Moter Trend - The relationship between a set of variables and miles per gallon

Feng Zhaoxu

08 Jun 2019

```
library(ggplot2)
library(GGally)
library(dplyr)
library(ggfortify)

data(mtcars)

mtcarsFactors <- mtcars
mtcarsFactors$am <- as.factor(mtcarsFactors$am)
levels(mtcarsFactors$am) <- c("automatic", "manual")

mtcarsFactors$cyl <- as.factor(mtcarsFactors$cyl)
mtcarsFactors$gear <- as.factor(mtcarsFactors$gear)
mtcarsFactors$vs <- as.factor(mtcarsFactors$vs)
levels(mtcarsFactors$vs) <- c("V", "S")
```

Exploratory data analyses

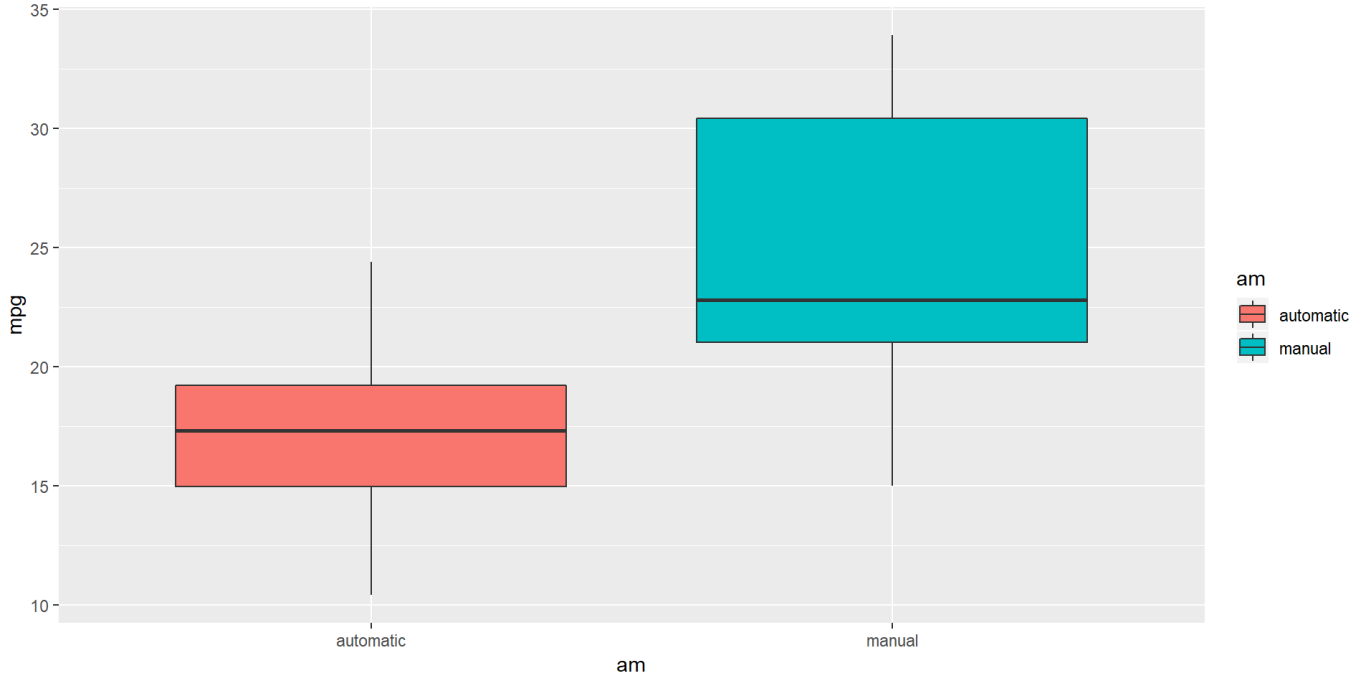
```
# Res 1
dim(mtcarsFactors)
```

```
## [1] 32 11
```

```
# Res 2
head(mtcarsFactors)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs        am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  V    manual    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  V    manual    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  S    manual    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  S automatic    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  V automatic    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  S automatic    3    1
```

```
# Figure 1
library(ggplot2)
p <- ggplot(mtcarsFactors, aes(am, mpg))
p + geom_boxplot(aes(fill = am))
```



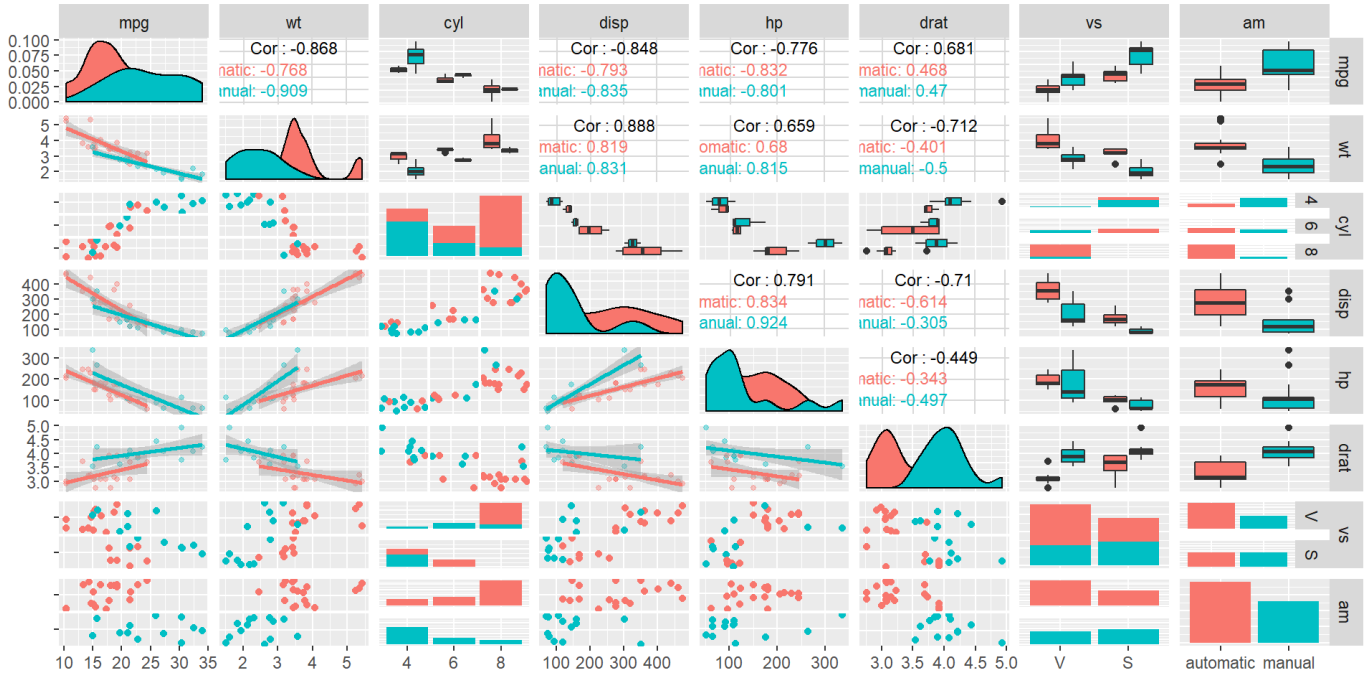

```
# Res 3
cors <- cor(mtcars$mpg, mtcars)
orderedCors <- cors[,order(-abs(cors[1,]))]
orderedCors
```

```
##        mpg         wt        cyl       disp         hp       drat         vs         am
carb       gear
##  1.0000000 -0.8676594 -0.8521620 -0.8475514 -0.7761684  0.6811719  0.6640389  0.5998324 -
0.5509251  0.4802848
##       qsec
##  0.4186840
```

```
# Res 4
amPos <- which(names(orderedCors)=="am")
subsetColumns <- names(orderedCors)[1:amPos]
subsetColumns
```

```
## [1] "mpg"  "wt"   "cyl"  "disp" "hp"   "drat" "vs"   "am"
```

```
# Figure 2
mtcarsFactors[,subsetColumns] %>%
   ggpairs(
     mapping = ggplot2::aes(color = am),
     upper = list(continuous = wrap("cor", size = 3)),
     lower = list(continuous = wrap("smooth", alpha=0.4, size=1), combo = wrap("dot"))
   )
```



Model selection

First we start with the basic model

```
# Res 5
basicFit <- lm(mpg ~ am, mtcarsFactors)
summary(basicFit)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcarsFactors)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## ammanual       7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

```
# Res 6
totalFit <- lm(mpg ~ ., mtcarsFactors)
summary(totalFit)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcarsFactors)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2015 -1.2319  0.1033  1.1953  4.3085
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.09262   17.13627   0.881   0.3895
## cyl6        -1.19940    2.38736  -0.502   0.6212
## cyl8         3.05492    4.82987   0.633   0.5346
## disp         0.01257    0.01774   0.708   0.4873
## hp          -0.05712    0.03175  -1.799   0.0879 .
## drat         0.73577    1.98461   0.371   0.7149
## wt          -3.54512    1.90895  -1.857   0.0789 .
## qsec         0.76801    0.75222   1.021   0.3201
## vsS          2.48849    2.54015   0.980   0.3396
## ammanual     3.34736    2.28948   1.462   0.1601
## gear4       -0.99922    2.94658  -0.339   0.7382
## gear5        1.06455    3.02730   0.352   0.7290
## carb         0.78703    1.03599   0.760   0.4568
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.616 on 19 degrees of freedom
## Multiple R-squared:  0.8845, Adjusted R-squared:  0.8116
## F-statistic: 12.13 on 12 and 19 DF,  p-value: 1.764e-06
```

```
# Res 7
bestFit <- step(totalFit,direction="both",trace=FALSE)
summary(bestFit)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcarsFactors)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## ammanual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Model examination

```
# Figure 3
autoplot(bestFit)
```