Effect of car transmission type on MPG

Executive Summary

In this report, we are to (1) examine whether an automatic or manual transmission is better for MPG and (2) quantify the MPG difference between the two types of transmissions. After performing a multiple linear regression to find the best-fit model, we conclude that the type of transmission does not actually directly impact MPG.

Data Loading & Preparation

The data is retrieved from the `mtcars` data set provided by the `datasets` package.

```
data(mtcars)
```

Exploratory Data Analysis

The `mpg` data was plotted against all other variables (see Figure 1 in the appendix). An initial overview of the `mpg ~ am` boxplot seem to suggest that cars with automatic transmissions seem to achieve less miles per gallon than cars with manual transmissions. However, the plots of the other data also show some correlation of `mpg` with the other variables available, suggesting the presence of one or more confounding variables.

Model fitting

Variable selection

A measurement of each variable's correlation with `mpg` reveals that there is an especially strong negative correlation with `cyl`, `disp` and `wt`:

```
cor(mtcars$mpg, mtcars)
```

```
##      mpg       cyl       disp         hp      drat         wt     qsec
## [1,]   1 -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##             vs        am      gear       carb
## [1,] 0.6640389 0.5998324 0.4802848 -0.5509251
```

Based of this initial exploration, we will focus our efforts on `cyl`, `disp` and `wt`.

Model testing

```
fit1 <- lm(mpg ~ am, data = mtcars)
fit2 <- lm(mpg ~ am + wt, data = mtcars)
fit3 <- lm(mpg ~ am + wt + cyl, data = mtcars)
fit4 <- lm(mpg ~ am + wt + cyl + disp, data = mtcars)
anova(fit1, fit2, fit3, fit4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + cyl
## Model 4: mpg ~ am + wt + cyl + disp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 63.4179 1.469e-08 ***
## 3     28 191.05  1     87.27 12.5055  0.001488 **
## 4     27 188.43  1      2.62  0.3756  0.545093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From this, we can see that `wt` and `cyl` are both significant whereas `disp` is not, based on the p-values in a nested model Anova analysis with an alpha of 0.05.

Now, let's test the significance of `am` to our model:

```
fit1 <- lm(mpg ~ wt + cyl, data = mtcars)
fit2 <- lm(mpg ~ wt + cyl + am, data = mtcars)
anova(fit1, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + cyl
## Model 2: mpg ~ wt + cyl + am
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     29 191.17
## 2     28 191.05  1   0.12491 0.0183 0.8933
```

Interestingly, it appears that `am` is not significant, with a p-value significantly exceeding our alpha of 0.05. If we plot `mpg` against `wt` and highlight `cyl` and `am` in the data points, we see that both `cyl` and `am` are proxy indicators of `wt`. The heavier a car, the more cylinders in the engine (see Figure 2). Also, cars with automatic transmissions are heavier than cars with manual transmissions (see Figure 3).

Therefore, the model most illustrative of effect on a car's MPG would be:

```
summary(lm(mpg ~ wt, data = mtcars))$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 37.285126   1.877627 19.857575 8.241799e-19
## wt          -5.344472   0.559101 -9.559044 1.293959e-10
```

See Figure 4 for residual and diagnostic plots.

Conclusion

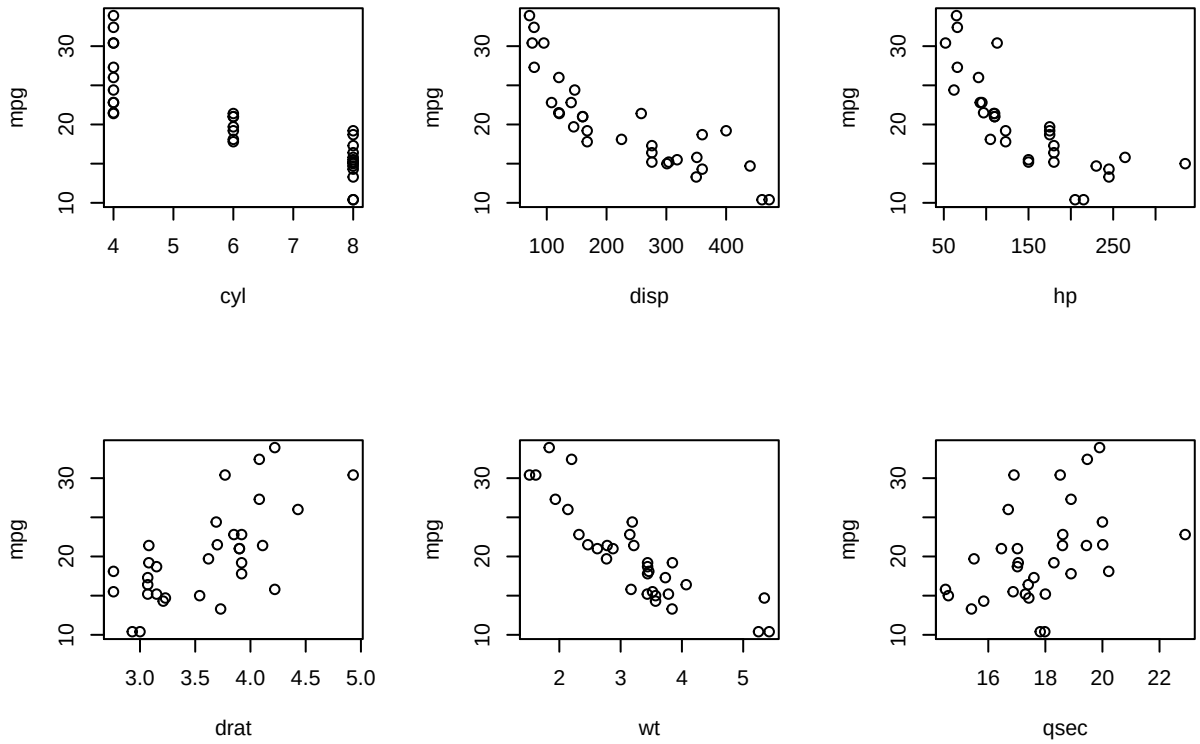
Based on our findings, we can conclude that manual transmissions are better than automatic transmissions for MPG. However, this is not because of the transmission type, but instead because of the weight of the car.

As for quantifying the MPG difference for weight, the true regressor, each 1000 lb increase in weight of a car results in about 5.3 less MPG.

Appendix

Figure 1

```
par(mfrow = c(2,3))
plot(mpg ~ ., data = mtcars)
```



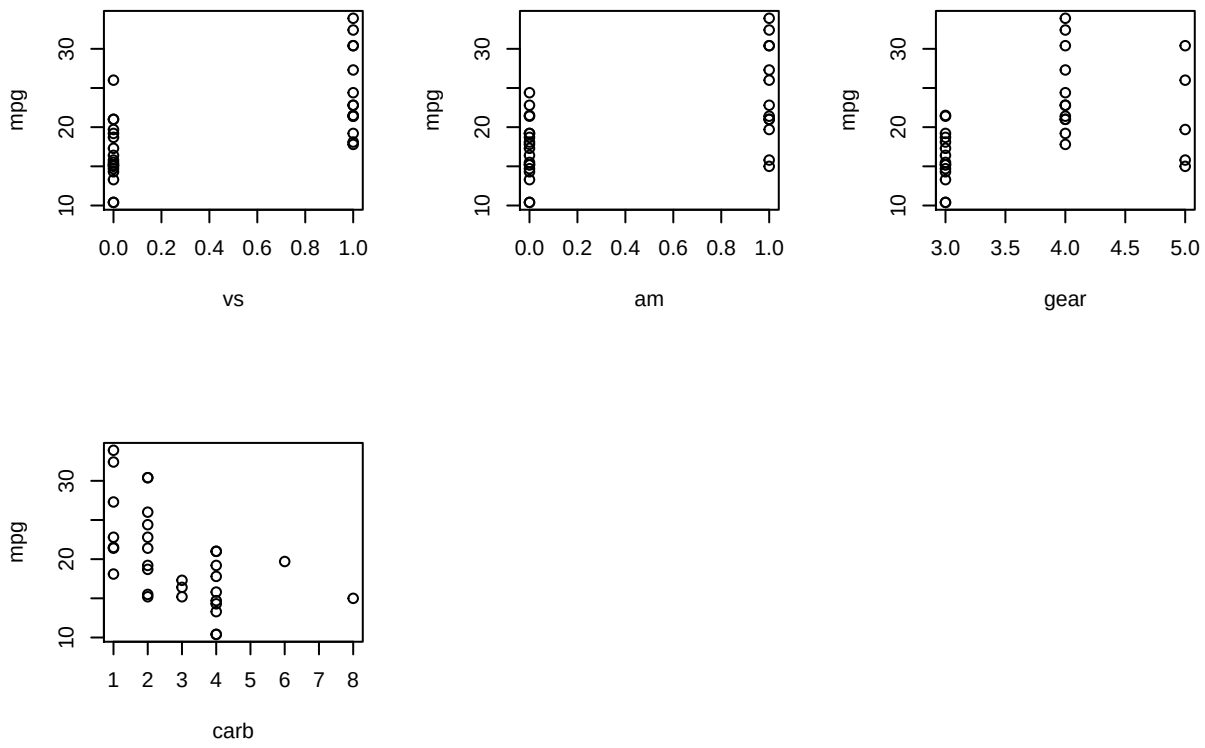


Figure 2

```
plot(mtcars$wt, mtcars$mpg, pch=as.numeric(mtcars$cyl))
legend("topright", legend = levels(as.factor(mtcars$cyl)), pch=c(4,6,8))
```

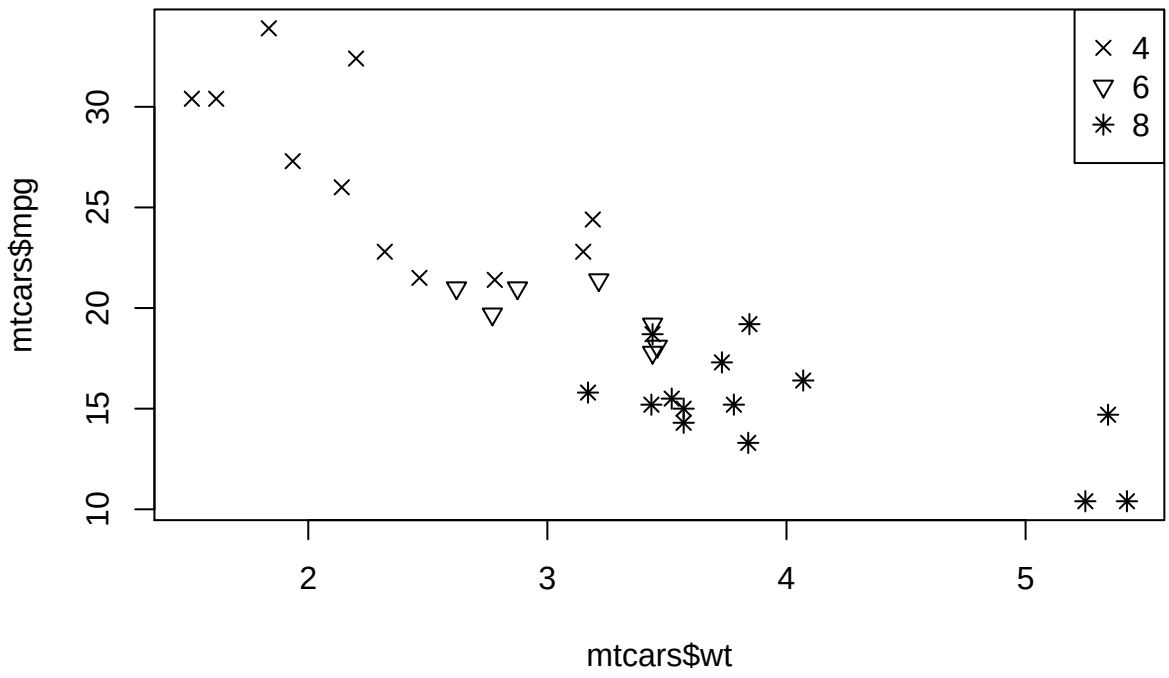


Figure 3

```
plot(mtcars$wt, mtcars$mpg, pch = mtcars$am)
legend("topright", legend = c("automatic", "manual"), pch=as.numeric(levels(as.factor(mtcars$am))))
```

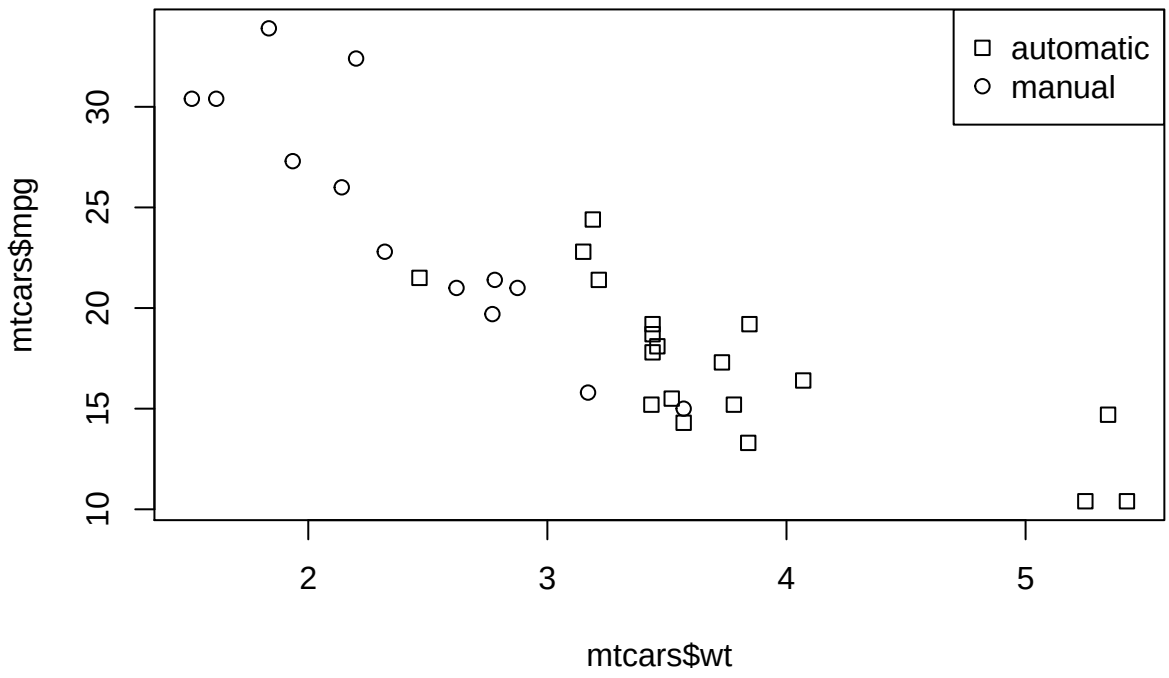
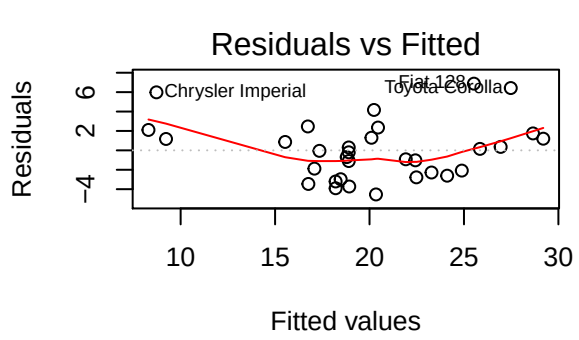
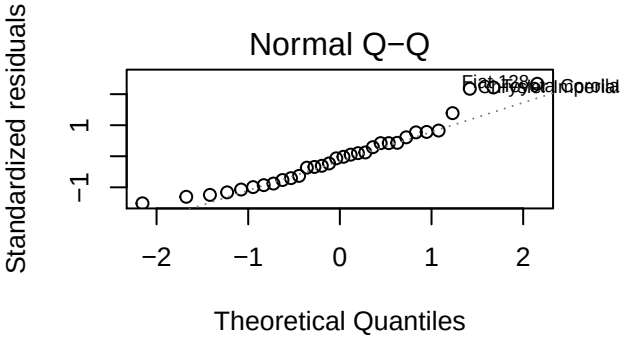
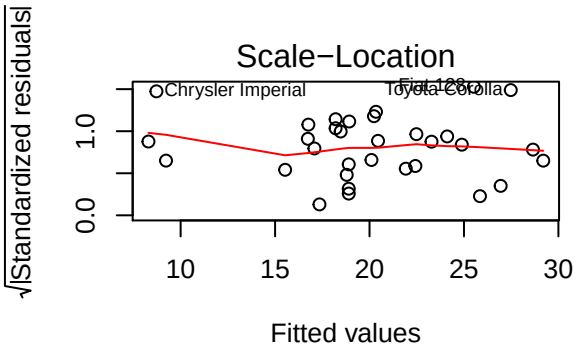


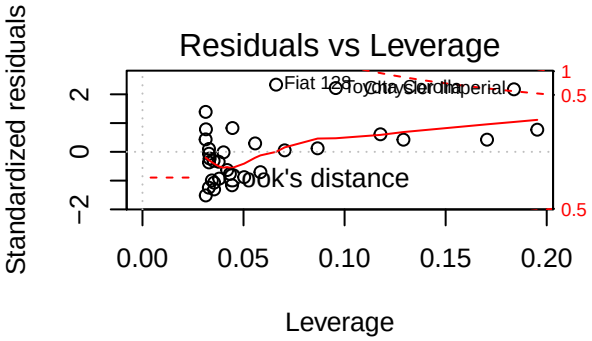
Figure 4

```
model <- lm(mpg ~ wt, data = mtcars)
par(mfrow = c(2, 2)); plot(model)
```


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128


Normal Q−Q
Standardized residuals
Theoretical Quantiles
Fiat 128


Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
Fiat 128
Chrysler Imperial
Toyota Corolla