

# HW1

Wu Cheng-en

2024-09-24

## Table of contents

<b>Base</b>	<b>1</b>
<b>Summary report of penguins_raw</b>	<b>1</b>
Brief view of the data . . . . .	1
Missing values . . . . .	4
Dividing them by species . . . . .	6
<b>Visualizing the data</b>	<b>7</b>

## Base

```
library(palmerpenguins)
```

## Summary report of penguins\_raw

### Brief view of the data

Penguins\_raw is a dataframe containing 344 data with 17 variables

```
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(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

```

```
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 (Pygoscelis adeliae)" ...
 $ 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 Stage" "Adult, 1 Egg Stage" ...
 $ 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"

```

## Missing values

We can see how many missing values in each variable

```
sum(is.na(penguins_raw$studyName))
```

```
[1] 0
```

```
sum(is.na(penguins_raw$`Sample Number`))
```

```
[1] 0
```

```
sum(is.na(penguins_raw$Species))
```

```
[1] 0
```

```
sum(is.na(penguins_raw$Region))
```

```
[1] 0
```

```
sum(is.na(penguins_raw$Island))
```

```
[1] 0
```

```
sum(is.na(penguins_raw$Stage))
```

```
[1] 0
```

```
sum(is.na(penguins_raw$`Individual ID`))
```

```
[1] 0
```

```
sum(is.na(penguins_raw$`Clutch Completion`))
```

```
[1] 0
```

```
sum(is.na(penguins_raw$`Date Egg`))
```

```
[1] 0
```

```
sum(is.na(penguins_raw$`Culmen Length (mm)`))
```

```
[1] 2
```

```
sum(is.na(penguins_raw$`Culmen Depth (mm)`))
```

```
[1] 2
```

```
sum(is.na(penguins_raw$`Flipper Length (mm)`))
```

```
[1] 2
```

```
sum(is.na(penguins_raw$`Body Mass (g)`))
```

```
[1] 2
```

```
sum(is.na(penguins_raw$Sex))
```

```
[1] 11
```

```
sum(is.na(penguins_raw$`Delta 15 N (o/oo)`)
```

```
[1] 14
```

```
sum(is.na(penguins_raw$`Delta 13 C (o/oo)`)
```

```
[1] 13
```

```
sum(is.na(penguins_raw))
```

```
[1] 336
```

## Dividing them by species

Since there are three species (Adelie,Gentoo,Chinstrap), we can divide them by species for further analyzation

```
v1 <- subset(penguins_raw,penguins_raw$Species == "Adelie Penguin (Pygoscelis adeliae)")
v2 <- subset(penguins_raw,penguins_raw$Species == "Gentoo penguin (Pygoscelis papua)")
v3 <- subset(penguins_raw,penguins_raw$Species == "Chinstrap penguin (Pygoscelis antarctica)")
v1
```

```
# A tibble: 152 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
7 PAL0708          7 Adelie Penguin~ Anvers Torge~ Adul~ N4A1
8 PAL0708          8 Adelie Penguin~ Anvers Torge~ Adul~ N4A2
9 PAL0708          9 Adelie Penguin~ Anvers Torge~ Adul~ N5A1
10 PAL0708         10 Adelie Penguin~ Anvers Torge~ Adul~ N5A2
```

```
# i 142 more rows
```

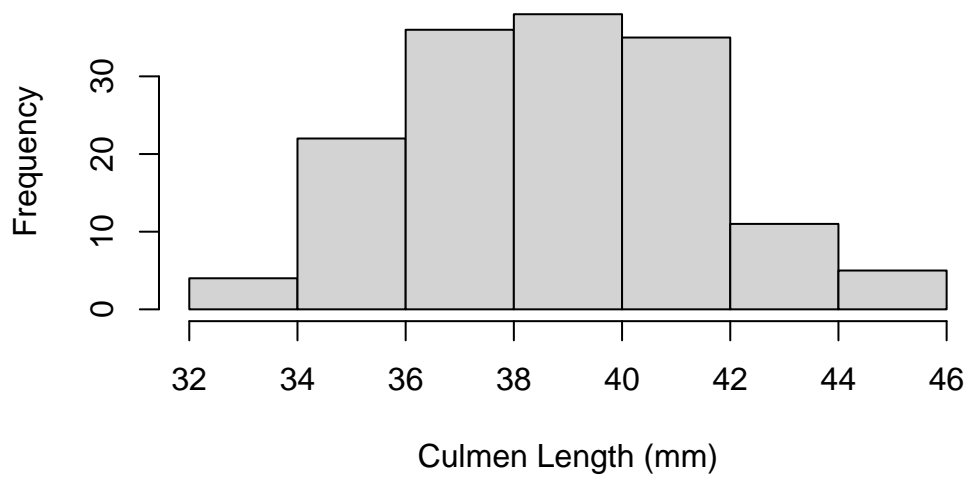
```
# i 10 more variables: `Clutch Completion` <chr>, `Date Egg` <date>,
#   `Culmen Length (mm)` <dbl>, `Culmen Depth (mm)` <dbl>,
#   `Flipper Length (mm)` <dbl>, `Body Mass (g)` <dbl>, Sex <chr>,
#   `Delta 15 N (o/oo)` <dbl>, `Delta 13 C (o/oo)` <dbl>, Comments <chr>
```

## Visualizing the data

1.Create three histograms visualizing the distributions of the Culmen Length (mm) for each species.

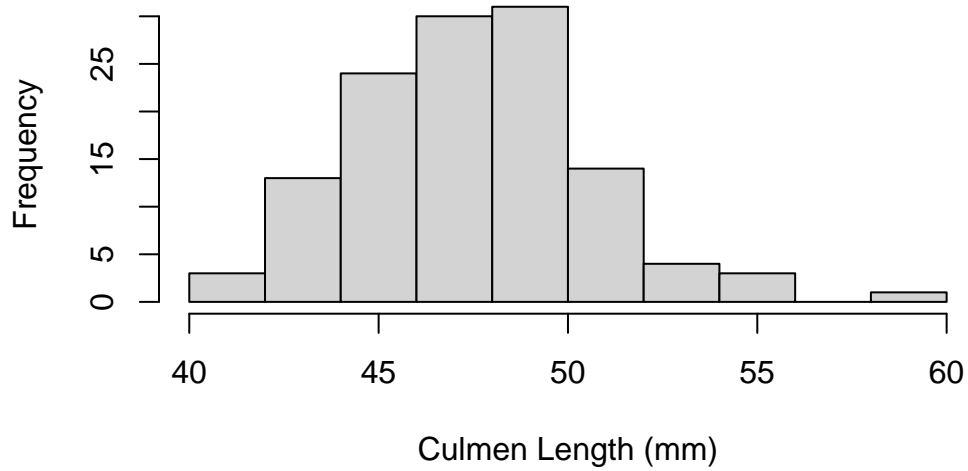
```
hist(v1$`Culmen Length (mm)`,main = "Culmen Length (mm) of Adelie Penguin", xlab = "Culmen L
```

**Culmen Length (mm) of Adelie Penguin**



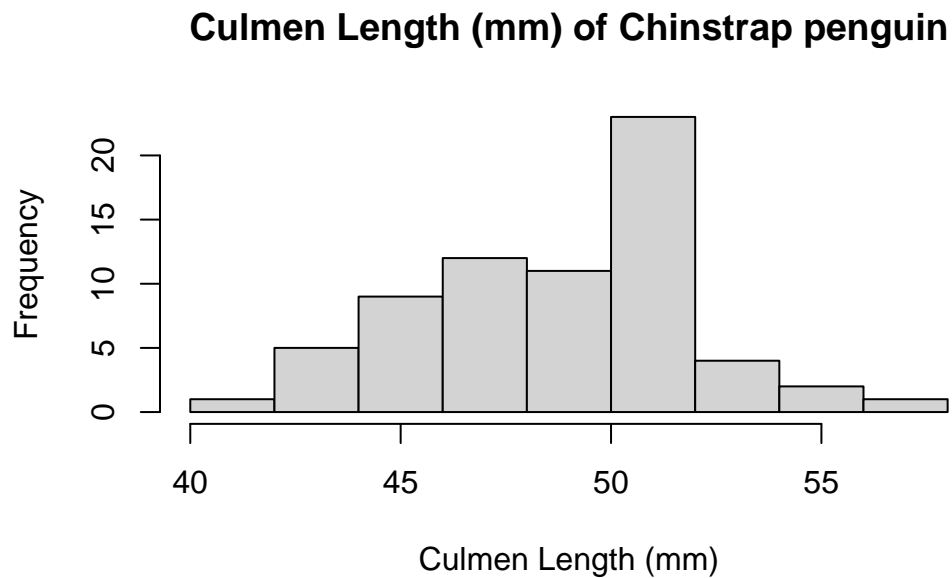
```
hist(v2$`Culmen Length (mm)`,main = "Culmen Length (mm) of Gentoo penguin", xlab = "Culmen L
```

**Culmen Length (mm) of Gentoo penguin**





```
hist(v3$`Culmen Length (mm)`,main = "Culmen Length (mm) of Chinstrap penguin", xlab = "Culmen Length (mm)")
```



2. create a box plot of the body mass of each penguin species

```
boxplot(body_mass_g ~ species,  
        data = penguins,  
        main="Body Mass (in g) by Penguin Species",  
        xlab="Penguin Species",  
        ylab="Body Mass (in g)",  
        col = c('coral', 'coral2', 'coral3'))
```

**Body Mass (in g) by Penguin Species**

