Pulse Rate Study

Zhan Li

```
library(tidyverse)
library(broom)
```

Question 1

(a)

```
url <- "http://ritsokiguess.site/STAC32/pulsemarch.csv"
data <- read_csv(url)
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Sex = col_character(),
##   Before = col_double(),
##   After = col_double()
## )
```

```
data
```

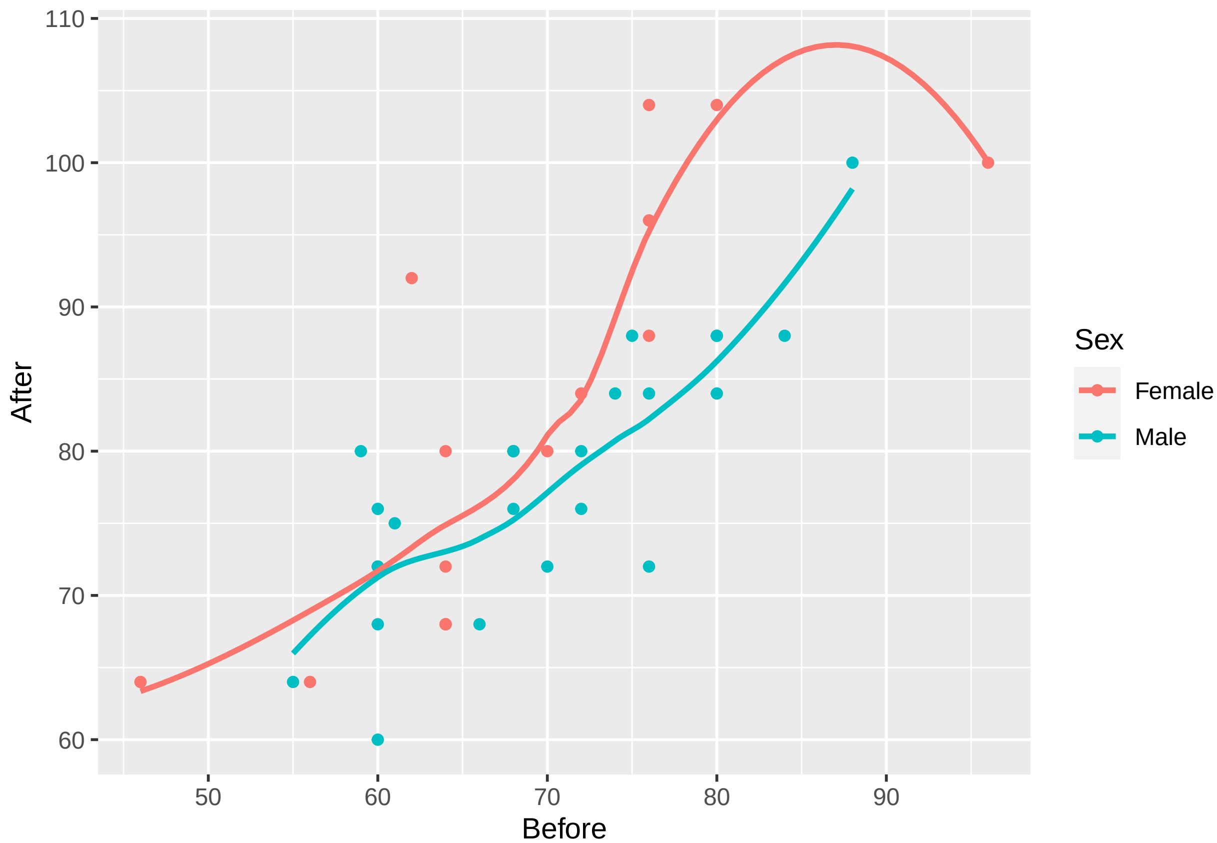
```
## # A tibble: 40 x 3
##    Sex    Before After
##    <chr>   <dbl> <dbl>
##  1 Female     72    84
##  2 Male       60    72
##  3 Female     68    80
##  4 Male       70    72
##  5 Male       68    80
##  6 Male       61    75
##  7 Male       80    84
##  8 Male       72    76
##  9 Female     64    80
## 10 Female     62    92
## # ... with 30 more rows
```

- use "read_csv" because it's a csv file.

(b)

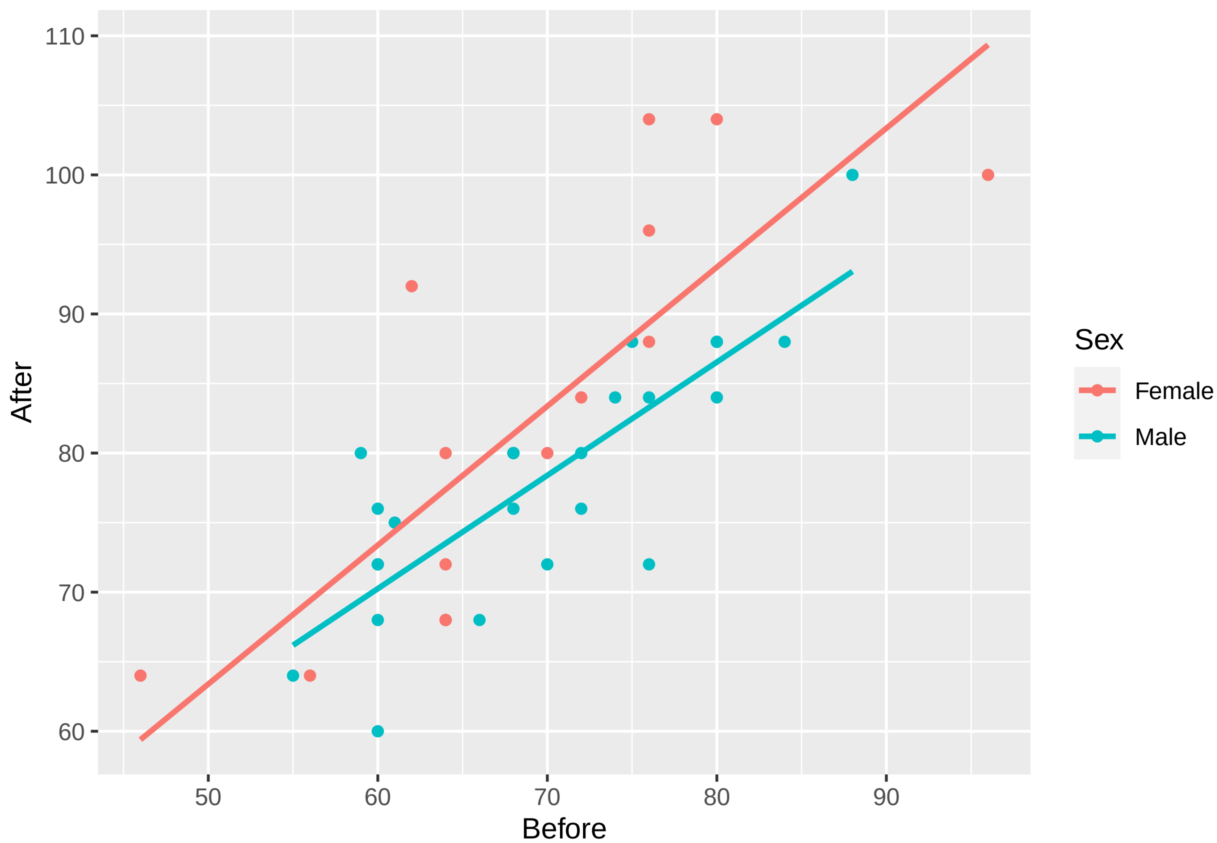
```
ggplot(data, aes(x = Before, y = After, colour = Sex)) + geom_point() + geom_smooth(se = F)
```

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




```
ggplot(data, aes(x = Before, y = After, colour = Sex)) + geom_point() + geom_smooth(method = "lm", se =
```

```
## `geom_smooth()` using formula 'y ~ x'
```



- The X variable is Before and Y variable is After because the Before pulse is our explanatory variable and the After pulse is our response variable.
- Use blue to represent males and red to represent females.
- First, we don't assume linear, but we can see from the first graph that both males and females can pass having a linear relationship between before and after.
- Add a linear regression line in the second graph.

(c)

- As seen from the graph, females generally have a higher after pulse rate than males do.
- Also, for most points all the graph–both males and females, the after pulse rates are higher than the before pulse rates.
- From the linear regression line, the slope of the female pulse rate is slightly larger than the male's. This means that the rate of pulse increase is a little higher for females.

(d)

```
fit <- lm(After ~ Before + Sex, data = data)
summary(fit)
```

```
##
## Call:
## lm(formula = After ~ Before + Sex, data = data)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -11.8653  -4.6319  -0.4271   3.3856  16.0047 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  19.8003     7.9217   2.499   0.0170 *  
## Before        0.9064     0.1127   8.046  1.2e-09 ***
## SexMale      -4.8191     2.2358  -2.155   0.0377 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 6.918 on 37 degrees of freedom
## Multiple R-squared:  0.6468, Adjusted R-squared:  0.6277 
## F-statistic: 33.87 on 2 and 37 DF,  p-value: 4.355e-09
```

```
glance(fit)
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
##       <dbl>         <dbl> <dbl>     <dbl>   <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1     0.647         0.628  6.92      33.9 4.36e-9     2  -133.  273.  280.
## # ... with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
tidy(fit)
```

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic       p.value
##   <chr>          <dbl>     <dbl>     <dbl>         <dbl>
## 1 (Intercept)   19.8       7.92       2.50 0.0170       
## 2 Before         0.906     0.113      8.05 0.00000000120
## 3 SexMale       -4.82      2.24      -2.16 0.0377
```

(e)

Under the linear regression model:

- The "SexMale" tells me the after pulse of males is about 4.8191 lower than that for females at the same before pulse.
- The slope of "Before" says the increasing before pulse by 1 increases the after pulse by about 0.9064 for both males and females.
- r squared is 0.6467726, which is not too terrible.
- The second slope does surprise me because I was expecting the slope to be greater than 1 since the after pulse should increase at a faster rate as the before pulse increases.

(f)

- As seen from the output, the estimate value of SexMale is -4.8191363.
- The "SexMale" tells me the after pulse of males is about 4.8191 lower than that for females at the same before pulse.

Question 2

(a)

```
cv <- function(x){
  sd(x)/mean(x)
}
```

(b)

```
x <- 1:5
cv(x)
```

```
## [1] 0.5270463
```

- Hence, the coefficient of variation of the set of integers 1 through 5 is 0.5270463.

(c)

```
x <- c(-2.8, -1.8, -0.8, 1.2, 4.2)
cv(x)
```

```
## [1] 6.248491e+16
```

- The coefficient is 6.2484909×10^{16}.
- It doesn't make sense because the mean of x is 0. But we can't divide by 0 because that'd be undefined.

```
mean(x)
```

```
## [1] 4.440892e-17
```

(d)

```
cv <- function(x){
  stopifnot(x>=0)
  sd(x)/mean(x)
}
```

- Use "stopifnot" to give an error message if the vector is negative.
- Test the function using our previous inputs:

```
x <- 1:5
cv(x)
```

```
## [1] 0.5270463
```

```
x <- c(-2.8, -1.8, -0.8, 1.2, 4.2)
cv(x)
```

```
## Error in cv(x): x >= 0 are not all TRUE
```

- Notice the first input outputs the same value we got from part (b)
- But the second input outputs error message since there are negative numbers, which is exactly what we wanted.