

# MPG Analysis

*Cliff Lau*

## Executive Summary

For this analysis, we investigated the impact of Mile Per Gallon (MPG) from automatic and manual transmission. Based on our analysis, we found no evidence to support using either automatic or manual transmission will lead to a different MPG.

## Analysis

We will be using the `mtcars` dataset from R as well as package `ggplot2` and `dplyr` for our analysis.

```
library(ggplot2)
library(dplyr)

data(mtcars)
```

For the `mtcars` dataset, we're mostly interested in the `mpg` and `am` variables. But we should also consider some other confounding variables that might affect the `mpg` and mask the true effect of auto vs manual.

First, we'll rename the `am` column and change it to factor.

```
mtcars <-
  tbl_df(mtcars) %>%
  mutate(am = as.factor(ifelse(am == 0, "automatic", "manual")),
         cyl = as.factor(cyl))
```

Next, we performed some summary analysis on the dataset:

```
glimpse(mtcars)
```

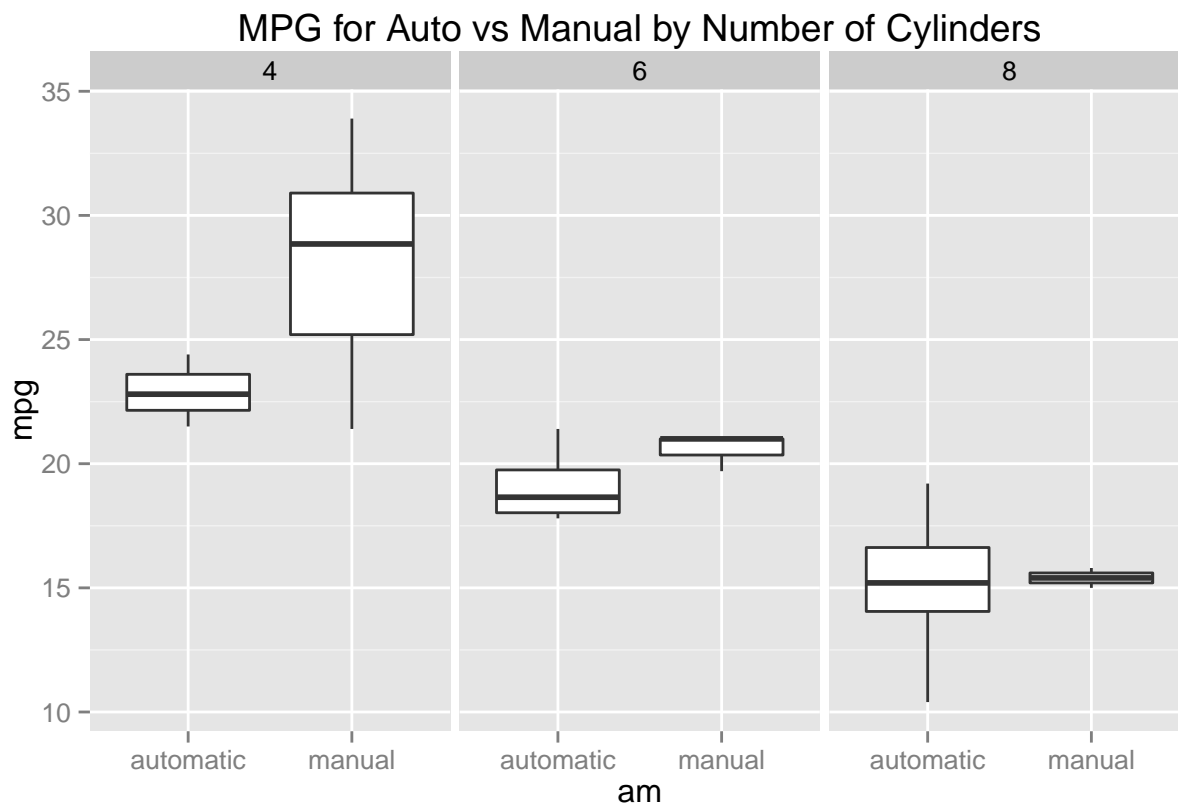
```
## Observations: 32
## Variables:
## $ mpg   (dbl) 21.0, 21.0, 22.8, 21.4, 18.7, 18.1, 14.3, 24.4, 22.8, 19....
## $ cyl   (fctr) 6, 6, 4, 6, 8, 6, 8, 4, 4, 6, 6, 8, 8, 8, 8, 8, 4, 4,...
## $ disp  (dbl) 160.0, 160.0, 108.0, 258.0, 360.0, 225.0, 360.0, 146.7, 1...
## $ hp    (dbl) 110, 110, 93, 110, 175, 105, 245, 62, 95, 123, 123, 180, ...
## $ drat  (dbl) 3.90, 3.90, 3.85, 3.08, 3.15, 2.76, 3.21, 3.69, 3.92, 3.9...
## $ wt    (dbl) 2.620, 2.875, 2.320, 3.215, 3.440, 3.460, 3.570, 3.190, 3...
## $ qsec  (dbl) 16.46, 17.02, 18.61, 19.44, 17.02, 20.22, 15.84, 20.00, 2...
## $ vs    (dbl) 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, ...
## $ am    (fctr) manual, manual, manual, automatic, automatic, automatic,...
## $ gear  (dbl) 4, 4, 4, 3, 3, 3, 3, 4, 4, 4, 4, 3, 3, 3, 3, 3, 4, 4, ...
## $ carb  (dbl) 4, 4, 1, 1, 2, 1, 4, 2, 2, 4, 4, 3, 3, 3, 4, 4, 4, 1, 2, ...
```

```
summary(mtcars)
```

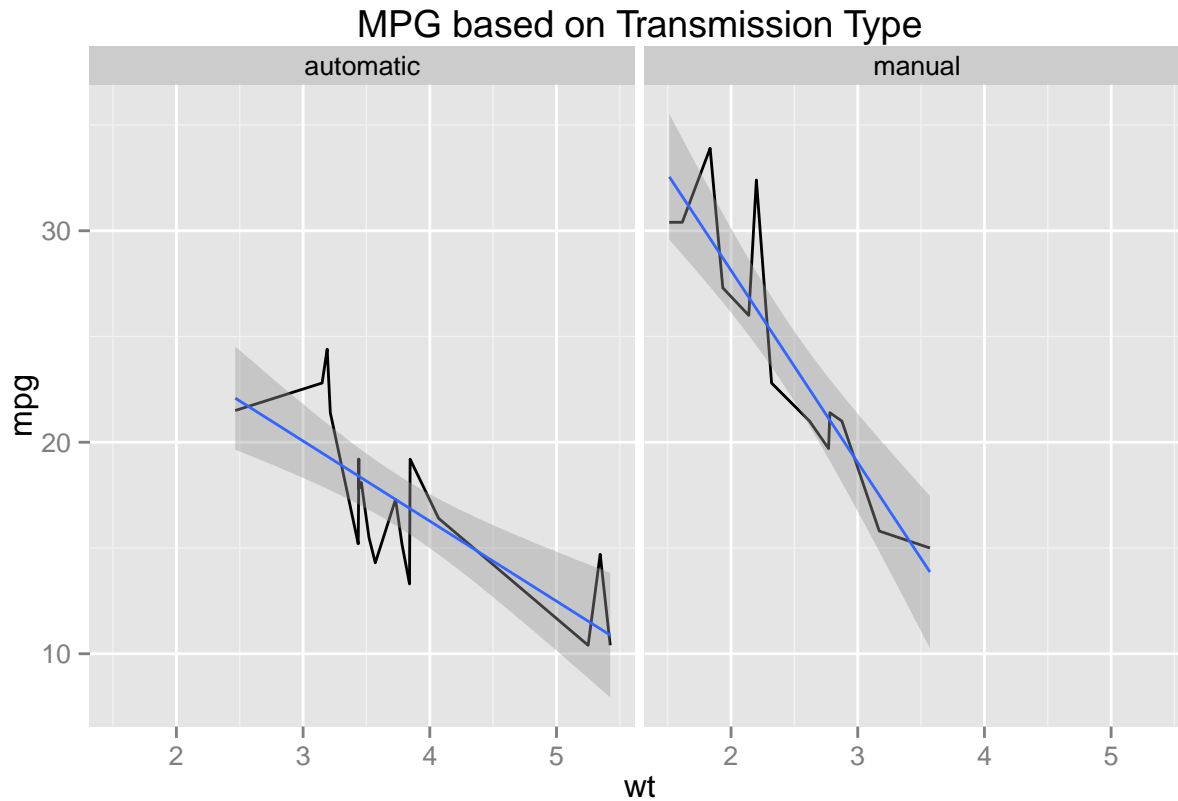
```
##      mpg      cyl      disp      hp      drat
## Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   Min.   :2.760
## 1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.080
## Median :19.20   8:14   Median :196.3   Median :123.0   Median :3.695
## Mean   :20.09           Mean   :230.7   Mean   :146.7   Mean   :3.597
## 3rd Qu.:22.80           3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.920
## Max.   :33.90           Max.   :472.0   Max.   :335.0   Max.   :4.930
##      wt      qsec      vs      am
## Min.   :1.513   Min.   :14.50   Min.   :0.0000   automatic:19
## 1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000   manual   :13
## Median :3.325   Median :17.71   Median :0.0000
## Mean   :3.217   Mean   :17.85   Mean   :0.4375
## 3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
## Max.   :5.424   Max.   :22.90   Max.   :1.0000
##      gear      carb
## Min.   :3.000   Min.   :1.000
## 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
```

As well as some exploratory graphs:

```
ggplot(mtcars, aes(am, mpg, cyl)) +
  geom_boxplot() +
  facet_grid(.~cyl) +
  ggtitle("MPG for Auto vs Manual by Number of Cylinders")
```



```
ggplot(mtcars, aes(wt, mpg, am))+
  geom_line() +
  geom_smooth(method = "lm") +
  facet_grid( .~am) +
  ggtitle("MPG based on Transmission Type")
```



Based on the plots above, we might suspect that there is no difference in auto and manual transmission. To confirm, we'll fit a linear model with the 3 predictors: am, cyl, and wt:

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

```
##
## Call:
## lm(formula = mpg ~ am + cyl + wt, data = mtcars)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-4.4898	-1.3116	-0.5039	1.4162	5.7758

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	33.7536	2.8135	11.997	2.5e-12 ***
## ammanual	0.1501	1.3002	0.115	0.90895
## cyl6	-4.2573	1.4112	-3.017	0.00551 **
## cyl8	-6.0791	1.6837	-3.611	0.00123 **
## wt	-3.1496	0.9080	-3.469	0.00177 **

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.603 on 27 degrees of freedom
## Multiple R-squared:  0.8375, Adjusted R-squared:  0.8134
## F-statistic: 34.79 on 4 and 27 DF,  p-value: 2.73e-10
```

Based on the summary statistics above, we can see that the `am` predictor is not significant based on the p-value. Therefore, we can conclude that automatic vs manual makes no difference in terms of MPG.