Coursera Regression Models: Course Project

Michael Nichols March 10, 2018 —

Executive Summary

Findings of the evaluation of the mtcars dataset to explore the relationship between a set of variables and miles per gallon (MPG) (outcome): 1) Manual transmission cars tend to have better MPG than automatic transmission cars, but other variables must also be considered in order to draw worthwhile solutions ~ most significantly, weight and quarter mile time. 2) On average, manual transmission cars average 2.94 MPG more than automatic transmission cars, considering most significant other variables. ###Setup

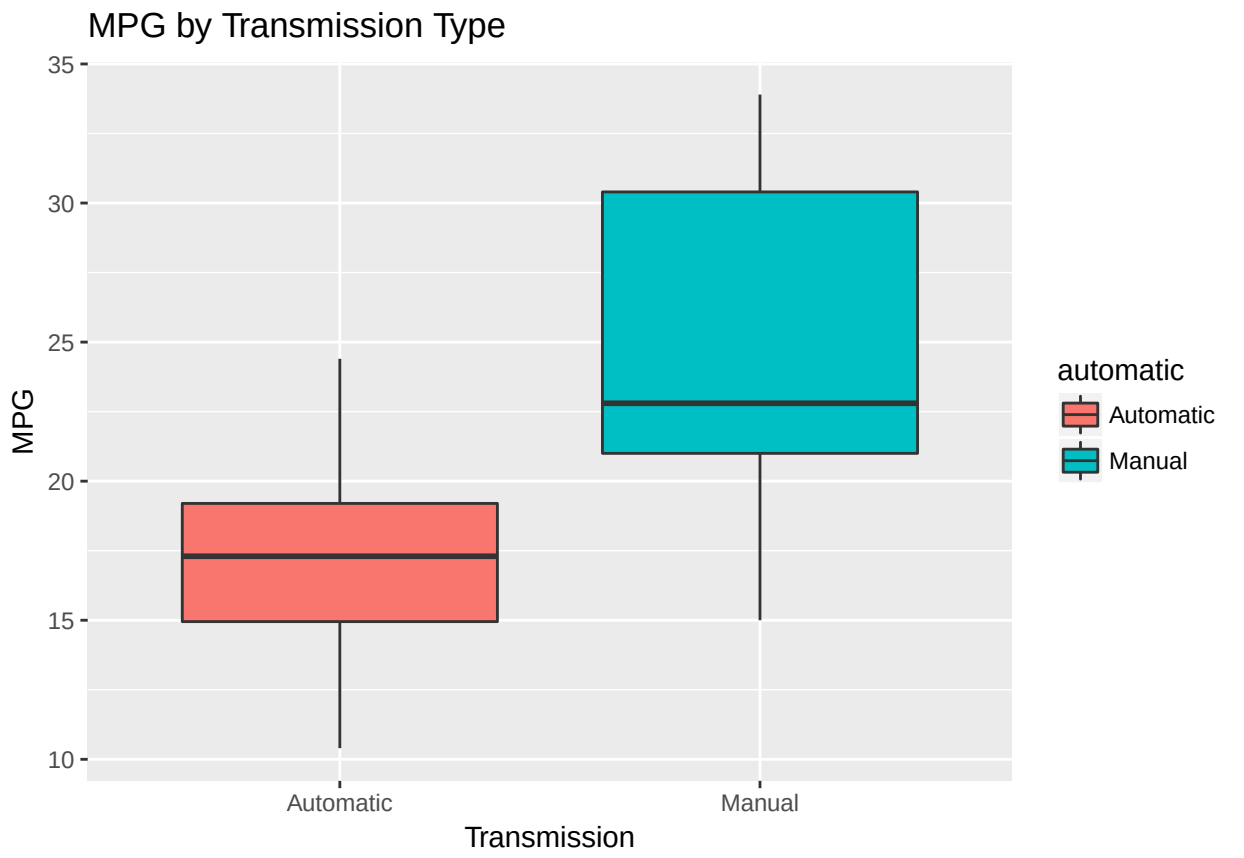
```
#load R packages
library(ggplot2) # for exploratory data analyis / visuals
#load dataset
data(mtcars)
#display top 6 rows
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
#update 'am' variable as automatic vs transmission and set to factor variable
#from the RStudio's documentation: am = Transmission (0 = automatic, 1 = manual)
mtcars$automatic <- as.factor(mtcars$am)
levels(mtcars$automatic) <-c("Automatic", "Manual")
#remove original transmission variable
mtcars$am <- NULL
```

Exploratory Data Analysis

```
#Create boxplot to show average mpg by transmission type.
ggplot(mtcars, aes(x = automatic, y = mpg, fill = automatic)) +
  geom_boxplot(notch = F) +
  scale_x_discrete("Transmission") +
  scale_y_continuous("MPG") +
  ggtitle("MPG by Transmission Type")
```



Clear tendency for manual transmission cars to have a higher MPG. However, this does not consider other variables that may also be playing a factor. ### Analysis

```
#simple linear regression
fit <- lm(mpg ~ automatic, data = mtcars)
summary(fit)$coef
```

```
##                 Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)     17.147368   1.124603 15.247492 1.133983e-15
## automaticManual  7.244939   1.764422  4.106127 2.850207e-04
```

```
summary(fit)$r.squared
```

```
## [1] 0.3597989
```

Interpretation: Coefficient and intercepts indicate on average manual transmission cars have 7.245 mpg more than automatic transmission cars. R^2 value = 0.3598; indicating the model explains 35.98% of the variance, which is not sufficient. Without further variance explained, we need to conduct further tests.

```
#conduct basic t test to determine whether one transmission is better in terms of MPG
ttestMPG <- t.test(mpg ~ automatic, data = mtcars)
round(ttestMPG$p.value,4)
```

```
## [1] 0.0014
```

```
#Based on the pvalue above (0.0014) from the initial basic test, we would reject the null hypothesis th

#Identify most significant other variables, by testing correlation to mpg of all variables.
#revert transmission type to numeric for correlation test
mtcars$automatic <- as.numeric(mtcars$automatic)
```

```r
sort(abs(cor(mtcars)[1,]), decreasing = TRUE)
```

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

```r
#Based on the correlations between each variable and mpg, transmission type is 6th most significant. We
```

```r
fitAll <- lm(mpg ~ ., data = mtcars)
#Run the step function (stepwise regression) to find the best choice of predictors.
bestFit <- step(fitAll, trace = 0, direction = "both")
#provide the summary
summary(bestFit)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  6.681943  8.0101071  0.834189 4.112351e-01
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
## automatic    2.935837  1.4109045  2.080819 4.671551e-02
```

```r
summary(bestFit)$r.squared
```

```
## [1] 0.8496636
```

The resulting bestFit model evaluates mpg based on wt (weight), qsec (1/4 mile times), and automatic (transmimssion type).

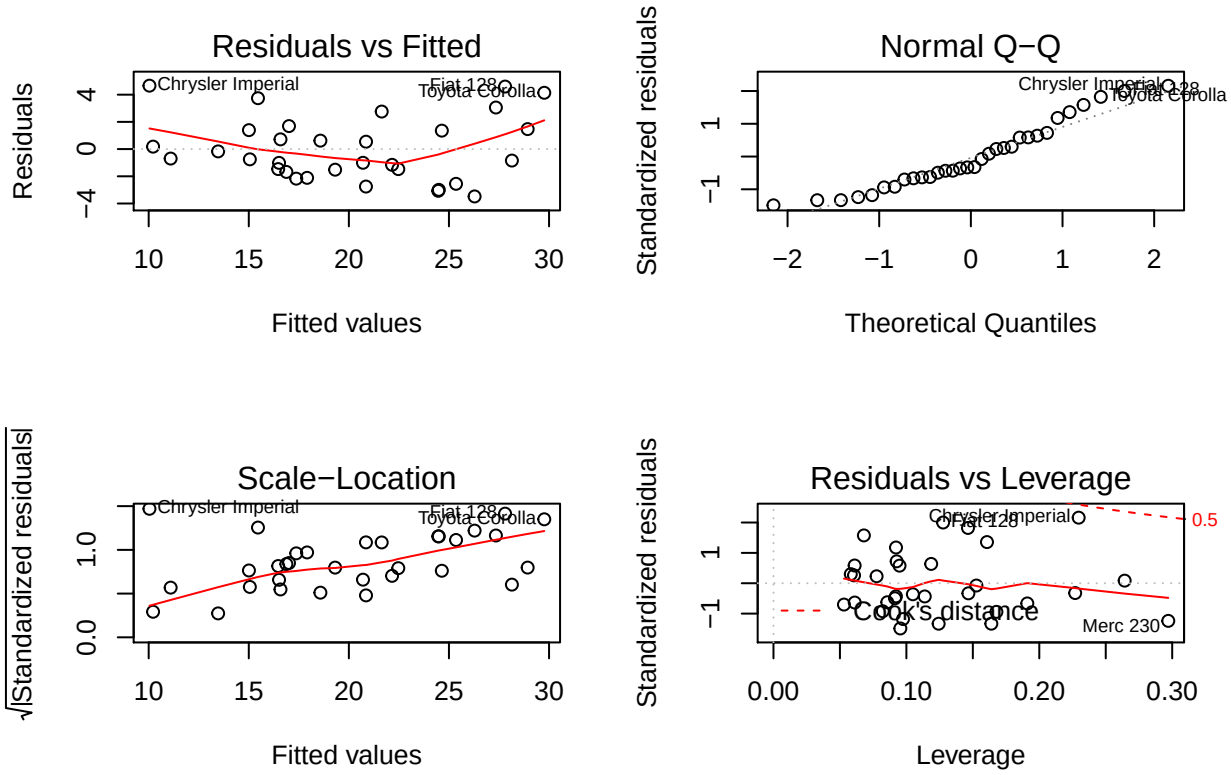
Best Fit Model R^2: 84.97%, meaning the model captures 84.97% of the variance, far more respectable than the 36% captured by transmission type alone.

```r
summary(bestFit)$coef[4]
```

```
## [1] 2.935837
```

```r
#On average, manual transmission cars average 2.94 MPG more than automatic transmission cars, consideri

#Display residual diagnostics using the best fit model (mpg ~ wt + qsec + automatic).
par(mfrow=c(2,2))
plot(bestFit)
```

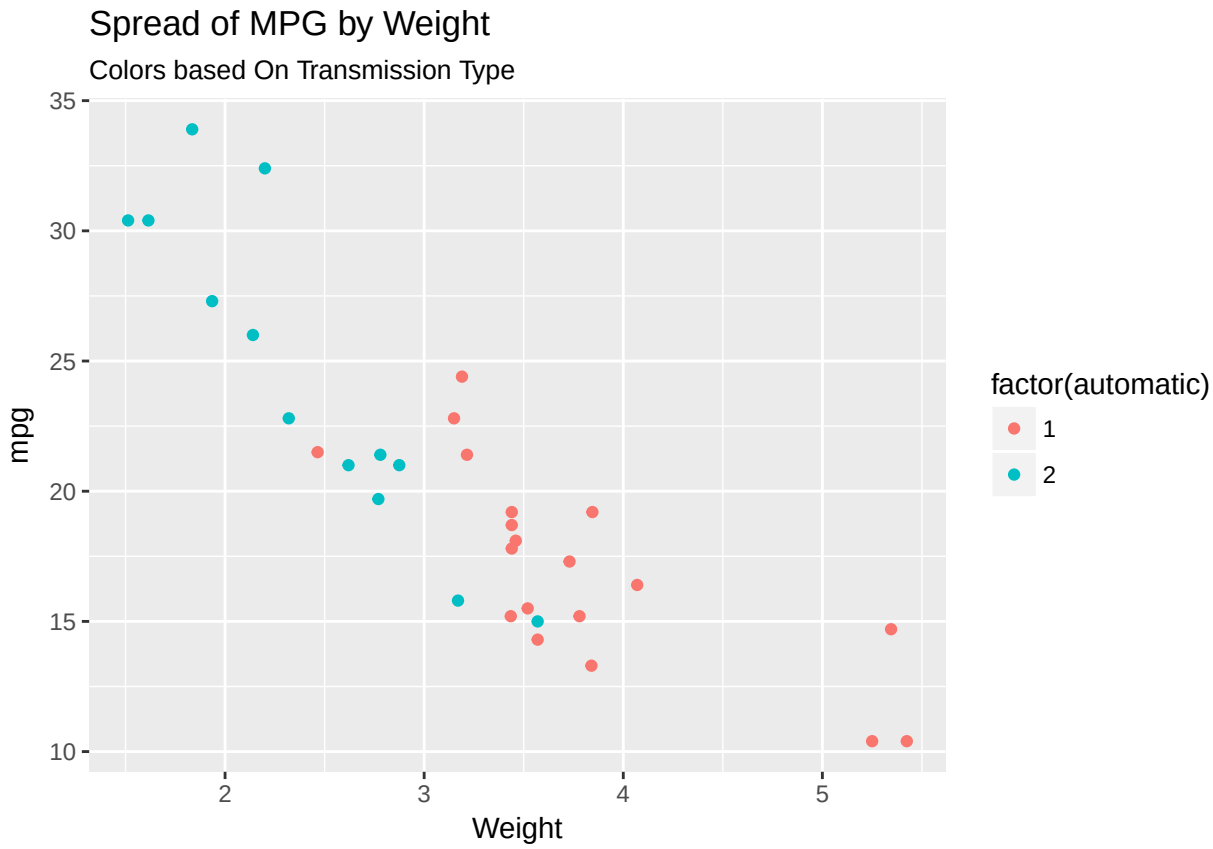


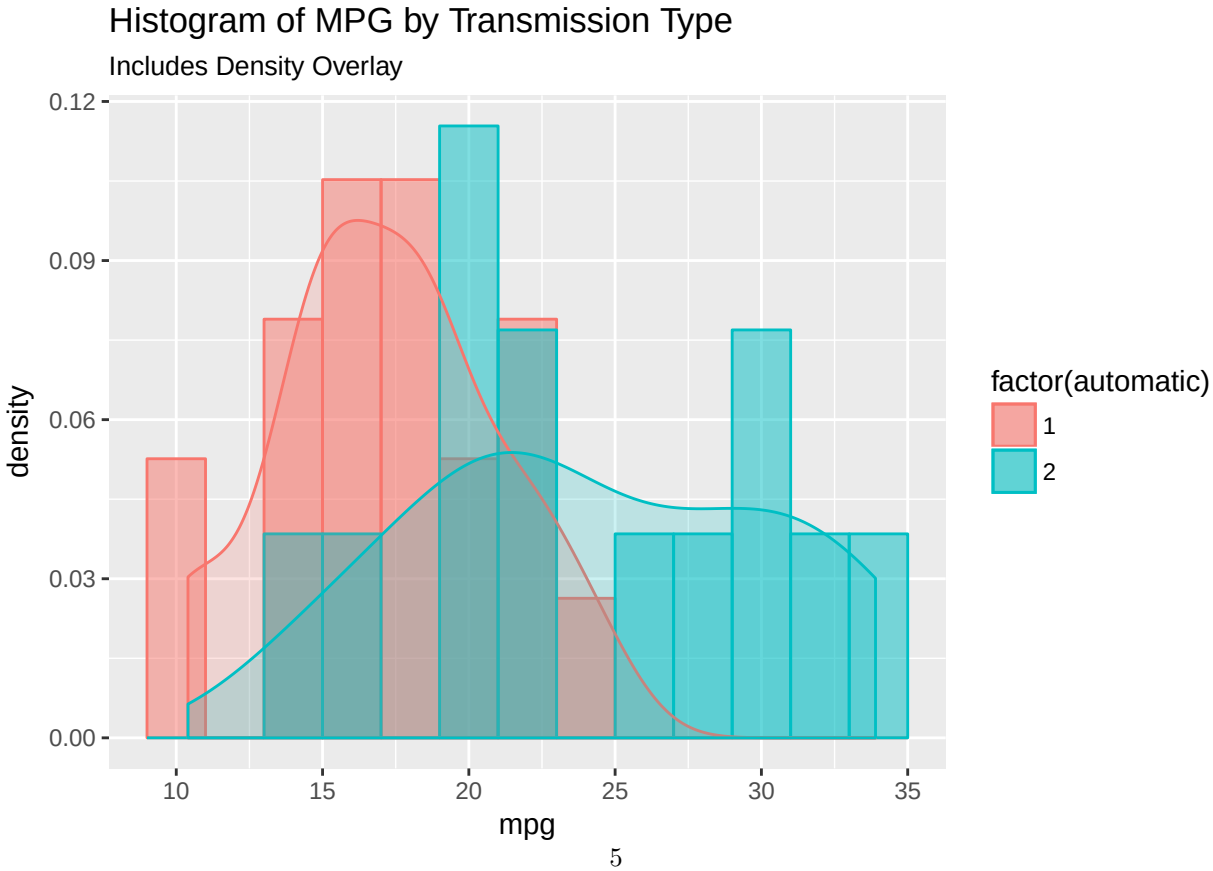

```
#Interpretation:
# - Residual vs fitted plot shows slight bend,indicating some non-linear relationship was not explained
# - Residuals appear normally distributed in the Normal Q-Q plot, as desired.
```

Conclusion

To recap, manual cars tend to have better MPG than automatic cars. However, by incorporating other variables, most significantly weight and quarter-mile times, you can imperfectly, yet much more effectively predict MPG.

Appendix


Spread of MPG by Weight
Colors based On Transmission Type
mpg
Weight
factor(automatic)
1
2


Histogram of MPG by Transmission Type
Includes Density Overlay
density
mpg
factor(automatic)
1
2