet_wrap(~Species)
```



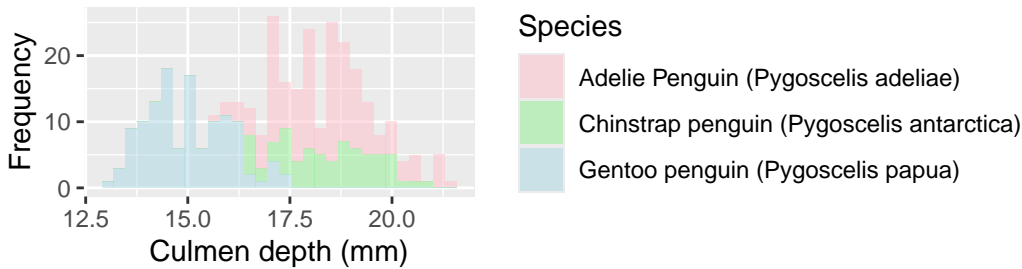
Bar plot

```
library(gridExtra)
p1 <- ggplot(data = penguins_data, aes(x = `Culmen Length (mm)`) +
  geom_histogram(aes(fill = Species), alpha = 0.5) +
  scale_fill_manual(values = c("pink", "lightgreen", "lightblue")) +
  labs(x = "Culmen length (mm)", y = "Frequency", title = "Penguin culmen lengths")
p2 <- ggplot(data = penguins_data, aes(x = `Culmen Depth (mm)`) +
  geom_histogram(aes(fill = Species), alpha = 0.5) +
  scale_fill_manual(values = c("pink", "lightgreen", "lightblue")) +
  labs(x = "Culmen depth (mm)", y = "Frequency", title = "Penguin culmen depths")
grid.arrange(p1, p2, nrow=2, ncol=1)
```


Penguin culmen lengths

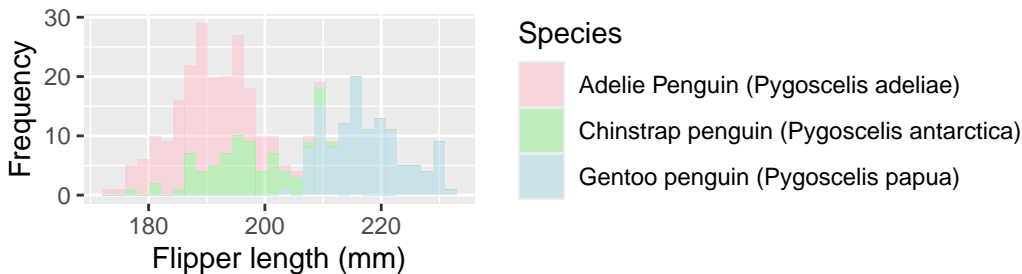


Penguin culmen depths

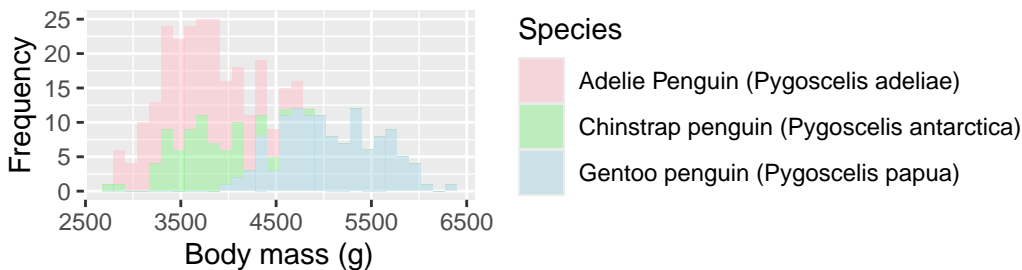


```
p3 <- ggplot(data = penguins_data, aes(x = `Flipper Length (mm)`) +
  geom_histogram(aes(fill = Species), alpha = 0.5) +
  scale_fill_manual(values = c("pink", "lightgreen", "lightblue")) +
  labs(x = "Flipper length (mm)", y = "Frequency", title = "Penguin flipper lengths")
p4 <- ggplot(data = penguins_data, aes(x = `Body Mass (g)`) +
  geom_histogram(aes(fill = Species), alpha = 0.5) +
  scale_fill_manual(values = c("pink", "lightgreen", "lightblue")) +
  labs(x = "Body mass (g)", y = "Frequency", title = "Penguin body mass")
grid.arrange(p3, p4, nrow=2, ncol=1)
```

Penguin flipper lengths

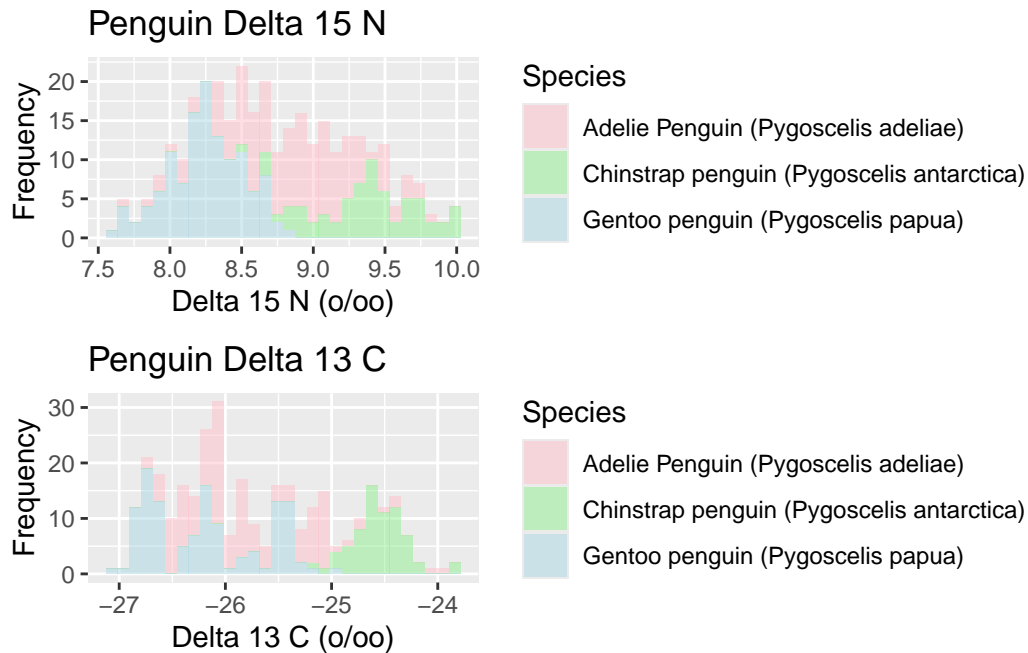


Penguin body mass



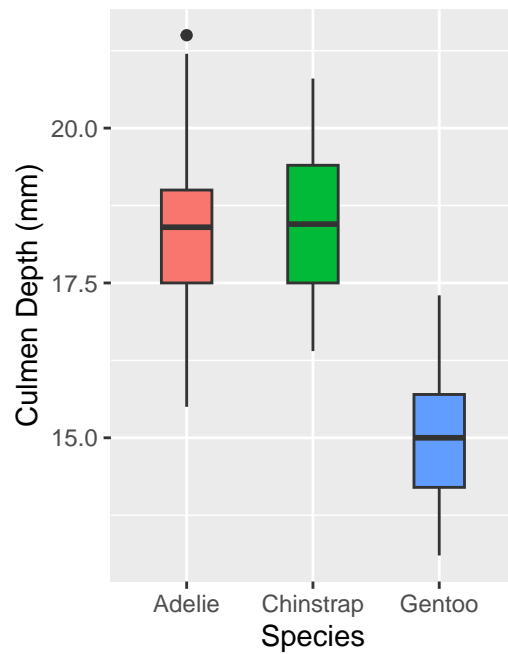
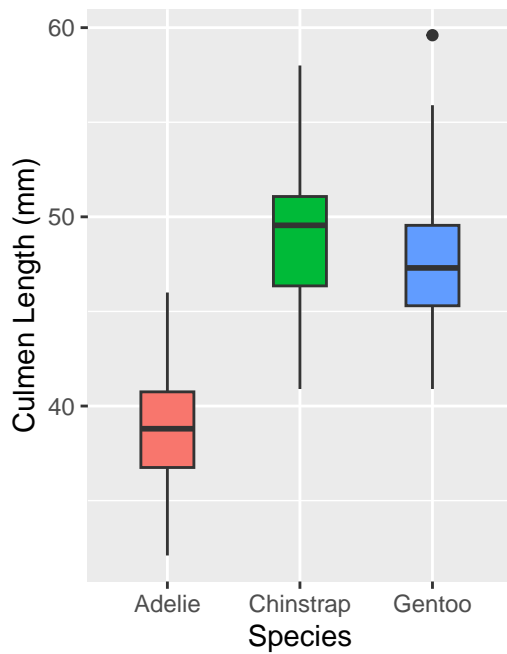
```
p5 <- ggplot(data = penguins_data, aes(x = `Delta 15 N (o/oo)`) +
  geom_histogram(aes(fill = Species), alpha = 0.5) +
  scale_fill_manual(values = c("pink", "lightgreen", "lightblue")) +
  labs(x = "Delta 15 N (o/oo)", y = "Frequency", title = "Penguin Delta 15 N")
p6 <- ggplot(data = penguins_data, aes(x = `Delta 13 C (o/oo)`) +
  geom_histogram(aes(fill = Species), alpha = 0.5) +
  scale_fill_manual(values = c("pink", "lightgreen", "lightblue")) +
```

```
labs(x = "Delta 13 C (o/oo)", y = "Frequency", title = "Penguin Delta 13 C")
grid.arrange(p5,p6,nrow=2,ncol=1)
```



Boxplot

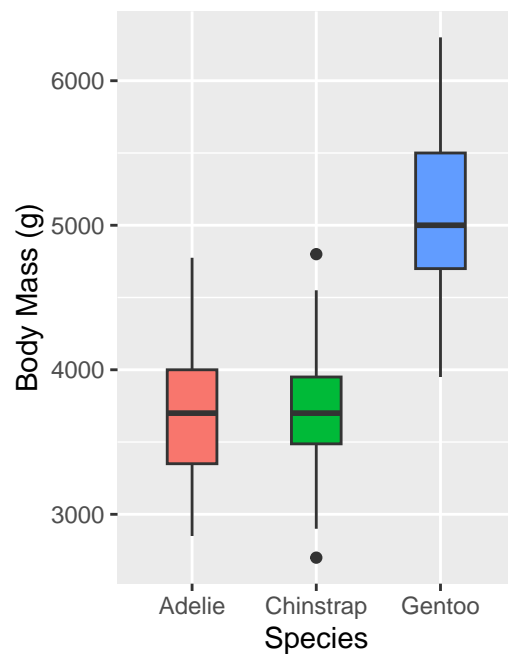
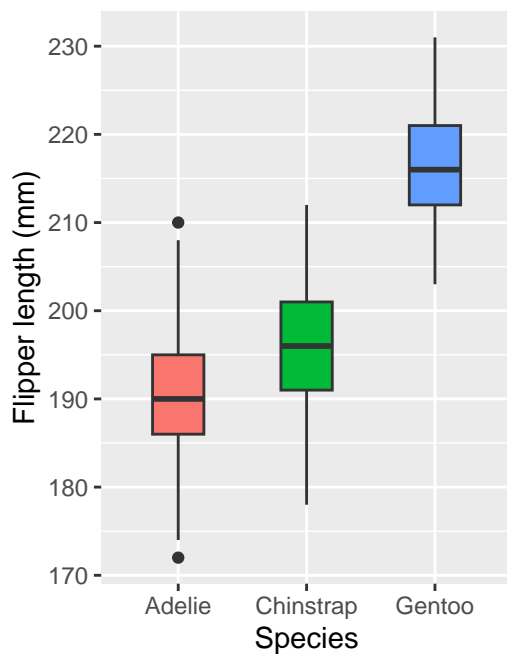
```
p1 <- ggplot(data = penguins_data, aes(x = Species, y = `Culmen Length (mm)`)) +
  geom_boxplot(aes(fill = Species), width = 0.4, show.legend = FALSE) +
  scale_color_manual(values = c("pink","lightgreen","lightblue")) +
  labs(x = "Species",y = "Culmen Length (mm)") +
  scale_x_discrete(labels = c("Adelie Penguin (Pygoscelis adeliae)" = "Adelie",
                              "Chinstrap penguin (Pygoscelis antarctica)" = "Chinstrap",
                              "Gentoo penguin (Pygoscelis papua)" = "Gentoo"))
p2 <- ggplot(data = penguins_data, aes(x = Species, y = `Culmen Depth (mm)`)) +
  geom_boxplot(aes(fill = Species), width = 0.4, show.legend = FALSE) +
  scale_color_manual(values = c("pink","lightgreen","lightblue")) +
  labs(x = "Species",y = "Culmen Depth (mm)") +
  scale_x_discrete(labels = c("Adelie Penguin (Pygoscelis adeliae)" = "Adelie",
                              "Chinstrap penguin (Pygoscelis antarctica)" = "Chinstrap",
                              "Gentoo penguin (Pygoscelis papua)" = "Gentoo"))
grid.arrange(p1,p2,nrow=1,ncol=2)
```



```
p3 <- ggplot(data = penguins_data, aes(x = Species, y = `Flipper Length (mm)`) +
  geom_boxplot(aes(fill = Species), width = 0.4, show.legend = FALSE) +
  scale_color_manual(values = c("pink","lightgreen","lightblue")) +
  labs(x = "Species",y = "Flipper length (mm)") +
  scale_x_discrete(labels = c("Adelie Penguin (Pygoscelis adeliae)" = "Adelie",
    "Chinstrap penguin (Pygoscelis antarctica)" = "Chinstrap",
    "Gentoo penguin (Pygoscelis papua)" = "Gentoo"))

p4 <- ggplot(data = penguins_data, aes(x = Species, y = `Body Mass (g)`) +
  geom_boxplot(aes(fill = Species), width = 0.4, show.legend = FALSE) +
  scale_color_manual(values = c("pink","lightgreen","lightblue")) +
  labs(x = "Species",y = "Body Mass (g)") +
  scale_x_discrete(labels = c("Adelie Penguin (Pygoscelis adeliae)" = "Adelie",
    "Chinstrap penguin (Pygoscelis antarctica)" = "Chinstrap",
    "Gentoo penguin (Pygoscelis papua)" = "Gentoo")))

grid.arrange(p3,p4,nrow=1,ncol=2)
```



```
p5 <- ggplot(data = penguins_data, aes(x = Species, y = `Delta 15 N (o/oo)`)) +
  geom_boxplot(aes(fill = Species), width = 0.4, show.legend = FALSE) +
  scale_color_manual(values = c("pink","lightgreen","lightblue")) +
  labs(x = "Species",y = "Delta 15 N (o/oo)") +
  scale_x_discrete(labels = c("Adelie Penguin (Pygoscelis adeliae)" = "Adelie",
                              "Chinstrap penguin (Pygoscelis antarctica)" = "Chinstrap",
                              "Gentoo penguin (Pygoscelis papua)" = "Gentoo"))

p6 <- ggplot(data = penguins_data, aes(x = Species, y = `Delta 13 C (o/oo)`)) +
  geom_boxplot(aes(fill = Species), width = 0.4, show.legend = FALSE) +
  scale_color_manual(values = c("pink","lightgreen","lightblue")) +
  labs(x = "Species", y = "Delta 13 C (o/oo)") +
  scale_x_discrete(labels = c("Adelie Penguin (Pygoscelis adeliae)" = "Adelie",
                              "Chinstrap penguin (Pygoscelis antarctica)" = "Chinstrap",
                              "Gentoo penguin (Pygoscelis papua)" = "Gentoo"))

grid.arrange(p5,p6,nrow=1,ncol=2)
```

