

penguin Classifier

```
library(tidyverse)
theme_set(theme_bw())
```

Data

```
library(palmerpenguins)
penguins
```

```
# A tibble: 344 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
7 Adelie Torgersen     38.9           17.8            181          3625
8 Adelie Torgersen     39.2           19.6            195          4675
9 Adelie Torgersen     34.1           18.1            193          3475
10 Adelie Torgersen     42            20.2            190          4250
# i 334 more rows
# i 2 more variables: sex <fct>, year <int>
```

Explore Data

```
penguins |> count(species)
```

```
# A tibble: 3 x 2
  species      n
  <fct>    <int>
1 Adelie  222
2 Adelie   67
3 Adelie   15
```

```

      <fct>      <int>
1 Adelie        152
2 Chinstrap     68
3 Gentoo        124

```

```
penguins |> count(island)
```

```

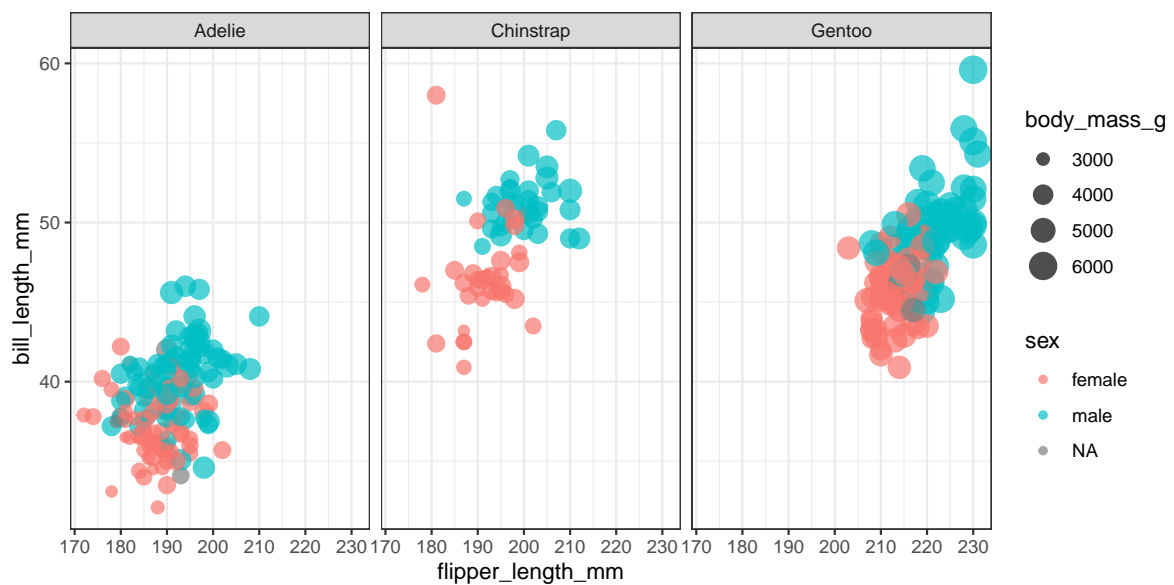
# A tibble: 3 x 2
  island      n
  <fct>    <int>
1 Biscoe   168
2 Dream    124
3 Torgersen 52

```

```

penguins |>
  ggplot(aes(flipper_length_mm, bill_length_mm, color=sex, size=body_mass_g))+
  geom_point(alpha=0.7)+
  facet_wrap(~species)

```



```

penguins_df <- penguins %>%
  filter(!is.na(sex)) %>%
  select(-year, -island)
penguins_df

```

```

# A tibble: 333 x 6
  species bill_length_mm bill_depth_mm flipper_length_mm body_mass_g sex

```

| | <fct> | <dbl> | <dbl> | <int> | <int> | <fct> |
|----|--------|-------|-------|-------|-------|--------|
| 1 | Adelie | 39.1 | 18.7 | 181 | 3750 | male |
| 2 | Adelie | 39.5 | 17.4 | 186 | 3800 | female |
| 3 | Adelie | 40.3 | 18 | 195 | 3250 | female |
| 4 | Adelie | 36.7 | 19.3 | 193 | 3450 | female |
| 5 | Adelie | 39.3 | 20.6 | 190 | 3650 | male |
| 6 | Adelie | 38.9 | 17.8 | 181 | 3625 | female |
| 7 | Adelie | 39.2 | 19.6 | 195 | 4675 | male |
| 8 | Adelie | 41.1 | 17.6 | 182 | 3200 | female |
| 9 | Adelie | 38.6 | 21.2 | 191 | 3800 | male |
| 10 | Adelie | 34.6 | 21.1 | 198 | 4400 | male |

i 323 more rows

Build a model

```
library(tidymodels)

set.seed(123)
penguin_split <- initial_split(penguins_df, strata=sex)
penguin_train <- training(penguin_split)
penguin_test <- testing(penguin_split)

set.seed(234)
penguin_boot <- bootstraps(penguin_train)
penguin_boot
```

```
# Bootstrap sampling
# A tibble: 25 x 2
  splits          id
  <list>         <chr>
1 <split [249/92]> Bootstrap01
2 <split [249/99]> Bootstrap02
3 <split [249/94]> Bootstrap03
4 <split [249/97]> Bootstrap04
5 <split [249/93]> Bootstrap05
6 <split [249/90]> Bootstrap06
7 <split [249/86]> Bootstrap07
8 <split [249/92]> Bootstrap08
9 <split [249/90]> Bootstrap09
10 <split [249/89]> Bootstrap10
# i 15 more rows
```