Lab 1b. Linear Regression

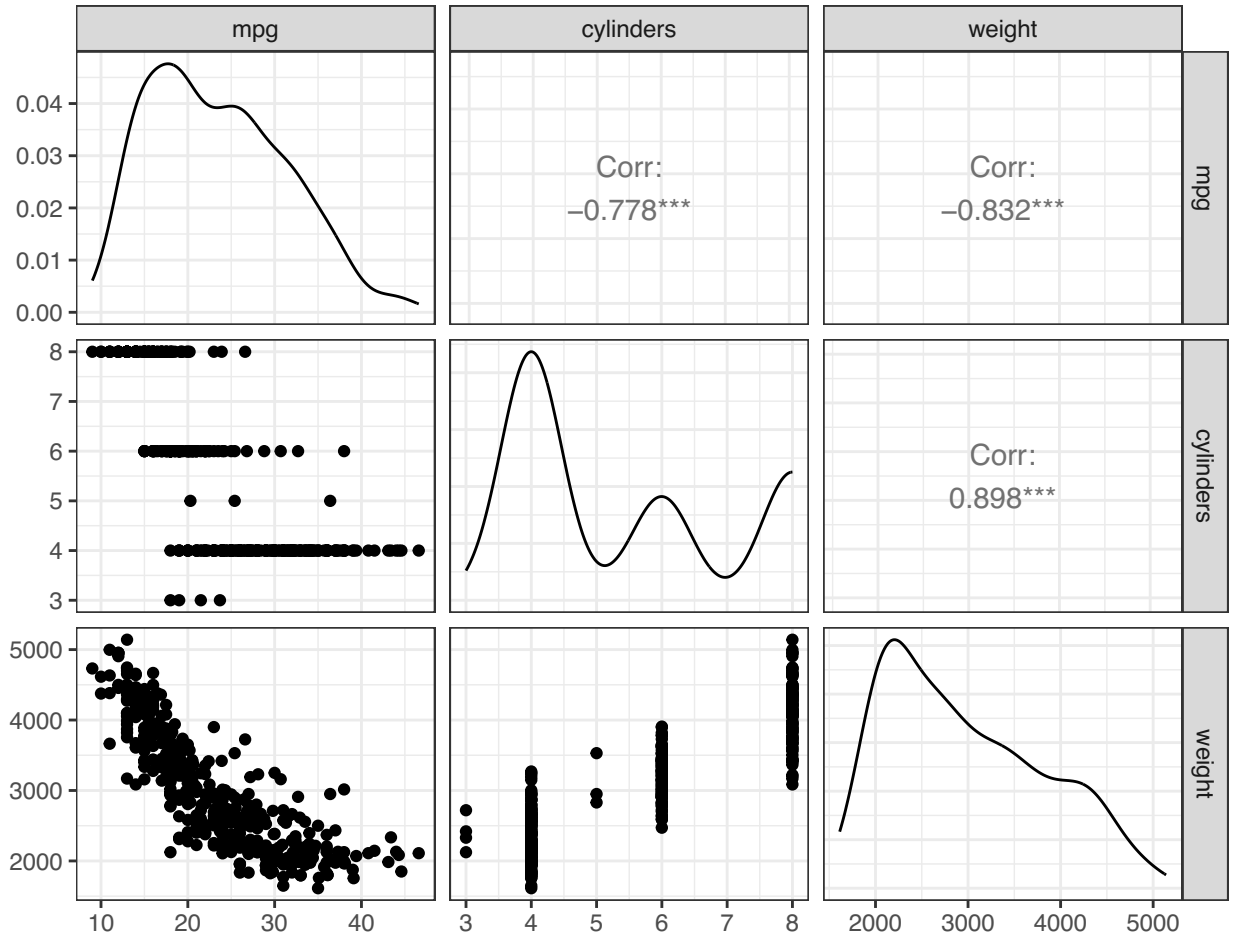
ISL Chapter 3

Yerim Oh

Example

The Auto was taken from the StatLib library which is maintained at Carnegie Mellon University. The dataset was used in the 1983 American Statistical Association Exposition. The original dataset has 397 observations, of which 5 have missing values for the variable "horsepower". These rows are removed here.

```
##   mpg cylinders displacement horsepower weight acceleration year origin
## 1  18         8          307        130   3504         12.0   70      1
## 2  15         8          350        165   3693         11.5   70      1
## 3  18         8          318        150   3436         11.0   70      1
## 4  16         8          304        150   3433         12.0   70      1
## 5  17         8          302        140   3449         10.5   70      1
## 6  15         8          429        198   4341         10.0   70      1
##                       name
## 1 chevrolet chevelle malibu
## 2         buick skylark 320
## 3        plymouth satellite
## 4             amc rebel sst
## 5               ford torino
## 6          ford galaxie 500
```


mpg
cylinders
weight
mpg
cylinders
weight
Corr:
−0.778***
Corr:
−0.832***
Corr:
0.898***
0.04
0.03
0.02
0.01
0.00
8
7
6
5
4
3
5000
4000
3000
2000
10
20
30
40
3
4
5
6
7
8
2000
3000
4000
5000

1. Fit a model for `mpg` using `cylinders`, `weight` as explanatory variables. Print a summary.

```
model_fit <- lm(mpg ~ cylinders+weight, data = Auto)
summary(model_fit)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + weight, data = Auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.6469  -2.8282  -0.2905   2.1606  16.5856
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 46.2923105  0.7939685  58.305   <2e-16 ***
## cylinders   -0.7213779  0.2893780  -2.493   0.0131 *
## weight      -0.0063471  0.0005811 -10.922   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.304 on 389 degrees of freedom
## Multiple R-squared:  0.6975, Adjusted R-squared:  0.6959
## F-statistic: 448.4 on 2 and 389 DF,  p-value: < 2.2e-16
```

2. Test the hypothesis that `cylinders` is not related to `mpg`.

$H_0: \beta_1 = 0$ t-statistics $= \frac{-0.72}{0.29} \approx -2.493$
$H_A: \beta_1 \neq 0$ p-value = 0.0131

We have a moderate evidence against the null hypothesis that the cylinders do not effect the mpg of the car

- H_0: $\beta_{cylinders} = 0$
- H_A: $\beta_{cylinders} \neq 0$
- t-stat = -2.493
- p-value = 0.0131
  - We have strong evidence against the null hypothesis that the cylinders do not affect the miles per gas.

3. Test the hypothesis that `weight` is not related to `mpg`.

$H_0: \beta_2 = 0$, $H_A: \beta_2 \neq 0$
t-stat = -10.922, p-value = less than 2e-16

We have a very strong evidence against the null hypothesis that the vehicle weight do not effect the mpg

- H_0: $\beta_{weight} = 0$
- H_A: $\beta_{weight} \neq 0$
- t-stat = -10.922
- p-value = less than 2e-16
  - We have a very strong evidence against the null hypothesis that the weight does not affect the miles per gas.

4. Test the hypothesis neither `cylinders` or `weight` is related to `mpg`.

$H_0: \beta_1 = \beta_2 = 0$ H_A: at least one of β_1 or β_2 is non zero
F-statistics: 448.4 p-value = less than 2.2e-16

We have a very strong evidence against the null hypothesis that neither cylinder or weight is related to car

- H_0: $\beta_{cylinders} = \beta_{weight} = 0$
- H_A: $\beta_{cylinders} \neq 0$ or $\beta_{weight} \neq 0$; At least one of $\beta_{cylinders}$ or β_{weight} is not equal to 0
- F-stat = 448.4
- p-value = less than 2.2e-16
  - We have a very strong evidence against the null hypothesis that neither cylinders or weight is related to mpg.

5. Do your results align with your intuition based on the visualization of the data?

Both variables seem to be related to the `mpg` variable. The conclusions therefore align with the data visualization.