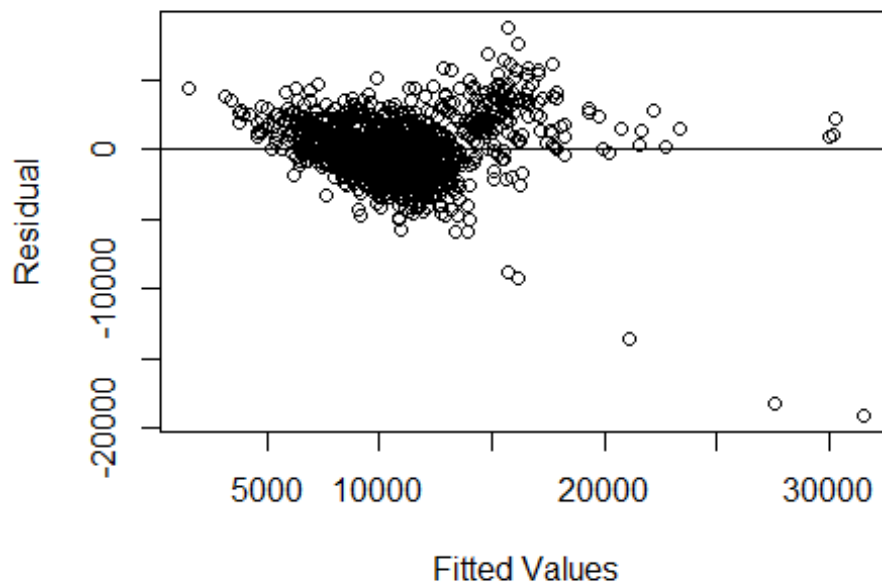# Residuals.R

Shraddha Somani

```r
# Residuals
# Part A: Residuals and Transforms
ToyotaPrices <- read.csv("D:/DADM/Assignment/ToyotaPrices.csv")
ToyotaPrices_PKWT = subset(ToyotaPrices, select = c(Price, KM, Weight,
Tow_Bar))
head(ToyotaPrices_PKWT)

##     Price     KM Weight Tow_Bar
## 1 13500 46986   1165       0
## 2 13750 72937   1165       0
## 3 13950 41711   1165       0
## 4 14950 48000   1165       0
## 5 13750 38500   1170       0
## 6 12950 61000   1170       0

# Exercise 1: Residual Plot
# Q1 (A) - Obtain plot
# Obtain the Residuals vs Fitted Plot of the fitted model with an added
horizontal line at y = 0
# Do the points look randomly distributed about the line?
fit = lm(Price ~ KM + Weight + Tow_Bar, ToyotaPrices_PKWT)
plot(fit$fitted.values, fit$residuals, xlab = "Fitted Values", ylab =
"Residual")
abline(h = 0, v = NULL)
```

```
# Analysis:
# - Yes, the points look randomly distributed about the line.

# Q1 (B) - Obtain plot using the z-scores of the resiudals
# Repeat the Residuals vs Fitted Plot using z-scores of the residuals. Add
empirical rule horizontal lines at +2 and ???2.
# Use these lines to judge whether or not the residuals are normal or there
are outliers. Point out any outliers. Point out any floor or ceiling effects.
Do you think the residuals are normal?
a = fit$residuals
mean(a)

## [1] 1.60432e-12

sd(a)

## [1] 2141.851

b = (a - mean(a))/sd(a)
c = plot(b, fit$fitted.values, xlab = "Z-Score of Residual", ylab = "Fitted
Values")
abline(h = NULL, v = -2)
abline(h = NULL, v = 2)
```
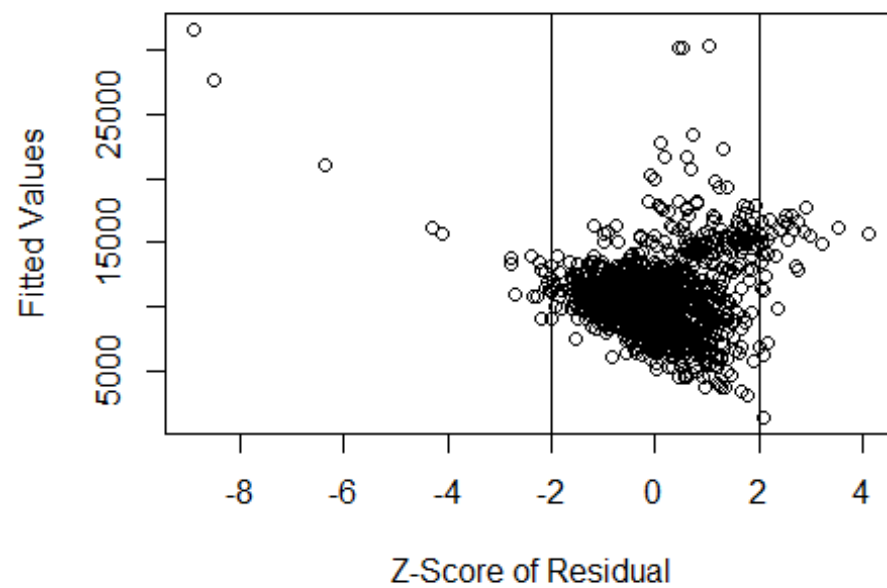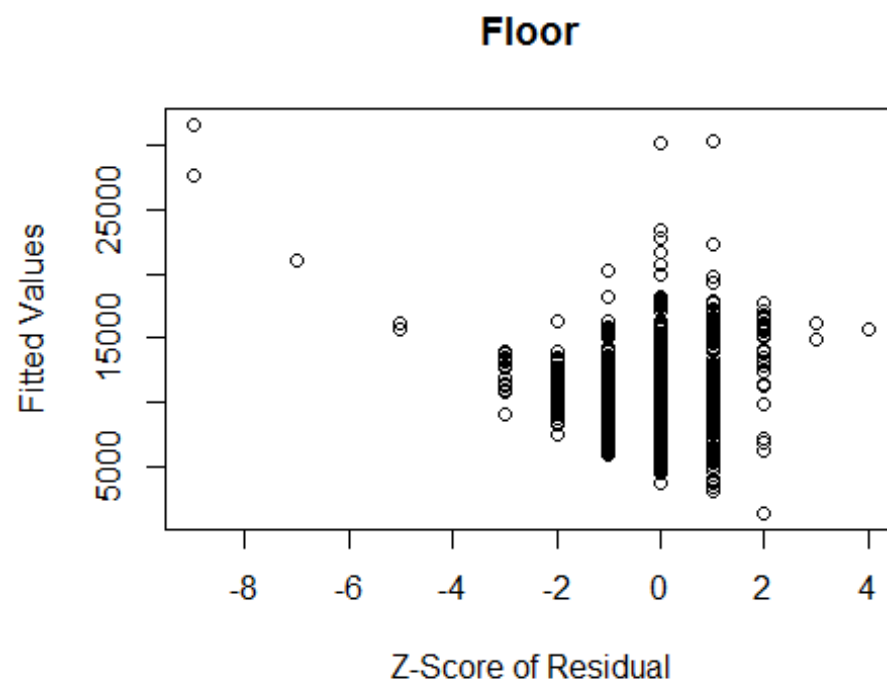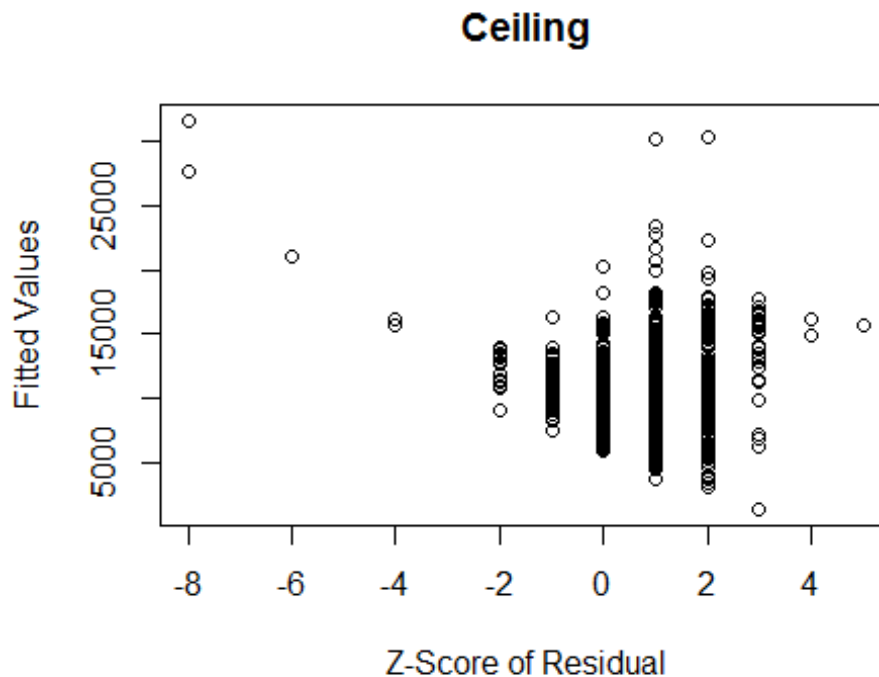
```
d = plot(floor(b), fit$fitted.values, xlab = "Z-Score of Residual", ylab =
"Fitted Values", main = "Floor")
```

Floor

```r
e = plot(ceiling(b), fit$fitted.values, xlab = "Z-Score of Residual", ylab =
"Fitted Values", main = "Ceiling")
```

Ceiling


```r
# Analysis:
# - The residuals are normal.
# - There are outliers present.
# - floor(x) = is the largest integer less than or equal to x. Eg
floor(3.456) = 3
# - ceiling(x) = is the smallest integer greater than or equal to x. Eg
ceiling(3.456) = 4
# - Floor and ceiling plot is same, however there's a difference between x
axis range and outliers.
# - There are more number of outliers present in ceiling compared to that of
floor.


# Exercise 2: Residual Normal QQ-Plot
# Q2 (A) - Obtain the normal probability QQ-Plot of the residuals.
qqnorm(fit$residuals)
qqline(fit$residuals)
```
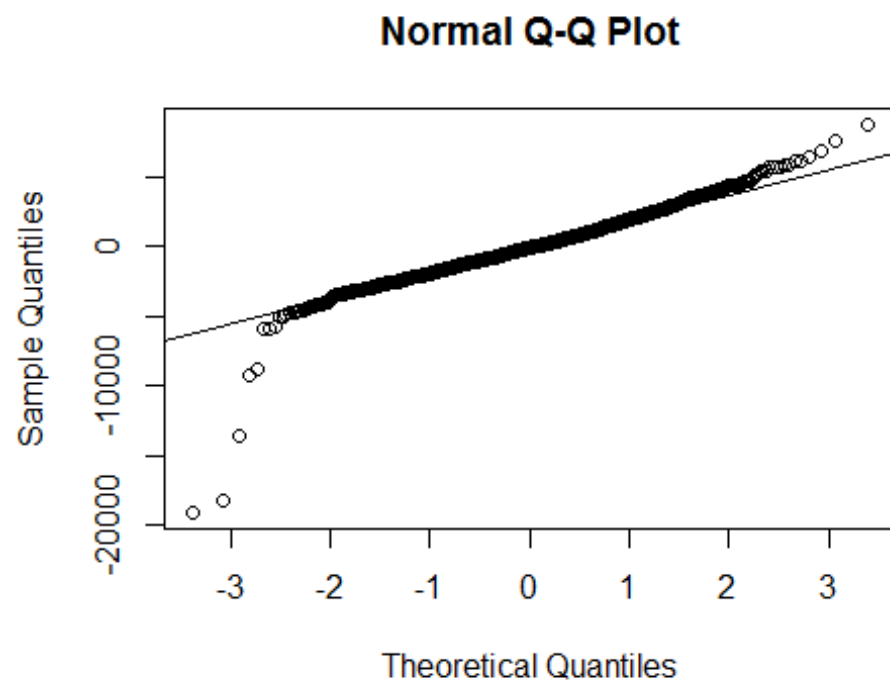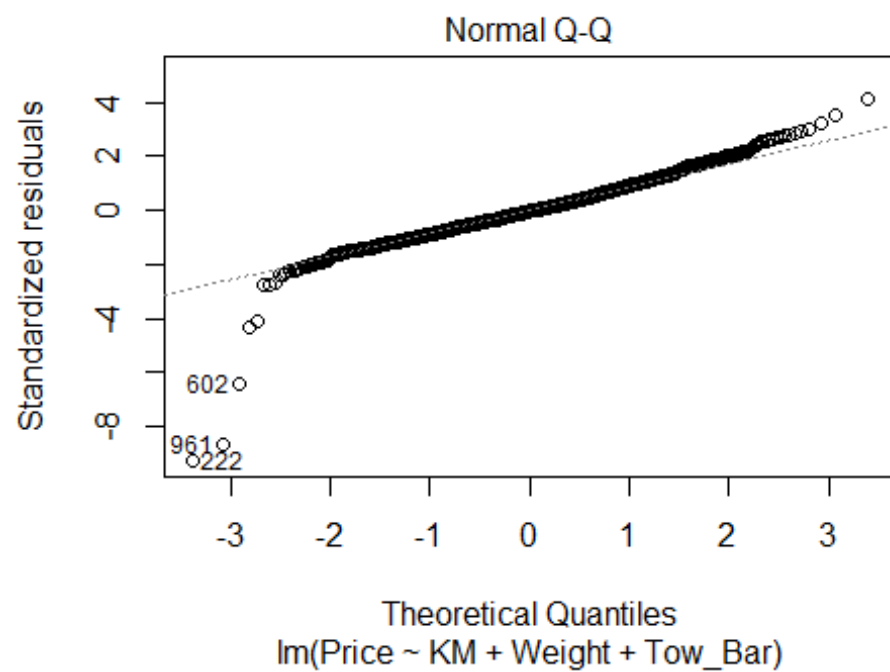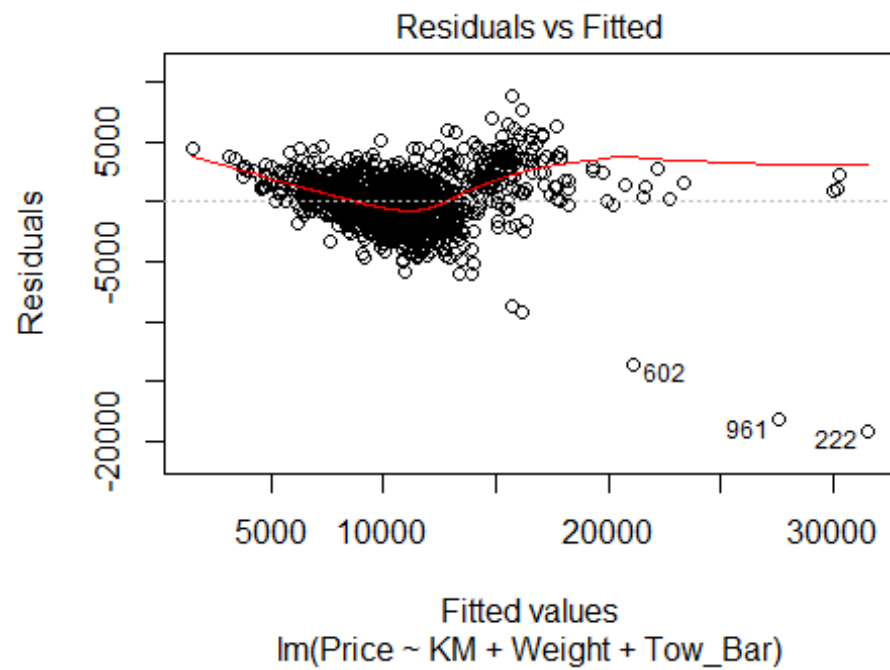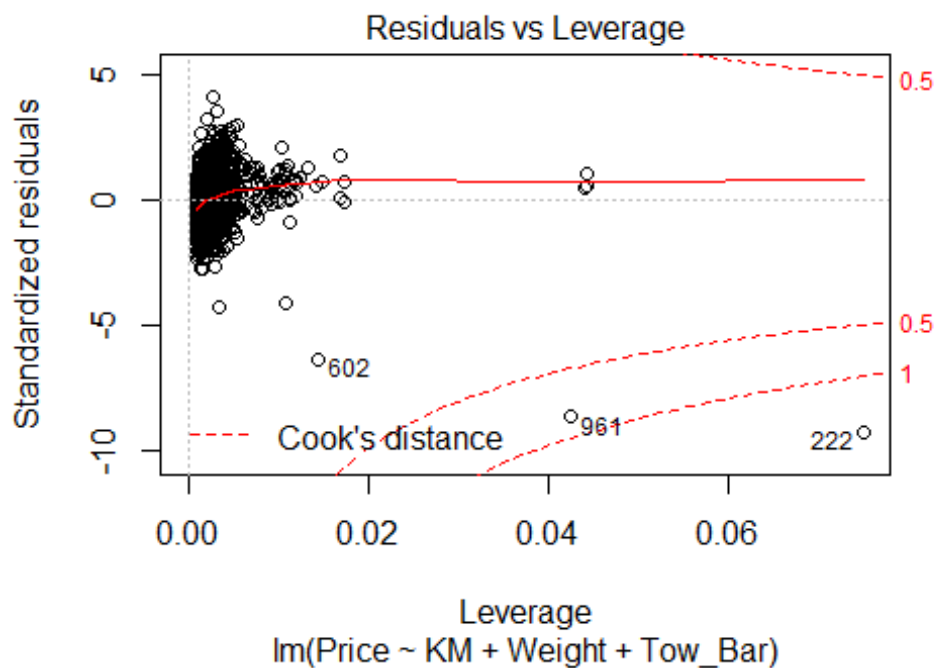
Normal Q-Q Plot


```
#Q2 (B) - Do the residuals look normal?
# Analysis:
# - Yes, the residual look normal. But there are few outliers present.


# Exercise 3: Composite goodness-of-fit plots
plot(fit)
```

Residuals vs Fitted

Residuals

5000
-5000
-20000

602
961
222

5000   10000   20000   30000

Fitted values
lm(Price ~ KM + Weight + Tow_Bar)

Normal Q-Q

Standardized residuals

4
2
0
-4
-8

602
961
222

-3   -2   -1   0   1   2   3

Theoretical Quantiles
lm(Price ~ KM + Weight + Tow_Bar)

Scale-Location

Im(Price ~ KM + Weight + Tow_Bar)

Residuals vs Leverage

Im(Price ~ KM + Weight + Tow_Bar)
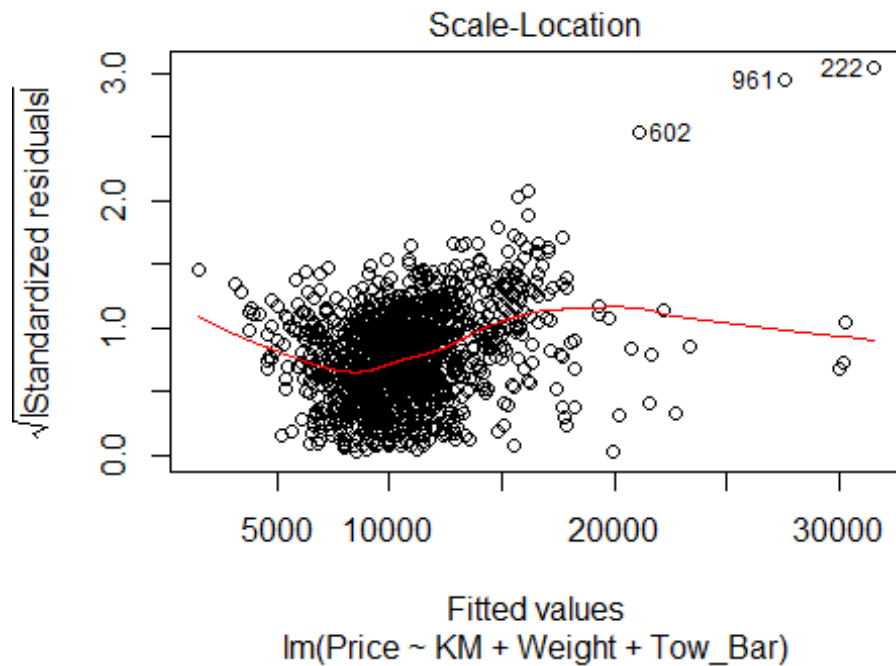
Q3 (A) - Obtain the composite goodness-of-fit plots
Obtain the composite goodness-of-fit plots for the fitted model. What plots
involve the residuals? Do the Residuals vs Fitted Plot and the Normal QQ-Plot
Look about the same as those obtained earlier?

```
# Analysis:
# - Plots involving the residual are: Residual vs Fitted Plot, Residual vs
Leverage.
# - Yes, the Residuals vs Fitted Plot and the Normal QQ-Plot look about the
same as those obtained earlier.

# Q3 (B) - Outliers
# We examine the Residuals vs Leverage Plot in the composite goodness-of-fit
plots. Outliers points will be identified by their row name. Are there any
outliers? If so, what are their row names.
# Analysis:
# - Yes, there are outliers present.
# - Row Names - 602, 961 and 222
```