ggplot_hook_practice

Shaley

2023-11-01

Setting up my enviroment

Notes: I start to set up my enviroment by installing the 'tidyverse' and 'palmerpenguins' packages

```r
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```r
install.packages("palmerpenguins")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

next we load our installed packages using the 'library' function

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(palmerpenguins)
```

Now we get familiar with our data using the 'head' function

```r
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>
```

we can also familiriazie with our data using the `glimpse` or `colnames` functions

```
glimpse(penguins)
```

```
## Rows: 344
## Columns: 8
## $ species           <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adel~
## $ island            <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgerse~
## $ bill_length_mm    <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, ~
## $ bill_depth_mm     <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, ~
## $ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186~
## $ body_mass_g       <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, ~
## $ sex               <fct> male, female, female, NA, female, male, female, male~
## $ year              <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007~
```

```
colnames(penguins)
```

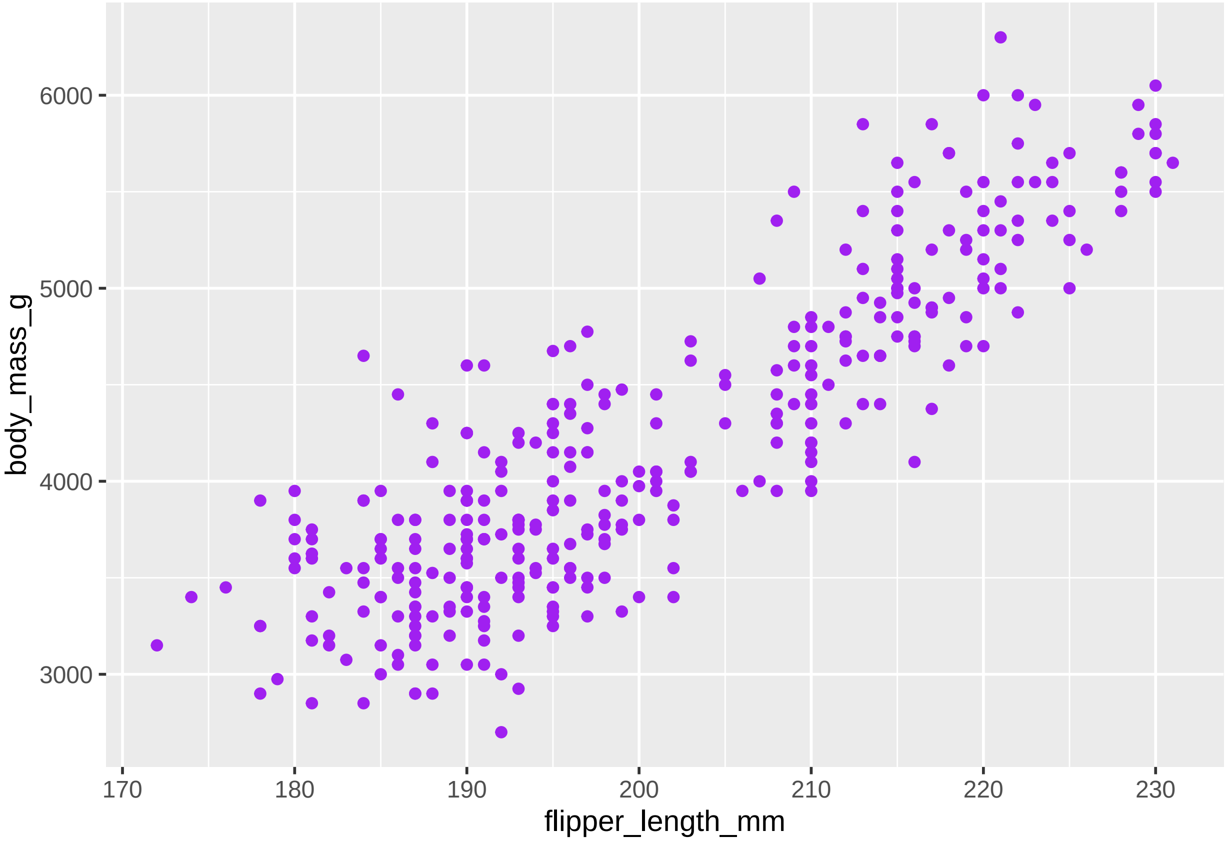
```
## [1] "species"           "island"            "bill_length_mm"
## [4] "bill_depth_mm"     "flipper_length_mm" "body_mass_g"
## [7] "sex"               "year"
```

Visualization

Here, we plot the flipper_length in mm against the body_mass in gram and give the color purple to our plot

```
ggplot(data=penguins,aes(x=flipper_length_mm,y=body_mass_g))+
  geom_point(color="purple")
```

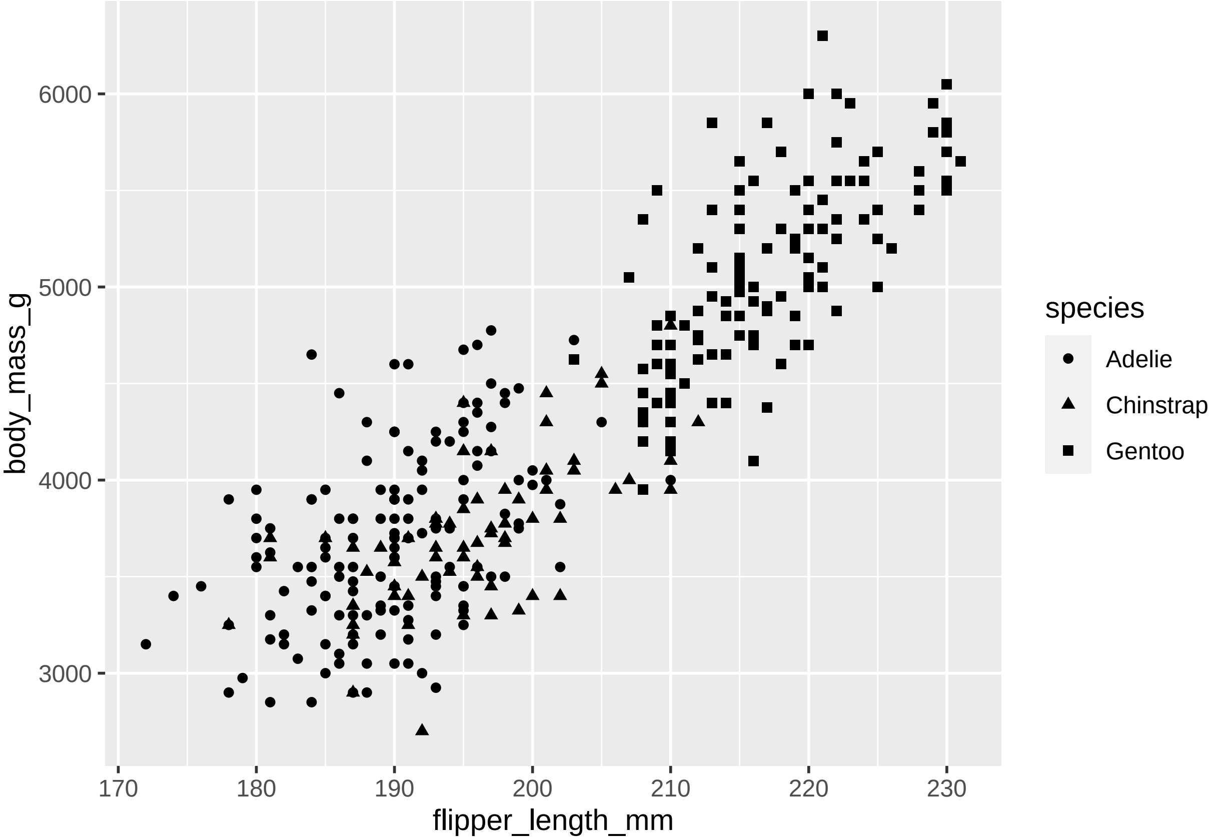
```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



Distinguish each species with shapes Here we write a code to distingusih each species with different shapes

```
ggplot(data=penguins,aes(x=flipper_length_mm,y=body_mass_g))+
  geom_point(aes(shape=species))
```

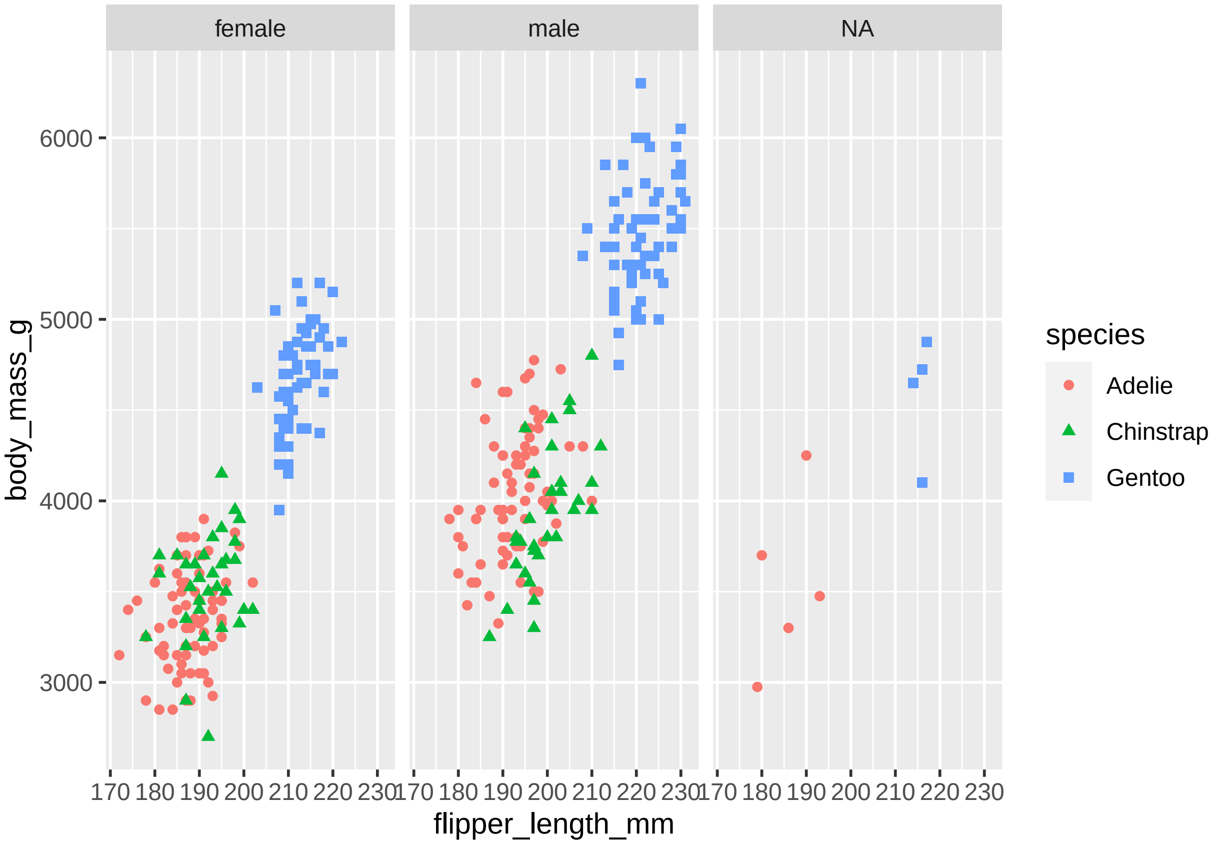
```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



Now let's plot different plots for the different species based on their sex using the `facet_wrap` function

```
ggplot(data=penguins,aes(x=flipper_length_mm,y=body_mass_g))+
  geom_point(aes(color=species,
                 shape=species)) +
  facet_wrap(~sex)
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



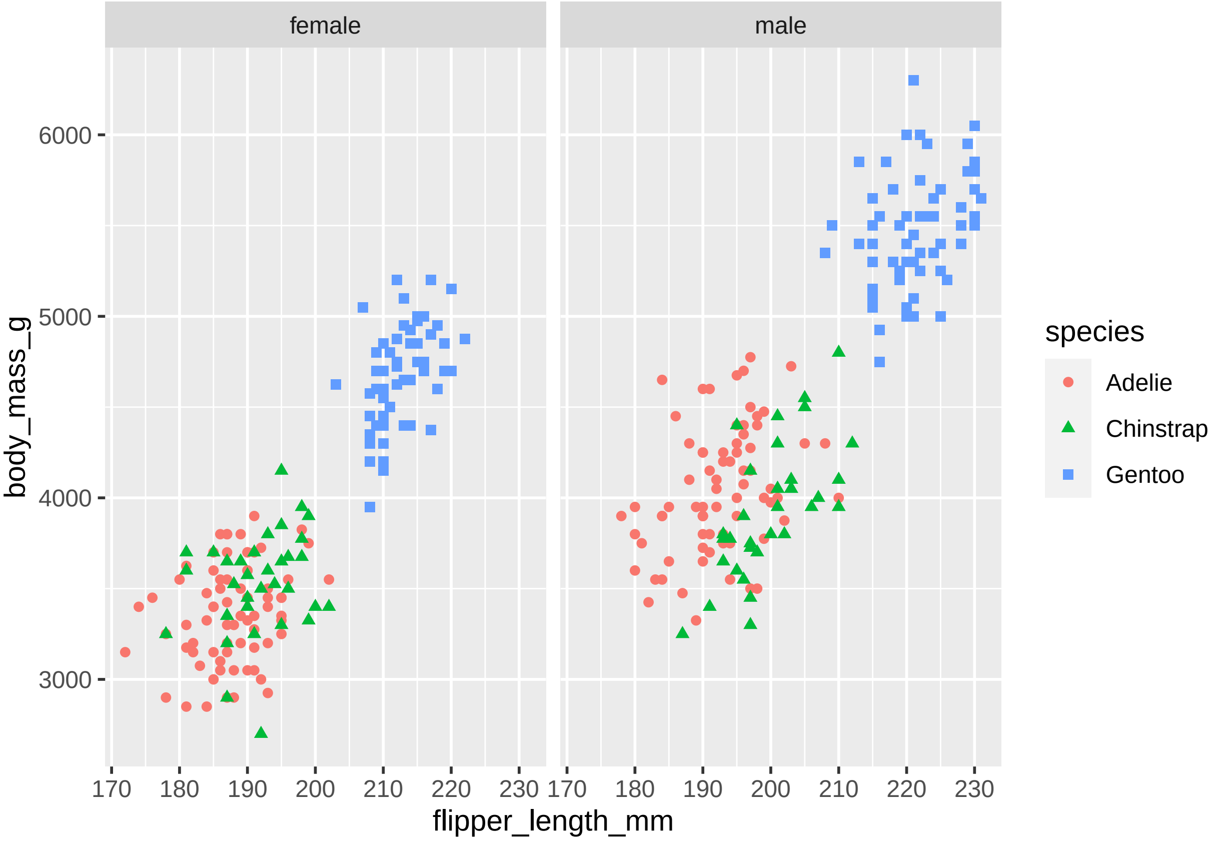
if u noticed in the above visuals that missing values were included in our visuals, let's drop the missing values using the `drop_na` function

```
penguins %>%   drop_na(sex) %>%   ggplot(data=penguins,aes(x=flipper_length_mm,y=body_mass_g))+
geom_point(aes(color=species,                    shape=species)) +   facet_wrap(~sex)
```

Note! the above code will return an error msg, this is because we already used the **pipe** function(%>%) to call the dataset and then we are calling the dataset again inside the ggplot function.

To correct this we remove the `data=penguins` and run the code again

```
penguins %>%
  drop_na(sex) %>%
  ggplot(aes(x=flipper_length_mm,y=body_mass_g))+
  geom_point(aes(color=species,
                 shape=species)) +
  facet_wrap(~sex)
```


female
male
body_mass_g
6000
5000
4000
3000
170
180
190
200
210
220
230
170
180
190
200
210
220
230
flipper_length_mm
species
Adelie
Chinstrap
Gentoo