HW1

Lee Tzuhui

2024-09-14

Table of contents


```
#str(penguins_raw)
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
```

studyName: 研究名稱，代表進行該數據收集的研究項目或計畫。

Sample Number: 樣本編號，用來唯一識別每個樣本的編號。

Species: 企鵝的種類，常見的物種包括 Adelie、Chinstrap 和 Gentoo。

Region: 企鵝所處的地理區域，通常指研究中涉及的地理範圍。

Island: 企鵝棲息的島嶼，如 Biscoe、Dream 和 Torgersen。

Stage: 企鵝生命或繁殖階段，可能代表研究時所處的繁殖周期階段。

Individual ID: 個體識別碼，每隻企鵝的唯一身份編號。

Clutch Completion: 是否完成了鳥巢中的蛋孵化工作，可能為「Yes」或「No」。

Date Egg: 蛋的日期，可能指第一次發現蛋的日期或孵化日期。

Culmen Length (mm): 喙峰長度（以毫米為單位），指喙從基部到尖端的長度。

Culmen Depth (mm): 喙峰深度（以毫米為單位），指喙的高度。

Flipper Length (mm): 翼的長度（以毫米為單位）。

Body Mass (g): 體重（以克為單位）。

Sex: 企鵝的性別，通常為 male（雄性）或 female（雌性）。有時可能會有缺失值。

Delta 15 N (o/oo): 氮-15 同位素比值，與食物鏈中的營養位置有關，單位是千分比（o/oo）。

Delta 13 C (o/oo): 碳-13 同位素比值，通常用來判斷企鵝的食物來源或棲息環境，單位是千分比（o/oo）。

Comments: 研究者的評論欄，可能包含對樣本的觀察或其他補充資訊。

```
#   skim()
skim(penguins_raw)
```

Table 1: Data summary

Name	penguins_raw
Number of rows	344
Number of columns	17
_____	
Column type frequency:	
character	9
Date	1
numeric	7
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
studyName	0	1.00	7	7	0	3	0
Species	0	1.00	33	41	0	3	0
Region	0	1.00	6	6	0	1	0
Island	0	1.00	5	9	0	3	0
Stage	0	1.00	18	18	0	1	0
Individual ID	0	1.00	4	6	0	190	0
Clutch Completion	0	1.00	2	3	0	2	0
Sex	11	0.97	4	6	0	2	0
Comments	290	0.16	18	68	0	10	0

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
Date Egg	0	1	2007-11-09	2009-12-01	2008-11-09	50

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Sample Number	0	1.00	63.15	40.43	1.00	29.00	58.00	95.25	152.00	▇▆▆▃▂
Culmen Length (mm)	2	0.99	43.92	5.46	32.10	39.23	44.45	48.50	59.60	▃▇▇▆▁
Culmen Depth (mm)	2	0.99	17.15	1.97	13.10	15.60	17.30	18.70	21.50	▅▇▇▆▂
Flipper Length (mm)	2	0.99	200.92	14.06	172.00	190.00	197.00	213.00	231.00	▂▇▃▅▂
Body Mass (g)	2	0.99	4201.75	801.95	2700.00	3550.00	4050.00	4750.00	6300.00	▃▇▆▃▂
Delta 15 N (o/oo)	14	0.96	8.73	0.55	7.63	8.30	8.65	9.17	10.03	▃▇▆▃▂
Delta 13 C (o/oo)	13	0.96	-25.69	0.79	-27.02	-26.32	-25.83	-25.06	-23.79	▂▇▆▃▂

```
# Species, Island, and Sex
multi_var_count <- penguins_raw %>%
  filter(!is.na(Species), !is.na(Island),!is.na(Sex)) %>%
  group_by(Species, Island, Sex) %>%
  summarise(Count = n(), .groups = 'drop')


knitr::kable(multi_var_count, caption = "Species, Island   Sex    ")
```

Table 5: Species, Island 和 Sex 的分佈表

Species	Island	Sex	Count
Adelie Penguin (Pygoscelis adeliae)	Biscoe	FEMALE	22
Adelie Penguin (Pygoscelis adeliae)	Biscoe	MALE	22

Species	Island	Sex	Count
Adelie Penguin (Pygoscelis adeliae)	Dream	FEMALE	27
Adelie Penguin (Pygoscelis adeliae)	Dream	MALE	28
Adelie Penguin (Pygoscelis adeliae)	Torgersen	FEMALE	24
Adelie Penguin (Pygoscelis adeliae)	Torgersen	MALE	23
Chinstrap penguin (Pygoscelis antarctica)	Dream	FEMALE	34
Chinstrap penguin (Pygoscelis antarctica)	Dream	MALE	34
Gentoo penguin (Pygoscelis papua)	Biscoe	FEMALE	58
Gentoo penguin (Pygoscelis papua)	Biscoe	MALE	61

```
library(dplyr)

# 1.
string_vars <- sapply(penguins_raw, is.character)

# 2.
for (var_name in names(penguins_raw)[string_vars]) {
  cat("    -", var_name, ":\n")

  #   count()          NA
  var_count <- penguins_raw %>%
    filter(!is.na(!!sym(var_name)), !!sym(var_name) != "") %>%
    count(!!sym(var_name))

  #
  if(nrow(var_count) == 0) {
    cat("     \n")
  } else {
    print(var_count)
  }
  cat("\n")  #
}
```

```
    - studyName :
# A tibble: 3 x 2
  studyName     n
  <chr>     <int>
1 PAL0708     110
2 PAL0809     114
3 PAL0910     120
```

```
   - Species :
# A tibble: 3 x 2
  Species                                      n
  <chr>                                    <int>
1 Adelie Penguin (Pygoscelis adeliae)        152
2 Chinstrap penguin (Pygoscelis antarctica)   68
3 Gentoo penguin (Pygoscelis papua)          124

   - Region :
# A tibble: 1 x 2
  Region     n
  <chr>  <int>
1 Anvers   344

   - Island :
# A tibble: 3 x 2
  Island        n
  <chr>     <int>
1 Biscoe      168
2 Dream       124
3 Torgersen    52

   - Stage :
# A tibble: 1 x 2
  Stage                  n
  <chr>              <int>
1 Adult, 1 Egg Stage   344

   - Individual ID :
# A tibble: 190 x 2
   `Individual ID`     n
   <chr>           <int>
 1 N100A1              1
 2 N100A2              1
 3 N10A1               1
 4 N10A2               1
 5 N11A1               2
 6 N11A2               2
 7 N12A1               2
 8 N12A2               2
 9 N13A1               3
10 N13A2               3
# i 180 more rows
```

```
   - Clutch Completion :
# A tibble: 2 x 2
  `Clutch Completion`     n
  <chr>               <int>
1 No                     36
2 Yes                   308

   - Sex :
# A tibble: 2 x 2
  Sex        n
  <chr>  <int>
1 FEMALE   165
2 MALE     168

   - Comments :
# A tibble: 10 x 2
   Comments                                                              n
   <chr>                                                             <int>
 1 Adult not sampled.                                                    1
 2 Adult not sampled. Nest never observed with full clutch.              1
 3 Nest never observed with full clutch.                                34
 4 Nest never observed with full clutch. Not enough blood for isotopes.  1
 5 No blood sample obtained for sexing.                                  2
 6 No blood sample obtained.                                             2
 7 No delta15N data received from lab.                                   1
 8 Not enough blood for isotopes.                                        7
 9 Sexing primers did not amplify.                                       4
10 Sexing primers did not amplify. Not enough blood for isotopes.        1
```