My Report

Ting-Hsuan CHU

2024-09-18

Table of contents

Summary Staistic

```
library(extrafont)
```

```
Registering fonts with R
```

```
library(Hmisc)
```

```
Warning: package 'Hmisc' was built under R version 4.3.3
```

```
Attaching package: 'Hmisc'

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

    format.pval, units
```

```r
library(palmerpenguins)
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

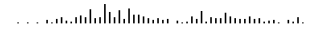
lowest : 32.1 33.1 33.5 34 34.1, highest: 55.1 55.8 55.9 58 59.6

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

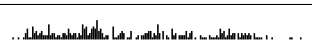
Value	FEMALE	MALE
Frequency	165	168
Proportion	0.495	0.505

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
```

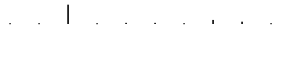
Δ 13 C (o/oo):



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.                                  Adult not sampled. Nest never observed with ful
highest: No blood sample obtained.                           No delta15N data received from lab.
```

Table 1

```
library(table1)
summary(penguins_raw)
```

```
 studyName         Sample Number      Species             Region
Length:344         Min.   :  1.00   Length:344         Length:344
Class :character   1st Qu.: 29.00   Class :character   Class :character
Mode  :character   Median : 58.00   Mode  :character   Mode  :character
                   Mean   : 63.15
                   3rd Qu.: 95.25
                   Max.   :152.00

   Island             Stage           Individual ID      Clutch Completion
Length:344         Length:344         Length:344         Length:344
Class :character   Class :character   Class :character   Class :character
Mode  :character   Mode  :character   Mode  :character   Mode  :character



   Date Egg          Culmen Length (mm) Culmen Depth (mm) Flipper Length (mm)
Min.   :2007-11-09   Min.   :32.10      Min.   :13.10     Min.   :172.0
1st Qu.:2007-11-28   1st Qu.:39.23      1st Qu.:15.60     1st Qu.:190.0
Median :2008-11-09   Median :44.45      Median :17.30     Median :197.0
Mean   :2008-11-27   Mean   :43.92      Mean   :17.15     Mean   :200.9
3rd Qu.:2009-11-16   3rd Qu.:48.50      3rd Qu.:18.70     3rd Qu.:213.0
Max.   :2009-12-01   Max.   :59.60      Max.   :21.50     Max.   :231.0
                     NA's   :2          NA's   :2         NA's   :2
Body Mass (g)      Sex            Delta 15 N (o/oo) Delta 13 C (o/oo)
Min.   :2700   Length:344         Min.   : 7.632    Min.   :-27.02
1st Qu.:3550   Class :character   1st Qu.: 8.300    1st Qu.:-26.32
Median :4050   Mode  :character   Median : 8.652    Median :-25.83
Mean   :4202                      Mean   : 8.733    Mean   :-25.69
3rd Qu.:4750                      3rd Qu.: 9.172    3rd Qu.:-25.06
Max.   :6300                      Max.   :10.025    Max.   :-23.79
NA's   :2                         NA's   :14        NA's   :13
  Comments
Length:344
Class :character
Mode  :character
```

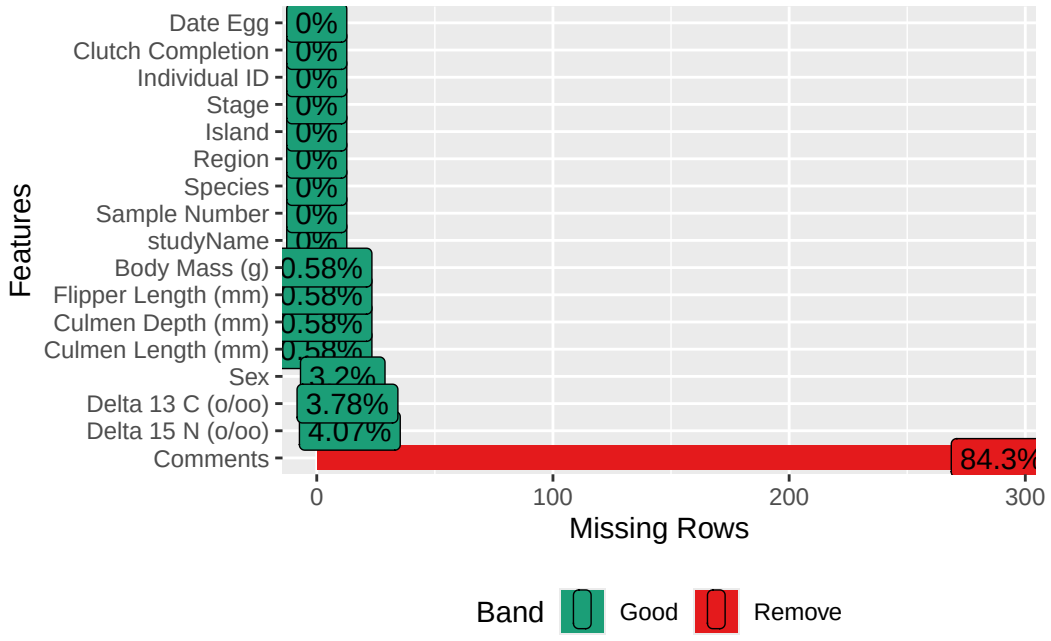
```
# help(table1)
```

Missing Values

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

```
Warning: package 'DataExplorer' was built under R version 4.4.0
```

```
plot_missing(penguins_raw)
```



After deleting Missing Value

```
clean_Penguins <- na.omit(penguins_raw)
summary(clean_Penguins)
```

```
 studyName         Sample Number      Species             Region
Length:34          Min.   :  1.00   Length:34          Length:34
Class :character   1st Qu.: 23.25   Class :character   Class :character
Mode  :character   Median : 39.00   Mode  :character   Mode  :character
                   Mean   : 51.71
                   3rd Qu.: 68.25
                   Max.   :140.00
   Island             Stage           Individual ID      Clutch Completion
Length:34          Length:34          Length:34          Length:34
Class :character   Class :character   Class :character   Class :character
Mode  :character   Mode  :character   Mode  :character   Mode  :character


   Date Egg          Culmen Length (mm) Culmen Depth (mm) Flipper Length (mm)
Min.   :2007-11-10   Min.   :35.90      Min.   :13.70     Min.   :172.0
1st Qu.:2007-11-27   1st Qu.:39.33      1st Qu.:16.60     1st Qu.:190.0
Median :2008-11-06   Median :44.50      Median :17.90     Median :195.5
Mean   :2008-09-15   Mean   :44.71      Mean   :17.65     Mean   :196.7
3rd Qu.:2009-08-19   3rd Qu.:49.60      3rd Qu.:19.15     3rd Qu.:201.5
Max.   :2009-12-01   Max.   :58.00      Max.   :20.00     Max.   :225.0
Body Mass (g)      Sex            Delta 15 N (o/oo) Delta 13 C (o/oo)
Min.   :2700   Length:34          Min.   : 7.992    Min.   :-26.84
1st Qu.:3344   Class :character   1st Qu.: 8.627    1st Qu.:-26.08
Median :3738   Mode  :character   Median : 9.041    Median :-25.16
Mean   :3877                      Mean   : 9.019    Mean   :-25.33
3rd Qu.:4238                      3rd Qu.: 9.374    3rd Qu.:-24.62
Max.   :5700                      Max.   :10.025    Max.   :-23.89
  Comments
Length:34
Class :character
Mode  :character
```

```r
clean_Penguins$Sex <- as.factor(clean_Penguins$Sex)
clean_Penguins$Stage <- as.factor(clean_Penguins$Stage)
str(clean_Penguins)
```

```
tibble [34 x 17] (S3: tbl_df/tbl/data.frame)
 $ studyName          : chr [1:34] "PAL0708" "PAL0708" "PAL0708" "PAL0708" ...
 $ Sample Number      : num [1:34] 7 8 29 30 39 69 70 121 122 131 ...
 $ Species            : chr [1:34] "Adelie Penguin (Pygoscelis adeliae)" "Adelie Penguin (Pyg
 $ Region             : chr [1:34] "Anvers" "Anvers" "Anvers" "Anvers" ...
 $ Island             : chr [1:34] "Torgersen" "Torgersen" "Biscoe" "Biscoe" ...
 $ Stage              : Factor w/ 1 level "Adult, 1 Egg Stage": 1 1 1 1 1 1 1 1 1 1 ...
 $ Individual ID      : chr [1:34] "N4A1" "N4A2" "N18A1" "N18A2" ...
 $ Clutch Completion  : chr [1:34] "No" "No" "No" "No" ...
 $ Date Egg           : Date[1:34], format: "2007-11-15" "2007-11-15" ...
 $ Culmen Length (mm) : num [1:34] 38.9 39.2 37.9 40.5 37.6 35.9 41.8 36.2 37.7 38.5 ...
 $ Culmen Depth (mm)  : num [1:34] 17.8 19.6 18.6 18.9 19.3 16.6 19.4 17.2 19.8 17.9 ...
 $ Flipper Length (mm): num [1:34] 181 195 172 180 181 190 198 187 198 190 ...
 $ Body Mass (g)      : num [1:34] 3625 4675 3150 3950 3300 ...
 $ Sex                : Factor w/ 2 levels "FEMALE","MALE": 1 2 1 2 1 1 2 1 2 1 ...
 $ Delta 15 N (o/oo)  : num [1:34] 9.19 9.46 8.38 8.9 9.41 ...
 $ Delta 13 C (o/oo)  : num [1:34] -25.2 -24.9 -25.2 -25.1 -25 ...
 $ Comments           : chr [1:34] "Nest never observed with full clutch." "Nest never obser
 - 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 ""
```

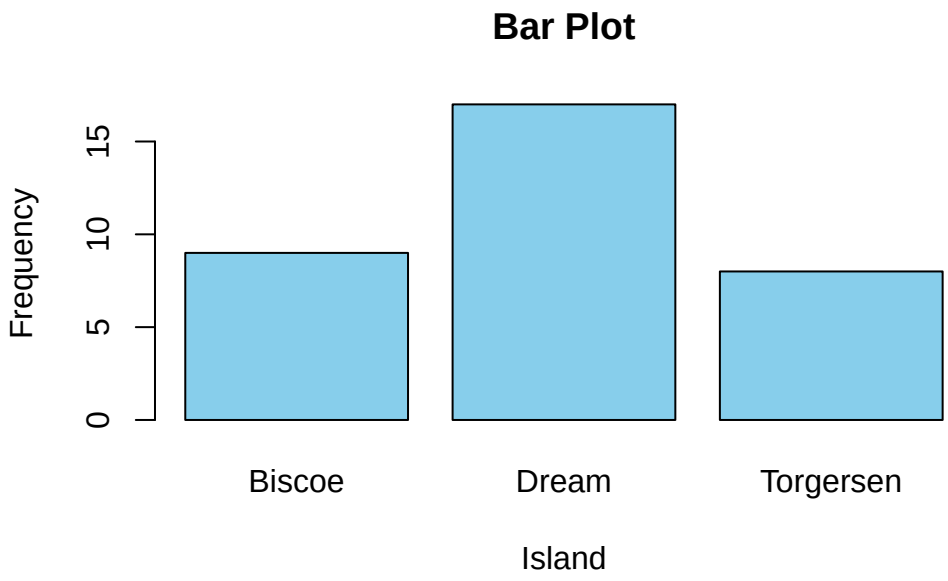
```
  .. .. ..- attr(*, "class")= chr [1:2] "collector_date" "collector"
  .. ..$ Culmen Length (mm) : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ Culmen Depth (mm)  : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ Flipper 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"
 - attr(*, "na.action")= 'omit' Named int [1:310] 1 2 3 4 5 6 9 10 11 12 ...
  ..- attr(*, "names")= chr [1:310] "1" "2" "3" "4" ...
```

```
table1(~ Species+`Culmen Length (mm)`+Island| Sex, data=clean_Penguins)
```

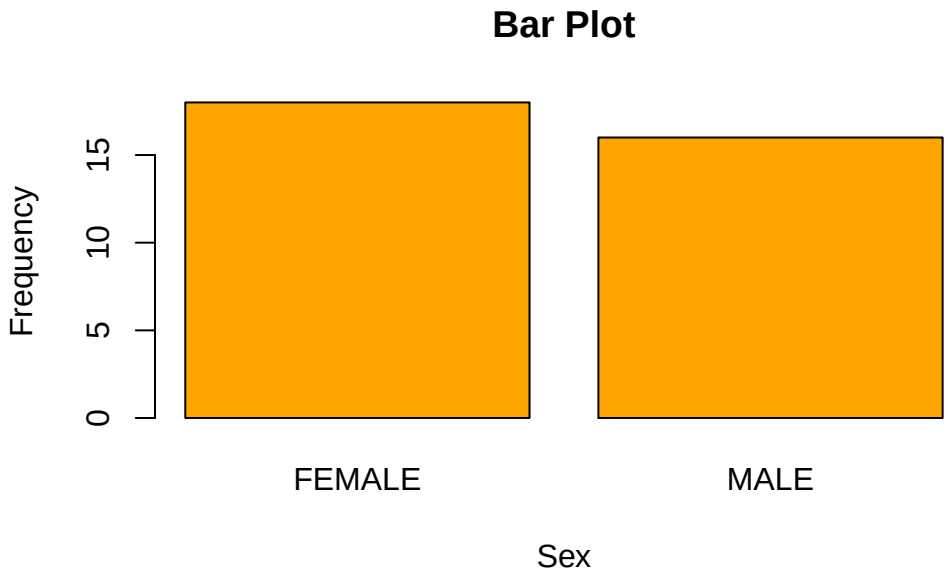
	FEMALE	MALE	Overall
	(N=18)	(N=16)	(N=34)
Species			
Adelie Penguin (Pygoscelis adeliae)	7 (38.9%)	6 (37.5%)	13 (38.2%)
Chinstrap penguin (Pygoscelis antarctica)	7 (38.9%)	7 (43.8%)	14 (41.2%)
Gentoo penguin (Pygoscelis papua)	4 (22.2%)	3 (18.8%)	7 (20.6%)
Culmen Length (mm)			
Mean (SD)	43.0 (5.82)	46.7 (5.29)	44.7 (5.80)
Median [Min, Max]	43.0 [35.9, 58.0]	49.6 [37.7, 53.5]	44.5 [35.9, 58.0]
Island			
Biscoe	5 (27.8%)	4 (25.0%)	9 (26.5%)
Dream	9 (50.0%)	8 (50.0%)	17 (50.0%)
Torgersen	4 (22.2%)	4 (25.0%)	8 (23.5%)

Descriptive statistics

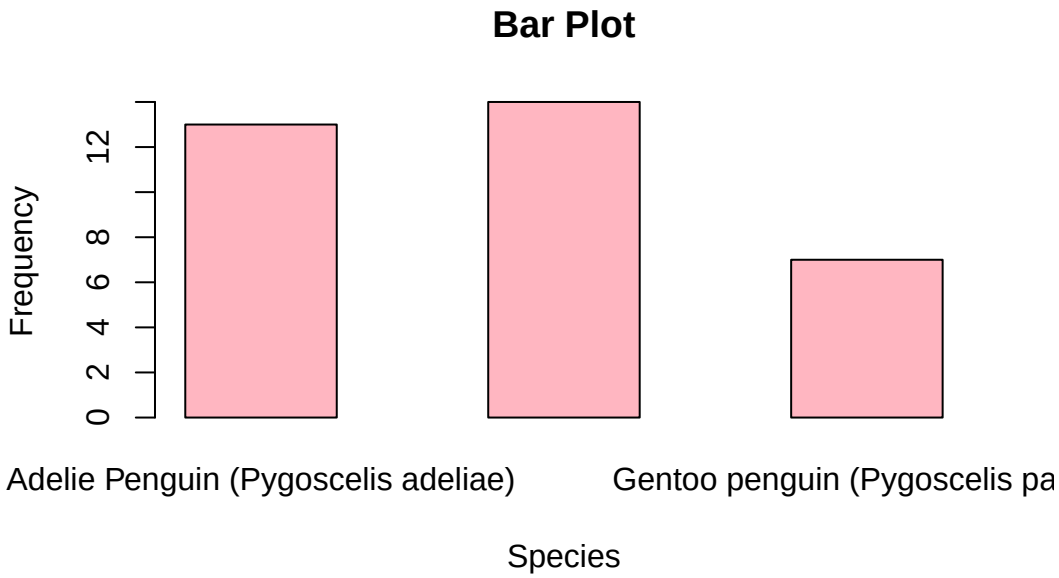
```
barplot(table(clean_Penguins$Island), main = "Bar Plot", col = "skyblue", xlab = "Island", yl
```



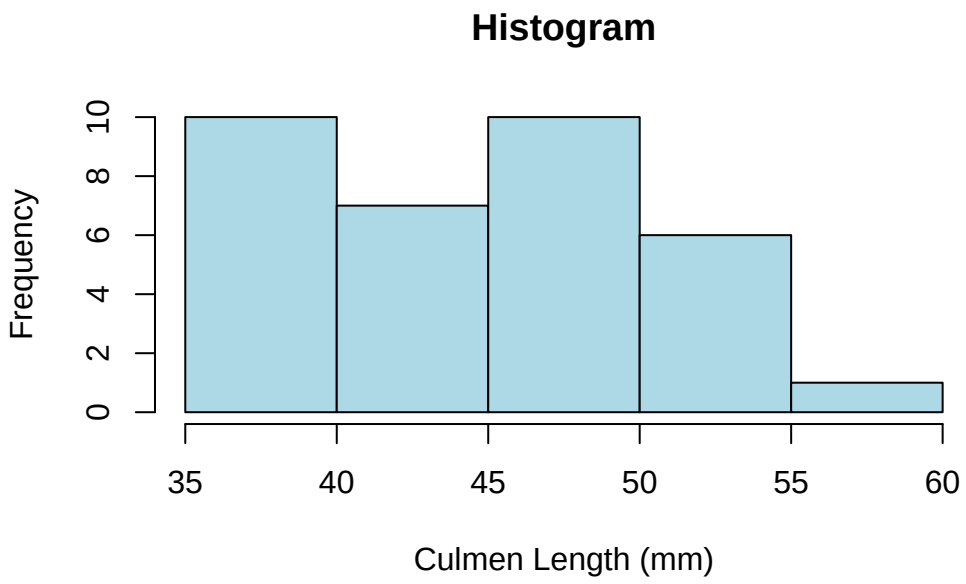

```
barplot(table(clean_Penguins$Sex), main = "Bar Plot", col = "orange", xlab = "Sex", ylab = "F
```



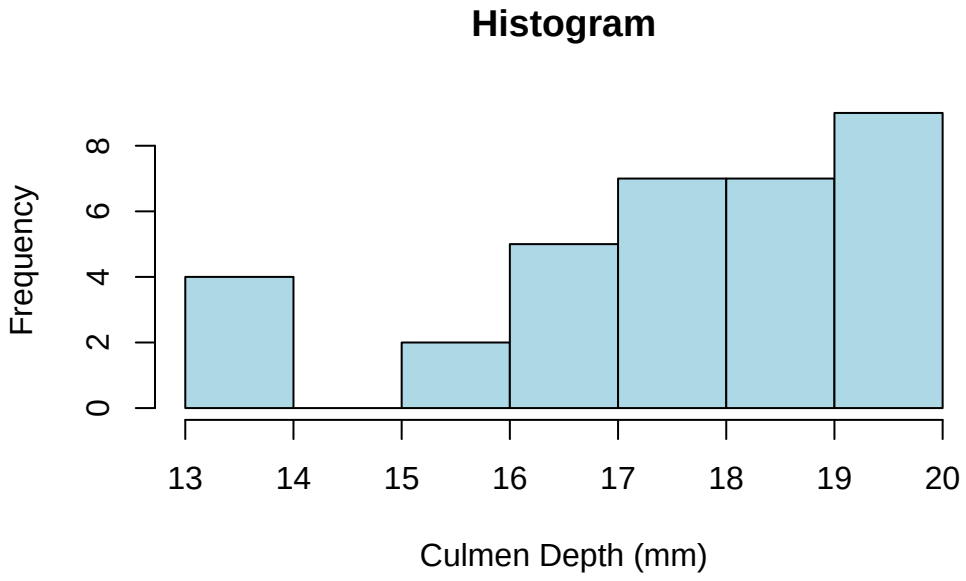

```r
barplot(table(clean_Penguins$Species),space =1, main = "Bar Plot", col = "lightpink", xlab =
```



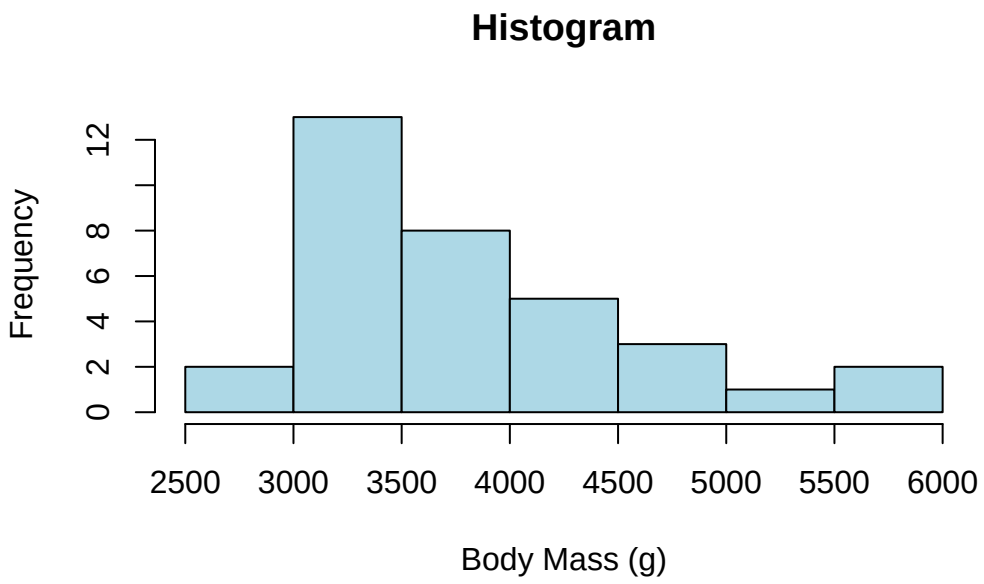
```
hist(clean_Penguins$`Culmen Length (mm)`, main = "Histogram", xlab = "Culmen Length (mm)", co
```



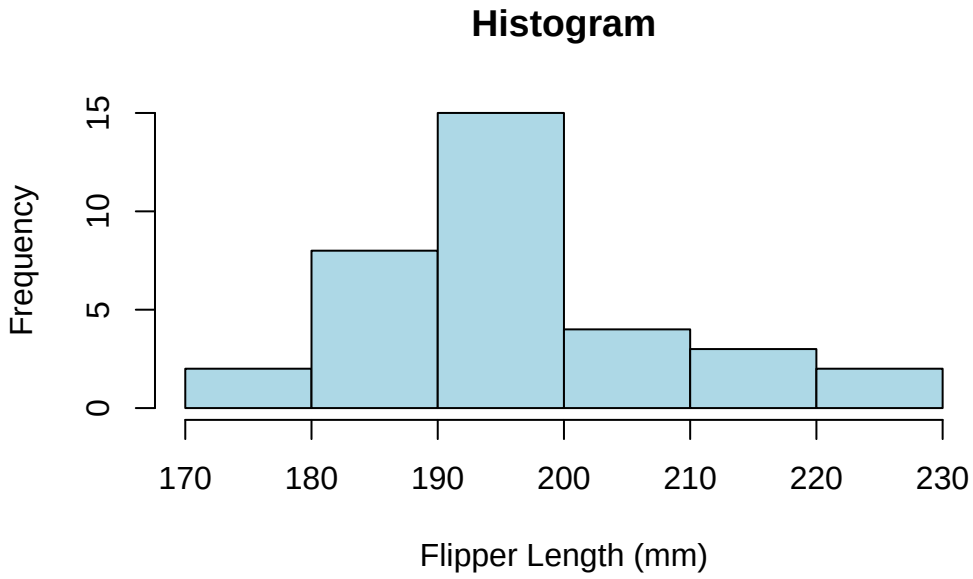

```
hist(clean_Penguins$`Culmen Depth (mm)`, main = "Histogram", xlab = "Culmen Depth (mm)", col
```



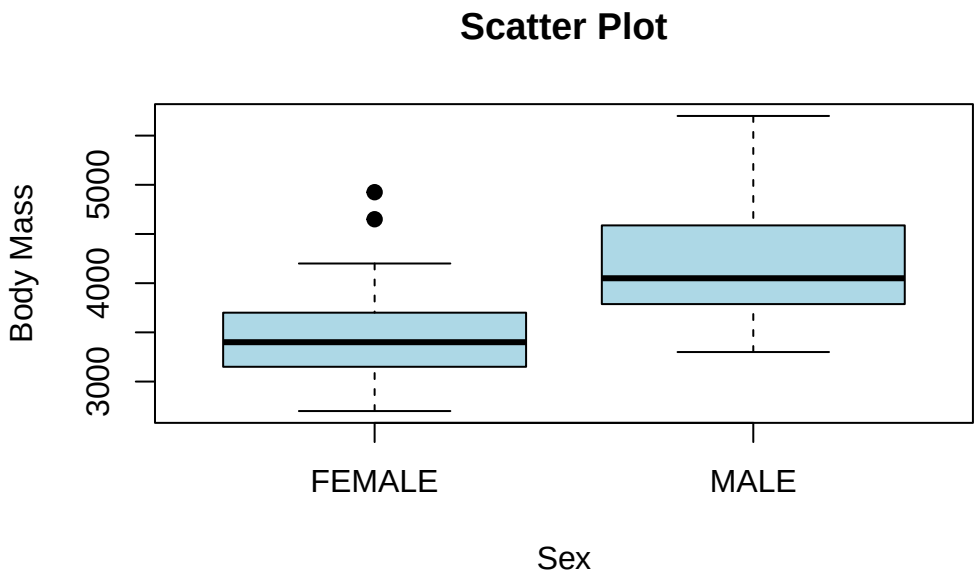
```r
hist(clean_Penguins$`Body Mass (g)`, main = "Histogram", xlab = "Body Mass (g)", col = "light
```




```r
hist(clean_Penguins$`Flipper Length (mm)`, main = "Histogram", xlab = "Flipper Length (mm)",
```

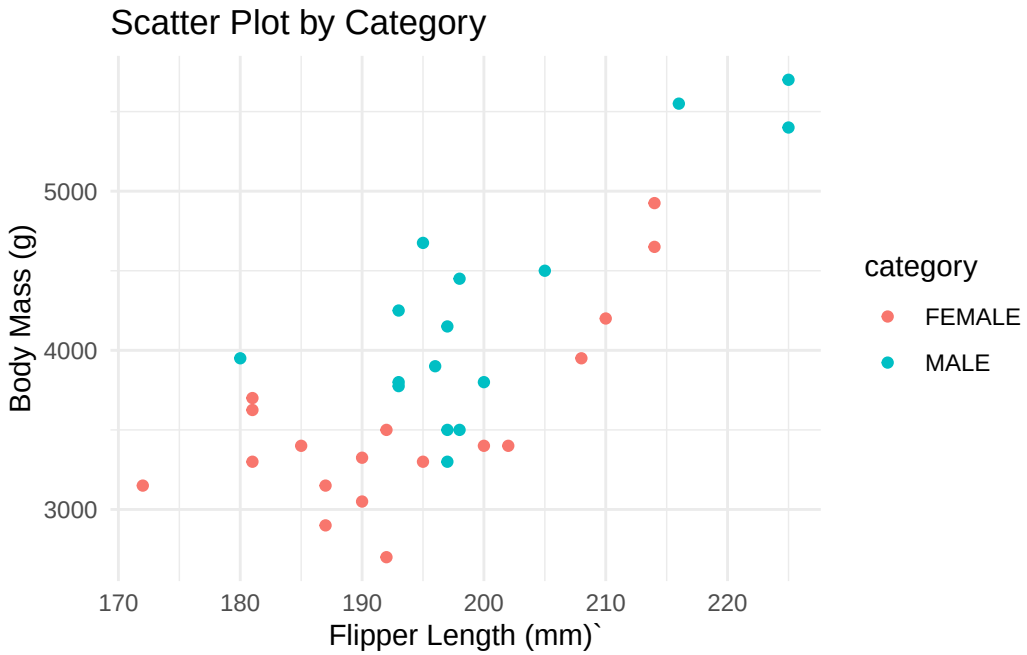


```r
plot(clean_Penguins$Sex, clean_Penguins$`Body Mass (g)`, main = "Scatter Plot", xlab = "Sex"
```

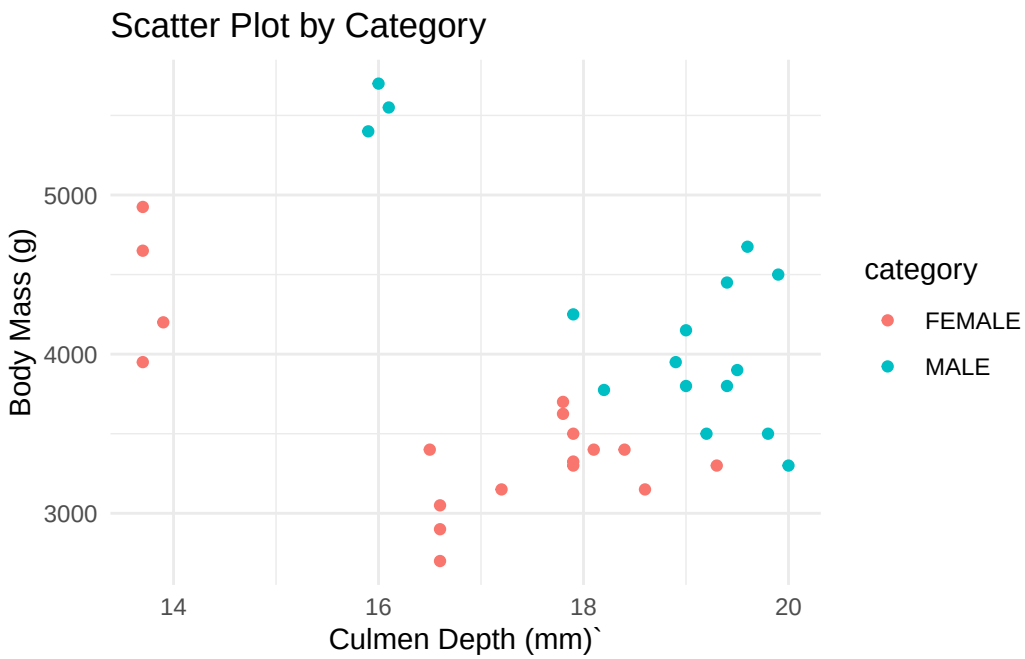



```r
#plot(clean_Penguins$`Culmen Length (mm)`, clean_Penguins$`Body Mass (g)`, main = "Scatter Pl
#plot(clean_Penguins$`Culmen Depth (mm)`, clean_Penguins$`Body Mass (g)`, main = "Scatter Plo
#plot(clean_Penguins$`Flipper Length (mm)`, clean_Penguins$`Body Mass (g)`, main = "Scatter P
library(ggplot2)
data1 <- data.frame(
  x = clean_Penguins$`Flipper Length (mm)`,
  y = clean_Penguins$`Body Mass (g)`,
  category = clean_Penguins$Sex
)
data2 <- data.frame(
  x = clean_Penguins$`Culmen Depth (mm)`,
  y = clean_Penguins$`Body Mass (g)`,
  category = clean_Penguins$Sex
)
data3 <- data.frame(
  x = clean_Penguins$`Culmen Length (mm)`,
  y = clean_Penguins$`Body Mass (g)`,
  category = clean_Penguins$Sex
)
ggplot(data1, aes(x = x, y = y, color = category)) +
```

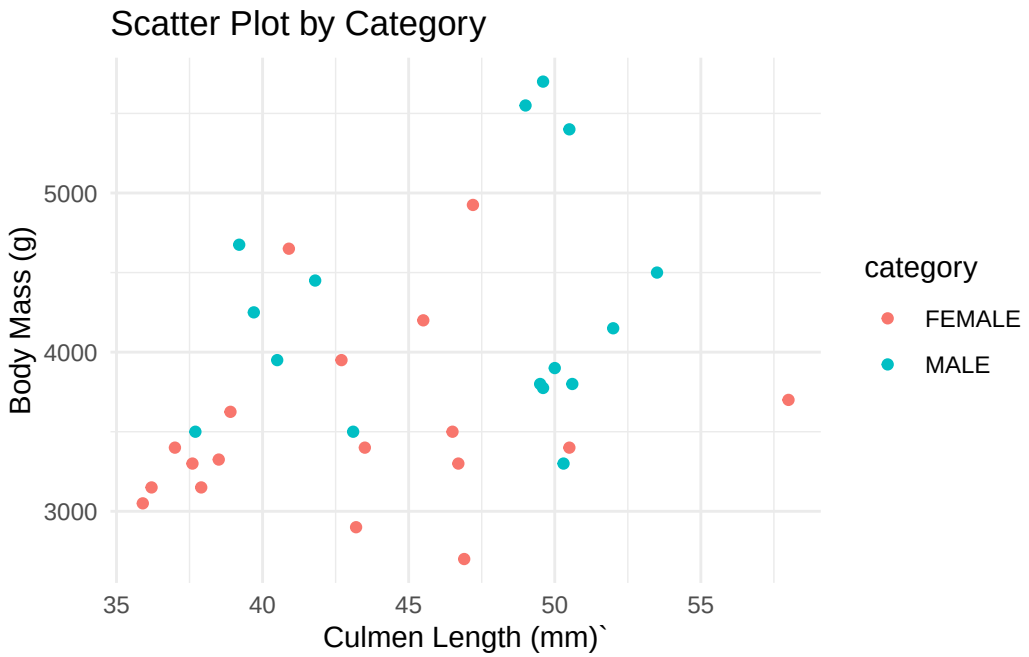
```
  geom_point() +
  labs(title = "Scatter Plot by Category", x = "Flipper Length (mm)`", y = "Body Mass (g)") +
  theme_minimal()
```




```
ggplot(data2, aes(x = x, y = y, color = category)) +
  geom_point() +
  labs(title = "Scatter Plot by Category", x = "Culmen Depth (mm)`", y = "Body Mass (g)") +
  theme_minimal()
```




```
ggplot(data3, aes(x = x, y = y, color = category)) +
  geom_point() +
  labs(title = "Scatter Plot by Category", x = "Culmen Length (mm)`", y = "Body Mass (g)") +
  theme_minimal()
```




```
colnames(clean_Penguins)
```

```
 [1] "studyName"           "Sample Number"       "Species"
 [4] "Region"              "Island"              "Stage"
 [7] "Individual ID"       "Clutch Completion"   "Date Egg"
[10] "Culmen Length (mm)"  "Culmen Depth (mm)"   "Flipper Length (mm)"
[13] "Body Mass (g)"       "Sex"                 "Delta 15 N (o/oo)"
[16] "Delta 13 C (o/oo)"   "Comments"
```

```
clean_Penguins_numeric <- clean_Penguins[, c(10,11,12,13,15,16)]
cor_matrix <- cor(clean_Penguins_numeric, use = "complete.obs")
print(cor_matrix)
```

```
                   Culmen Length (mm) Culmen Depth (mm) Flipper Length (mm)
Culmen Length (mm)          1.00000000        0.06930721           0.3873545
Culmen Depth (mm)           0.06930721        1.00000000          -0.4878841
Flipper Length (mm)         0.38735449       -0.48788408           1.0000000
Body Mass (g)               0.32411747       -0.30617689           0.7624367
Delta 15 N (o/oo)           0.23791352        0.68799862          -0.3576174
Delta 13 C (o/oo)           0.40254292        0.51959433          -0.3845557
                   Body Mass (g) Delta 15 N (o/oo) Delta 13 C (o/oo)
```

```
Culmen Length (mm)      0.3241175          0.2379135          0.4025429
Culmen Depth (mm)      -0.3061769          0.6879986          0.5195943
Flipper Length (mm)     0.7624367         -0.3576174         -0.3845557
Body Mass (g)           1.0000000         -0.4327395         -0.3773092
Delta 15 N (o/oo)      -0.4327395          1.0000000          0.6098007
Delta 13 C (o/oo)      -0.3773092          0.6098007          1.0000000
```

```
# correlation   Flipper Length Body Mass
#  Culmen Depth Delta 15 N Delta 15 N Delta 13 N      0.6
```

Shapiro-Wilk normality test

```
x1<- as.numeric(unlist(clean_Penguins_numeric[, 4]))
y1 <- as.numeric(unlist(clean_Penguins_numeric[, 3]))
shapiro_test1 <- shapiro.test(x1)
shapiro_test2 <- shapiro.test(y1)
print(shapiro_test1)
```

```
    Shapiro-Wilk normality test

data:  x1
W = 0.92762, p-value = 0.02668
```

```
print(shapiro_test2)
```

```
    Shapiro-Wilk normality test

data:  y1
W = 0.96177, p-value = 0.2736
```

Shapiro-wilk normality test的虛無假設為來自常態分配， 第一個test檢定Body Mass可以發現，
在設定alpha=0.05下， 其p-value=0.027<0.05,reject 虛無假設， 意味著Body Mass不是來自常態分配，
而從上面的histogram也可以看出其分配有右偏的情形。 第二個test檢定Flipper Length
(mm)可以發現， 在設定alpha=0.05下， 其p-value=0.27>0.05,do not reject 虛無假設，
意味著Flipper Length來自常態分配， 而從上面的histogram也可以看出其趨勢量集中在中間的情形。

Pearson's product-moment correlation

```
#
cor_test_result <- cor.test(x1, y1)
print(cor_test_result)
```

```
    Pearson's product-moment correlation

data:  x1 and y1
t = 6.6655, df = 32, p-value = 1.604e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5716627 0.8750009
sample estimates:
      cor
0.7624367
```

對照上面的correlaton plot，結果相同

Pearson's Chi-squared test

```
table_data <- table(clean_Penguins$Sex, clean_Penguins$Island)
chisq_result <- chisq.test(table_data)
```

```
Warning in chisq.test(table_data): Chi-squared approximation may be incorrect
```

```
print(chisq_result)
```

```
    Pearson's Chi-squared test

data:  table_data
X-squared = 0.052469, df = 2, p-value = 0.9741
```

```r
table_data <- table(clean_Penguins$Species, clean_Penguins$Island)
chisq_result <- chisq.test(table_data)
```

```
Warning in chisq.test(table_data): Chi-squared approximation may be incorrect
```

```r
print(chisq_result)
```

```

    Pearson's Chi-squared test

data:  table_data
X-squared = 43.915, df = 4, p-value = 6.683e-09
```

Chi-squared test的虛無假設為兩變數獨立(不相關)， 第一個test檢定Sex和Island可以發現，
在設定alpha=0.05下，其p-value=0.9741 > 0.05, do not #reject虛無假設，意味著Sex和Island獨立，
沒有太大的關聯 第二個test檢定Species和Island可以發現， 意味著Species和Island有關聯，符合直觀，
地域性確實可能會影響一個物種種類的分布情形