H24101256

chen–fu LIU

2024–09–17

Table of contents


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

```
# 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 Staistic

```
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
```

```
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)	Gentoo penguin (Pygoscelis papua)
Frequency	152	68	124
Proportion	0.442	0.198	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

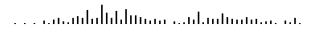
```
lowest : 32.1 33.1 33.5 34   34.1, highest: 55.1 55.8 55.9 58   59.6
```

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.
```

中文

Table 1

```r
library(table1)
str(penguins_raw)
```

```
tibble [344 x 17] (S3: tbl_df/tbl/data.frame)
 $ studyName          : chr [1:344] "PAL0708" "PAL0708" "PAL0708" "PAL0708" ...
 $ Sample Number      : num [1:344] 1 2 3 4 5 6 7 8 9 10 ...
 $ Species            : chr [1:344] "Adelie Penguin (Pygoscelis adeliae)" "Adelie Penguin (Py
 $ Region             : chr [1:344] "Anvers" "Anvers" "Anvers" "Anvers" ...
 $ Island             : chr [1:344] "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...
 $ Stage              : chr [1:344] "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" "Adult, 1 Egg S
 $ Individual ID      : chr [1:344] "N1A1" "N1A2" "N2A1" "N2A2" ...
 $ Clutch Completion  : chr [1:344] "Yes" "Yes" "Yes" "Yes" ...
 $ 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                : chr [1:344] "MALE" "FEMALE" "FEMALE" 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"
```

```r
penguins_raw <- penguins_raw
table1(~ Species+`Culmen Length (mm)`+`Culmen Depth (mm)`+`Body Mass (g)`+`Delta 15 N (o/oo)
```

	Biscoe	Dream	Torgersen	Overall
	(N=168)	(N=124)	(N=52)	(N=344)
Species				

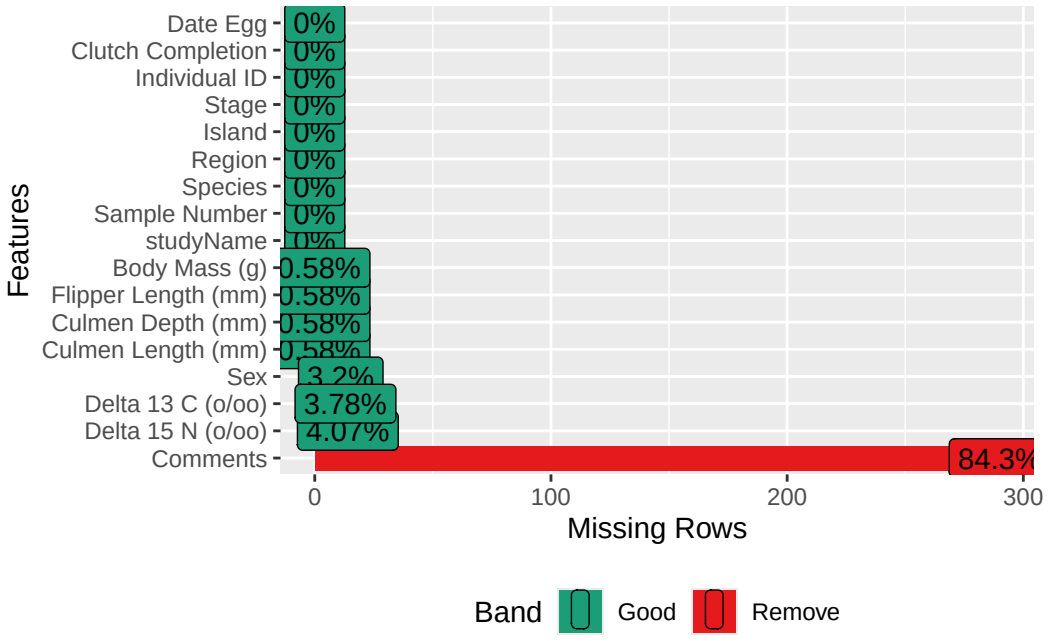
	Biscoe	Dream	Torgersen	Overall
Adelie Penguin (Pygoscelis adeliae)	44 (26.2%)	56 (45.2%)	52 (100%)	152 (44.2%)
Gentoo penguin (Pygoscelis papua)	124 (73.8%)	0 (0%)	0 (0%)	124 (36.0%)
Chinstrap penguin (Pygoscelis antarctica)	0 (0%)	68 (54.8%)	0 (0%)	68 (19.8%)
Culmen Length (mm)				
Mean (SD)	45.3 (4.77)	44.2 (5.95)	39.0 (3.03)	43.9 (5.46)
Median [Min, Max]	45.8 [34.5, 59.6]	44.7 [32.1, 58.0]	38.9 [33.5, 46.0]	44.5 [32.1, 59.6]
Missing	1 (0.6%)	0 (0%)	1 (1.9%)	2 (0.6%)
Culmen Depth (mm)				
Mean (SD)	15.9 (1.82)	18.3 (1.13)	18.4 (1.34)	17.2 (1.97)
Median [Min, Max]	15.5 [13.1, 21.1]	18.4 [15.5, 21.2]	18.4 [15.9, 21.5]	17.3 [13.1, 21.5]
Missing	1 (0.6%)	0 (0%)	1 (1.9%)	2 (0.6%)
Body Mass (g)				
Mean (SD)	4720 (783)	3710 (417)	3710 (445)	4200 (802)
Median [Min, Max]	4780 [2850, 6300]	3690 [2700, 4800]	3700 [2900, 4700]	4050 [2700, 6300]
Missing	1 (0.6%)	0 (0%)	1 (1.9%)	2 (0.6%)
Delta 15 N (o/oo)				
Mean (SD)	8.40 (0.394)	9.18 (0.441)	8.79 (0.461)	8.73 (0.552)
Median [Min, Max]	8.36 [7.63, 9.80]	9.23 [8.01, 10.0]	8.87 [7.70, 9.59]	8.65 [7.63, 10.0]
Missing	2 (1.2%)	5 (4.0%)	7 (13.5%)	14 (4.1%)

```
# help(table1)
```

Missing Values

```
library(DataExplorer)

plot_missing(penguins_raw)
```


Features
Date Egg
Clutch Completion
Individual ID
Stage
Island
Region
Species
Sample Number
studyName
Body Mass (g)
Flipper Length (mm)
Culmen Depth (mm)
Culmen Length (mm)
Sex
Delta 13 C (o/oo)
Delta 15 N (o/oo)
Comments
0%
0%
0%
0%
0%
0%
0%
0%
0%
0.58%
0.58%
0.58%
0.58%
3.2%
3.78%
4.07%
84.3%
0
100
200
300
Missing Rows
Band
Good
Remove