

R Notebook

```
library(readxl)
Tng_Ctr_Hour2 <- read_excel("C:/RBS/Business Forecasting/Group Project/Final Project/Tng_Ctr_Hour2.xlsx")
View(Tng_Ctr_Hour2)
```

```
library(data.table)
library(ggplot2)
library(stringr)
library(TTR)
library(fpp)
```

```
## Loading required package: forecast
```

```
## Registered S3 method overwritten by 'quantmod':
##   method           from
##   as.zoo.data.frame zoo
```

```
## Loading required package: fma
```

```
## Loading required package: expsmooth
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
```

```
## Loading required package: tseries
```

```
library(fpp2)
```

```
##
## Attaching package: 'fpp2'
```

```
## The following objects are masked from 'package:fpp':
##
##   ausair, ausbeer, austa, austourists, debitcards, departures,
##   elecequip, euretail, guinearice, oil, sunspotarea, usmelec
```

```
library(ggplot2)
library(stats)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##   between, first, last

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(graphics)
library(ggfortify)
```

```
## Registered S3 methods overwritten by 'ggfortify':
##   method                from
##   autoplot.Arima         forecast
##   autoplot.acf           forecast
##   autoplot.ar            forecast
##   autoplot.bats          forecast
##   autoplot.decomposed.ts forecast
##   autoplot.ets           forecast
##   autoplot.forecast      forecast
##   autoplot.stl           forecast
##   autoplot.ts            forecast
##   fitted.ar             forecast
##   fortify.ts             forecast
##   residuals.ar          forecast
```

```
setDT(Tng_Ctr_Hour2)
Tng_Ctr_Hour2[,Quarter:=factor(Quarter)]
Tng_Ctr_Hour2[,Month:=factor(Month)]
Tng_Ctr_Hour2[,Year:=factor(Year)]
```

Multiple Regression Analysis

```
setDT(Tng_Ctr_Hour2)
class(Tng_Ctr_Hour2)
```

```
## [1] "data.table" "data.frame"
```

```
Tng_Ctr_Hour2[,RPM:=shift(RPM, n=2)]
Tng_Ctr_Hour2[,NJURN:=shift(NJURN, n=1)]
Tng_Ctr_Hour2[,Cons_Sent:=shift(Cons_Sent, n=2)]
head(Tng_Ctr_Hour2)
```

```
##      Year   Yr Quarter Month Device_Hrs DH_Prev_Year DH_YoY_Change
## 1: 2015-01 2015     Q1   Jan     960.42           NA           NA
## 2: 2015-02 2015     Q1   Feb     944.08           NA           NA
## 3: 2015-03 2015     Q1   Mar    1429.12           NA           NA
## 4: 2015-04 2015     Q2   Apr    1097.00           NA           NA
## 5: 2015-05 2015     Q2   May     915.85           NA           NA
## 6: 2015-06 2015     Q2   Jun     783.45           NA           NA
##      DH_YoY_Ch_Per Total_Inst_Hrs Total_Inst_Hrs_Prev_Year Inst_Hrs_YoY_Change
## 1:                NA          1700.67                NA                NA
## 2:                NA          1614.00                NA                NA
## 3:                NA          2532.90                NA                NA
## 4:                NA          2152.25                NA                NA
## 5:                NA          1695.43                NA                NA
## 6:                NA          1675.91                NA                NA
##      Total_Inst_Hrs_YoY_Change_Per2 Cons_Sent NJURN      RPM CPIUrban CPIMedian
## 1:                                NA      NA   NA      NA  234.747  1.947530
## 2:                                NA      NA   6.8      NA  235.342  1.954495
## 3:                                NA     98.1   6.7 65975447  235.976  2.433336
## 4:                                NA     95.4   6.3 59784666  236.222  2.962651
## 5:                                NA     93.0   5.8 75751609  237.001  2.501291
## 6:                                NA     95.9   6.0 73090871  237.657  2.972400
```

```
head(Tng_Ctr_Hour2)
```

```
##      Year   Yr Quarter Month Device_Hrs DH_Prev_Year DH_YoY_Change
## 1: 2015-01 2015     Q1   Jan     960.42           NA           NA
## 2: 2015-02 2015     Q1   Feb     944.08           NA           NA
## 3: 2015-03 2015     Q1   Mar    1429.12           NA           NA
## 4: 2015-04 2015     Q2   Apr    1097.00           NA           NA
## 5: 2015-05 2015     Q2   May     915.85           NA           NA
## 6: 2015-06 2015     Q2   Jun     783.45           NA           NA
##      DH_YoY_Ch_Per Total_Inst_Hrs Total_Inst_Hrs_Prev_Year Inst_Hrs_YoY_Change
## 1:                NA          1700.67                NA                NA
## 2:                NA          1614.00                NA                NA
## 3:                NA          2532.90                NA                NA
## 4:                NA          2152.25                NA                NA
## 5:                NA          1695.43                NA                NA
## 6:                NA          1675.91                NA                NA
##      Total_Inst_Hrs_YoY_Change_Per2 Cons_Sent NJURN      RPM CPIUrban CPIMedian
## 1:                                NA      NA   NA      NA  234.747  1.947530
## 2:                                NA      NA   6.8      NA  235.342  1.954495
## 3:                                NA     98.1   6.7 65975447  235.976  2.433336
## 4:                                NA     95.4   6.3 59784666  236.222  2.962651
## 5:                                NA     93.0   5.8 75751609  237.001  2.501291
## 6:                                NA     95.9   6.0 73090871  237.657  2.972400
```

```
Tng_Ctr_Hour2[, Quarter := factor(Quarter, ordered = T)]
Tng_Ctr_Hour2[, Month := factor(Month, ordered = T)]
summary(Tng_Ctr_Hour2)
```

```
##      Year      Yr      Quarter      Month      Device_Hrs
## 2015-01: 1   Length:81   Q1:21   Apr      : 7   Min.      : 222.8
## 2015-02: 1   Class :character   Q2:21   Aug      : 7   1st Qu.: 899.0
## 2015-03: 1   Mode  :character   Q3:21   Feb      : 7   Median :1008.0
## 2015-04: 1                                   Q4:18   Jan      : 7   Mean    : 990.1
## 2015-05: 1                                   Jul      : 7   3rd Qu.:1101.7
## 2015-06: 1                                   Jun      : 7   Max.    :1519.9
## (Other):75                                   (Other):39
## DH_Prev_Year      DH_YoY_Change      DH_YoY_Ch_Per      Total_Inst_Hrs
## Length:81         Length:81         Length:81         Min.      : 504.6
## Class :character   Class :character   Class :character   1st Qu.:1937.3
## Mode  :character   Mode  :character   Mode  :character   Median :2203.2
##                                     Mean    :2165.7
##                                     3rd Qu.:2446.8
##                                     Max.    :3084.1
##
## Total_Inst_Hrs_Prev_Year Inst_Hrs_YoY_Change Total_Inst_Hrs_YoY_Change_Per2
## Length:81              Length:81              Length:81
## Class :character        Class :character        Class :character
## Mode  :character        Mode  :character        Mode  :character
##
##
##
## Cons_Sent      NJURN      RPM      CPIUrban
## Min.      : 71.80   Min.      : 2.900   Min.      : 2908236   Min.      :234.7
## 1st Qu.: 89.45   1st Qu.: 4.075   1st Qu.: 68377378   1st Qu.:241.2
## Median : 93.80   Median : 4.850   Median : 77274227   Median :250.8
## Mean    : 92.00   Mean    : 5.607   Mean    : 70811208   Mean    :250.2
## 3rd Qu.: 97.95   3rd Qu.: 6.225   3rd Qu.: 85417423   3rd Qu.:257.4
## Max.    :101.40   Max.    :16.600   Max.    :101794185   Max.    :274.1
## NA's     :2      NA's      :1      NA's      :2
## CPIMedian
## Min.      :0.9755
## 1st Qu.:2.1551
## Median :2.5922
## Mean    :2.5862
## 3rd Qu.:2.9557
## Max.    :5.5690
##
```

```
str(Tng_Ctr_Hour2)
```

```
## Classes 'data.table' and 'data.frame': 81 obs. of 17 variables:
## $ Year      : Factor w/ 81 levels "2015-01","2015-02",...: 1 2 3 4 5 6 7 8 9 10
## $ Yr        : chr "2015" "2015" "2015" "2015" ...
## $ Quarter   : Ord.factor w/ 4 levels "Q1"<"Q2"<"Q3"<...: 1 1 1 2 2 2 3 3 3 4 ...
## $ Month     : Ord.factor w/ 12 levels "Apr"<"Aug"<"Dec"<...: 5 4 8 1 9 7 6 2 12
```

```
## $ Device_Hrs          : num  960 944 1429 1097 916 ...
## $ DH_Prev_Year        : chr   "NA" "NA" "NA" "NA" ...
## $ DH_YoY_Change       : chr   "NA" "NA" "NA" "NA" ...
## $ DH_YoY_Ch_Per       : chr   "NA" "NA" "NA" "NA" ...
## $ Total_Inst_Hrs      : num  1701 1614 2533 2152 1695 ...
## $ Total_Inst_Hrs_Prev_Year : chr   "NA" "NA" "NA" "NA" ...
## $ Inst_Hrs_YoY_Change : chr   "NA" "NA" "NA" "NA" ...
## $ Total_Inst_Hrs_YoY_Change_Per2: chr   "NA" "NA" "NA" "NA" ...
## $ Cons_Sent           : num   NA NA 98.1 95.4 93 95.9 90.7 96.1 93.1 91.9 ...
## $ NJURN               : num   NA 6.8 6.7 6.3 5.8 6 5.9 6.2 5.5 5.1 ...
## $ RPM                 : num   NA NA 65975447 59784666 75751609 ...
## $ CPIUrban            : num   235 235 236 236 237 ...
## $ CPIMedian           : num   1.95 1.95 2.43 2.96 2.5 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

```
Mach1 = subset(Tng_Ctr_Hour2, select = c(Yr, Month, Device_Hrs, Cons_Sent, NJURN, RPM, CPIUrban, CPIMedian))
```

```
Mach1$Yr = as.numeric(Mach1$Yr)
```

```
class(Mach1$Yr)
```

```
## [1] "numeric"
```

```
str(Mach1)
```

```
## Classes 'data.table' and 'data.frame':  81 obs. of  8 variables:
## $ Yr          : num  2015 2015 2015 2015 2015 ...
## $ Month       : Ord.factor w/ 12 levels "Apr"<"Aug"<"Dec"<...: 5 4 8 1 9 7 6 2 12 11 ...
## $ Device_Hrs : num  960 944 1429 1097 916 ...
## $ Cons_Sent  : num   NA NA 98.1 95.4 93 95.9 90.7 96.1 93.1 91.9 ...
## $ NJURN      : num   NA 6.8 6.7 6.3 5.8 6 5.9 6.2 5.5 5.1 ...
## $ RPM        : num   NA NA 65975447 59784666 75751609 ...
## $ CPIUrban   : num   235 235 236 236 237 ...
## $ CPIMedian  : num   1.95 1.95 2.43 2.96 2.5 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

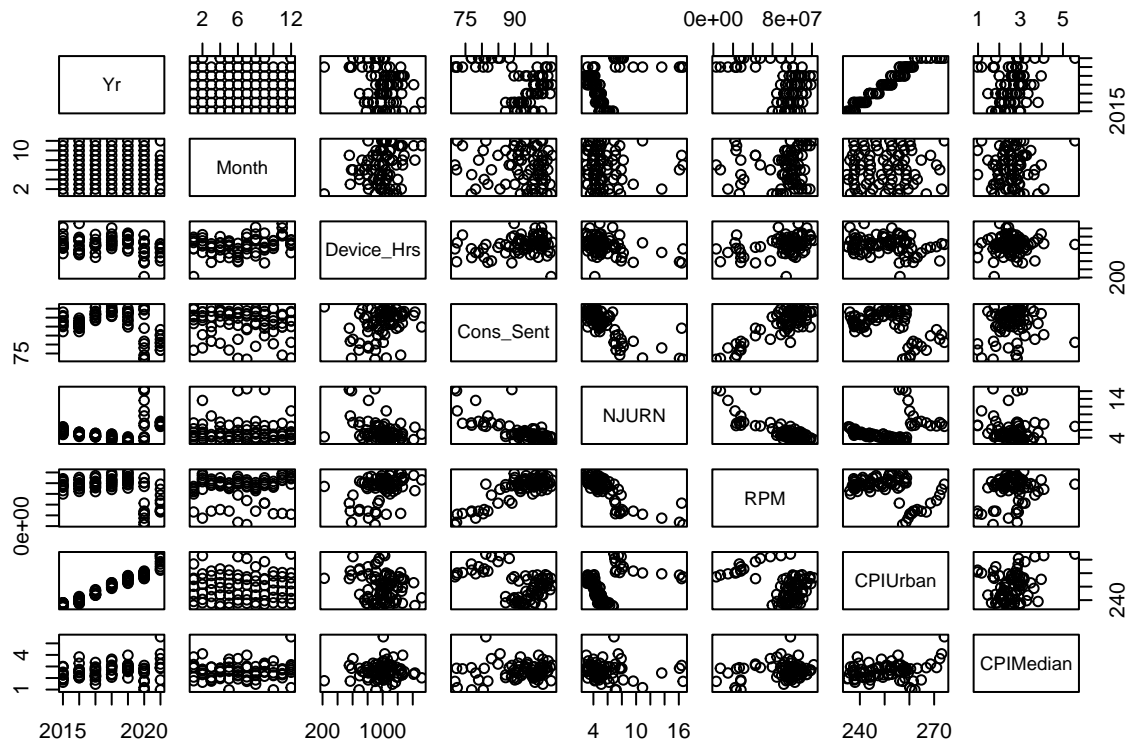
```
summary(Mach1)
```

```
##           Yr           Month      Device_Hrs      Cons_Sent      NJURN
## Min.      :2015    Apr       : 7    Min.      : 222.8    Min.      : 71.80    Min.      : 2.900
## 1st Qu.:2016    Aug       : 7    1st Qu.: 899.0    1st Qu.: 89.45    1st Qu.: 4.075
## Median :2018    Feb       : 7    Median :1008.0    Median : 93.80    Median : 4.850
## Mean   :2018    Jan       : 7    Mean   : 990.1    Mean   : 92.00    Mean   : 5.607
## 3rd Qu.:2020    Jul       : 7    3rd Qu.:1101.7    3rd Qu.: 97.95    3rd Qu.: 6.225
## Max.    :2021    Jun       : 7    Max.    :1519.9    Max.    :101.40    Max.    :16.600
##           (Other):39              NA's      :2          NA's      :1
##           RPM           CPIUrban      CPIMedian
## Min.      : 2908236    Min.      :234.7    Min.      :0.9755
## 1st Qu.: 68377378    1st Qu.:241.2    1st Qu.:2.1551
## Median : 77274227    Median :250.8    Median :2.5922
## Mean   : 70811208    Mean   :250.2    Mean   :2.5862
```

```
## 3rd Qu.: 85417423 3rd Qu.:257.4 3rd Qu.:2.9557
## Max. :101794185 Max. :274.1 Max. :5.5690
## NA's :2
```

```
TrainMach = Mach1
```

```
pairs(TrainMach)
```



```
TrainMach$Yr[TrainMach$Yr >= 2020] <- NA
```

```
summary(TrainMach)
```

```
##      Yr      Month      Device_Hrs      Cons_Sent      NJURN
## Min.   :2015   Apr    : 7   Min.    : 222.8   Min.    : 71.80   Min.    : 2.900
## 1st Qu.:2016   Aug    : 7   1st Qu.: 899.0   1st Qu.: 89.45   1st Qu.: 4.075
## Median :2017   Feb    : 7   Median :1008.0 Median : 93.80   Median : 4.850
## Mean   :2017   Jan    : 7   Mean   : 990.1   Mean   : 92.00   Mean   : 5.607
## 3rd Qu.:2018   Jul    : 7   3rd Qu.:1101.7 3rd Qu.: 97.95   3rd Qu.: 6.225
## Max.   :2019   Jun    : 7   Max.   :1519.9 Max.   :101.40   Max.   :16.600
## NA's   :21     (Other):39      NA's   :2      NA's   :1
##      RPM      CPIUrban      CPIMedian
## Min.   : 2908236   Min.   :234.7   Min.   :0.9755
## 1st Qu.: 68377378 1st Qu.:241.2 1st Qu.:2.1551
## Median : 77274227 Median :250.8  Median :2.5922
```

```
## Mean : 70811208 Mean :250.2 Mean :2.5862
## 3rd Qu.: 85417423 3rd Qu.:257.4 3rd Qu.:2.9557
## Max. :101794185 Max. :274.1 Max. :5.5690
## NA's :2
```

```
TrainMach = na.omit(TrainMach)
```

```
summary(TrainMach)
```

```
##      Yr      Month      Device_Hrs      Cons_Sent      NJURN
## Min. :2015 Apr : 5 Min. : 693.3 Min. : 87.20 Min. :2.900
## 1st Qu.:2016 Aug : 5 1st Qu.: 963.8 1st Qu.: 92.15 1st Qu.:3.900
## Median :2017 Dec : 5 Median :1041.0 Median : 95.80 Median :4.500
## Mean :2017 Jul : 5 Mean :1051.1 Mean : 95.08 Mean :4.497
## 3rd Qu.:2018 Jun : 5 3rd Qu.:1139.1 3rd Qu.: 98.17 3rd Qu.:5.075
## Max. :2019 Mar : 5 Max. :1519.9 Max. :101.40 Max. :6.700
##      (Other):28
##      RPM      CPIUrban      CPIMedian
## Min. : 59784666 Min. :236.0 Min. :1.465
## 1st Qu.: 74902332 1st Qu.:239.7 1st Qu.:2.233
## Median : 80129218 Median :244.7 Median :2.588
## Mean : 80977423 Mean :246.1 Mean :2.605
## 3rd Qu.: 87364404 3rd Qu.:252.4 3rd Qu.:2.959
## Max. :101794185 Max. :258.2 Max. :3.975
##
```

```
LM_Mach = lm(Device_Hrs ~ .-Yr, TrainMach)
summary(LM_Mach )
```

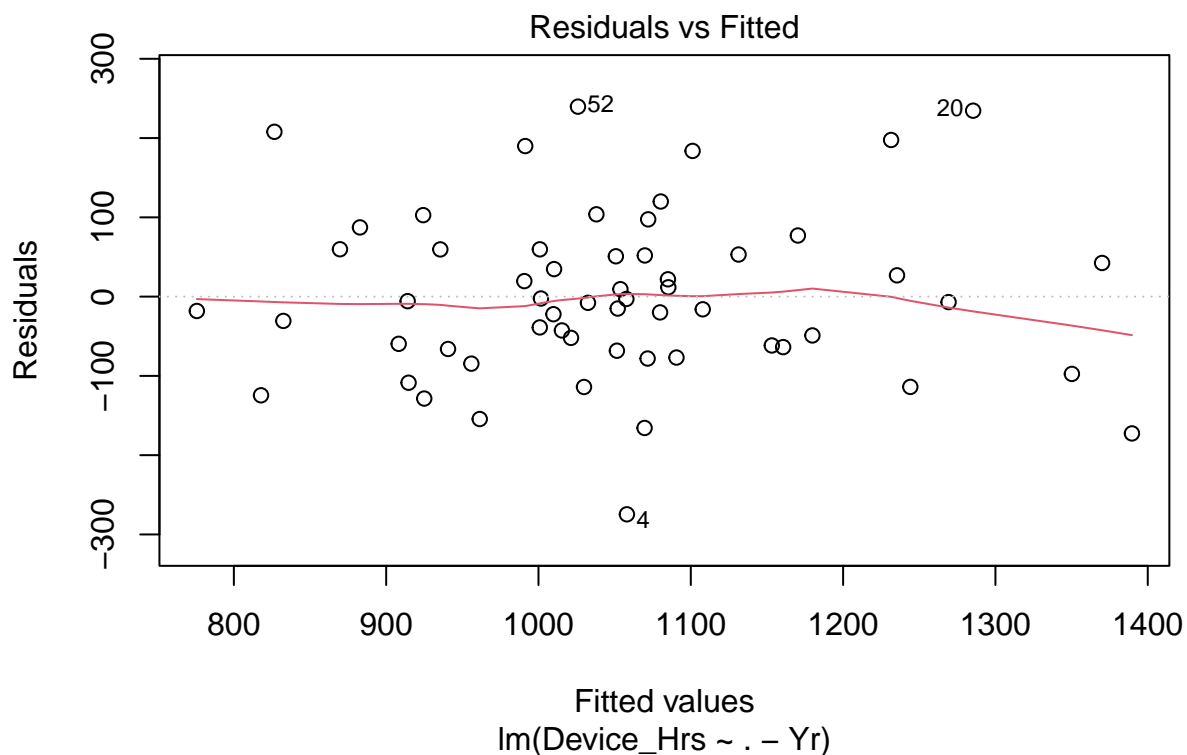
```
##
## Call:
## lm(formula = Device_Hrs ~ . - Yr, data = TrainMach)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -274.597  -65.467   -7.367   57.931  239.647
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.411e+03  2.004e+03  -1.203  0.23597
## Month.L      -1.455e+02  2.274e+02  -0.640  0.52574
## Month.Q      -8.353e+01  1.482e+02  -0.564  0.57610
## Month.C      -5.199e+02  2.641e+02  -1.969  0.05577 .
## Month^4       4.060e+01  1.092e+02   0.372  0.71182
## Month^5      -2.869e+02  1.689e+02  -1.699  0.09688 .
## Month^6       2.007e+02  2.315e+02   0.867  0.39101
## Month^7      -2.171e+02  7.985e+01  -2.719  0.00957 **
## Month^8      -1.228e+02  8.432e+01  -1.457  0.15284
## Month^9       1.789e+02  9.859e+01   1.814  0.07693 .
## Month^10      3.397e+02  1.964e+02   1.730  0.09120 .
## Month^11     -1.534e+02  8.742e+01  -1.755  0.08674 .
## Cons_Sent     2.953e+00  6.002e+00   0.492  0.62537
```

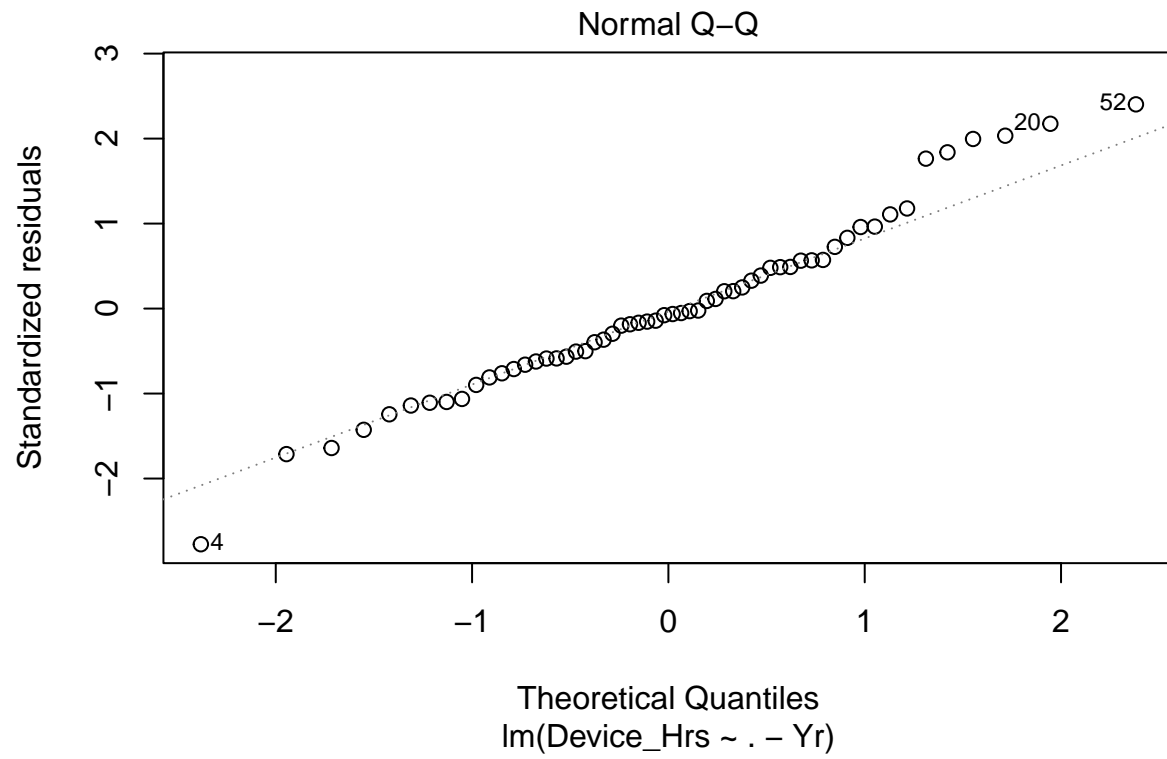
```
## NJURN      1.245e+02  7.230e+01  1.722  0.09252 .
## RPM        2.597e-05  1.893e-05  1.372  0.17763
## CPIUrban   2.644e+00  1.086e+01  0.243  0.80889
## CPIMedian  -4.993e+01  4.008e+01  -1.246  0.21996
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 124.3 on 41 degrees of freedom
## Multiple R-squared:  0.6188, Adjusted R-squared:  0.4701
## F-statistic:  4.16 on 16 and 41 DF,  p-value: 0.0001143
```

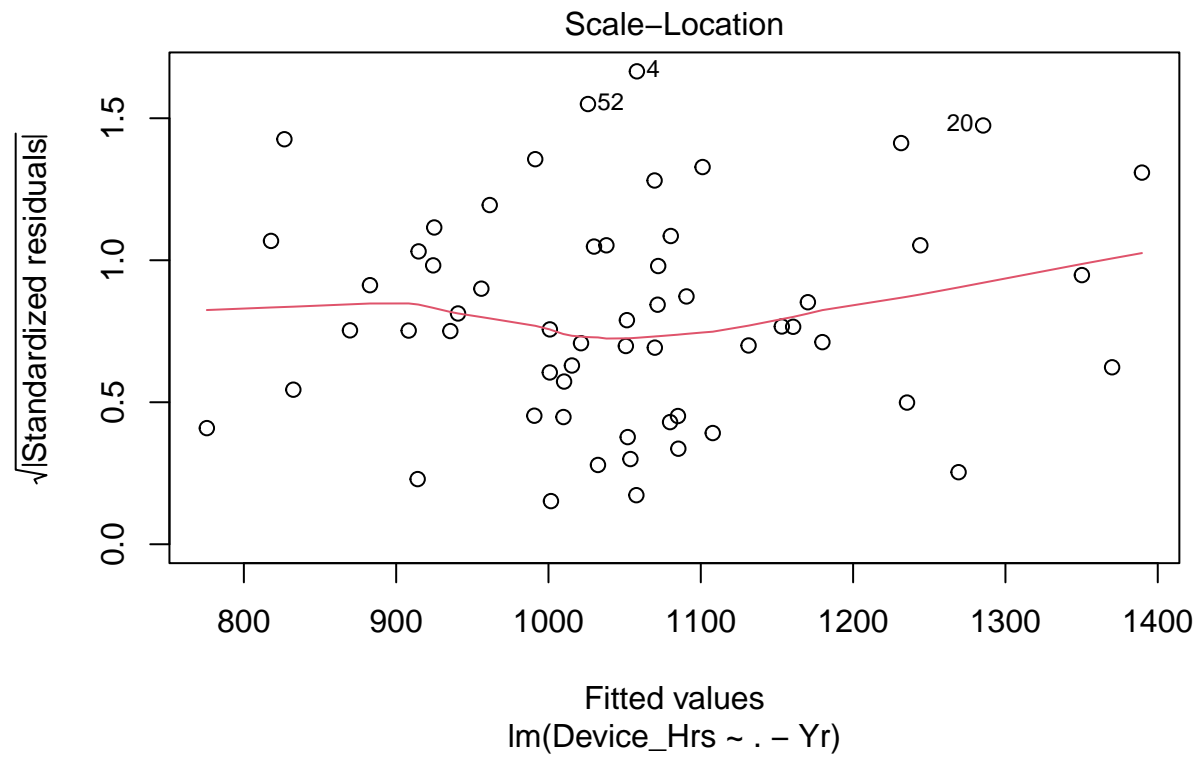
We show a relatively strong correlation between the model and Device Hours

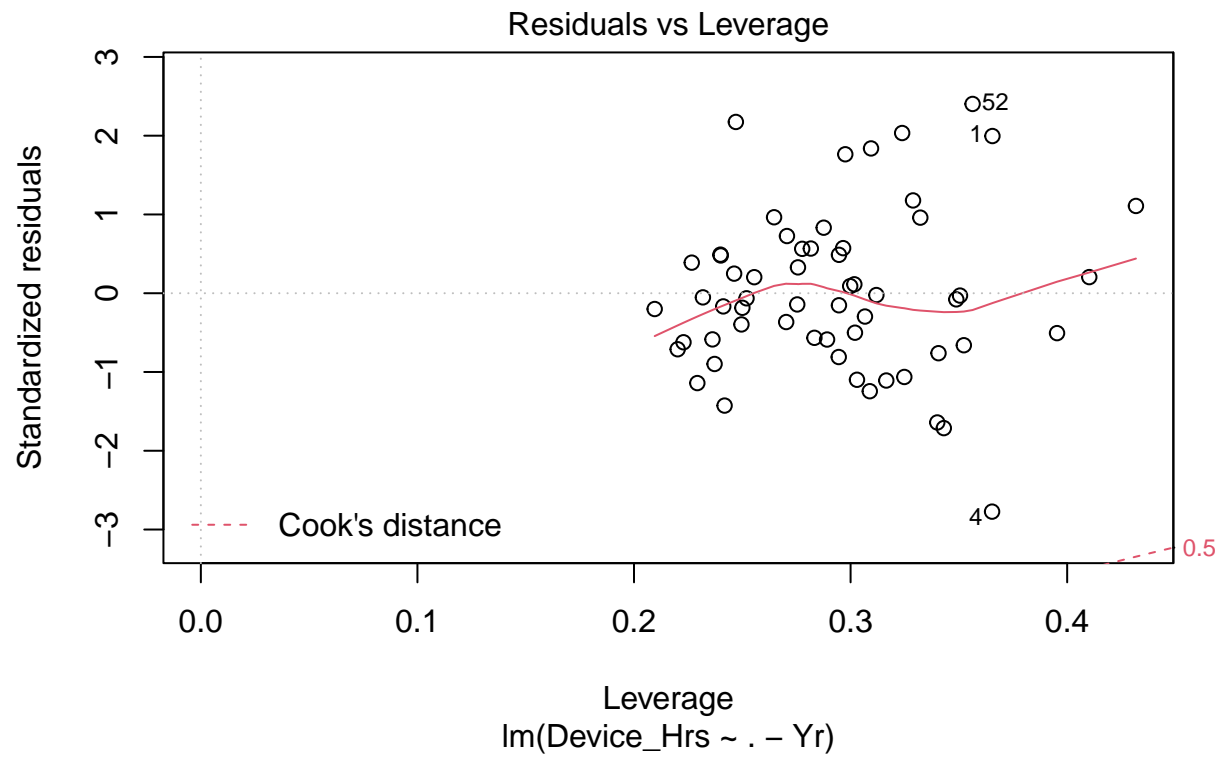
We see some correlation between Consumer Sentiment and NJ Unemployment, and Consumer Sentiment and Revenue Passenger Miles. Its possible there is some redundancy. The CPI Urban and CPI Median don't show a very strong correlation because CPI Urban is the CPI level, while CPI Median is the percentage change. That is why we're including both.

```
plot(LM_Mach)
```





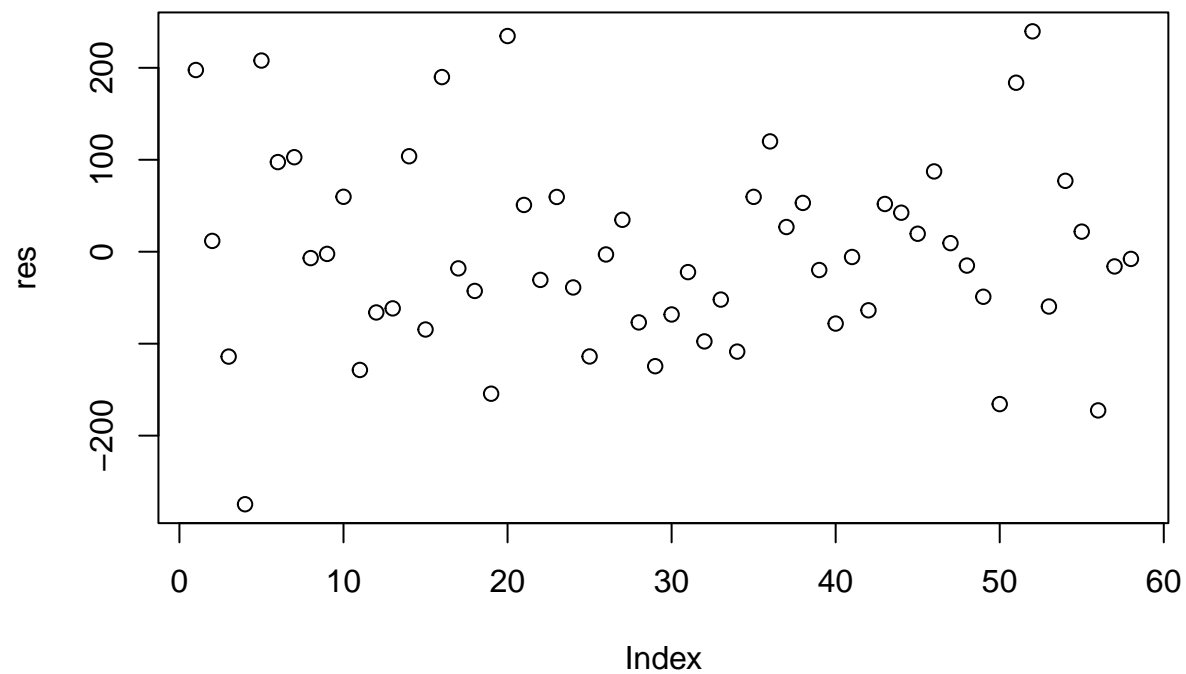




Our model looks like it is generally normal

Let's look at the residuals

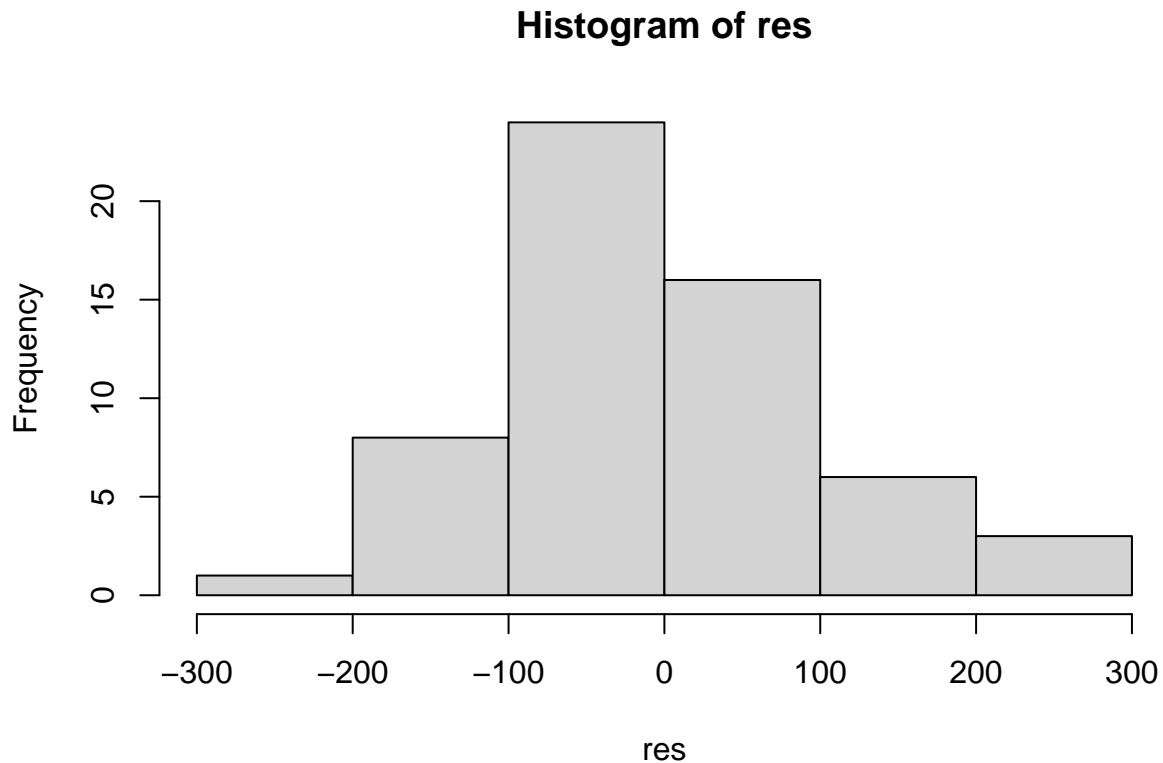
```
res= residuals(LM_Mach)
plot(res)
```



There doesn't appear to be any pattern, which leads us to believe the model is not missing any factors.

Lets take other looks at the residuals

```
hist(res)
```



This looks slightly skewed left, so not completely normal. Maybe we can get a stronger model by removing some independent variables and simplifying the model.

Stepwise

Because we have multiple independent variables, it is important to run a stepwise to see if we can get a stronger model.

```
Mach.Step = step(LM_Mach)
```

```
## Start:  AIC=573.32
## Device_Hrs ~ (Yr + Month + Cons_Sent + NJURN + RPM + CPIUrban +
##      CPIMedian) - Yr
##
##           Df Sum of Sq    RSS   AIC
## - CPIUrban  1      916 634457 571.40
## - Cons_Sent  1     3740 637281 571.66
## <none>                 633541 573.32
## - CPIMedian  1     23977 657519 573.48
## - RPM        1     29073 662614 573.92
## - NJURN      1     45846 679388 575.37
## - Month     11    818714 1452255 599.43
##
## Step:  AIC=571.4
```

```
## Device_Hrs ~ Month + Cons_Sent + NJURN + RPM + CPIMedian
```

```
##
```

	Df	Sum of Sq	RSS	AIC
## - Cons_Sent	1	5146	639603	569.87
## <none>			634457	571.40
## - CPIMedian	1	24619	659076	571.61
## - NJURN	1	44968	679425	573.38
## - RPM	1	73873	708330	575.79
## - Month	11	983547	1618004	603.70

```
##
```

```
## Step: AIC=569.87
```

```
## Device_Hrs ~ Month + NJURN + RPM + CPIMedian
```

```
##
```

	Df	Sum of Sq	RSS	AIC
## <none>			639603	569.87
## - CPIMedian	1	23613	663216	569.98
## - NJURN	1	51057	690660	572.33
## - RPM	1	94918	734521	575.90
## - Month	11	984794	1624397	601.93

```
Mach.Step
```

```
##
```

```
## Call:
```

```
## lm(formula = Device_Hrs ~ Month + NJURN + RPM + CPIMedian, data = TrainMach)
```

```
##
```

```
## Coefficients:
```

	Month.L	Month.Q	Month.C	Month^4	Month^5
## (Intercept)	-1.949e+03	-2.188e+02	-1.262e+02	-5.971e+02	7.313e+01
## Month^6	2.667e+02	-2.324e+02	-1.138e+02	2.045e+02	3.890e+02
## NJURN		RPM	CPIMedian		
##	1.279e+02	3.157e-05	-4.940e+01		

```
summary(Mach.Step)
```

```
##
```

```
## Call:
```

```
## lm(formula = Device_Hrs ~ Month + NJURN + RPM + CPIMedian, data = TrainMach)
```

```
##
```

```
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-265.886	-71.441	-9.159	59.434	235.481

```
##
```

```
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	-1.949e+03	1.294e+03	-1.506	0.139368
## Month.L	-2.188e+02	1.461e+02	-1.497	0.141649
## Month.Q	-1.262e+02	1.144e+02	-1.104	0.275915
## Month.C	-5.971e+02	1.761e+02	-3.390	0.001506 **
## Month^4	7.313e+01	7.917e+01	0.924	0.360762
## Month^5	-3.253e+02	1.336e+02	-2.435	0.019107 *
## Month^6	2.667e+02	1.646e+02	1.621	0.112353

```
## Month^7      -2.324e+02  6.424e+01  -3.618 0.000775 ***
## Month^8      -1.138e+02  7.854e+01  -1.449 0.154609
## Month^9       2.045e+02  6.463e+01   3.164 0.002858 **
## Month^10     3.890e+02  1.423e+02   2.733 0.009077 **
## Month^11     -1.600e+02  8.506e+01  -1.881 0.066716 .
## NJURN        1.279e+02  6.906e+01   1.853 0.070795 .
## RPM          3.157e-05  1.250e-05   2.526 0.015296 *
## CPIMedian    -4.940e+01  3.921e+01  -1.260 0.214483
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 122 on 43 degrees of freedom
## Multiple R-squared:  0.6152, Adjusted R-squared:  0.4899
## F-statistic:  4.91 on 14 and 43 DF,  p-value: 2.613e-05
```

```
Test = lm(formula = Device_Hrs ~ Month + Month:RPM + Month:NJURN + Month:CPIUrban + RPM, data = Mach1)
```

```
summary(Test)
```

```
##
## Call:
## lm(formula = Device_Hrs ~ Month + Month:RPM + Month:NJURN + Month:CPIUrban +
##     RPM, data = Mach1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -215.97  -54.94   -1.57    41.82   354.51
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.094e+02  2.453e+03   0.208 0.836842
## Month.L        -1.775e+03  8.950e+03  -0.198 0.844113
## Month.Q         1.584e+04  6.954e+03   2.278 0.029745 *
## Month.C        -9.461e+03  5.864e+03  -1.613 0.116774
## Month^4         2.643e+03  8.327e+03   0.317 0.753024
## Month^5        -5.298e+03  1.089e+04  -0.487 0.629938
## Month^6        -3.839e+02  1.036e+04  -0.037 0.970669
## Month^7        -1.258e+04  1.026e+04  -1.226 0.229346
## Month^8        -3.852e+03  8.652e+03  -0.445 0.659292
## Month^9         1.012e+03  6.573e+03   0.154 0.878638
## Month^10        7.006e+02  7.821e+03   0.090 0.929196
## Month^11       -1.148e+03  7.190e+03  -0.160 0.874138
## RPM            8.269e-06  9.542e-06   0.867 0.392863
## Month.L:RPM     -4.383e-07  3.040e-05  -0.014 0.988587
## Month.Q:RPM     -7.143e-05  2.880e-05  -2.480 0.018766 *
## Month.C:RPM      3.160e-05  2.511e-05   1.259 0.217544
## Month^4:RPM     -3.799e-06  3.154e-05  -0.120 0.904900
## Month^5:RPM      9.975e-06  3.809e-05   0.262 0.795122
## Month^6:RPM     -2.074e-05  3.564e-05  -0.582 0.564761
## Month^7:RPM      4.457e-05  3.772e-05   1.181 0.246390
## Month^8:RPM      1.275e-05  3.419e-05   0.373 0.711685
## Month^9:RPM     -1.287e-05  2.874e-05  -0.448 0.657564
## Month^10:RPM    -2.562e-06  3.465e-05  -0.074 0.941531
```

```
## Month^11:RPM      3.277e-05  3.608e-05  0.908 0.370761
## MonthApr:NJURN    -4.327e+02  1.605e+02 -2.697 0.011212 *
## MonthAug:NJURN     1.446e+02  3.392e+02  0.426 0.672832
## MonthDec:NJURN     -9.442e+00  2.914e+02 -0.032 0.974359
## MonthFeb:NJURN      8.097e+01  3.965e+02  0.204 0.839501
## MonthJan:NJURN      2.602e+02  7.183e+02  0.362 0.719645
## MonthJul:NJURN      4.718e+02  2.891e+02  1.632 0.112782
## MonthJun:NJURN      1.705e+01  1.204e+02  0.142 0.888287
## MonthMar:NJURN      1.310e+02  2.010e+02  0.651 0.519584
## MonthMay:NJURN     -2.656e+01  2.490e+01 -1.067 0.294253
## MonthNov:NJURN      1.908e+02  3.980e+02  0.479 0.635031
## MonthOct:NJURN     -3.457e+02  7.831e+02 -0.441 0.661924
## MonthSep:NJURN     -1.302e+02  1.447e+02 -0.900 0.375278
## MonthApr:CPIUrban -4.202e+01  9.738e+00 -4.316 0.000151 ***
## MonthAug:CPIUrban  3.036e+00  6.107e+00  0.497 0.622675
## MonthDec:CPIUrban  2.874e+00  1.002e+01  0.287 0.776258
## MonthFeb:CPIUrban  4.448e+00  9.477e+00  0.469 0.642158
## MonthJan:CPIUrban  1.619e+01  2.330e+01  0.695 0.492353
## MonthJul:CPIUrban  4.556e+00  8.447e+00  0.539 0.593473
## MonthJun:CPIUrban  5.835e+00  1.121e+01  0.521 0.606342
## MonthMar:CPIUrban -1.151e+01  1.247e+01 -0.923 0.363114
## MonthMay:CPIUrban  1.072e+01  7.642e+00  1.403 0.170677
## MonthNov:CPIUrban  1.154e+01  2.416e+01  0.477 0.636395
## MonthOct:CPIUrban -2.551e+01  4.828e+01 -0.528 0.600939
## MonthSep:CPIUrban  4.517e+00  5.424e+00  0.833 0.411320
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 150.2 on 31 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.804, Adjusted R-squared:  0.5069
## F-statistic: 2.706 on 47 and 31 DF, p-value: 0.002178
```

```
Test.Step = step(Test)
```

```
## Start: AIC=813.94
## Device_Hrs ~ Month + Month:RPM + Month:NJURN + Month:CPIUrban +
## RPM
##
##           Df Sum of Sq    RSS    AIC
## <none>                 698922 813.94
## - Month:NJURN      12    295569 994491 817.80
## - Month:RPM         11    383682 1082604 826.51
## - Month:CPIUrban   12    546495 1245417 835.58
```

```
Test.Step
```

```
##
## Call:
## lm(formula = Device_Hrs ~ Month + Month:RPM + Month:NJURN + Month:CPIUrban +
## RPM, data = Mach1)
##
## Coefficients:
```



```
##      (Intercept)      Month.L      Month.Q      Month.C
##      5.094e+02      -1.775e+03      1.584e+04      -9.461e+03
##      Month^4      Month^5      Month^6      Month^7
##      2.643e+03      -5.298e+03      -3.839e+02      -1.258e+04
##      Month^8      Month^9      Month^10      Month^11
##      -3.852e+03      1.012e+03      7.006e+02      -1.148e+03
##      RPM      Month.L:RPM      Month.Q:RPM      Month.C:RPM
##      8.269e-06      -4.383e-07      -7.143e-05      3.160e-05
##      Month^4:RPM      Month^5:RPM      Month^6:RPM      Month^7:RPM
##      -3.799e-06      9.975e-06      -2.074e-05      4.457e-05
##      Month^8:RPM      Month^9:RPM      Month^10:RPM      Month^11:RPM
##      1.275e-05      -1.287e-05      -2.562e-06      3.277e-05
##      MonthApr:NJURN      MonthAug:NJURN      MonthDec:NJURN      MonthFeb:NJURN
##      -4.327e+02      1.446e+02      -9.442e+00      8.097e+01
##      MonthJan:NJURN      MonthJul:NJURN      MonthJun:NJURN      MonthMar:NJURN
##      2.602e+02      4.718e+02      1.705e+01      1.310e+02
##      MonthMay:NJURN      MonthNov:NJURN      MonthOct:NJURN      MonthSep:NJURN
##      -2.656e+01      1.908e+02      -3.457e+02      -1.302e+02
##      MonthApr:CPIUrban      MonthAug:CPIUrban      MonthDec:CPIUrban      MonthFeb:CPIUrban
##      -4.202e+01      3.036e+00      2.874e+00      4.448e+00
##      MonthJan:CPIUrban      MonthJul:CPIUrban      MonthJun:CPIUrban      MonthMar:CPIUrban
##      1.619e+01      4.556e+00      5.835e+00      -1.151e+01
##      MonthMay:CPIUrban      MonthNov:CPIUrban      MonthOct:CPIUrban      MonthSep:CPIUrban
##      1.072e+01      1.154e+01      -2.551e+01      4.517e+00
```

```
summary(Test.Step)
```

```
##
## Call:
## lm(formula = Device_Hrs ~ Month + Month:RPM + Month:NJURN + Month:CPIUrban +
##      RPM, data = Mach1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -215.97  -54.94   -1.57    41.82   354.51
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.094e+02  2.453e+03   0.208  0.836842
## Month.L        -1.775e+03  8.950e+03  -0.198  0.844113
## Month.Q         1.584e+04  6.954e+03   2.278  0.029745 *
## Month.C        -9.461e+03  5.864e+03  -1.613  0.116774
## Month^4         2.643e+03  8.327e+03   0.317  0.753024
## Month^5        -5.298e+03  1.089e+04  -0.487  0.629938
## Month^6        -3.839e+02  1.036e+04  -0.037  0.970669
## Month^7        -1.258e+04  1.026e+04  -1.226  0.229346
## Month^8        -3.852e+03  8.652e+03  -0.445  0.659292
## Month^9         1.012e+03  6.573e+03   0.154  0.878638
## Month^10        7.006e+02  7.821e+03   0.090  0.929196
## Month^11       -1.148e+03  7.190e+03  -0.160  0.874138
## RPM             8.269e-06  9.542e-06   0.867  0.392863
## Month.L:RPM     -4.383e-07  3.040e-05  -0.014  0.988587
## Month.Q:RPM     -7.143e-05  2.880e-05  -2.480  0.018766 *
## Month.C:RPM     3.160e-05  2.511e-05   1.259  0.217544
```

```

## Month^4:RPM      -3.799e-06  3.154e-05  -0.120  0.904900
## Month^5:RPM      9.975e-06  3.809e-05   0.262  0.795122
## Month^6:RPM     -2.074e-05  3.564e-05  -0.582  0.564761
## Month^7:RPM      4.457e-05  3.772e-05   1.181  0.246390
## Month^8:RPM      1.275e-05  3.419e-05   0.373  0.711685
## Month^9:RPM     -1.287e-05  2.874e-05  -0.448  0.657564
## Month^10:RPM    -2.562e-06  3.465e-05  -0.074  0.941531
## Month^11:RPM     3.277e-05  3.608e-05   0.908  0.370761
## MonthApr:NJURN   -4.327e+02  1.605e+02  -2.697  0.011212 *
## MonthAug:NJURN    1.446e+02  3.392e+02   0.426  0.672832
## MonthDec:NJURN   -9.442e+00  2.914e+02  -0.032  0.974359
## MonthFeb:NJURN    8.097e+01  3.965e+02   0.204  0.839501
## MonthJan:NJURN    2.602e+02  7.183e+02   0.362  0.719645
## MonthJul:NJURN    4.718e+02  2.891e+02   1.632  0.112782
## MonthJun:NJURN    1.705e+01  1.204e+02   0.142  0.888287
## MonthMar:NJURN    1.310e+02  2.010e+02   0.651  0.519584
## MonthMay:NJURN   -2.656e+01  2.490e+01  -1.067  0.294253
## MonthNov:NJURN    1.908e+02  3.980e+02   0.479  0.635031
## MonthOct:NJURN   -3.457e+02  7.831e+02  -0.441  0.661924
## MonthSep:NJURN   -1.302e+02  1.447e+02  -0.900  0.375278
## MonthApr:CPIUrban -4.202e+01  9.738e+00  -4.316  0.000151 ***
## MonthAug:CPIUrban  3.036e+00  6.107e+00   0.497  0.622675
## MonthDec:CPIUrban  2.874e+00  1.002e+01   0.287  0.776258
## MonthFeb:CPIUrban  4.448e+00  9.477e+00   0.469  0.642158
## MonthJan:CPIUrban  1.619e+01  2.330e+01   0.695  0.492353
## MonthJul:CPIUrban  4.556e+00  8.447e+00   0.539  0.593473
## MonthJun:CPIUrban  5.835e+00  1.121e+01   0.521  0.606342
## MonthMar:CPIUrban -1.151e+01  1.247e+01  -0.923  0.363114
## MonthMay:CPIUrban  1.072e+01  7.642e+00   1.403  0.170677
## MonthNov:CPIUrban  1.154e+01  2.416e+01   0.477  0.636395
## MonthOct:CPIUrban -2.551e+01  4.828e+01  -0.528  0.600939
## MonthSep:CPIUrban  4.517e+00  5.424e+00   0.833  0.411320
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 150.2 on 31 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.804, Adjusted R-squared:  0.5069
## F-statistic: 2.706 on 47 and 31 DF, p-value: 0.002178

```

Stepwise reveals that the best way to predict training hours is to use the month and the Revenue Passenger Miles. The other variables were not strong predictors.

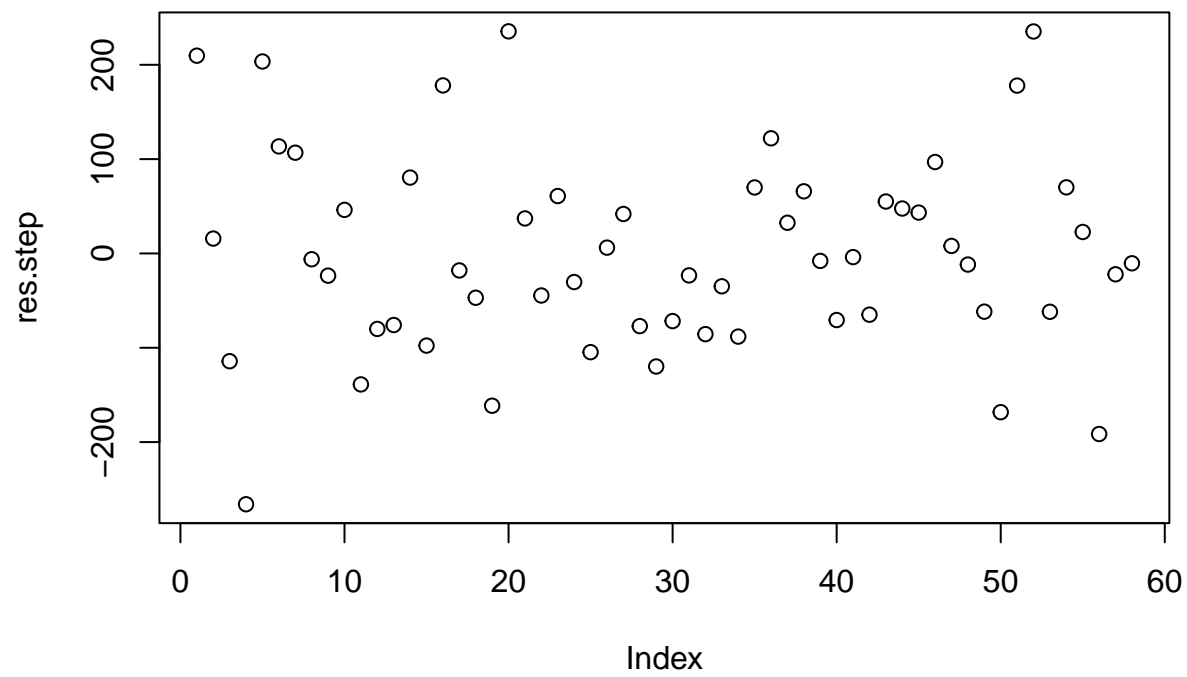
This model has a decently strong Adjusted R2, a high F-statistic, and a very low p-value.

Lets look at the residuals of the new model.

```

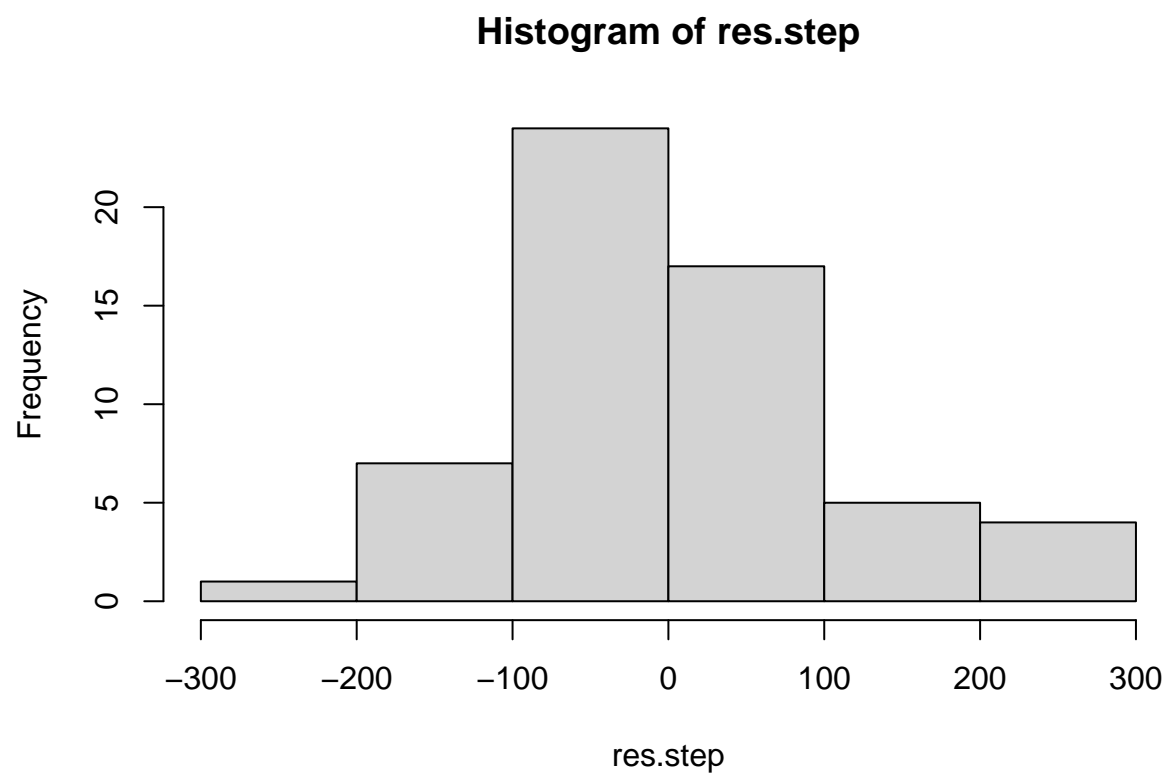
res.step= residuals(Mach.Step)
plot(res.step)

```



This also looks normally distributed.

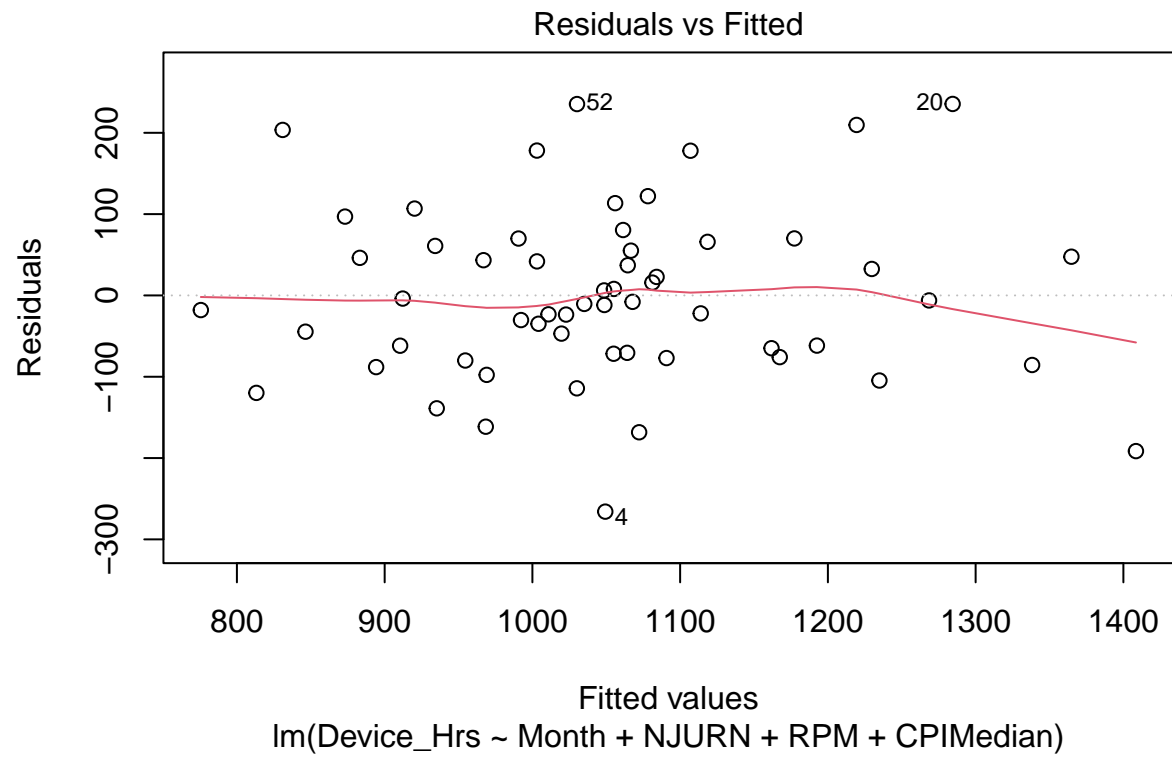
```
hist(res.step)
```

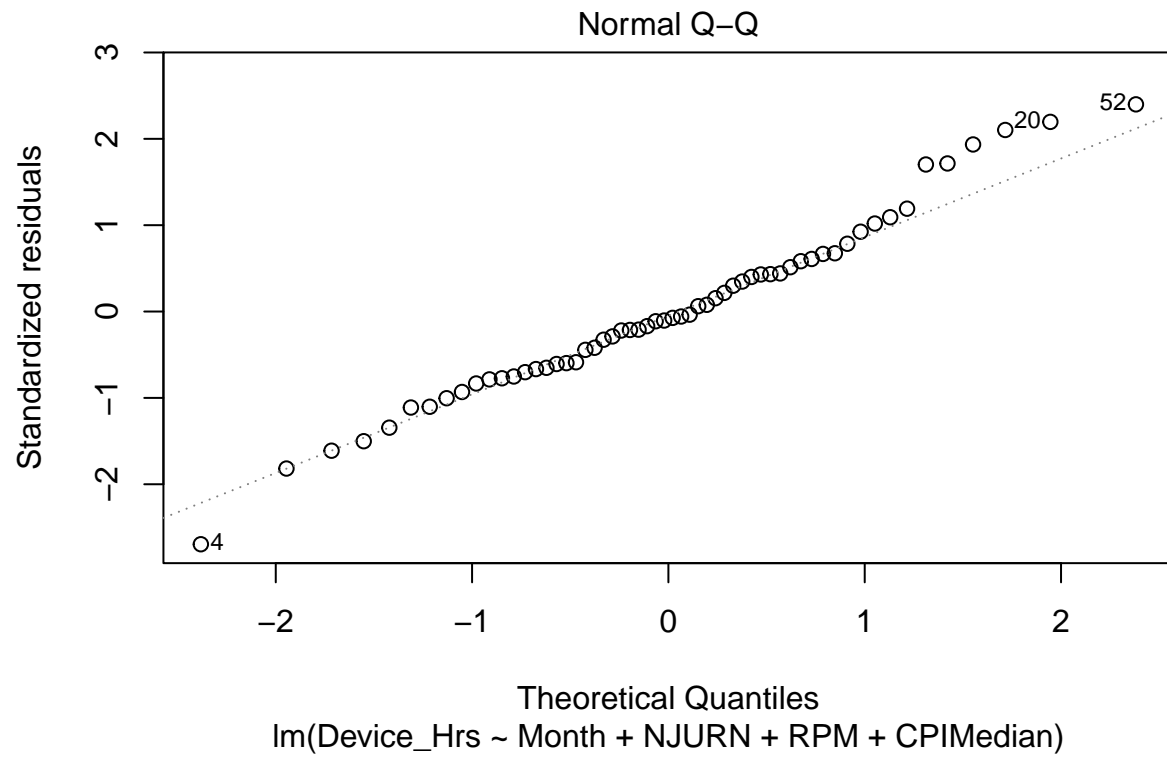


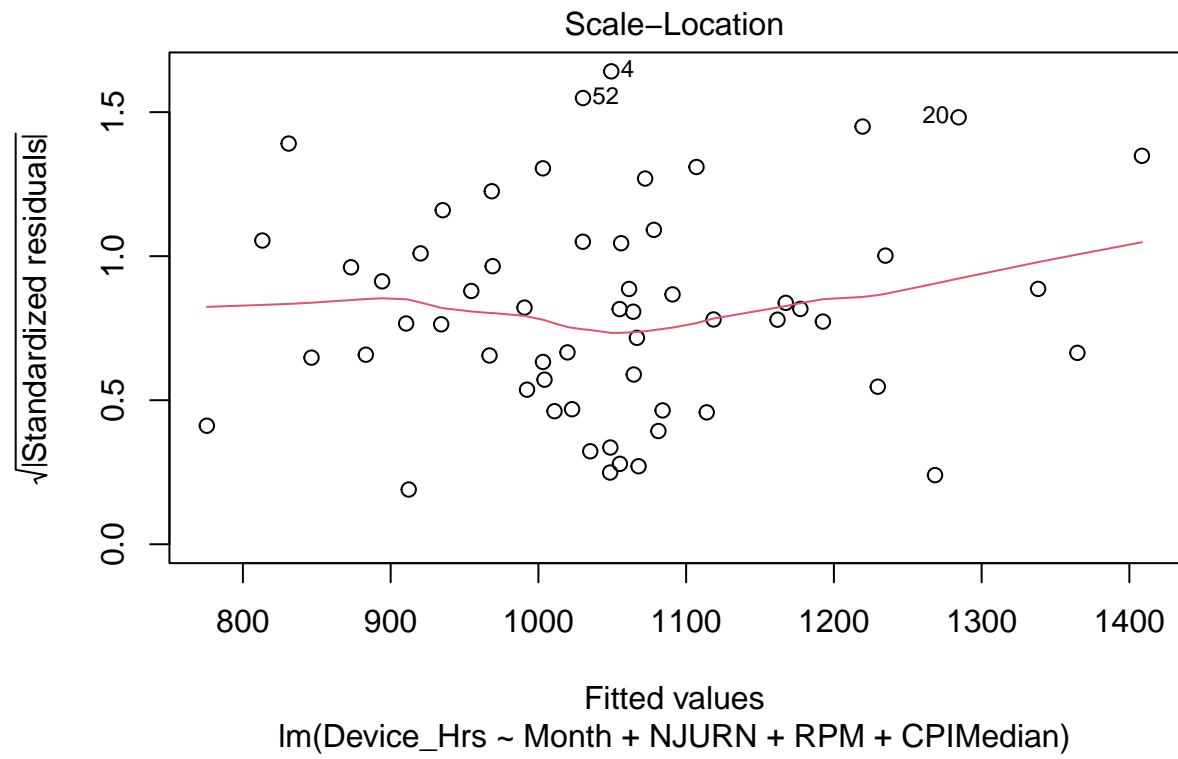
this looks to be more normally distributed, though it still slightly skewed to the left.

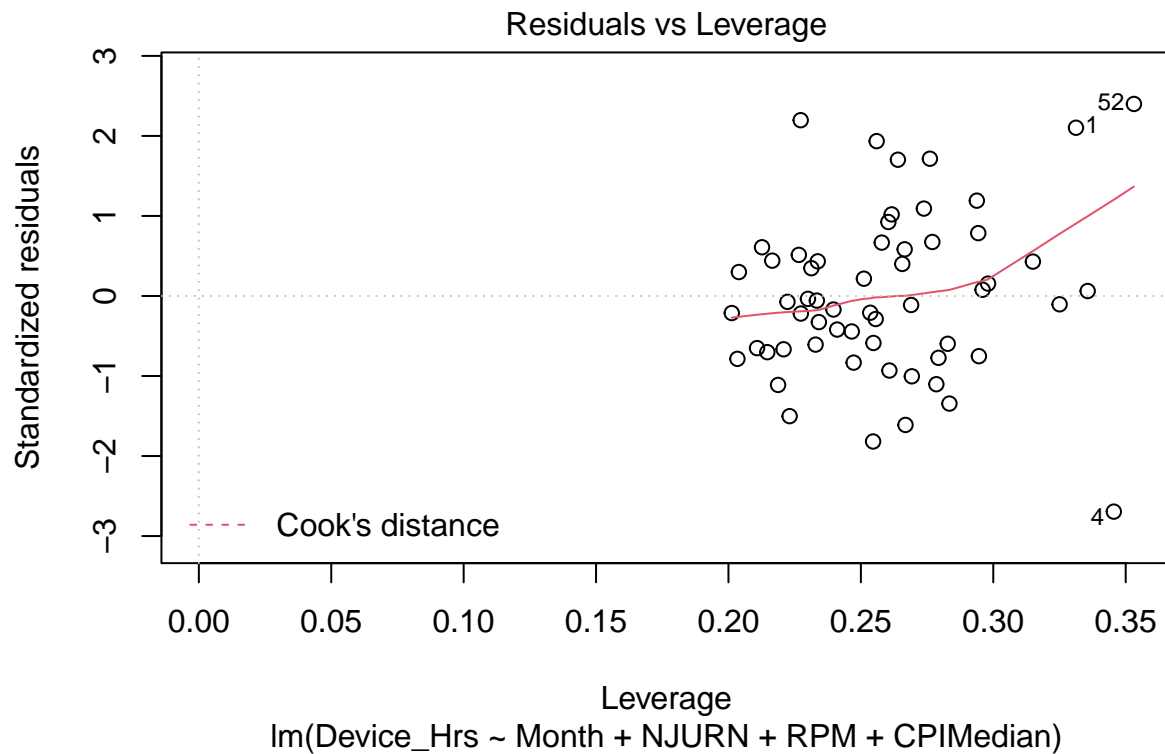
Final residual check.

```
plot(Mach.Step)
```





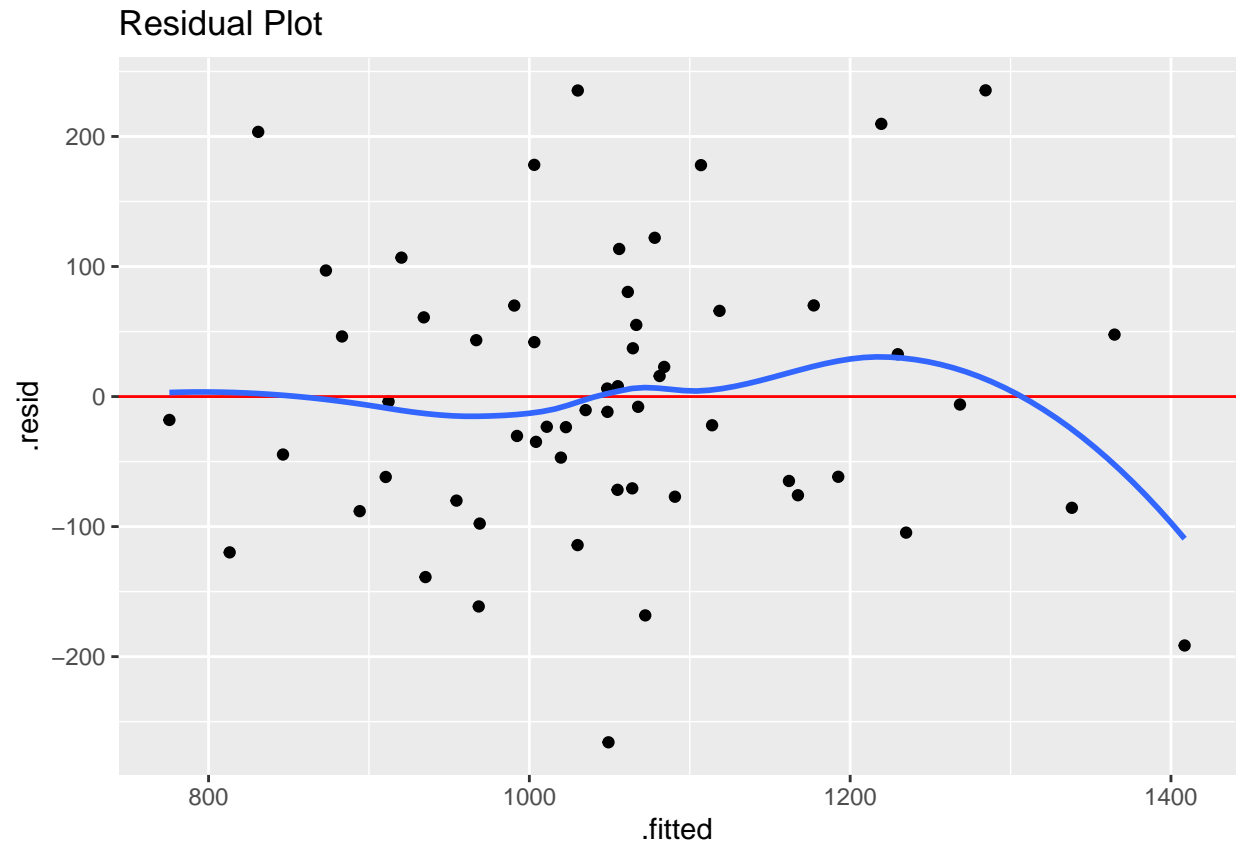




Normal QQ plot shows that the model is relatively normal.

```
p1= ggplot(Mach.Step) +
  aes(x=.fitted, y=.resid)+
  geom_point() + geom_abline(intercept = 0, slope = 0, color="red") + geom_smooth(method = "loess", se = TRUE)
  labs(title= "Residual Plot")
p1
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

We show some skewedness at the right side of the graph, but generally it appears strong..