R Notebook

packges

```
pacman::p_load(tidyverse,
               gapminder,
               palmerpenguins,
               janitor,
               gtsummary,
               visdat)
```

Recapitulation

EDA

```
head(penguins)
```

```
## # A tibble: 6 x 8
##   species island bill_length_mm bill_depth_mm flipper_length_~ body_mass_g sex
##   <fct>   <fct>           <dbl>         <dbl>            <int>       <int> <fct>
## 1 Adelie  Torge~           39.1          18.7              181        3750 male
## 2 Adelie  Torge~           39.5          17.4              186        3800 fema~
## 3 Adelie  Torge~           40.3          18                195        3250 fema~
## 4 Adelie  Torge~           NA            NA                 NA          NA <NA>
## 5 Adelie  Torge~           36.7          19.3              193        3450 fema~
## 6 Adelie  Torge~           39.3          20.6              190        3650 male
## # ... with 1 more variable: year <int>
```

```
dim(penguins)
```

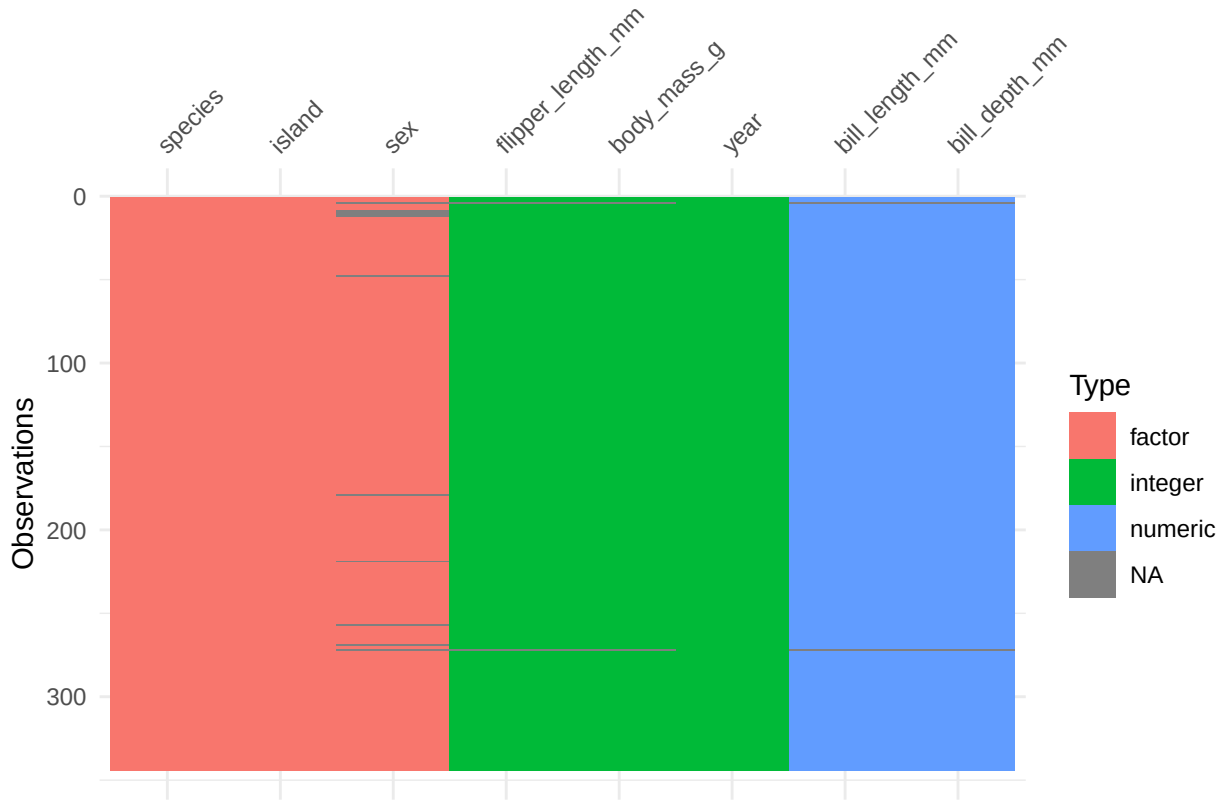
```
## [1] 344   8
```

```
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
```

```
visdat::vis_dat(penguins)
```

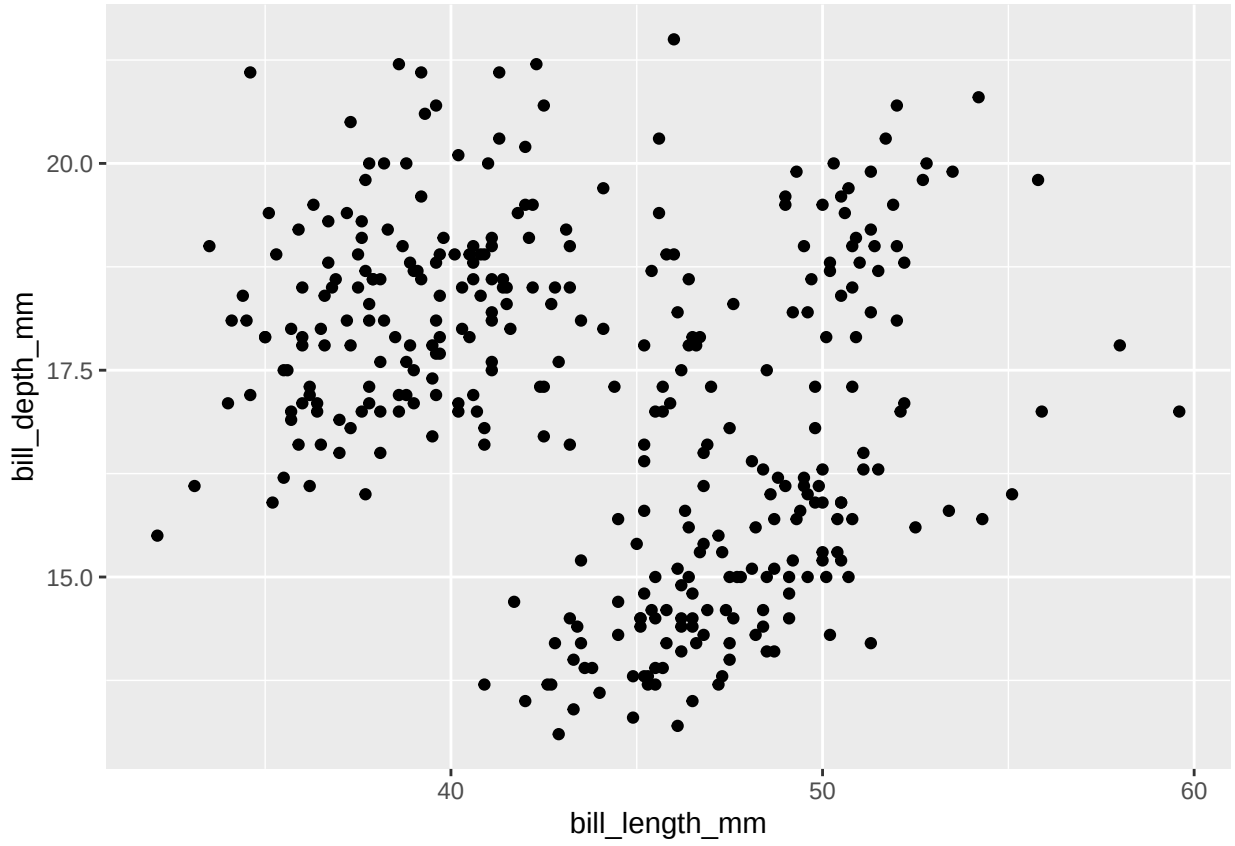


Data visualizations

Points bill_length_mm vs bill_depth_mm

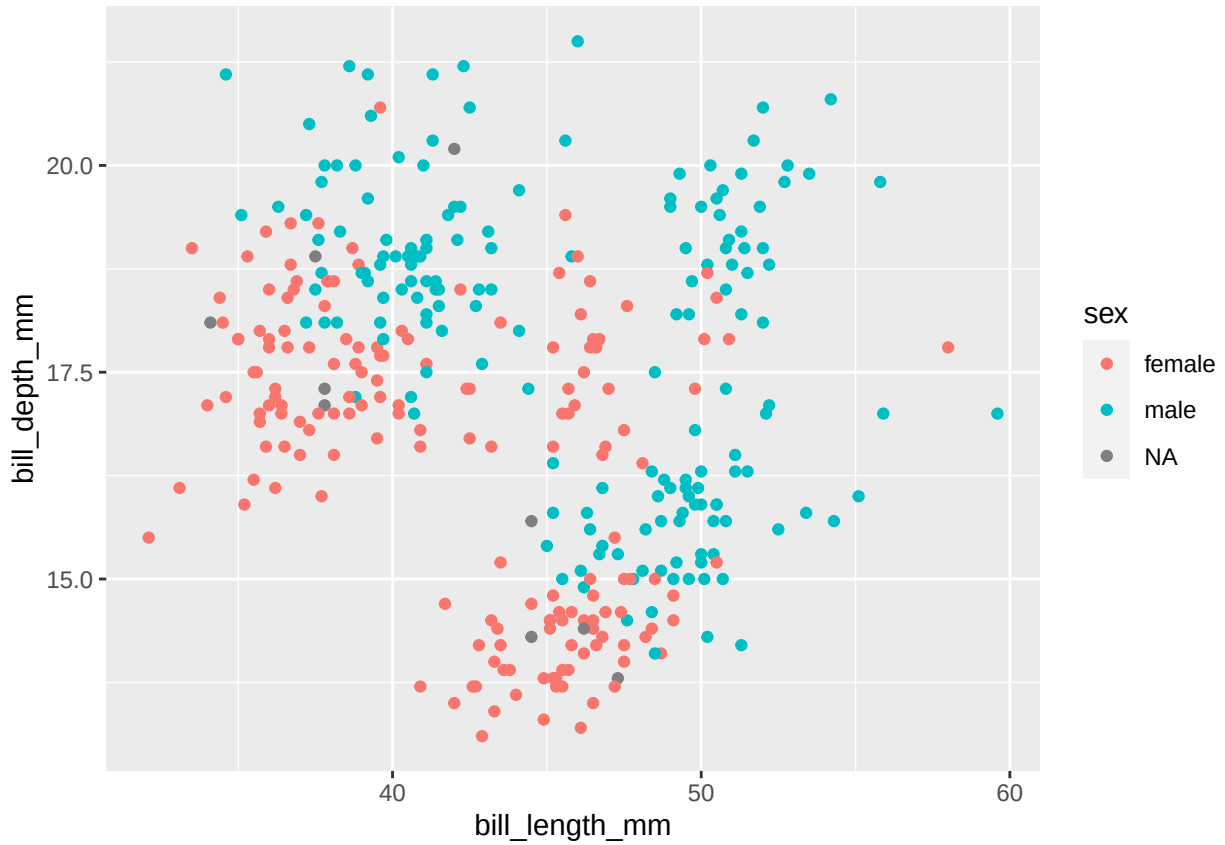
```
penguins %>%
  ggplot(aes(x = bill_length_mm,
             y = bill_depth_mm)) +
  geom_point()
```

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

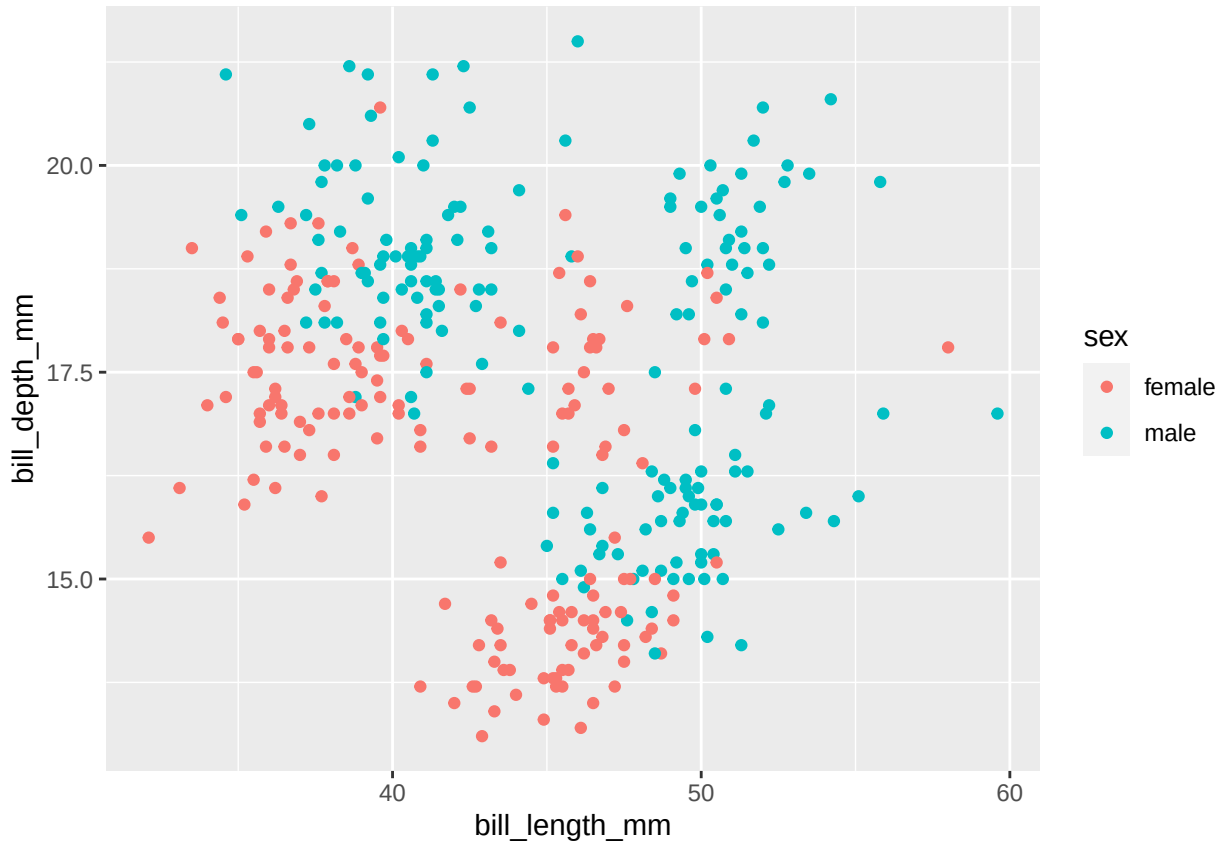



```
penguins %>%
  ggplot(aes(x = bill_length_mm,
             y = bill_depth_mm)) +
  geom_point() +
  aes(color = sex)
```

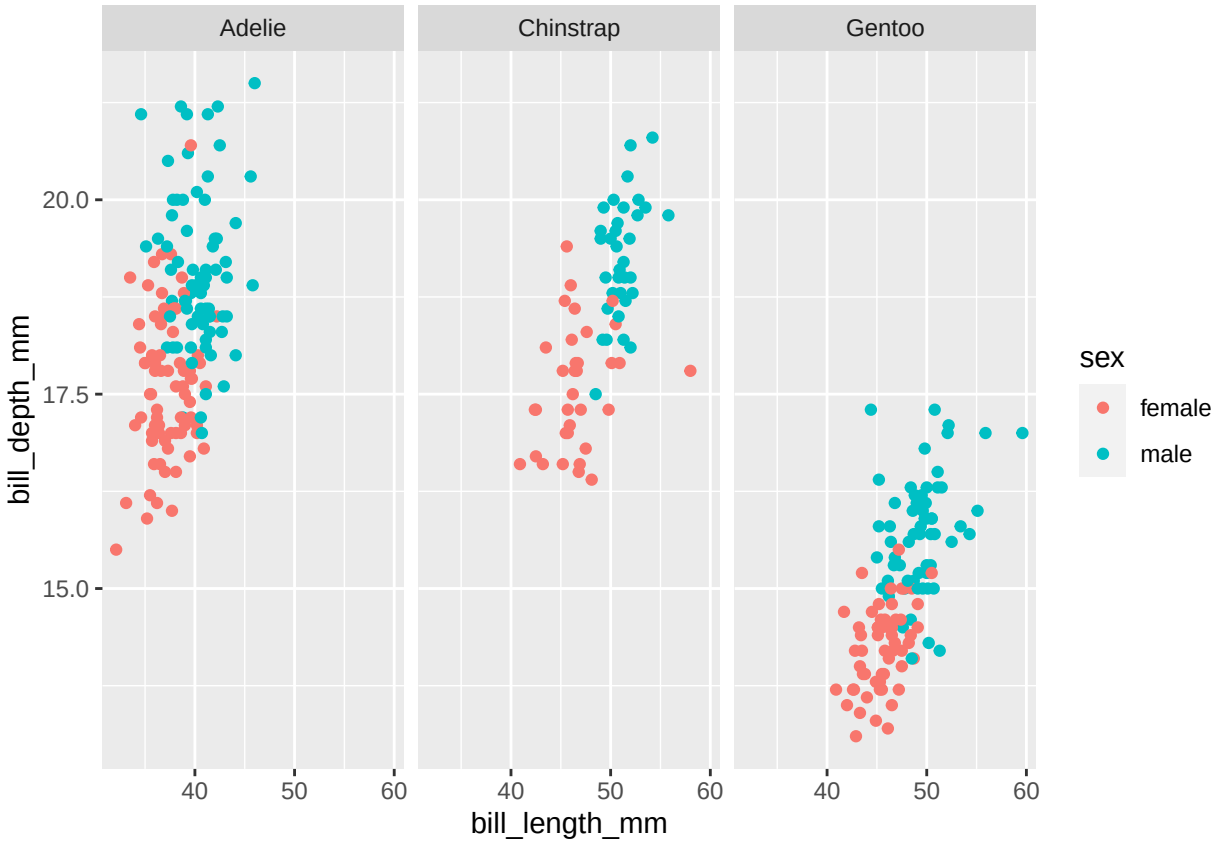
```
## Warning: Removed 2 rows containing missing values (geom_point).
```




```
penguins %>% # my date
  drop_na() %>%  # i will dropn the NA values
  ggplot(aes(x = bill_length_mm,
             y = bill_depth_mm)) +
  geom_point() +
  aes(color = sex)
```

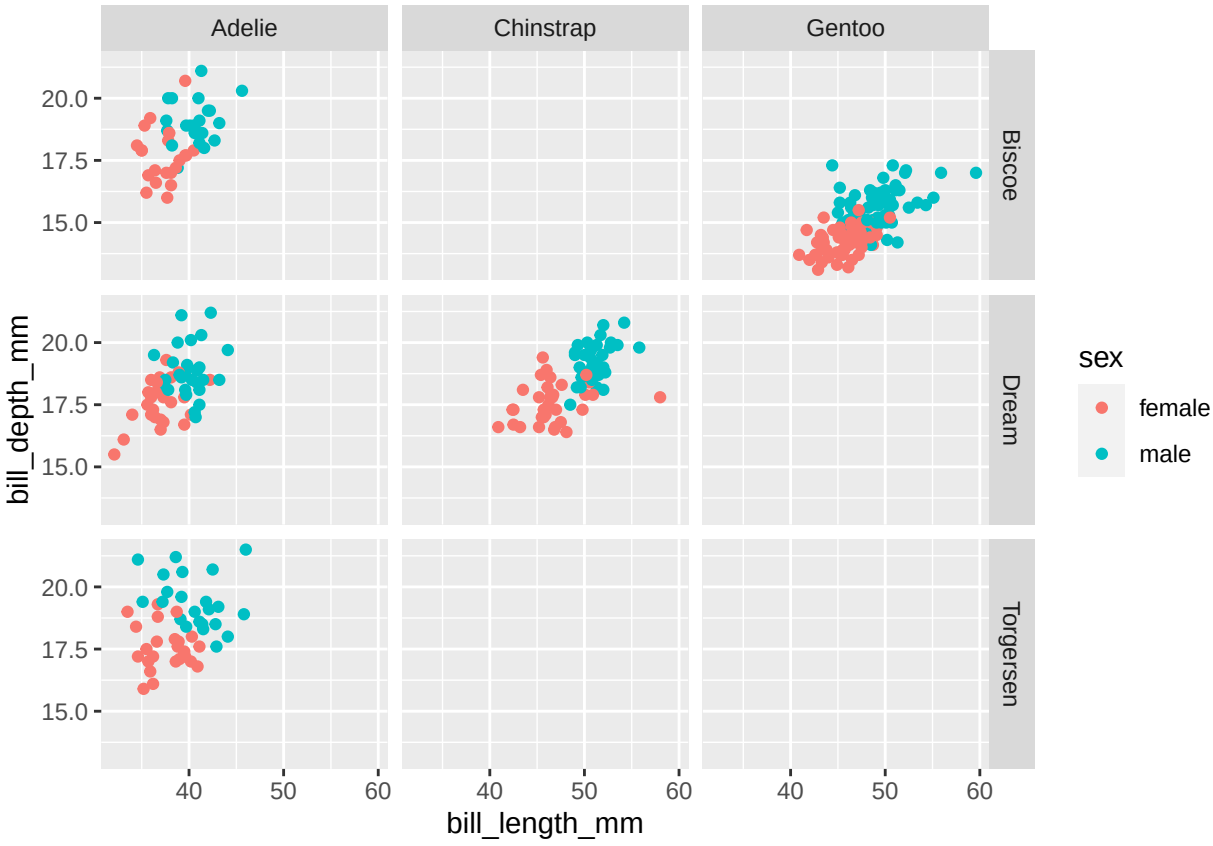



```
penguins %>% # my date
  drop_na() %>%  # i will dropn the NA values
  ggplot(aes(x = bill_length_mm,
             y = bill_depth_mm)) +
  geom_point() +
  aes(color = sex) +
  facet_wrap(~species)
```



Facetting

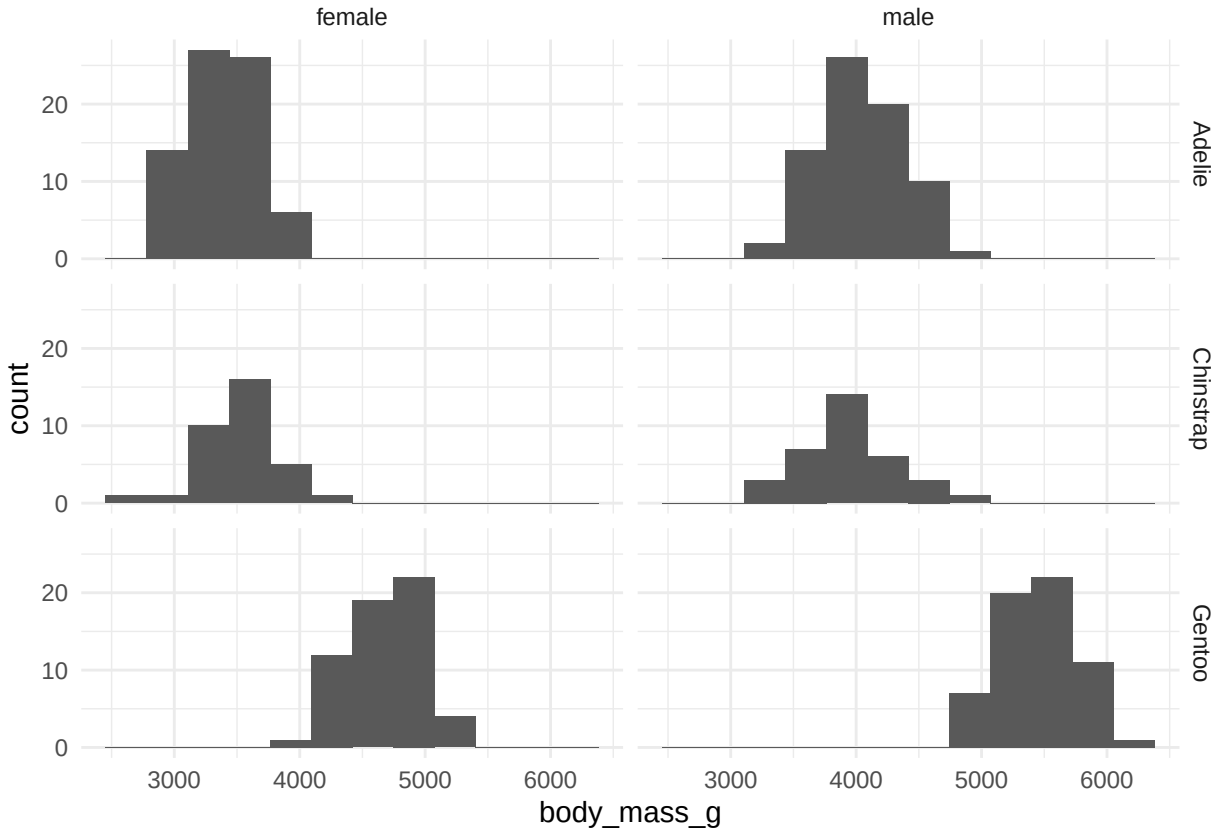
```
penguins %>% # my date
  drop_na() %>%  # i will drop the NA values
  ggplot(aes(x = bill_length_mm,
             y = bill_depth_mm)) +
  geom_point() +
  aes(color = sex) +
  facet_grid(island ~ species)
```



One variable continuous

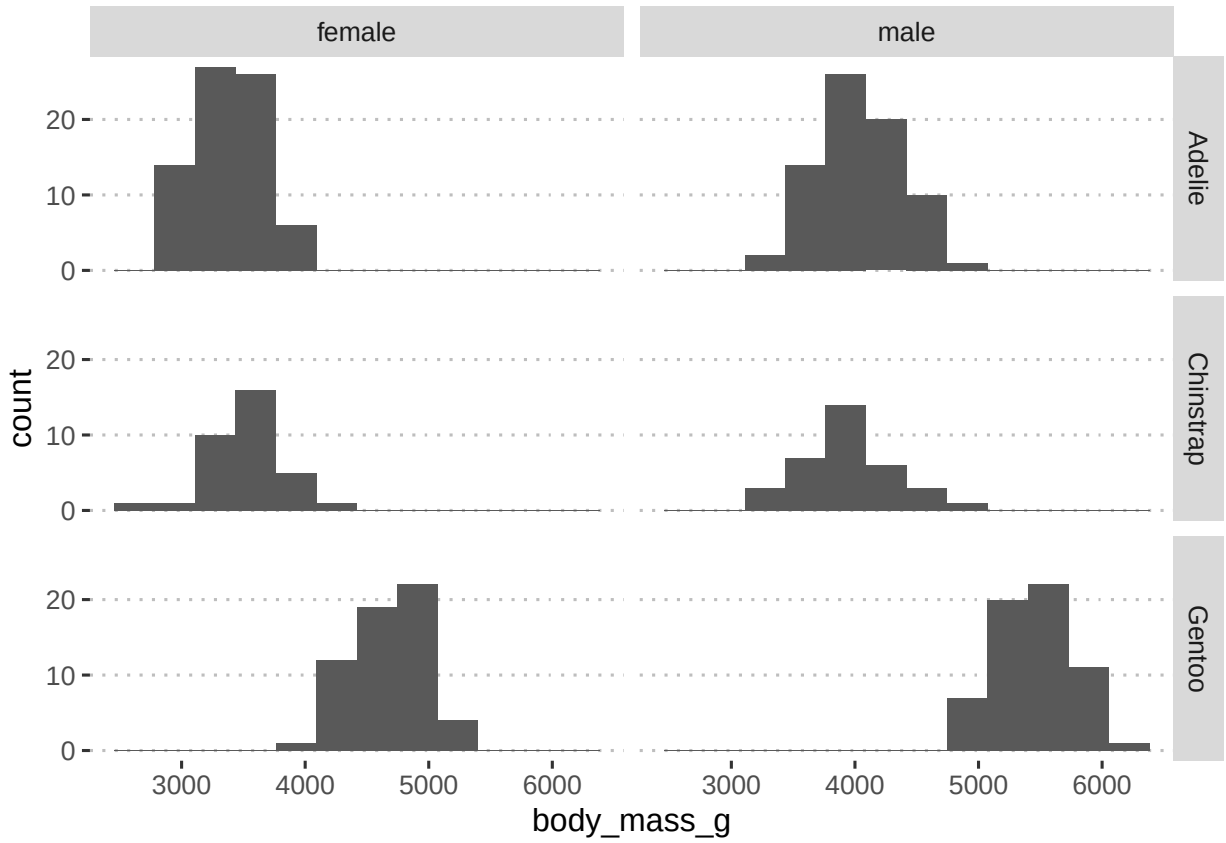
make an histogram with body_mass_g

```
penguins %>%
  drop_na() %>%
  ggplot(aes(x = body_mass_g)) +
  geom_histogram(bins = 12) +
  facet_grid(species ~ sex) +
  theme_minimal()
```



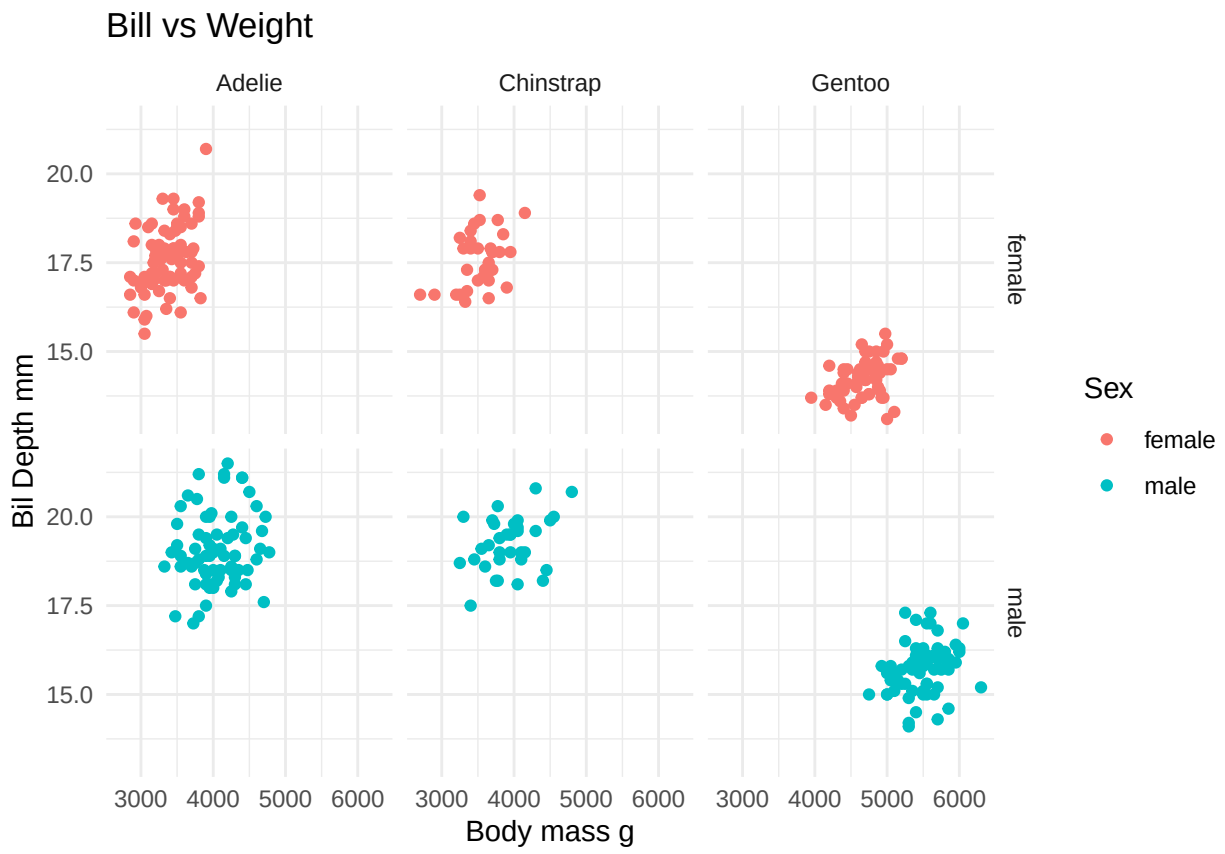

```
pacman::p_load(ggthemes)
```

```
penguins %>%
  drop_na() %>%
  ggplot(aes(x = body_mass_g)) +
  geom_histogram(bins = 12) +
  facet_grid(species ~ sex) +
  ggpubr::theme_pubclean()
```



Two variables continuous

```
penguins %>%
  drop_na() %>%
  ggplot(aes(x = body_mass_g,
             y = bill_depth_mm)) +
  geom_point() +
  aes(color = sex) +
  facet_grid(sex ~ species) +
  theme_minimal() +
  labs(
    title = "Bill vs Weight",
    y = "Bil Depth mm",
    x = "Body mass g",
    color = "Sex"
  )
```



Explore correlations

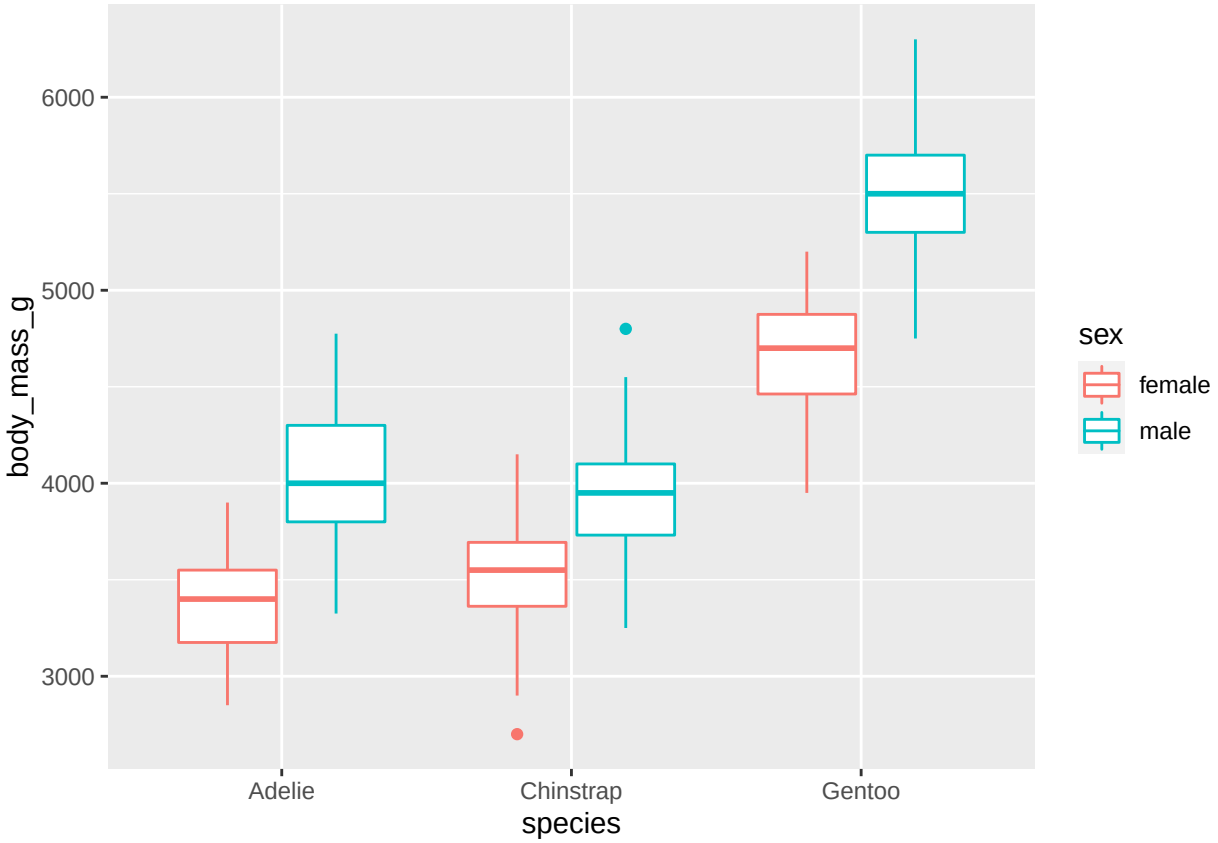
```
penguins %>%
  drop_na() %>%
  ggplot(aes(x = body_mass_g,
             y = bill_depth_mm)) +
  geom_point() +
  aes(color = sex) +
  theme_minimal() +
  labs(
    title = "Bill vs Weight",
    y = "Bil Depth mm",
    x = "Body mass g",
    color = "Sex"
  ) +
  geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Two variables continuous vs nominal

body_mass_g by species

```
penguins %>%
  drop_na() %>%
  ggplot(aes(x = species,
             y = body_mass_g)) +
  geom_boxplot() +
  aes(color = sex)
```



Trends

```
gapminder
```

```
## # A tibble: 1,704 x 6
##    country     continent  year lifeExp      pop gdpPercap
##    <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
##  1 Afghanistan Asia       1952    28.8  8425333      779.
##  2 Afghanistan Asia       1957    30.3  9240934      821.
##  3 Afghanistan Asia       1962    32.0 10267083      853.
##  4 Afghanistan Asia       1967    34.0 11537966      836.
##  5 Afghanistan Asia       1972    36.1 13079460      740.
##  6 Afghanistan Asia       1977    38.4 14880372      786.
##  7 Afghanistan Asia       1982    39.9 12881816      978.
##  8 Afghanistan Asia       1987    40.8 13867957      852.
##  9 Afghanistan Asia       1992    41.7 16317921      649.
## 10 Afghanistan Asia       1997    41.8 22227415      635.
## # ... with 1,694 more rows
```

```
gapminder <- gapminder::gapminder
```