Motor Trend MPG Analysis

Michael Krämer

25.10.2015

Abstract

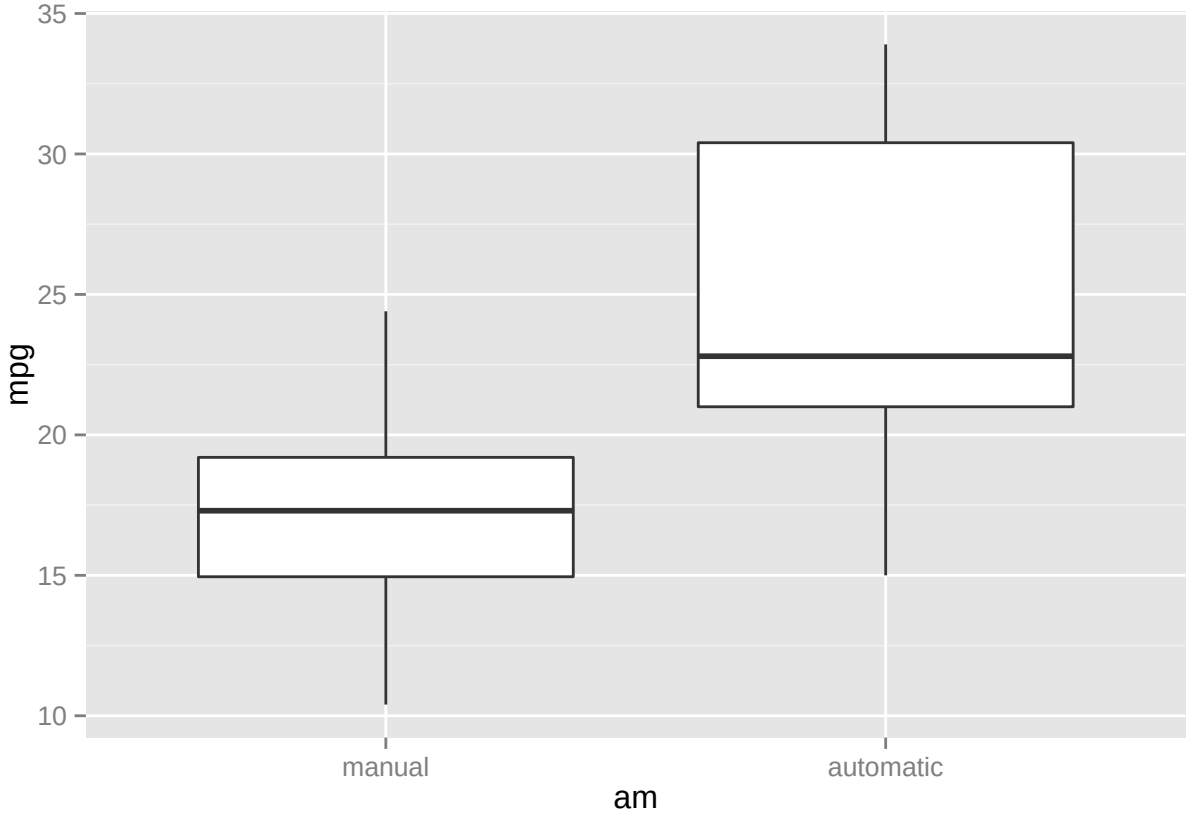
The report shows changes on mpg for cars, especially the influence of the transmission type.

General data description

The data is the mtcars data set from the dataset library. The variables can be seen in R help.

Exploratory analysis

The interest is primaryly into the mpg of the cars of the data set. Since the transmission type is of interest, too, let's have a look at this.



Regression model

The coefficients of all variables against mpg.

```r
summary(lm(mpg ~ ., data=data))$coefficients
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## amautomatic  2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

As seen from the t-values, the weight (wt) seems to be the best candidate as a predictor for mpg. Since the transmission type is of interest, too, let's fit a model with only these 2 predictors.

```r
summary(lm(mpg ~ wt + am, data=data))$coefficients
```

```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
## amautomatic -0.02361522  1.5456453 -0.01527855 9.879146e-01
```

Unfortunately, the relation to am shows by the near-zero t value almost no relation in this regression context. I assume this has to do with the transmission type only having 2 levels. The data shows still that the cars with automatic transmission in the sample have larger mpg as seen in the figure. However, the data supports the statement that this is more due to their weight.

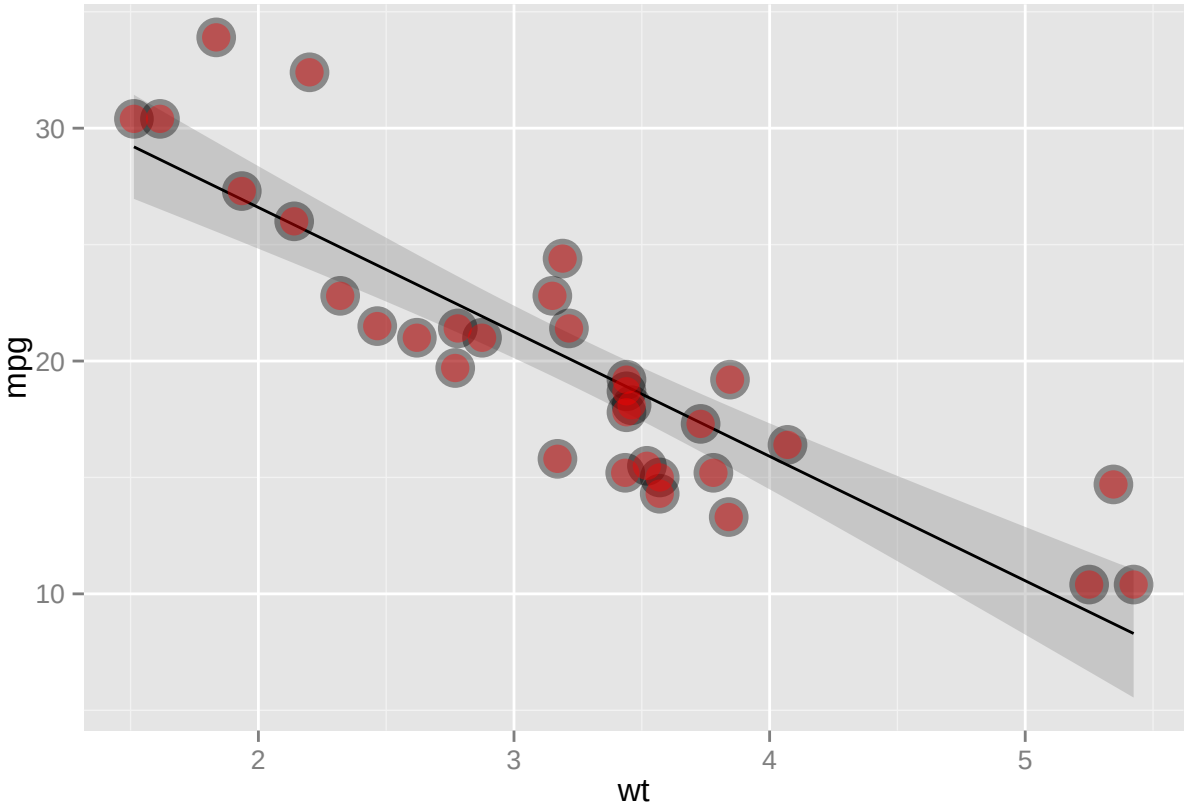
```r
fit <- lm(mpg ~ wt, data=data)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## wt           -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

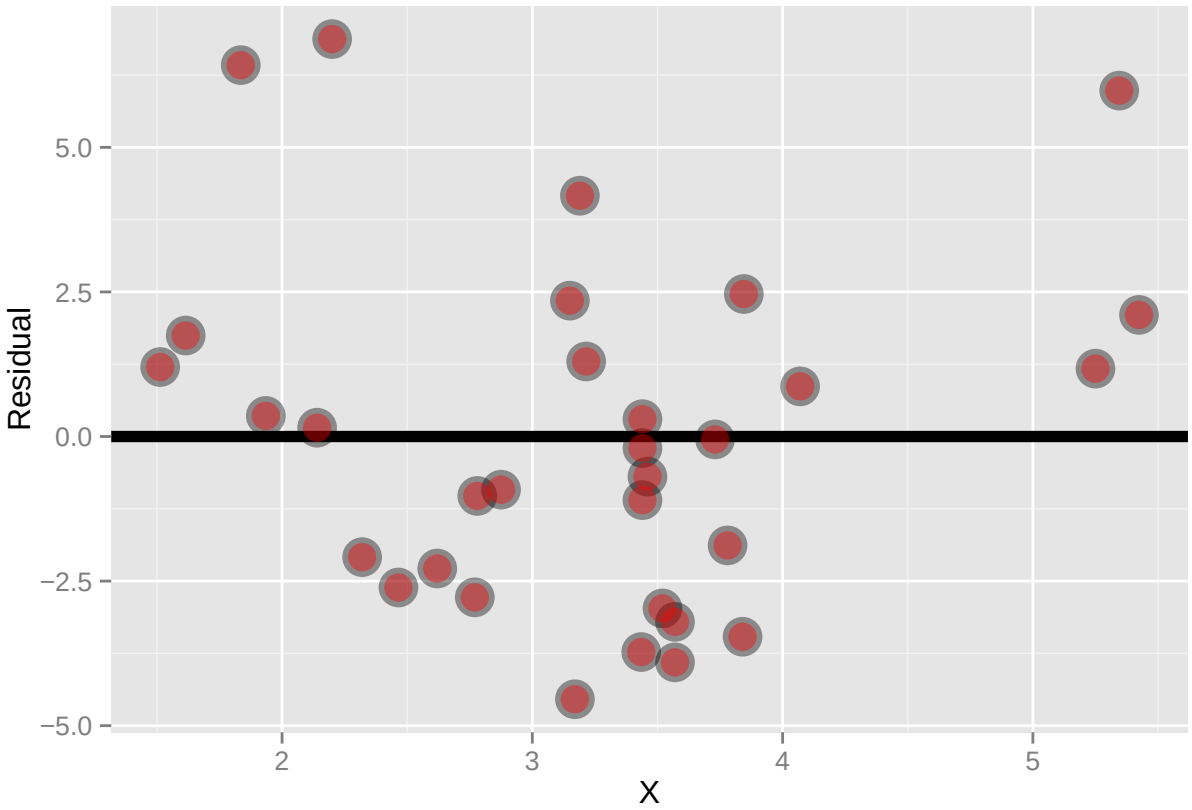
The model with only weight as a predictor seems a good fit. R-squared shows it explains around 75% of the variance of mpg.

Residuals

The fitted linear model with only weight as a predictor.



And the appropriate residuals of that.



The residuals show no particular pattern which is a good sign.

Conclusions

I was not able to show a particular relation from transmission type to mpg, since I think the difference seen in the data is explained better by other influences like the weight.

Sources

The sources for the report can be found on Github at https://github.com/mkraemerx/datasciencecoursera/tree/master/07Regression