

HW1

Fa-Hsiang Yin

2024-09-17

Table of contents

Summary Staistic	1
Missing Values	4
Male v.s.Female: Body Measurement	6
Table1	6
Graph	7
Species v.s. Island	8

```
library(palmerpenguins)
library(dplyr)
library(ggplot2)
library(Hmisc)
theme_set(theme_minimal())
```

Summary Staistic

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

penguins_raw
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

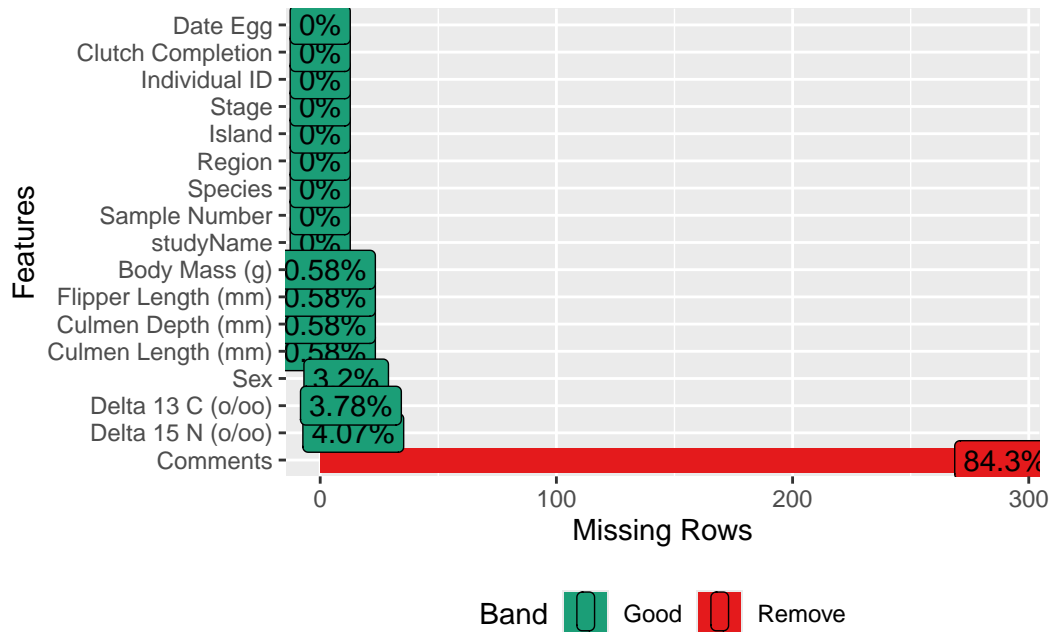
n	missing	distinct
54	290	10

lowest : Adult not sampled.
highest: No blood sample obtained.

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

Missing Values

```
library(Hmisc)
library(DataExplorer)
plot_missing(penguins_raw)
```



Note that the variable "Comment" is additional information for certain situations causing the missing values, we can just ignore it.

We also notice that missing values can be roughly categorized into three situations as follows:

1. Missing body measurements : includes culmen length, culmen depth, flipper length and body mass.
2. Missing blood isotope measurement : missing delta 15 N or both delta 15 N and delta 13 C.
3. Missing gender info.

When situation 1 occurs, situation 2 and 3 will also occur. (Comment: Adult not sampled.)

```
mypenguins <- penguins_raw[,-17]
```

Male v.s.Female: Body Measurement

Table1

```
library(tidyverse)
library(table1)
pen_table<- mypenguins%>%
  drop_na(Sex)%>%
  `names<-`(replace(names(.),1:16,gsub(" ", "_",names(.),
    fixed = TRUE)))%>%
  `names<-`(replace(names(.),c(10,11,12,13,15,16),
    c('Culmen_Length','Culmen_Depth',
      'Flipper_Length','Body_Mass',
      'Delta_15_N','Delta_13_C'))))
table1(~ Culmen_Length+Culmen_Depth+Flipper_Length+Body_Mass+Delta_15_N+Delta_13_C|Sex,pen
```

	FEMALE	MALE	Overall
	(N=165)	(N=168)	(N=333)
Culmen_Length			
Mean (SD)	42.1 (4.90)	45.9 (5.37)	44.0 (5.47)
Median [Min, Max]	42.8 [32.1, 58.0]	46.8 [34.6, 59.6]	44.5 [32.1, 59.6]
Culmen_Depth			
Mean (SD)	16.4 (1.80)	17.9 (1.86)	17.2 (1.97)
Median [Min, Max]	17.0 [13.1, 20.7]	18.5 [14.1, 21.5]	17.3 [13.1, 21.5]
Flipper_Length			
Mean (SD)	197 (12.5)	205 (14.5)	201 (14.0)
Median [Min, Max]	193 [172, 222]	201 [178, 231]	197 [172, 231]
Body_Mass			
Mean (SD)	3860 (666)	4550 (788)	4210 (805)
Median [Min, Max]	3650 [2700, 5200]	4300 [3250, 6300]	4050 [2700, 6300]
Delta_15_N			
Mean (SD)	8.68 (0.552)	8.81 (0.546)	8.74 (0.552)
Median [Min, Max]	8.62 [7.63, 9.81]	8.71 [7.77, 10.0]	8.66 [7.63, 10.0]
Missing	2 (1.2%)	7 (4.2%)	9 (2.7%)
Delta_13_C			

	FEMALE	MALE	Overall
Mean (SD)	-25.7 (0.798)	-25.7 (0.798)	-25.7 (0.797)
Median [Min, Max]	-25.9 [-27.0, -23.9]	-25.8 [-26.9, -23.8]	-25.8 [-27.0, -23.8]
Missing	2 (1.2%)	6 (3.6%)	8 (2.4%)

Graph

```

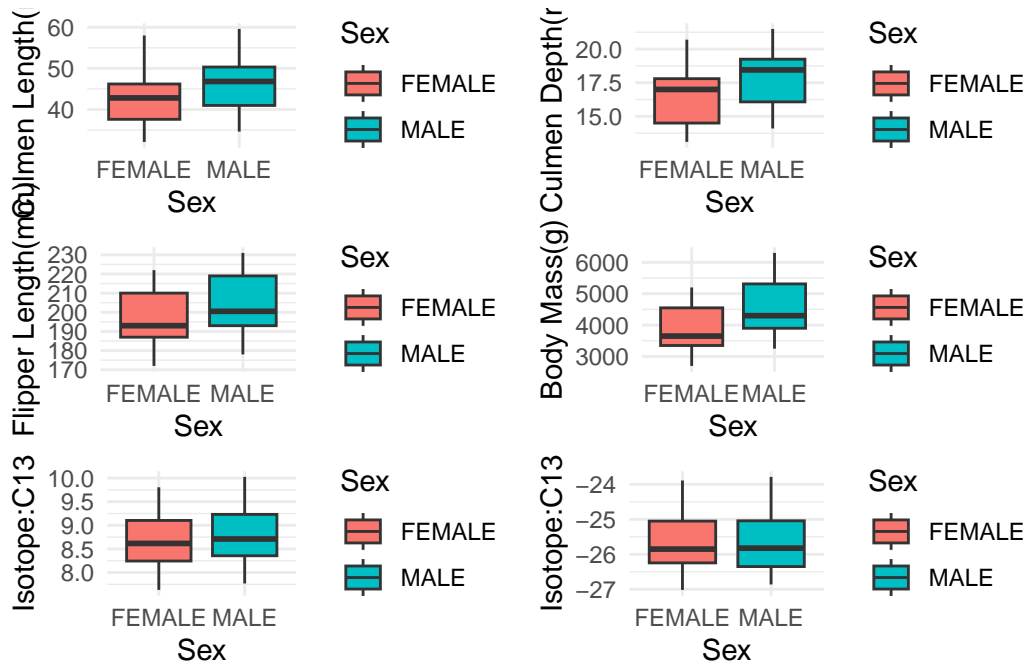
library(ggplot2)
library(gridExtra)
library(gapminder)
library(dplyr)
p1 <- ggplot(pen_table, aes(x = Sex,
                           y = Culmen_Length, fill = Sex)) +
  geom_boxplot() +
  scale_y_continuous(name = "Culmen Length(mm)") +
  scale_x_discrete(name = "Sex")
p2 <- ggplot(pen_table, aes(x = Sex,
                           y = Culmen_Depth, fill = Sex)) +
  geom_boxplot() +
  scale_y_continuous(name = "Culmen Depth(mm)") +
  scale_x_discrete(name = "Sex")
p3 <- ggplot(pen_table, aes(x = Sex,
                           y = Flipper_Length, fill = Sex)) +
  geom_boxplot() +
  scale_y_continuous(name = "Flipper Length(mm)") +
  scale_x_discrete(name = "Sex")
p4 <- ggplot(pen_table, aes(x = Sex,
                           y = Body_Mass, fill = Sex)) +
  geom_boxplot() +
  scale_y_continuous(name = "Body Mass(g)") +
  scale_x_discrete(name = "Sex")
p5 <- ggplot(pen_table, aes(x = Sex,
                           y = Delta_15_N, fill = Sex)) +
  geom_boxplot() +
  scale_y_continuous(name = "Isotope:C13") +
  scale_x_discrete(name = "Sex")
p6 <- ggplot(pen_table, aes(x = Sex,
                           y = Delta_13_C, fill = Sex)) +

```

```

geom_boxplot() +
  scale_y_continuous(name = "Isotope:C13") +
  scale_x_discrete(name = "Sex")
grid.arrange(p1,p2,p3,p4,p5,p6,
             ncol = 2, nrow = 3)

```



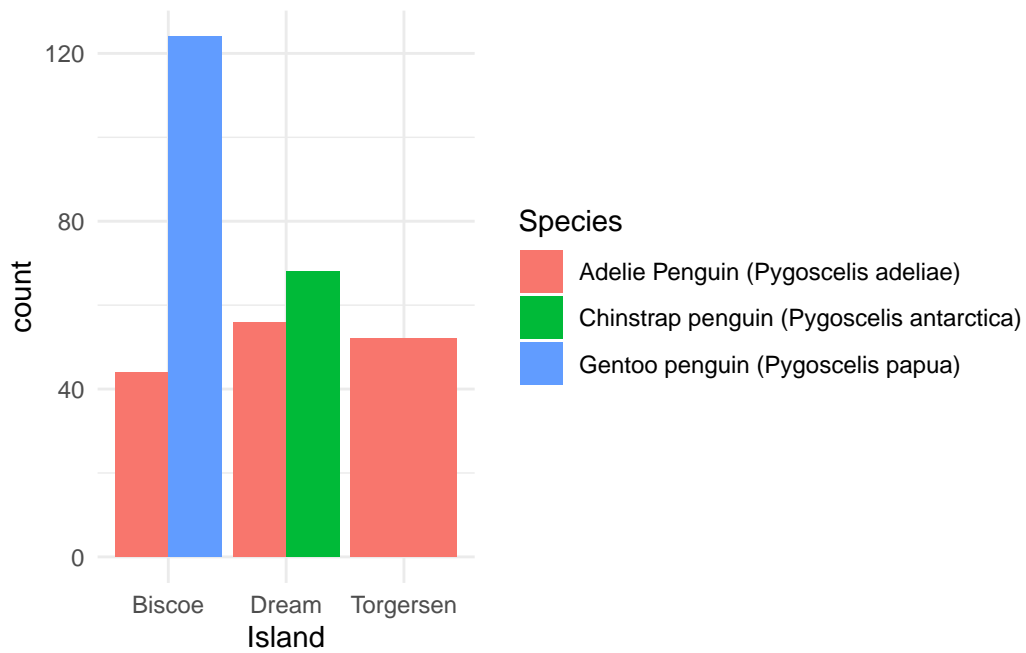
We can see that the body measurements of male penguins appear to be larger than those of female penguins.

Species v.s. Island

```

p <- ggplot(data = mypenguins) +
  geom_bar(mapping = aes(x = Island, fill = Species), position = "dodge")
p

```

Of the three species, only the Adelie Penguin (*Pygoscelis adeliae*) has been observed on all three islands.