**Installation**

palmerpenguins is yet to make it to CRAN, so you need to download and combine the data to make a CSV file and upload into dataframe per the instruction as given on the README file of this github repository.

**Accessing Data**

After successful installation, you can find out that there are two datasets attached with the package – penguins and penguins\_raw. You can check out their help page (?penguins\_raw) to understand more about respective datasets.

**Loading Library**

library(tidyverse)

library(palmerpenguins)

**Meta – Glimpse of penguins dataset**

penguins dataset has got the following 7 columns and 344 columns

names(penguins)

## [1] "species" "island" "bill\_length\_mm"

## [4] "bill\_depth\_mm" "flipper\_length\_mm" "body\_mass\_g"

## [7] "sex"

Of the 7 columns, 3 are categorical (species,island,sex) and the rest are numeric.

glimpse(penguins)

## Rows: 344

## Columns: 7

## $ species Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Ade…

## $ island Torgersen, Torgersen, Torgersen, Torgersen, Torgers…

## $ bill\_length\_mm 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1,…

## $ bill\_depth\_mm 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1,…

## $ flipper\_length\_mm 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 18…

## $ body\_mass\_g 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475,…

## $ sex male, female, female, NA, female, male, female, mal…

**Penguins Data Column Definition**

**species**  
a factor denoting penguin species (Adélie, Chinstrap and Gentoo)

**island**  
a factor denoting island in Palmer Archipelago, Antarctica (Biscoe, Dream or Torgersen)

**bill\_length\_mm**  
a number denoting bill length (millimeters)

**bill\_depth\_mm**  
a number denoting bill depth (millimeters)

**flipper\_length\_mm**  
an integer denoting flipper length (millimeters)

**body\_mass\_g**  
an integer denoting body mass (grams)

**sex**  
a factor denoting penguin sex (female, male)

**Missing Values**

A good thing about penguins over iris is that, It’s got missing values NA. It’s quite an important thing to be present while using for educational purposes!

penguins %>%

#group\_by(species) %>%

select(everything()) %>%

summarise\_all(funs(sum([is.na](http://is.na)(.)))) %>%

pivot\_longer(cols = 1:7, names\_to = 'columns', values\_to = 'NA\_count') %>%

arrange(desc(NA\_count)) %>%

ggplot(aes(y = columns, x = NA\_count)) + geom\_col(fill = 'darkorange') +

geom\_label(aes(label = NA\_count)) +

# scale\_fill\_manual(values = c("darkorange","purple","cyan4")) +

theme\_minimal() +

labs(title = 'Penguins - NA Count')

## Warning: `funs()` is deprecated as of dplyr 0.8.0.

## Please use a list of either functions or lambdas:

##

## # Simple named list:

## list(mean = mean, median = median)

##

## # Auto named with `tibble::lst()`:

## tibble::lst(mean, median)

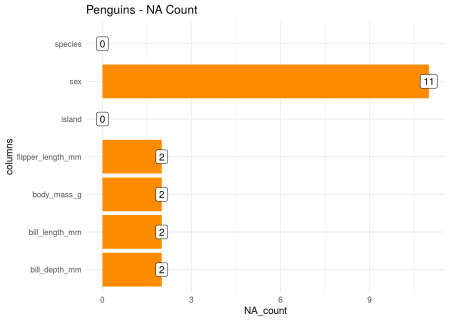
##

## # Using lambdas

## list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))

## This warning is displayed once every 8 hours.

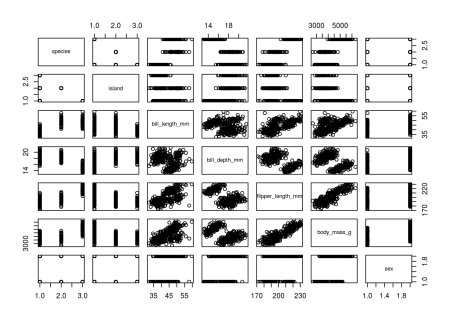
## Call `lifecycle::last\_warnings()` to see where this warning was generated.



**Simple Scatter Plot**

Like iris, You can simply make a scatter plot using base-R’s plot()

plot(penguins)

  
### Bar Plot

In this Bar plot, We can visualize the count of each species in the penguins dataset

penguins %>%

count(species) %>%

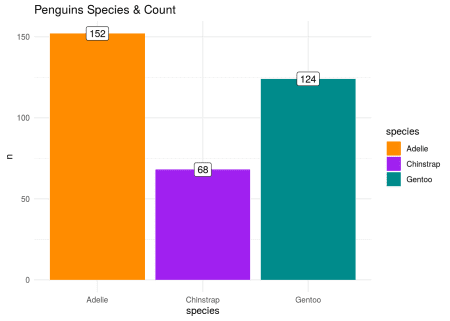
ggplot() + geom\_col(aes(x = species, y = n, fill = species)) +

geom\_label(aes(x = species, y = n, label = n)) +

scale\_fill\_manual(values = c("darkorange","purple","cyan4")) +

theme\_minimal() +

labs(title = 'Penguins Species & Count')

  
### Bar Plot for each Species

In this Bar Plot, We can visualize Species distribution of each Sex (with faceted plot)

penguins %>%

drop\_na() %>%

count(sex, species) %>%

ggplot() + geom\_col(aes(x = species, y = n, fill = species)) +

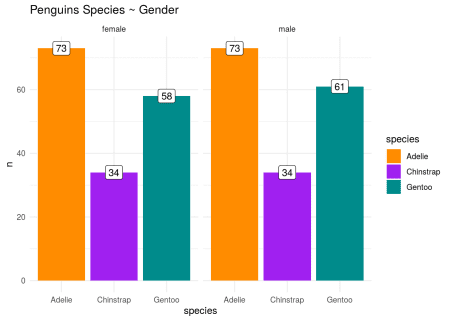
geom\_label(aes(x = species, y = n, label = n)) +

scale\_fill\_manual(values = c("darkorange","purple","cyan4")) +

facet\_wrap(~sex) +

theme\_minimal() +

labs(title = 'Penguins Species ~ Gender')



**Correlation Matrix**

penguins %>%

select\_if(is.numeric) %>%

drop\_na() %>%

cor()

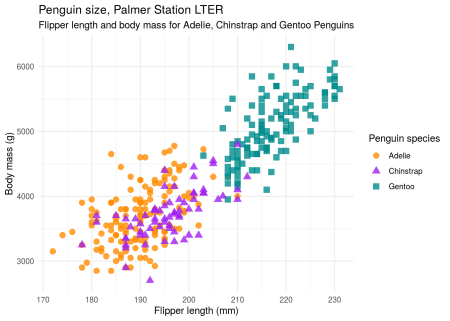
## bill\_length\_mm bill\_depth\_mm flipper\_length\_mm body\_mass\_g

## bill\_length\_mm 1.0000000 -0.2350529 0.6561813 0.5951098

## bill\_depth\_mm -0.2350529 1.0000000 -0.5838512 -0.4719156

## flipper\_length\_mm 0.6561813 -0.5838512 1.0000000 0.8712018

## body\_mass\_g 0.5951098 -0.4719156 0.8712018 1.0000000



**Scatter Plot – Penguins Size Relation wrt Species**

In this scatter plot, we’ll try to visualize relationship between flipper\_length\_mm and body\_mass\_g with respect to each species.

library(tidyverse)

ggplot(data = penguins,

aes(x = flipper\_length\_mm,

y = body\_mass\_g)) +

geom\_point(aes(color = species,

shape = species),

size = 3,

alpha = 0.8) +

#theme\_minimal() +

scale\_color\_manual(values = c("darkorange","purple","cyan4")) +

labs(title = "Penguin size, Palmer Station LTER",

subtitle = "Flipper length and body mass for Adelie, Chinstrap and Gentoo Penguins",

x = "Flipper length (mm)",

y = "Body mass (g)",

color = "Penguin species",

shape = "Penguin species") +

theme\_minimal()

**Scatter Plot – Penguins Size Relation wrt Island**

library(tidyverse)

ggplot(data = penguins,

aes(x = flipper\_length\_mm,

y = body\_mass\_g)) +

geom\_point(aes(color = island,

shape = species),

size = 3,

alpha = 0.8) +

#theme\_minimal() +

scale\_color\_manual(values = c("darkorange","purple","cyan4")) +

labs(title = "Penguin size, Palmer Station LTER",

subtitle = "Flipper length and body mass for each island",

x = "Flipper length (mm)",

y = "Body mass (g)",

color = "Penguin island",

shape = "Penguin species") +

theme\_minimal()

