My Report

H24100014溫翎傑

2024-09-16

Table of contents

Statistical Thinking

Reference: https://www.fharrell.com/post/rflow/

Summary Staistic

```r
library(palmerpenguins)
library(Hmisc)
```

```
Attaching package: 'Hmisc'

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

    format.pval, units
```

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

<div align="center">

penguins
8 Variables   344 Observations

</div>

---

species

n	missing	distinct
344	0	3

Value	Adelie	Chinstrap	Gentoo
Frequency	152	68	124
Proportion	0.442	0.198	0.360

---

island

n	missing	distinct
344	0	3

Value	Biscoe	Dream	Torgersen
Frequency	168	124	52
Proportion	0.488	0.360	0.151

---

bill_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
```

---

bill_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

---

year

	n	missing	distinct	Info	Mean	Gmd
	344	0	3	0.888	2008	0.8919

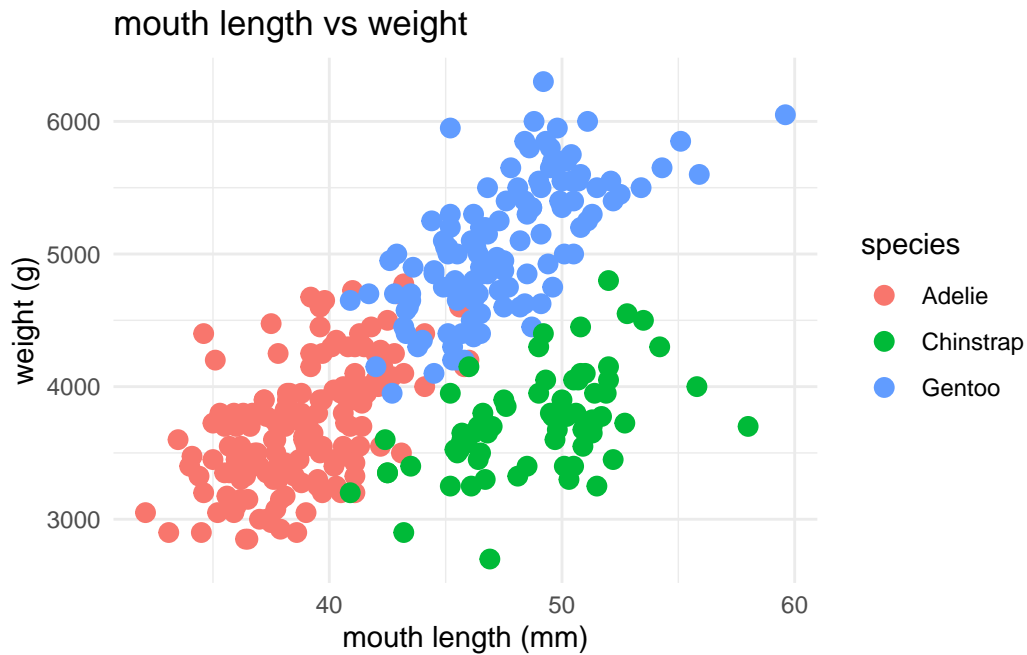| Value	2007	2008	2009
Frequency	110	114	120
Proportion	0.320	0.331	0.349

For the frequency table, variable is rounded to the nearest 0

中文

Table 1

```r
library(palmerpenguins)
library(ggplot2)
ggplot(penguins, aes(x = bill_length_mm, y = body_mass_g, color = species)) +
  geom_point(size = 3) +
  labs(title = "mouth length vs weight",
       x = "mouth length (mm)",
       y = "weight (g)") +
  theme_minimal()
```

mouth length vs weight

Missing Values

```r
library(palmerpenguins)
library(DataExplorer)
library(Hmisc)
latex(describe(penguins), file = "", caption.placement = "top")
```

penguins
8 Variables   344 Observations

species

n	missing	distinct
344	0	3

Value	Adelie	Chinstrap	Gentoo
Frequency	152	68	124
Proportion	0.442	0.198	0.360

island

n	missing	distinct
344	0	3

Value	Biscoe	Dream	Torgersen
Frequency	168	124	52
Proportion	0.488	0.360	0.151

---

bill_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

---

bill_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

---

year

n	missing	distinct	Info	Mean	Gmd
344	0	3	0.888	2008	0.8919

```
Value       2007  2008  2009
Frequency    110   114   120
Proportion 0.320 0.331 0.349
```

For the frequency table, variable is rounded to the nearest 0

```
plot_missing(penguins)
```