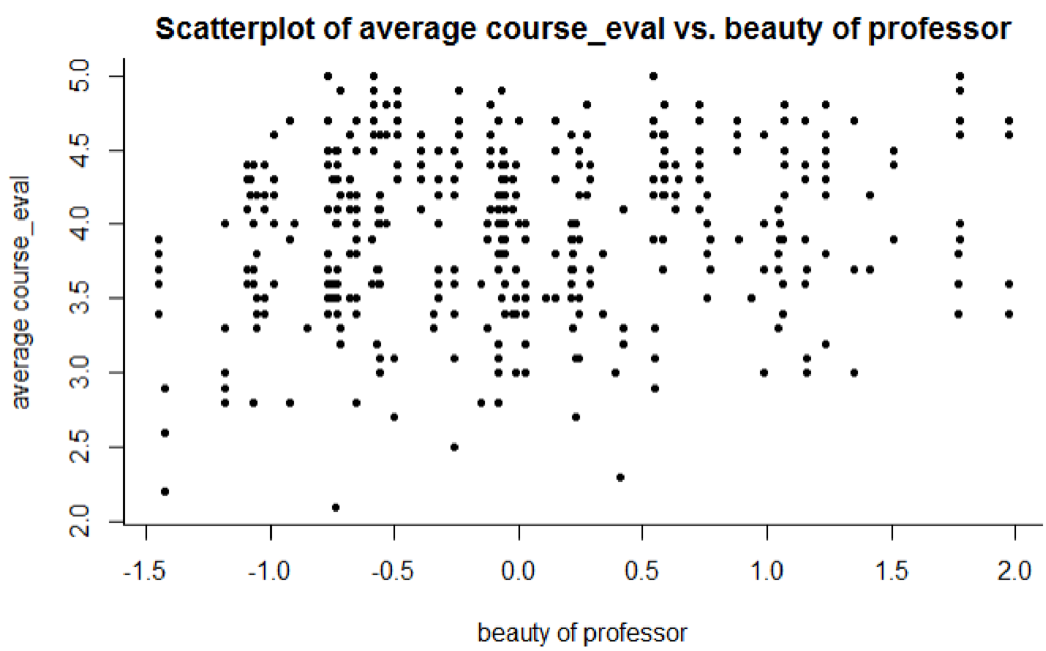
Empirical Exercise

Group 6

Question a

Here comes the scatterplot of average course evaluations (Course_Eval) on the professor's beauty(Beauty). It does have a relationship between them.

```
plot(df$beauty, df$course_eval, col = "black", pch =16, cex = 0.7, bty = "l",
        main = "Scatterplot of average course_eval vs. beauty of professor",
        xlab = "beauty of professor", ylab = "average course_eval")
```



Question b

The regression of average course evaluations(Course_Eval) on the professor's beauty(Beauty) is

Course_Eval =3.99827 + 0.133*Beauty

The estimated intercept is 3.99827, and the estimated slope is 0.133.

```
mod1 <- lm(course_eval ~ beauty, data = df)
summary(mod1)
```

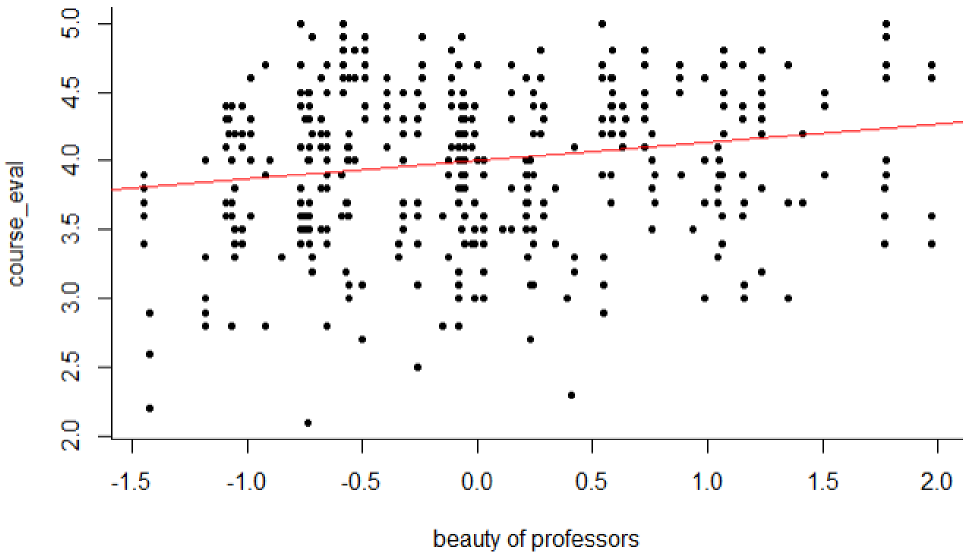
```
Call:
lm(formula = course_eval ~ beauty, data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-1.80015 -0.36304  0.07254  0.40207  1.10373

Coefficients:
            Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)  3.99827    0.02535   157.727   < 2e-16 ***
beauty       0.13300    0.03218    4.133    4.25e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5455 on 461 degrees of freedom
Multiple R-squared:  0.03574,  Adjusted R-squared:  0.03364
F-statistic: 17.08 on 1 and 461 DF,  p-value: 4.247e-05
```

```
plot(df$beauty, df$course_eval,
     col = "black", pch =16, cex = 0.7, bty = "l",
     xlab = "beauty of professors", ylab = "course_eval")
# add a straight line with an intercept a and slop b
abline(coef(mod1)[1], coef(mod1)[2], col="red")
# add a text on the plot
text(23, 660, "course_eval = 3.99827 + 0.133 beauty",
     cex.lab = 0.9, font.lab = 3)
```



We can see from the summary below, the mean of the independent variable beauty is equal to 0, and the estimated intercept

$$\widehat{intercept} = \overline{\mathbf{course\_eval}} - \mathbf{0.133} * \overline{\mathbf{beauty}} = \overline{\mathbf{course\_eval}} - \mathbf{0} * \mathbf{0.133}$$
$$= \overline{\mathbf{course\_eval}}$$

Thus the eatimated intercept is equal to the mean of Coures_eval.

```
df <- classdata[c("course_eval", "beauty")]
summary(df)
```

```
  course_eval        beauty
 Min.    :2.100   Min.    :-1.45049
 1st Qu.:3.600   1st Qu.:-0.65627
 Median :4.000   Median :-0.06801
 Mean    :3.998   Mean    : 0.00000
 3rd Qu.:4.400   3rd Qu.: 0.54560
 Max.    :5.000   Max.    : 1.97002
```

Question c

```
sd(df$beauty)
```

```
[1] 0.7886477
```

So the standard deviation of *Beauty* is 0.7886477;

1) Professor Watson has an average value of *Beauty*, which is equal to 0,
   so his **Course_Eval =3.99827 + 0 * 0.133 *0.79= 3.99827;**

2) Professor Stock's value of *Beauty* is one standard deviation above the average.
   so his **Course_Eval =3.99827 + 1 * 0.133*0.79 = 4.10316.**

Question d

```
sd(df$course_eval)
sd(df$beauty)
```

```
[1] 0.5548656
[1] 0.7886477
```

The standard deviation of *course evaluation* is 0.5548656, and the standard deviation of *Beauty* is 0.7886477. One additional standard deviation on *Beauty* will increase *course evaluation* is 0.133*0.7886477=0.105, only nearly 20% on the standard deviation of *course evaluation.* So the estimated effect of Beauty on Course_Eval is small.

And we suppose that the null hypothesis is that estimated slope is equal to a population value, and the alternative hypothesis is the opposite. We can find that the t value is equal to

4.133, which is much larger than 1.96, equivalently, P-value is 4.25e-05, is much smaller than 0.05, so we reject the null hypothesis. It means that the size of estimated slope is small on the Course_Eval

Question e

Because $R^2 = 0.03574$, the goodness of fit is low because the *Beauty* only explains 3.574% of the variance in *course evaluation*, so *Beauty* can't explain *course evaluation* well.