Objective

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome).

They are particularly interested in the following two questions:

- “Is an automatic or manual transmission better for MPG”
- “Quantify the MPG difference between automatic and manual transmissions”

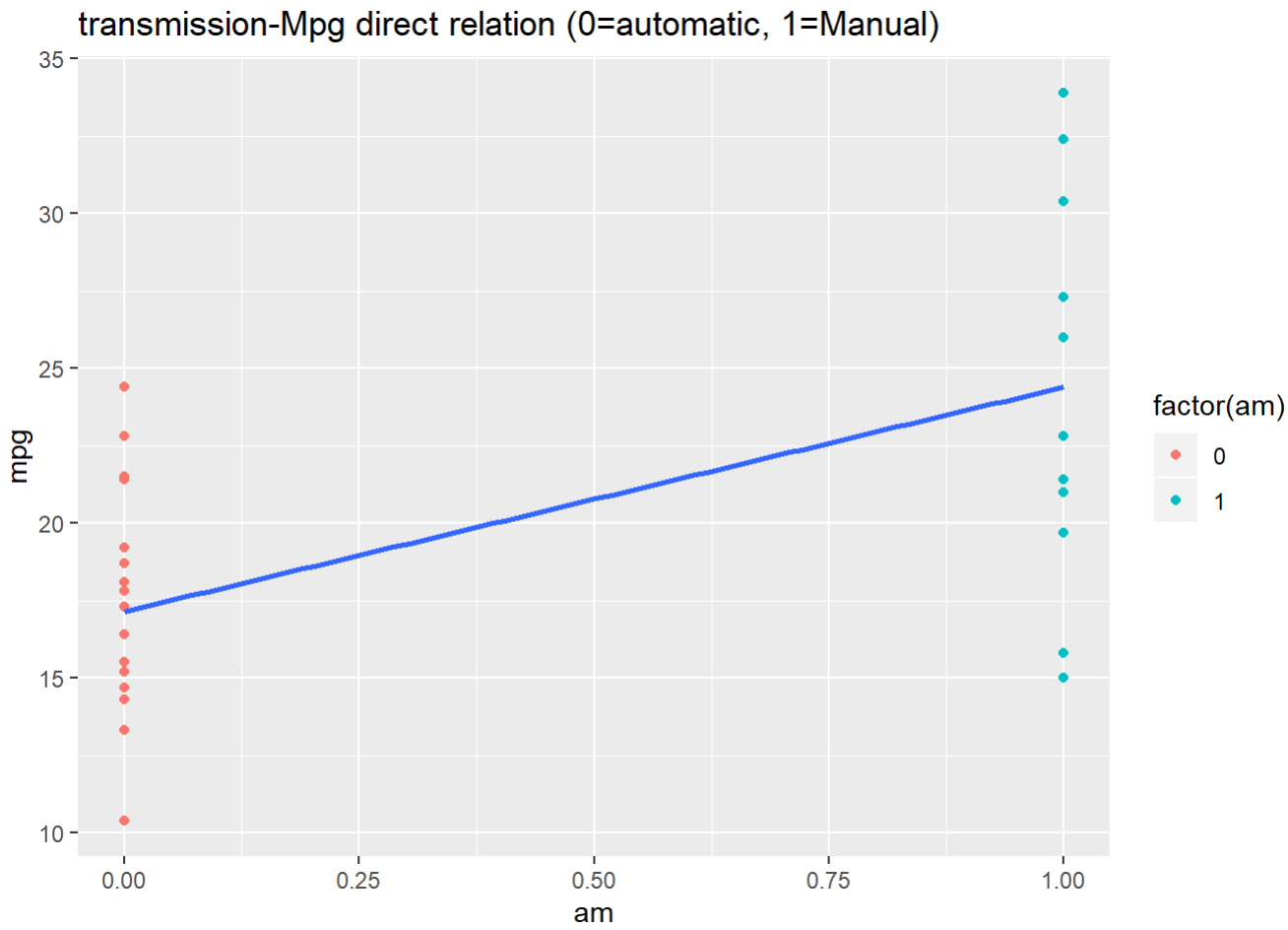
```
library(tidyr)
library(ggplot2)

#Load dataset
data(mtcars)

# overview mtcars
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
```

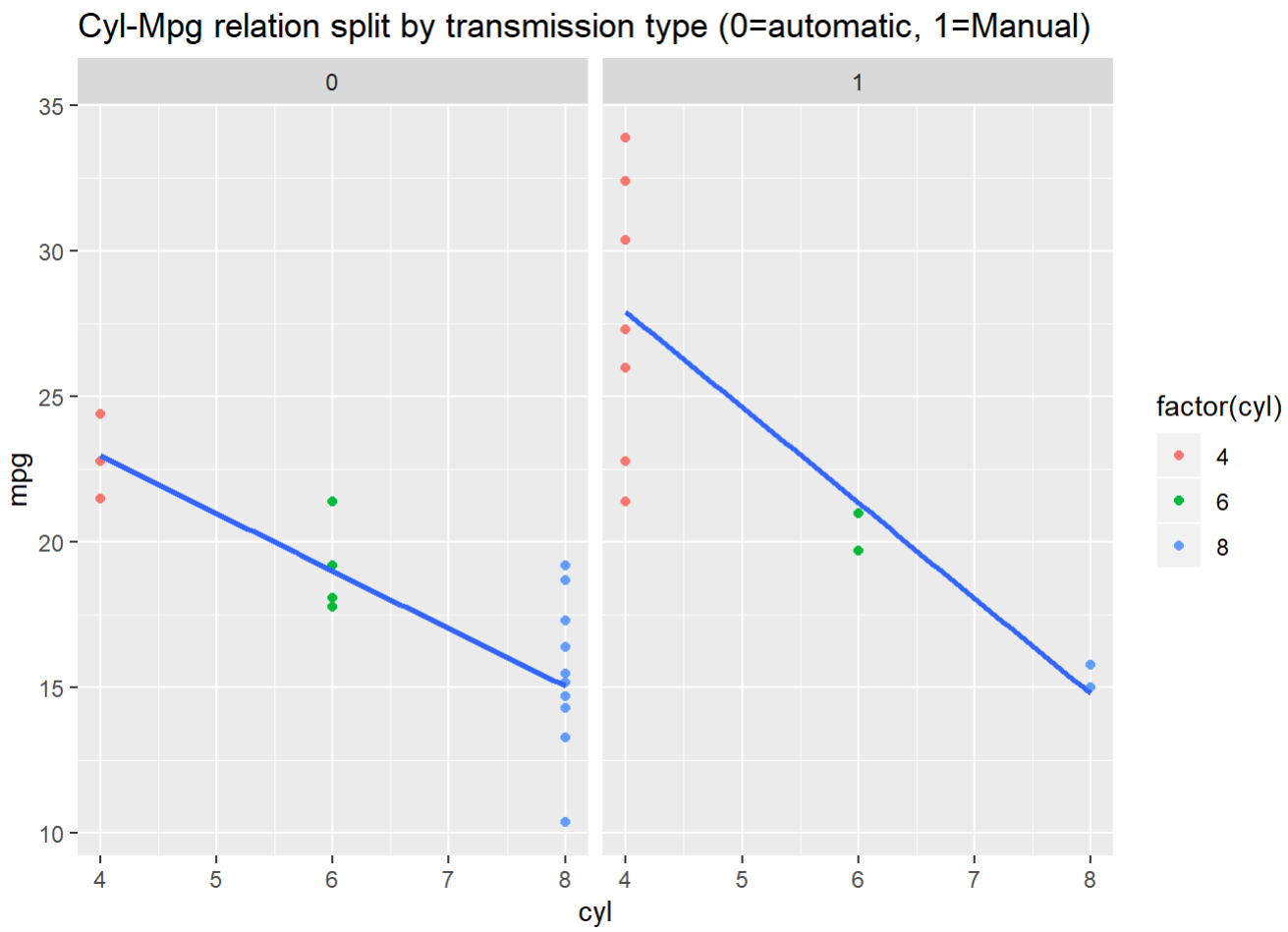
EDA phase

We will focus on transmission relationship with other variables

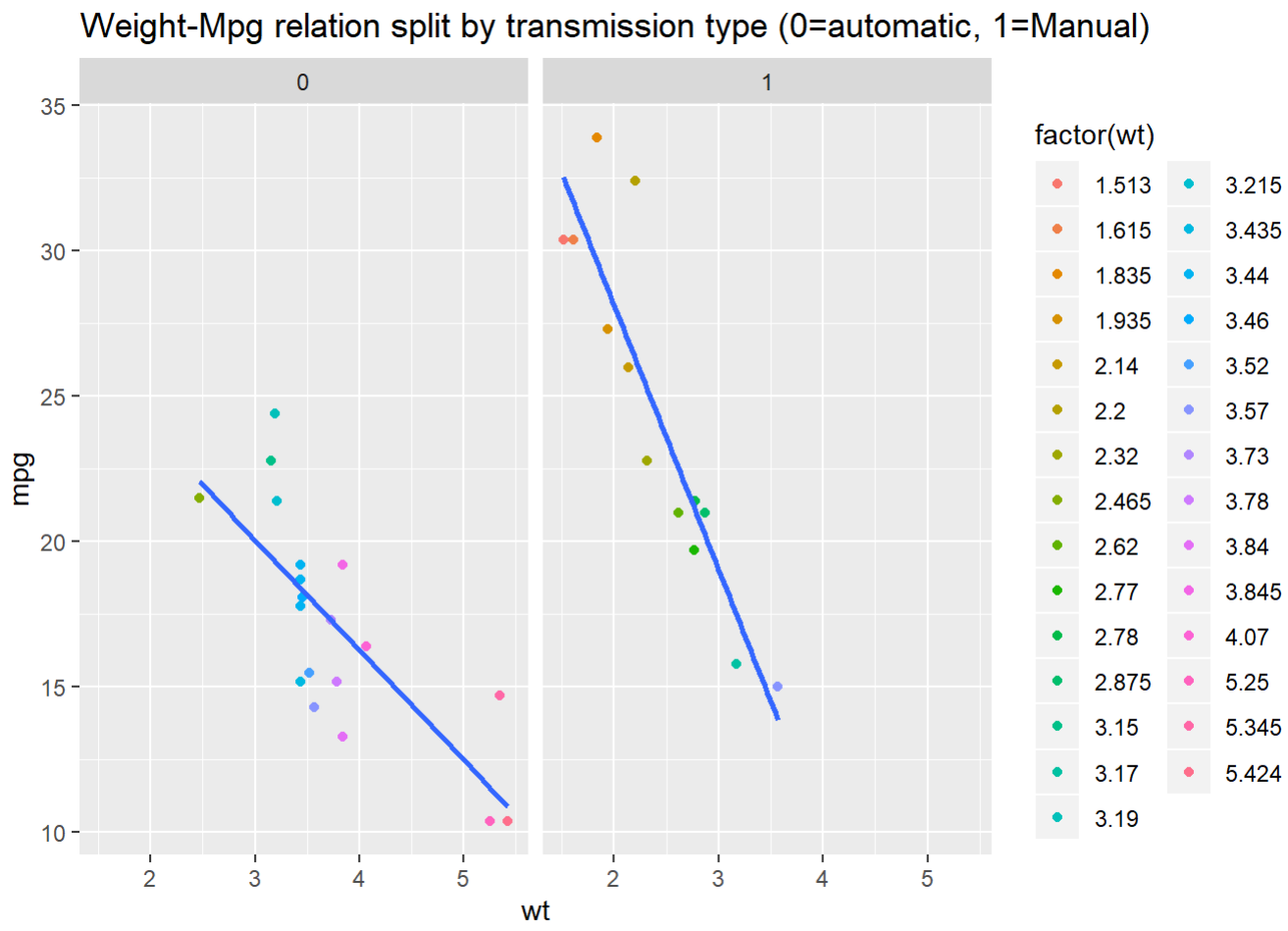
```
# check direct relation between mpg and transmission type, ignoring other factors
mtcars %>%
  ggplot(aes(am,mpg))+
  geom_point(aes(colour = factor(am)),na.rm = TRUE)+
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("transmission-Mpg direct relation (0=automatic, 1=Manual)")
```



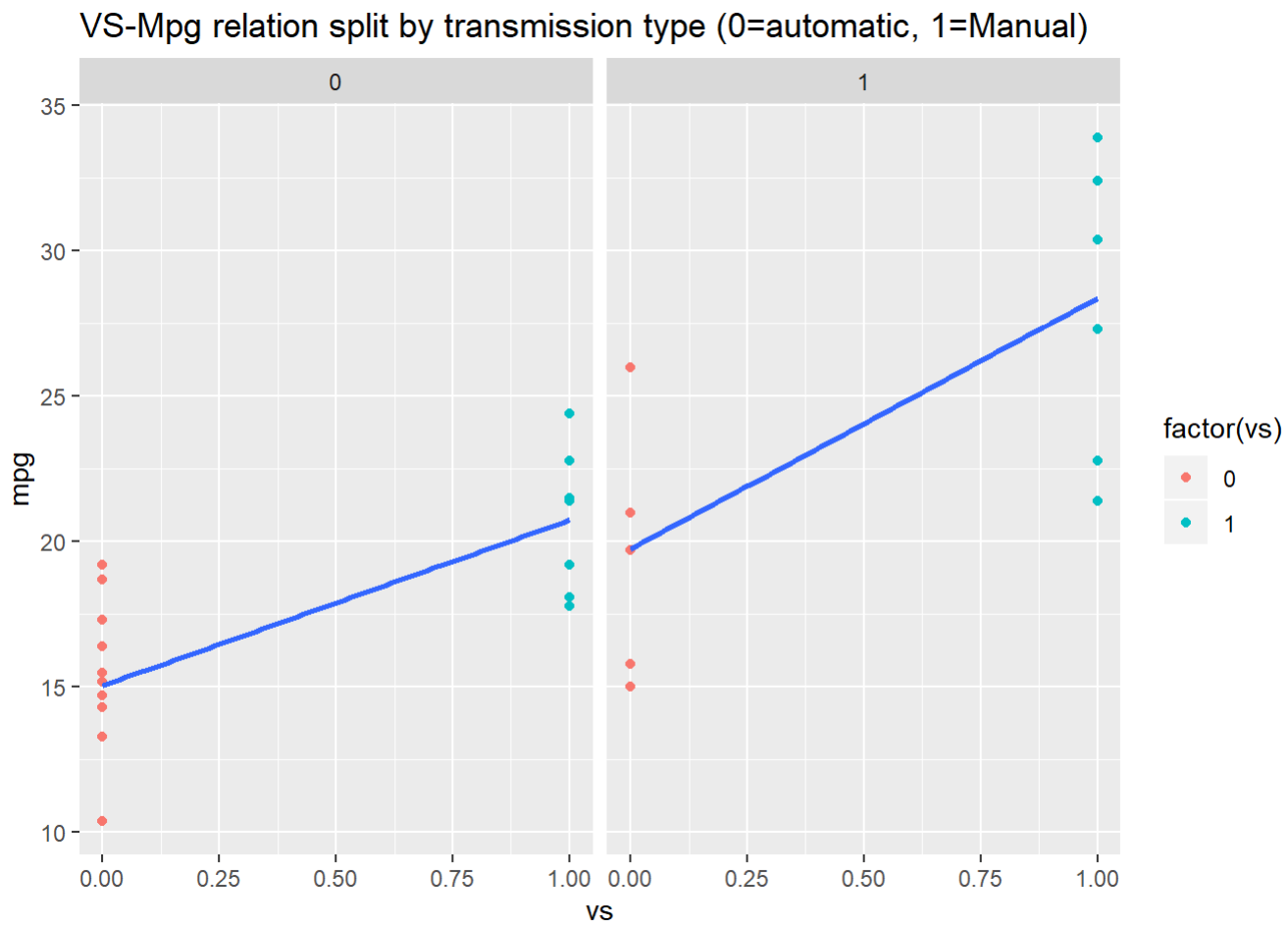

```
# select couple of variables and check relationship
# check relation between mpg and cyl, split by transmission type
mtcars %>%
  ggplot(aes(cyl,mpg))+
  geom_point(aes(colour = factor(cyl)),na.rm = TRUE)+
  facet_grid(cols = vars(as.factor(am)))+
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("Cyl-Mpg relation split by transmission type (0=automatic, 1=Manual)")
```




```
# check relation between mpg and wt, split by transmission type
mtcars %>%
  ggplot(aes(wt,mpg))+
  geom_point(aes(colour = factor(wt)),na.rm = TRUE)+
  facet_grid(cols = vars(as.factor(am)))+
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("Weight-Mpg relation split by transmission type (0=automatic, 1=Manual)")
```




```
# check relation between mpg and vs, split by transmission type
mtcars %>%
  ggplot(aes(vs,mpg))+
  geom_point(aes(colour = factor(vs)),na.rm = TRUE)+
  facet_grid(cols = vars(as.factor(am)))+
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("VS-Mpg relation split by transmission type (0=automatic, 1=Manual)")
```



explore linear models phase

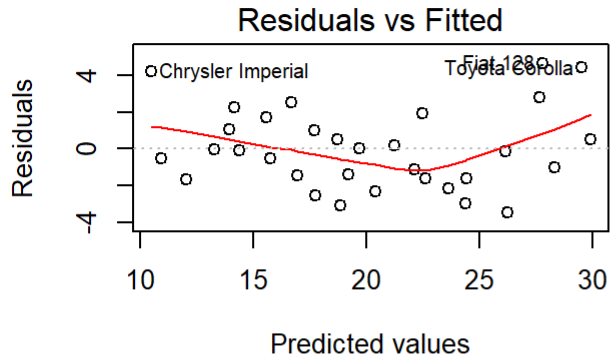
```
# fit1 is simple, will be only between mpg and transmission
fit1 <- lm(mpg~am-1 , data = mtcars)
summary(fit1)$coefficients
```

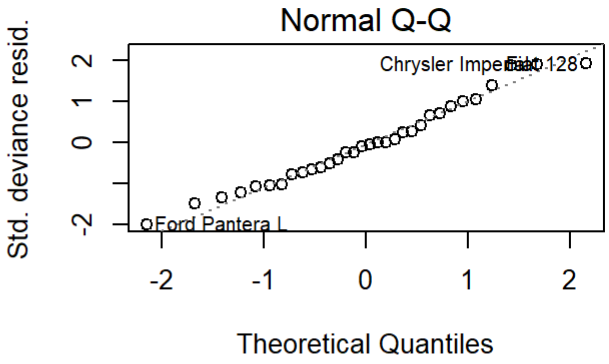
```
##    Estimate Std. Error  t value     Pr(>|t|)
## am 24.39231   3.956183 6.165616 7.666189e-07
```

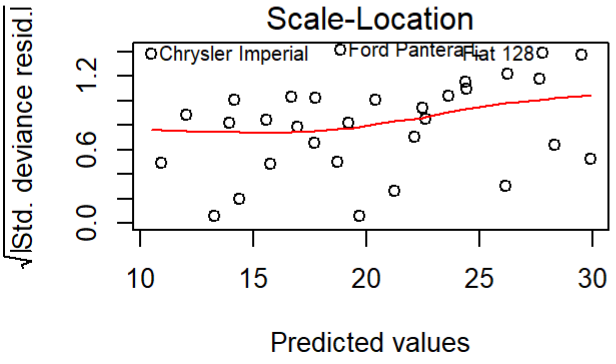
```
# fit2 is multi-model between mpg & all variables
fit2 <- glm(mpg~. , data = mtcars)
summary(fit2)$coefficients
```

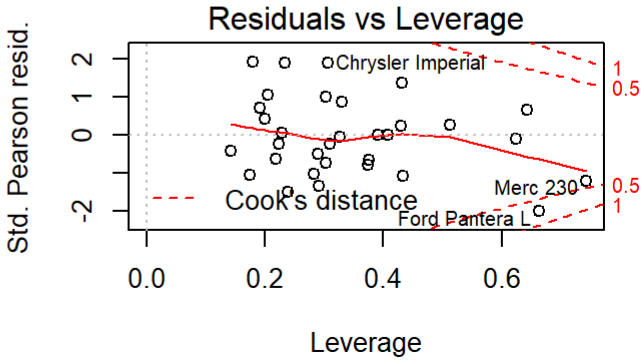
```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am           2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

```
# plot model
par(mfrow = c(2, 2))
plot(fit2)
```





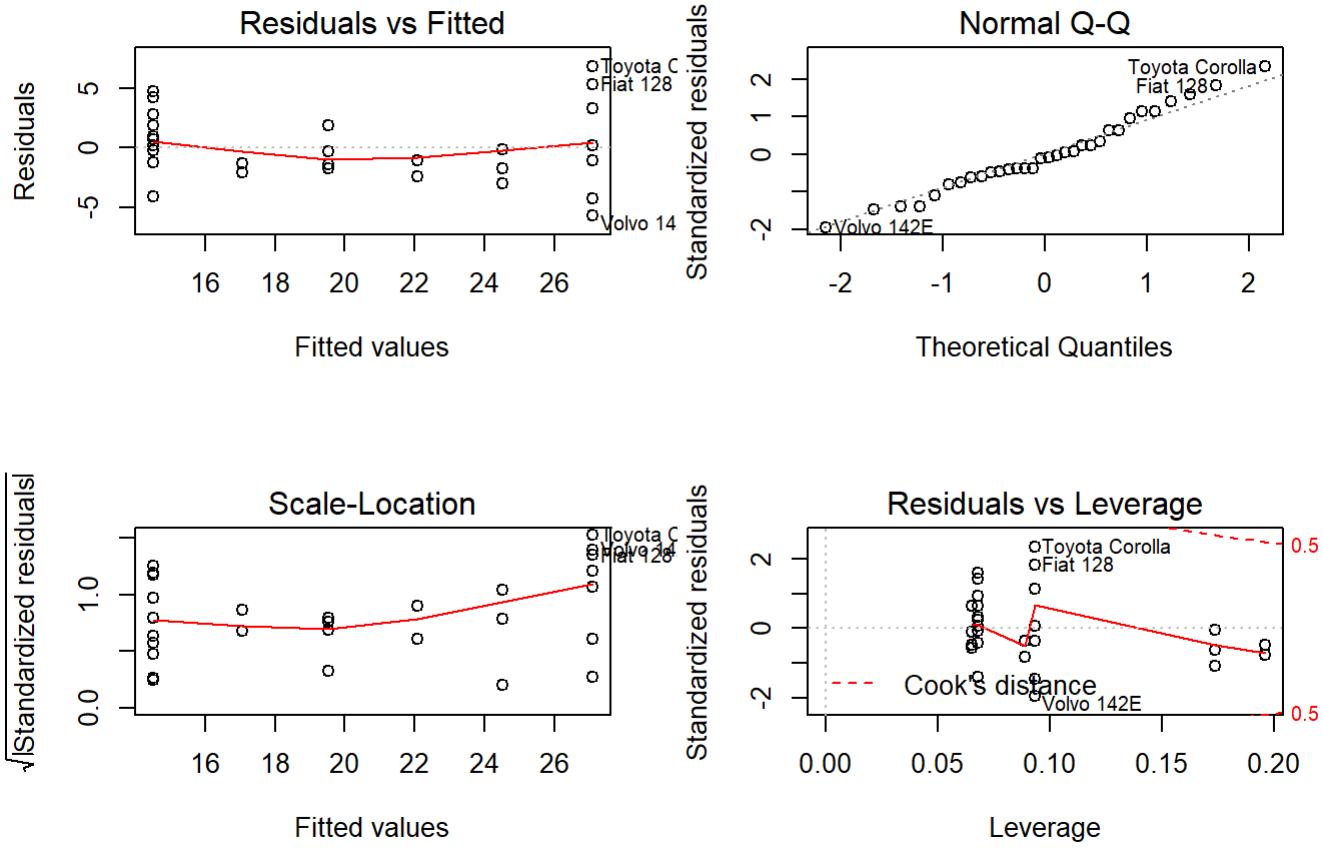





```
# will add one variable at a time to mpg and transmission
fit3 <- lm(mpg~cyl+am , data = mtcars)
summary(fit3)$coefficients
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 34.522443  2.6031842 13.261621 7.694408e-14
## cyl         -2.500958  0.3608282 -6.931159 1.284560e-07
## am           2.567035  1.2914280  1.987749 5.635445e-02
```

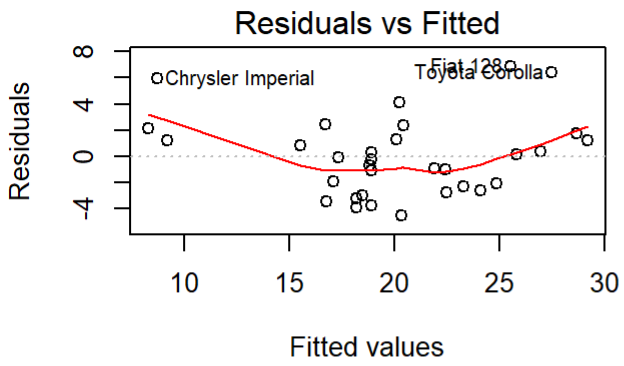
```
# plot model
par(mfrow = c(2, 2))
plot(fit3)
```

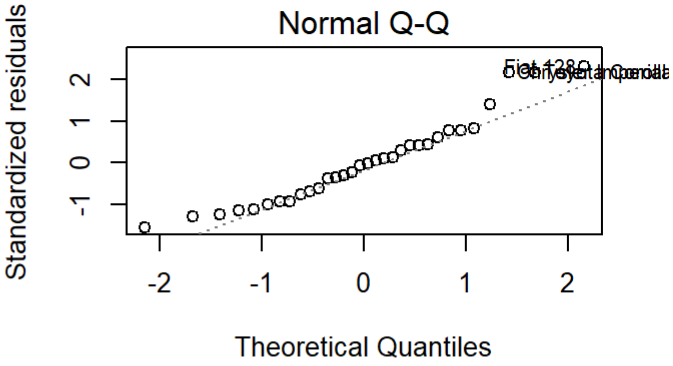


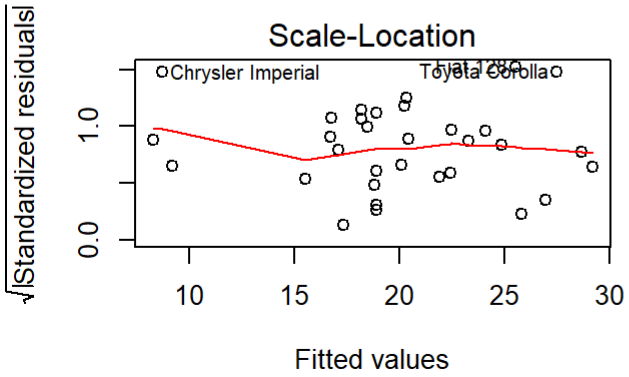

```
fit4 <- lm(mpg~wt+am , data = mtcars)
summary(fit4)$coefficients
```

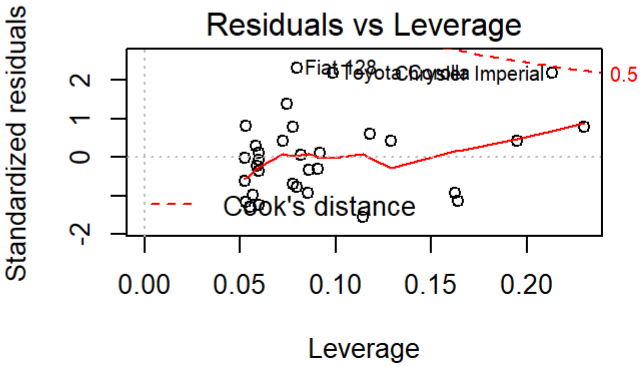
```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
## am          -0.02361522  1.5456453 -0.01527855 9.879146e-01
```

```
# plot model
par(mfrow = c(2, 2))
plot(fit4)
```





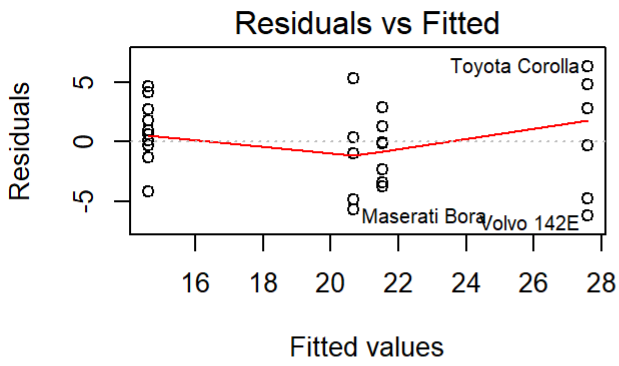


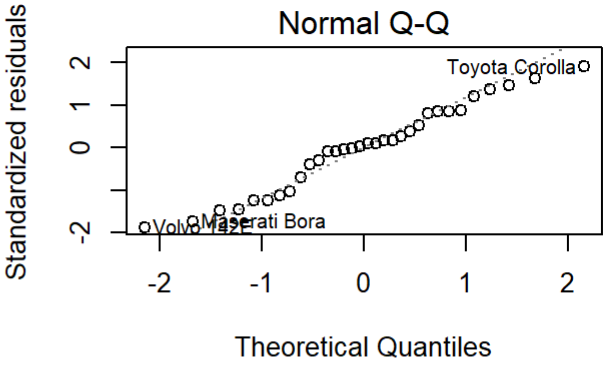


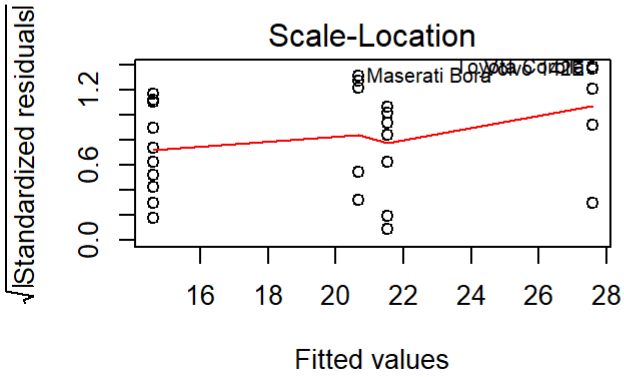

```
fit5 <- lm(mpg~vs+am , data = mtcars)
summary(fit5)$coefficients
```

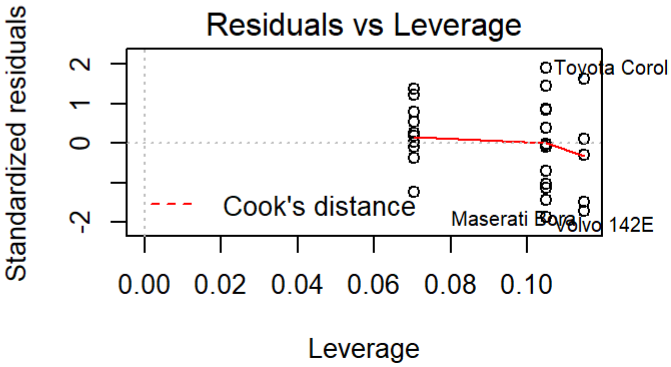
```
##             Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 14.594444  0.9261514 15.758162 9.352153e-16
## vs           6.929365  1.2621316  5.490208 6.500962e-06
## am           6.066667  1.2748423  4.758759 4.958115e-05
```

```
# plot model
par(mfrow = c(2, 2))
plot(fit5)
```









Summary

- Automatic Transmission has better impact on mpg
- Residuals show that considering other variables in the process of fitting a model is much better than considering only mpg & am