

HW1

Kai-Chien Chen

2024-09-16

Table of contents

Import Data	1
Summary Statistics	2
Missing value	5
Some Plots	6

Import Data

```
# install.packages("palmerpenguins")  
library(palmerpenguins)
```

Warning: package 'palmerpenguins' was built under R version 4.2.3

```
data <- penguins_raw  
head(data)
```

```
# A tibble: 6 x 17  
  studyName `Sample Number` Species      Region Island Stage `Individual ID`  
  <chr>          <dbl> <chr>          <chr>  <chr>  <chr> <chr>  
1 PAL0708          1 Adelie Penguin ~ Anvers Torge~ Adul~ N1A1  
2 PAL0708          2 Adelie Penguin ~ Anvers Torge~ Adul~ N1A2  
3 PAL0708          3 Adelie Penguin ~ Anvers Torge~ Adul~ N2A1
```

```

4 PAL0708          4 Adelie Penguin ~ Anvers Torge~ Adul~ N2A2
5 PAL0708          5 Adelie Penguin ~ Anvers Torge~ Adul~ N3A1
6 PAL0708          6 Adelie Penguin ~ Anvers Torge~ Adul~ N3A2
# i 10 more variables: `Clutch Completion` <chr>, `Date Egg` <date>,
#   `Culmen Length (mm)` <dbl>, `Culmen Depth (mm)` <dbl>,
#   `Flipper Length (mm)` <dbl>, `Body Mass (g)` <dbl>, Sex <chr>,
#   `Delta 15 N (o/oo)` <dbl>, `Delta 13 C (o/oo)` <dbl>, Comments <chr>

```

Summary Statistics

```
library(Hmisc)
```

Warning: package 'Hmisc' was built under R version 4.2.2

Loading required package: lattice

Loading required package: survival

Loading required package: Formula

Loading required package: ggplot2

Warning: package 'ggplot2' was built under R version 4.2.2

Attaching package: 'Hmisc'

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

```
format.pval, units
```

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

17 Variables data
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 34.0 34.1, highest: 55.1 55.8 55.9 58.0 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 : 7.63220 7.63452 7.63884 7.68528 7.68870, highest: 9.93727 9.98044 10.02019 10.02372 10.02544

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 : -27.01854 -26.95470 -26.89644 -26.86485 -26.86352
highest: -24.16566 -24.10255 -23.90309 -23.89017 -23.78767

Comments

n	missing	distinct
54	290	10

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

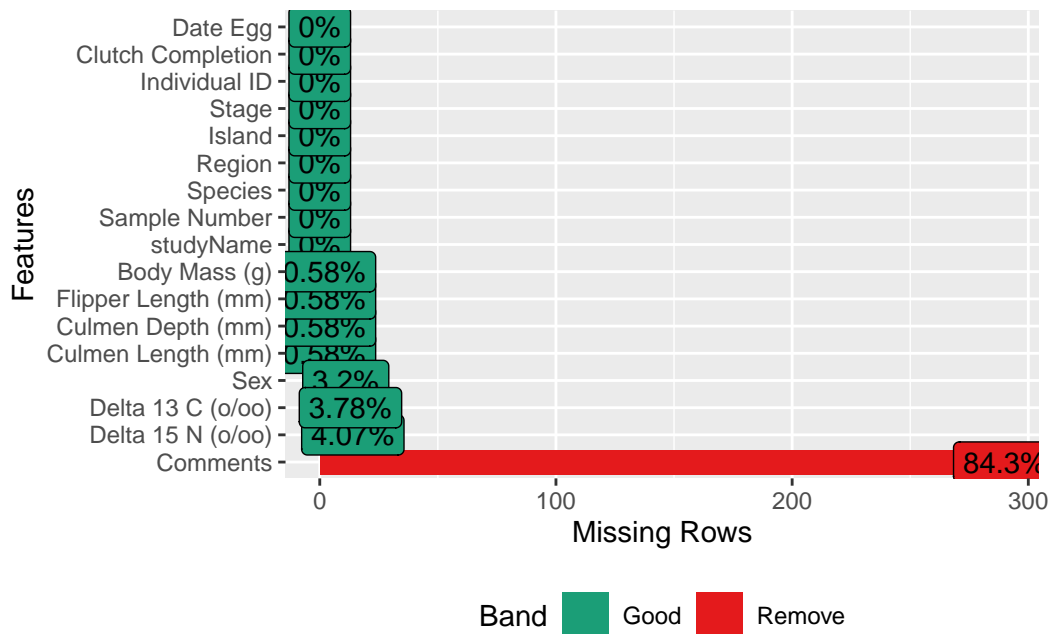
Adult not sampled. Nest never observed with full blood sample.
No delta15N data received from lab.

Missing value

```
library(Hmisc)
library(DataExplorer)
```

Warning: package 'DataExplorer' was built under R version 4.2.3

```
plot_missing(data)
```

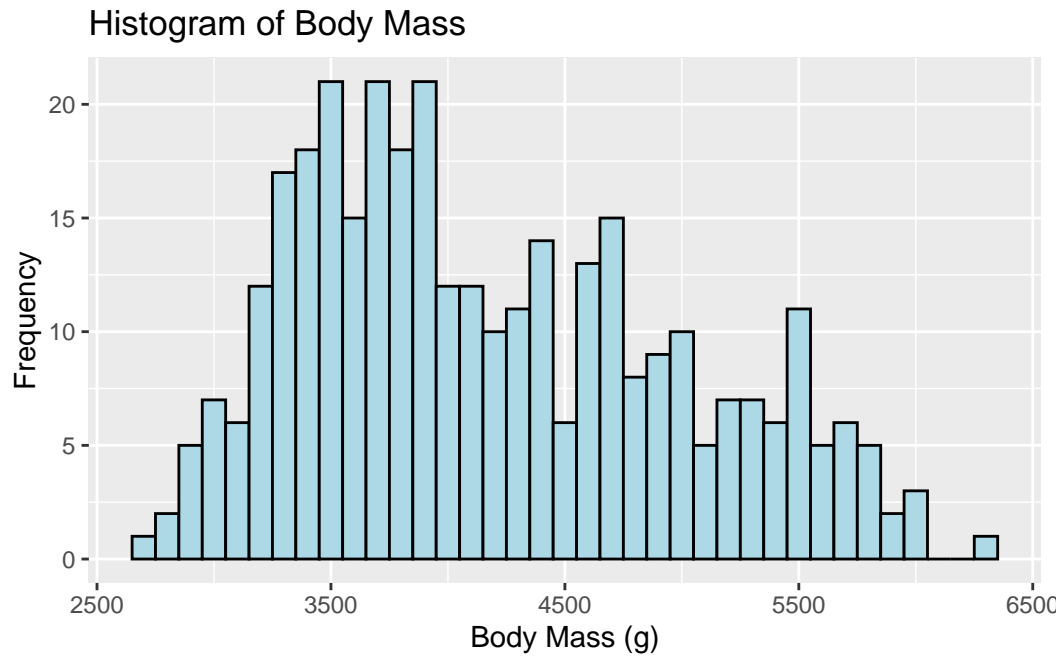


Some Plots

體重之直方圖

```
library(ggplot2)
ggplot(data, aes(x = `Body Mass (g)`) +
  geom_histogram(binwidth = 100, fill = "lightblue", color = "black") +
  labs(title = "Histogram of Body Mass", x = "Body Mass (g)", y = "Frequency")
```

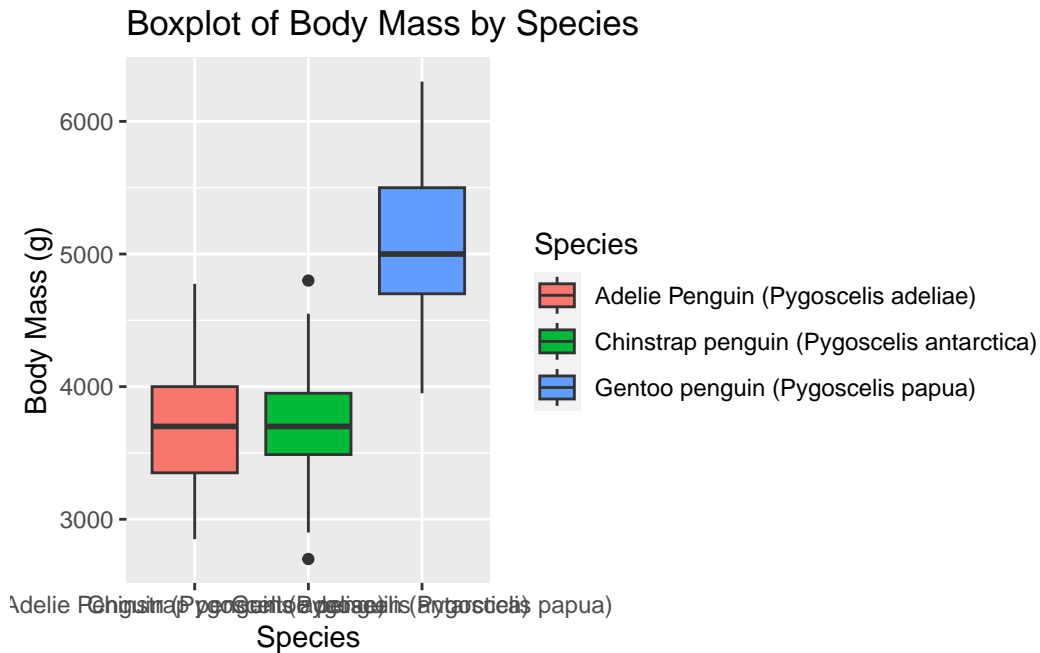
Warning: Removed 2 rows containing non-finite values (`stat_bin()`).



依物種做箱型圖

```
ggplot(data, aes(x = Species, y = `Body Mass (g)`, fill = Species)) +  
  geom_boxplot() +  
  labs(title = "Boxplot of Body Mass by Species", y = "Body Mass (g)")
```

Warning: Removed 2 rows containing non-finite values (`stat_boxplot()`).



相關性分析

```
library(dplyr)
```

Warning: package 'dplyr' was built under R version 4.2.3

Attaching package: 'dplyr'

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

src, summarize

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

filter, lag

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

intersect, setdiff, setequal, union


```
#
numeric_vars <- c("Culmen Length (mm)", "Culmen Depth (mm)", "Flipper Length (mm)", "Body Mass (g)", "Delta 15 N (o/oo)", "Delta 13 C (o/oo)")

correlation_matrix <- cor(data %>% select(all_of(numeric_vars)), use = "complete.obs")

#
library(corrplot)
```

corrplot 0.94 loaded

```
corrplot(correlation_matrix, method = "circle")
```

