HW1

謝沛恩

9/16/24

Table of contents

Summary Staistic

先觀察資料欄位

```
library(palmerpenguins)
```

```
Warning: package 'palmerpenguins' was built under R version 4.2.3
```

```
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
```

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

```
[1] 336
```

Columns introduction

總共有17個變數，344筆資料

```
colnames(penguins_raw)
```

```
 [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"
```

連續變數（數值）：

Culmen Length (mm)：企鵝鳥喙長度

Culmen Depth (mm)：企鵝鳥喙寬度

Flipper Length (mm)：企鵝腳蹼長度

Body Mass (g)：企鵝體重

Delta 15 N (o/oo) - The ratio of isotope δ15 N

Delta 13 C (o/oo) - The ratio of isotope δ13 C

類別變數（分類）：

studyName：研究名稱

Sample Number：編號

Species：物種，有Chinstrap、Adélie、Gentoo三種

Region：地區

Island：島嶼，有Dream、Torgersen、Biscoe三個

Stage：狀態階段

Individual ID：各企鵝的個人ID

Clutch Completion：卵窩完整與否(是/否)

Sex：性別，男女

Comments：評論

時間變數： Data egg ：生蛋日期

Continuous variables

```
#
continuous_vars <- c('Culmen Length (mm)', 'Culmen Depth (mm)',
                     'Flipper Length (mm)', 'Body Mass (g)',
                     'Delta 15 N (o/oo)', 'Delta 13 C (o/oo)')

# Step 1: Summary
summary_stats <- lapply(penguins_raw[continuous_vars], summary)
print("Summary statistics for continuous variables:")
```

```
[1] "Summary statistics for continuous variables:"
```

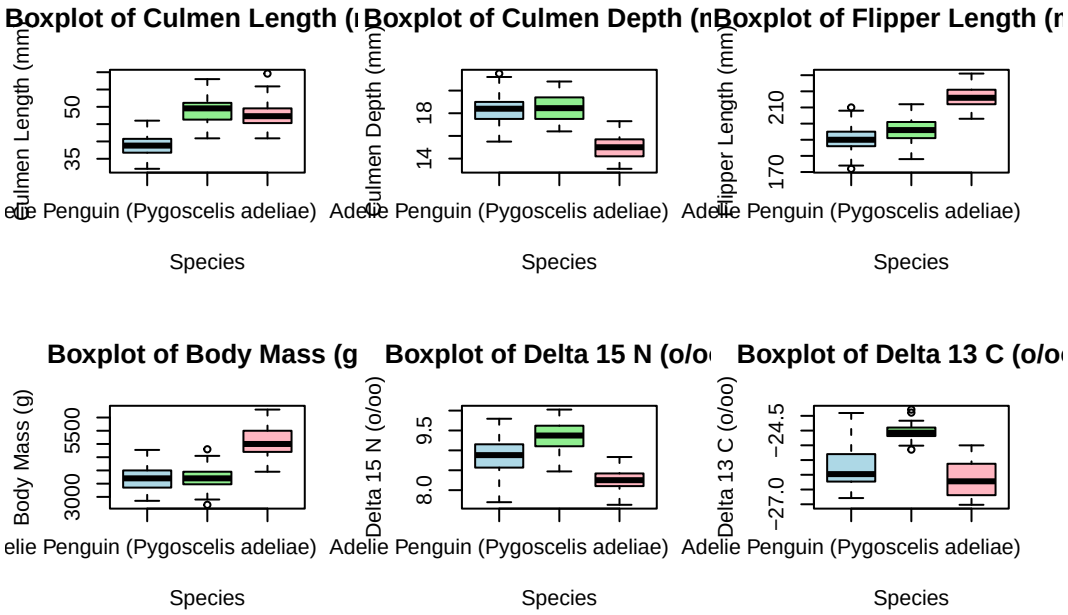
```
print(summary_stats)
```

```
$`Culmen Length (mm)`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  32.10   39.23   44.45   43.92   48.50   59.60       2

$`Culmen Depth (mm)`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  13.10   15.60   17.30   17.15   18.70   21.50       2

$`Flipper Length (mm)`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  172.0   190.0   197.0   200.9   213.0   231.0       2

$`Body Mass (g)`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   2700    3550    4050    4202    4750    6300       2

$`Delta 15 N (o/oo)`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  7.632   8.300   8.652   8.733   9.172  10.025      14

$`Delta 13 C (o/oo)`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 -27.02  -26.32  -25.83  -25.69  -25.06  -23.79      13
```

從此可看出連續變數基本的敘述統計，像是平均數、最大最小值 、缺失值個數等。

Step 2:繪製每個連續變數的 boxplot，並根據 Species 分組

```
par(mfrow = c(2, 3))  #

for (var in continuous_vars) {
  boxplot(penguins_raw[[var]] ~ penguins_raw$Species,
          main = paste("Boxplot of", var),
          xlab = "Species",
          ylab = var,
          col = c("lightblue", "lightgreen", "lightpink"))
}
```



Categorical variables

次數分配表

透過次數分配表，可看出各類別的次數與比例。

```r
categorical_columns1 <- c("studyName",'Species',"Island",'Region','Stage',
                          "Sex", 'Clutch Completion')
#
for (col in categorical_columns1) {
  print(paste("Category Distribution for", col, ":"))

  #
  value_counts <- table(penguins_raw[[col]])

  #
  percentage <- (value_counts / nrow(penguins_raw)) * 100

  #
  result <- data.frame(Count = value_counts, `Percentage (%)` = percentage)

  print(result)
  cat("\n")
}
```

```
[1] "Category Distribution for studyName :"
  Count.Var1 Count.Freq Percentage.....Var1 Percentage.....Freq
1    PAL0708        110             PAL0708            31.97674
2    PAL0809        114             PAL0809            33.13953
3    PAL0910        120             PAL0910            34.88372

[1] "Category Distribution for Species :"
                                Count.Var1 Count.Freq
1       Adelie Penguin (Pygoscelis adeliae)        152
2 Chinstrap penguin (Pygoscelis antarctica)         68
3         Gentoo penguin (Pygoscelis papua)        124
                        Percentage.....Var1 Percentage.....Freq
1       Adelie Penguin (Pygoscelis adeliae)            44.18605
2 Chinstrap penguin (Pygoscelis antarctica)            19.76744
3         Gentoo penguin (Pygoscelis papua)            36.04651

[1] "Category Distribution for Island :"
  Count.Var1 Count.Freq Percentage.....Var1 Percentage.....Freq
1     Biscoe        168              Biscoe            48.83721
2      Dream        124               Dream            36.04651
3  Torgersen         52           Torgersen            15.11628

[1] "Category Distribution for Region :"
```

```
  Count.Var1 Count.Freq Percentage.....Var1 Percentage.....Freq
1     Anvers        344              Anvers                 100
```

```
[1] "Category Distribution for Stage :"
          Count.Var1 Count.Freq Percentage.....Var1 Percentage.....Freq
1 Adult, 1 Egg Stage        344  Adult, 1 Egg Stage                 100
```

```
[1] "Category Distribution for Sex :"
  Count.Var1 Count.Freq Percentage.....Var1 Percentage.....Freq
1     FEMALE        165              FEMALE            47.96512
2       MALE        168                MALE            48.83721
```

```
[1] "Category Distribution for Clutch Completion :"
  Count.Var1 Count.Freq Percentage.....Var1 Percentage.....Freq
1         No         36                  No            10.46512
2        Yes        308                 Yes            89.53488
```

```
table(penguins_raw$`Individual ID`)#
```

```
N100A1 N100A2  N10A1  N10A2  N11A1  N11A2  N12A1  N12A2  N13A1  N13A2  N14A1
     1      1      1      1      2      2      2      2      3      3      2
 N14A2  N15A1  N15A2  N16A1  N16A2  N17A1  N17A2  N18A1  N18A2  N19A1  N19A2
     2      2      2      1      1      2      2      3      3      2      2
  N1A1   N1A2  N20A1  N20A2  N21A1  N21A2  N22A1  N22A2  N23A1  N23A2  N24A1
     2      2      2      2      3      3      3      3      3      3      3
 N24A2  N25A1  N25A2  N26A1  N26A2  N27A1  N27A2  N28A1  N28A2  N29A1  N29A2
     3      2      2      1      1      2      2      3      3      3      3
  N2A1   N2A2  N30A1  N30A2  N31A1  N31A2  N32A1  N32A2  N33A1  N33A2  N34A1
     2      2      2      2      1      1      3      3      1      1      3
 N34A2  N35A1  N35A2  N36A1  N36A2  N37A1  N37A2  N38A1  N38A2  N39A1  N39A2
     3      3      3      3      3      2      2      3      3      3      3
  N3A1   N3A2  N40A1  N40A2  N41A1  N41A2  N42A1  N42A2  N43A1  N43A2  N44A1
     1      1      2      2      2      2      2      2      1      1      2
 N44A2  N45A1  N45A2  N46A1  N46A2  N47A1  N47A2  N48A1  N48A2  N49A1  N49A2
     2      1      1      2      2      2      2      1      1      2      2
  N4A1   N4A2  N50A1  N50A2  N51A1  N51A2  N53A1  N53A2  N54A1  N54A2  N55A1
     2      2      2      2      2      2      2      2      1      1      2
 N55A2  N56A1  N56A2  N58A1  N58A2   N5A1   N5A2  N60A1  N60A2  N61A1  N61A2
     2      2      2      2      2      2      2      2      2      3      3
 N62A1  N62A2  N63A1  N63A2  N64A1  N64A2  N65A1  N65A2  N66A1  N66A2  N67A1
     2      2      2      2      2      2      1      1      2      2      3
```

```
 N67A2  N68A1  N68A2  N69A1  N69A2   N6A1   N6A2  N70A1  N70A2  N71A1  N71A2
     3      1      1      3      3      3      3      1      1      2      2
 N72A1  N72A2  N73A1  N73A2  N74A1  N74A2  N75A1  N75A2  N76A1  N76A2  N77A1
     3      3      2      2      1      1      1      1      1      1      1
 N77A2  N78A1  N78A2  N79A1  N79A2   N7A1   N7A2  N80A1  N80A2  N81A1  N81A2
     1      1      1      1      1      2      2      1      1      1      1
 N82A1  N82A2  N83A1  N83A2  N84A1  N84A2  N85A1  N85A2  N86A1  N86A2  N87A1
     1      1      1      1      1      1      2      2      1      1      1
 N87A2  N88A1  N88A2  N89A1  N89A2   N8A1   N8A2  N90A1  N90A2  N92A1  N92A2
     1      1      1      1      1      3      3      1      1      1      1
 N93A1  N93A2  N94A1  N94A2  N95A1  N95A2  N96A1  N96A2  N98A1  N98A2  N99A1
     1      1      1      1      1      1      1      1      1      1      1
 N99A2   N9A1   N9A2
     1      1      1
```

```
table(penguins_raw$Comments)#
```

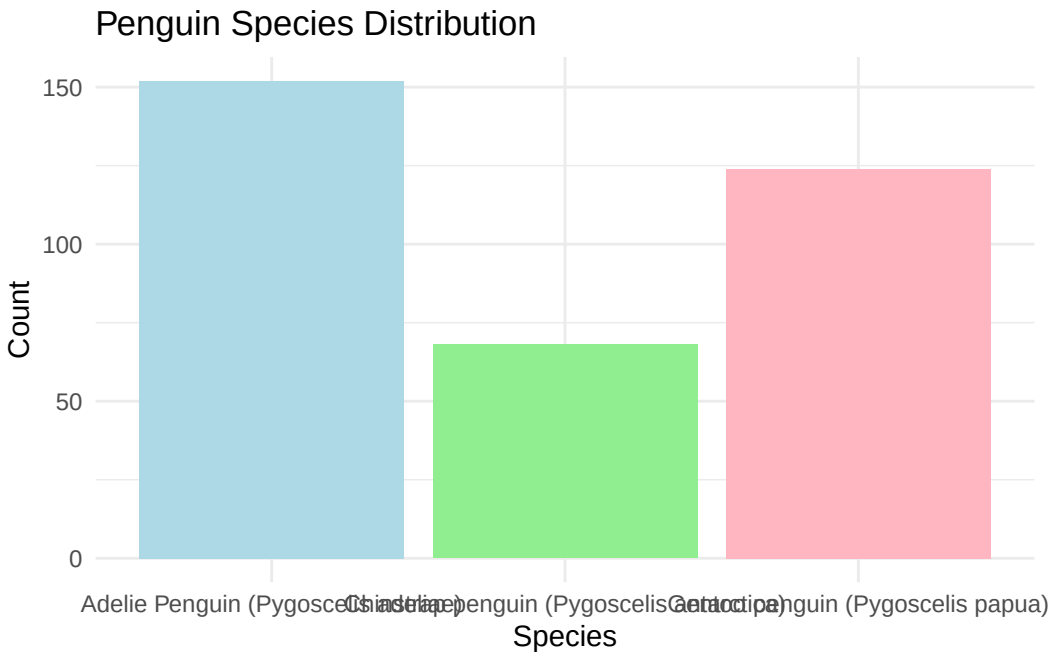
```
                                                  Adult not sampled.
                                                                   1
           Adult not sampled. Nest never observed with full clutch.
                                                                   1
                              Nest never observed with full clutch.
                                                                  34
Nest never observed with full clutch. Not enough blood for isotopes.
                                                                   1
                             No blood sample obtained for sexing.
                                                                   2
                                       No blood sample obtained.
                                                                   2
                              No delta15N data received from lab.
                                                                   1
                                   Not enough blood for isotopes.
                                                                   7
                                  Sexing primers did not amplify.
                                                                   4
     Sexing primers did not amplify. Not enough blood for isotopes.
                                                                   1
```

繪畫Species、studyName、Island、Sex、Clutch Completion的長條圖， 並以Species做分類。

```
library(ggplot2)
```

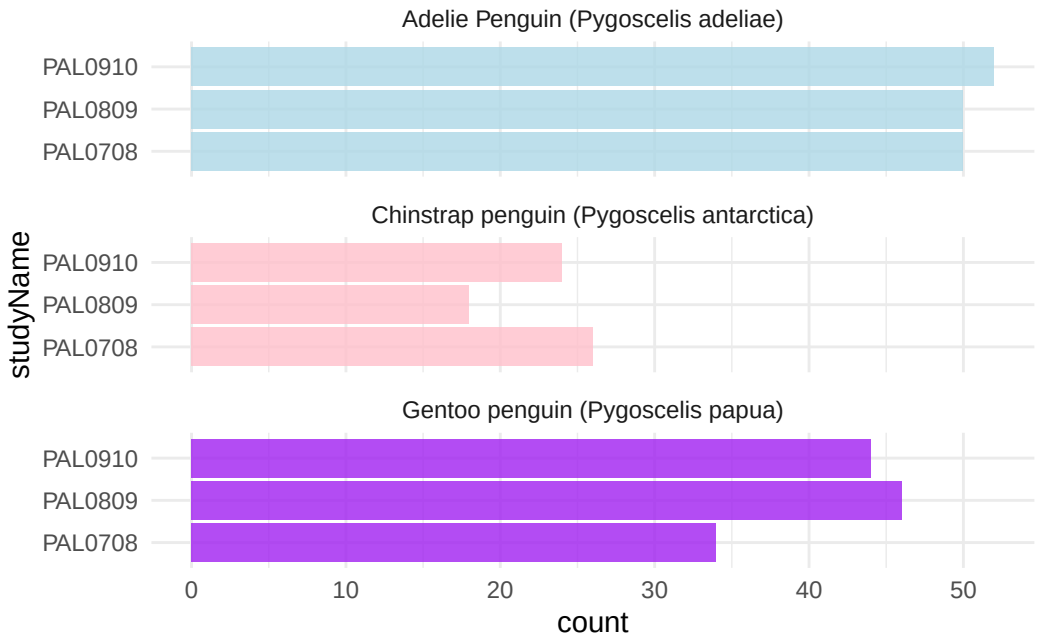
```
Warning: package 'ggplot2' was built under R version 4.2.3
```

```r
#Species bar chart
ggplot(penguins_raw, aes(x = Species)) +
  geom_bar(fill = c("lightblue", "lightgreen", "lightpink")) +
  labs(title = "Penguin Species Distribution", x = "Species", y = "Count") +
  theme_minimal()
```

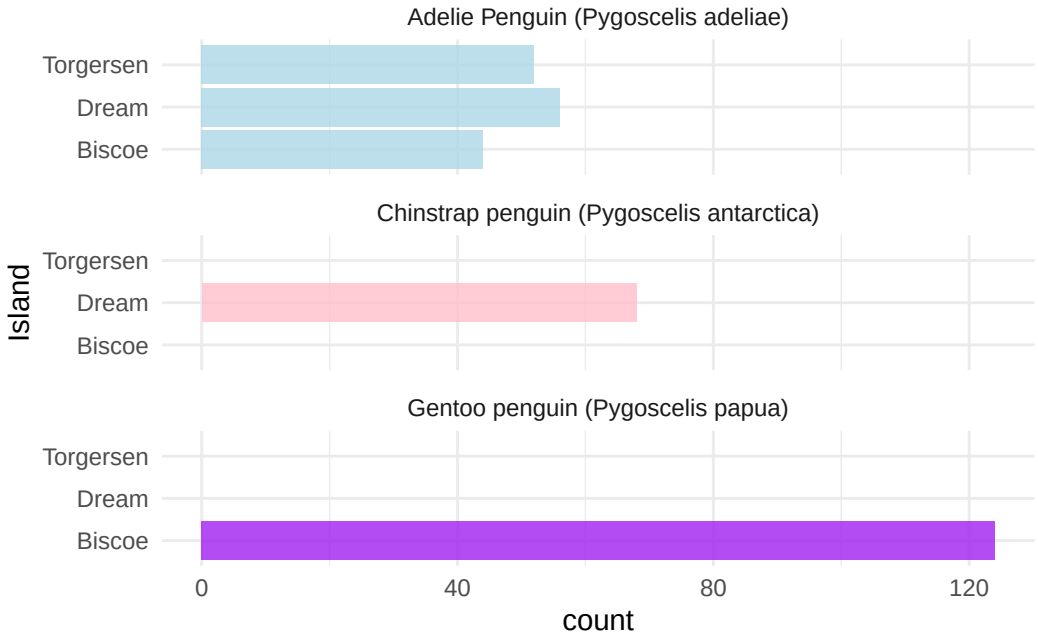



```r
#studyName
ggplot(penguins_raw, aes(x = studyName, fill = Species)) +
  geom_bar(alpha = 0.8) +
  scale_fill_manual(values = c("lightblue","pink","purple"),
                    guide = FALSE) +
  theme_minimal() +
  facet_wrap(~Species, ncol = 1) +
  coord_flip()
```

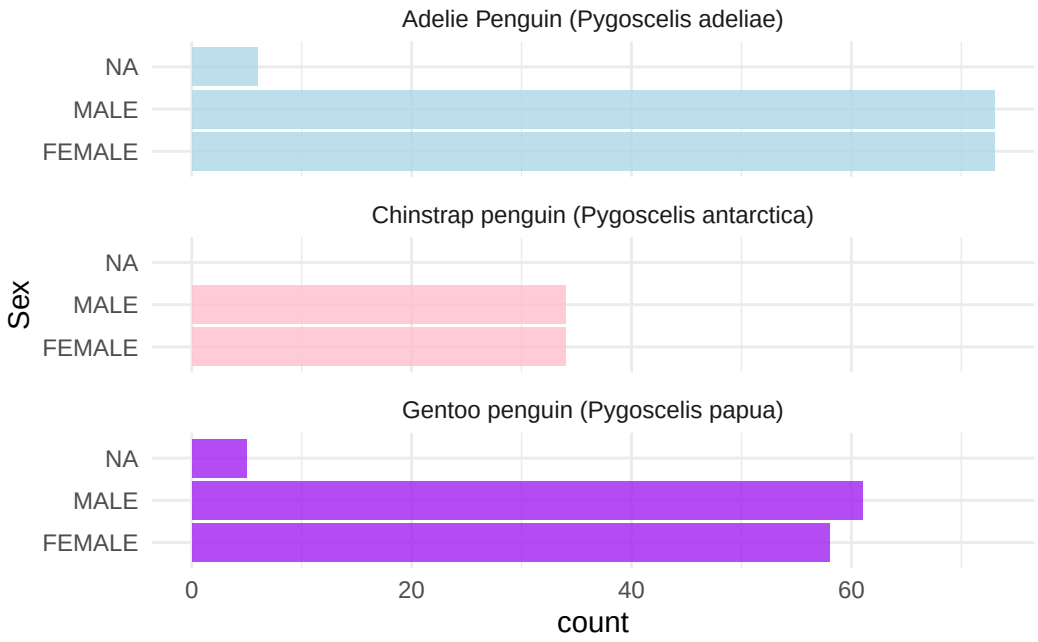
```
Warning: The `guide` argument in `scale_*()` cannot be `FALSE`. This was deprecated in
ggplot2 3.3.4.
i Please use "none" instead.
```



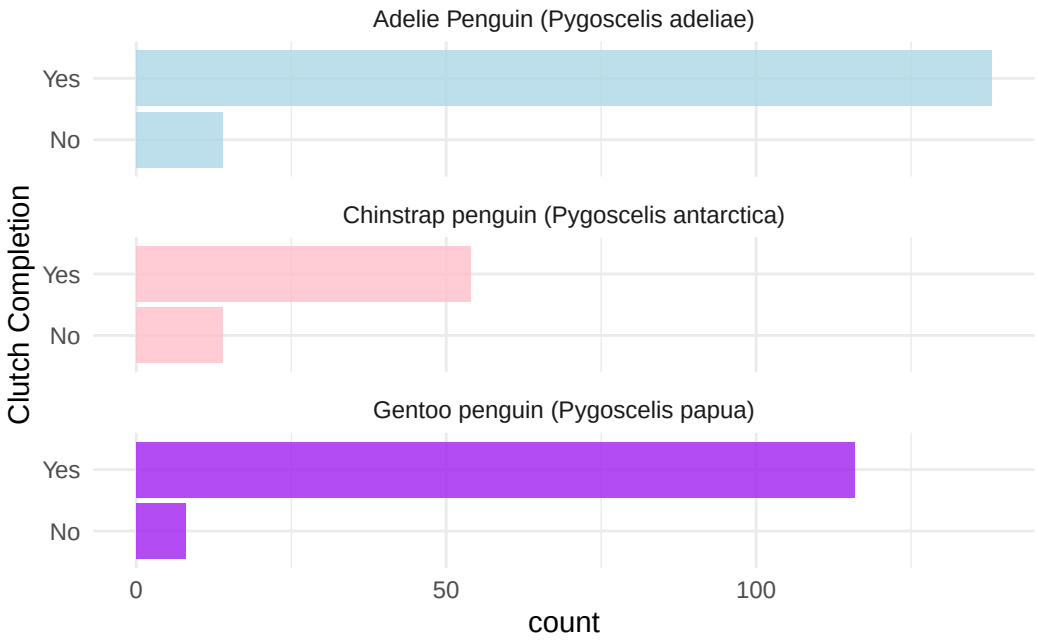

```
#Island
ggplot(penguins_raw, aes(x = Island, fill = Species)) +
  geom_bar(alpha = 0.8) +
  scale_fill_manual(values = c("lightblue","pink","purple"),
                    guide = FALSE) +
  theme_minimal() +
  facet_wrap(~Species, ncol = 1) +
  coord_flip()
```




```
#Sex
ggplot(penguins_raw, aes(x = Sex, fill = Species)) +
  geom_bar(alpha = 0.8) +
  scale_fill_manual(values = c("lightblue","pink","purple"),
                    guide = FALSE) +
  theme_minimal() +
  facet_wrap(~Species, ncol = 1) +
  coord_flip()
```




```
#Clutch Completion
ggplot(penguins_raw, aes(x = `Clutch Completion`, fill = Species)) +
  geom_bar(alpha = 0.8) +
  scale_fill_manual(values = c("lightblue","pink","purple"),
                    guide = FALSE) +
  theme_minimal() +
  facet_wrap(~Species, ncol = 1) +
  coord_flip()
```


Adelie Penguin (Pygoscelis adeliae)
Chinstrap penguin (Pygoscelis antarctica)
Gentoo penguin (Pygoscelis papua)
Clutch Completion
Yes
No
0
50
100
count