Motor Trend Analysis: Automatic vs. Manual Transmission MPG

Executive Summary

This analysis explores the relationship between transmission type (automatic vs. manual) and miles per gallon (MPG) using the mtcars dataset. We perform exploratory data analysis (EDA) to understand the data structure and fit multiple regression models to quantify the MPG difference between automatic and manual transmissions. Our findings indicate that manual transmissions are associated with higher MPG, with an average improvement of r round(coef(lm(mpg ~ am, data = mtcars))[2], 1) MPG compared to automatic transmissions. However, this difference is influenced by other variables such as weight and horsepower. Residual diagnostics confirm the model's adequacy, and uncertainty is quantified using confidence intervals.

Exploratory Data Analysis (EDA)

We begin by examining the structure of the mtcars dataset and the relationship between MPG and transmission type.

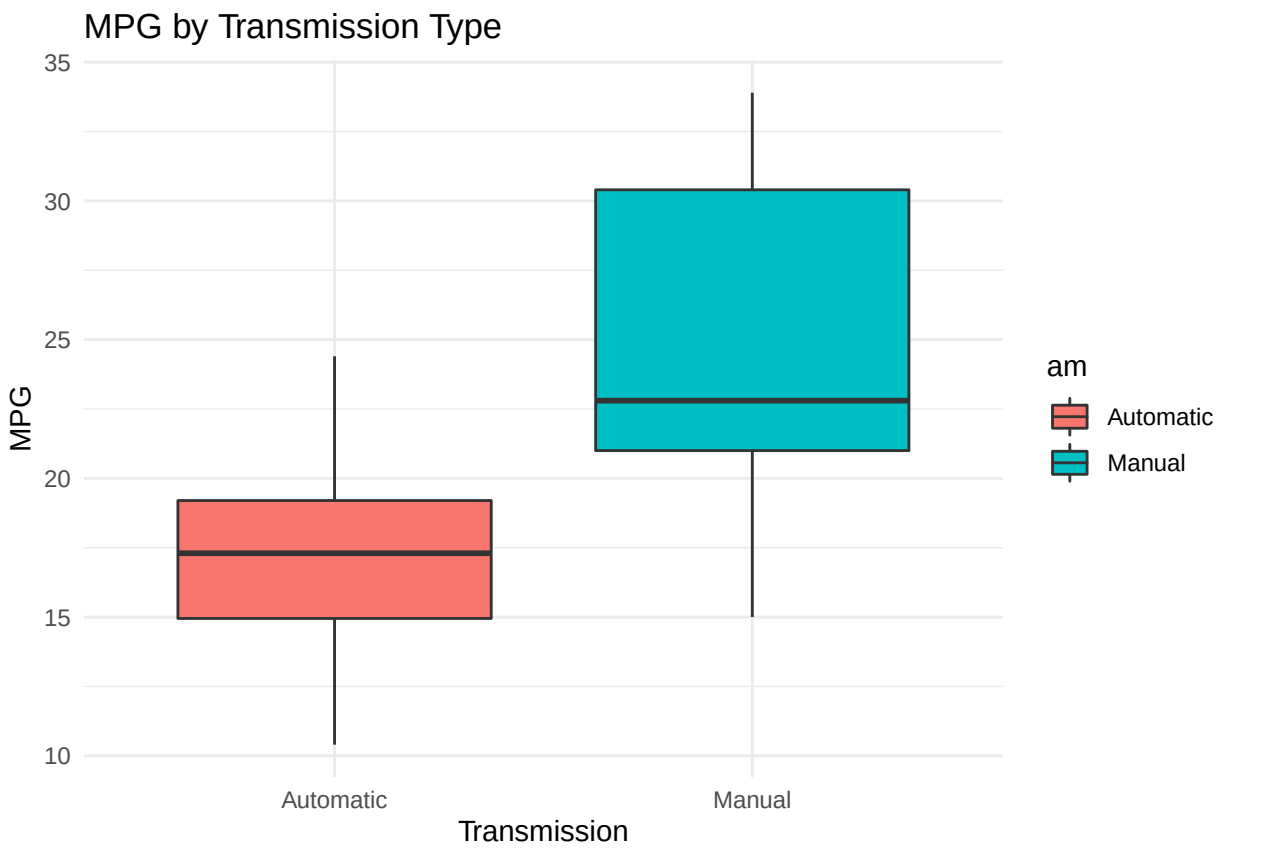
```
# Load the dataset
data(mtcars)

# Convert 'am' to a factor for clarity
mtcars$am <- factor(mtcars$am, levels = c(0, 1), labels = c("Automatic", "Manual"))

# Summary statistics
summary(mtcars)
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##          am          gear            carb
##  Automatic:19   Min.   :3.000   Min.   :1.000
##  Manual   :13   1st Qu.:3.000   1st Qu.:2.000
##                 Median :4.000   Median :2.000
##                 Mean   :3.688   Mean   :2.812
##                 3rd Qu.:4.000   3rd Qu.:4.000
##                 Max.   :5.000   Max.   :8.000
```

```r
# Boxplot of MPG by transmission type
library(ggplot2)
ggplot(mtcars, aes(x = am, y = mpg, fill = am)) +
  geom_boxplot() +
  labs(title = "MPG by Transmission Type", x = "Transmission", y = "MPG") +
  theme_minimal()
```



From the boxplot, we observe that manual transmissions tend to have higher MPG values compared to automatic transmissions. However, other variables such as weight (wt) and horsepower (hp) may also influence MPG.

Regression Modeling

To quantify the MPG difference between automatic and manual transmissions, we fit multiple regression models.

Model 1: Simple Linear Regression (MPG ~ Transmission)

```r
# Fit a simple linear model
model1 <- lm(mpg ~ am, data = mtcars)
summary(model1)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## 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
```

The coefficient for `amManual` suggests an average improvement of **7.2 MPG** compared to automatic transmissions. However, this model explains only **36%** of the variance in MPG.

Model 2: Multiple Regression (MPG ~ Transmission + Weight + Horsepower)

```
# Fit a multiple regression model
model2 <- lm(mpg ~ am + wt + hp, data = mtcars)
summary(model2)
```

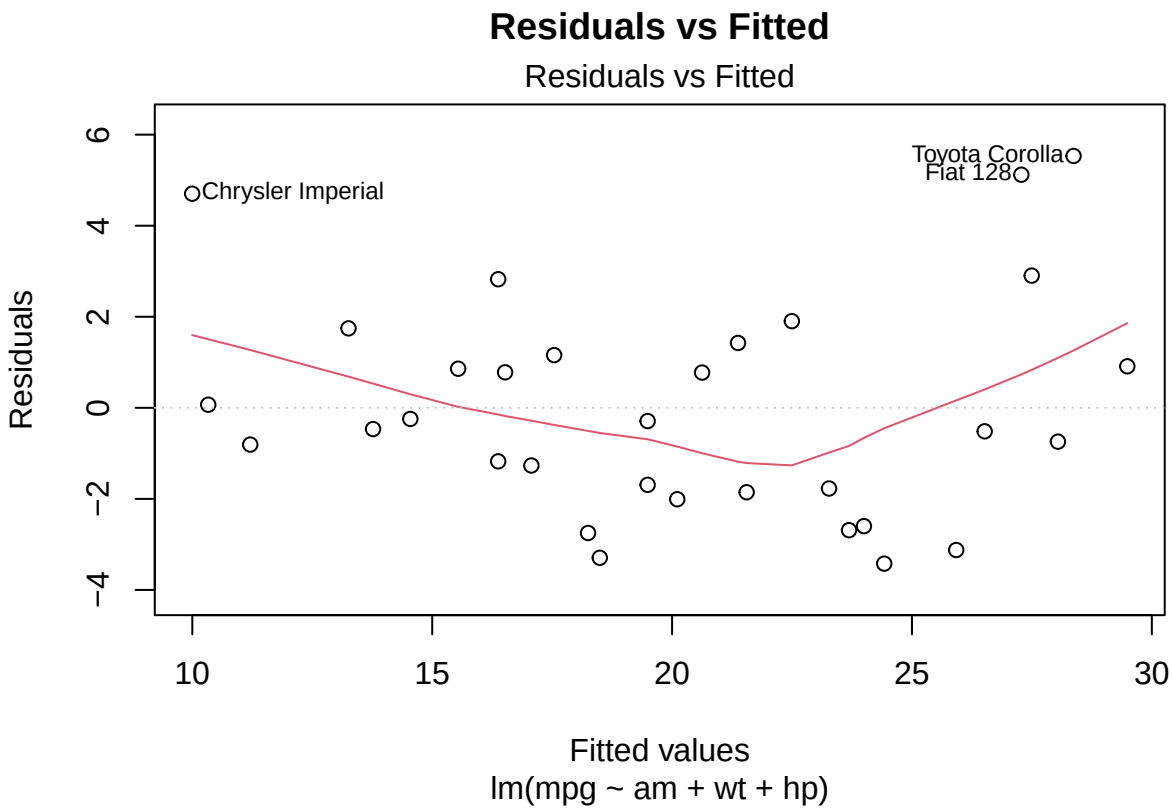
```
##
## Call:
## lm(formula = mpg ~ am + wt + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## amManual     2.083710   1.376420   1.514 0.141268
## wt          -2.878575   0.904971  -3.181 0.003574 **
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

After adjusting for weight and horsepower, the MPG difference between manual and automatic transmissions decreases to 2.1 MPG. This suggests that part of the observed difference is due to confounding variables.

Model Diagnostics

We perform residual analysis to check the adequacy of the final model.

```
# Residual plot
plot(model2, which = 1, main = "Residuals vs Fitted")
```



The residuals appear randomly distributed around zero, indicating no major violations of linear regression assumptions.

Quantifying Uncertainty

We calculate a 95% confidence interval for the MPG difference between manual and automatic transmissions.

```
confint(model2, "amManual", level = 0.95)
```

```
##                2.5 %   97.5 %
## amManual -0.7357587 4.903179
```

The confidence interval suggests that manual transmissions improve MPG by -0.7 to 4.9 MPG compared to automatic transmissions, after adjusting for weight and horsepower.

Conclusion

Manual transmissions are associated with higher MPG compared to automatic transmissions, with an average improvement of 2.1 MPG after adjusting for weight and horsepower. However, this difference is smaller than initially observed in the simple model, highlighting the importance of considering confounding variables. Residual diagnostics confirm the model's adequacy, and the confidence interval quantifies the uncertainty in our estimate.

Appendix

Figure 1: MPG by Transmission Type

```
ggplot(mtcars, aes(x = am, y = mpg, fill = am)) +
  geom_boxplot() +
  labs(title = "MPG by Transmission Type", x = "Transmission", y = "MPG") +
  theme_minimal()
```

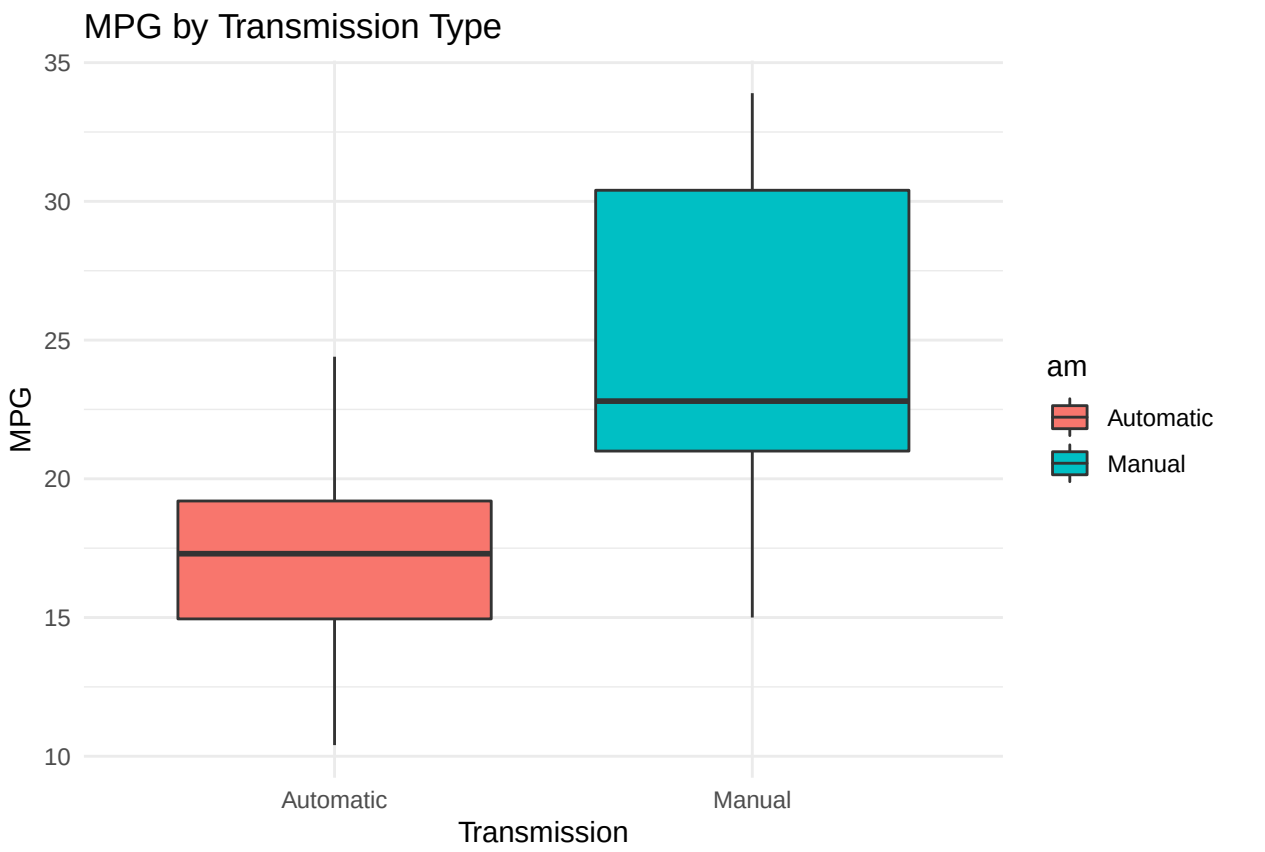
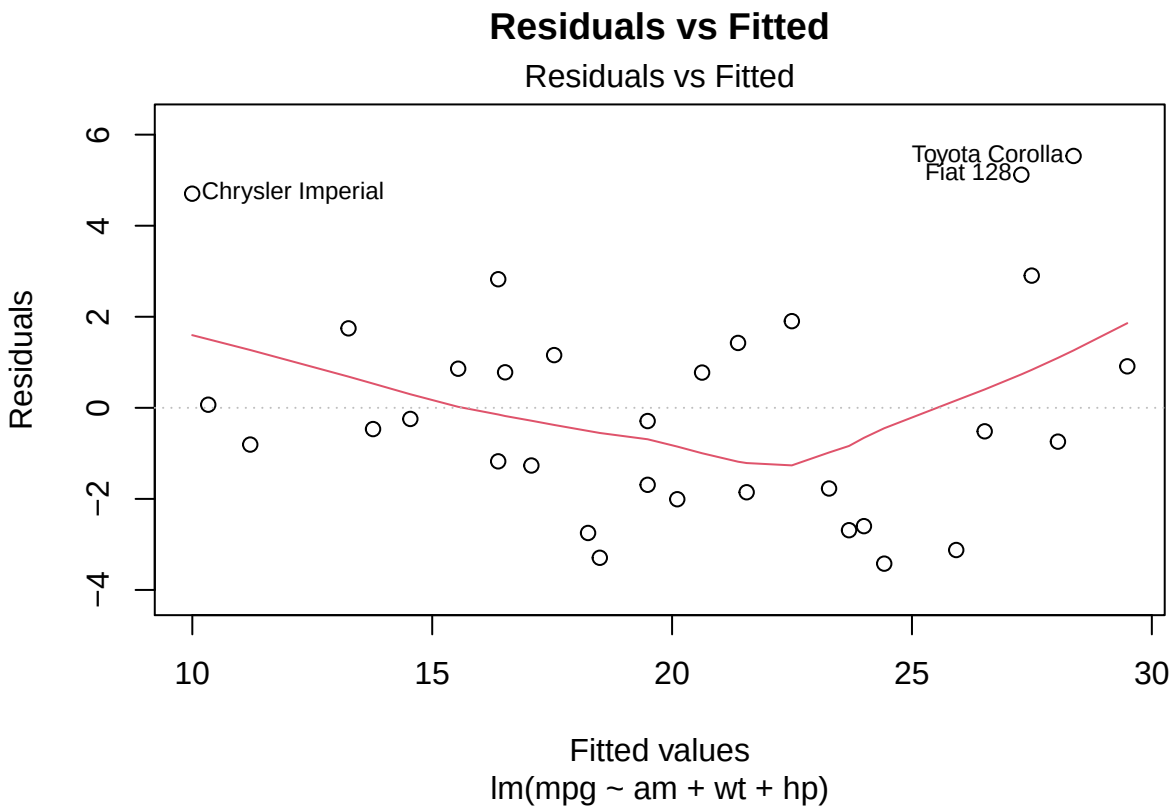


Figure 2: Residuals vs Fitted Plot

```
plot(model2, which = 1, main = "Residuals vs Fitted")
```


Residuals vs Fitted
Residuals vs Fitted
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals
6
4
2
0
−2
−4
10
15
20
25
30
Fitted values
lm(mpg ~ am + wt + hp)