

Assignment1114

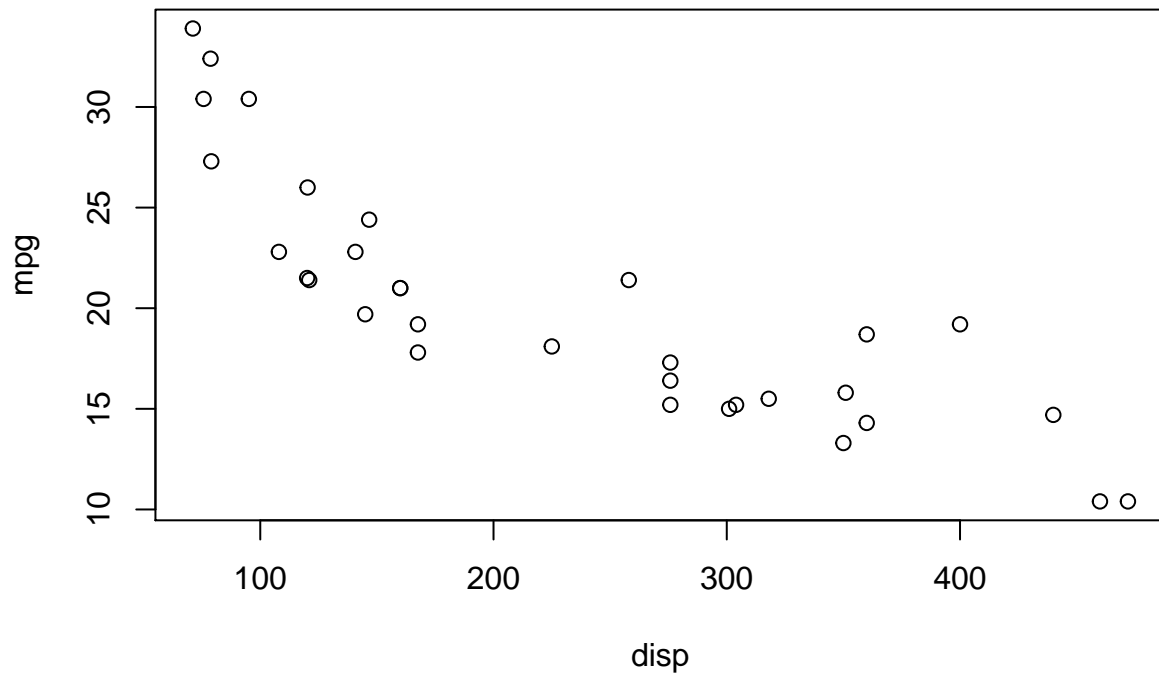
Suz

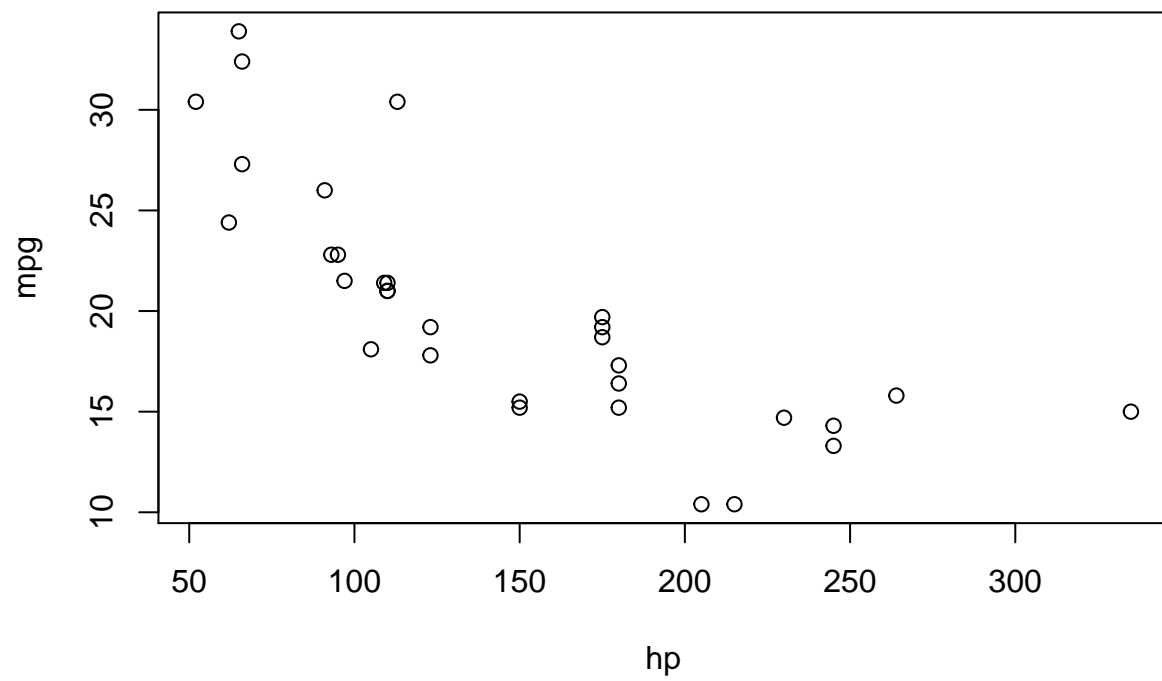
2019 11 14

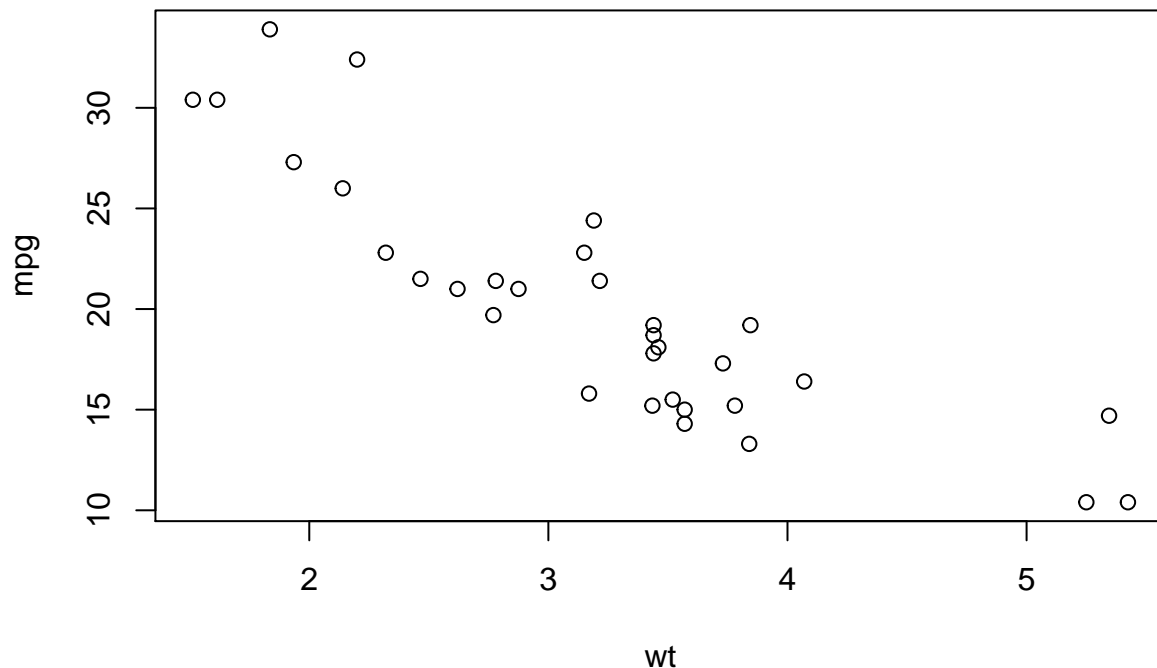
#show equation #don't show code for summary >> {r echo=FALSE}

Dependant Variable (Response): mpg Independent Variables (Predictor): disp, hp, wt

First to look into the relationship between the variables we look into the plots:







From this, we can assume that mpg and the other variables are in a negative relationship. Next we create a linear model regression with the variables, with mpg as the response variable to see how much each variables contribute in predicting mpg.

A linear model can be created using the `lm()` function. The minimum parameter it takes is the formular, which is constructed like this:

(eq1) *lm(reponse variable ~ predictor variable)*

Use tilda(~) to divide the reponse variable to the Left hand side, the individual variables to the Right hand side. For multiple individual variables, you can use '+' to add another variable.

(eq2) *lm(reponse variable ~ predictor variable1+predictor variable2)*

If you want to also add the interaction coefficient of each variables you can use '·'.

(eq3) *lm(reponse variable ~ predictor variable1+predictor variable2+predictor variable1:predictor variable2)*

To simplify eq3 above use '*'.

(eq4) *lm(reponse variable ~ predictor variable1 * predictor variable2)*

```
stp1<-lm(mpg~disp+hp+wt, data=mtcars)
```

The summary:

```
##
## Call:
## lm(formula = mpg ~ disp + hp + wt, data = mtcars)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.891 -1.640 -0.172  1.061  5.861
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.105505   2.110815  17.579 < 2e-16 ***
## disp        -0.000937   0.010350  -0.091  0.92851
## hp          -0.031157   0.011436  -2.724  0.01097 *
## wt          -3.800891   1.066191  -3.565  0.00133 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.639 on 28 degrees of freedom
## Multiple R-squared:  0.8268, Adjusted R-squared:  0.8083
## F-statistic: 44.57 on 3 and 28 DF,  p-value: 8.65e-11
```

The mathematical equation for this is:

$$\text{mpg} = 37.105505 + (-0.000937) * \text{disp} + (-0.031157) * \text{hp} + (-3.800891) * \text{wt}$$

With this equation we are able to predict the mileage per gallon depending on the disp, hp, wt variables.

However, as disp does not have a significance under 0.05, we leave it out in the next step.

```
stp2<-lm(mpg~hp+wt, data=mtcars)
```

The summary:

```
##
## Call:
## lm(formula = mpg ~ hp + wt, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.941 -1.600 -0.182  1.050  5.854
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.22727    1.59879  23.285 < 2e-16 ***
## hp          -0.03177    0.00903  -3.519  0.00145 **
## wt          -3.87783    0.63273  -6.129 1.12e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.593 on 29 degrees of freedom
## Multiple R-squared:  0.8268, Adjusted R-squared:  0.8148
## F-statistic: 69.21 on 2 and 29 DF,  p-value: 9.109e-12
```

Both variables have a significance level of 99% (p-value < 0.01).

Now we will see how much the variables effect mpg not individually but paired.

```
stp3<-lm(mpg~hp*wt,data=mtcars)
```

The summary:

```
##
## Call:
## lm(formula = mpg ~ hp * wt, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0632 -1.6491 -0.7362  1.4211  4.5513
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  49.80842    3.60516   13.816 5.01e-14 ***
## hp          -0.12010    0.02470   -4.863 4.04e-05 ***
## wt          -8.21662    1.26971   -6.471 5.20e-07 ***
## hp:wt         0.02785    0.00742    3.753 0.000811 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.153 on 28 degrees of freedom
## Multiple R-squared:  0.8848, Adjusted R-squared:  0.8724
## F-statistic: 71.66 on 3 and 28 DF,  p-value: 2.981e-13
```

We can see that hp and wt together have a positive relationship to mpg, but the significance level is lower than when calculated individually. Therefore, the interaction between hp and wt can be statistically less significant to consider into the equation.

The final equation of step 3 model looks like this:

$$\text{mpg} = 49.80 + (-0.12010) * \text{hp} + (-8.21662) * \text{wt} + (0.02785) * (\text{hp}:\text{wt})$$

A change in the value of wt would make the biggest change in the value of mpg.