Lec 5 Exercise

Yanzhuo Cao

2023-05-19

Auto data

Let's load the `auto.txt` to start the exploration

```
Auto = read.table("auto.txt", header=T)
```

Model building

Please construct two linear models `lm1` and `lm2`. In `lm1`, regress acceleration on weight only. In `lm2`, regress acceleration on both weight and horsepower. Compare the coefficients for weight in `lm1` and `lm2`. What do you find? (Hint: do heavier cars require more or less time to accelerate from 0 to 60 mph?)

```
lm1 <- lm(acceleration ~ weight, data = Auto)
lm2 <- lm(acceleration ~ weight + horsepower, data = Auto)
coef_lm1 <- coef(lm1)["weight"]
coef_lm2 <- coef(lm2)["weight"]
coef_lm1
```

```
##       weight
## -0.001353896
```

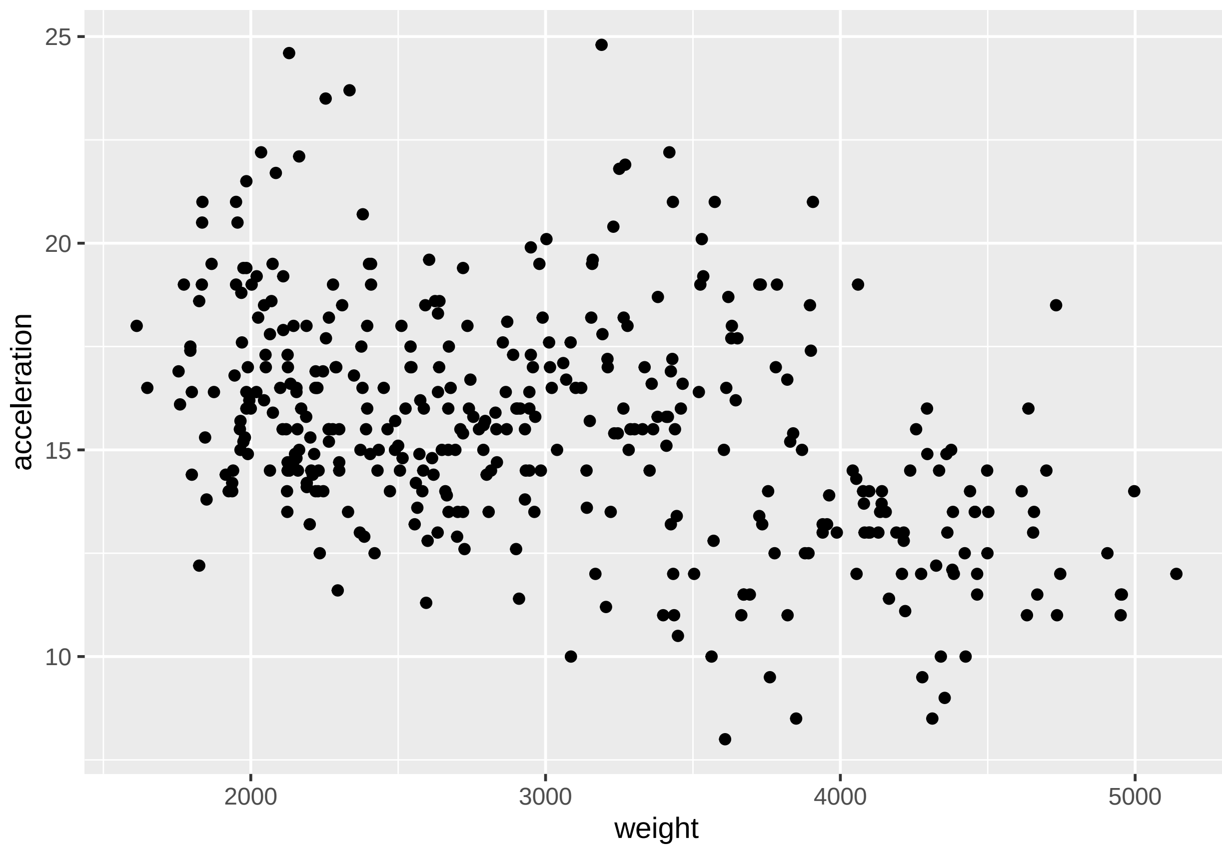
```
coef_lm2
```

```
##     weight
## 0.00230182
```

Answer to the question: When only consider the weight, we can find that the coefficient of weight is negative, which means that the heavier cars will require less time to accelerate. However, if we consider both weight and horsepower, we can find that the coefficient of weight is positive, which means that the heavier cars will require more time to accelerate.

Effect of weight Not Controlling for Other Predictors

```
library(ggplot2)
ggplot(Auto, aes(x=weight, y=acceleration)) + geom_point()
```

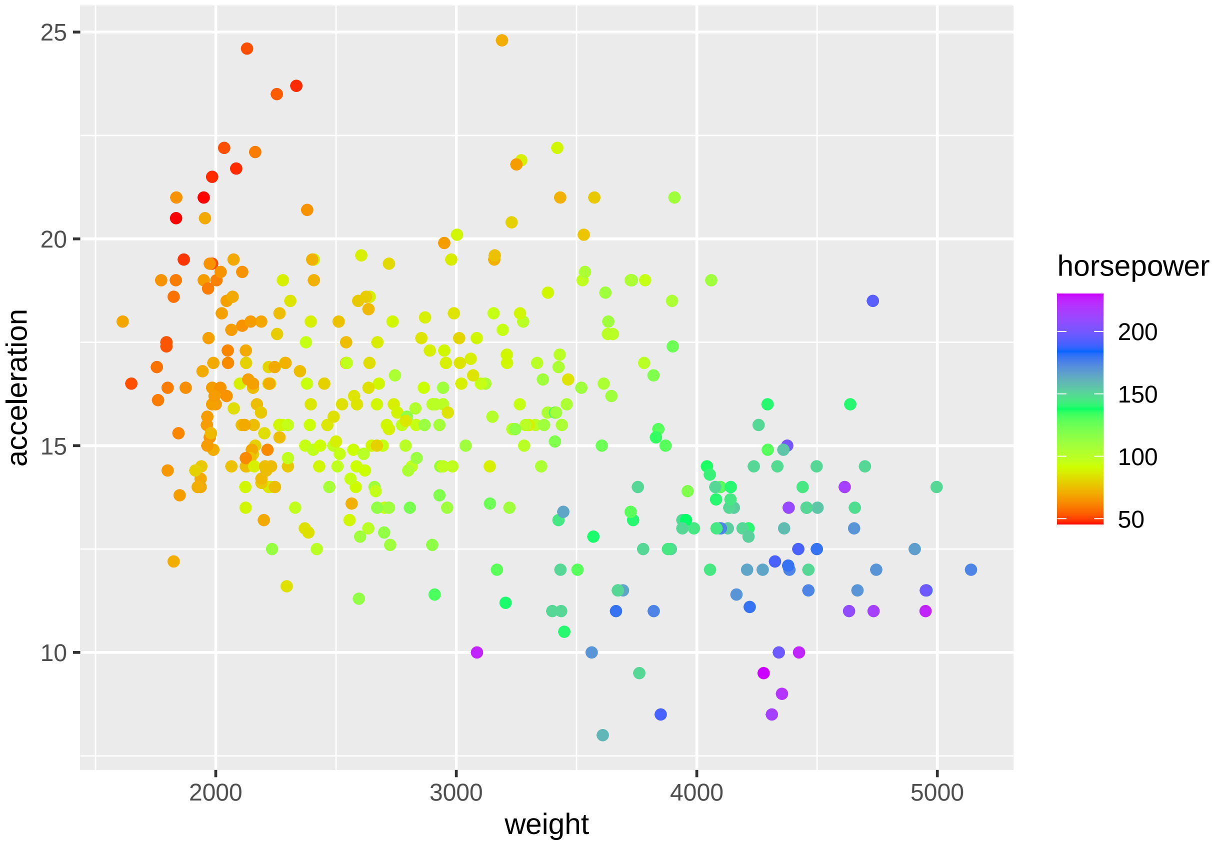


From the scatter plot above, are weight and acceleration are positively or negatively associated? Do heavier vehicles generally require more or less time to accelerate from 0 to 60 mph?

Answer to the question: From the scatter plot above, I think that weight and acceleration are negatively associated. Heavier vehicles generally require less time to accelerate from 0 to 60 mph

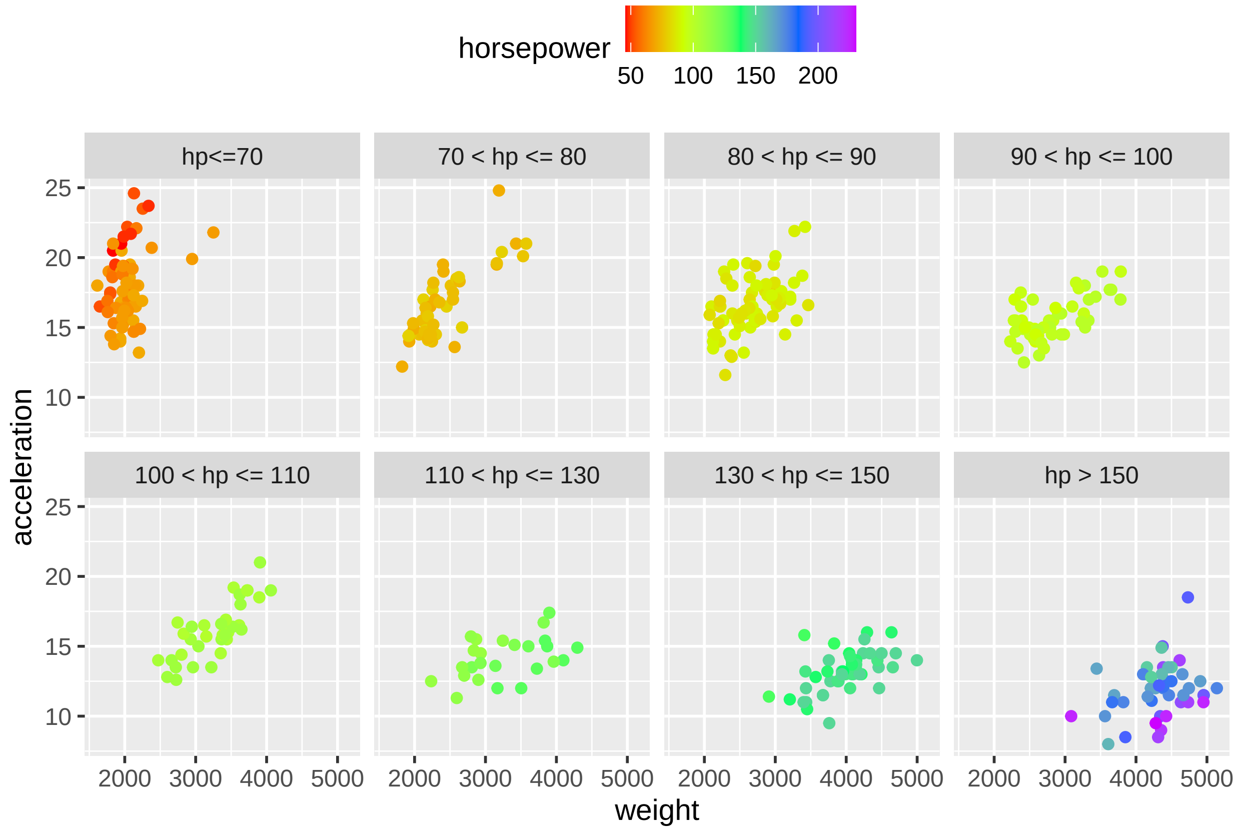
Effect of weight Controlling for horsepower

```
ggplot(Auto, aes(x=weight, y=acceleration, col=horsepower)) +
geom_point() + scale_color_gradientn(colours = rainbow(5))
```



You can also visualize this by creating subplots for data with similar horsepower

```
Auto$hp = cut(Auto$horsepower,
breaks=c(45,70, 80, 90,100,110, 130, 150,230),
labels=c("hp<=70", "70 < hp <= 80", "80 < hp <= 90",
"90 < hp <= 100", "100 < hp <= 110",
"110 < hp <= 130",
"130 < hp <= 150", "hp > 150"))
ggplot(Auto, aes(x=weight, y=acceleration, col=horsepower)) +
geom_point() + scale_color_gradientn(colours = rainbow(5)) +
facet_wrap(~hp, nrow=2) + theme(legend.position="top")
```



Consider car models of similar horsepower (similar color), are weight and acceleration positively or negatively correlated?

Answer to the question: Consider car models of similar horsepower, weight and acceleration are positively correlated.

Now, can you explain why the association between acceleration and weight is flipped from positive to negative when horsepower is ignored?

Let's revisit the problem with another approach

Step 1. Regress acceleration on horsepower. Let RY be the residuals of this model.

```
# Step 1
model1 <- lm(acceleration ~ horsepower, data = Auto)
summary(model1)  # Display model coefficients and statistics
```

```
##
## Call:
## lm(formula = acceleration ~ horsepower, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.9947 -1.2913 -0.1748  1.1229  7.6053
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 20.70193    0.29274   70.72   <2e-16 ***
```

```
## horsepower  -0.04940    0.00263  -18.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.002 on 390 degrees of freedom
## Multiple R-squared:  0.475,  Adjusted R-squared:  0.4736
## F-statistic: 352.8 on 1 and 390 DF,  p-value: < 2.2e-16
```

```r
RY <- residuals(model1)
```

Step 2. Regress weight on horsepower. Let RWT be the residuals of this model.

```r
# Step 2
model2 <- lm(weight ~ horsepower, data = Auto)
summary(model2)  # Display model coefficients and statistics
```

```
##
## Call:
## lm(formula = weight ~ horsepower, data = Auto)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2191.1  -297.7   -80.1   330.8  1150.8
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 984.5003    62.5143   15.75   <2e-16 ***
## horsepower   19.0782     0.5616   33.97   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 427.4 on 390 degrees of freedom
## Multiple R-squared:  0.7474, Adjusted R-squared:  0.7468
## F-statistic:  1154 on 1 and 390 DF,  p-value: < 2.2e-16
```

```r
RWT <- residuals(model2)
```

Step 3. Regress RY on RWT.

```r
# Step 3
model3 <- lm(RY ~ RWT)
summary(model3)  # Display model coefficients and statistics
```

```
##
## Call:
## lm(formula = RY ~ RWT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2802 -1.1236 -0.2544  0.9128  7.1814
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.352e-17  8.804e-02    0.00        1
## RWT         2.302e-03  2.065e-04   11.15   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.743 on 390 degrees of freedom
## Multiple R-squared:  0.2416, Adjusted R-squared:  0.2397
## F-statistic: 124.3 on 1 and 390 DF,  p-value: < 2.2e-16
```

Does any of the coefficient above look familiar? Explain your findings!

Answer to the question: In Step 1, the regression model between acceleration and horsepower indicates that the coefficient of horsepower is -0.04940. This is a negative correlation, which is consistent with our intuitive understanding of the acceleration performance of the car.

In Step 2, the regression model between weight and horsepower shows a coefficient of 19.0782 for horsepower. This is a positive relationship, implying an association between higher horsepower and heavier cars.

In Step 3, the regression model of RY (residual from step 1) to RWT (residual from step 2) shows a coefficient of 2.302e-03 for RWT. However, the coefficient of this association is very small, close to zero, and thus not very practical.

Taken together, these coefficients provide some information on the relationship between horsepower, weight, and acceleration. They support our common sense understanding of car performance that horsepower has a positive effect on acceleration, while weight has a negative effect on acceleration. At the same time, the coefficients in Step 3 indicate that the association between residuals is relatively weak after accounting for the effects of horsepower and weight, indicating that the explanatory power of the model may be limited.