part1\_10

gyc

2022-08-08

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

## 팔머 펭귄을 분석해보자

팔머 펭귄 데이터셋이 궁금해?? 여기를 가봐 <https://bit.ly/36rDgFx>

library(palmerpenguins)  
library(psych)  
penguins

## # A tibble: 344 × 8  
## species island bill\_length\_mm bill\_depth\_mm flipper\_…¹ body\_…² sex year  
## <fct> <fct> <dbl> <dbl> <int> <int> <fct> <int>  
## 1 Adelie Torgersen 39.1 18.7 181 3750 male 2007  
## 2 Adelie Torgersen 39.5 17.4 186 3800 fema… 2007  
## 3 Adelie Torgersen 40.3 18 195 3250 fema… 2007  
## 4 Adelie Torgersen NA NA NA NA <NA> 2007  
## 5 Adelie Torgersen 36.7 19.3 193 3450 fema… 2007  
## 6 Adelie Torgersen 39.3 20.6 190 3650 male 2007  
## 7 Adelie Torgersen 38.9 17.8 181 3625 fema… 2007  
## 8 Adelie Torgersen 39.2 19.6 195 4675 male 2007  
## 9 Adelie Torgersen 34.1 18.1 193 3475 <NA> 2007  
## 10 Adelie Torgersen 42 20.2 190 4250 <NA> 2007  
## # … with 334 more rows, and abbreviated variable names ¹​flipper\_length\_mm,  
## # ²​body\_mass\_g  
## # ℹ Use `print(n = ...)` to see more rows

## 이제부터 그림을 그려볼꺼야

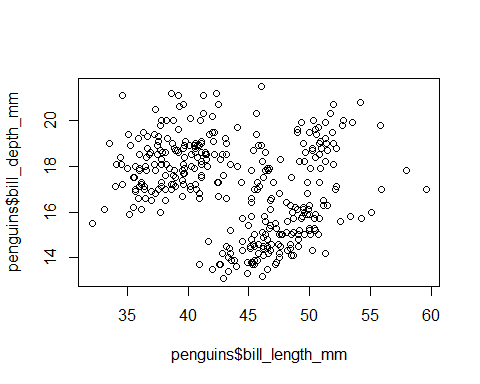
이런 순서로 해보자.

1. 부리 길이<->부리 깊이의 산점도를 그린다.

* 음의 상관이 있는 것 처럼 보이는가?
* 이것은 simpson’s paradox로 인한 것이다. 2.펭귄의 종별로 색깔을 입힌다.
* 종별로 구분하니 양의 상관이 있는 것처럼 보이지 않는가?
* 심슨의 역설은 우리가 조심해야 할 통계의 함정이다.

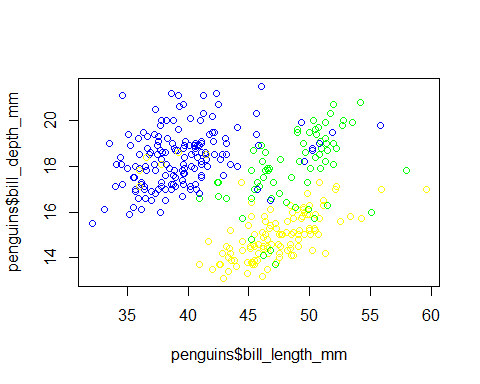
## 1번 문제

plot(penguins$bill\_length\_mm,penguins$bill\_depth\_mm)

 - 점들이 너무 퍼져 있어서 음인지 양인지도 분간이 어려우므로 부리길이와 깊이는 아무 상관이 없어 보인다.

## 2문제 문제

pg\_nax<-na.omit(penguins)  
mycolor <- ifelse(pg\_nax$species=='Adelie','blue',ifelse(pg\_nax$species=='Gentoo','yellow','green'))  
plot(penguins$bill\_length\_mm,penguins$bill\_depth\_mm,col=mycolor)

 - 종별로 색깔을 구별해서보니 - Adelie는 약한 양의 상관 - Gentoo는 강한 양의 상관 - Chinstrap는 중간쯤 양의 상관이 보이는 것을 확인

## 2번 문제 다른방법 => as.numeric()

pg\_nax<-na.omit(penguins)  
mycolor<-as.numeric(pg\_nax$species)  
plot(penguins$bill\_length\_mm,penguins$bill\_depth\_mm,col=mycolor)

