

Palmerpenguins Report

F24116097 呂采蓁

2024-09-17

Table of contents

Statistical Thinking	1
Summary Staistic	2
Missing Values	4
organize data	4
Table	6
scatter plot	8
box plot	10
bar plot	14

Statistical Thinking

The dataset contains information about penguin species, their physical attributes, and environmental data.

Reference:<https://allisonhorst.github.io/palmerpenguins/>

Summary Staistic

```
library(Hmisc)
```

Warning: package 'Hmisc' was built under R version 4.3.2

Attaching package: 'Hmisc'

The following objects are masked from 'package:base':

format.pval, units

```
library(palmerpenguins)
```

```
str(penguins_raw)
```

```
tibble [344 x 17] (S3: tbl_df/tbl/data.frame)
 $ studyName      : chr [1:344] "PAL0708" "PAL0708" "PAL0708" "PAL0708" ...
 $ Sample Number  : num [1:344] 1 2 3 4 5 6 7 8 9 10 ...
 $ Species        : chr [1:344] "Adelie Penguin (Pygoscelis adeliae)" "Adelie Penguin (Pygoscelis adeliae)" ...
 $ Region         : chr [1:344] "Anvers" "Anvers" "Anvers" "Anvers" ...
 $ Island         : chr [1:344] "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...
 $ Stage          : chr [1:344] "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" ...
 $ Individual ID   : chr [1:344] "N1A1" "N1A2" "N2A1" "N2A2" ...
 $ Clutch Completion : chr [1:344] "Yes" "Yes" "Yes" "Yes" ...
 $ Date Egg       : Date[1:344], format: "2007-11-11" "2007-11-11" ...
 $ Culmen Length (mm) : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
 $ Culmen Depth (mm) : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
 $ Flipper Length (mm): num [1:344] 181 186 195 NA 193 190 181 195 193 190 ...
 $ Body Mass (g)     : num [1:344] 3750 3800 3250 NA 3450 ...
 $ Sex             : chr [1:344] "MALE" "FEMALE" "FEMALE" NA ...
 $ Delta 15 N (o/oo) : num [1:344] NA 8.95 8.37 NA 8.77 ...
 $ Delta 13 C (o/oo) : num [1:344] NA -24.7 -25.3 NA -25.3 ...
 $ Comments         : chr [1:344] "Not enough blood for isotopes." NA NA "Adult not sampled" ...
- attr(*, "spec")=List of 3
 ..$ cols      :List of 17
 .. ..$ studyName      : list()
 .. ..$- attr(*, "class")= chr [1:2] "collector_character" "collector"
```

```

.. ..$ Sample Number      : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Species            : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Region             : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Island             : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Stage              : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Individual ID      : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Clutch Completion  : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Date Egg           :List of 1
.. .. ..$ format: chr ""
.. .. ..- attr(*, "class")= chr [1:2] "collector_date" "collector"
.. ..$ Culmen Length (mm) : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Culmen Depth (mm) : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Flipper Length (mm): list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Body Mass (g)      : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Sex                : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
.. ..$ Delta 15 N (o/oo)  : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Delta 13 C (o/oo)  : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
.. ..$ Comments           : list()
.. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
..$ default: list()
.. ..- attr(*, "class")= chr [1:2] "collector_guess" "collector"
..$ skip      : num 1
..- attr(*, "class")= chr "col_spec"

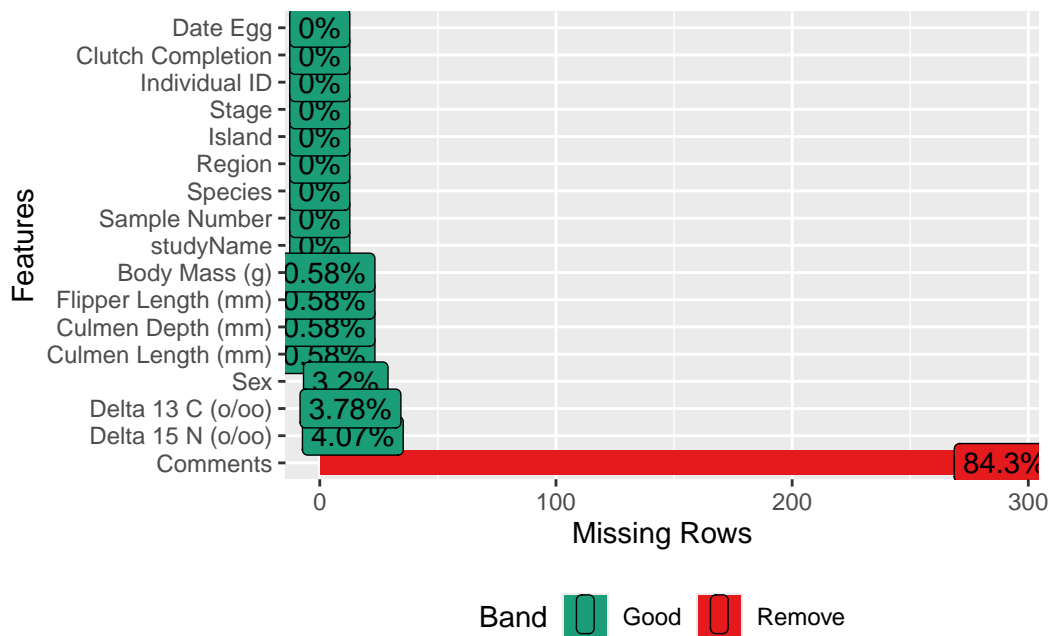
```

Missing Values

```
library(DataExplorer)
```

Warning: package 'DataExplorer' was built under R version 4.3.3

```
plot_missing(penguins_raw)
```



organize data

```
library(dplyr)
```

Warning: package 'dplyr' was built under R version 4.3.2

Attaching package: 'dplyr'

The following objects are masked from 'package:Hmisc':

src, summarize

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
data <- penguins_raw %>%
  rename(
    sample_number = `Sample Number`,
    species = `Species`,
    region = `Region`,
    island = `Island`,
    stage = `Stage`,
    individual_id = `Individual ID`,
    clutch_completion = `Clutch Completion`,
    date_egg = `Date Egg`,
    culmen_length_mm = `Culmen Length (mm)`,
    culmen_depth_mm = `Culmen Depth (mm)`,
    flipper_length_mm = `Flipper Length (mm)`,
    body_mass_g = `Body Mass (g)`,
    sex = `Sex`,
    delta_15_N = `Delta 15 N (o/oo)`,
    delta_13_C = `Delta 13 C (o/oo)`,
    comments = `Comments`
  )
```

```
summary(data)
```

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	delta_13_C
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

Table

```
library(tidyverse)
```

Warning: package 'tidyverse' was built under R version 4.3.3

Warning: package 'ggplot2' was built under R version 4.3.3

Warning: package 'tibble' was built under R version 4.3.2

Warning: package 'tidyr' was built under R version 4.3.2

Warning: package 'readr' was built under R version 4.3.3

Warning: package 'purrr' was built under R version 4.3.2

Warning: package 'stringr' was built under R version 4.3.2

Warning: package 'forcats' was built under R version 4.3.3

Warning: package 'lubridate' was built under R version 4.3.3

-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --

v forcats 1.0.0 v readr 2.1.5

v ggplot2 3.5.1 v stringr 1.5.1

v lubridate 1.9.3 v tibble 3.2.1

v purrr 1.0.2 v tidyr 1.3.0

-- Conflicts ----- tidyverse_conflicts() --

x dplyr::filter() masks stats::filter()

x dplyr::lag() masks stats::lag()

x dplyr::src() masks Hmisc::src()

x dplyr::summarize() masks Hmisc::summarize()

i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become

```
penguins_raw %>%  
  count(Species)
```

A tibble: 3 x 2

Species	n
<chr>	<int>
1 Adelie Penguin (Pygoscelis adeliae)	152
2 Chinstrap penguin (Pygoscelis antarctica)	68
3 Gentoo penguin (Pygoscelis papua)	124

```
penguins_raw %>%
  group_by(Species) %>%
  summarize(across(where(is.numeric), mean, na.rm = TRUE))
```

Warning: There was 1 warning in `summarize()`.

- i In argument: `across(where(is.numeric), mean, na.rm = TRUE)`.
- i In group 1: `Species = "Adelie Penguin (Pygoscelis adeliae)"`.

Caused by warning:

- ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0. Supply arguments directly to `.fns` through an anonymous function instead.

```
# Previously
across(a:b, mean, na.rm = TRUE)

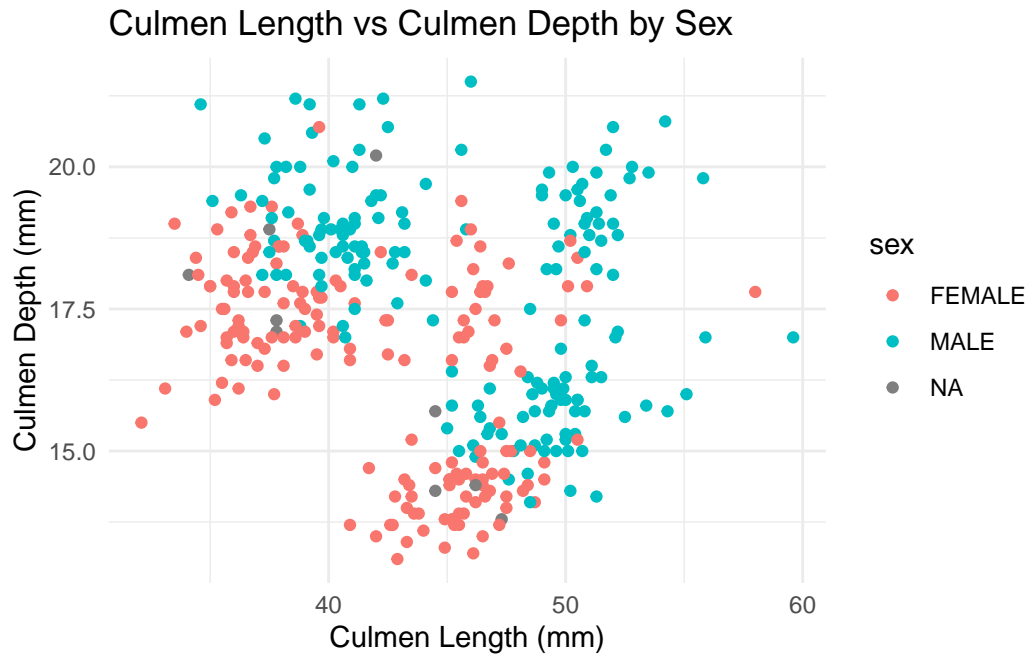
# Now
across(a:b, \(x) mean(x, na.rm = TRUE))

# A tibble: 3 x 8
  Species          `Sample Number` `Culmen Length (mm)` `Culmen Depth (mm)`
  <chr>              <dbl>          <dbl>          <dbl>
1 Adelie Penguin (Pygo~      76.5            38.8            18.3
2 Chinstrap penguin (P~      34.5            48.8            18.4
3 Gentoo penguin (Pygo~      62.5            47.5            15.0
# i 4 more variables: `Flipper Length (mm)` <dbl>, `Body Mass (g)` <dbl>,
#   `Delta 15 N (o/oo)` <dbl>, `Delta 13 C (o/oo)` <dbl>
```

scatter plot

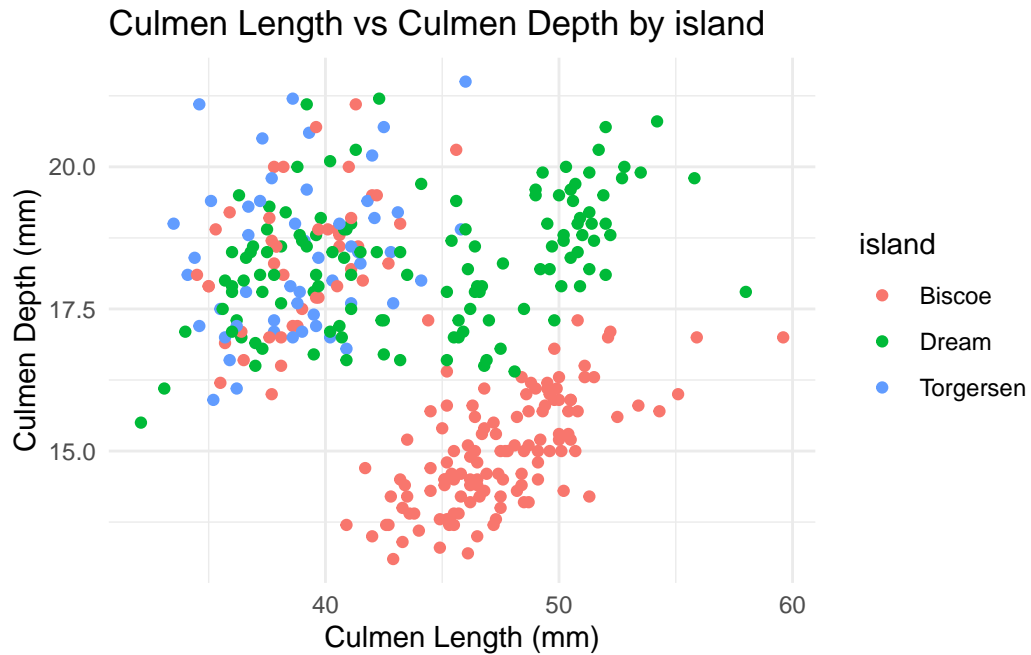
```
ggplot(data, aes(x=culmen_length_mm, y=culmen_depth_mm, color=sex)) +
  geom_point() +
  labs(title="Culmen Length vs Culmen Depth by Sex", x="Culmen Length (mm)", y="Culmen Depth (mm)")
  theme_minimal()
```

Warning: Removed 2 rows containing missing values or values outside the scale range (`geom_point()`).



```
ggplot(data, aes(x=culmen_length_mm, y=culmen_depth_mm, color=island)) +  
  geom_point() +  
  labs(title="Culmen Length vs Culmen Depth by island", x="Culmen Length (mm)", y="Culmen Depth (mm)") +  
  theme_minimal()
```

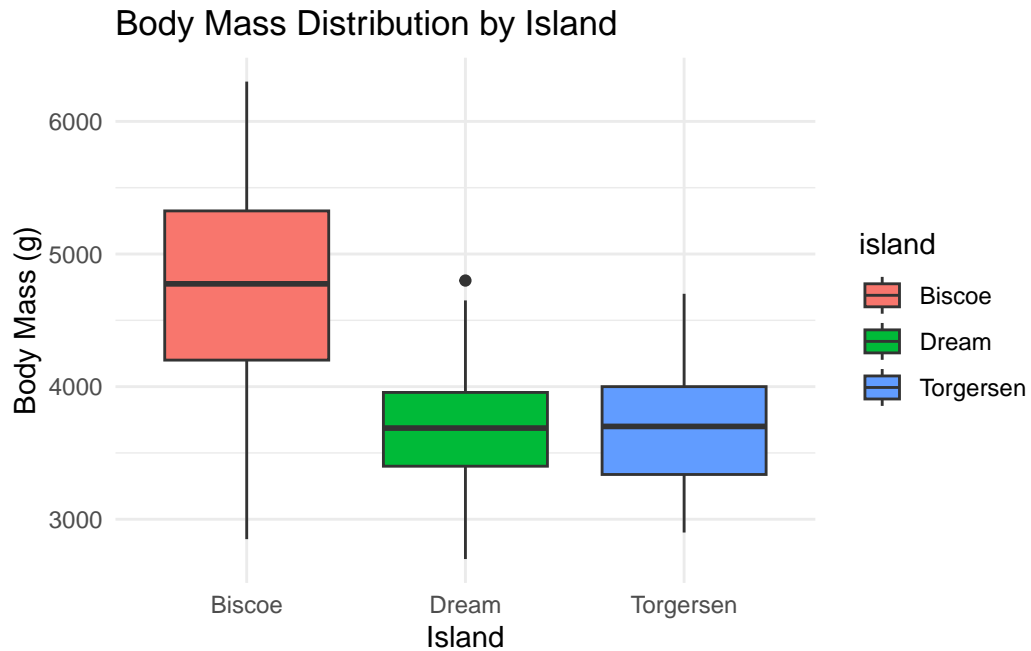
Warning: Removed 2 rows containing missing values or values outside the scale range (``geom_point()``).



box plot

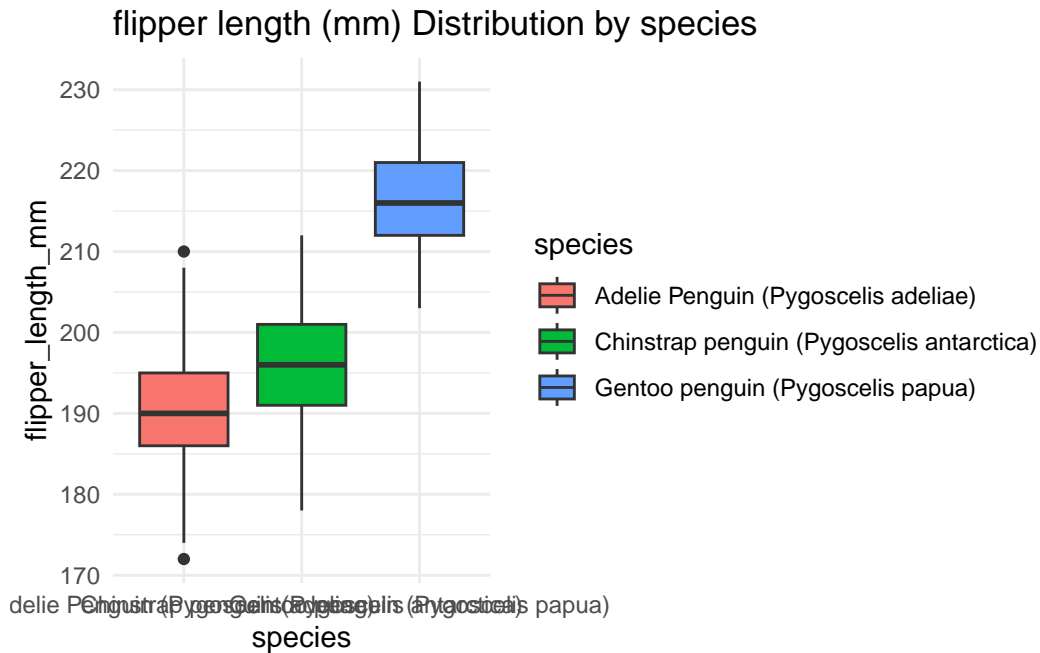
```
library(ggplot2)
ggplot(data, aes(x=island, y=body_mass_g, fill=island)) +
  geom_boxplot() +
  labs(title="Body Mass Distribution by Island", x="Island", y="Body Mass (g)") +
  theme_minimal()
```

Warning: Removed 2 rows containing non-finite outside the scale range (``stat_boxplot()``).



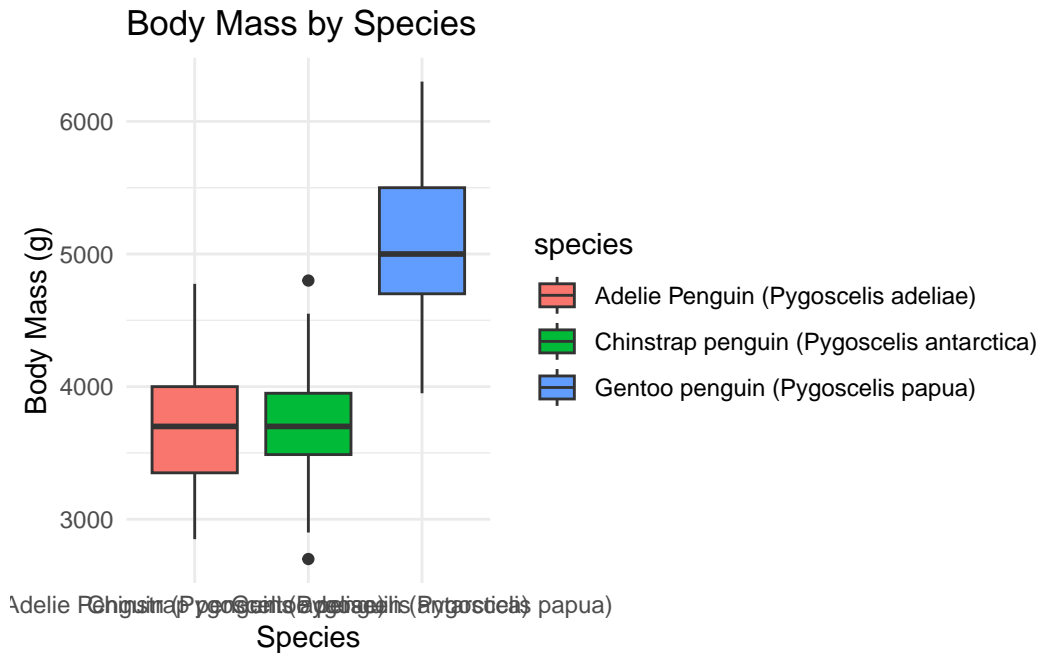
```
ggplot(data, aes(x=species, y=flipper_length_mm, fill=species)) +  
  geom_boxplot() +  
  labs(title="flipper length (mm) Distribution by species", x="species", y="flipper_length_mm") +  
  theme_minimal()
```

Warning: Removed 2 rows containing non-finite outside the scale range (``stat_boxplot()``).



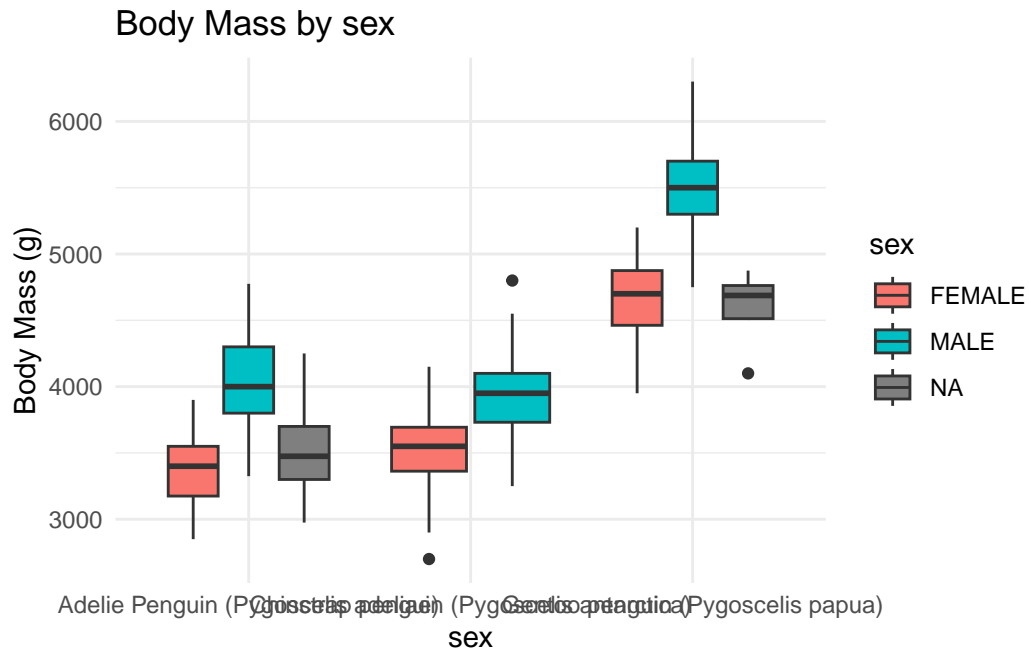
```
ggplot(data, aes(x=species, y=body_mass_g, fill=species)) +
  geom_boxplot() +
  labs(title="Body Mass by Species", x="Species", y="Body Mass (g)") +
  theme_minimal()
```

Warning: Removed 2 rows containing non-finite outside the scale range (`stat_boxplot()`).



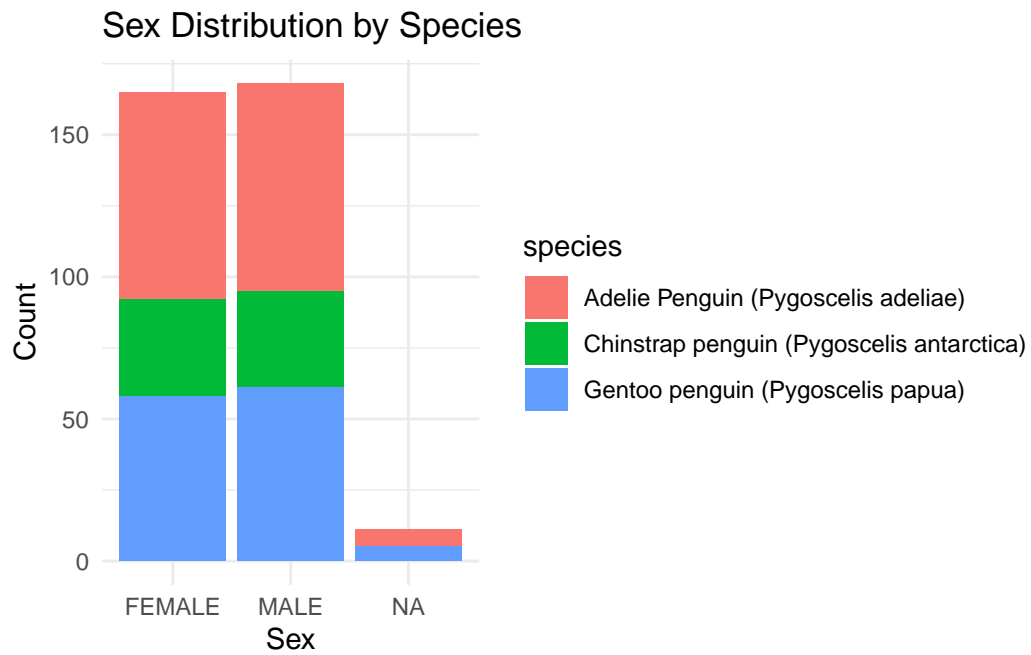
```
ggplot(data, aes(x=species, y=body_mass_g, fill=sex)) +
  geom_boxplot() +
  labs(title="Body Mass by sex", x="sex", y="Body Mass (g)") +
  theme_minimal()
```

Warning: Removed 2 rows containing non-finite outside the scale range (`stat_boxplot()`).

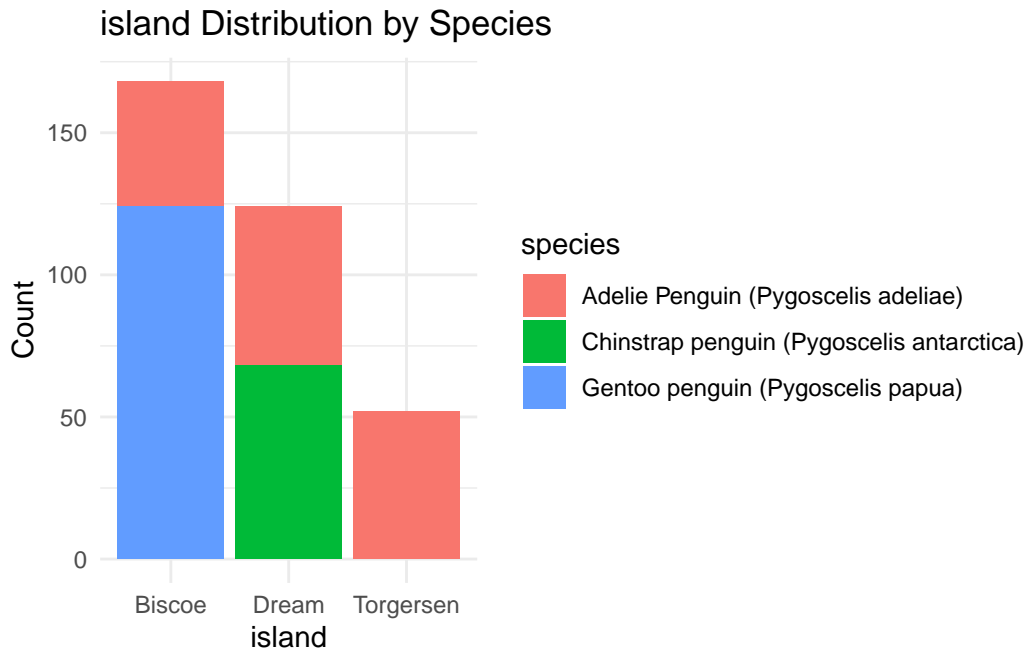


bar plot

```
ggplot(data, aes(x=sex, fill=species)) +
  geom_bar() +
  labs(title="Sex Distribution by Species", x="Sex", y="Count") +
  theme_minimal()
```



```
ggplot(data, aes(x=island, fill=species)) +  
  geom_bar() +  
  labs(title="island Distribution by Species", x="island", y="Count") +  
  theme_minimal()
```



```
citation("palmerpenguins")
```

To cite palmerpenguins in publications use:

Horst AM, Hill AP, Gorman KB (2020). palmerpenguins: Palmer Archipelago (Antarctica) penguin data. R package version 0.1.0. <https://allisonhorst.github.io/palmerpenguins/>. doi: 10.5281/zenodo.3960218.

A BibTeX entry for LaTeX users is

```
@Manual{,
  title = {palmerpenguins: Palmer Archipelago (Antarctica) penguin data},
  author = {Allison Marie Horst and Alison Presmanes Hill and Kristen B Gorman},
  year = {2020},
  note = {R package version 0.1.0},
  doi = {10.5281/zenodo.3960218},
  url = {https://allisonhorst.github.io/palmerpenguins/},
}
```