Lab 12

Yunting Chiu

2021-03-26

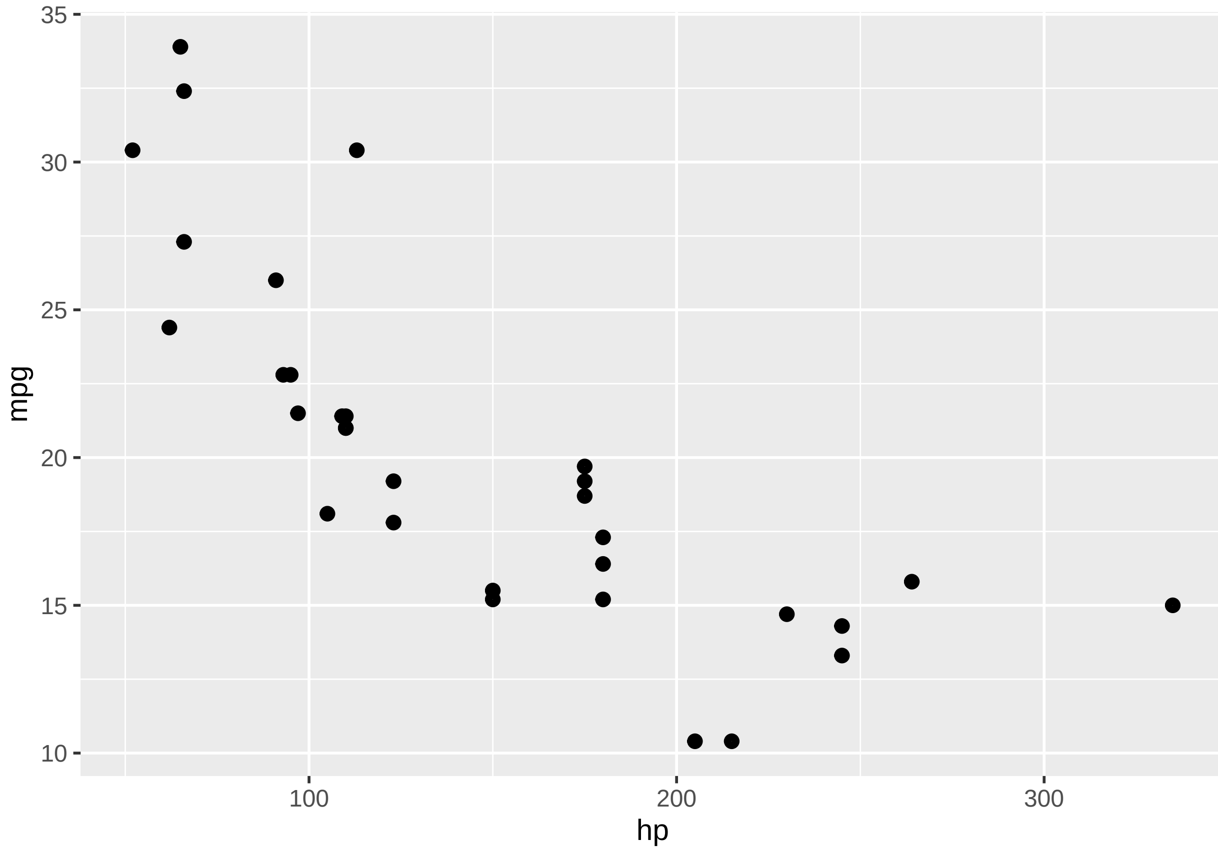
Install the data

```
mtcars %>%
  head() %>%
  select(mpg, hp, am)
```

```
##                    mpg  hp am
## Mazda RX4         21.0 110  1
## Mazda RX4 Wag     21.0 110  1
## Datsun 710        22.8  93  1
## Hornet 4 Drive    21.4 110  0
## Hornet Sportabout 18.7 175  0
## Valiant           18.1 105  0
```

We will focus on three variables: `mpg`, `hp`, `am`

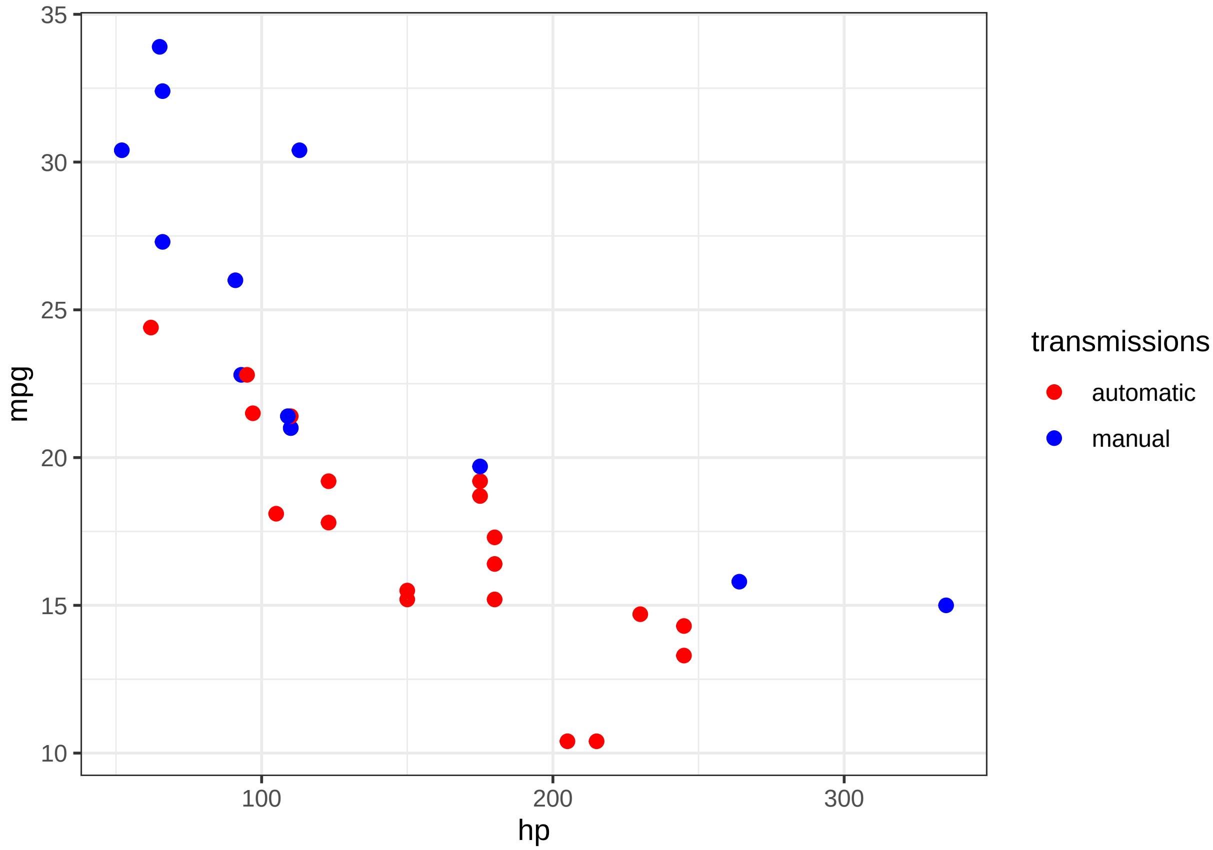
```
mtcars %>%
  ggplot(aes(x = hp, y = mpg)) +
  geom_point(shape = 21, size = .1, stroke = 2)
```



We could also label the points based on the transmission type `am`.
`am` is a dummy variable:

- 0 for automatic transmissions
- 1 for manual transmissions

```
mtcars %>%
  ggplot(aes(x = hp, y = mpg, col = as.factor(am))) +
  geom_point(shape = 21, size = .1, stroke = 2) +
   scale_color_manual(name="transmissions",
                      labels=c("automatic","manual"),
                      values=c("red","blue")) +
  theme_bw()
```



simple linear regression

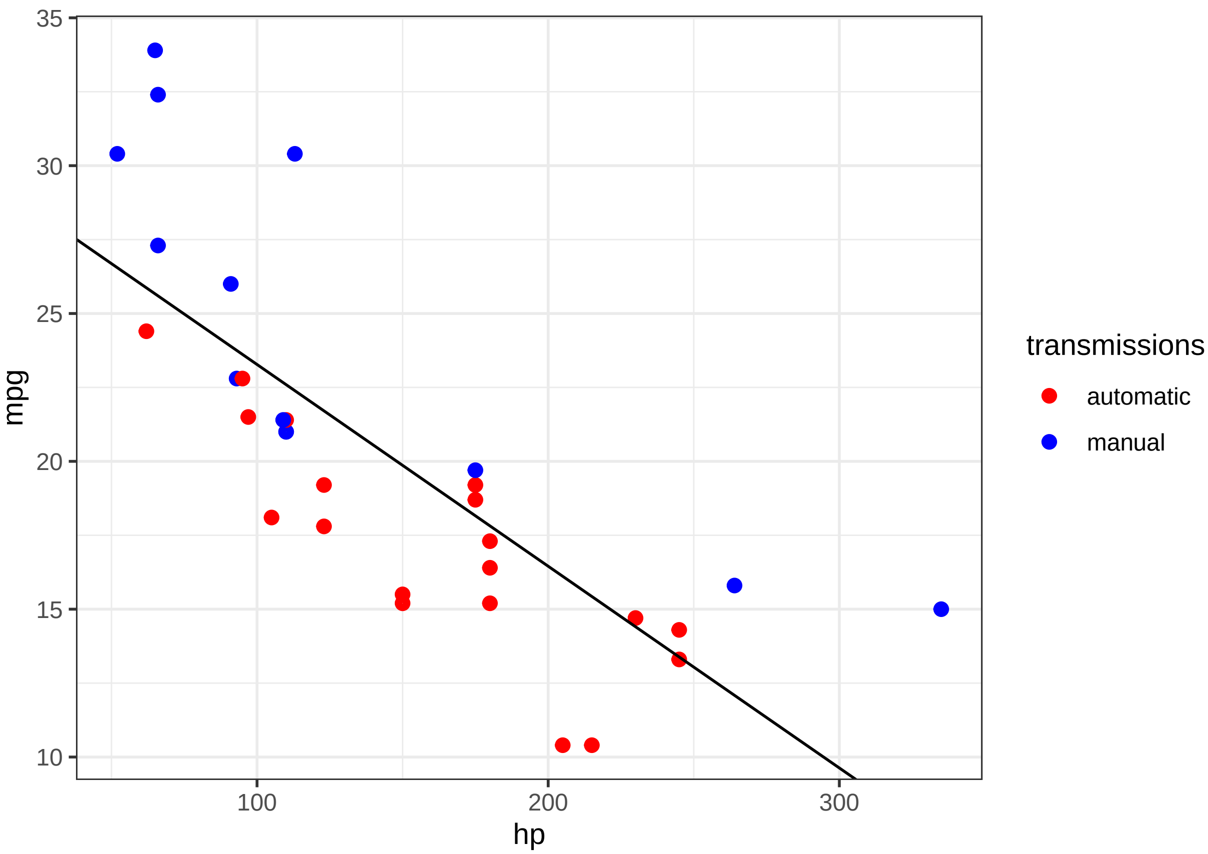
$$Y = \beta 0 + \beta 1 X1 + ei$$

```
mpg_hp_slr = lm(mpg ~ hp, data = mtcars)
summary(mpg_hp_slr)
```

```
##
## Call:
## lm(formula = mpg ~ hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7121 -2.1122 -0.8854  1.5819  8.2360
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
## hp          -0.06823    0.01012  -6.742 1.79e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.863 on 30 degrees of freedom
## Multiple R-squared:  0.6024, Adjusted R-squared:  0.5892
## F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```

Add the fitted line to the plot

```
mtcars %>%
  ggplot(aes(x = hp, y = mpg, col = as.factor(am))) +
  # geom_smooth(method='lm', formula = y~x, se = FALSE) +
    geom_point(shape = 21, size = .1, stroke = 2) +
   scale_color_manual(name="transmissions",
                       labels=c("automatic","manual"),
                       values=c("red","blue")) +
  theme_bw() +
  geom_abline(aes(intercept = 30.09886, slope = -0.06823))
```



Obviously, the red points are below the line, meaning that the model overestimates the fuel efficiency of automatic transmission. On the other hand, the blue points are above the line, meaning that the model underestimates the fuel efficiency of manual transmission. Thus, we need to constantly adjust the model.

Multiple regression model

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + ei$$

The model looks like

```
mpg_hp_add = lm(mpg ~ hp + am, data = mtcars)
summary(mpg_hp_add)
```

```
##
## Call:
```

```
## lm(formula = mpg ~ hp + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3843 -2.2642  0.1366  1.6968  5.8657
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26.584914   1.425094  18.655  < 2e-16 ***
## hp          -0.058888   0.007857  -7.495 2.92e-08 ***
## am           5.277085   1.079541   4.888 3.46e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.909 on 29 degrees of freedom
## Multiple R-squared:  0.782,  Adjusted R-squared:  0.767
## F-statistic: 52.02 on 2 and 29 DF,  p-value: 2.55e-10
```

As we mentioned, X2(`am`) is a dummy variable, it's only takes the values 0 and 1. We can write two separate versions, one for manual transmissions and the other for automatic transmissions.

For automatic transmissions - X2 = 0

$$Y = \beta 0 + \beta 1X1 + ei$$

For manual transmissions - X2 = 1

$$Y = (\beta 0 + \beta 2) + \beta 1X1 + ei$$

These models have the same slope $\beta 1$, but different intercepts, which differ by $\beta 2$. So the change in `mpg` is the same for both models, but the average mpg difference between the two transmission types is $\beta 2$.

```
summary(mpg_hp_add)$coefficients[1] # b0
```

```
## [1] 26.58491
```

```
summary(mpg_hp_add)$coefficients[2] # b1
```

```
## [1] -0.0588878
```

```
summary(mpg_hp_add)$coefficients[3] # b2
```

```
## [1] 5.277085
```

The estimated slope and intercepts can then be calculated by combining these.

```
intercept_auto = coef(mpg_hp_add)[1]
intercept_auto
```

```
## (Intercept)
##    26.58491
```

```
intercept_manu = coef(mpg_hp_add)[1] + coef(mpg_hp_add)[3]
intercept_manu
```

```
## (Intercept)
##      31.862
```

```
slope_auto = coef(mpg_hp_add)[2]
slope_auto
```

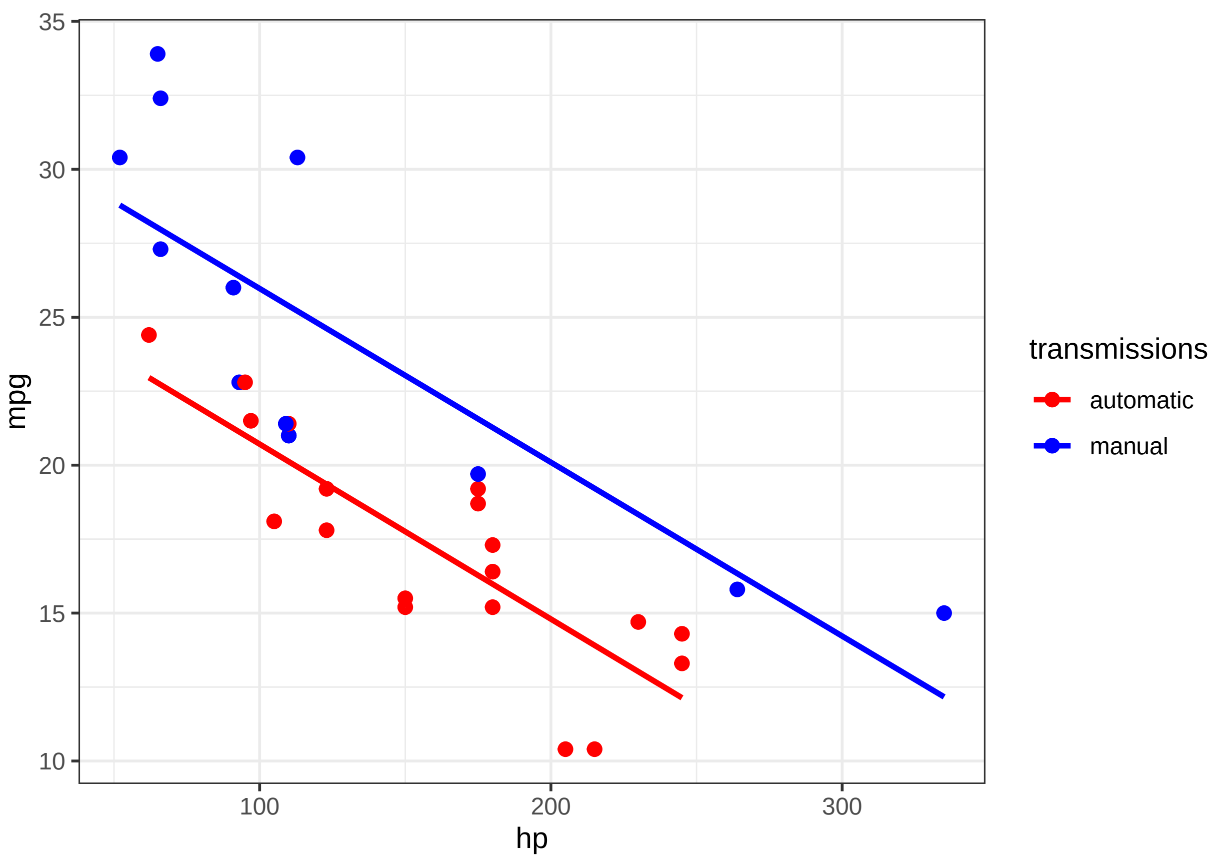
```
##          hp
## -0.0588878
```

```
slope_manu = coef(mpg_hp_add)[2]
slope_manu
```

```
##          hp
## -0.0588878
```

Re-Plot

```
mtcars %>%
  ggplot(aes(x = hp, y = mpg, col = as.factor(am))) +
  geom_smooth(method='lm', formula = y~x, se = FALSE) +
    geom_point(shape = 21, size = .1, stroke = 2) +
   scale_color_manual(name="transmissions",
                     labels=c("automatic","manual"),
                     values=c("red","blue")) +
  theme_bw()
```



The above plot makes it abundantly clear that $\beta 2$ is significant, but let us test it mathematically.

- Hypothesis Test: $\beta 2 = 0$ vs $\beta 2 \neq 0$

t-test

```
summary(mpg_hp_add)$coefficients["am", ]
```

```
##     Estimate   Std. Error      t value     Pr(>|t|)
## 5.277085e+00 1.079541e+00 4.888270e+00 3.460318e-05
```

F test

```
anova(mpg_hp_slr, mpg_hp_add)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ hp
## Model 2: mpg ~ hp + am
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 447.67
## 2     29 245.44  1    202.24 23.895 3.46e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the t-test and f-test, the p-values are the same, but the F test statistic is the t test statistic squared.

Interpretations

- bo = 26.584914 is the estimated average `mpg` for an automatic transmission car with 0 `hp`.
- b0+b2 = 26.584914 + 5.277085 = 31.862 is the estimated average `mpg` for a manual transmission car with 0 `hp`.
- b2 = 5.277085 is the estimated difference in average `mpg` for cars with manual transmissions as compared to those with automatic transmission, for any `hp`.
- b1 = -0.058888 is the estimated change in average `mpg` corresponds to increase in one unit of `hp`, no matter manual or automatic transmissions.