hw\_2

JiminLee\_2229027

2023-09-15

getwd()

## [1] "/Users/jimin/Desktop/데스크탑 - 이지민의 MacBook Air/지민/ewha/2023-2/BD/HW"

### 1. palmerpenguines library에는 Antartica의 Palmer station근처에 서식하는 펭귄에 대한 자료인 penguins가 내장되어 있다.

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 1.0.2  
## ✔ tibble 3.2.1 ✔ dplyr 1.1.3  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.2  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

install.packages("palmerpenguins", dependencies=TRUE, repos = "http://cran.us.r-project.org")

## also installing the dependency 'recipes'

## Warning in readRDS(dest): lzma decoder corrupt data

##   
## There is a binary version available but the source version is later:  
## binary source needs\_compilation  
## recipes 1.0.6 1.0.8 FALSE  
##   
##   
## The downloaded binary packages are in  
## /var/folders/zl/ljg24f\_15m93w7f2rt14c2n80000gn/T//RtmpIXueMD/downloaded\_packages

## installing the source package 'recipes'

## Warning in install.packages("palmerpenguins", dependencies = TRUE, repos =  
## "http://cran.us.r-project.org"): installation of package 'recipes' had non-zero  
## exit status

library('palmerpenguins')

data("penguins")  
head(penguins)

## # A tibble: 6 × 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>

#### 1) penguins 자료의 개수는?

nrow(penguins)

## [1] 344

#### 2) penguins 자료에는 어떤 변수들이 있는가?

variable.names(penguins) # or colnames()

## [1] "species" "island" "bill\_length\_mm"   
## [4] "bill\_depth\_mm" "flipper\_length\_mm" "body\_mass\_g"   
## [7] "sex" "year"

#### 3) species는 어떤 것들이 있으며 각각 몇 개의 자료가 있는가?

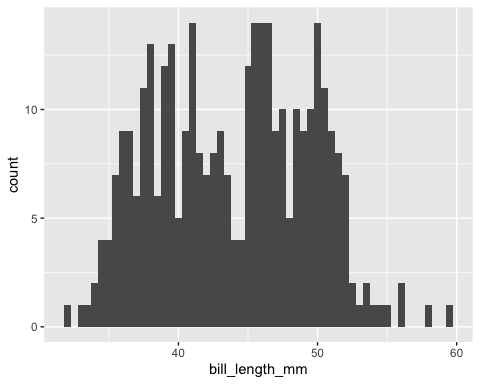
summary(penguins$species) # or table() method

## Adelie Chinstrap Gentoo   
## 152 68 124

#### 4) penguins 자료에 있는 4개의 연속변수 각각의 분포를 수업시간에 배운 내용을 바탕으로 살펴보시요.

# bill\_length\_mm, bill\_depth\_mm, flipper\_length\_mm, body\_mass\_g are continous variable  
ggplot(penguins, aes(x=bill\_length\_mm)) + geom\_histogram(binwidth=0.5)

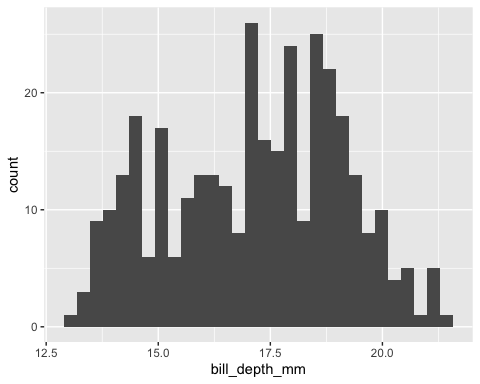
## Warning: Removed 2 rows containing non-finite values (stat\_bin).



ggplot(penguins, aes(x=bill\_depth\_mm)) + geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

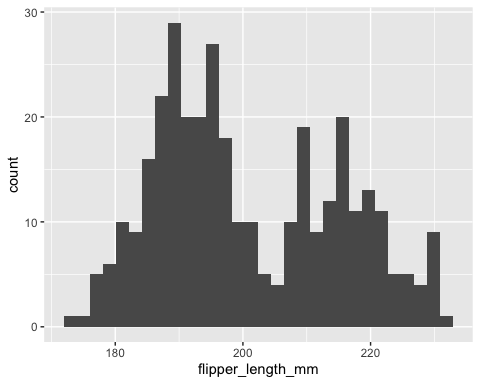
## Warning: Removed 2 rows containing non-finite values (stat\_bin).



ggplot(penguins, aes(x=flipper\_length\_mm)) + geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

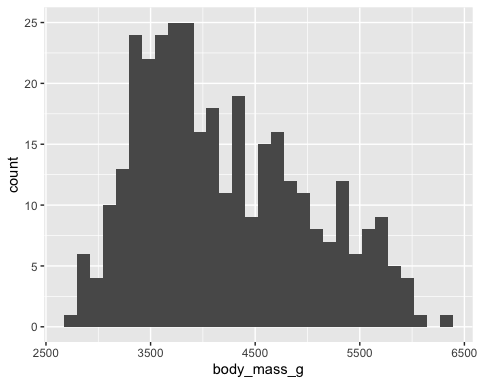
## Warning: Removed 2 rows containing non-finite values (stat\_bin).



ggplot(penguins, aes(x=body\_mass\_g)) + geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

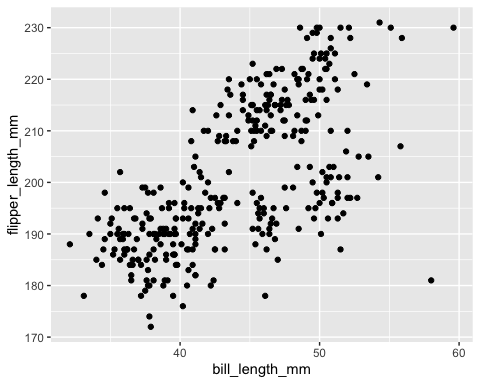
## Warning: Removed 2 rows containing non-finite values (stat\_bin).



#### 5) bill\_length\_mm과 flipper\_length\_mm의 산점도를 그리시오.

ggplot(penguins, aes(x=bill\_length\_mm, y=flipper\_length\_mm))+geom\_point()

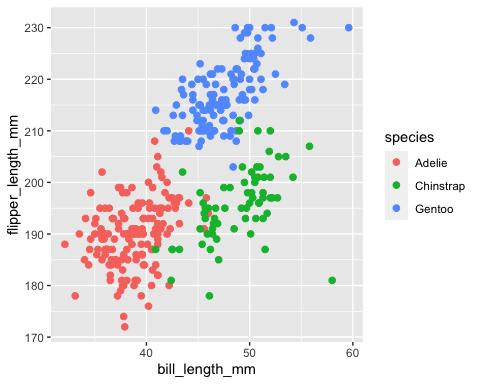
## Warning: Removed 2 rows containing missing values (geom\_point).



#### 6) 5)에서 그린 산점도에 species 별로 다른 색을 칠하고 Species가 구별되는지 살펴보시오.

ggplot(penguins, aes(x=bill\_length\_mm, y=flipper\_length\_mm, colour=species)) + geom\_point(size=2)

## Warning: Removed 2 rows containing missing values (geom\_point).



#### 7) 산점도를 species 별로 나누어서 그려보시오.

ggplot(penguins, aes(x=bill\_length\_mm, y=flipper\_length\_mm, colour=species)) + geom\_point() + facet\_wrap(~species)

## Warning: Removed 2 rows containing missing values (geom\_point).

