HW1

Chan Ya-Jun 詹雅鈞 H24101303

2024-09-16

Table of contents

查看資料筆數與變數名稱

```
library(palmerpenguins)
data(package = 'palmerpenguins')

colnames(penguins)
```

```
[1] "species"           "island"            "bill_length_mm"
[4] "bill_depth_mm"     "flipper_length_mm" "body_mass_g"
[7] "sex"               "year"
```

```
nrow(penguins)
```

```
[1] 344
```

```
head(penguins)
```

```
# A tibble: 6 x 8
  species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
  <fct>   <fct>              <dbl>         <dbl>             <int>       <int>
1 Adelie  Torgersen           39.1          18.7               181        3750
2 Adelie  Torgersen           39.5          17.4               186        3800
3 Adelie  Torgersen           40.3          18                 195        3250
```

```
4 Adelie  Torgersen           NA             NA                  NA          NA
5 Adelie  Torgersen           36.7           19.3               193        3450
6 Adelie  Torgersen           39.3           20.6               190        3650
# i 2 more variables: sex <fct>, year <int>
```

```
summary(penguins)
```

```
      species          island    bill_length_mm  bill_depth_mm
 Adelie   :152   Biscoe   :168   Min.   :32.10   Min.   :13.10
 Chinstrap: 68   Dream    :124   1st Qu.:39.23   1st Qu.:15.60
 Gentoo   :124   Torgersen: 52   Median :44.45   Median :17.30
                                 Mean   :43.92   Mean   :17.15
                                 3rd Qu.:48.50   3rd Qu.:18.70
                                 Max.   :59.60   Max.   :21.50
                                 NA's   :2       NA's   :2
 flipper_length_mm  body_mass_g       sex           year
 Min.   :172.0     Min.   :2700   female:165   Min.   :2007
 1st Qu.:190.0     1st Qu.:3550   male  :168   1st Qu.:2007
 Median :197.0     Median :4050   NA's  : 11   Median :2008
 Mean   :200.9     Mean   :4202                Mean   :2008
 3rd Qu.:213.0     3rd Qu.:4750                3rd Qu.:2009
 Max.   :231.0     Max.   :6300                Max.   :2009
 NA's   :2         NA's   :2
```

刪除資料的缺失值

```
penguins<-na.omit(penguins)
nrow(penguins)
```

```
[1] 333
```

類別變數contingency table

```
table( penguins$species, penguins$island)
```

```
           Biscoe Dream Torgersen
  Adelie       44    55        47
  Chinstrap     0    68         0
  Gentoo      119     0         0
```

```r
table( penguins$species, penguins$sex)
```

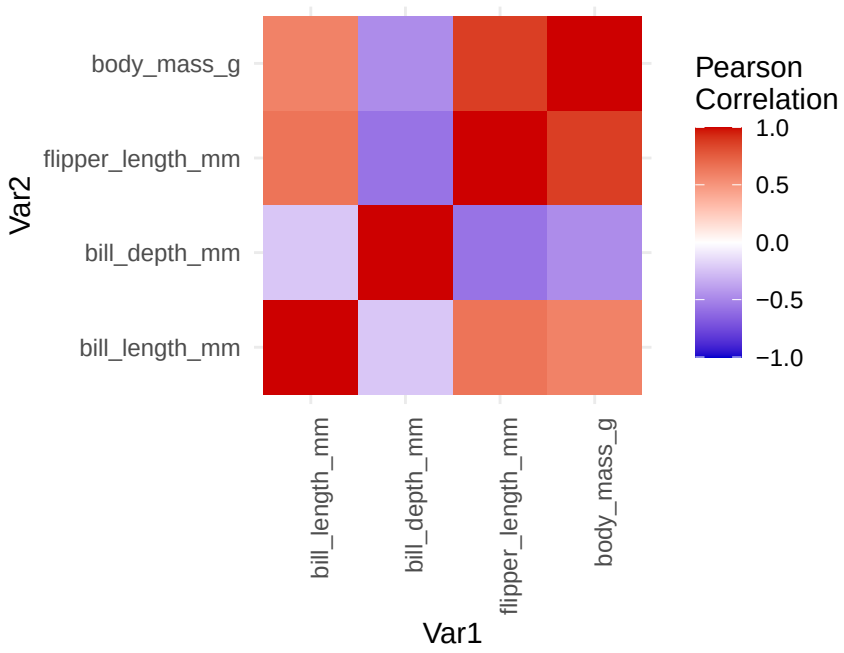
```
           female male
  Adelie       73   73
  Chinstrap    34   34
  Gentoo       58   61
```

連續變數相關係數+heatmap

```r
library(ggplot2)
round(cor(penguins[,3:6]),3)
```

```
                  bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
bill_length_mm             1.000        -0.229             0.653       0.589
bill_depth_mm             -0.229         1.000            -0.578      -0.472
flipper_length_mm          0.653        -0.578             1.000       0.873
body_mass_g                0.589        -0.472             0.873       1.000
```

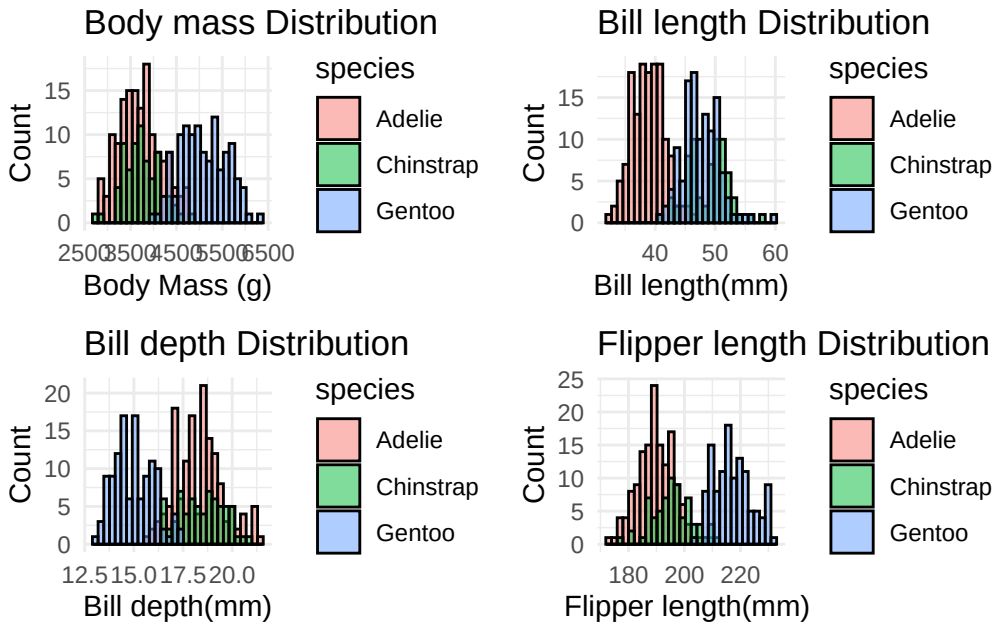
```r
library(reshape2)
melted_corr <- melt(cor(penguins[,3:6])) #reshape
ggplot(data = melted_corr, aes(x=Var1, y=Var2, fill=value)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue3", high = "red3", mid = "white",
                       midpoint = 0, limit = c(-1,1), space = "Lab",
                       name="Pearson\nCorrelation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 1,
                                   size = 9, hjust = 1)) +
  coord_fixed()
```



不同類別的連續變數分布

```
library(patchwork)

p1 <- ggplot(penguins, aes(x = body_mass_g, fill = species)) +
  geom_histogram(bins = 30, color = "black", alpha = 0.5, position = "identity") +
  labs(title = "Body mass Distribution",
       x = "Body Mass (g)",
       y = "Count") +
  theme_minimal()

p2 <- ggplot(penguins, aes(x = bill_length_mm , fill = species)) +
  geom_histogram(bins = 30, color = "black", alpha = 0.5, position = "identity") +
  labs(title = "Bill length Distribution",
       x = "Bill length(mm)",
       y = "Count") +
  theme_minimal()

p3 <- ggplot(penguins, aes(x = bill_depth_mm    , fill = species)) +
  geom_histogram(bins = 30, color = "black", alpha = 0.5, position = "identity") +
  labs(title = "Bill depth Distribution",
       x = "Bill depth(mm)",
```

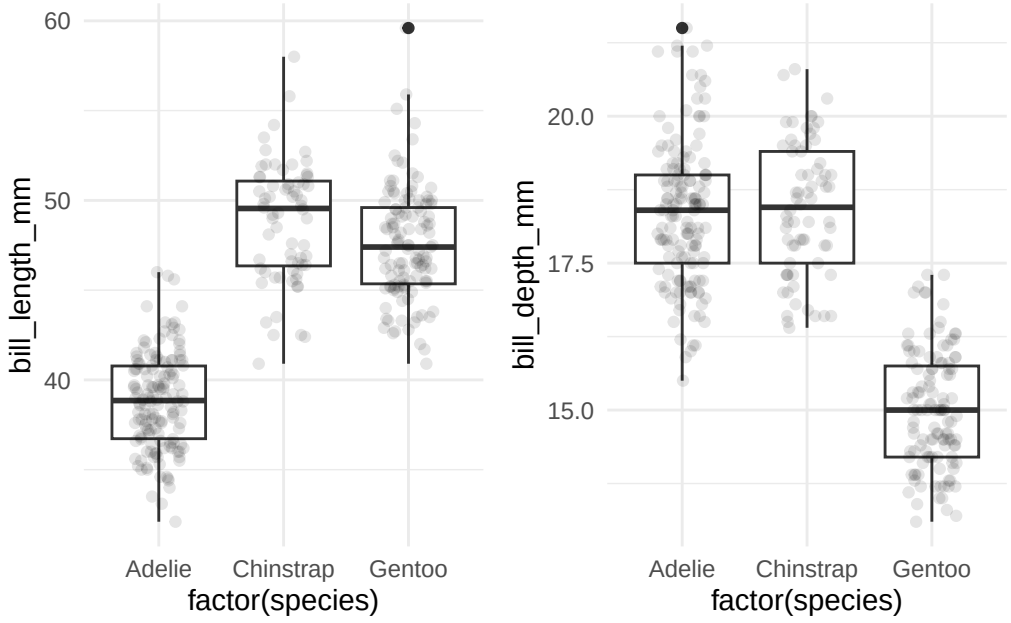
```
       y = "Count") +
  theme_minimal()

p4 <- ggplot(penguins, aes(x = flipper_length_mm    , fill = species)) +
  geom_histogram(bins = 30, color = "black", alpha = 0.5, position = "identity") +
  labs(title = "Flipper length Distribution",
       x = "Flipper length(mm) ",
       y = "Count") +
  theme_minimal()

(p1 | p2) / (p3| p4)
```



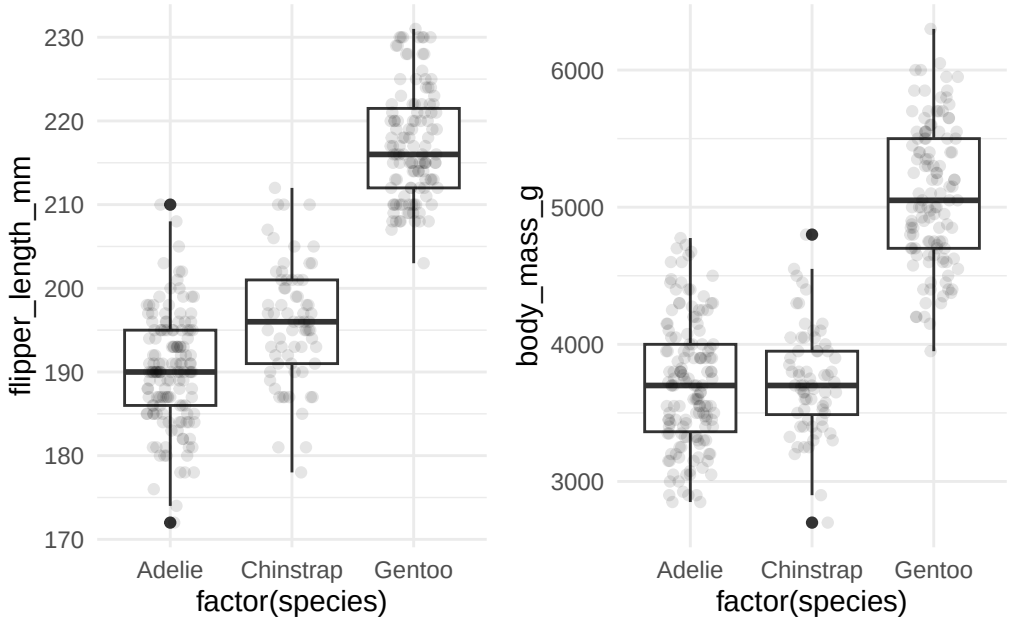
Box plot

```
p1<-ggplot(penguins, aes(x = factor(species), y = bill_length_mm  )) +
  geom_boxplot() +
  geom_jitter(width = 0.2, height = 0, alpha = 0.1) +
  theme_minimal()

p2<-ggplot(penguins, aes(x = factor(species), y = bill_depth_mm  )) +
  geom_boxplot() +
```

```
  geom_jitter(width = 0.2, height = 0, alpha = 0.1) +
  theme_minimal()

p3<-ggplot(penguins, aes(x = factor(species), y = flipper_length_mm   )) +
  geom_boxplot() +
  geom_jitter(width = 0.2, height = 0, alpha = 0.1) +
  theme_minimal()

p4<-ggplot(penguins, aes(x = factor(species), y = body_mass_g    )) +
  geom_boxplot() +
  geom_jitter(width = 0.2, height = 0, alpha = 0.1) +
  theme_minimal()
(p1|p2)
```

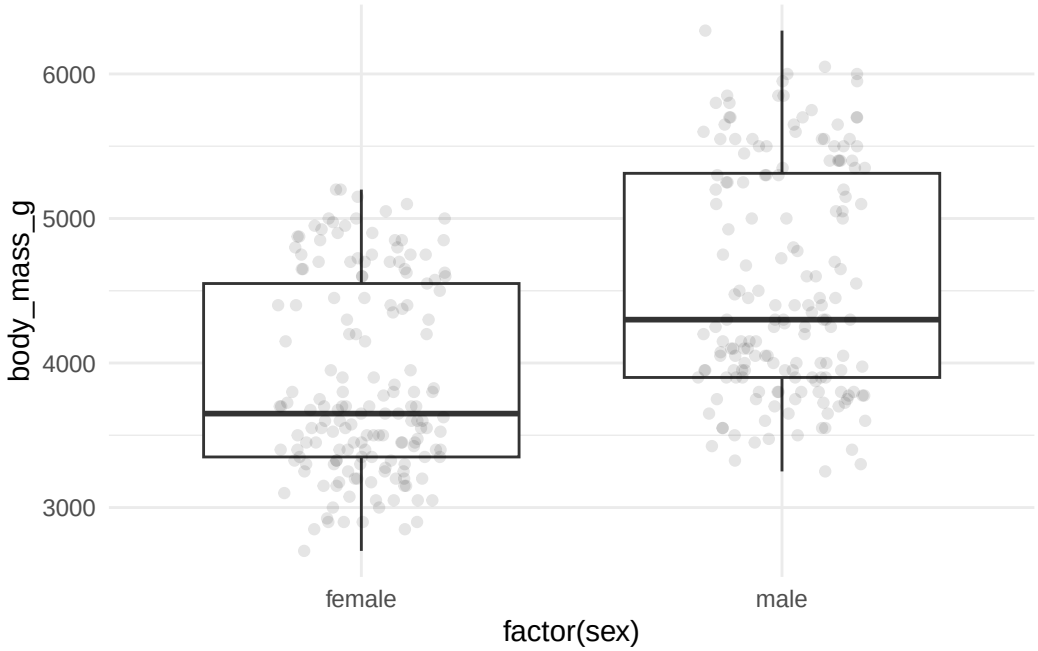



```
(p3|p4)
```



檢定不同性別的體重是否存在顯著差異

```
ggplot(penguins, aes(x = factor(sex), y = body_mass_g    )) +
  geom_boxplot() +
  geom_jitter(width = 0.2, height = 0, alpha = 0.1) +
  theme_minimal()
```




```
t.test(body_mass_g ~ sex, data = penguins)
```

```
    Welch Two Sample t-test

data:  body_mass_g by sex
t = -8.5545, df = 323.9, p-value = 4.794e-16
alternative hypothesis: true difference in means between group female and group male is not e
95 percent confidence interval:
 -840.5783 -526.2453
sample estimates:
mean in group female   mean in group male
            3862.273             4545.685
```

Scatterplot+lm

```
library(cowplot)
```

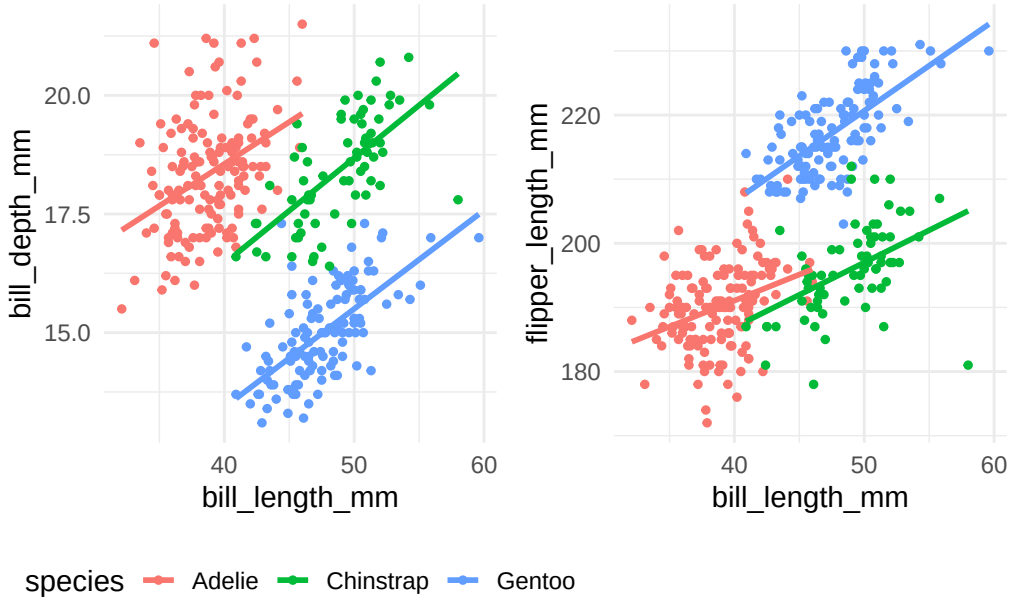
```
Attaching package: 'cowplot'
```

```
The following object is masked from 'package:patchwork':

    align_plots
```

```
p1<-ggplot(penguins, aes(x = bill_length_mm , y = bill_depth_mm , color = species)) +
  geom_point(size = 1) +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal()+theme( legend.key.width = unit(1, "lines") ,legend.position = "bottom")

p2<-ggplot(penguins, aes(x = bill_length_mm , y = flipper_length_mm  , color = species)) +
  geom_point(size = 1) +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal()  +theme( legend.key.width = unit(1, "lines") ,legend.position = "none")
p3<-ggplot(penguins, aes(x = bill_length_mm , y = body_mass_g   , color = species)) +
  geom_point(size = 1) +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal()  +theme(legend.position = "bottom")

p4<-ggplot(penguins, aes(x = bill_depth_mm , y = flipper_length_mm  , color = species)) +
  geom_point(size = 1) +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal()+theme(legend.position = "none")

p5<-ggplot(penguins, aes(x = bill_depth_mm  , y = body_mass_g   , color = species)) +
  geom_point(size = 1) +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal()+theme(legend.position = "bottom")

p6<-ggplot(penguins, aes(x = flipper_length_mm   , y = body_mass_g   , color = species)) +
  geom_point(size = 1) +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal()+theme(legend.position = "none")

p1|p2
```

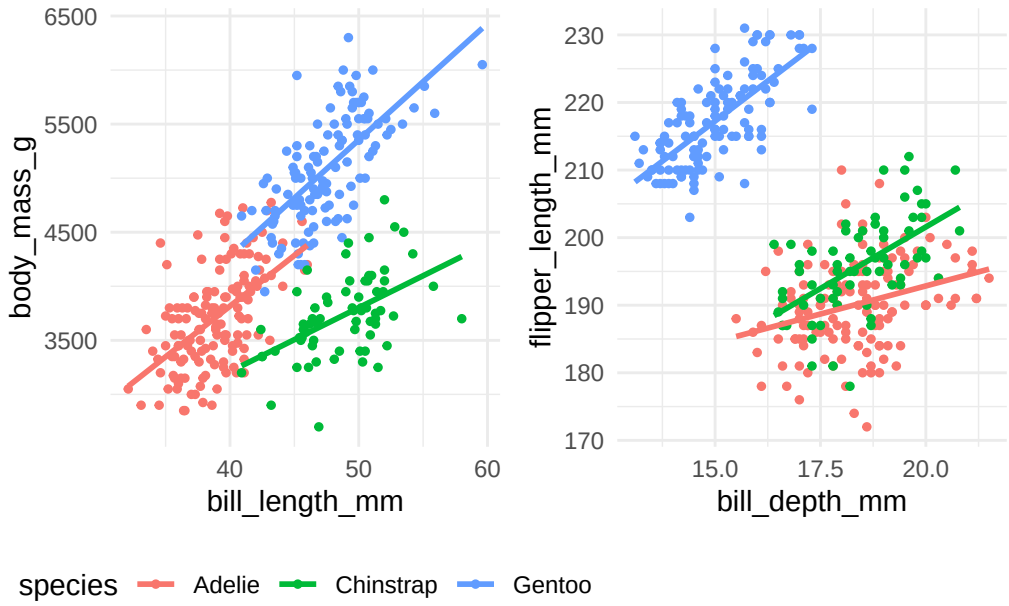
```
`geom_smooth()` using formula = 'y ~ x'
```

```
`geom_smooth()` using formula = 'y ~ x'
```



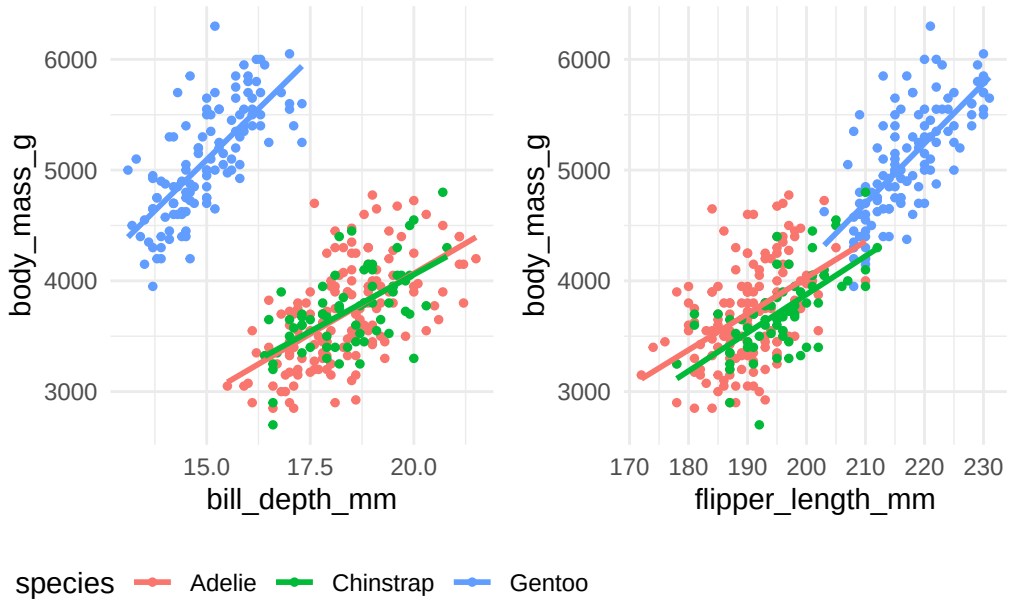

```
p3|p4
```

```
`geom_smooth()` using formula = 'y ~ x'
`geom_smooth()` using formula = 'y ~ x'
```




```
p5|p6
```

```
`geom_smooth()` using formula = 'y ~ x'
`geom_smooth()` using formula = 'y ~ x'
```


body_mass_g
6000
5000
4000
3000
15.0
17.5
20.0
bill_depth_mm
body_mass_g
6000
5000
4000
3000
170
180
190
200
210
220
230
flipper_length_mm
species
Adelie
Chinstrap
Gentoo