DATA 603 Final Project Report

Inwoo Choi, Gaurav Malik, Rupal Gandhi, Brad Robinson

DECEMBER 3, 2019

Introduction

- Our group is investigating a data set of fracture parameters of gas wells in Alberta and B.C. Our objective is to determine the best fracture parameters to maximize the 36 month cumulative production of gas well. One specific question we are investigating is what is the optimum number of fracs on a horizontal well to maximize the 36 month cumulative production. Our expectation is that more fracs will clearly lead to better 36 month cumulative production.

- An obviously important industry to Calgary and Alberta is the Oil and Gas industry. In fact, 10% of the GDP of Canada is created by the Oil and Gas Industry. As an example of the importance of this industry - most homes and businesses in North America and Europe are heated with “natural gas” i.e. Methane gas. In North America about one third of electricity is generated by natural gas. Despite comments about a “climate crisis” caused by carbon dioxide emissions, home heating and electricity generation with natural gas will not change in the span of our careers. Probably, this will not change in our lifetimes.

- Large volumes of natural gas are produced and consumed every day in North America - approximately, 110 billioncubic feet or, in S.I., 110 million gigajoules. The gross sales of this product is US$330,000,000 per day or US$120,000,000,000 annually. It is an important export industry of Canada to the United States. In fact, it is an important export industry of Texas to Mexico.

- It seems that there is much profit to be made producing oil and natural gas but, the net profit margins can be razor thin. The cost to drill and stimulate a typical horizontal well could be easily $6,000,000 per well. The cost to frac a well could be $4,000,000 of the overall cost. A net profit is not guaranteed.

- Significant amounts of data are collected in the drilling and stimulation of a well. Understanding the most effective parameters to stimulate a well could conceivably result in a reduction of 10% in cost but, a 10% increase in ultimate production. Understanding potential optimum parameters takes data scientists to examine the data.

- In particular, it is important to our group since the Oil and Gas industry is a potential employer. Since all of our group have connections to the oil and gas industry we were also curious about which frac parameters are most effective. We received data from Ross Smith Energy Group which is a data analytics firm serving the oil and gas industry. We are using the data they use to assist companies in finding the best frac parameters.

Methodology

We have 11 predictor variables we are considering from the Ross Smith Energy Group dataset. In particular:

1. “RSFracJobType” i.e. is the frac fluid “slickwater” (water with surfactant and friction inhibitor), gelled frac fluid or fluid “energized” with Nitrogen.
2. “SpudToSales_DAYS” ie. the length of time the Montney reservoir is open to drilling and frac fluids - which generally will impair the permeability of the reservoir.
3. PerfInterval_FT" i.e. the length of the horizontal well that is open to the Montney reservoir. Ross Smith Energy Group works with many U.S. based companies so, the units are in feet which are used in the U.S. Oil and Gas industry.
4. “FracStages” i.e. the number of fracs that are created in the PerfInterval.
5. “AverageStageSpacing_FT” i.e. the spacing between fracs in feet.
6. “ProppantLoading_LBSPerGAL” i.e. the weight in pounds of frac sand per gallon of frac fluid. Again, U.S. Oil and Gas industry units are derived from the British Imperial system of measurements.
7. “RSProppantType” i.e. whether the sand used to prop open the well used expensive artificially created ceramic proppant, sand coated with a sticky resin or just frac sand. Frac sand is natural but it has specific properties such as roundness, purity of SiO_2 content and others. It has an engineered specification.
8. “ProppantIntensity_LBSPerFT” i.e. the weight in pounds of sand per foot of PerfInterval. The more sand injected into the rock the better the potential oil and gas production.
9. “FluidIntensity_BBLPerFT” i.e. the amount of frac fluid used per foot of PerfInterval. The more intense this volume the farther and more complex the created fractures could be.
10. “FrictionReducer_LBS” i.e. a chemical added to water to create a frac fluid called “slickwater” to make the water easier to pump.
11. Surfactant_LBS" i.e. a chemical added to water to create a frac fluid called “slickwater” which emulsifies the fluid so it can carry the frac sand i.e. proppant.

There is much science and engineering that goes into these parameters. They are all intended to maximize the volume of produced oil and gas in the initial production of a well. That is, our dependent variable 36 month cumulative production of oil and gas. This is expressed as “BOE” or barrels of oil equivalent. The produced gas in thousands of cubic feet are “converted” to a barrel of oil at a conversion of 6000 cubic feet (the energy equivalent to a barrel of oil) to one barrel of oil.

We will use a Multi-Linear regression to investigate these variables. The general approach we will take is to:

1. Investigate multi-colinearity by pairs plots and the VIF function. Variables demonstrating co-linearity will be removed from the multi-linear regression.

2. Stepwise to see which variables should be in our model at a p-value of alpha =0.05

3. Create t-tests to confirm and refine the stepwise results.

4. Investigate for interactive terms

5. Investigate for quadratic terms

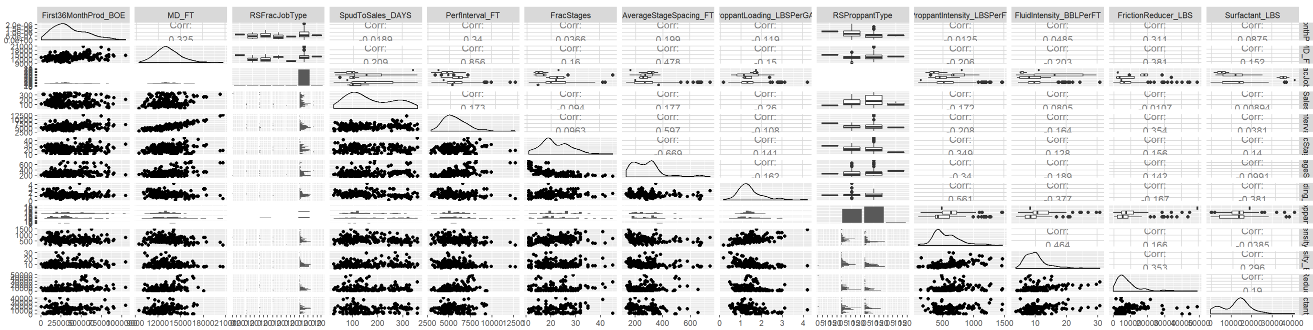
6. Investigate to determine if a log transformation or BoxCox transformation improves the model

7. Investigate for the reliability of the model by demonstrating Linearity, Independence i.e. uncorrelated error terms, Equal Variance, Normality, Outliers and a test for Multi-collinearity

1. Multi-Collinearity

- Here we look at en exploratory pair plot in order to determine how we can best proceed with the linear regression. Our dependent variable/response variable is “First36MonthProd_BOE” as we are attempting to see what factors will lead to more production.

```
ggpairs(data1, columns = c("First36MonthProd_BOE","MD_FT", "RSFracJobType","SpudToSales_DAYS", "PerfInterval_FT", "FracStage
s", "AverageStageSpacing_FT", "ProppantLoading_LBSPerGAL", "RSProppantType", "ProppantIntensity_LBSPerFT","FluidIntensity_BB
LPerFT", "FrictionReducer_LBS", "Surfactant_LBS"))
```



- Pairwise plot suggest there may be issues with multicollinearity between MD_FT and PerfInterval_FT but this will be determined via VIF test later.
    - So far everything seems to be loosely positively correlated to First36MonthProd_BOE but we see instances where there may be funneling in terms of residuals. Again this can be confirmed later in the tests.

```
vif(lm(First36MonthProd_BOE~MD_FT+PerfInterval_FT, data=data1))
```

```
##          MD_FT PerfInterval_FT
##       3.744725        3.744725
```

- VIF test for MD_FT and PerInterval_FT suggest they are inter-correlated. PerfInterval_FT is more important factor in determining production so we will keep PerfInterval_FT and remove MD_FT for further tests.

1. Initial Multi-Linear Regression Model

First let’s create a simple lenear regression and perform VIF test again to check multi-colinearity.

```
initmodel = lm(First36MonthProd_BOE~
                #MD_FT+
                 factor(RSFracJobType)+
                 SpudToSales_DAYS+
                 PerfInterval_FT+
                 FracStages+
                 AverageStageSpacing_FT+
                 ProppantLoading_LBSPerGAL+
                 factor(RSProppantType)+
                 ProppantIntensity_LBSPerFT+
                 FluidIntensity_BBLPerFT+
                 FrictionReducer_LBS+
                 Surfactant_LBS
                 , data=data1)
summary(initmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ factor(RSFracJobType) + SpudToSales_DAYS +
##     PerfInterval_FT + FracStages + AverageStageSpacing_FT + ProppantLoading_LBSPerGAL +
##     factor(RSProppantType) + ProppantIntensity_LBSPerFT + FluidIntensity_BBLPerFT +
##     FrictionReducer_LBS + Surfactant_LBS, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -421902 -108222  -33441   81408  531467
##
## Coefficients:
##                                      Estimate Std. Error t value
## (Intercept)                         2.289e+05  3.118e+05   0.734
## factor(RSFracJobType)ENERGIZED      1.903e+05  1.919e+05   0.992
## factor(RSFracJobType)SLICKCROSSLINKHYB -3.184e+05  2.233e+05  -1.426
## factor(RSFracJobType)SLICKWATER     1.005e+05  1.748e+05   0.575
## SpudToSales_DAYS                    1.760e+02  1.843e+02   0.955
## PerfInterval_FT                     9.582e+01  3.175e+01   3.018
## FracStages                         -1.565e+04  7.512e+03  -2.083
## AverageStageSpacing_FT             -6.293e+02  5.775e+02  -1.090
## ProppantLoading_LBSPerGAL          -1.522e+05  6.876e+04  -2.213
## factor(RSProppantType)SAND + RESIN  2.348e+05  1.951e+05   1.204
## factor(RSProppantType)SAND ONLY     3.898e+04  1.955e+05   0.199
## ProppantIntensity_LBSPerFT          4.161e+02  1.720e+02   2.419
## FluidIntensity_BBLPerFT            -2.930e+04  8.521e+03  -3.439
## FrictionReducer_LBS                 4.717e+00  2.275e+00   2.073
## Surfactant_LBS                      5.976e+00  2.326e+00   2.569
##                                    Pr(>|t|)
## (Intercept)                        0.464278
## factor(RSFracJobType)ENERGIZED     0.323110
## factor(RSFracJobType)SLICKCROSSLINKHYB 0.156239
## factor(RSFracJobType)SLICKWATER    0.566467
## SpudToSales_DAYS                   0.341261
## PerfInterval_FT                    0.003050 **
## FracStages                         0.039125 *
## AverageStageSpacing_FT             0.277795
## ProppantLoading_LBSPerGAL          0.028576 *
## factor(RSProppantType)SAND + RESIN 0.230917
## factor(RSProppantType)SAND ONLY    0.842282
## ProppantIntensity_LBSPerFT         0.016924 *
## FluidIntensity_BBLPerFT            0.000781 ***
## FrictionReducer_LBS                0.040070 *
## Surfactant_LBS                     0.011311 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 169900 on 133 degrees of freedom
##   (95 observations deleted due to missingness)
## Multiple R-squared:  0.4501, Adjusted R-squared:  0.3922
## F-statistic: 7.777 on 14 and 133 DF,  p-value: 8.105e-12
```

```
#vif(initmodel)
```

With our model created we can investigate further for multi-collinearity:

```
vif(initmodel)
```

```
##                                 GVIF Df GVIF^(1/(2*Df))
## factor(RSFracJobType)       1.835602  3        1.106530
## SpudToSales_DAYS            1.308702  1        1.143985
## PerfInterval_FT             6.588466  1        2.566801
## FracStages                 10.593078  1        3.254701
## AverageStageSpacing_FT     13.405211  1        3.661313
## ProppantLoading_LBSPerGAL   7.674621  1        2.770311
## factor(RSProppantType)      1.707187  2        1.143063
## ProppantIntensity_LBSPerFT  8.610077  1        2.934293
## FluidIntensity_BBLPerFT     7.741483  1        2.782352
## FrictionReducer_LBS         1.876350  1        1.369799
## Surfactant_LBS              1.682000  1        1.296920
```

FracStages and AverageStageSpacing_FT demonstrate a VIF value over 10 so there is co-linearity present. Since FracStages is an important predictor variable we want to investigate, we removed AverageStageSpacing_FT.

```
init_2 = lm(First36MonthProd_BOE~
                #MD_FT+
                 factor(RSFracJobType)+
                 SpudToSales_DAYS+
                 PerfInterval_FT+
                 FracStages+
                #AverageStageSpacing_FT
                 ProppantLoading_LBSPerGAL+
                 factor(RSProppantType)+
                 ProppantIntensity_LBSPerFT+
                 FluidIntensity_BBLPerFT+
                 FrictionReducer_LBS+
                 Surfactant_LBS
                 , data=data1)
summary(init_2)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ factor(RSFracJobType) + SpudToSales_DAYS +
##     PerfInterval_FT + FracStages + ProppantLoading_LBSPerGAL +
##     factor(RSProppantType) + ProppantIntensity_LBSPerFT + FluidIntensity_BBLPerFT +
##     FrictionReducer_LBS + Surfactant_LBS, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -419731 -104040  -23649   83553  534741
##
## Coefficients:
##                                        Estimate Std. Error t value
## (Intercept)                           9.193e+04  2.856e+05   0.322
## factor(RSFracJobType)ENERGIZED         1.886e+05  1.920e+05   0.982
## factor(RSFracJobType)SLICKCROSSLINKHYB -3.023e+05  2.230e+05  -1.355
## factor(RSFracJobType)SLICKWATER        1.114e+05  1.746e+05   0.638
## SpudToSales_DAYS                       1.669e+02  1.842e+02   0.906
## PerfInterval_FT                        6.489e+01  1.424e+01   4.557
## FracStages                            -8.016e+03  2.712e+03  -2.956
## ProppantLoading_LBSPerGAL             -1.411e+05  6.805e+04  -2.074
## factor(RSProppantType)SAND + RESIN     1.924e+05  1.913e+05   1.006
## factor(RSProppantType)SAND ONLY       -3.279e+02  1.923e+05  -0.002
## ProppantIntensity_LBSPerFT             3.677e+02  1.663e+02   2.211
## FluidIntensity_BBLPerFT               -2.722e+04  8.310e+03  -3.276
## FrictionReducer_LBS                    4.028e+00  2.187e+00   1.842
## Surfactant_LBS                         6.034e+00  2.327e+00   2.592
##                                       Pr(>|t|)
## (Intercept)                            0.74803
## factor(RSFracJobType)ENERGIZED         0.32768
## factor(RSFracJobType)SLICKCROSSLINKHYB  0.17754
## factor(RSFracJobType)SLICKWATER        0.52450
## SpudToSales_DAYS                       0.36659
## PerfInterval_FT                       1.16e-05 ***
## FracStages                             0.00368 **
## ProppantLoading_LBSPerGAL              0.04003 *
## factor(RSProppantType)SAND + RESIN     0.31629
## factor(RSProppantType)SAND ONLY        0.99864
## ProppantIntensity_LBSPerFT             0.02875 *
## FluidIntensity_BBLPerFT                0.00134 **
## FrictionReducer_LBS                    0.06773 .
## Surfactant_LBS                         0.01059 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 170100 on 134 degrees of freedom
##   (95 observations deleted due to missingness)
## Multiple R-squared:  0.4452, Adjusted R-squared:  0.3914
## F-statistic: 8.272 on 13 and 134 DF,  p-value: 4.626e-12
```

```
vif(init_2)
```

```
##                                 GVIF Df GVIF^(1/(2*Df))
## factor(RSFracJobType)      1.793791  3        1.102289
## SpudToSales_DAYS           1.306006  1        1.142806
## PerfInterval_FT            1.323612  1        1.150484
## FracStages                 1.378230  1        1.173981
## ProppantLoading_LBSPerGAL  7.506917  1        2.739875
## factor(RSProppantType)     1.631027  2        1.130096
## ProppantIntensity_LBSPerFT 8.035040  1        2.834615
## FluidIntensity_BBLPerFT    7.351833  1        2.711426
## FrictionReducer_LBS        1.731204  1        1.315752
## Surfactant_LBS             1.681128  1        1.296583
```

ProppantIntensity_LBSPerFT is removed from the model since its VIF value is over 5. The model is re-run

```
init_3 = lm(First36MonthProd_BOE~
                #MD_FT+
                 factor(RSFracJobType)+
                 SpudToSales_DAYS+
                 PerfInterval_FT+
                 FracStages+
                #AverageStageSpacing_FT
                 ProppantLoading_LBSPerGAL+
                 factor(RSProppantType)+
                #ProppantIntensity_LBSPerFT
                 FluidIntensity_BBLPerFT+
                 FrictionReducer_LBS+
                 Surfactant_LBS
                 , data=data1)
summary(init_3)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ factor(RSFracJobType) + SpudToSales_DAYS +
##     PerfInterval_FT + FracStages + ProppantLoading_LBSPerGAL +
##     factor(RSProppantType) + FluidIntensity_BBLPerFT + FrictionReducer_LBS +
##     Surfactant_LBS, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -428926 -101371  -28278   91491  534751
##
## Coefficients:
##                                      Estimate Std. Error t value
## (Intercept)                        -8.382e+04  2.782e+05  -0.301
## factor(RSFracJobType)ENERGIZED      1.600e+05  1.943e+05   0.823
## factor(RSFracJobType)SLICKCROSSLINKHYB -3.336e+05  2.257e+05  -1.478
## factor(RSFracJobType)SLICKWATER     1.112e+05  1.771e+05   0.628
## SpudToSales_DAYS                    1.164e+02  1.854e+02   0.628
## PerfInterval_FT                     6.238e+01  1.440e+01   4.333
## FracStages                         -6.007e+03  2.591e+03  -2.318
## ProppantLoading_LBSPerGAL          -8.359e+03  3.248e+04  -0.257
## factor(RSProppantType)SAND + RESIN  2.022e+05  1.940e+05   1.042
## factor(RSProppantType)SAND ONLY     1.313e+04  1.950e+05   0.067
## FluidIntensity_BBLPerFT            -1.151e+04  4.366e+03  -2.635
## FrictionReducer_LBS                 3.777e+00  2.215e+00   1.705
## Surfactant_LBS                      6.973e+00  2.321e+00   3.004
##                                    Pr(>|t|)
## (Intercept)                         0.76368
## factor(RSFracJobType)ENERGIZED      0.41173
## factor(RSFracJobType)SLICKCROSSLINKHYB  0.14180
## factor(RSFracJobType)SLICKWATER     0.53121
## SpudToSales_DAYS                    0.53134
## PerfInterval_FT                    2.85e-05 ***
## FracStages                          0.02195 *
## ProppantLoading_LBSPerGAL           0.79727
## factor(RSProppantType)SAND + RESIN  0.29918
## factor(RSProppantType)SAND ONLY     0.94640
## FluidIntensity_BBLPerFT             0.00940 **
## FrictionReducer_LBS                 0.09049 .
## Surfactant_LBS                      0.00317 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 172500 on 135 degrees of freedom
##   (95 observations deleted due to missingness)
## Multiple R-squared:  0.425,  Adjusted R-squared:  0.3739
## F-statistic: 8.315 on 12 and 135 DF,  p-value: 1.336e-11
```

```
vif(init_3)
```

```
##                               GVIF Df GVIF^(1/(2*Df))
## factor(RSFracJobType)     1.725073  3        1.095136
## SpudToSales_DAYS          1.285900  1        1.133975
## PerfInterval_FT           1.315207  1        1.146825
## FracStages                1.223422  1        1.106084
## ProppantLoading_LBSPerGAL 1.661806  1        1.289110
## factor(RSProppantType)    1.625719  2        1.129175
## FluidIntensity_BBLPerFT   1.972725  1        1.404537
## FrictionReducer_LBS       1.726552  1        1.313983
## Surfactant_LBS            1.625055  1        1.274776
```

The predictor variables now have VIF values under 2, so we proceed with a stepwise procedure to determine which variables we should keep in the model.

2. STEPWISE LINEAR REGRESSION

A P-value of entry was selected as 0.1 and a p-value of removal was selected as a fairly high 0.5:

```
initstep = ols_step_both_p(init_3, pent=0.1, prem=0.5, details = FALSE)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. factor(RSFracJobType)
## 2. SpudToSales_DAYS
## 3. PerfInterval_FT
## 4. FracStages
## 5. ProppantLoading_LBSPerGAL
## 6. factor(RSProppantType)
## 7. FluidIntensity_BBLPerFT
## 8. FrictionReducer_LBS
## 9. Surfactant_LBS
##
## We are selecting variables based on p value...
##
## Variables Entered/Removed:
##
## - PerfInterval_FT added
## - factor(RSProppantType) added
```

```
## Note: model has aliased coefficients
##       sums of squares computed by model comparison
```

```
## - FrictionReducer_LBS added
## - FluidIntensity_BBLPerFT added
## - Surfactant_LBS added
## - FracStages added
## - factor(RSFracJobType) added
##
## No more variables to be added/removed.
##
##
## Final Model Output
## ------------------
##
##                              Model Summary
## ----------------------------------------------------------------------
## R                       0.649       RMSE                   167962.442
## R-Squared               0.421       Coef. Var                  47.108
## Adj. R-Squared          0.382       MSE                28211382079.766
## Pred R-Squared           -Inf       MAE                    124060.512
## ----------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## -----------------------------------------------------------------------------
##                    Sum of
##                   Squares         DF        Mean Square      F         Sig.
## -----------------------------------------------------------------------------
## Regression    2.998798e+12         10    299879835006.237     10.63    0.0000
## Residual      4.118862e+12        146     28211382079.766
## Total          7.11766e+12        156
## -----------------------------------------------------------------------------
##
##                                                           Parameter Estimates
## ------------------------------------------------------------------------------------------------------------------------
--------
##                                  model           Beta    Std. Error    Std. Beta      t        Sig           lower
upper
## ------------------------------------------------------------------------------------------------------------------------
--------
##                            (Intercept)     -81196.040    261402.824                 -0.311    0.757    -597818.375    435
426.295
##                        PerfInterval_FT         60.768        13.284        0.319     4.574    0.000         34.514
87.022
##     factor(RSProppantType)SAND + RESIN     199052.749    188618.929        0.459     1.055    0.293    -173723.452    571
828.950
##        factor(RSProppantType)SAND ONLY      11023.827    188900.303        0.025     0.058    0.954    -362308.467    384
356.120
##                    FrictionReducer_LBS          3.784         2.141        0.146     1.768    0.079         -0.447
8.015
##                FluidIntensity_BBLPerFT     -10339.599      3947.841       -0.217    -2.619    0.010      -18141.898     -2
537.300
##                         Surfactant_LBS          6.764         2.117        0.242     3.196    0.002          2.581
10.948
##                             FracStages      -6209.923      2272.829       -0.176    -2.732    0.007      -10701.819     -1
718.027
##         factor(RSFracJobType)ENERGIZED     169057.153    187647.552        0.152     0.901    0.369    -201799.271    539
913.577
## factor(RSFracJobType)SLICKCROSSLINKHYB    -319551.893    219398.672       -0.168    -1.456    0.147    -753159.505    114
055.719
##        factor(RSFracJobType)SLICKWATER     119934.383    171472.364        0.131     0.699    0.485    -218954.279    458
823.044
## ------------------------------------------------------------------------------------------------------------------------
--------
```

```
summary(initstep$model)
```

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -429063  -98010  -30471   81748  545751
##
## Coefficients:
##                                        Estimate Std. Error t value
## (Intercept)                          -8.120e+04  2.614e+05  -0.311
## PerfInterval_FT                       6.077e+01  1.328e+01   4.574
## factor(RSProppantType)SAND + RESIN    1.991e+05  1.886e+05   1.055
## factor(RSProppantType)SAND ONLY       1.102e+04  1.889e+05   0.058
## FrictionReducer_LBS                   3.784e+00  2.141e+00   1.768
## FluidIntensity_BBLPerFT              -1.034e+04  3.948e+03  -2.619
## Surfactant_LBS                        6.764e+00  2.117e+00   3.196
## FracStages                           -6.210e+03  2.273e+03  -2.732
## factor(RSFracJobType)ENERGIZED        1.691e+05  1.876e+05   0.901
## factor(RSFracJobType)SLICKCROSSLINKHYB -3.196e+05  2.194e+05  -1.456
## factor(RSFracJobType)SLICKWATER       1.199e+05  1.715e+05   0.699
##                                      Pr(>|t|)
## (Intercept)                           0.75654
## PerfInterval_FT                      1.01e-05 ***
## factor(RSProppantType)SAND + RESIN    0.29302
## factor(RSProppantType)SAND ONLY       0.95354
## FrictionReducer_LBS                   0.07921 .
## FluidIntensity_BBLPerFT               0.00975 **
## Surfactant_LBS                        0.00171 **
## FracStages                            0.00707 **
## factor(RSFracJobType)ENERGIZED        0.36911
## factor(RSFracJobType)SLICKCROSSLINKHYB  0.14740
## factor(RSFracJobType)SLICKWATER       0.48539
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 168000 on 146 degrees of freedom
##   (86 observations deleted due to missingness)
## Multiple R-squared:  0.4213, Adjusted R-squared:  0.3817
## F-statistic: 10.63 on 10 and 146 DF,  p-value: 2.093e-13
```

After the stepwise procedure SpudToSales_DAYS and ProppandLoading_LBSPerGal were removed from the model. A test for multi-collinearity was run again:

```
vif(lm(First36MonthProd_BOE~PerfInterval_FT+factor(RSProppantType)+FrictionReducer_LBS+FluidIntensity_BBLPerFT+Surfactant_LBS+FracStages+factor(RSFracJobType), data=data1))
```

```
##                             GVIF Df GVIF^(1/(2*Df))
## PerfInterval_FT         1.225631  1        1.107082
## factor(RSProppantType)  1.391940  2        1.086188
## FrictionReducer_LBS     1.725177  1        1.313460
## FluidIntensity_BBLPerFT 1.734314  1        1.316934
## Surfactant_LBS          1.450909  1        1.204537
## FracStages              1.047091  1        1.023274
## factor(RSFracJobType)   1.636646  3        1.085573
```

All values look to be around 1 so, multi-collinearity is not an issue with these predictor variables.

3. T-tests to REFINE THE ADDITIVE MODEL

A t-test was performed again to determine how our model is performing:

Null hypothesis: $H_0 : \beta_i = 0$ for all i=1,2,3,4,5,6,7

Alternate hypothesis: $H_A : atleast\ one\ \beta_i \neq 0$

```
init_4=lm(First36MonthProd_BOE~PerfInterval_FT+factor(RSProppantType)+FrictionReducer_LBS+FluidIntensity_BBLPerFT+Surfactant
_LBS+FracStages+factor(RSFracJobType), data=data1)
summary(init_4)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + factor(RSProppantType) +
##     FrictionReducer_LBS + FluidIntensity_BBLPerFT + Surfactant_LBS +
##     FracStages + factor(RSFracJobType), data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -429063  -98010  -30471   81748  545751
##
## Coefficients:
##                                       Estimate Std. Error t value
## (Intercept)                          -8.120e+04  2.614e+05  -0.311
## PerfInterval_FT                       6.077e+01  1.328e+01   4.574
## factor(RSProppantType)SAND + RESIN    1.991e+05  1.886e+05   1.055
## factor(RSProppantType)SAND ONLY       1.102e+04  1.889e+05   0.058
## FrictionReducer_LBS                   3.784e+00  2.141e+00   1.768
## FluidIntensity_BBLPerFT              -1.034e+04  3.948e+03  -2.619
## Surfactant_LBS                        6.764e+00  2.117e+00   3.196
## FracStages                           -6.210e+03  2.273e+03  -2.732
## factor(RSFracJobType)ENERGIZED        1.691e+05  1.876e+05   0.901
## factor(RSFracJobType)SLICKCROSSLINKHYB -3.196e+05  2.194e+05  -1.456
## factor(RSFracJobType)SLICKWATER       1.199e+05  1.715e+05   0.699
##                                      Pr(>|t|)
## (Intercept)                           0.75654
## PerfInterval_FT                      1.01e-05 ***
## factor(RSProppantType)SAND + RESIN    0.29302
## factor(RSProppantType)SAND ONLY       0.95354
## FrictionReducer_LBS                   0.07921 .
## FluidIntensity_BBLPerFT               0.00975 **
## Surfactant_LBS                        0.00171 **
## FracStages                            0.00707 **
## factor(RSFracJobType)ENERGIZED        0.36911
## factor(RSFracJobType)SLICKCROSSLINKHYB  0.14740
## factor(RSFracJobType)SLICKWATER       0.48539
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 168000 on 146 degrees of freedom
##   (86 observations deleted due to missingness)
## Multiple R-squared:  0.4213, Adjusted R-squared:  0.3817
## F-statistic: 10.63 on 10 and 146 DF,  p-value: 2.093e-13
```

p-value is <0.05 of t-test for full model suggest there are not enough evidnece in support of null hypothesis. We reject null hypothesis and accept alternate hypothesis, $H_A : atleast\ one\ \beta_i \neq 0$

This suggest there is linear relationship between response variable First36MonthProd_BOE and predictors.

p-value of co-efficient of FrictionReducer_LBS is 0.07921 which is in between 0.10 and 0.05. This p-value is just little greater than 0.05 but below 0.10 so we will consider this predictor for now. Also p-value of co-efficients of RSProppantType and RSFracJobType is >0.05. This suggest these variables are not significant for this modeling. We will drop these variables and check the model again.

The FracStages variable was not included in this model so, which is very important to understand. Otherwise,all terms are significant but the model is very poor compared to the additive model. We will discard this interactive model.

RSProppantType is removed from the init_4 model because its p-values ranged from 0.29 to 0.95. A t-test performed for the p-values of predictor variables of the new model:

```
init_5=lm(First36MonthProd_BOE~PerfInterval_FT+FrictionReducer_LBS+FluidIntensity_BBLPerFT+Surfactant_LBS+FracStages+factor
(RSFracJobType), data=data1)
summary(init_5)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + FrictionReducer_LBS +
##     FluidIntensity_BBLPerFT + Surfactant_LBS + FracStages + factor(RSFracJobType),
##     data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -353312 -137799  -39143   93557  593438
##
## Coefficients:
##                                        Estimate Std. Error t value
## (Intercept)                          -1.311e+05  2.056e+05  -0.638
## PerfInterval_FT                       6.350e+01  1.498e+01   4.238
## FrictionReducer_LBS                   3.312e+00  2.215e+00   1.496
## FluidIntensity_BBLPerFT              -3.547e+03  4.232e+03  -0.838
## Surfactant_LBS                        6.980e+00  2.387e+00   2.924
## FracStages                           -6.244e+03  2.561e+03  -2.438
## factor(RSFracJobType)ENERGIZED        2.394e+05  2.114e+05   1.132
## factor(RSFracJobType)SLICKCROSSLINKHYB -2.401e+05  2.473e+05  -0.971
## factor(RSFracJobType)SLICKWATER       2.048e+05  1.930e+05   1.061
##                                      Pr(>|t|)
## (Intercept)                           0.52473
## PerfInterval_FT                      3.96e-05 ***
## FrictionReducer_LBS                   0.13689
## FluidIntensity_BBLPerFT               0.40336
## Surfactant_LBS                        0.00399 **
## FracStages                            0.01595 *
## factor(RSFracJobType)ENERGIZED        0.25937
## factor(RSFracJobType)SLICKCROSSLINKHYB  0.33313
## factor(RSFracJobType)SLICKWATER       0.29026
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 189600 on 148 degrees of freedom
##   (86 observations deleted due to missingness)
## Multiple R-squared:  0.2526, Adjusted R-squared:  0.2122
## F-statistic: 6.252 on 8 and 148 DF,  p-value: 5.845e-07
```

The RSFracJobType variable is removed from the model because its p-values are 0.26 to 0.333. The t-test is performed again to determine which variables should remain in the model.

```
init_6=lm(First36MonthProd_BOE~PerfInterval_FT+FrictionReducer_LBS+FluidIntensity_BBLPerFT+Surfactant_LBS+FracStages, data=d
ata1)
summary(init_6)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + FrictionReducer_LBS +
##     FluidIntensity_BBLPerFT + Surfactant_LBS + FracStages, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -468662 -126996  -49158   97857  616268
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             71955.237 105583.915   0.681 0.496600
## PerfInterval_FT            60.720     15.203   3.994 0.000101 ***
## FrictionReducer_LBS         4.207      2.223   1.893 0.060277 .
## FluidIntensity_BBLPerFT  -886.415   3897.870  -0.227 0.820412
## Surfactant_LBS              4.203      2.169   1.938 0.054525 .
## FracStages              -5896.964   2595.541  -2.272 0.024501 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 194000 on 151 degrees of freedom
##   (86 observations deleted due to missingness)
## Multiple R-squared:  0.2011, Adjusted R-squared:  0.1747
## F-statistic: 7.604 on 5 and 151 DF,  p-value: 2.132e-06
```

FluidIntensity_BBLPerFT has a p-value of 0.8204 so we remove it from the model and run the t-test again:

```
init_7=lm(First36MonthProd_BOE~PerfInterval_FT+FrictionReducer_LBS+Surfactant_LBS+FracStages, data=data1)
summary(init_7)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + FracStages, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -449155 -126601  -50180  100284  618903
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         63902.068  97891.251   0.653   0.5149
## PerfInterval_FT        61.558     15.002   4.103  6.6e-05 ***
## FrictionReducer_LBS     4.158      2.077   2.002   0.0471 *
## Surfactant_LBS          3.620      1.992   1.817   0.0711 .
## FracStages          -5887.381   2581.477  -2.281   0.0240 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 193200 on 153 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.1996, Adjusted R-squared:  0.1787
## F-statistic: 9.539 on 4 and 153 DF,  p-value: 6.513e-07
```

The model init_7 is much better but, the p-value of Surfactant_LBS is 0.0771>0.05 - level of significance, but it is just little higher. So we accept this model.

Simple linear model at this stage is:

$$\begin{aligned}\widehat{First36MonthProd\ BOE} =& 63902.068 + 61.558 \cdot PerfInterval\ FT + 4.158 \cdot FrictionReducer\ LBS \\ &+ 3.620 \cdot Surfactant\ LBS - 5887.381 \cdot FracStages\end{aligned}$$

4. INTERACTIVE TERMS ARE INVESTIGATED

Here we constructed the interactive and proceed to eliminate all interactive terms that are non-significant. We also see that removing some interactive terms will cause an entire variable to become non-significant. We will elminate those as well.

```
initint = lm(First36MonthProd_BOE~(PerfInterval_FT+FrictionReducer_LBS+Surfactant_LBS+FracStages)^2, data=data1)
summary(initint)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ (PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + FracStages)^2, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -554077 -127862  -43119  112171  629475
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        -4.683e+04  3.532e+05  -0.133    0.895
## PerfInterval_FT                     8.019e+01  6.423e+01   1.248    0.214
## FrictionReducer_LBS                 1.284e+01  1.534e+01   0.837    0.404
## Surfactant_LBS                      1.773e+00  1.080e+01   0.164    0.870
## FracStages                         -1.894e+03  1.488e+04  -0.127    0.899
## PerfInterval_FT:FrictionReducer_LBS -1.587e-03  1.932e-03  -0.821    0.413
## PerfInterval_FT:Surfactant_LBS      1.331e-03  1.817e-03   0.732    0.465
## PerfInterval_FT:FracStages         -1.010e+00  2.624e+00  -0.385    0.701
## FrictionReducer_LBS:Surfactant_LBS  -4.144e-04  2.660e-04  -1.558    0.121
## FrictionReducer_LBS:FracStages      3.245e-01  3.302e-01   0.983    0.327
## Surfactant_LBS:FracStages          -1.175e-01  2.920e-01  -0.402    0.688
##
## Residual standard error: 194800 on 147 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2178, Adjusted R-squared:  0.1646
## F-statistic: 4.093 on 10 and 147 DF,  p-value: 5.513e-05
```

```
initint2 = lm(First36MonthProd_BOE~PerfInterval_FT+FrictionReducer_LBS+Surfactant_LBS+FracStages+PerfInterval_FT*FrictionRed
ucer_LBS, data=data1)
summary(initint2)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + FracStages + PerfInterval_FT * FrictionReducer_LBS,
##     data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -443862 -126842  -42779  108647  621670
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         3.586e+04  1.212e+05   0.296 0.767820
## PerfInterval_FT                     6.633e+01  1.931e+01   3.435 0.000764
## FrictionReducer_LBS                 9.071e+00  1.264e+01   0.717 0.474171
## Surfactant_LBS                      3.368e+00  2.097e+00   1.606 0.110306
## FracStages                         -5.960e+03  2.595e+03  -2.297 0.023009
## PerfInterval_FT:FrictionReducer_LBS -7.087e-04  1.799e-03  -0.394 0.694111
##
## (Intercept)
## PerfInterval_FT                     ***
## FrictionReducer_LBS
## Surfactant_LBS
## FracStages                          *
## PerfInterval_FT:FrictionReducer_LBS
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 193700 on 152 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2004, Adjusted R-squared:  0.1741
## F-statistic: 7.621 on 5 and 152 DF,  p-value: 2.049e-06
```

```
initint3 = lm(First36MonthProd_BOE~PerfInterval_FT+FrictionReducer_LBS+Surfactant_LBS+FracStages+PerfInterval_FT*Surfactant_
LBS, data=data1)
summary(initint3)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + FracStages + PerfInterval_FT * Surfactant_LBS,
##     data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -447876 -126527  -50290  100208  619293
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    6.209e+04  1.419e+05   0.438  0.66228
## PerfInterval_FT                6.186e+01  2.293e+01   2.699  0.00775 **
## FrictionReducer_LBS            4.157e+00  2.084e+00   1.995  0.04784 *
## Surfactant_LBS                 3.779e+00  9.212e+00   0.410  0.68224
## FracStages                    -5.884e+03  2.597e+03  -2.265  0.02490 *
## PerfInterval_FT:Surfactant_LBS -2.822e-05  1.595e-03  -0.018  0.98591
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 193800 on 152 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.1996, Adjusted R-squared:  0.1733
## F-statistic: 7.582 on 5 and 152 DF,  p-value: 2.201e-06
```

```
initint4 = lm(First36MonthProd_BOE~PerfInterval_FT+FrictionReducer_LBS+Surfactant_LBS+FracStages+FrictionReducer_LBS*Surfact
ant_LBS, data=data1)
summary(initint4)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + FracStages + FrictionReducer_LBS * Surfactant_LBS,
##     data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -507621 -128751  -48366   96540  617772
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         2.855e+04  1.017e+05   0.281   0.7792
## PerfInterval_FT                     6.062e+01  1.499e+01   4.044 8.33e-05
## FrictionReducer_LBS                 7.421e+00  3.318e+00   2.237   0.0267
## Surfactant_LBS                      6.089e+00  2.792e+00   2.181   0.0307
## FracStages                         -5.210e+03  2.632e+03  -1.980   0.0496
## FrictionReducer_LBS:Surfactant_LBS -2.795e-04  2.218e-04  -1.260   0.2096
##
## (Intercept)
## PerfInterval_FT                    ***
## FrictionReducer_LBS                *
## Surfactant_LBS                     *
## FracStages                         *
## FrictionReducer_LBS:Surfactant_LBS
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 192800 on 152 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2079, Adjusted R-squared:  0.1818
## F-statistic: 7.978 on 5 and 152 DF,  p-value: 1.058e-06
```

```
initint5 = lm(First36MonthProd_BOE~PerfInterval_FT+FrictionReducer_LBS+Surfactant_LBS+FracStages+PerfInterval_FT*FracStages,
data=data1)
summary(initint5)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + FracStages + PerfInterval_FT * FracStages,
##     data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -448977 -126820  -49863  100926  620295
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 3.824e+04  3.203e+05   0.119   0.9051
## PerfInterval_FT             6.594e+01  5.423e+01   1.216   0.2259
## FrictionReducer_LBS         4.145e+00  2.089e+00   1.985   0.0490 *
## Surfactant_LBS              3.630e+00  2.002e+00   1.813   0.0718 .
## FracStages                 -4.776e+03  1.345e+04  -0.355   0.7230
## PerfInterval_FT:FracStages -1.886e-01  2.241e+00  -0.084   0.9330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 193800 on 152 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.1997, Adjusted R-squared:  0.1733
## F-statistic: 7.583 on 5 and 152 DF,  p-value: 2.194e-06
```

```
initint6 = lm(First36MonthProd_BOE~PerfInterval_FT+FrictionReducer_LBS+Surfactant_LBS+FracStages+PerfInterval_FT*FracStages,
data=data1)
summary(initint6)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + FracStages + PerfInterval_FT * FracStages,
##     data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -448977 -126820  -49863  100926  620295
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                3.824e+04  3.203e+05   0.119   0.9051
## PerfInterval_FT            6.594e+01  5.423e+01   1.216   0.2259
## FrictionReducer_LBS        4.145e+00  2.089e+00   1.985   0.0490 *
## Surfactant_LBS             3.630e+00  2.002e+00   1.813   0.0718 .
## FracStages                -4.776e+03  1.345e+04  -0.355   0.7230
## PerfInterval_FT:FracStages -1.886e-01  2.241e+00  -0.084   0.9330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 193800 on 152 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.1997, Adjusted R-squared:  0.1733
## F-statistic: 7.583 on 5 and 152 DF,  p-value: 2.194e-06
```

```
initint7 = lm(First36MonthProd_BOE~PerfInterval_FT+FrictionReducer_LBS+Surfactant_LBS+FracStages+FrictionReducer_LBS*FracSta
ges, data=data1)
summary(initint7)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + FracStages + FrictionReducer_LBS * FracStages,
##     data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -441540 -126767  -45211   98368  621793
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    9.335e+04  1.255e+05   0.744   0.4583
## PerfInterval_FT                6.160e+01  1.504e+01   4.095 6.85e-05 ***
## FrictionReducer_LBS            1.899e+00  6.352e+00   0.299   0.7653
## Surfactant_LBS                 3.507e+00  2.019e+00   1.737   0.0845 .
## FracStages                    -7.183e+03  4.306e+03  -1.668   0.0974 .
## FrictionReducer_LBS:FracStages  1.017e-01  2.701e-01   0.376   0.7072
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 193700 on 152 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2004, Adjusted R-squared:  0.1741
## F-statistic: 7.617 on 5 and 152 DF,  p-value: 2.062e-06
```

None of interactive term in above different interactive models is significant. Predictors are not interacting with each other.

MODEL INTERPRETATION

Our simple linear regression model is:

$$\widehat{First36MonthProd\ BOE} = 63902.068 + 61.558 \cdot PerfInterval\ FT + 4.158 \cdot FrictionReducer\ LBS + 3.620 \cdot Surfactant\ LBS - 5887.381 \cdot FracStages$$

1. Considering FracStage and holding the other variables constant for each additional fracture stimulation the First36Month_BOE could be expected to decrease by 5887.381 BOE. Considering a fracture stimulation costs up to $100,000 a decrease in cumulative 36 month production of 5887 BOE is an unexpected result.

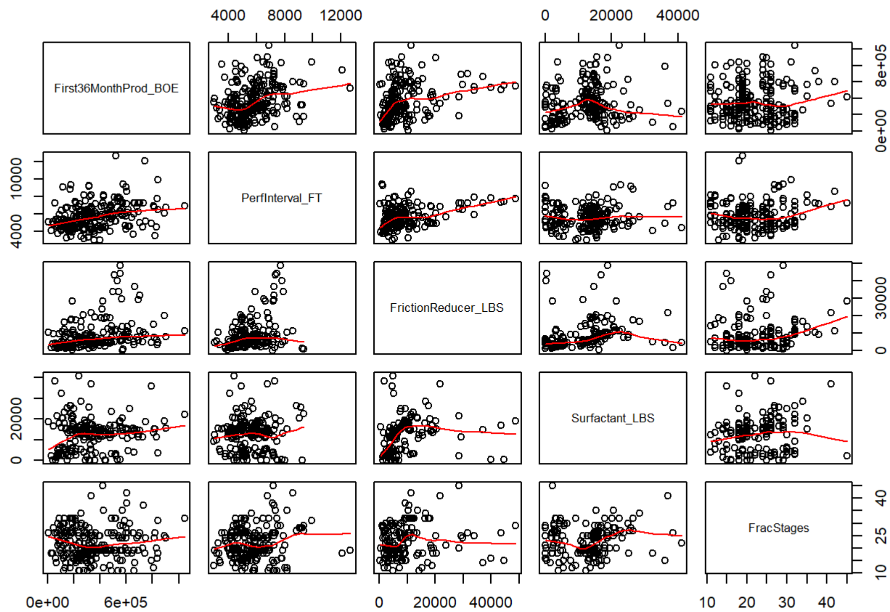
2. Considering the PerfInterval_FT and holding the other variables constant for each additional foot of horizontal distance the First36Month_BOE could increase by 61.558 BOE. An additional 1000 feet of horizontal distance may equate to 61558 BOE.

3. Considering FrictionReducer_LBS and holding the other variables constant for each additional pound of friction reducer the First36Month_BOE may increase by 4.158 BOE.

4. Considering Surfactant and holding the other variables constant for each additional pound of surfactant the First36Month_BOE may improve by 3.620 BOE or effectively zero.

5. INVESTIGATE QUADRATIC TERMS

```
pairs(First36MonthProd_BOE~PerfInterval_FT+FrictionReducer_LBS+Surfactant_LBS+FracStages , data=data1, panel=panel.smooth)
```



From an examination of pairs plots it is not clear that a quadratic terms should be added to the model.

6. INVESTIGATE LOG and BOXCOX TRANSFORMATIONS

```
log_7=lm(log(First36MonthProd_BOE)~PerfInterval_FT+FrictionReducer_LBS+Surfactant_LBS+FracStages, data=data1)
summary(log_7)
```

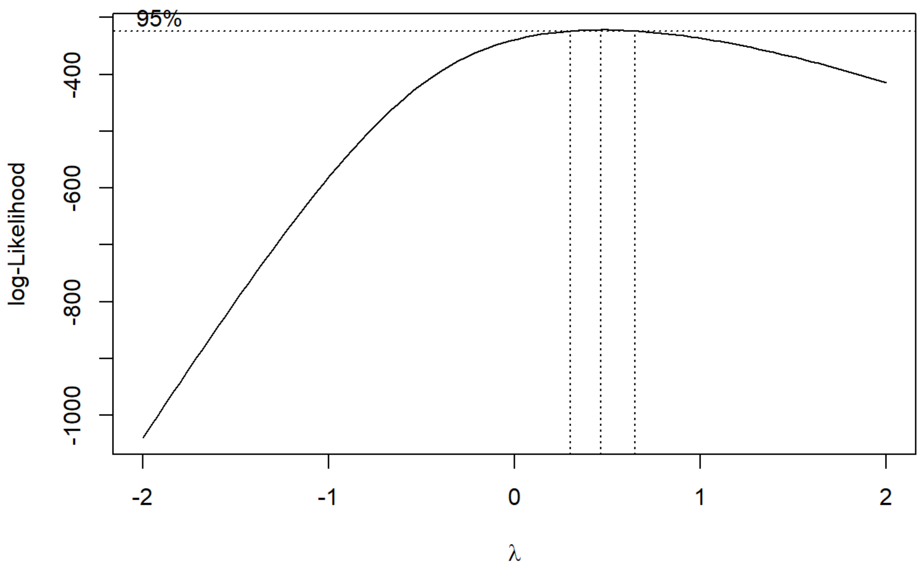
```
##
## Call:
## lm(formula = log(First36MonthProd_BOE) ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + FracStages, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5573 -0.3086 -0.0250  0.4493  1.2965
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1.160e+01  3.502e-01  33.125  < 2e-16 ***
## PerfInterval_FT      2.142e-04  5.367e-05   3.992 0.000101 ***
## FrictionReducer_LBS  1.470e-05  7.429e-06   1.979 0.049584 *
## Surfactant_LBS       1.208e-05  7.125e-06   1.695 0.092062 .
## FracStages          -2.291e-02  9.235e-03  -2.481 0.014191 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.691 on 153 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.1944, Adjusted R-squared:  0.1734
## F-statistic: 9.233 on 4 and 153 DF,  p-value: 1.04e-06
```

A log transformation did not improve the R^2adj and the p-value of Surfactant_LBS increases to 0.092062 The log transformation model is not considered further. Compared to the untransformed model the RMSE does decrease from 192700 to 0.691. The R^2adj does not improve.

A boxcox examination demonstrates that the square root of First36MonthProd_BOE may improve the model.

```
boxcox(init_7, objective.name = "Shapiro-Wilk")
```

```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##  extra argument 'objective.name' will be disregarded
```




```
init_sqrt=lm((First36MonthProd_BOE)^0.5~PerfInterval_FT+FrictionReducer_LBS+Surfactant_LBS+FracStages, data=data1)
summary(init_sqrt)
```

```
##
## Call:
## lm(formula = (First36MonthProd_BOE)^0.5 ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + FracStages, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -461.63 -102.87  -21.53  112.10  401.63
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         317.050651  83.785009   3.784 0.000221 ***
## PerfInterval_FT       0.054611   0.012840   4.253 3.65e-05 ***
## FrictionReducer_LBS   0.003799   0.001778   2.137 0.034176 *
## Surfactant_LBS        0.003128   0.001705   1.835 0.068419 .
## FracStages           -5.599610   2.209483  -2.534 0.012271 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 165.3 on 153 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2151, Adjusted R-squared:  0.1945
## F-statistic: 10.48 on 4 and 153 DF,  p-value: 1.572e-07
```

Compared to the untransformed model the R^2adj improves to 0.1945 and the RMSE improves from 192700 to 165.3. Also, the p-values of co-efficients of the predictor variables are significant. We will accept this as best fit model.

Best fit model so far is:

$$\sqrt{First36\widehat{Month}Prod\ BOE} = 317.050651 + 0.054611 \cdot PerfIntervalFT + 0.003799 \cdot FrictionReducerLBS + 0.003128 \cdot SurfactantLBS - 5.599610 \cdot FracStages$$

Lets check for quadratic, cubic and higher dimensions if it improve our model.

Null Hypothesis: H_0 : homoscedasticity is present

Alternative Hypothesis: H_a : heteroscedasticity is present

```
init_sqrt11=lm((First36MonthProd_BOE)^0.5~PerfInterval_FT+I(PerfInterval_FT^2)+FrictionReducer_LBS + Surfactant_LBS +FracSta
ges,data=data1)

summary(init_sqrt11)
```

```
##
## Call:
## lm(formula = (First36MonthProd_BOE)^0.5 ~ PerfInterval_FT + I(PerfInterval_FT^2) +
##     FrictionReducer_LBS + Surfactant_LBS + FracStages, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -462.40 -101.36  -21.71  112.58  401.90
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.865e+02  2.654e+02   1.080   0.2820
## PerfInterval_FT       6.564e-02  9.166e-02   0.716   0.4750
## I(PerfInterval_FT^2) -9.625e-07  7.922e-06  -0.121   0.9035
## FrictionReducer_LBS   3.815e-03  1.789e-03   2.133   0.0345 *
## Surfactant_LBS        3.096e-03  1.730e-03   1.790   0.0755 .
## FracStages           -5.578e+00  2.224e+00  -2.509   0.0132 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 165.9 on 152 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2151, Adjusted R-squared:  0.1893
## F-statistic: 8.333 on 5 and 152 DF,  p-value: 5.518e-07
```

```
bptest(init_sqrt11)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  init_sqrt11
## BP = 7.2151, df = 5, p-value = 0.2051
```

Quadratic power of perfinterval not improving model. We can not accept this model.

```
init_sqrt12=lm((First36MonthProd_BOE)^0.5~PerfInterval_FT+FrictionReducer_LBS+I(FrictionReducer_LBS^2) + Surfactant_LBS +Fra
cStages,data=data1)

summary(init_sqrt12)
```

```
##
## Call:
## lm(formula = (First36MonthProd_BOE)^0.5 ~ PerfInterval_FT + FrictionReducer_LBS +
##     I(FrictionReducer_LBS^2) + Surfactant_LBS + FracStages, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -475.63 -101.02  -12.86  110.42  400.42
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               2.880e+02  8.372e+01   3.440 0.000751 ***
## PerfInterval_FT           5.529e-02  1.268e-02   4.361 2.37e-05 ***
## FrictionReducer_LBS       1.429e-02  5.002e-03   2.857 0.004881 **
## I(FrictionReducer_LBS^2) -2.614e-07  1.167e-07  -2.240 0.026571 *
## Surfactant_LBS            1.832e-03  1.779e-03   1.030 0.304758
## FracStages               -6.222e+00  2.199e+00  -2.830 0.005288 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 163.2 on 152 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2401, Adjusted R-squared:  0.2151
## F-statistic: 9.607 on 5 and 152 DF,  p-value: 5.509e-08
```

In this model p-value of regression coefficient of surfactant is higher. making it insignificant. so we will not accept this model.

```
init_sqrt13=lm((First36MonthProd_BOE)^0.5~PerfInterval_FT+FrictionReducer_LBS + Surfactant_LBS+I(Surfactant_LBS^2) +FracStag
es,data=data1)

summary(init_sqrt13)
```

```
##
## Call:
## lm(formula = (First36MonthProd_BOE)^0.5 ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + I(Surfactant_LBS^2) + FracStages, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -487.17 -102.79  -19.96   95.71  409.33
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          2.089e+02  8.798e+01   2.374  0.01885 *
## PerfInterval_FT      5.864e-02  1.253e-02   4.681 6.26e-06 ***
## FrictionReducer_LBS  3.039e-03  1.741e-03   1.745  0.08303 .
## Surfactant_LBS       1.533e-02  4.132e-03   3.710  0.00029 ***
## I(Surfactant_LBS^2) -3.994e-07  1.240e-07  -3.223  0.00155 **
## FracStages          -4.353e+00  2.179e+00  -1.997  0.04758 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 160.5 on 152 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2653, Adjusted R-squared:  0.2411
## F-statistic: 10.98 on 5 and 152 DF,  p-value: 4.93e-09
```

```
init_sqrt14=lm((First36MonthProd_BOE)^0.5~PerfInterval_FT+FrictionReducer_LBS + Surfactant_LBS+I(Surfactant_LBS^2)+I(Surfact
ant_LBS^3) +FracStages,data=data1)

summary(init_sqrt14)
```

```
##
## Call:
## lm(formula = (First36MonthProd_BOE)^0.5 ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + I(Surfactant_LBS^2) + I(Surfactant_LBS^3) +
##     FracStages, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -469.19 -103.07  -22.67   89.90  419.01
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1.475e+02  9.351e+01   1.578  0.11675
## PerfInterval_FT      5.873e-02  1.243e-02   4.725 5.23e-06 ***
## FrictionReducer_LBS  3.733e-03  1.769e-03   2.110  0.03652 *
## Surfactant_LBS       3.320e-02  1.058e-02   3.138  0.00205 **
## I(Surfactant_LBS^2) -1.774e-06  7.603e-07  -2.333  0.02095 *
## I(Surfactant_LBS^3)  2.521e-11  1.376e-11   1.832  0.06891 .
## FracStages          -3.677e+00  2.194e+00  -1.676  0.09575 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 159.2 on 151 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2812, Adjusted R-squared:  0.2527
## F-statistic: 9.847 on 6 and 151 DF,  p-value: 3.724e-09
```

```
init_sqrt15=lm((First36MonthProd_BOE)^0.5~PerfInterval_FT+FrictionReducer_LBS + Surfactant_LBS+I(Surfactant_LBS^2)+I(Surfact
ant_LBS^3)+I(Surfactant_LBS^4) +FracStages,data=data1)

summary(init_sqrt15)
```

```
##
## Call:
## lm(formula = (First36MonthProd_BOE)^0.5 ~ PerfInterval_FT + FrictionReducer_LBS +
##     Surfactant_LBS + I(Surfactant_LBS^2) + I(Surfactant_LBS^3) +
##     I(Surfactant_LBS^4) + FracStages, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -467.58 -101.64  -21.85   90.91  412.22
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1.434e+02  9.675e+01   1.482   0.1405
## PerfInterval_FT      5.896e-02  1.254e-02   4.702 5.78e-06 ***
## FrictionReducer_LBS  3.762e-03  1.783e-03   2.110   0.0365 *
## Surfactant_LBS       3.612e-02  1.974e-02   1.830   0.0693 .
## I(Surfactant_LBS^2) -2.155e-06  2.297e-06  -0.938   0.3496
## I(Surfactant_LBS^3)  4.185e-11  9.561e-11   0.438   0.6622
## I(Surfactant_LBS^4) -2.245e-16  1.277e-15  -0.176   0.8606
## FracStages          -3.745e+00  2.235e+00  -1.676   0.0958 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 159.8 on 150 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2814, Adjusted R-squared:  0.2479
## F-statistic: 8.391 on 7 and 150 DF,  p-value: 1.225e-08
```

Above models with higher surfactant powers we determined cubic power of surfactant produce a better model. with p values of all regression coefficent wither below 0.05 or just little higher than 0.05.

At this point we accept this model to be best fit for further exploration.

Best model so far:

```
coefficients(init_sqrt14)
```

```
##        (Intercept)     PerfInterval_FT FrictionReducer_LBS 
##       1.475207e+02        5.872815e-02        3.732632e-03 
##      Surfactant_LBS I(Surfactant_LBS^2) I(Surfactant_LBS^3) 
##       3.319713e-02       -1.774125e-06        2.520911e-11 
##          FracStages 
##      -3.677415e+00
```

$$\sqrt{First36\widehat{Month}Prod\,BOE} = 147.5207 + 0.05872815 \cdot PerfIntervalFT + \\ 0.003732632 \cdot FrictionReducerLBS + 0.003128 \cdot SurfactantLBS - 1.774125e^{-06} \cdot SurfactantLBS^2 \\ + 2.520911e^{-11} \cdot SurfactantLBS^3 - 3.677415 \cdot FracStages$$

```
init_sqrt16=lm((First36MonthProd_BOE)^0.5~PerfInterval_FT+FrictionReducer_LBS + Surfactant_LBS+I(Surfactant_LBS^2)+I(Surfact
ant_LBS^3) +FracStages+I(FracStages^2),data=data1)

summary(init_sqrt16)
```

```
## 
## Call:
## lm(formula = (First36MonthProd_BOE)^0.5 ~ PerfInterval_FT + FrictionReducer_LBS + 
##     Surfactant_LBS + I(Surfactant_LBS^2) + I(Surfactant_LBS^3) + 
##     FracStages + I(FracStages^2), data = data1)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -468.08 -103.43  -22.89   89.97  417.82 
## 
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)    
## (Intercept)          1.701e+02  1.778e+02   0.957  0.34018    
## PerfInterval_FT      5.863e-02  1.249e-02   4.694 5.99e-06 ***
## FrictionReducer_LBS  3.657e-03  1.845e-03   1.982  0.04927 *  
## Surfactant_LBS       3.316e-02  1.062e-02   3.124  0.00214 ** 
## I(Surfactant_LBS^2) -1.767e-06  7.643e-07  -2.312  0.02216 *  
## I(Surfactant_LBS^3)  2.502e-11  1.386e-11   1.805  0.07314 .  
## FracStages          -5.591e+00  1.296e+01  -0.431  0.66690    
## I(FracStages^2)      3.944e-02  2.633e-01   0.150  0.88114    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 159.8 on 150 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2813, Adjusted R-squared:  0.2478 
## F-statistic: 8.389 on 7 and 150 DF,  p-value: 1.23e-08
```

This model makes FracStages insignificant. we will not accept this model.

```
init_sqrt17=lm((First36MonthProd_BOE)^0.5~PerfInterval_FT+I(PerfInterval_FT^2)+FrictionReducer_LBS+I(FrictionReducer_LBS^2)
 +Surfactant_LBS+I(Surfactant_LBS^2) +FracStages+I(FracStages^2),data=data1)

summary(init_sqrt17)
```

```
##
## Call:
## lm(formula = (First36MonthProd_BOE)^0.5 ~ PerfInterval_FT + I(PerfInterval_FT^2) +
##     FrictionReducer_LBS + I(FrictionReducer_LBS^2) + Surfactant_LBS +
##     I(Surfactant_LBS^2) + FracStages + I(FracStages^2), data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -491.46 -106.14  -17.88   97.84  422.67
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               3.778e+02  3.027e+02   1.248  0.21394
## PerfInterval_FT          -1.306e-02  9.288e-02  -0.141  0.88836
## I(PerfInterval_FT^2)      6.274e-06  8.051e-06   0.779  0.43705
## FrictionReducer_LBS       1.131e-02  5.632e-03   2.008  0.04650 *
## I(FrictionReducer_LBS^2) -2.050e-07  1.288e-07  -1.592  0.11357
## Surfactant_LBS            1.308e-02  4.622e-03   2.830  0.00529 **
## I(Surfactant_LBS^2)      -3.524e-07  1.330e-07  -2.649  0.00895 **
## FracStages               -3.427e+00  1.336e+01  -0.256  0.79795
## I(FracStages^2)          -3.493e-02  2.748e-01  -0.127  0.89901
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 160.6 on 149 degrees of freedom
##   (85 observations deleted due to missingness)
## Multiple R-squared:  0.2787, Adjusted R-squared:  0.2399
## F-statistic: 7.196 on 8 and 149 DF,  p-value: 4.831e-08
```

After exploring different models with quadratic, cubic and other higher powers we come up with best fit model to be.

Model name: init_sqrt14

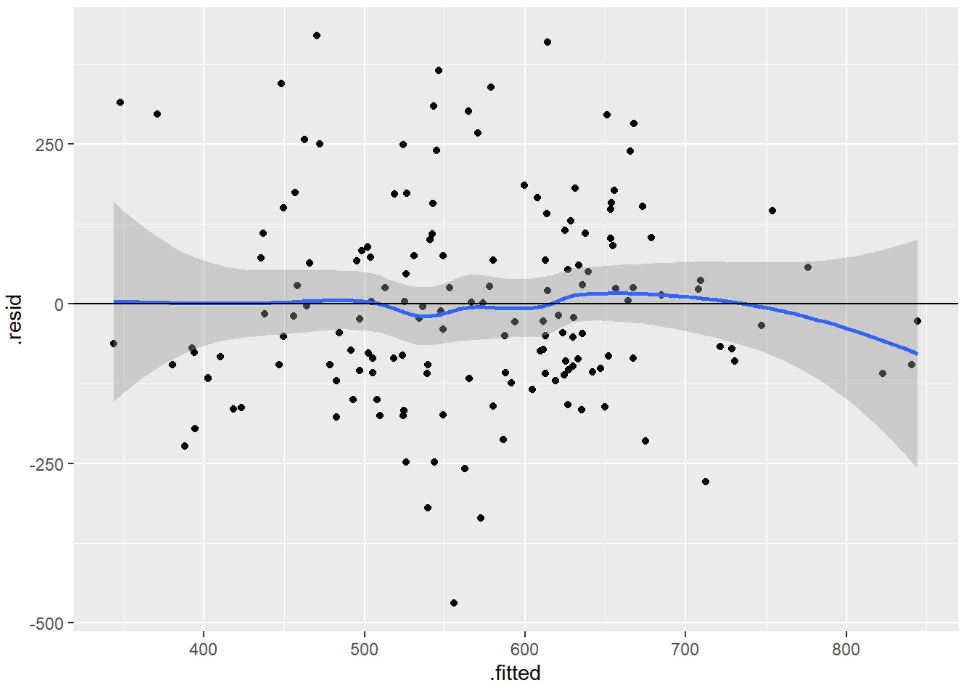
$$\begin{aligned}\sqrt{First36\widehat{Month}Prod\;BOE} =& 147.5207 + 0.05872815 \cdot PerfIntervalFT + \\ & 0.003732632 \cdot FrictionReducerLBS + 0.003128 \cdot SurfactantLBS - 1.774125e^{-06} \cdot SurfactantLBS^2 \\ & + 2.520911e^{-11} \cdot SurfactantLBS^3 - 3.677415 \cdot FracStages\end{aligned}$$

INVESTIGATION OF FUNDAMENTAL ASSUMPTIONS

LINEARITY ASSUMPTION

```
ggplot(init_sqrt14, aes(x=.fitted, y=.resid)) +
  geom_point() + geom_smooth()+
  geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



There is no apparent pattern of the residuals in the plot above so, we can conclude there is a straight-line relationship between the predictor variables and the response variable - particularly, with a square root transformation.

INDEPENDENCE ASSUMPTION

Again, examining the plot above we can conclude the error terms are uncorrelated. There are no obvious trends or patterns and, little significant “clumping” of residuals.

EQUAL VARIANCE ASSUMPTION

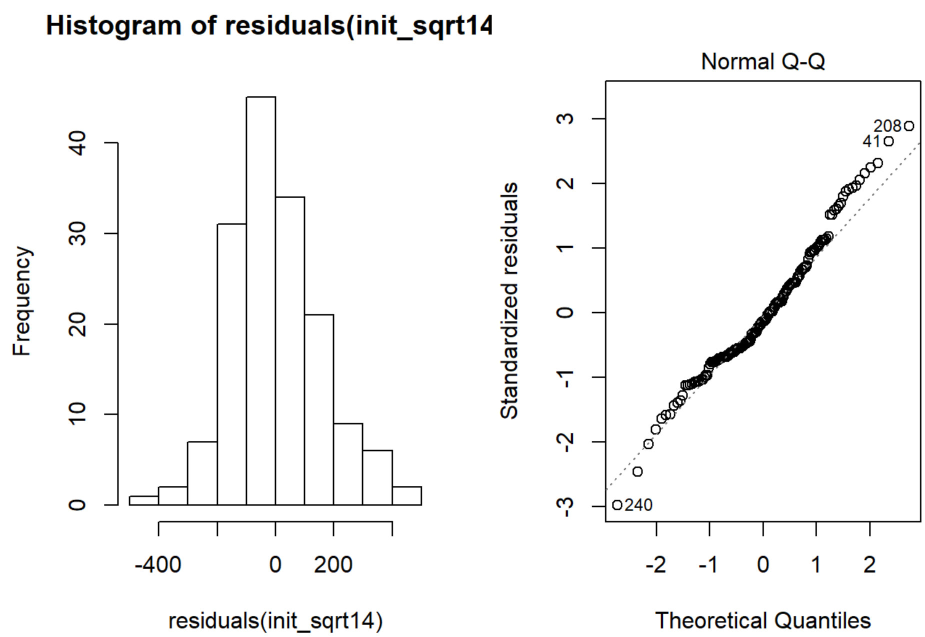
Again examining the plot above we can see there are no obvious trends or “funneling” of the data. Also, by performing a Breusch-Pagan test (below) we obtain a p-value of 0.05129. We fail to reject null hypothesis H_0 that the data is homoscedastic. We can conclude the error terms have a constant variance.

```
bptest(init_sqrt14)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  init_sqrt14
## BP = 12.522, df = 6, p-value = 0.05129
```

NORMALITY ASSUMPTION

```
par(mfrow=c(1,2))
hist(residuals(init_sqrt14))
plot(init_sqrt14, which=2)
```



Null Hypothesis:

H_0 : sample data is normally distributed

Alternative Hypothesis: H_a : sample data is not normally distributed

```
shapiro.test(residuals(init_sqrt14))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(init_sqrt14)
## W = 0.97764, p-value = 0.01147
```

This model failed normality assumption from Shapiro-Wilk test . So we will not be able to accept this model.

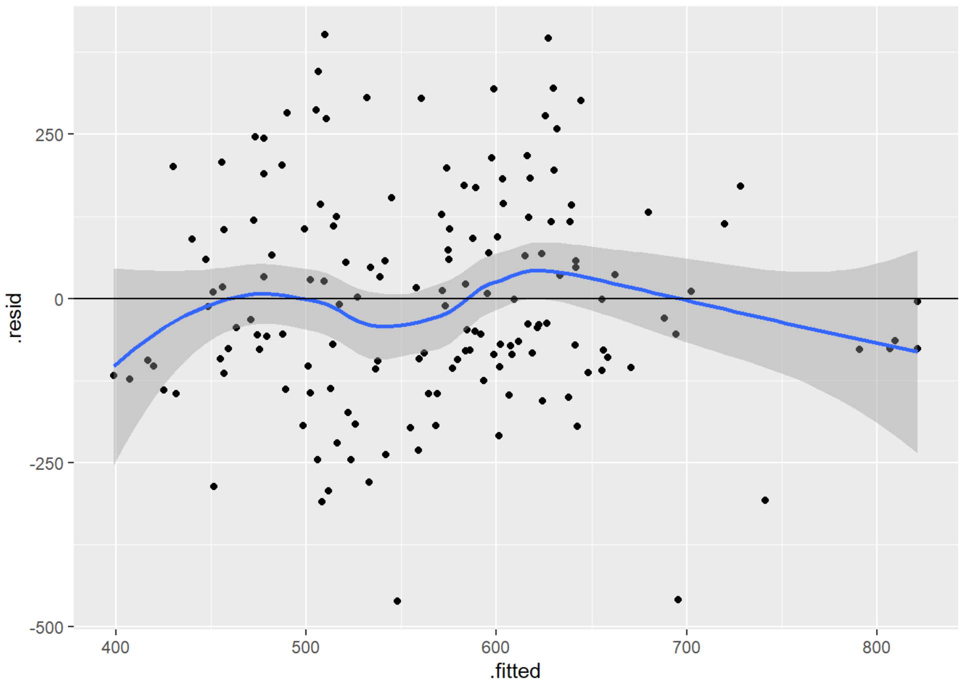
We will accept model without powers of predictors and check for assumptions. Model name: init_sqrt

$$\sqrt{First36\widehat{Month}Prod\ BOE} = 317.050651 + 0.054611 \cdot PerfIntervalFT + 0.003799 \cdot FrictionReducerLBS \\ + 0.003128 \cdot SurfactantLBS - 5.599610 \cdot FracStages$$

LINEARITY ASSUMPTION

```
ggplot(init_sqrt, aes(x=.fitted, y=.resid)) +
  geom_point() + geom_smooth()+
  geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



There is no apparent pattern of the residuals in the plot above so, we can conclude there is a straight-line relationship between the predictor variables and the response variable - particularly, with a square root transformation.

INDEPENDENCE ASSUMPTION

Again, examining the plot above we can conclude the error terms are uncorrelated. There are no obvious trends or patterns and, little significant “clumping” of residuals.

EQUAL VARIANCE ASSUMPTION

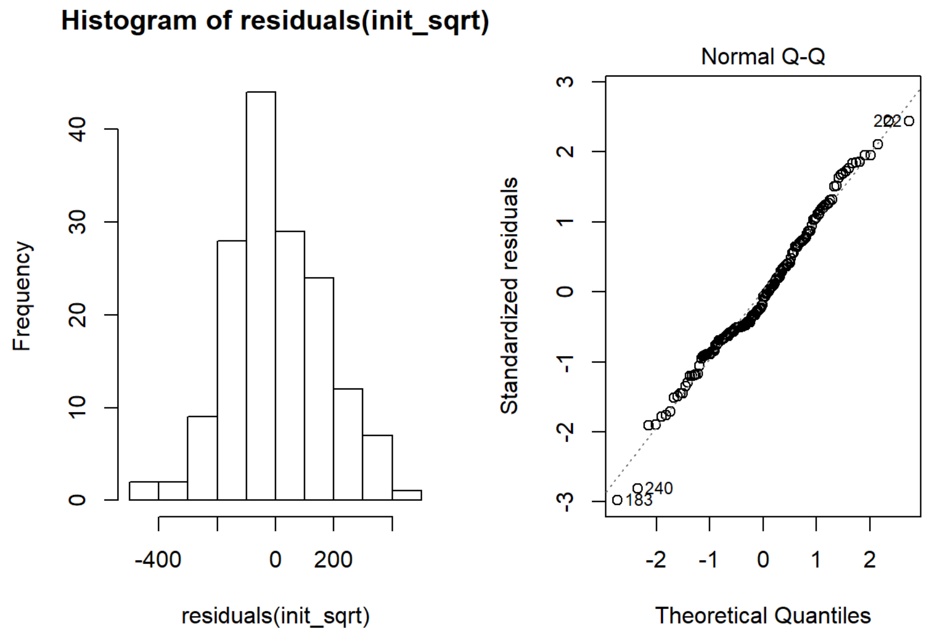
Again examining the plot above we can see there are no obvious trends or “funneling” of the data. Also, by performing a Breusch-Pagan test (below) we obtain a p-value of 0.1322. We fail to reject null hypothesis H_0 that the data is homoscedastic. We can conclude the error terms have a constant variance.

```
bptest(init_sqrt)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  init_sqrt
## BP = 7.0724, df = 4, p-value = 0.1321
```

NORMALITY ASSUMPTION

```
par(mfrow=c(1,2))
hist(residuals(init_sqrt))
plot(init_sqrt, which=2)
```




```
shapiro.test(residuals(init_sqrt))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(init_sqrt)
## W = 0.98642, p-value = 0.1268
```

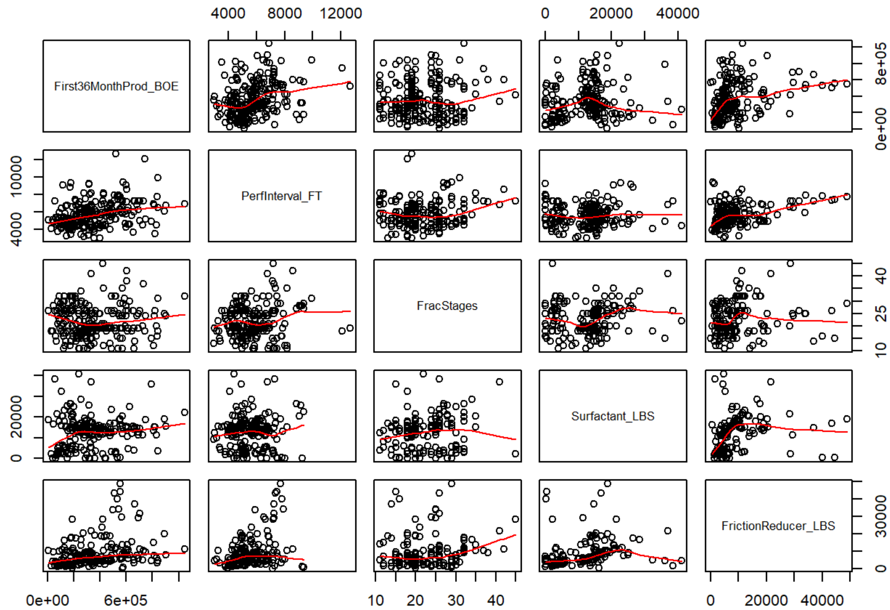
The histogram of the residuals is nearly normally distributed and, a Q-Qplot is nearly linear. Further, a Shapiro-Wilk test gives a p-value of 0.1543. So, we fail to reject null hypothesis H_0 : that the residuals of the regression are normally distributed. The data satisfies the normality assumption.

MULTI-COLLINEARITY

```
vif(init_sqrt)
```

```
##     PerfInterval_FT FrictionReducer_LBS      Surfactant_LBS
##            1.190174            1.229802            1.058995
##          FracStages
##            1.021441
```

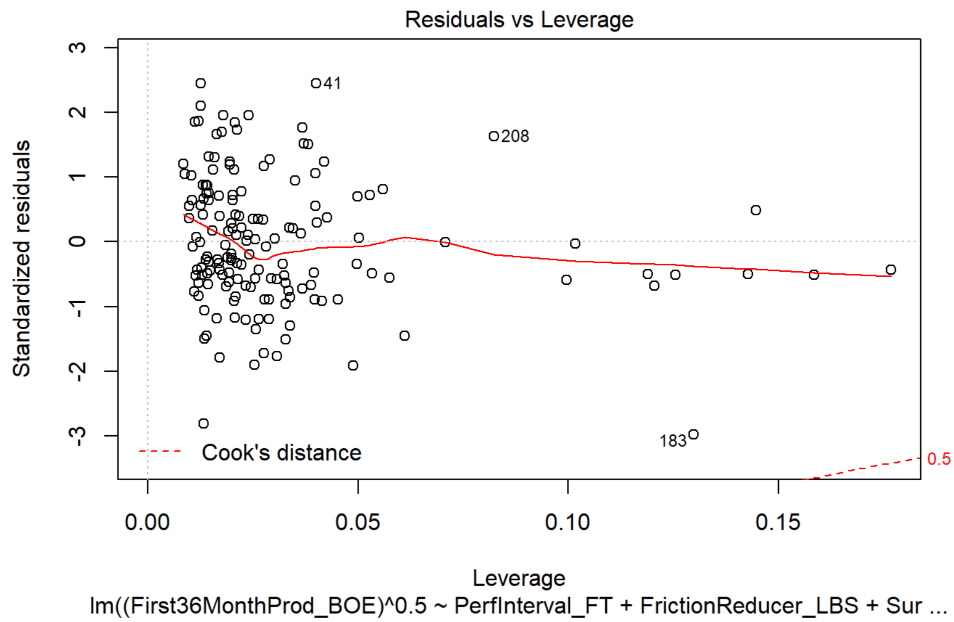
```
pairs(First36MonthProd_BOE~PerfInterval_FT+FracStages+Surfactant_LBS+FrictionReducer_LBS, data=data1, panel=panel.smooth)
```



A VIF test of the model and the pairs plot do not demonstrate multi-collinearity. The data used in the square root transformed model meets this assumption.

OUTLIERS and LEVERAGE

```
plot(init_sqrt,which=5)
```



A plot of Cook’s Distance does not indicate significant leverage of any data points. The data used meets the assumption that there is not any influential data. Removing any data points will not improve the model.

MODEL CONCLUSIONS

```
summary(init_7)$adj.r.squared
```

```
## [1] 0.1786897
```

```
summary(init_sqrt)$adj.r.squared
```

```
## [1] 0.1945466
```

```
sigma(init_7)
```

```
## [1] 193155
```

```
sigma(init_sqrt)
```

```
## [1] 165.3212
```

- The fundamental assumptions about the data such as linearity, independence, normality, homoscedasticity, not colinear and, with a lack of influential data points are satisfied. The R^2adj improves to 0.194 by making a square root transformation compared to 0.1787 of the additive model. The RMSE decreases from 193155 from the additive model to just 165.3 of the square root transformed model demonstrating a much better model fit. Overall, we can conclude our model may predict 19.4% of the variance of a liquids rich gas well after a fracture stimulation.

DISCUSSION OF RESULTS

$$\sqrt{First36\widehat{Month}Prod\ BOE} = 317.050651 + 0.054611 \cdot PerfIntervalFT + 0.003799 \cdot FrictionReducerLBS + 0.003128 \cdot SurfactantLBS - 5.599610 \cdot FracStages$$

- Our model had an R^2adj of 0.1945 so, a limited predictive ability. We did use a data set provided by Ross Smith Energy Group and, they do these types of predictions and models professionally. So, given the results we obtained we were surprised and disppointed. Considering the expense of drilling, fracturing, and oil field chemicals a much better predictive model is needed. The most important predictors were not captured which are the geology such as the quality of reservoir in terms of porosity and permeability. Geomechanics are also very important such as the Young’s modulus and Poisson’s Ratio of the rock. For a truly predictive model these variables need to be incorporated.
- Nevertheless, the results are important because they inform Montney gas play operators on steps they should take. Likely, the minimum of fractures and chemicals to stimulate the well would improve the economic results to Montney play operators. Not spending extra on expensive proppants or other inputs would similarly improve the economic results. The intensity of the effort in terms of fracs, chemicals and proppant was not predictive.
- What is predictive is defining where the highest quality reservoir exists (porosity and permeability) and where the most brittle and fracable rock exists (Young’s modulus and Poisson’s ratio). This should be the focus of Montney play operators and the data scientists who support these operators.

Going forward we will produce an alternative model that produces better R-squared Adj, but for generating this model we ignored significance of predictors in simple additive model that was generated through stepwise, backwards, and forwards regression methods. For purpose of this project we are accepting best fit model as presented in report but proposing this alternative model.

1. Full additive model and Multicollinearity test.

First we will test for multicollinearity of the full additive model before proceeding with stepwise, backwards, forwards regression. We would have to make sure that these regression methods do not select for variables that are multicollinear early on in our analysis before we move onto formulating our model with interactive terms. Using the VIF function, we will get GVIF values as we have dummy variables. For one coefficient variables, $GVIF^{(1/(2*Df))}$ of 2 would equal VIF of 4. In this instance we will slowly remove variables with the highest $GVIF^{(1/(2*Df))}$ until all variables have $GVIF^{(1/(2*Df))}$ of less than 1.5. We chose 1.5 using some foreknowledge as we have found in preliminary testing that $GVIF^{(1/(2*Df))}$ over 1.5 seems to cause problems with stepwise, backwards, and forwards regression methods where multicollinear values will crowd out any other predictor variables despite the fact that they may still be useful.

Let's proceed with the test:

```
testfull = lm(First36MonthProd_BOE~
                 MD_FT+
                 factor(RSFracJobType)+
                 SpudToSales_DAYS+
                 PerfInterval_FT+
                 FracStages+
                 AverageStageSpacing_FT+
                 ProppantLoading_LBSPerGAL+
                 factor(RSProppantType)+
                 ProppantIntensity_LBSPerFT+
                 FluidIntensity_BBLPerFT+
                 FrictionReducer_LBS+
                 Surfactant_LBS
                 , data=data1)
vif(testfull)
```

```
##                                 GVIF Df GVIF^(1/(2*Df))
## MD_FT                       3.323646  1        1.823087
## factor(RSFracJobType)       1.896737  3        1.112588
## SpudToSales_DAYS            1.328031  1        1.152402
## PerfInterval_FT             7.509321  1        2.740314
## FracStages                 10.856960  1        3.294990
## AverageStageSpacing_FT     13.531963  1        3.678582
## ProppantLoading_LBSPerGAL   8.349051  1        2.889472
## factor(RSProppantType)      1.735987  2        1.147854
## ProppantIntensity_LBSPerFT  8.907449  1        2.984535
## FluidIntensity_BBLPerFT     8.398636  1        2.898040
## FrictionReducer_LBS         2.049106  1        1.431470
## Surfactant_LBS              1.682525  1        1.297122
```

We see that **AverageStageSpacing_FT** has the highest $GVIF^{(1/(2*Df))}$ of 3.678582 which is greater than 1.5 so we will remove that first.

```
testfull2 = lm(First36MonthProd_BOE~
                 MD_FT+
                 factor(RSFracJobType)+
                 SpudToSales_DAYS+
                 PerfInterval_FT+
                 FracStages+
                 #AverageStageSpacing_FT+
                 ProppantLoading_LBSPerGAL+
                 factor(RSProppantType)+
                 ProppantIntensity_LBSPerFT+
                 FluidIntensity_BBLPerFT+
                 FrictionReducer_LBS+
                 Surfactant_LBS
                 , data=data1)
vif(testfull2)
```

```
##                                   GVIF Df GVIF^(1/(2*Df))
## MD_FT                       3.292514  1        1.814529
## factor(RSFracJobType)       1.845799  3        1.107552
## SpudToSales_DAYS            1.326570  1        1.151768
## PerfInterval_FT             2.709359  1        1.646013
## FracStages                  1.425532  1        1.193956
## ProppantLoading_LBSPerGAL   8.241393  1        2.870783
## factor(RSProppantType)      1.668702  2        1.136566
## ProppantIntensity_LBSPerFT  8.414681  1        2.900807
## FluidIntensity_BBLPerFT     8.103969  1        2.846747
## FrictionReducer_LBS         1.934207  1        1.390758
## Surfactant_LBS              1.681526  1        1.296737
```

ProppantIntensity_LBSPerFT has the highest $GVIF^{(1/(2*Df))}$ of 2.870783. We will remove this.

```
testfull3 = lm(First36MonthProd_BOE~
                 MD_FT+
                 factor(RSFracJobType)+
                 SpudToSales_DAYS+
                 PerfInterval_FT +
                 FracStages+
                 #AverageStageSpacing_FT+
                 ProppantLoading_LBSPerGAL+
                 factor(RSProppantType)+
                 #ProppantIntensity_LBSPerFT+
                 FluidIntensity_BBLPerFT+
                 FrictionReducer_LBS+
                 Surfactant_LBS
                 , data=data1)
vif(testfull3)
```

```
##                                GVIF Df GVIF^(1/(2*Df))
## MD_FT                      3.143967  1        1.773124
## factor(RSFracJobType)      1.800841  3        1.103009
## SpudToSales_DAYS           1.298002  1        1.139299
## PerfInterval_FT            2.594010  1        1.610593
## FracStages                 1.311097  1        1.145031
## ProppantLoading_LBSPerGAL  1.766738  1        1.329187
## factor(RSProppantType)     1.663802  2        1.135731
## FluidIntensity_BBLPerFT    2.098633  1        1.448666
## FrictionReducer_LBS        1.907851  1        1.381250
## Surfactant_LBS             1.629925  1        1.276685
```

MD_FT has the highest $GVIF^{(1/(2*Df))}$ of 1.773124. This will be removed.

```
testfull4 = lm(First36MonthProd_BOE~
                 #MD_FT+
                 factor(RSFracJobType)+
                 SpudToSales_DAYS+
                 PerfInterval_FT +
                 FracStages+
                 #AverageStageSpacing_FT+
                 ProppantLoading_LBSPerGAL+
                 factor(RSProppantType)+
                 #ProppantIntensity_LBSPerFT+
                 FluidIntensity_BBLPerFT+
                 FrictionReducer_LBS+
                 Surfactant_LBS
                 , data=data1)
vif(testfull4)
```

```
##                                GVIF Df GVIF^(1/(2*Df))
## factor(RSFracJobType)      1.725073  3        1.095136
## SpudToSales_DAYS           1.285900  1        1.133975
## PerfInterval_FT            1.315207  1        1.146825
## FracStages                 1.223422  1        1.106084
## ProppantLoading_LBSPerGAL  1.661806  1        1.289110
## factor(RSProppantType)     1.625719  2        1.129175
## FluidIntensity_BBLPerFT    1.972725  1        1.404537
## FrictionReducer_LBS        1.726552  1        1.313983
## Surfactant_LBS             1.625055  1        1.274776
```

All $GVIF^{(1/(2*Df))}$ values are less than 1.5 so we can now proceed with the regression attemps.

```
initmodel = testfull4
```

2. Stepwise, Backwards, Forwards Regression

Let's set up the hypothesis for testing the overall model:
$H_0: \ All\ \beta_i = 0$ Where the i refers to the variables: PerfInterval_FT, FracStages, RSProppantType, …
$H_a: \text{At least one } \beta_i \neq 0$

a) Stepwise Regression

```
initstep = ols_step_both_p(initmodel, pent=0.1, prem=0.3, details = FALSE)
```

```
## Stepwise Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. factor(RSFracJobType)
## 2. SpudToSales_DAYS
## 3. PerfInterval_FT
## 4. FracStages
## 5. ProppantLoading_LBSPerGAL
## 6. factor(RSProppantType)
## 7. FluidIntensity_BBLPerFT
## 8. FrictionReducer_LBS
## 9. Surfactant_LBS
##
## We are selecting variables based on p value...
##
## Variables Entered/Removed:
##
## - PerfInterval_FT added
## - factor(RSProppantType) added
```

```
## Note: model has aliased coefficients
##       sums of squares computed by model comparison
```

```
## - FrictionReducer_LBS added
## - FluidIntensity_BBLPerFT added
## - Surfactant_LBS added
## - FracStages added
## - factor(RSFracJobType) added
##
## No more variables to be added/removed.
##
##
## Final Model Output
## ------------------
##
##                             Model Summary
## ----------------------------------------------------------------------
## R                       0.649       RMSE                   167962.442
## R-Squared               0.421       Coef. Var                  47.108
## Adj. R-Squared          0.382       MSE                28211382079.766
## Pred R-Squared           -Inf       MAE                    124060.512
## ----------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                    ANOVA
## -----------------------------------------------------------------------------
##                    Sum of
##                   Squares         DF        Mean Square      F         Sig.
## -----------------------------------------------------------------------------
## Regression    2.998798e+12         10    299879835006.237     10.63    0.0000
## Residual      4.118862e+12        146     28211382079.766
## Total          7.11766e+12        156
## -----------------------------------------------------------------------------
##
##                                                       Parameter Estimates
## ------------------------------------------------------------------------------------------------------------------------
--------
##                                  model           Beta    Std. Error    Std. Beta      t        Sig           lower
upper
## ------------------------------------------------------------------------------------------------------------------------
--------
##                            (Intercept)     -81196.040    261402.824                -0.311    0.757    -597818.375    435
426.295
##                         PerfInterval_FT         60.768        13.284        0.319     4.574    0.000         34.514
87.022
##     factor(RSProppantType)SAND + RESIN    199052.749    188618.929        0.459     1.055    0.293    -173723.452    571
828.950
##        factor(RSProppantType)SAND ONLY     11023.827    188900.303        0.025     0.058    0.954    -362308.467    384
356.120
##                     FrictionReducer_LBS          3.784         2.141        0.146     1.768    0.079         -0.447
8.015
##                 FluidIntensity_BBLPerFT     -10339.599      3947.841       -0.217    -2.619    0.010      -18141.898     -2
537.300
##                          Surfactant_LBS          6.764         2.117        0.242     3.196    0.002          2.581
10.948
##                              FracStages      -6209.923      2272.829       -0.176    -2.732    0.007      -10701.819     -1
718.027
##         factor(RSFracJobType)ENERGIZED    169057.153    187647.552        0.152     0.901    0.369    -201799.271    539
913.577
## factor(RSFracJobType)SLICKCROSSLINKHYB    -319551.893    219398.672       -0.168    -1.456    0.147    -753159.505    114
055.719
##        factor(RSFracJobType)SLICKWATER    119934.383    171472.364        0.131     0.699    0.485    -218954.279    458
823.044
## ------------------------------------------------------------------------------------------------------------------------
--------
```

```
summary(initstep$model)
```

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -429063  -98010  -30471   81748  545751
##
## Coefficients:
##                                       Estimate Std. Error t value
## (Intercept)                          -8.120e+04  2.614e+05  -0.311
## PerfInterval_FT                       6.077e+01  1.328e+01   4.574
## factor(RSProppantType)SAND + RESIN     1.991e+05  1.886e+05   1.055
## factor(RSProppantType)SAND ONLY        1.102e+04  1.889e+05   0.058
## FrictionReducer_LBS                   3.784e+00  2.141e+00   1.768
## FluidIntensity_BBLPerFT              -1.034e+04  3.948e+03  -2.619
## Surfactant_LBS                        6.764e+00  2.117e+00   3.196
## FracStages                           -6.210e+03  2.273e+03  -2.732
## factor(RSFracJobType)ENERGIZED         1.691e+05  1.876e+05   0.901
## factor(RSFracJobType)SLICKCROSSLINKHYB -3.196e+05  2.194e+05  -1.456
## factor(RSFracJobType)SLICKWATER        1.199e+05  1.715e+05   0.699
##                                      Pr(>|t|)
## (Intercept)                           0.75654
## PerfInterval_FT                      1.01e-05 ***
## factor(RSProppantType)SAND + RESIN     0.29302
## factor(RSProppantType)SAND ONLY        0.95354
## FrictionReducer_LBS                   0.07921 .
## FluidIntensity_BBLPerFT               0.00975 **
## Surfactant_LBS                        0.00171 **
## FracStages                            0.00707 **
## factor(RSFracJobType)ENERGIZED         0.36911
## factor(RSFracJobType)SLICKCROSSLINKHYB  0.14740
## factor(RSFracJobType)SLICKWATER        0.48539
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 168000 on 146 degrees of freedom
##   (86 observations deleted due to missingness)
## Multiple R-squared:  0.4213, Adjusted R-squared:  0.3817
## F-statistic: 10.63 on 10 and 146 DF,  p-value: 2.093e-13
```

b) Backwards Regression

```
initback=ols_step_backward_p(initmodel, prem = 0.1, details=FALSE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . factor(RSFracJobType)
## 2 . SpudToSales_DAYS
## 3 . PerfInterval_FT
## 4 . FracStages
## 5 . ProppantLoading_LBSPerGAL
## 6 . factor(RSProppantType)
## 7 . FluidIntensity_BBLPerFT
## 8 . FrictionReducer_LBS
## 9 . Surfactant_LBS
##
## We are eliminating variables based on p value...
##
## Variables Removed:
##
## - ProppantLoading_LBSPerGAL
## - SpudToSales_DAYS
##
## No more variables satisfy the condition of p value = 0.1
##
##
## Final Model Output
## ------------------
##
##                              Model Summary
## --------------------------------------------------------------------
## R                       0.649       RMSE                  167962.442
## R-Squared               0.421       Coef. Var                 47.108
## Adj. R-Squared          0.382       MSE                28211382079.766
## Pred R-Squared           -Inf       MAE                   124060.512
## --------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## ---------------------------------------------------------------------------
##                    Sum of
##                   Squares         DF        Mean Square      F         Sig.
## ---------------------------------------------------------------------------
## Regression    2.998798e+12         10    299879835006.238     10.63    0.0000
## Residual      4.118862e+12        146     28211382079.766
## Total          7.11766e+12        156
## ---------------------------------------------------------------------------
##
##                                                          Parameter Estimates
## ----------------------------------------------------------------------------------------------------------------------------
--------
##                                     model           Beta    Std. Error    Std. Beta      t        Sig           lower
upper
## ----------------------------------------------------------------------------------------------------------------------------
--------
##                               (Intercept)     -81196.040    261402.824                 -0.311    0.757    -597818.375    435
426.295
##          factor(RSFracJobType)ENERGIZED     169057.153    187647.552        0.152     0.901    0.369    -201799.271    539
913.577
## factor(RSFracJobType)SLICKCROSSLINKHYB    -319551.893    219398.672       -0.168    -1.456    0.147    -753159.505    114
055.719
##         factor(RSFracJobType)SLICKWATER     119934.383    171472.364        0.131     0.699    0.485    -218954.279    458
823.044
##                           PerfInterval_FT         60.768        13.284        0.319     4.574    0.000         34.514
87.022
##                                FracStages      -6209.923      2272.829       -0.176    -2.732    0.007     -10701.819     -1
718.027
##     factor(RSProppantType)SAND + RESIN     199052.749    188618.929        0.459     1.055    0.293    -173723.452    571
828.950
##        factor(RSProppantType)SAND ONLY      11023.827    188900.303        0.025     0.058    0.954    -362308.467    384
356.120
##                   FluidIntensity_BBLPerFT     -10339.599      3947.841       -0.217    -2.619    0.010     -18141.898     -2
537.300
```

```
##                  FrictionReducer_LBS          3.784         2.141        0.146     1.768    0.079         -0.447
8.015
##                      Surfactant_LBS          6.764         2.117        0.242     3.196    0.002          2.581
10.948
## ------------------------------------------------------------------------------------------------------------------------
--------
```

```
summary(initback$model)
```

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -429063  -98010  -30471   81748  545751
##
## Coefficients:
##                                       Estimate Std. Error t value
## (Intercept)                          -8.120e+04  2.614e+05  -0.311
## factor(RSFracJobType)ENERGIZED        1.691e+05  1.876e+05   0.901
## factor(RSFracJobType)SLICKCROSSLINKHYB -3.196e+05  2.194e+05  -1.456
## factor(RSFracJobType)SLICKWATER       1.199e+05  1.715e+05   0.699
## PerfInterval_FT                       6.077e+01  1.328e+01   4.574
## FracStages                           -6.210e+03  2.273e+03  -2.732
## factor(RSProppantType)SAND + RESIN    1.991e+05  1.886e+05   1.055
## factor(RSProppantType)SAND ONLY       1.102e+04  1.889e+05   0.058
## FluidIntensity_BBLPerFT              -1.034e+04  3.948e+03  -2.619
## FrictionReducer_LBS                   3.784e+00  2.141e+00   1.768
## Surfactant_LBS                        6.764e+00  2.117e+00   3.196
##                                      Pr(>|t|)
## (Intercept)                           0.75654
## factor(RSFracJobType)ENERGIZED        0.36911
## factor(RSFracJobType)SLICKCROSSLINKHYB  0.14740
## factor(RSFracJobType)SLICKWATER       0.48539
## PerfInterval_FT                      1.01e-05 ***
## FracStages                            0.00707 **
## factor(RSProppantType)SAND + RESIN    0.29302
## factor(RSProppantType)SAND ONLY       0.95354
## FluidIntensity_BBLPerFT               0.00975 **
## FrictionReducer_LBS                   0.07921 .
## Surfactant_LBS                        0.00171 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 168000 on 146 degrees of freedom
##   (86 observations deleted due to missingness)
## Multiple R-squared:  0.4213, Adjusted R-squared:  0.3817
## F-statistic: 10.63 on 10 and 146 DF,  p-value: 2.093e-13
```

c) Forward Regression

```
initfor=ols_step_forward_p(initmodel,penter = 0.1, details=FALSE)
```

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. factor(RSFracJobType)
## 2. SpudToSales_DAYS
## 3. PerfInterval_FT
## 4. FracStages
## 5. ProppantLoading_LBSPerGAL
## 6. factor(RSProppantType)
## 7. FluidIntensity_BBLPerFT
## 8. FrictionReducer_LBS
## 9. Surfactant_LBS
##
## We are selecting variables based on p value...
##
## Variables Entered:
##
## - PerfInterval_FT
## - factor(RSProppantType)
```

```
## Note: model has aliased coefficients
##       sums of squares computed by model comparison
```

```
## - FrictionReducer_LBS
## - FluidIntensity_BBLPerFT
## - Surfactant_LBS
## - FracStages
## - factor(RSFracJobType)
##
## No more variables to be added.
##
## Final Model Output
## ------------------
##
##                              Model Summary
## ---------------------------------------------------------------------
## R                        0.649       RMSE                  167962.442
## R-Squared                0.421       Coef. Var                 47.108
## Adj. R-Squared           0.382       MSE                28211382079.766
## Pred R-Squared            -Inf       MAE                   124060.512
## ---------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                     ANOVA
## ----------------------------------------------------------------------------
##                    Sum of
##                   Squares         DF        Mean Square      F         Sig.
## ----------------------------------------------------------------------------
## Regression    2.998798e+12        10    299879835006.237     10.63    0.0000
## Residual      4.118862e+12       146     28211382079.766
## Total          7.11766e+12       156
## ----------------------------------------------------------------------------
##
##                                                    Parameter Estimates
## ------------------------------------------------------------------------------------------------------------------------
--------
##                                  model           Beta    Std. Error    Std. Beta      t        Sig           lower
upper
## ------------------------------------------------------------------------------------------------------------------------
--------
##                            (Intercept)     -81196.040    261402.824                -0.311    0.757    -597818.375    435
426.295
##                         PerfInterval_FT         60.768        13.284        0.319     4.574    0.000         34.514
87.022
##     factor(RSProppantType)SAND + RESIN    199052.749    188618.929        0.459     1.055    0.293    -173723.452    571
828.950
##        factor(RSProppantType)SAND ONLY     11023.827    188900.303        0.025     0.058    0.954    -362308.467    384
356.120
##                     FrictionReducer_LBS          3.784         2.141        0.146     1.768    0.079         -0.447
8.015
##                 FluidIntensity_BBLPerFT    -10339.599      3947.841       -0.217    -2.619    0.010     -18141.898     -2
537.300
##                          Surfactant_LBS          6.764         2.117        0.242     3.196    0.002          2.581
10.948
##                              FracStages      -6209.923      2272.829       -0.176    -2.732    0.007     -10701.819     -1
718.027
##         factor(RSFracJobType)ENERGIZED    169057.153    187647.552        0.152     0.901    0.369    -201799.271    539
913.577
## factor(RSFracJobType)SLICKCROSSLINKHYB   -319551.893    219398.672       -0.168    -1.456    0.147    -753159.505    114
055.719
##        factor(RSFracJobType)SLICKWATER    119934.383    171472.364        0.131     0.699    0.485    -218954.279    458
823.044
## ------------------------------------------------------------------------------------------------------------------------
--------
```

```
summary(initfor$model)
```

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -429063  -98010  -30471   81748  545751
##
## Coefficients:
##                                        Estimate Std. Error t value
## (Intercept)                           -8.120e+04  2.614e+05  -0.311
## PerfInterval_FT                        6.077e+01  1.328e+01   4.574
## factor(RSProppantType)SAND + RESIN      1.991e+05  1.886e+05   1.055
## factor(RSProppantType)SAND ONLY         1.102e+04  1.889e+05   0.058
## FrictionReducer_LBS                    3.784e+00  2.141e+00   1.768
## FluidIntensity_BBLPerFT               -1.034e+04  3.948e+03  -2.619
## Surfactant_LBS                         6.764e+00  2.117e+00   3.196
## FracStages                            -6.210e+03  2.273e+03  -2.732
## factor(RSFracJobType)ENERGIZED          1.691e+05  1.876e+05   0.901
## factor(RSFracJobType)SLICKCROSSLINKHYB -3.196e+05  2.194e+05  -1.456
## factor(RSFracJobType)SLICKWATER         1.199e+05  1.715e+05   0.699
##                                       Pr(>|t|)
## (Intercept)                            0.75654
## PerfInterval_FT                       1.01e-05 ***
## factor(RSProppantType)SAND + RESIN      0.29302
## factor(RSProppantType)SAND ONLY         0.95354
## FrictionReducer_LBS                    0.07921 .
## FluidIntensity_BBLPerFT                0.00975 **
## Surfactant_LBS                         0.00171 **
## FracStages                             0.00707 **
## factor(RSFracJobType)ENERGIZED          0.36911
## factor(RSFracJobType)SLICKCROSSLINKHYB  0.14740
## factor(RSFracJobType)SLICKWATER         0.48539
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 168000 on 146 degrees of freedom
##   (86 observations deleted due to missingness)
## Multiple R-squared:  0.4213, Adjusted R-squared:  0.3817
## F-statistic: 10.63 on 10 and 146 DF,  p-value: 2.093e-13
```

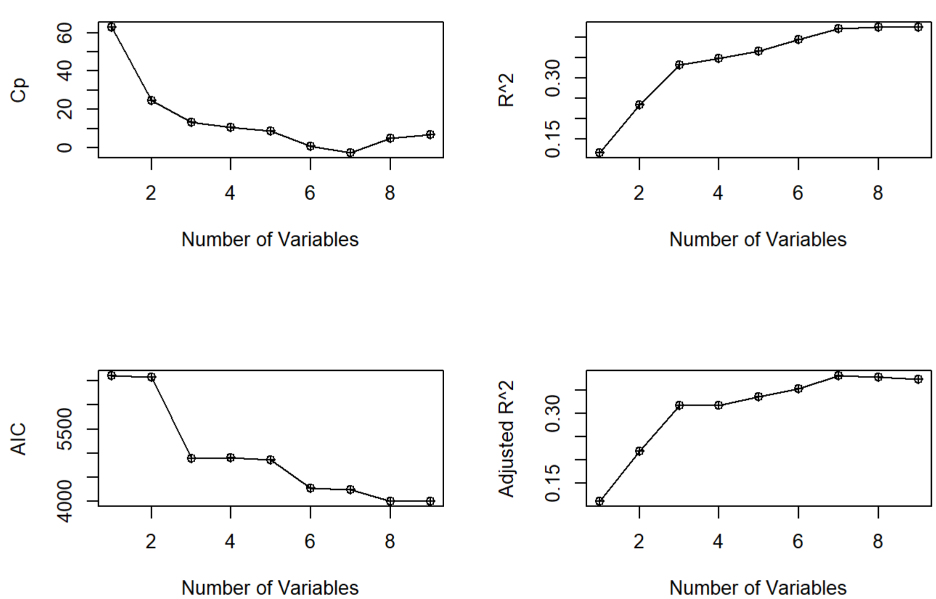
Stepwise, backwards, and forwards regression methods all yielded the same model. We see that the P-val for the overall model is less than 0.05 so we can reject the null that all β_i is 0. The most insignificant values seem to be the factors **RSProppantType**, and **RSFracJobType** but we will not remove them yet to see if they will contribute to the interactive term. The model has R^2_{adj} of 0.3817 and an RMSE of 168000.

d) AIC, Cp, R^2_{adj}

```
ks=ols_step_best_subset(initmodel, details=TRUE)
```

```
## Warning in b * sx: longer object length is not a multiple of shorter object
## length

## Warning in b * sx: longer object length is not a multiple of shorter object
## length

## Warning in b * sx: longer object length is not a multiple of shorter object
## length

## Warning in b * sx: longer object length is not a multiple of shorter object
## length

## Warning in b * sx: longer object length is not a multiple of shorter object
## length

## Warning in b * sx: longer object length is not a multiple of shorter object
## length

## Warning in b * sx: longer object length is not a multiple of shorter object
## length

## Warning in b * sx: longer object length is not a multiple of shorter object
## length
```

```
par(mfrow=c(2,2)) # split the plotting panel into a 2 x 2 grid
plot(ks$cp,type = "o",pch=10, xlab="Number of Variables",ylab= "Cp")
plot(ks$rsq,type = "o",pch=10, xlab="Number of Variables",ylab= "R^2")
#plot(ks$rss, xlab="Number of Variables",ylab= "RMSE")
plot(ks$aic,type = "o",pch=10, xlab="Number of Variables",ylab= "AIC")
plot(ks$adjr,type = "o",pch=10, xlab="Number of Variables",ylab= "Adjusted R^2")
```




```
best.subset<-regsubsets(First36MonthProd_BOE~
                 #MD_FT+
                 factor(RSFracJobType)+
                 SpudToSales_DAYS+
                 PerfInterval_FT +
                 FracStages+
                 #AverageStageSpacing_FT+
                 ProppantLoading_LBSPerGAL+
                 factor(RSProppantType)+
                 #ProppantIntensity_LBSPerFT+
                 FluidIntensity_BBLPerFT+
                 FrictionReducer_LBS+
                 Surfactant_LBS
                 , data=data1, nv=10 )
```

```
## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax,
## force.in = force.in, : 4 linear dependencies found
```

```
## Reordering variables and trying again:
```

```
#by default, regsubsets() only reports results up to the best 8-variable model
#Model selection by exhaustive search, forward or backward stepwise, or sequential replacement
#The summary() command outputs the best set of variables for each model size using RMSE.
summary(best.subset)
```

```
## Subset selection object
## Call: regsubsets.formula(First36MonthProd_BOE ~ factor(RSFracJobType) +
##     SpudToSales_DAYS + PerfInterval_FT + FracStages + ProppantLoading_LBSPerGAL +
##     factor(RSProppantType) + FluidIntensity_BBLPerFT + FrictionReducer_LBS +
##     Surfactant_LBS, data = data1, nv = 10)
## 16 Variables  (and intercept)
##                                        Forced in Forced out
## factor(RSFracJobType)CROSSLINK             FALSE      FALSE
## factor(RSFracJobType)ENERGIZED             FALSE      FALSE
## factor(RSFracJobType)SLICKCROSSLINKHYB     FALSE      FALSE
## SpudToSales_DAYS                           FALSE      FALSE
## PerfInterval_FT                            FALSE      FALSE
## FracStages                                 FALSE      FALSE
## ProppantLoading_LBSPerGAL                  FALSE      FALSE
## factor(RSProppantType)SAND + RESIN         FALSE      FALSE
## factor(RSProppantType)SAND ONLY            FALSE      FALSE
## FluidIntensity_BBLPerFT                    FALSE      FALSE
## FrictionReducer_LBS                        FALSE      FALSE
## Surfactant_LBS                             FALSE      FALSE
## factor(RSFracJobType)SLICKLINGELHYB        FALSE      FALSE
## factor(RSFracJobType)SLICKWATER            FALSE      FALSE
## factor(RSFracJobType)UNKNOWN               FALSE      FALSE
## factor(RSProppantType)UNKNOWN              FALSE      FALSE
## 1 subsets of each size up to 11
## Selection Algorithm: exhaustive
##           factor(RSFracJobType)CROSSLINK factor(RSFracJobType)ENERGIZED
## 1  ( 1 )  " "                            " "
## 2  ( 1 )  " "                            " "
## 3  ( 1 )  " "                            " "
## 4  ( 1 )  " "                            " "
## 5  ( 1 )  " "                            " "
## 6  ( 1 )  " "                            " "
## 7  ( 1 )  " "                            " "
## 8  ( 1 )  " "                            " "
## 9  ( 1 )  "*"                            " "
## 10  ( 1 ) "*"                            "*"
## 11  ( 1 ) "*"                            " "
##           factor(RSFracJobType)SLICKCROSSLINKHYB
## 1  ( 1 )  " "
## 2  ( 1 )  " "
## 3  ( 1 )  " "
## 4  ( 1 )  "*"
## 5  ( 1 )  "*"
## 6  ( 1 )  "*"
## 7  ( 1 )  "*"
## 8  ( 1 )  "*"
## 9  ( 1 )  "*"
## 10  ( 1 ) "*"
## 11  ( 1 ) "*"
##           factor(RSFracJobType)SLICKLINGELHYB
## 1  ( 1 )  " "
## 2  ( 1 )  " "
## 3  ( 1 )  " "
## 4  ( 1 )  " "
## 5  ( 1 )  " "
## 6  ( 1 )  " "
## 7  ( 1 )  " "
## 8  ( 1 )  " "
## 9  ( 1 )  " "
## 10  ( 1 ) " "
## 11  ( 1 ) " "
##           factor(RSFracJobType)SLICKWATER factor(RSFracJobType)UNKNOWN
## 1  ( 1 )  " "                             " "
## 2  ( 1 )  " "                             " "
## 3  ( 1 )  " "                             " "
## 4  ( 1 )  " "                             " "
## 5  ( 1 )  " "                             " "
## 6  ( 1 )  " "                             " "
## 7  ( 1 )  " "                             " "
## 8  ( 1 )  " "                             " "
## 9  ( 1 )  " "                             " "
## 10  ( 1 ) " "                             " "
## 11  ( 1 ) "*"                             " "
##           SpudToSales_DAYS PerfInterval_FT FracStages
```

```
## 1  ( 1 )  " "              " "             " "
## 2  ( 1 )  " "              "*"             " "
## 3  ( 1 )  " "              "*"             "*"
## 4  ( 1 )  " "              "*"             " "
## 5  ( 1 )  " "              "*"             "*"
## 6  ( 1 )  " "              "*"             "*"
## 7  ( 1 )  " "              "*"             "*"
## 8  ( 1 )  "*"              "*"             "*"
## 9  ( 1 )  "*"              "*"             "*"
## 10  ( 1 ) "*"              "*"             "*"
## 11  ( 1 ) "*"              "*"             "*"
##           ProppantLoading_LBSPerGAL factor(RSProppantType)SAND + RESIN
## 1  ( 1 )  " "                       " "
## 2  ( 1 )  " "                       " "
## 3  ( 1 )  " "                       " "
## 4  ( 1 )  " "                       " "
## 5  ( 1 )  " "                       " "
## 6  ( 1 )  " "                       " "
## 7  ( 1 )  " "                       "*"
## 8  ( 1 )  " "                       "*"
## 9  ( 1 )  " "                       "*"
## 10  ( 1 ) " "                       "*"
## 11  ( 1 ) "*"                       "*"
##           factor(RSProppantType)SAND ONLY factor(RSProppantType)UNKNOWN
## 1  ( 1 )  "*"                             " "
## 2  ( 1 )  "*"                             " "
## 3  ( 1 )  "*"                             " "
## 4  ( 1 )  "*"                             " "
## 5  ( 1 )  "*"                             " "
## 6  ( 1 )  "*"                             " "
## 7  ( 1 )  " "                             " "
## 8  ( 1 )  " "                             " "
## 9  ( 1 )  " "                             " "
## 10  ( 1 ) " "                             " "
## 11  ( 1 ) " "                             " "
##           FluidIntensity_BBLPerFT FrictionReducer_LBS Surfactant_LBS
## 1  ( 1 )  " "                     " "                 " "
## 2  ( 1 )  " "                     " "                 " "
## 3  ( 1 )  " "                     " "                 " "
## 4  ( 1 )  " "                     " "                 "*"
## 5  ( 1 )  " "                     " "                 "*"
## 6  ( 1 )  "*"                     " "                 "*"
## 7  ( 1 )  "*"                     "*"                 "*"
## 8  ( 1 )  "*"                     "*"                 "*"
## 9  ( 1 )  "*"                     "*"                 "*"
## 10  ( 1 ) "*"                     "*"                 "*"
## 11  ( 1 ) "*"                     "*"                 "*"
```

```
reg.summary<-summary(best.subset)

# for the output interpretation
rsquare<-c(reg.summary$rsq)
cp<-c(reg.summary$cp)
AdjustedR<-c(reg.summary$adjr2)
RMSE<-c(reg.summary$rss)
cbind(rsquare,cp,RMSE,AdjustedR)
```

```
##         rsquare         cp         RMSE AdjustedR
##  [1,] 0.1851843 41.6304539 5.691184e+12 0.1796034
##  [2,] 0.3021244 16.9892853 4.874401e+12 0.2924985
##  [3,] 0.3290574 12.8534492 4.686284e+12 0.3150794
##  [4,] 0.3516287  9.7112601 4.528631e+12 0.3334925
##  [5,] 0.3870415  3.6435518 4.281286e+12 0.3654585
##  [6,] 0.4048260  1.5919116 4.157068e+12 0.3794995
##  [7,] 0.4187428  0.4214126 4.059865e+12 0.3896799
##  [8,] 0.4214915  1.7951931 4.040666e+12 0.3881961
##  [9,] 0.4232143  3.4027140 4.028633e+12 0.3855978
## [10,] 0.4246879  5.0670001 4.018340e+12 0.3826943
## [11,] 0.4249627  7.0044029 4.016421e+12 0.3784523
```

2. Building the Interactive Model

Here we construct the full interactive model and proceed to eliminate all interactive terms that are non-significant. We will also see that removing some interactive terms will cause and entire variable to become non-significant. We will elminate those as well further down.

```
initint = lm(First36MonthProd_BOE~(PerfInterval_FT+
                factor(RSProppantType)+
                FrictionReducer_LBS+
                FluidIntensity_BBLPerFT+
                Surfactant_LBS+
                FracStages+
                factor(RSFracJobType))^2, data=data1)
summary(initint)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ (PerfInterval_FT + factor(RSProppantType) +
##     FrictionReducer_LBS + FluidIntensity_BBLPerFT + Surfactant_LBS +
##     FracStages + factor(RSFracJobType))^2, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -400468  -78949  -10779   56834  456917
##
## Coefficients: (20 not defined because of singularities)
##                                                                      Estimate
## (Intercept)                                                         3.032e+05
## PerfInterval_FT                                                    -7.623e+01
## factor(RSProppantType)SAND + RESIN                                  1.253e+05
## factor(RSProppantType)SAND ONLY                                    -2.636e+04
## FrictionReducer_LBS                                                 5.485e+01
## FluidIntensity_BBLPerFT                                             1.671e+02
## Surfactant_LBS                                                     -3.363e+01
## FracStages                                                         -1.012e+04
## factor(RSFracJobType)ENERGIZED                                     -1.040e+08
## factor(RSFracJobType)SLICKCROSSLINKHYB                             -5.950e+05
## factor(RSFracJobType)SLICKWATER                                     1.540e+05
## PerfInterval_FT:factor(RSProppantType)SAND + RESIN                  3.936e+01
## PerfInterval_FT:factor(RSProppantType)SAND ONLY                            NA
## PerfInterval_FT:FrictionReducer_LBS                                -2.331e-03
## PerfInterval_FT:FluidIntensity_BBLPerFT                             3.018e+00
## PerfInterval_FT:Surfactant_LBS                                      4.795e-03
## PerfInterval_FT:FracStages                                          2.332e+00
## PerfInterval_FT:factor(RSFracJobType)ENERGIZED                     -2.402e+04
## PerfInterval_FT:factor(RSFracJobType)SLICKCROSSLINKHYB              8.616e+01
## PerfInterval_FT:factor(RSFracJobType)SLICKWATER                            NA
## factor(RSProppantType)SAND + RESIN:FrictionReducer_LBS              4.413e-01
## factor(RSProppantType)SAND ONLY:FrictionReducer_LBS                        NA
## factor(RSProppantType)SAND + RESIN:FluidIntensity_BBLPerFT         -8.258e+03
## factor(RSProppantType)SAND ONLY:FluidIntensity_BBLPerFT                    NA
## factor(RSProppantType)SAND + RESIN:Surfactant_LBS                   9.276e+00
## factor(RSProppantType)SAND ONLY:Surfactant_LBS                             NA
## factor(RSProppantType)SAND + RESIN:FracStages                      -9.729e+03
## factor(RSProppantType)SAND ONLY:FracStages                                 NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)ENERGIZED   2.010e+08
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)ENERGIZED             NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)SLICKCROSSLINKHYB  NA
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)SLICKCROSSLINKHYB     NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)SLICKWATER         NA
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)SLICKWATER            NA
## FrictionReducer_LBS:FluidIntensity_BBLPerFT                        -2.598e+00
## FrictionReducer_LBS:Surfactant_LBS                                 -1.045e-03
## FrictionReducer_LBS:FracStages                                      3.045e-01
## FrictionReducer_LBS:factor(RSFracJobType)ENERGIZED                  1.241e+04
## FrictionReducer_LBS:factor(RSFracJobType)SLICKCROSSLINKHYB                 NA
## FrictionReducer_LBS:factor(RSFracJobType)SLICKWATER                        NA
## FluidIntensity_BBLPerFT:Surfactant_LBS                              2.206e+00
## FluidIntensity_BBLPerFT:FracStages                                 -5.312e+02
## FluidIntensity_BBLPerFT:factor(RSFracJobType)ENERGIZED             -8.297e+06
## FluidIntensity_BBLPerFT:factor(RSFracJobType)SLICKCROSSLINKHYB             NA
## FluidIntensity_BBLPerFT:factor(RSFracJobType)SLICKWATER                    NA
## Surfactant_LBS:FracStages                                          -4.104e-01
## Surfactant_LBS:factor(RSFracJobType)ENERGIZED                       7.673e+02
## Surfactant_LBS:factor(RSFracJobType)SLICKCROSSLINKHYB                      NA
## Surfactant_LBS:factor(RSFracJobType)SLICKWATER                             NA
## FracStages:factor(RSFracJobType)ENERGIZED                                  NA
## FracStages:factor(RSFracJobType)SLICKCROSSLINKHYB                          NA
## FracStages:factor(RSFracJobType)SLICKWATER                                 NA
##                                                                    Std. Error
## (Intercept)                                                         5.049e+05
## PerfInterval_FT                                                     6.910e+01
## factor(RSProppantType)SAND + RESIN                                  3.535e+05
## factor(RSProppantType)SAND ONLY                                     2.622e+05
## FrictionReducer_LBS                                                 3.036e+01
## FluidIntensity_BBLPerFT                                             2.693e+04
## Surfactant_LBS                                                      1.773e+01
## FracStages                                                          1.640e+04
## factor(RSFracJobType)ENERGIZED                                      9.045e+07
```

```
## factor(RSFracJobType)SLICKCROSSLINKHYB                              6.482e+05
## factor(RSFracJobType)SLICKWATER                                     1.737e+05
## PerfInterval_FT:factor(RSProppantType)SAND + RESIN                  3.487e+01
## PerfInterval_FT:factor(RSProppantType)SAND ONLY                            NA
## PerfInterval_FT:FrictionReducer_LBS                                 3.056e-03
## PerfInterval_FT:FluidIntensity_BBLPerFT                             4.629e+00
## PerfInterval_FT:Surfactant_LBS                                      2.630e-03
## PerfInterval_FT:FracStages                                          2.645e+00
## PerfInterval_FT:factor(RSFracJobType)ENERGIZED                      2.216e+04
## PerfInterval_FT:factor(RSFracJobType)SLICKCROSSLINKHYB              1.221e+02
## PerfInterval_FT:factor(RSFracJobType)SLICKWATER                            NA
## factor(RSProppantType)SAND + RESIN:FrictionReducer_LBS              8.806e+00
## factor(RSProppantType)SAND ONLY:FrictionReducer_LBS                        NA
## factor(RSProppantType)SAND + RESIN:FluidIntensity_BBLPerFT          1.373e+04
## factor(RSProppantType)SAND ONLY:FluidIntensity_BBLPerFT                    NA
## factor(RSProppantType)SAND + RESIN:Surfactant_LBS                   5.576e+00
## factor(RSProppantType)SAND ONLY:Surfactant_LBS                             NA
## factor(RSProppantType)SAND + RESIN:FracStages                       6.548e+03
## factor(RSProppantType)SAND ONLY:FracStages                                 NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)ENERGIZED   1.804e+08
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)ENERGIZED             NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)SLICKCROSSLINKHYB  NA
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)SLICKCROSSLINKHYB     NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)SLICKWATER         NA
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)SLICKWATER            NA
## FrictionReducer_LBS:FluidIntensity_BBLPerFT                         7.307e-01
## FrictionReducer_LBS:Surfactant_LBS                                  5.264e-04
## FrictionReducer_LBS:FracStages                                      4.857e-01
## FrictionReducer_LBS:factor(RSFracJobType)ENERGIZED                  1.123e+04
## FrictionReducer_LBS:factor(RSFracJobType)SLICKCROSSLINKHYB                 NA
## FrictionReducer_LBS:factor(RSFracJobType)SLICKWATER                        NA
## FluidIntensity_BBLPerFT:Surfactant_LBS                              9.126e-01
## FluidIntensity_BBLPerFT:FracStages                                  9.048e+02
## FluidIntensity_BBLPerFT:factor(RSFracJobType)ENERGIZED              7.558e+06
## FluidIntensity_BBLPerFT:factor(RSFracJobType)SLICKCROSSLINKHYB             NA
## FluidIntensity_BBLPerFT:factor(RSFracJobType)SLICKWATER                    NA
## Surfactant_LBS:FracStages                                           4.236e-01
## Surfactant_LBS:factor(RSFracJobType)ENERGIZED                       7.726e+02
## Surfactant_LBS:factor(RSFracJobType)SLICKCROSSLINKHYB                      NA
## Surfactant_LBS:factor(RSFracJobType)SLICKWATER                             NA
## FracStages:factor(RSFracJobType)ENERGIZED                                  NA
## FracStages:factor(RSFracJobType)SLICKCROSSLINKHYB                          NA
## FracStages:factor(RSFracJobType)SLICKWATER                                 NA
##                                                                    t value
## (Intercept)                                                          0.600
## PerfInterval_FT                                                     -1.103
## factor(RSProppantType)SAND + RESIN                                   0.354
## factor(RSProppantType)SAND ONLY                                     -0.101
## FrictionReducer_LBS                                                  1.807
## FluidIntensity_BBLPerFT                                              0.006
## Surfactant_LBS                                                      -1.897
## FracStages                                                          -0.617
## factor(RSFracJobType)ENERGIZED                                      -1.149
## factor(RSFracJobType)SLICKCROSSLINKHYB                              -0.918
## factor(RSFracJobType)SLICKWATER                                      0.887
## PerfInterval_FT:factor(RSProppantType)SAND + RESIN                   1.129
## PerfInterval_FT:factor(RSProppantType)SAND ONLY                         NA
## PerfInterval_FT:FrictionReducer_LBS                                 -0.763
## PerfInterval_FT:FluidIntensity_BBLPerFT                              0.652
## PerfInterval_FT:Surfactant_LBS                                       1.823
## PerfInterval_FT:FracStages                                           0.881
## PerfInterval_FT:factor(RSFracJobType)ENERGIZED                      -1.084
## PerfInterval_FT:factor(RSFracJobType)SLICKCROSSLINKHYB               0.706
## PerfInterval_FT:factor(RSFracJobType)SLICKWATER                         NA
## factor(RSProppantType)SAND + RESIN:FrictionReducer_LBS               0.050
## factor(RSProppantType)SAND ONLY:FrictionReducer_LBS                     NA
## factor(RSProppantType)SAND + RESIN:FluidIntensity_BBLPerFT          -0.602
## factor(RSProppantType)SAND ONLY:FluidIntensity_BBLPerFT                 NA
## factor(RSProppantType)SAND + RESIN:Surfactant_LBS                    1.664
## factor(RSProppantType)SAND ONLY:Surfactant_LBS                          NA
## factor(RSProppantType)SAND + RESIN:FracStages                       -1.486
## factor(RSProppantType)SAND ONLY:FracStages                              NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)ENERGIZED    1.114
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)ENERGIZED          NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)SLICKCROSSLINKHYB    NA
```

```
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)SLICKCROSSLINKHYB         NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)SLICKWATER           NA
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)SLICKWATER              NA
## FrictionReducer_LBS:FluidIntensity_BBLPerFT                              -3.556
## FrictionReducer_LBS:Surfactant_LBS                                       -1.985
## FrictionReducer_LBS:FracStages                                            0.627
## FrictionReducer_LBS:factor(RSFracJobType)ENERGIZED                        1.105
## FrictionReducer_LBS:factor(RSFracJobType)SLICKCROSSLINKHYB                   NA
## FrictionReducer_LBS:factor(RSFracJobType)SLICKWATER                          NA
## FluidIntensity_BBLPerFT:Surfactant_LBS                                    2.417
## FluidIntensity_BBLPerFT:FracStages                                       -0.587
## FluidIntensity_BBLPerFT:factor(RSFracJobType)ENERGIZED                   -1.098
## FluidIntensity_BBLPerFT:factor(RSFracJobType)SLICKCROSSLINKHYB               NA
## FluidIntensity_BBLPerFT:factor(RSFracJobType)SLICKWATER                      NA
## Surfactant_LBS:FracStages                                                -0.969
## Surfactant_LBS:factor(RSFracJobType)ENERGIZED                             0.993
## Surfactant_LBS:factor(RSFracJobType)SLICKCROSSLINKHYB                        NA
## Surfactant_LBS:factor(RSFracJobType)SLICKWATER                               NA
## FracStages:factor(RSFracJobType)ENERGIZED                                    NA
## FracStages:factor(RSFracJobType)SLICKCROSSLINKHYB                            NA
## FracStages:factor(RSFracJobType)SLICKWATER                                   NA
##                                                                        Pr(>|t|)
## (Intercept)                                                            0.549328
## PerfInterval_FT                                                        0.272086
## factor(RSProppantType)SAND + RESIN                                     0.723597
## factor(RSProppantType)SAND ONLY                                        0.920087
## FrictionReducer_LBS                                                    0.073183
## FluidIntensity_BBLPerFT                                                0.995059
## Surfactant_LBS                                                         0.060172
## FracStages                                                             0.538544
## factor(RSFracJobType)ENERGIZED                                         0.252604
## factor(RSFracJobType)SLICKCROSSLINKHYB                                 0.360436
## factor(RSFracJobType)SLICKWATER                                        0.376970
## PerfInterval_FT:factor(RSProppantType)SAND + RESIN                     0.261178
## PerfInterval_FT:factor(RSProppantType)SAND ONLY                              NA
## PerfInterval_FT:FrictionReducer_LBS                                    0.446972
## PerfInterval_FT:FluidIntensity_BBLPerFT                                0.515676
## PerfInterval_FT:Surfactant_LBS                                         0.070656
## PerfInterval_FT:FracStages                                             0.379775
## PerfInterval_FT:factor(RSFracJobType)ENERGIZED                         0.280635
## PerfInterval_FT:factor(RSFracJobType)SLICKCROSSLINKHYB                 0.481750
## PerfInterval_FT:factor(RSFracJobType)SLICKWATER                              NA
## factor(RSProppantType)SAND + RESIN:FrictionReducer_LBS                 0.960114
## factor(RSProppantType)SAND ONLY:FrictionReducer_LBS                          NA
## factor(RSProppantType)SAND + RESIN:FluidIntensity_BBLPerFT             0.548487
## factor(RSProppantType)SAND ONLY:FluidIntensity_BBLPerFT                      NA
## factor(RSProppantType)SAND + RESIN:Surfactant_LBS                      0.098684
## factor(RSProppantType)SAND ONLY:Surfactant_LBS                               NA
## factor(RSProppantType)SAND + RESIN:FracStages                          0.139826
## factor(RSProppantType)SAND ONLY:FracStages                                   NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)ENERGIZED      0.267317
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)ENERGIZED               NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)SLICKCROSSLINKHYB    NA
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)SLICKCROSSLINKHYB       NA
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)SLICKWATER           NA
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)SLICKWATER              NA
## FrictionReducer_LBS:FluidIntensity_BBLPerFT                            0.000532
## FrictionReducer_LBS:Surfactant_LBS                                     0.049325
## FrictionReducer_LBS:FracStages                                         0.531850
## FrictionReducer_LBS:factor(RSFracJobType)ENERGIZED                     0.271271
## FrictionReducer_LBS:factor(RSFracJobType)SLICKCROSSLINKHYB                   NA
## FrictionReducer_LBS:factor(RSFracJobType)SLICKWATER                          NA
## FluidIntensity_BBLPerFT:Surfactant_LBS                                 0.017106
## FluidIntensity_BBLPerFT:FracStages                                     0.558198
## FluidIntensity_BBLPerFT:factor(RSFracJobType)ENERGIZED                 0.274401
## FluidIntensity_BBLPerFT:factor(RSFracJobType)SLICKCROSSLINKHYB               NA
## FluidIntensity_BBLPerFT:factor(RSFracJobType)SLICKWATER                      NA
## Surfactant_LBS:FracStages                                              0.334503
## Surfactant_LBS:factor(RSFracJobType)ENERGIZED                          0.322615
## Surfactant_LBS:factor(RSFracJobType)SLICKCROSSLINKHYB                        NA
## Surfactant_LBS:factor(RSFracJobType)SLICKWATER                               NA
## FracStages:factor(RSFracJobType)ENERGIZED                                    NA
## FracStages:factor(RSFracJobType)SLICKCROSSLINKHYB                            NA
## FracStages:factor(RSFracJobType)SLICKWATER                                   NA
##
```

```
## (Intercept)
## PerfInterval_FT
## factor(RSProppantType)SAND + RESIN
## factor(RSProppantType)SAND ONLY
## FrictionReducer_LBS                                                    .
## FluidIntensity_BBLPerFT
## Surfactant_LBS                                                         .
## FracStages
## factor(RSFracJobType)ENERGIZED
## factor(RSFracJobType)SLICKCROSSLINKHYB
## factor(RSFracJobType)SLICKWATER
## PerfInterval_FT:factor(RSProppantType)SAND + RESIN
## PerfInterval_FT:factor(RSProppantType)SAND ONLY
## PerfInterval_FT:FrictionReducer_LBS
## PerfInterval_FT:FluidIntensity_BBLPerFT
## PerfInterval_FT:Surfactant_LBS                                         .
## PerfInterval_FT:FracStages
## PerfInterval_FT:factor(RSFracJobType)ENERGIZED
## PerfInterval_FT:factor(RSFracJobType)SLICKCROSSLINKHYB
## PerfInterval_FT:factor(RSFracJobType)SLICKWATER
## factor(RSProppantType)SAND + RESIN:FrictionReducer_LBS
## factor(RSProppantType)SAND ONLY:FrictionReducer_LBS
## factor(RSProppantType)SAND + RESIN:FluidIntensity_BBLPerFT
## factor(RSProppantType)SAND ONLY:FluidIntensity_BBLPerFT
## factor(RSProppantType)SAND + RESIN:Surfactant_LBS                      .
## factor(RSProppantType)SAND ONLY:Surfactant_LBS
## factor(RSProppantType)SAND + RESIN:FracStages
## factor(RSProppantType)SAND ONLY:FracStages
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)ENERGIZED
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)ENERGIZED
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)SLICKCROSSLINKHYB
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)SLICKCROSSLINKHYB
## factor(RSProppantType)SAND + RESIN:factor(RSFracJobType)SLICKWATER
## factor(RSProppantType)SAND ONLY:factor(RSFracJobType)SLICKWATER
## FrictionReducer_LBS:FluidIntensity_BBLPerFT                            ***
## FrictionReducer_LBS:Surfactant_LBS                                     *
## FrictionReducer_LBS:FracStages
## FrictionReducer_LBS:factor(RSFracJobType)ENERGIZED
## FrictionReducer_LBS:factor(RSFracJobType)SLICKCROSSLINKHYB
## FrictionReducer_LBS:factor(RSFracJobType)SLICKWATER
## FluidIntensity_BBLPerFT:Surfactant_LBS                                 *
## FluidIntensity_BBLPerFT:FracStages
## FluidIntensity_BBLPerFT:factor(RSFracJobType)ENERGIZED
## FluidIntensity_BBLPerFT:factor(RSFracJobType)SLICKCROSSLINKHYB
## FluidIntensity_BBLPerFT:factor(RSFracJobType)SLICKWATER
## Surfactant_LBS:FracStages
## Surfactant_LBS:factor(RSFracJobType)ENERGIZED
## Surfactant_LBS:factor(RSFracJobType)SLICKCROSSLINKHYB
## Surfactant_LBS:factor(RSFracJobType)SLICKWATER
## FracStages:factor(RSFracJobType)ENERGIZED
## FracStages:factor(RSFracJobType)SLICKCROSSLINKHYB
## FracStages:factor(RSFracJobType)SLICKWATER
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 159200 on 125 degrees of freedom
##   (86 observations deleted due to missingness)
## Multiple R-squared:  0.5549, Adjusted R-squared:  0.4445
## F-statistic: 5.026 on 31 and 125 DF,  p-value: 4.443e-11
```

We will eliminate all interactive terms with P-vals greater than 0.8 then by 0.4 then 0.2 then by 0.1 then finally by 0.05. These values for filtering was chosen arbitrarily but it would allow us to quickly filter out the model for non-significance

```
#Remove all P-val>0.8
initint1 = lm(First36MonthProd_BOE~PerfInterval_FT+
               factor(RSProppantType)+
               FrictionReducer_LBS+
               FluidIntensity_BBLPerFT+
               Surfactant_LBS+
               FracStages+
               factor(RSFracJobType)+

               PerfInterval_FT*factor(RSProppantType)+
               PerfInterval_FT*FrictionReducer_LBS+
               PerfInterval_FT*FluidIntensity_BBLPerFT+
               PerfInterval_FT*Surfactant_LBS+
               PerfInterval_FT*FracStages+
               PerfInterval_FT*factor(RSFracJobType)+

               factor(RSProppantType)*FluidIntensity_BBLPerFT+
               factor(RSProppantType)*Surfactant_LBS+
               factor(RSProppantType)*FracStages+
               factor(RSProppantType)*factor(RSFracJobType)+

               FrictionReducer_LBS*FluidIntensity_BBLPerFT+
               FrictionReducer_LBS*Surfactant_LBS+
               FrictionReducer_LBS*FracStages+
               FrictionReducer_LBS*factor(RSFracJobType)+

               FluidIntensity_BBLPerFT*Surfactant_LBS+
               FluidIntensity_BBLPerFT*FracStages+
               FluidIntensity_BBLPerFT*factor(RSFracJobType)+

               Surfactant_LBS*FracStages+
               Surfactant_LBS*factor(RSFracJobType)+

               FracStages+factor(RSFracJobType)
             , data=data1)
summary(initint1)

#Remove all P-val>0.4
initint2 = lm(First36MonthProd_BOE~PerfInterval_FT+
               factor(RSProppantType)+
               FrictionReducer_LBS+
               FluidIntensity_BBLPerFT+
               Surfactant_LBS+
               FracStages+
               factor(RSFracJobType)+

               PerfInterval_FT*factor(RSProppantType)+
               PerfInterval_FT*FrictionReducer_LBS+
               PerfInterval_FT*FluidIntensity_BBLPerFT+
               PerfInterval_FT*Surfactant_LBS+
               PerfInterval_FT*factor(RSFracJobType)+

               factor(RSProppantType)*Surfactant_LBS+
               factor(RSProppantType)*FracStages+
               factor(RSProppantType)*factor(RSFracJobType)+

               FrictionReducer_LBS*FluidIntensity_BBLPerFT+
               FrictionReducer_LBS*Surfactant_LBS+
               FrictionReducer_LBS*factor(RSFracJobType)+

               FluidIntensity_BBLPerFT*Surfactant_LBS+
               FluidIntensity_BBLPerFT*factor(RSFracJobType)+

               Surfactant_LBS*FracStages+
               Surfactant_LBS*factor(RSFracJobType)+

               FracStages+factor(RSFracJobType)
             , data=data1)
summary(initint2)

#Remove all P-val>0.2
initint3 = lm(First36MonthProd_BOE~PerfInterval_FT+
               factor(RSProppantType)+
               FrictionReducer_LBS+
```

```
              FluidIntensity_BBLPerFT+
              Surfactant_LBS+
              FracStages+
              factor(RSFracJobType)+

              PerfInterval_FT*factor(RSProppantType)+
              PerfInterval_FT*Surfactant_LBS+

              factor(RSProppantType)*Surfactant_LBS+
              factor(RSProppantType)*FracStages+

              FrictionReducer_LBS*FluidIntensity_BBLPerFT+
              FrictionReducer_LBS*Surfactant_LBS+

              FluidIntensity_BBLPerFT*Surfactant_LBS+

              FracStages+factor(RSFracJobType)
            , data=data1)
summary(initint3)

#Remove all P-val>0.1
initint4 = lm(First36MonthProd_BOE~PerfInterval_FT+
              factor(RSProppantType)+
              FrictionReducer_LBS+
              FluidIntensity_BBLPerFT+
              Surfactant_LBS+
              FracStages+
              factor(RSFracJobType)+

              PerfInterval_FT*Surfactant_LBS+

              factor(RSProppantType)*Surfactant_LBS+
              factor(RSProppantType)*FracStages+

              FrictionReducer_LBS*FluidIntensity_BBLPerFT+
              FrictionReducer_LBS*Surfactant_LBS+

              FracStages+factor(RSFracJobType)
            , data=data1)
summary(initint4)

#Remove all P-val>0.05
initint5 = lm(First36MonthProd_BOE~PerfInterval_FT+
              factor(RSProppantType)+
              FrictionReducer_LBS+
              FluidIntensity_BBLPerFT+
              Surfactant_LBS+
              FracStages+
              factor(RSFracJobType)+

              PerfInterval_FT*Surfactant_LBS+

              factor(RSProppantType)*Surfactant_LBS+
              factor(RSProppantType)*FracStages+

              FrictionReducer_LBS*FluidIntensity_BBLPerFT+

              FracStages+factor(RSFracJobType)
            , data=data1)
summary(initint5)

#Remove all P-val>0.05
initint6 = lm(First36MonthProd_BOE~PerfInterval_FT+
              factor(RSProppantType)+
              FrictionReducer_LBS+
              FluidIntensity_BBLPerFT+
              Surfactant_LBS+
              FracStages+
              factor(RSFracJobType)+

              factor(RSProppantType)*Surfactant_LBS+
              factor(RSProppantType)*FracStages+

              FrictionReducer_LBS*FluidIntensity_BBLPerFT+
```

```
               FracStages+factor(RSFracJobType)
             , data=data1)
summary(initint6)

initint7 = lm(First36MonthProd_BOE~PerfInterval_FT+
               factor(RSProppantType)+
               FrictionReducer_LBS+
               FluidIntensity_BBLPerFT+
               Surfactant_LBS+
               FracStages+
               factor(RSFracJobType)+

               factor(RSProppantType)*FracStages+

               FrictionReducer_LBS*FluidIntensity_BBLPerFT+

               FracStages+factor(RSFracJobType)
             , data=data1)
summary(initint7)

initint8 = lm(First36MonthProd_BOE~PerfInterval_FT+
               factor(RSProppantType)+
               FrictionReducer_LBS+
               FluidIntensity_BBLPerFT+
               Surfactant_LBS+
               FracStages+

               factor(RSProppantType)*FracStages+

               FrictionReducer_LBS*FluidIntensity_BBLPerFT
             , data=data1)
summary(initint8)

initint9 = lm(First36MonthProd_BOE~PerfInterval_FT+
               factor(RSProppantType)+
               FrictionReducer_LBS+
               FluidIntensity_BBLPerFT+
               Surfactant_LBS+
               FracStages+

               FrictionReducer_LBS*FluidIntensity_BBLPerFT
             , data=data1)
summary(initint9)

initint10 = lm(First36MonthProd_BOE~PerfInterval_FT+
               FrictionReducer_LBS+
               FluidIntensity_BBLPerFT+
               Surfactant_LBS+
               FracStages+

               FrictionReducer_LBS*FluidIntensity_BBLPerFT
             , data=data1)
summary(initint10)
```

```
initint11 = lm(First36MonthProd_BOE~PerfInterval_FT+
               FrictionReducer_LBS+
               FluidIntensity_BBLPerFT+
               FracStages+

               FrictionReducer_LBS*FluidIntensity_BBLPerFT
             , data=data1)
summary(initint11)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + FrictionReducer_LBS +
##     FluidIntensity_BBLPerFT + FracStages + FrictionReducer_LBS *
##     FluidIntensity_BBLPerFT, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -361288 -129287  -45680  108553  590430
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                -1.666e+05  1.027e+05  -1.622
## PerfInterval_FT                             7.372e+01  1.313e+01   5.616
## FrictionReducer_LBS                         1.659e+01  4.306e+00   3.854
## FluidIntensity_BBLPerFT                     1.826e+04  6.665e+03   2.739
## FracStages                                 -4.593e+03  2.309e+03  -1.989
## FrictionReducer_LBS:FluidIntensity_BBLPerFT -1.164e+00  3.741e-01  -3.111
##                                            Pr(>|t|)
## (Intercept)                                0.106587
## PerfInterval_FT                             7.6e-08 ***
## FrictionReducer_LBS                        0.000164 ***
## FluidIntensity_BBLPerFT                    0.006803 **
## FracStages                                 0.048253 *
## FrictionReducer_LBS:FluidIntensity_BBLPerFT 0.002178 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 187000 on 174 degrees of freedom
##   (63 observations deleted due to missingness)
## Multiple R-squared:  0.266,  Adjusted R-squared:  0.2449
## F-statistic: 12.61 on 5 and 174 DF,  p-value: 1.857e-10
```

All predictors are significant. R-squared Adj=0.2449 and RMSE=187000

```
coef(initint11)
```

```
##                                 (Intercept)
##                               -1.666365e+05
##                             PerfInterval_FT
##                                7.371738e+01
##                         FrictionReducer_LBS
##                                1.659353e+01
##                     FluidIntensity_BBLPerFT
##                                1.825567e+04
##                                  FracStages
##                               -4.592856e+03
## FrictionReducer_LBS:FluidIntensity_BBLPerFT
##                               -1.163774e+00
```

$$\begin{aligned}First36\widehat{Month}Prod_{BOE} = &-166636.5 + 73.71738 \cdot PerfIntervalFT \\ &+ 16.59353 \cdot FrictionReducerLBS + 18255.67 \cdot FluidIntensityBBLPerFT - 4592.856 \cdot FracStages \\ &- 1.163774 \cdot rictionReducerLBS \cdot FluidIntensityBBLPerFT\end{aligned}$$

We will save the interpretation for later as we will find out that this model does not meet some of the assumptions. More work needs to be done and interpretation will be carried out with the final resulting model.

---

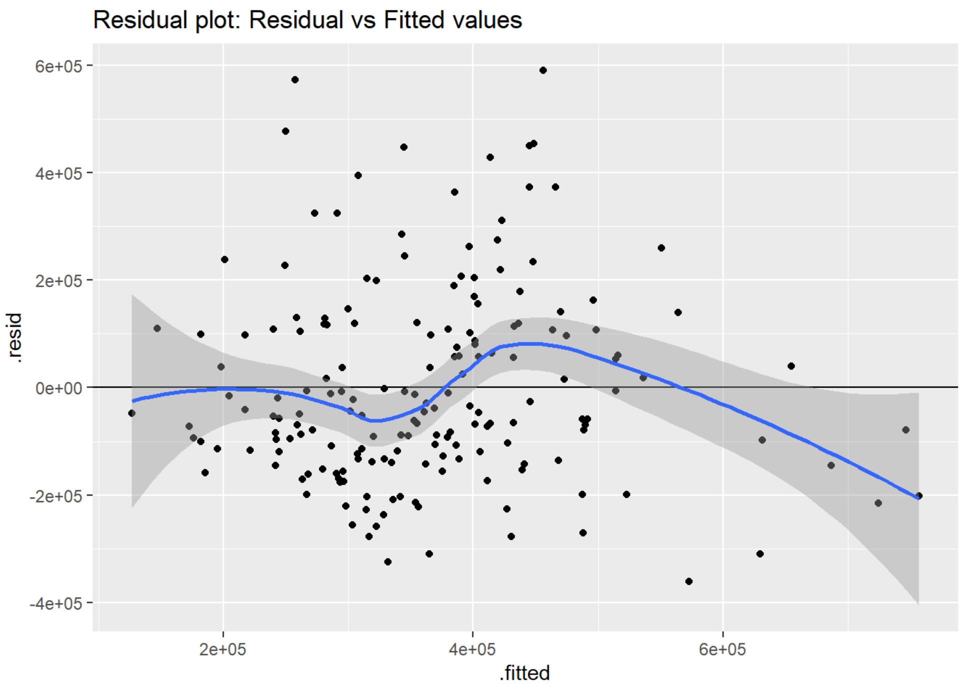
3. Testing the Assumptions of the interactive model.

a) Linearity and Homoscedasticity

First let's look at whether this regression model is linear and homoscedastic using Residual vs fitted values plot and the studentized Breusch-Pagan test.
Using the studentized Breusch-Pagan test Let:
H_0 : homoscedasticity is present
H_a : heteroscedasticity is present

```
ggplot(initint11, aes(x=.fitted, y=.resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth()+
  ggtitle("Residual plot: Residual vs Fitted values")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
bptest(initint11)
```

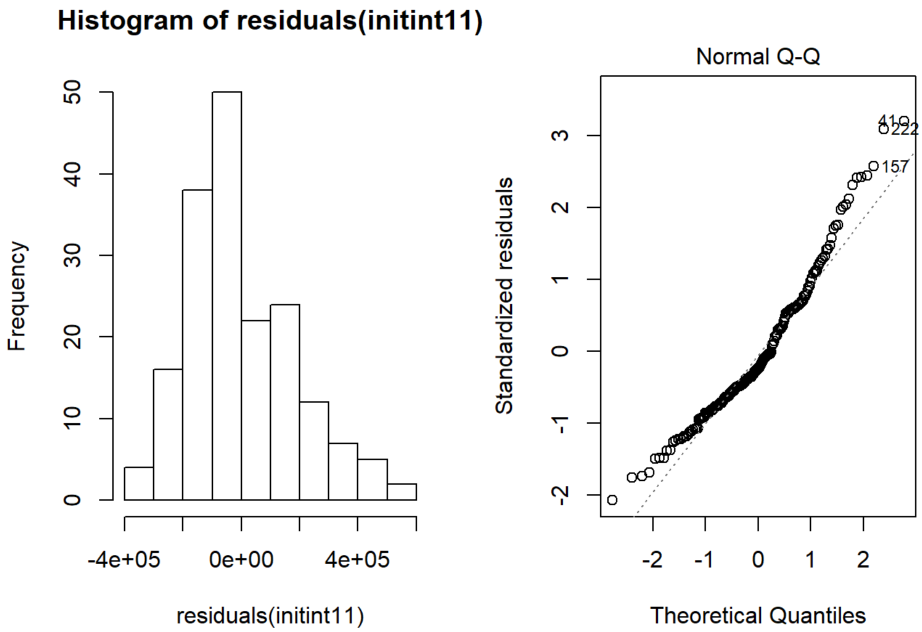
```
##
##  studentized Breusch-Pagan test
##
## data:  initint11
## BP = 6.4169, df = 5, p-value = 0.2677
```

The model roughly follows the linear trend except for the right tail end where it seems to diverge. The studentized Breusch-Pagan test yielded a P-value of 0.2677 which means that we can accept the null hypothesis and conclude that our model is largely homoscedastic.

b) Testing for Normality

Here we will plot the histogram and the Q-Q plot, and in addtion we will use the Shapiro-Wilk normality test to see if the data is in fact normal.
for the Shapiro-Wilk normality test let:
H_0 : The sample data is significantly normally distributed
H_a : The sample data is not significantly normally distributed

```
par(mfrow=c(1,2))
hist(residuals(initint11))
plot(initint11, which=2)
```




```
shapiro.test(residuals(initint11))
```

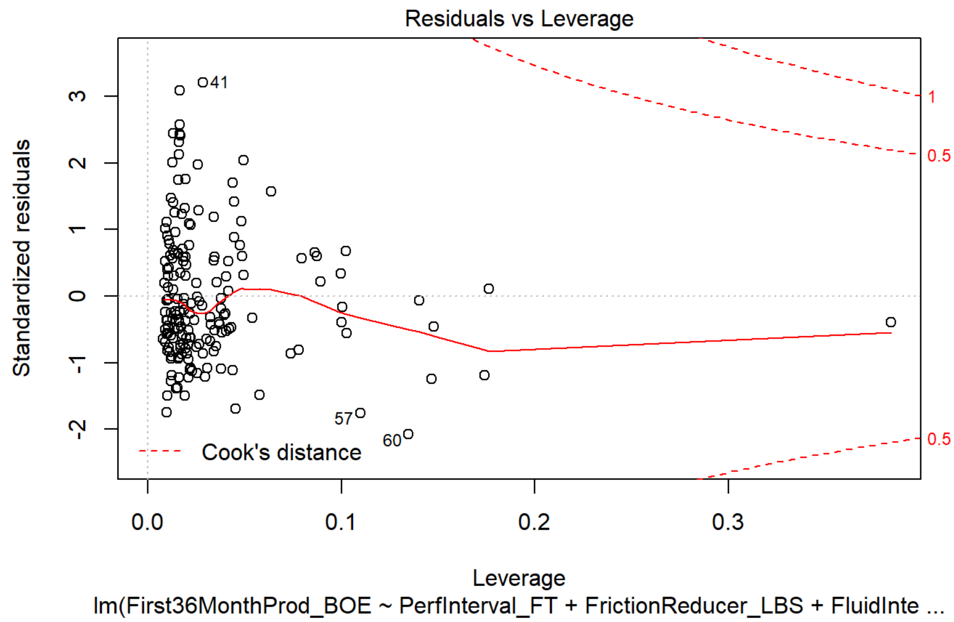
```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(initint11)
## W = 0.95317, p-value = 1.119e-05
```

It is quite apparent in both the histrogram and the Q-Q plot that the data is not normal. The sample data deviates quite strongly for the Q-Q plot and to confirm we have a P-value less than 0.05 using the Shapiro-Wilk normality test. This means we reject the null and conclude that the data is likely not normal. It's apparent that despite the model being significant, we see that this interactive model does not meet the assumption of normality. Therefore in a later section, we will do transformations to the dependent variable (**First36MonthProd_BOE**) to see if we can normalize the model.

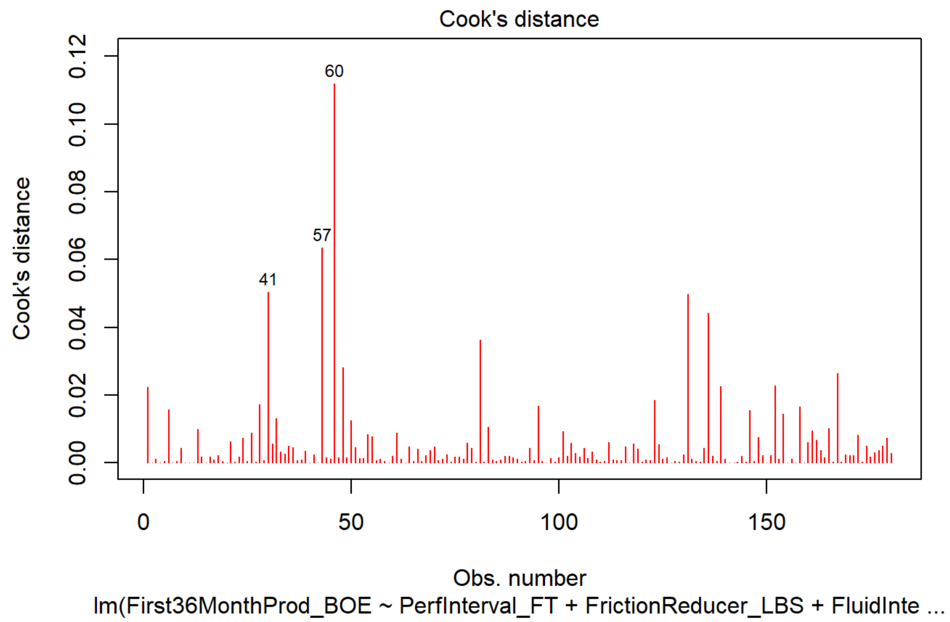
c) Looking for Outliers

Here we will utilize Residuals vs Leverage plot with Cook's distance in order to see if there are any outliers. We will also show another plot of Cook's distance vs obs.number as well for clarity.

```
plot(initint11,which=5)
```




```
plot(initint11,pch=18,col="red",which=c(4))
```



We can clearly see that there are no point with a Cook's distance greater than 0.5. So we can determine that there are no outliers that will influence our model.

---

4. Transformations of the Dependent Variable

We will limit to two transformations of **First36MonthProd_BOE** to see if we can normalize the data while still preserving the other assumptions. We will compare $\sqrt{First36MonthProd_{BOE}}$ and $\log\left(First36MonthProd_{BOE}\right)$.

I will not repeatedly list the hypothesis again as the analysis going forward will be repetitive.

a) Let's get the $\sqrt{First36MonthProd_{BOE}}$

```
sqrty = lm(sqrt(First36MonthProd_BOE)~PerfInterval_FT+
               FrictionReducer_LBS+
               FluidIntensity_BBLPerFT+
               FracStages+

               FrictionReducer_LBS*FluidIntensity_BBLPerFT
             , data=data1)

summary(sqrty)
```

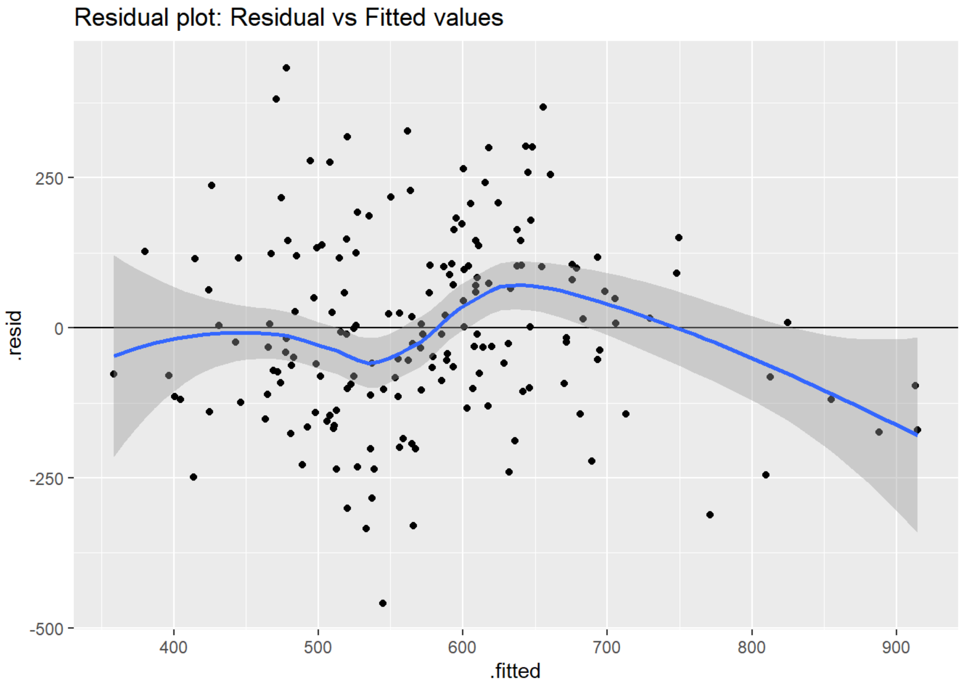
```
##
## Call:
## lm(formula = sqrt(First36MonthProd_BOE) ~ PerfInterval_FT + FrictionReducer_LBS +
##     FluidIntensity_BBLPerFT + FracStages + FrictionReducer_LBS *
##     FluidIntensity_BBLPerFT, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -458.43 -102.97  -11.03  104.71  433.30
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                 1.026e+02  8.727e+01   1.176
## PerfInterval_FT                             6.496e-02  1.115e-02   5.825
## FrictionReducer_LBS                         1.491e-02  3.658e-03   4.075
## FluidIntensity_BBLPerFT                     1.765e+01  5.662e+00   3.118
## FracStages                                 -4.575e+00  1.961e+00  -2.333
## FrictionReducer_LBS:FluidIntensity_BBLPerFT -1.057e-03  3.178e-04  -3.327
##                                            Pr(>|t|)
## (Intercept)                                 0.24117
## PerfInterval_FT                            2.70e-08 ***
## FrictionReducer_LBS                        6.98e-05 ***
## FluidIntensity_BBLPerFT                     0.00213 **
## FracStages                                  0.02082 *
## FrictionReducer_LBS:FluidIntensity_BBLPerFT  0.00107 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 158.8 on 174 degrees of freedom
##   (63 observations deleted due to missingness)
## Multiple R-squared:  0.2834, Adjusted R-squared:  0.2628
## F-statistic: 13.76 on 5 and 174 DF,  p-value: 2.515e-11
```

Already we see a higher R^2_{adj} of 0.2628 while the interactive model without transformations has R^2_{adj} of 0.2449. Comparing RMSE we see an RMSE of 158.8 in the transformed model which is significanlty lower than RMSE of 187000 which is seen in the basic interactive model without the transformation. Now let's look to see if the homoscedasticity and normality assumptions are met along with looking for outliers if any.

Let's look at whether this regression model is linear and homoscedastic using Residual vs fitted values plot and the studentized Breusch-Pagan test:

```
ggplot(sqrty, aes(x=.fitted, y=.resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth()+
  ggtitle("Residual plot: Residual vs Fitted values")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

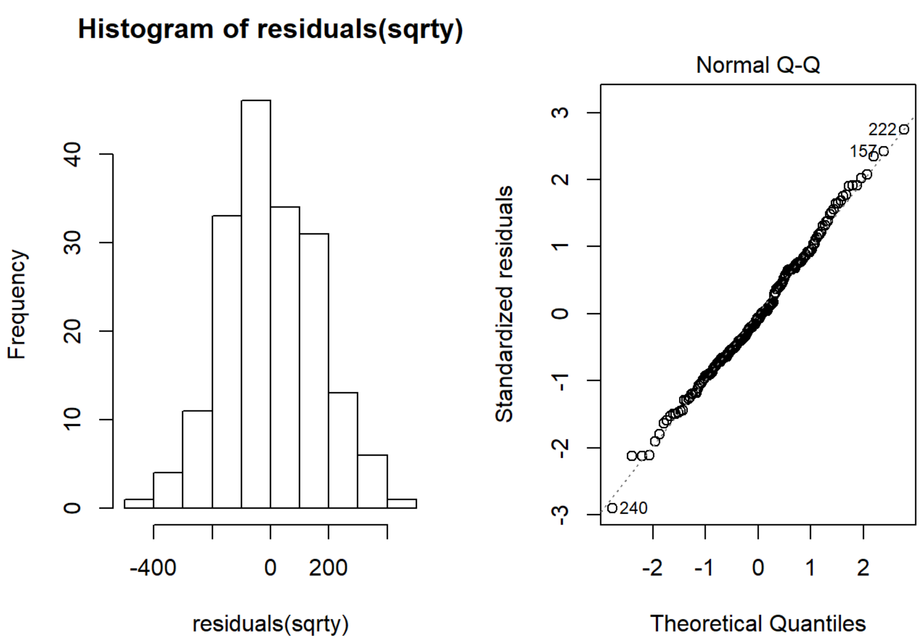



```
bptest(sqrty)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  sqrty
## BP = 4.818, df = 5, p-value = 0.4385
```

Again we see that this is sort of linear but breaks down a little near the right ends. The studentized Breusch-Pagan test reports a P-value of 0.4385 which means we accept the null and conclude that our model is largely homoscedastic.
Here we will plot the histogram and the Q-Q plot, and in addition we will use the Shapiro-Wilk normality test to see if the data is in fact normal:

```
par(mfrow=c(1,2))
hist(residuals(sqrty))
plot(sqrty, which=2)
```



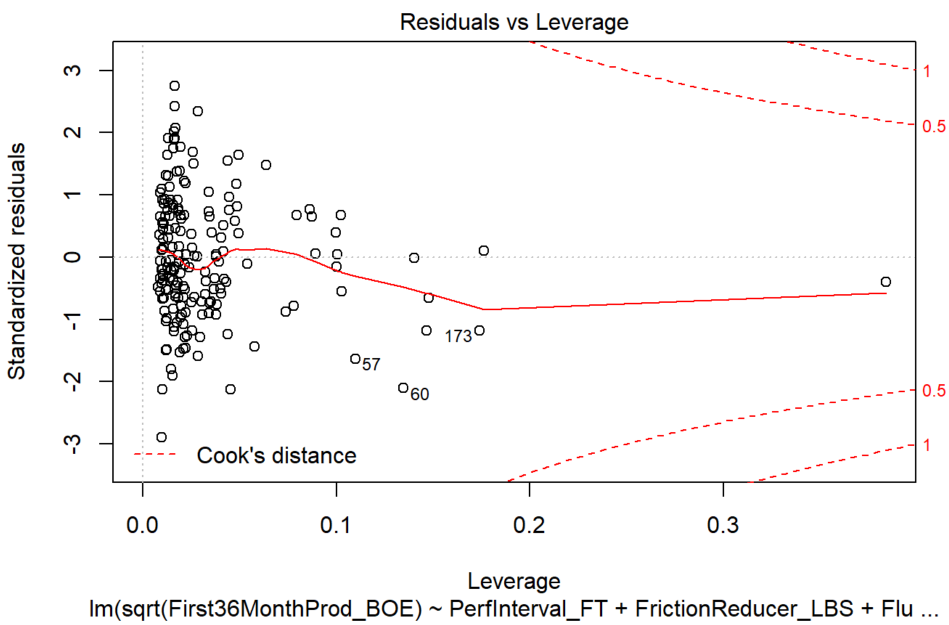

```
shapiro.test(residuals(sqrty))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(sqrty)
## W = 0.99475, p-value = 0.7783
```

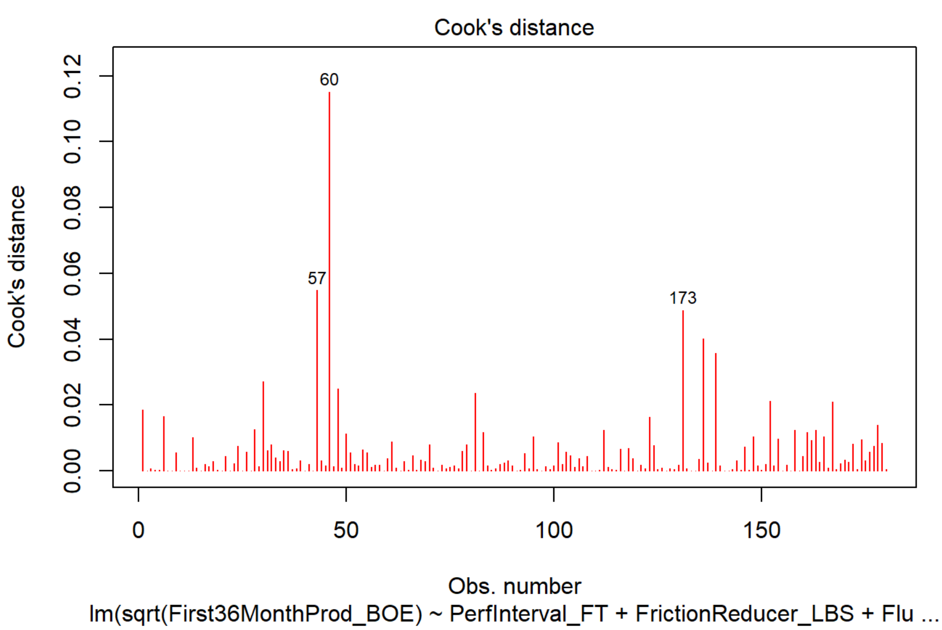
Both the historgram and the Q-Q plot indicates that the sample data is largely normal. To confirm, Shapiro-Wilk normality test reports a P-Value of 0.7783 which means we accept the null and conclude that our data is likely normal.

Here we will utilize Residuals vs Leverage plot with Cook's distance in order to see if there are any outliers. We will also show another plot of Cook's distance vs obs.number as well for clarity.

```
plot(sqrty,which=5)
```




```
plot(sqrty,pch=18,col="red",which=c(4))
```



b) $\log\left(First36MonthProd_{BOE}\right)$

```
logy = lm(log(First36MonthProd_BOE)~PerfInterval_FT+
               FrictionReducer_LBS+
               FluidIntensity_BBLPerFT+
               FracStages+

               FrictionReducer_LBS*FluidIntensity_BBLPerFT
             , data=data1)

summary(logy)
```

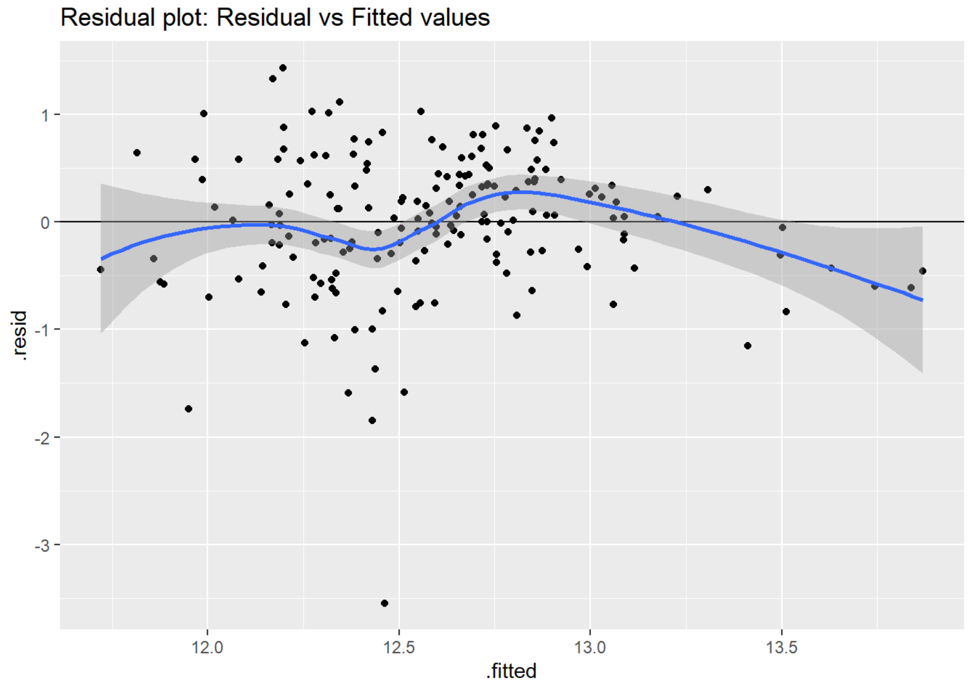
```
##
## Call:
## lm(formula = log(First36MonthProd_BOE) ~ PerfInterval_FT + FrictionReducer_LBS +
##     FluidIntensity_BBLPerFT + FracStages + FrictionReducer_LBS *
##     FluidIntensity_BBLPerFT, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5442 -0.3417  0.0504  0.4190  1.4334
##
## Coefficients:
##                                               Estimate Std. Error t value
## (Intercept)                                  1.073e+01  3.603e-01  29.798
## PerfInterval_FT                              2.522e-04  4.603e-05   5.479
## FrictionReducer_LBS                          5.761e-05  1.510e-05   3.815
## FluidIntensity_BBLPerFT                      7.551e-02  2.337e-02   3.231
## FracStages                                  -1.988e-02  8.097e-03  -2.455
## FrictionReducer_LBS:FluidIntensity_BBLPerFT -4.185e-06  1.312e-06  -3.190
##                                             Pr(>|t|)
## (Intercept)                                  < 2e-16 ***
## PerfInterval_FT                             1.48e-07 ***
## FrictionReducer_LBS                         0.000189 ***
## FluidIntensity_BBLPerFT                     0.001478 **
## FracStages                                  0.015078 *
## FrictionReducer_LBS:FluidIntensity_BBLPerFT 0.001686 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6557 on 174 degrees of freedom
##   (63 observations deleted due to missingness)
## Multiple R-squared:  0.259,  Adjusted R-squared:  0.2377
## F-statistic: 12.16 on 5 and 174 DF,  p-value: 4.074e-10
```

This model has an R^2_{adj} 0.2377 and an RMSE of 0.6557. The R^2_{adj} is slightly lower than the sqrtroot transformation, the log model is superior in terms of the RMSE where the value is only 0.5978. However, we will see that this model does not hold up in one of the assumptions shown in the following.

Let's look at whether this regression model is linear and homoscedastic using Residual vs fitted values plot and the studentized Breusch-Pagan test:

```
ggplot(logy, aes(x=.fitted, y=.resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_smooth()+
  ggtitle("Residual plot: Residual vs Fitted values")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



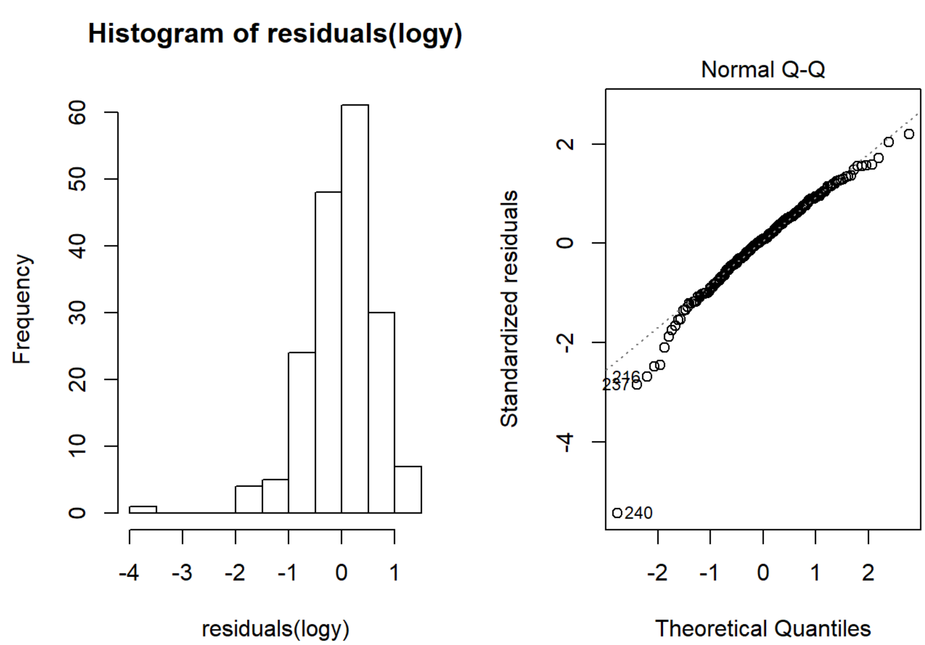

```
bptest(logy)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  logy
## BP = 3.3075, df = 5, p-value = 0.6527
```

The model looks the same in terms of linearity as the other two and the resulting P-value from the studentized Breusch-Pagan test shows that the data is largely homoscedastic.

Here we will plot the histogram and the Q-Q plot, and in addition we will use the Shapiro-Wilk normality test to see if the data is in fact normal:

```
par(mfrow=c(1,2))
hist(residuals(logy))
plot(logy, which=2)
```



```
shapiro.test(residuals(logy))
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  residuals(logy)
## W = 0.94118, p-value = 9.625e-07
```

We have a p-value less than 0.05 so we reject the null and we conclude that the sample data points are not normally distributed.

---

5. Best Model and interpretation

We can now conclude that the interactive model with the sqrt transformation applied is the best model. The sqrt transformed model meets the assumptions of homoscedasticity (P-val, 0.4385 >0.05) and normality (P-val, 0.7783>0.05) while the regular interactive model violates the normality assumption (P-val < 0.05). Comparing the sqrt transformed interactive model with the regular interactive model, we see that the transformed model has a higher R^2_{adj} of 0.2628 vs the 0.2449 of the regular interactive. In addition we see a much lower RMSE of 158.8 than the RMSE for the regular interactive model which is 187000.

The model is shown below:

```
coef(sqrty)
```

```
##                               (Intercept) 
##                             102.632502189 
##                           PerfInterval_FT 
##                               0.064956064 
##                       FrictionReducer_LBS 
##                               0.014906841 
##                  FluidIntensity_BBLPerFT 
##                              17.654444701 
##                                FracStages 
##                              -4.575202101 
## FrictionReducer_LBS:FluidIntensity_BBLPerFT 
##                              -0.001057213
```

$$\begin{aligned}\sqrt{First36\widehat{Month}Prod_{BOE}} =& 102.632502189 + 0.064956064 \cdot PerfInterval_{FT} \\ & + 0.014906841 \cdot FrictionReducer_{LBS} + 17.654444701 \cdot FluidIntensity_{BBLPerFT} - 4.575202101 \cdot FracS \\ & - 0.001057213 \cdot FrictionReducer_{LBS} \cdot FluidIntensity_{BBLPerFT}\end{aligned}$$

a) Interpretation

When every other variable is controlled for:
Perforation increase of 1 ft will increase the $\sqrt{First36MonthProd_{BOE}}$ by 0.064956064.
Friction Reducer increase of 1 lbs will increase the $\sqrt{First36MonthProd_{BOE}}$ by $ 0.013849628-0.001057213 *FluidIntensity_{BBLPerFT}$.
Fluid Intensity increase of 1 bbl/ft will increase the $\sqrt{First36MonthProd_{BOE}}$ by $17.653387488, -0.001057213 * FrictionReducer_{LBS}$.
Frac Stages increase of 1 will affect the $\sqrt{First36MonthProd_{BOE}}$ by -4.575202101

Here we are exploring different possibilities of predictors to be in the model for getting better production model. It also checks for quadratic, cubic, higher dimensions and binary dependent models if it improve our model. Due to scope of this project we did not come to any conclusion from the models we only provided a brief description/graphical representation of what the model entails.

```
#data = read.csv("wELLS.csv", header=TRUE)
#data = WELLS[-131,]
#calculating the 95th confidence level
data=data1
WELLS=data1
reg1<-lm(First36MonthProd_BOE~SpudToSales_DAYS+PerfInterval_FT+AverageStageSpacing_FT+ProppantLoading_LBSPerGAL+ProppantInte
nsity_LBSPerFT+FluidIntensity_BBLPerFT+FrictionReducer_LBS+Surfactant_LBS, data=data)
confint(reg1, level=.99)
```

```
##                                     0.5 %       99.5 %
## (Intercept)                -2.179041e+05 6.117953e+05
## SpudToSales_DAYS           -6.439147e+02 3.811827e+02
## PerfInterval_FT            -1.489546e+01 8.028063e+01
## AverageStageSpacing_FT     -4.850415e+01 1.182458e+03
## ProppantLoading_LBSPerGAL  -3.559704e+05 4.382608e+04
## ProppantIntensity_LBSPerFT -1.822431e+02 7.659021e+02
## FluidIntensity_BBLPerFT    -3.885345e+04 7.180037e+03
## FrictionReducer_LBS        -2.141721e+00 9.708161e+00
## Surfactant_LBS             -2.944920e+00 9.578919e+00
```

```
summary(reg1)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ SpudToSales_DAYS + PerfInterval_FT +
##     AverageStageSpacing_FT + ProppantLoading_LBSPerGAL + ProppantIntensity_LBSPerFT +
##     FluidIntensity_BBLPerFT + FrictionReducer_LBS + Surfactant_LBS,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -466860 -133928  -43545  109256  615306
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.969e+05  1.588e+05   1.240   0.2171
## SpudToSales_DAYS           -1.314e+02  1.963e+02  -0.669   0.5044
## PerfInterval_FT             3.269e+01  1.822e+01   1.794   0.0750 .
## AverageStageSpacing_FT      5.670e+02  2.357e+02   2.406   0.0175 *
## ProppantLoading_LBSPerGAL  -1.561e+05  7.654e+04  -2.039   0.0433 *
## ProppantIntensity_LBSPerFT  2.918e+02  1.815e+02   1.608   0.1102
## FluidIntensity_BBLPerFT    -1.584e+04  8.813e+03  -1.797   0.0745 .
## FrictionReducer_LBS         3.783e+00  2.269e+00   1.668   0.0976 .
## Surfactant_LBS              3.317e+00  2.398e+00   1.383   0.1688
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 196200 on 139 degrees of freedom
##   (95 observations deleted due to missingness)
## Multiple R-squared:  0.2342, Adjusted R-squared:  0.1901
## F-statistic: 5.313 on 8 and 139 DF,  p-value: 7.978e-06
```

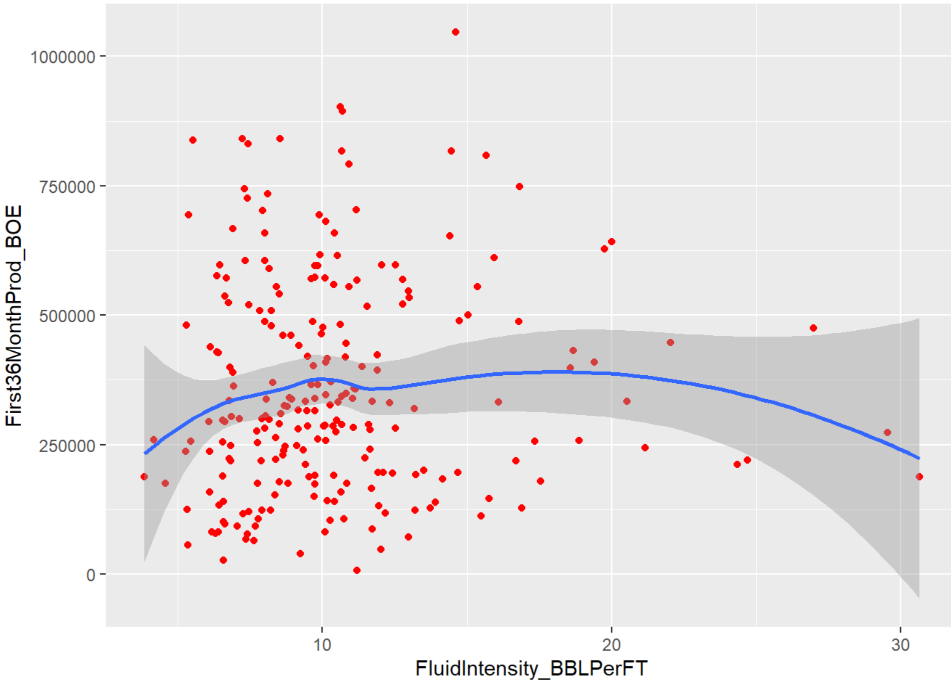
```
# This is for a simple linear model output when we are considering the "y" for the equation/model to be First36MonthProd_BOE
```

```
ggplot(data=WELLS,mapping= aes(x=FluidIntensity_BBLPerFT,y=First36MonthProd_BOE))+geom_point(color='red')+geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 7 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```




```
simplemodel=lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT+FrictionReducer_LBS,data=WELLS)
summary(simplemodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT +
##     FrictionReducer_LBS, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -365415 -145622  -49745  132476  675380
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            325081.535  40799.889   7.968 1.90e-13 ***
## FluidIntensity_BBLPerFT  -3236.102   3711.959  -0.872    0.384
## FrictionReducer_LBS          8.211      1.877   4.374 2.08e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 205300 on 177 degrees of freedom
##   (63 observations deleted due to missingness)
## Multiple R-squared:  0.09992,    Adjusted R-squared:  0.08975
## F-statistic: 9.825 on 2 and 177 DF,  p-value: 8.993e-05
```

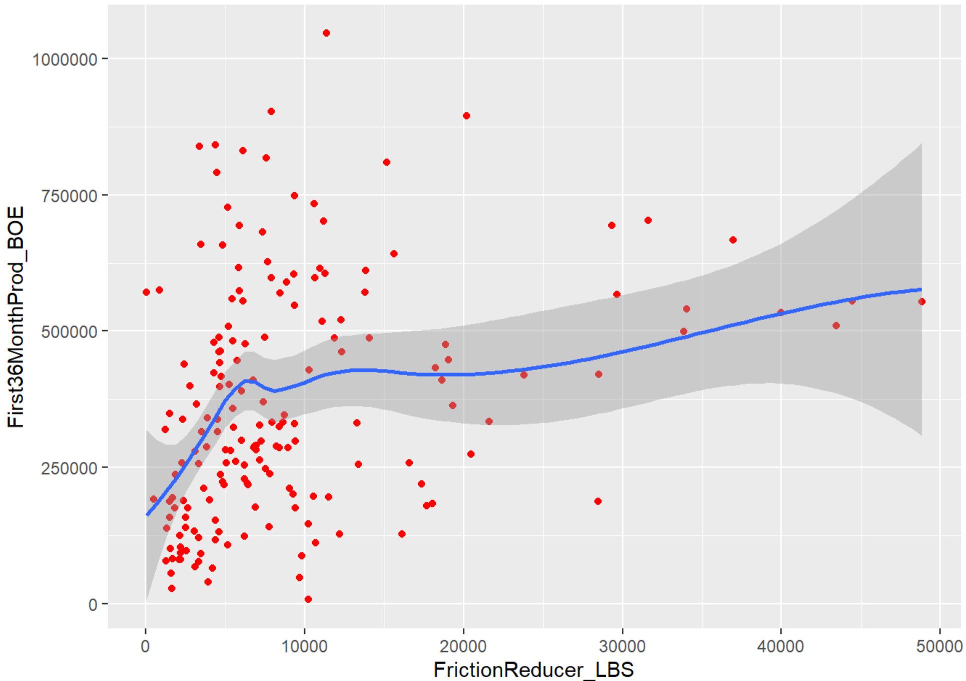
```
#This code here is for demonstrating the relation between Fluid Intensity , Friction Reducer and First 36 Month Prodcution o
f BOE. Along with the ggplot there is a summary provided that provides key information. As we can see the p-value and  F-sta
tistic have decent values but the R Adjusted squared indicates a poor value along with the RMSE value. This means that only
 a limited amount of information is being provided between these two variables.
```

```
ggplot(data=WELLS,mapping= aes(x=FrictionReducer_LBS,y=First36MonthProd_BOE))+geom_point(color='red')+geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 62 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 62 rows containing missing values (geom_point).
```




```
simplemodel=lm(First36MonthProd_BOE~FrictionReducer_LBS+ProppantIntensity_LBSPerFT,data=WELLS)
summary(simplemodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FrictionReducer_LBS + ProppantIntensity_LBSPerFT,
##     data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -371094 -149615  -48281  116700  682423
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                360796.058  43478.254   8.298 2.55e-14 ***
## FrictionReducer_LBS             8.181      1.764   4.637 6.81e-06 ***
## ProppantIntensity_LBSPerFT   -115.218     65.201  -1.767   0.0789 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 203500 on 178 degrees of freedom
##   (62 observations deleted due to missingness)
## Multiple R-squared:  0.1123, Adjusted R-squared:  0.1024
## F-statistic: 11.26 on 2 and 178 DF,  p-value: 2.478e-05
```

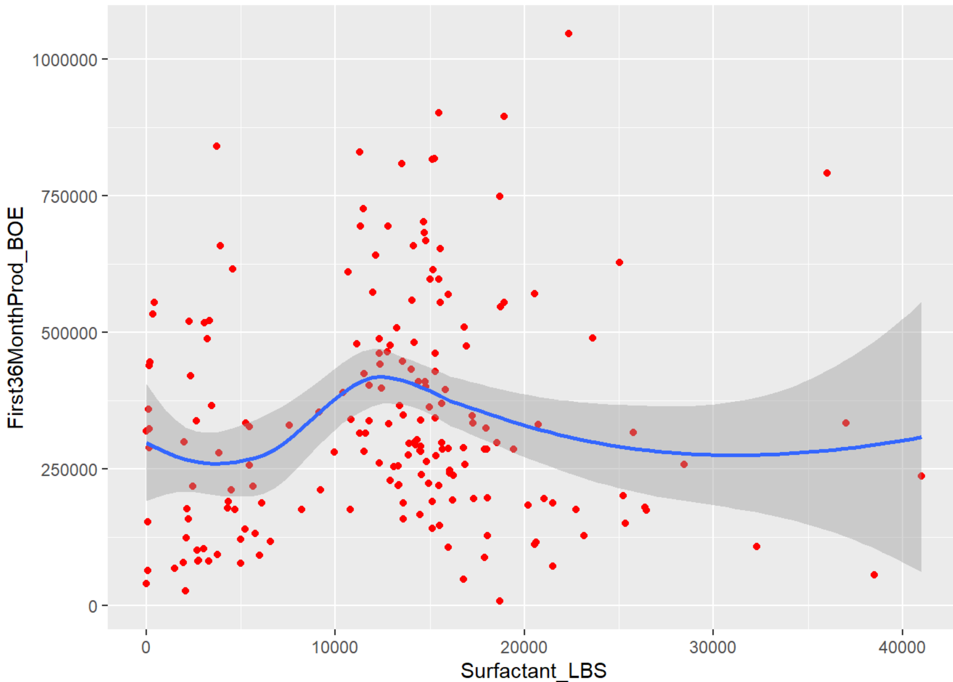
```
#This code here is for demonstrating the relation between Friction Reducer, Propant Intentsity and First 36 Month Prodcution
of BOE. Along with the ggplot there is a summary provided that provides key information. As we can see the p-value and  F-st
atistic have decent values but the R Adjusted squared indicates a poor value along with the RMSE value. This means that only
a limited amount of information is being provided between these two variables. Therefore the model shouldnt be accepted in i
ts entirety.
```

```
ggplot(data=WELLS,mapping= aes(x=Surfactant_LBS,y=First36MonthProd_BOE))+geom_point(color='red')+geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 53 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 53 rows containing missing values (geom_point).
```




```
simplemodel=lm(First36MonthProd_BOE~AverageStageSpacing_FT+Surfactant_LBS,data=WELLS)
summary(simplemodel)
```

```
## 
## Call:
## lm(formula = First36MonthProd_BOE ~ AverageStageSpacing_FT + 
##     Surfactant_LBS, data = WELLS)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -378897 -138331  -44455   99751  702091 
## 
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            1.406e+05  5.304e+04   2.651  0.00873 ** 
## AverageStageSpacing_FT 6.303e+02  1.636e+02   3.853  0.00016 ***
## Surfactant_LBS         3.055e+00  1.882e+00   1.623  0.10620    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 198100 on 187 degrees of freedom
##   (53 observations deleted due to missingness)
## Multiple R-squared:  0.08064,    Adjusted R-squared:  0.07081 
## F-statistic: 8.201 on 2 and 187 DF,  p-value: 0.0003854
```

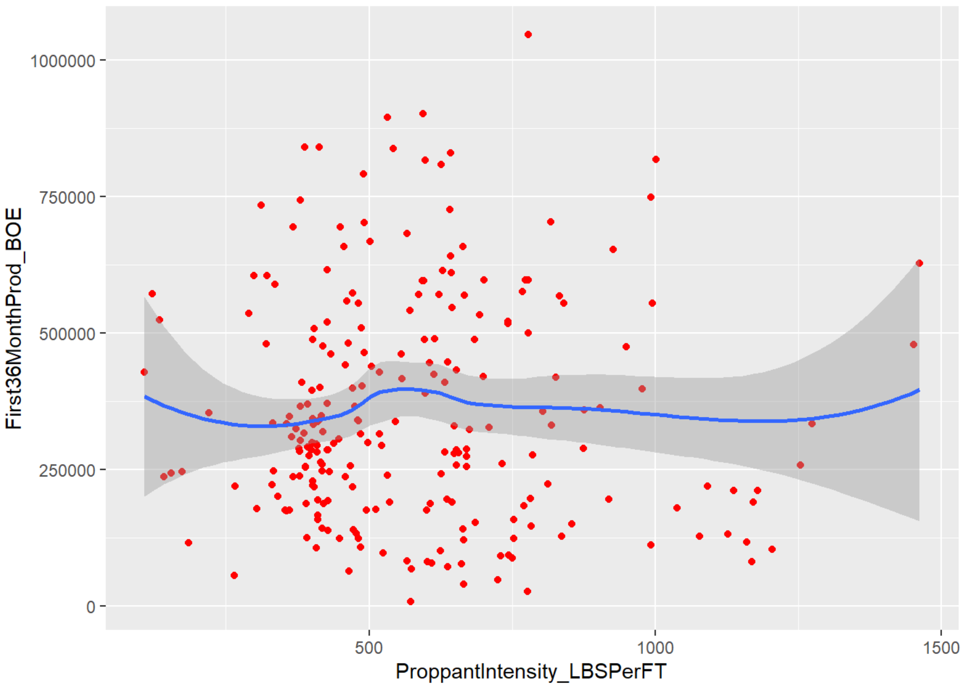
```
#This code here is for demonstrating the relation between Average Stage Spacing and Surfactant and First 36 Month Prodcution of BOE. Along with the ggplot there is a summary provided that provides key information. As we can see the p-value and  F-statistic have decent values but the R Adjusted squared indicates a poor value along with the RMSE value. This means that only a limited amount of information is being provided between these two variables. Therefore the model shouldnt be accepted in its entirety.
```

```
ggplot(data=WELLS,mapping= aes(x=ProppantIntensity_LBSPerFT,y=First36MonthProd_BOE))+geom_point(color='red')+geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 3 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```




```
simplemodel=lm(First36MonthProd_BOE~PerfInterval_FT+ProppantIntensity_LBSPerFT,data=WELLS)
summary(simplemodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT + ProppantIntensity_LBSPerFT,
##     data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -384027 -140514  -39310  120740  622835
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                41760.92   65682.68   0.636    0.526
## PerfInterval_FT               49.51       8.78   5.639 4.84e-08 ***
## ProppantIntensity_LBSPerFT    52.65      54.13   0.973    0.332
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 194100 on 237 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.1184, Adjusted R-squared:  0.111
## F-statistic: 15.92 on 2 and 237 DF,  p-value: 3.251e-07
```

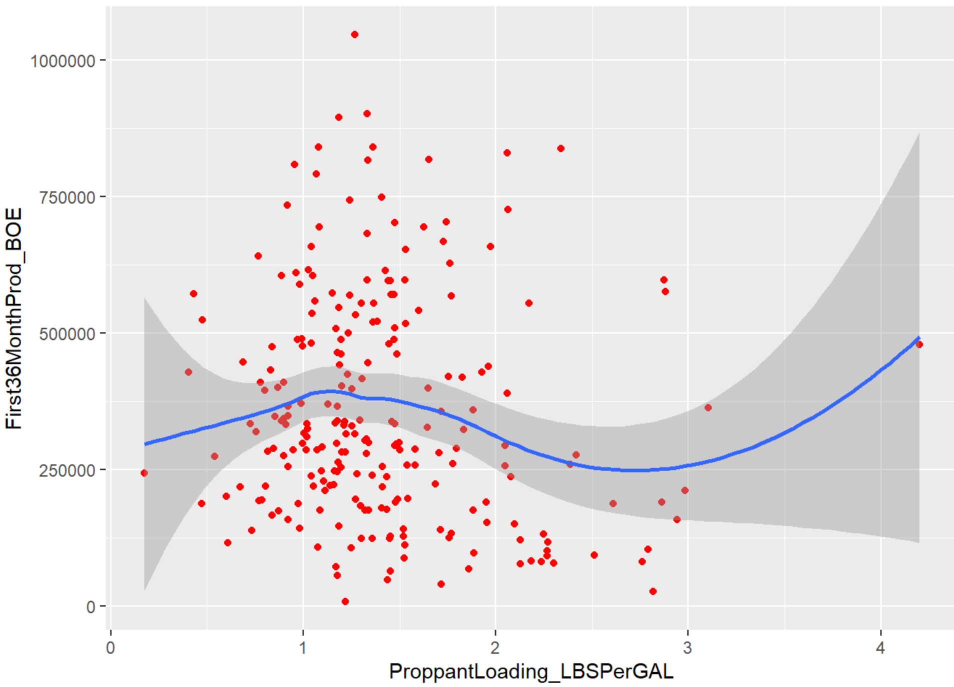
```
#This code here is for demonstrating the relation between Perf Inerval, Proppant Intensity and First 36 Month Prodcution of
 BOE. Along with the ggplot there is a summary provided that provides key information. As we can see the p-value and  F-stat
istic have good values but the R Adjusted squared indicates a poor value along with the RMSE value. This means that only a l
imited amount of information is being provided between these two variables. Therefore the model shouldnt be accepted in its
 entiret and another model should be found. These variables were chosen because the Perf Interval along with the Proppant In
tensity would have an influence over the prodiction rate.
```

```
ggplot(data=WELLS,mapping= aes(x=ProppantLoading_LBSPerGAL,y=First36MonthProd_BOE))+geom_point(color='red')+geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 7 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```



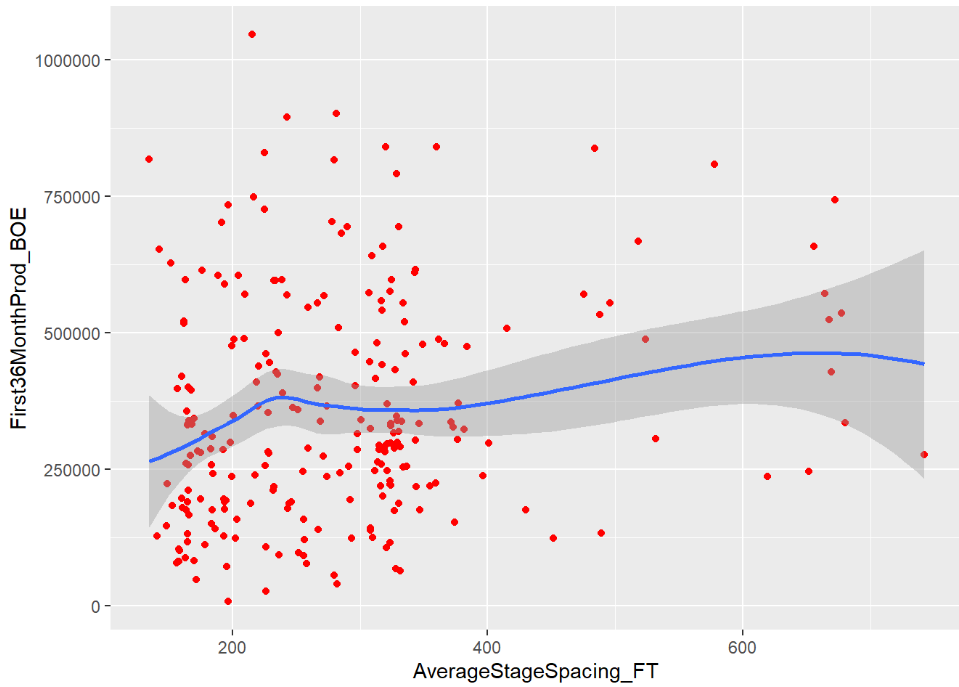

```
simplemodel=lm(First36MonthProd_BOE~Surfactant_LBS+ProppantLoading_LBSPerGAL,data=WELLS)
summary(simplemodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ Surfactant_LBS + ProppantLoading_LBSPerGAL,
##     data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -355865 -154941  -45623  112558  682239
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)               412917.428  60475.796   6.828 1.21e-10 ***
## Surfactant_LBS                 1.096      2.180   0.503   0.6157
## ProppantLoading_LBSPerGAL -57557.507  29818.765  -1.930   0.0551 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 204700 on 184 degrees of freedom
##   (56 observations deleted due to missingness)
## Multiple R-squared:  0.02912,    Adjusted R-squared:  0.01856
## F-statistic: 2.759 on 2 and 184 DF,  p-value: 0.06597
```

```
#This code here is for demonstrating the relation between Surfactant, Proppant Loading and First 36 Month Prodcution of BOE.
Along with the ggplot there is a summary provided that provides key information. As we can see the p-value   has a  good val
ue but the R Adjusted squared indicates a poor value along with the RMSE value and F statistic test. This means that another
model should be found since this one is not responsible. These variables were chosen because the Surfactant along with the P
roppant Loading would have an influence over the prodiction rate.
```

```
ggplot(data=WELLS,mapping= aes(x=AverageStageSpacing_FT,y=First36MonthProd_BOE))+geom_point(color='red')+geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



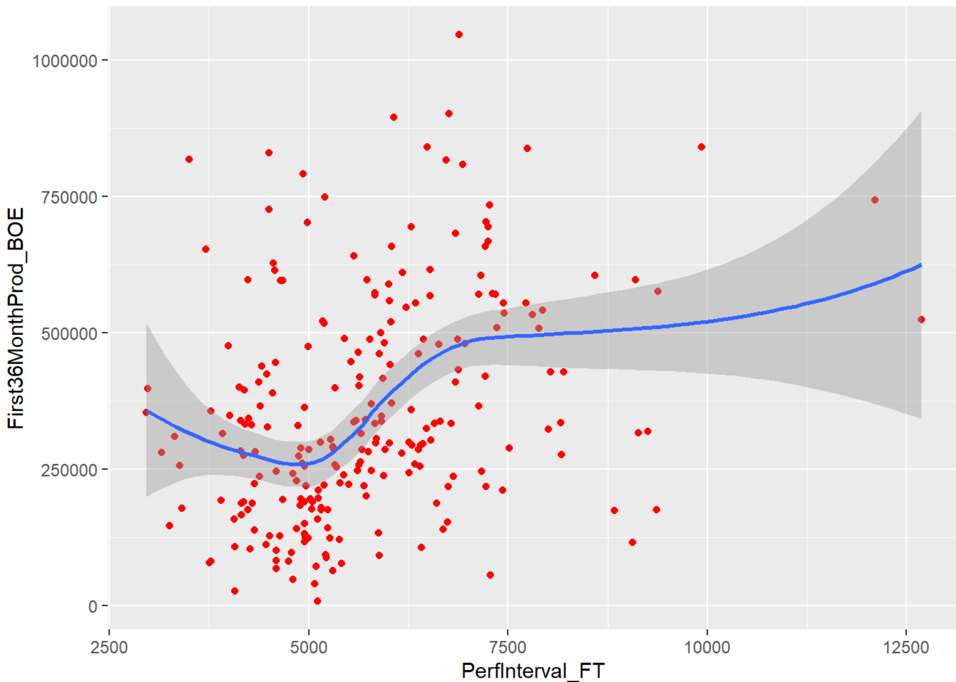

```
simplemodel=lm(First36MonthProd_BOE~AverageStageSpacing_FT,data=WELLS)
summary(simplemodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ AverageStageSpacing_FT, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -316768 -147486  -49221  108260  715853
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            255171.2    34158.5   7.470 1.47e-12 ***
## AverageStageSpacing_FT    351.2      111.2   3.159  0.00178 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 201100 on 241 degrees of freedom
## Multiple R-squared:  0.03977,    Adjusted R-squared:  0.03579
## F-statistic: 9.982 on 1 and 241 DF,  p-value: 0.001783
```

```
#This code here is for demonstrating the relation between Average spacing and First 36 Month Prodcution of BOE. Along with t
he ggplot there is a summary provided that provides key information. As we can see the p-value and  F-statistic have good va
lues but the R Adjusted squared indicates a poor value along with the RMSE value. This means that only a limited amount of i
nformation is being provided between these two variables and the model shouldnt be chosen and another one should be found. F
rom this model we can see that more variables are needed to determine a more holistic model.
```

```
ggplot(data=WELLS,mapping= aes(x=PerfInterval_FT,y=First36MonthProd_BOE))+geom_point(color='red')+geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
simplemodel=lm(First36MonthProd_BOE~PerfInterval_FT,data=WELLS)
summary(simplemodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ PerfInterval_FT, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -398378 -135463  -43976  113459  636078
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     81004.370  50454.095   1.606     0.11
## PerfInterval_FT    47.843      8.537   5.604 5.71e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 193000 on 241 degrees of freedom
## Multiple R-squared:  0.1153, Adjusted R-squared:  0.1116
## F-statistic:  31.4 on 1 and 241 DF,  p-value: 5.708e-08
```

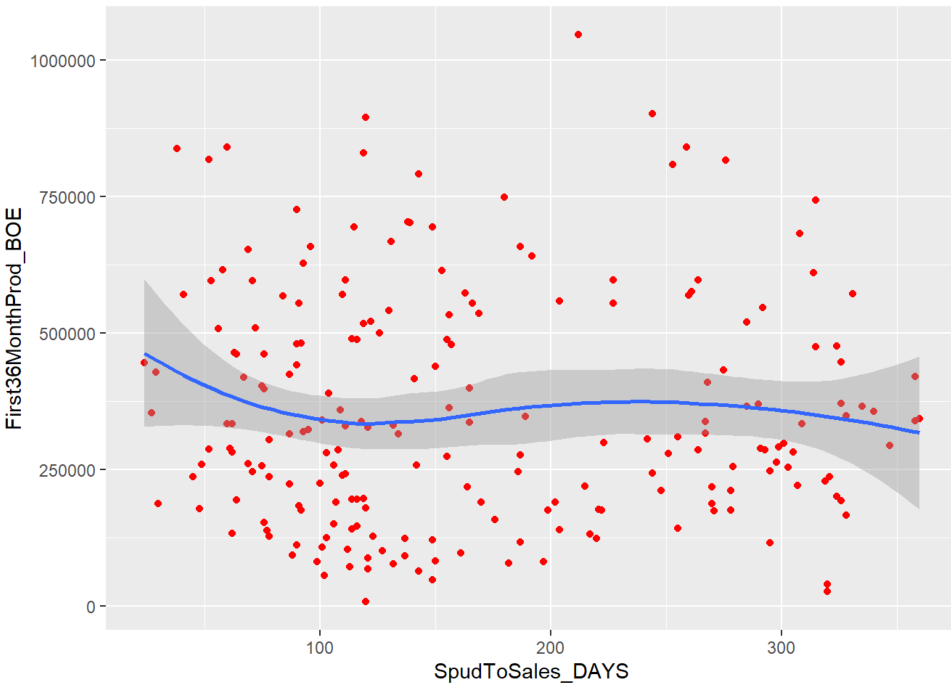
```
#This code here is for demonstrating the relation between Perf Inerval and First 36 Month Prodcution of BOE. Along with the
 ggplot there is a summary provided that provides key information. As we can see the p-value and  F-statistic have good valu
es but the R Adjusted squared indicates a poor value along with the RMSE value even though it is an improvement for the prio
r model. This means that only a limited amount of information is being provided between these two variables. Therefore the m
odel shouldnt be accepted in its entirety. These variables were chosen because the Perf Interval We can determine hat more v
ariables are needed for an anlysis.
```

```
ggplot(data=WELLS,mapping= aes(x=SpudToSales_DAYS,y=First36MonthProd_BOE))+geom_point(color='red')+geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 31 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 31 rows containing missing values (geom_point).
```



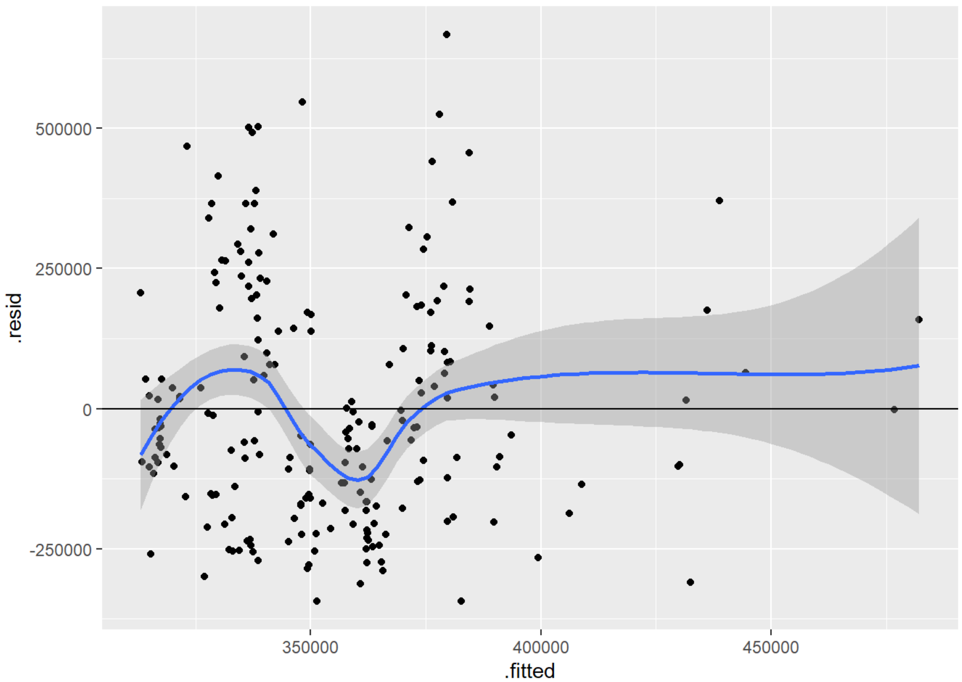

```
simplemodel=lm(First36MonthProd_BOE~ElevationGL_FT+SpudToSales_DAYS,data=WELLS)
summary(simplemodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ ElevationGL_FT + SpudToSales_DAYS,
##     data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -343866 -159257  -34856  140190  667030
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      126239.02  118277.40   1.067   0.2871
## ElevationGL_FT       93.27      46.19   2.019   0.0448 *
## SpudToSales_DAYS    -43.76     158.22  -0.277   0.7824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 209000 on 208 degrees of freedom
##   (32 observations deleted due to missingness)
## Multiple R-squared:  0.01926,    Adjusted R-squared:  0.009825
## F-statistic: 2.042 on 2 and 208 DF,  p-value: 0.1324
```

```
# We can clearly see that the overall model is not accurate as all major indicators like R adjusted squared, RMSE and F test
indicate a poor result. The model should be ignored and another one found.
```

```
ggplot(simplemodel, aes(x=.fitted, y=.resid)) + geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



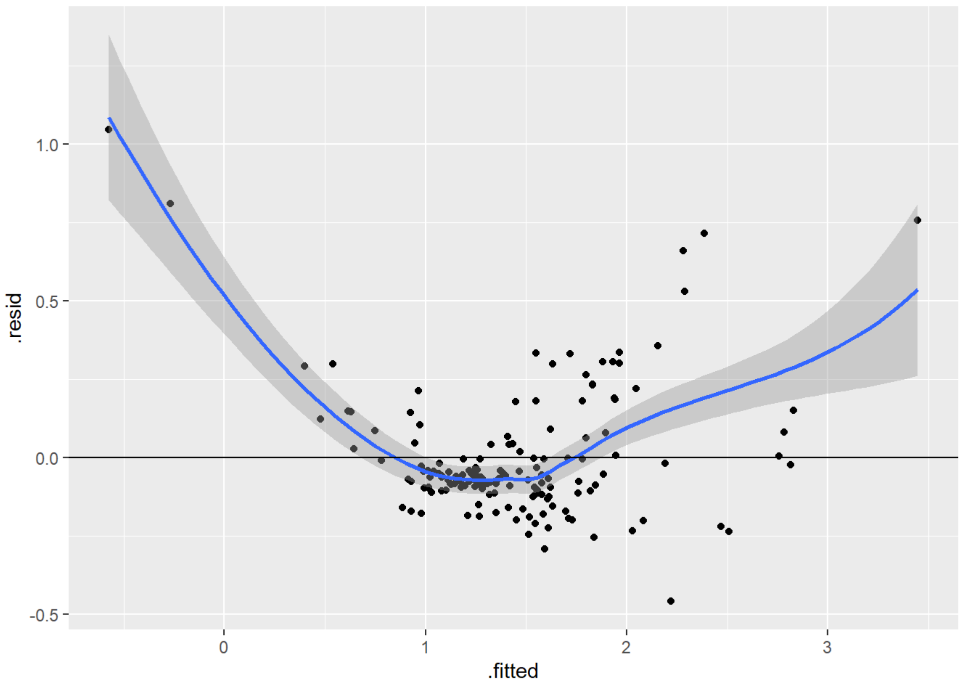

```
summary(simplemodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ ElevationGL_FT + SpudToSales_DAYS,
##     data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -343866 -159257  -34856  140190  667030
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      126239.02  118277.40   1.067   0.2871
## ElevationGL_FT       93.27      46.19   2.019   0.0448 *
## SpudToSales_DAYS    -43.76     158.22  -0.277   0.7824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 209000 on 208 degrees of freedom
##   (32 observations deleted due to missingness)
## Multiple R-squared:  0.01926,    Adjusted R-squared:  0.009825
## F-statistic: 2.042 on 2 and 208 DF,  p-value: 0.1324
```

```
# We can clearly see that the overall model is not accurate as all major indicators like R adjusted squared, RMSE and F test
indicate a poor result. The model should be ignored and another one found.
```

```
secondmodel<-lm(ProppantLoading_LBSPerGAL~Surfactant_LBS+I(FrictionReducer_LBS*First36MonthProd_BOE^8)+FluidIntensity_BBLPer
FT+ProppantIntensity_LBSPerFT, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



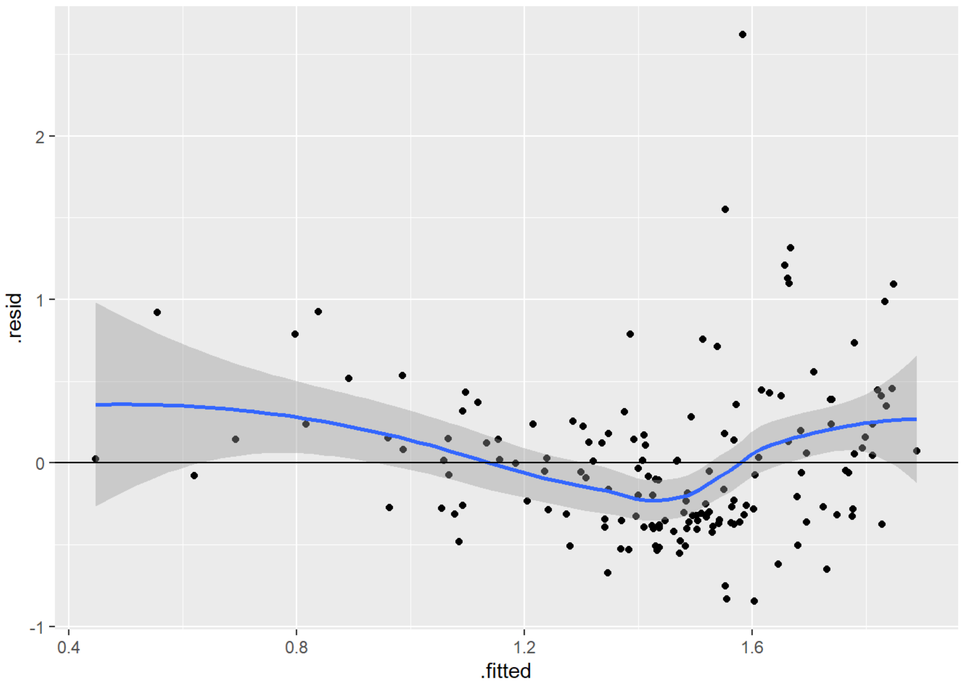

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = ProppantLoading_LBSPerGAL ~ Surfactant_LBS + I(FrictionReducer_LBS *
##     First36MonthProd_BOE^8) + FluidIntensity_BBLPerFT + ProppantIntensity_LBSPerFT,
##     data = WELLS)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45922 -0.10487 -0.06118  0.04295  1.04478
##
## Coefficients:
##                                                Estimate Std. Error
## (Intercept)                                   1.318e+00  5.717e-02
## Surfactant_LBS                               -1.016e-05  2.400e-06
## I(FrictionReducer_LBS * First36MonthProd_BOE^8) -2.601e-54  1.136e-53
## FluidIntensity_BBLPerFT                      -9.587e-02  4.523e-03
## ProppantIntensity_LBSPerFT                    2.087e-03  8.006e-05
##                                              t value Pr(>|t|)
## (Intercept)                                   23.050  < 2e-16 ***
## Surfactant_LBS                                -4.235 3.94e-05 ***
## I(FrictionReducer_LBS * First36MonthProd_BOE^8)  -0.229    0.819
## FluidIntensity_BBLPerFT                      -21.194  < 2e-16 ***
## ProppantIntensity_LBSPerFT                    26.073  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2122 on 152 degrees of freedom
##   (86 observations deleted due to missingness)
## Multiple R-squared:  0.8609, Adjusted R-squared:  0.8572
## F-statistic: 235.2 on 4 and 152 DF,  p-value: < 2.2e-16
```

```
# quadratic model 1
# Only analyze models with y value of 36 month production.
```

```
secondmodel<-lm(ProppantLoading_LBSPerGAL~Surfactant_LBS+I(ProppantIntensity_LBSPerFT*First36MonthProd_BOE^8)+FluidIntensity
_BBLPerFT+FrictionReducer_LBS, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



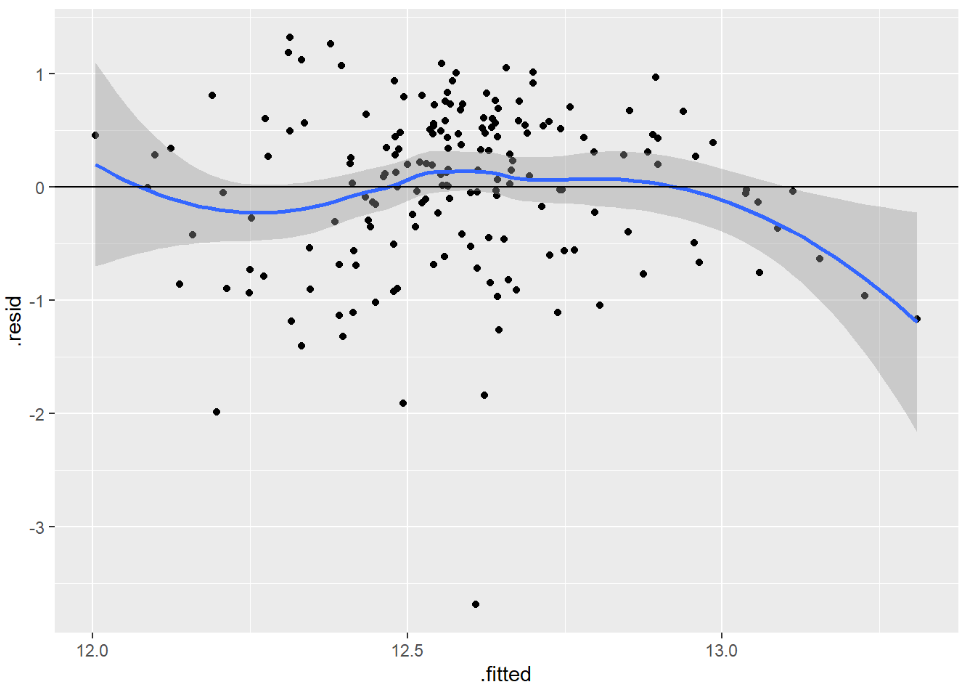

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = ProppantLoading_LBSPerGAL ~ Surfactant_LBS + I(ProppantIntensity_LBSPerFT *
##     First36MonthProd_BOE^8) + FluidIntensity_BBLPerFT + FrictionReducer_LBS,
##     data = WELLS)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.84768 -0.34529 -0.07377  0.19491  2.61895
##
## Coefficients:
##                                                         Estimate
## (Intercept)                                            2.145e+00
## Surfactant_LBS                                        -1.995e-05
## I(ProppantIntensity_LBSPerFT * First36MonthProd_BOE^8)  1.282e-52
## FluidIntensity_BBLPerFT                               -4.092e-02
## FrictionReducer_LBS                                   -4.958e-07
##                                                       Std. Error t value
## (Intercept)                                            1.114e-01  19.247
## Surfactant_LBS                                         5.571e-06  -3.581
## I(ProppantIntensity_LBSPerFT * First36MonthProd_BOE^8)  4.225e-52   0.303
## FluidIntensity_BBLPerFT                                9.885e-03  -4.140
## FrictionReducer_LBS                                    5.197e-06  -0.095
##                                                       Pr(>|t|)
## (Intercept)                                            < 2e-16 ***
## Surfactant_LBS                                         0.00046 ***
## I(ProppantIntensity_LBSPerFT * First36MonthProd_BOE^8)  0.76195
## FluidIntensity_BBLPerFT                               5.75e-05 ***
## FrictionReducer_LBS                                    0.92412
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4963 on 152 degrees of freedom
##   (86 observations deleted due to missingness)
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2193
## F-statistic: 11.95 on 4 and 152 DF,  p-value: 1.801e-08
```

```
# quadratic model 2
# Only analyze models with y value of 36 month production.
```

```
secondmodel<-lm(log(First36MonthProd_BOE)~log(I(TotalFluidPumped_BBL*FluidIntensity_BBLPerFT^.5*TotalWaterPumped_GAL^.5)+FrictionReducer_LBS+ProppantIntensity_LBSPerFT*ProppantLoading_LBSPerGAL+PerfInterval_FT), data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



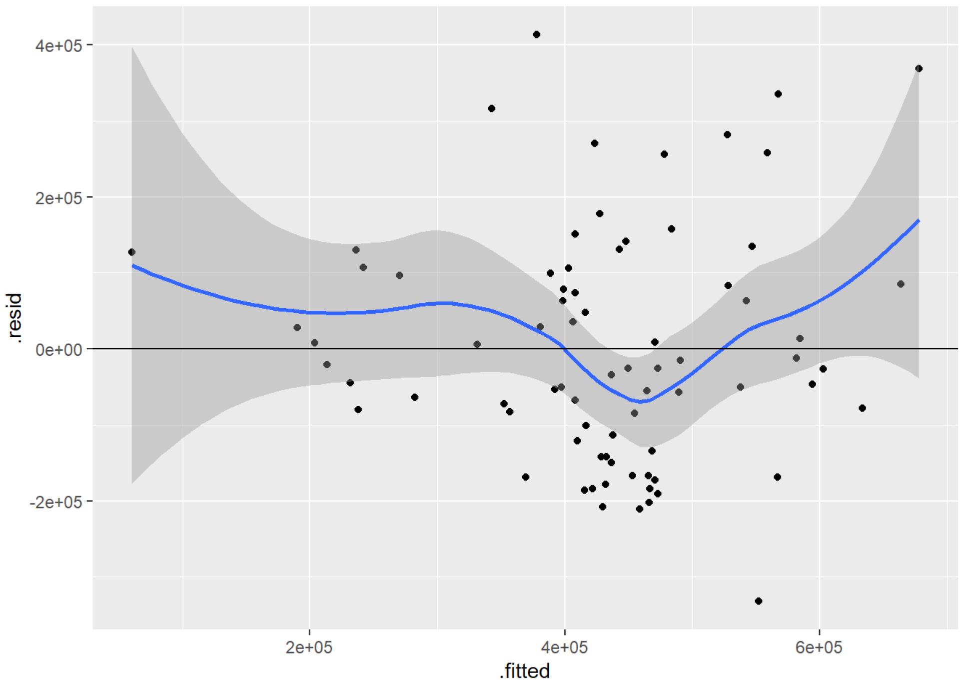

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = log(First36MonthProd_BOE) ~ log(I(TotalFluidPumped_BBL *
##     FluidIntensity_BBLPerFT^0.5 * TotalWaterPumped_GAL^0.5) +
##     FrictionReducer_LBS + ProppantIntensity_LBSPerFT * ProppantLoading_LBSPerGAL +
##     PerfInterval_FT), data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6895 -0.4612  0.1205  0.5085  1.3170
##
## Coefficients:
##
Estimate
## (Intercept)
6.90188
## log(I(TotalFluidPumped_BBL * FluidIntensity_BBLPerFT^0.5 * TotalWaterPumped_GAL^0.5) + FrictionReducer_LBS + ProppantInte
nsity_LBSPerFT * ProppantLoading_LBSPerGAL + PerfInterval_FT)  0.29249
##
Std. Error
## (Intercept)
1.38742
## log(I(TotalFluidPumped_BBL * FluidIntensity_BBLPerFT^0.5 * TotalWaterPumped_GAL^0.5) + FrictionReducer_LBS + ProppantInte
nsity_LBSPerFT * ProppantLoading_LBSPerGAL + PerfInterval_FT)    0.07139
##
t value
## (Intercept)
4.975
## log(I(TotalFluidPumped_BBL * FluidIntensity_BBLPerFT^0.5 * TotalWaterPumped_GAL^0.5) + FrictionReducer_LBS + ProppantInte
nsity_LBSPerFT * ProppantLoading_LBSPerGAL + PerfInterval_FT)   4.097
##
Pr(>|t|)
## (Intercept)
1.55e-06
## log(I(TotalFluidPumped_BBL * FluidIntensity_BBLPerFT^0.5 * TotalWaterPumped_GAL^0.5) + FrictionReducer_LBS + ProppantInte
nsity_LBSPerFT * ProppantLoading_LBSPerGAL + PerfInterval_FT) 6.37e-05
##
## (Intercept)
***
## log(I(TotalFluidPumped_BBL * FluidIntensity_BBLPerFT^0.5 * TotalWaterPumped_GAL^0.5) + FrictionReducer_LBS + ProppantInte
nsity_LBSPerFT * ProppantLoading_LBSPerGAL + PerfInterval_FT) ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.722 on 176 degrees of freedom
##   (65 observations deleted due to missingness)
## Multiple R-squared:  0.08707,    Adjusted R-squared:  0.08189
## F-statistic: 16.79 on 1 and 176 DF,  p-value: 6.374e-05
```

```
# quadratic model 3
# This quadratic model is for determining the first 36 month production having interactive terms of total fluid pumped, squa
re root of fluid intensity and square root of total water pumped. The reason these terms were chosen as interactive terms is
becuase if they change they will have the greatest impact on the overall production of oil. The reason that these interactiv
e terms were either raised to the power of a half or logged is through a trial and error approach that yeilded the best resu
lts. The log reduces the RMSE value. As can be seen the P value, RMSE and F stat are good except for the R adjusted squared
 that could be improved.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(ScaleInhibitor_LBS^2)+ProppantIntensity_L
BSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



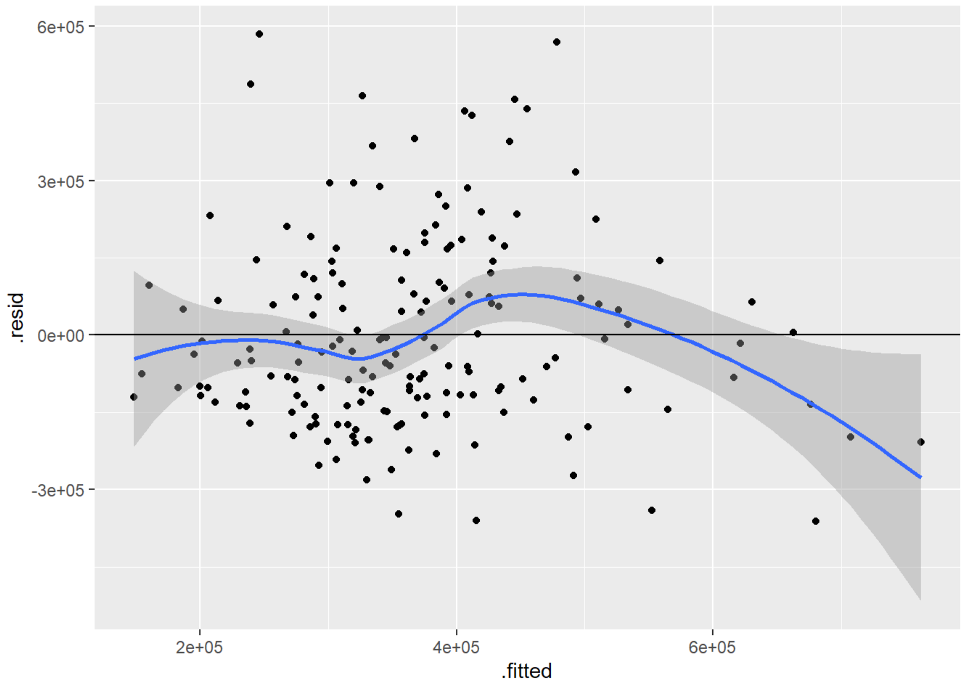

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + I(ScaleInhibitor_LBS^2) + ProppantIntensity_LBSPerFT *
##     ProppantLoading_LBSPerGAL, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -332583 -115233  -25947   96884  413330
##
## Coefficients:
##                                                       Estimate Std. Error
## (Intercept)                                         -5.065e+05  3.796e+05
## FluidIntensity_BBLPerFT                              2.440e+04  2.922e+04
## FrictionReducer_LBS                                  4.523e+01  1.498e+01
## I(ScaleInhibitor_LBS^2)                             -3.063e-03  1.117e-03
## ProppantIntensity_LBSPerFT                           3.074e+02  4.230e+02
## ProppantLoading_LBSPerGAL                            4.676e+05  2.556e+05
## FluidIntensity_BBLPerFT:FrictionReducer_LBS         -1.381e+00  9.465e-01
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL -2.917e+02  1.163e+02
##                                                     t value Pr(>|t|)
## (Intercept)                                          -1.334  0.18654
## FluidIntensity_BBLPerFT                               0.835  0.40649
## FrictionReducer_LBS                                   3.019  0.00357 **
## I(ScaleInhibitor_LBS^2)                              -2.742  0.00779 **
## ProppantIntensity_LBSPerFT                            0.727  0.46996
## ProppantLoading_LBSPerGAL                             1.829  0.07172 .
## FluidIntensity_BBLPerFT:FrictionReducer_LBS          -1.459  0.14922
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL  -2.507  0.01455 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 162200 on 68 degrees of freedom
##   (167 observations deleted due to missingness)
## Multiple R-squared:  0.3475, Adjusted R-squared:  0.2804
## F-statistic: 5.174 on 7 and 68 DF,  p-value: 9.285e-05
```

```
# quadratic model 4
# This model has a good P value and R Squared adjusted value but the F-satistic and and the RMSE values are very poor and co
mpared to the previous model needs improvement. The reason that the Scale Inhibitor was chosen as an interactive variable sq
uared is becuase it helps prevent the build up of particulates that would affect the flow of fluid. Therefore if we can cont
rol the scale inhibitor and determine its value then fluid flow can be increased which would help he production rate.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(PerfInterval_FT^2)+ProppantIntensity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



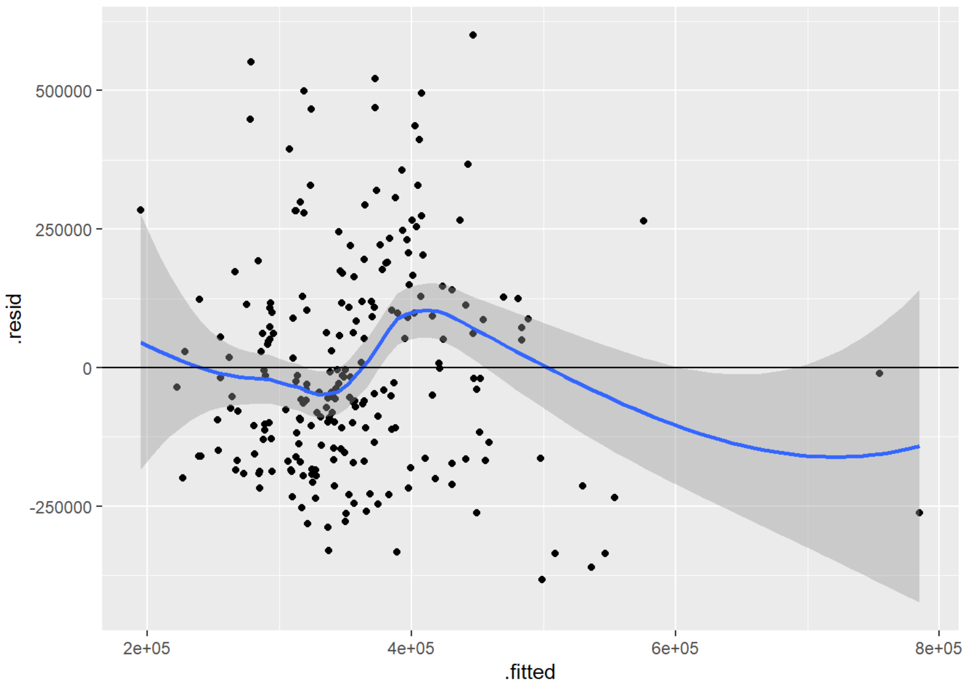

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + I(PerfInterval_FT^2) + ProppantIntensity_LBSPerFT *
##     ProppantLoading_LBSPerGAL, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -361523 -130387  -36775  103643  584129
##
## Coefficients:
##                                                      Estimate Std. Error
## (Intercept)                                         1.274e+05  1.859e+05
## FluidIntensity_BBLPerFT                             8.151e+03  1.410e+04
## FrictionReducer_LBS                                 1.308e+01  5.150e+00
## I(PerfInterval_FT^2)                                5.460e-03  1.113e-03
## ProppantIntensity_LBSPerFT                          5.206e+01  2.217e+02
## ProppantLoading_LBSPerGAL                          -6.086e+04  9.768e+04
## FluidIntensity_BBLPerFT:FrictionReducer_LBS        -8.268e-01  4.587e-01
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL -2.184e+00  7.712e+01
##                                                    t value Pr(>|t|)
## (Intercept)                                          0.685   0.4941
## FluidIntensity_BBLPerFT                              0.578   0.5641
## FrictionReducer_LBS                                  2.539   0.0120 *
## I(PerfInterval_FT^2)                                 4.904 2.16e-06 ***
## ProppantIntensity_LBSPerFT                           0.235   0.8147
## ProppantLoading_LBSPerGAL                           -0.623   0.5341
## FluidIntensity_BBLPerFT:FrictionReducer_LBS         -1.802   0.0732 .
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL  -0.028   0.9774
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 189600 on 172 degrees of freedom
##   (63 observations deleted due to missingness)
## Multiple R-squared:  0.2538, Adjusted R-squared:  0.2235
## F-statistic: 8.359 on 7 and 172 DF,  p-value: 8.884e-09
```

```
# quadratic model 5
# This model has a good P value and R Squared adjusted value but the F-satistic and and the RMSE values are very poor and co
mpared to the previous models and needs improvement. The reason that the Perf Interval  was chosen as an interactive variabl
e squared is becuase it determine what the rate of fluid flow will be. Therefore if we can control the perf interval  and de
termine its value then fluid flow can be increased which would help he production rate. The other varialbes are needed in th
is model as it relates all the values together.
```

```
secondmodel<-lm(First36MonthProd_BOE~I(PerfInterval_FT^2)+ProppantIntensity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point()+geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



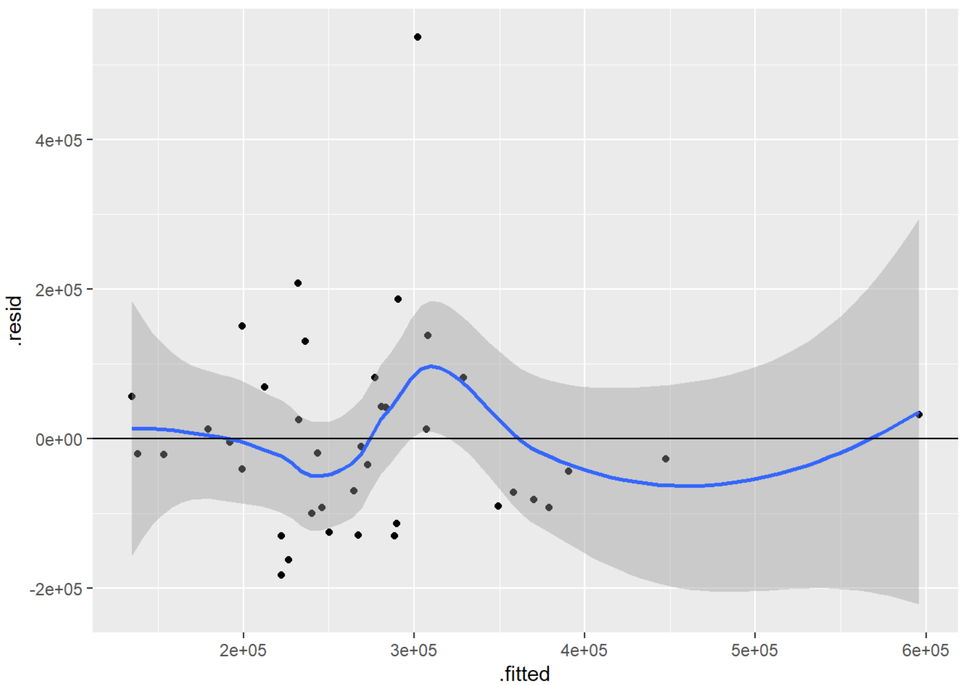

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ I(PerfInterval_FT^2) + ProppantIntensity_LBSPerFT *
##     ProppantLoading_LBSPerGAL, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -382306 -150406  -35999  113364  600011
##
## Coefficients:
##                                                       Estimate Std. Error
## (Intercept)                                          1.605e+05  8.726e+04
## I(PerfInterval_FT^2)                                 3.728e-03  6.767e-04
## ProppantIntensity_LBSPerFT                           2.510e+02  1.236e+02
## ProppantLoading_LBSPerGAL                           -1.047e+04  5.497e+04
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL -7.371e+01  6.630e+01
##                                                      t value Pr(>|t|)
## (Intercept)                                            1.839   0.0672 .
## I(PerfInterval_FT^2)                                   5.509 9.59e-08 ***
## ProppantIntensity_LBSPerFT                             2.031   0.0434 *
## ProppantLoading_LBSPerGAL                             -0.190   0.8492
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL  -1.112   0.2675
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 194600 on 231 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.1309, Adjusted R-squared:  0.1159
## F-statistic: 8.701 on 4 and 231 DF,  p-value: 1.473e-06
```

```
# quadratic model 6 attempt without using all the pieces
# This model is similar to the one prior for the interactive term but the only difference is that we are using fewer variabl
es to determine what effect it would have on the model. As can be seen from the summary table the moment we removed the othe
r variables the Pvalue, AdjustedR squared value, F-statistic value along with the RMSE all change. the R adjusted squared va
lue decreases as does the F test value and the RMSE value increases. This means that the model isnt as accurae and efficient
as the prior one.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+Proppant_LBS+I(AcidVolume_BBL^2)+ProppantIn
tensity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



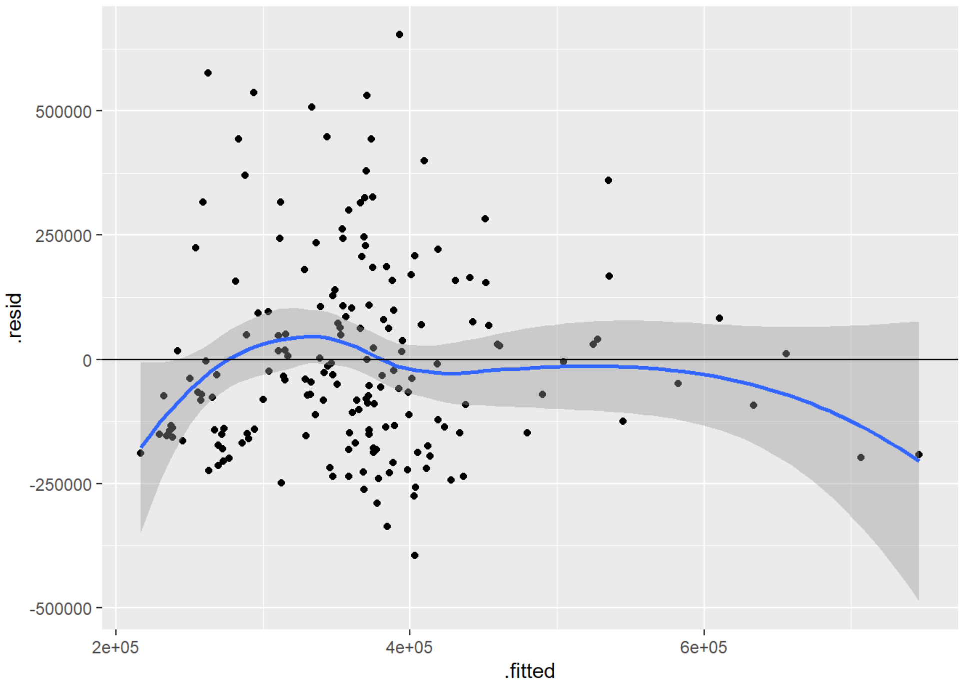

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + Proppant_LBS + I(AcidVolume_BBL^2) +
##     ProppantIntensity_LBSPerFT * ProppantLoading_LBSPerGAL, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -182858  -91691  -20445   49194  536267
##
## Coefficients:
##                                                       Estimate Std. Error
## (Intercept)                                          6.159e+04  5.031e+05
## FluidIntensity_BBLPerFT                              2.076e+04  3.787e+04
## FrictionReducer_LBS                                 -6.436e+01  4.674e+01
## Proppant_LBS                                         2.610e-02  3.459e-02
## I(AcidVolume_BBL^2)                                 -1.732e-05  2.424e-05
## ProppantIntensity_LBSPerFT                          -8.929e+02  7.038e+02
## ProppantLoading_LBSPerGAL                            2.399e+05  2.243e+05
## FluidIntensity_BBLPerFT:FrictionReducer_LBS          8.771e+00  4.303e+00
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL -1.789e+00  2.231e+02
##                                                     t value Pr(>|t|)
## (Intercept)                                           0.122   0.9034
## FluidIntensity_BBLPerFT                               0.548   0.5876
## FrictionReducer_LBS                                  -1.377   0.1787
## Proppant_LBS                                          0.754   0.4565
## I(AcidVolume_BBL^2)                                  -0.715   0.4804
## ProppantIntensity_LBSPerFT                           -1.269   0.2143
## ProppantLoading_LBSPerGAL                             1.070   0.2932
## FluidIntensity_BBLPerFT:FrictionReducer_LBS           2.039   0.0504 .
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL  -0.008   0.9937
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 147000 on 30 degrees of freedom
##   (204 observations deleted due to missingness)
## Multiple R-squared:  0.3067, Adjusted R-squared:  0.1218
## F-statistic: 1.659 on 8 and 30 DF,  p-value: 0.1501
```

```
# quadratic model 7
# In this model the acid volume is an interactive term. The reason that it is an interactive term is because having acid in
 the pipe will corrode the pipe and damage it causing a loss in production. Having it squared is becuase it increases the Ad
justed R value squared without affecting the other parameters to much. As can be seen from the model it is inefficient since
all parameters indicate it.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(TVD_FT^2)+ProppantIntensity_LBSPerFT*Prop
pantLoading_LBSPerGAL, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



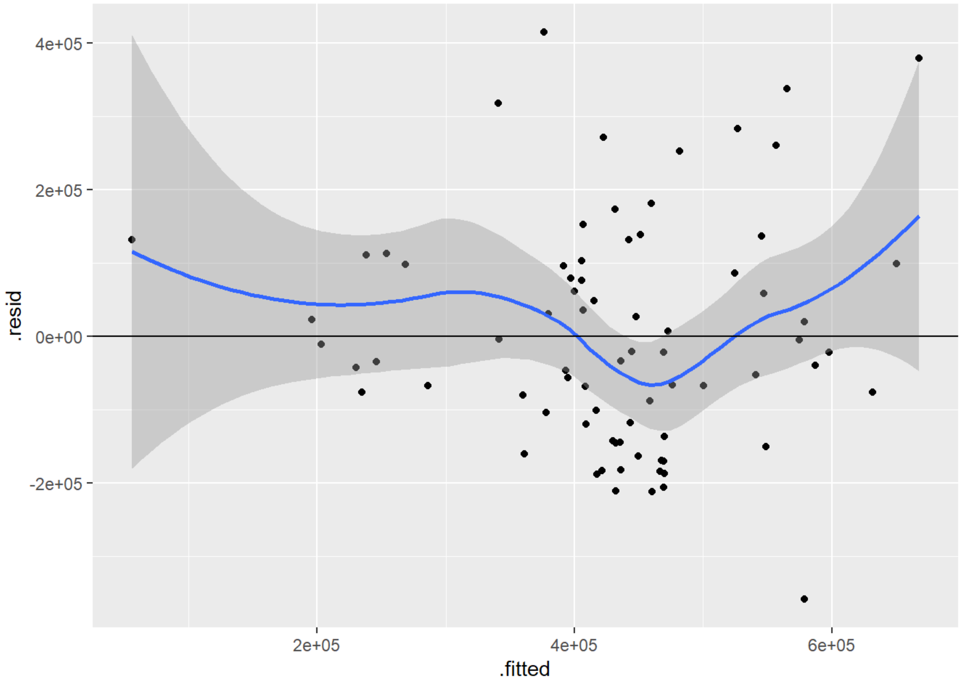

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + I(TVD_FT^2) + ProppantIntensity_LBSPerFT *
##     ProppantLoading_LBSPerGAL, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -395686 -148426  -39508   99888  653467
##
## Coefficients:
##                                                       Estimate Std. Error
## (Intercept)                                          1.770e+05  2.337e+05
## FluidIntensity_BBLPerFT                              1.332e+04  1.551e+04
## FrictionReducer_LBS                                  1.706e+01  5.400e+00
## I(TVD_FT^2)                                          2.099e-03  1.545e-03
## ProppantIntensity_LBSPerFT                          -1.500e+02  2.351e+02
## ProppantLoading_LBSPerGAL                           -3.123e+04  1.098e+05
## FluidIntensity_BBLPerFT:FrictionReducer_LBS         -9.152e-01  4.897e-01
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL  2.759e+01  8.214e+01
##                                                      t value Pr(>|t|)
## (Intercept)                                            0.757  0.44990
## FluidIntensity_BBLPerFT                                0.859  0.39176
## FrictionReducer_LBS                                    3.159  0.00187 **
## I(TVD_FT^2)                                            1.358  0.17621
## ProppantIntensity_LBSPerFT                            -0.638  0.52452
## ProppantLoading_LBSPerGAL                             -0.285  0.77632
## FluidIntensity_BBLPerFT:FrictionReducer_LBS           -1.869  0.06332 .
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL   0.336  0.73733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 201400 on 172 degrees of freedom
##   (63 observations deleted due to missingness)
## Multiple R-squared:  0.1585, Adjusted R-squared:  0.1243
## F-statistic:  4.63 on 7 and 172 DF,  p-value: 9.051e-05
```

```
# quadratic model 8
# The interactive term in this is TVD ft. The reason why this term is an interactive term is becuase the depth that the pipe
is drilled to will determine what the pressure is and that will determine what the flow rate is. The reason that it is squar
ed is becuase it improves the parameters. As can be seen from the model the R adjusted squared and RMSE values are not gooda
s is the F test results.The only good value is the pvalue and this makes the model not a good choice.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(TVD_FT*ScaleInhibitor_LBS^2)+ProppantInte
nsity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



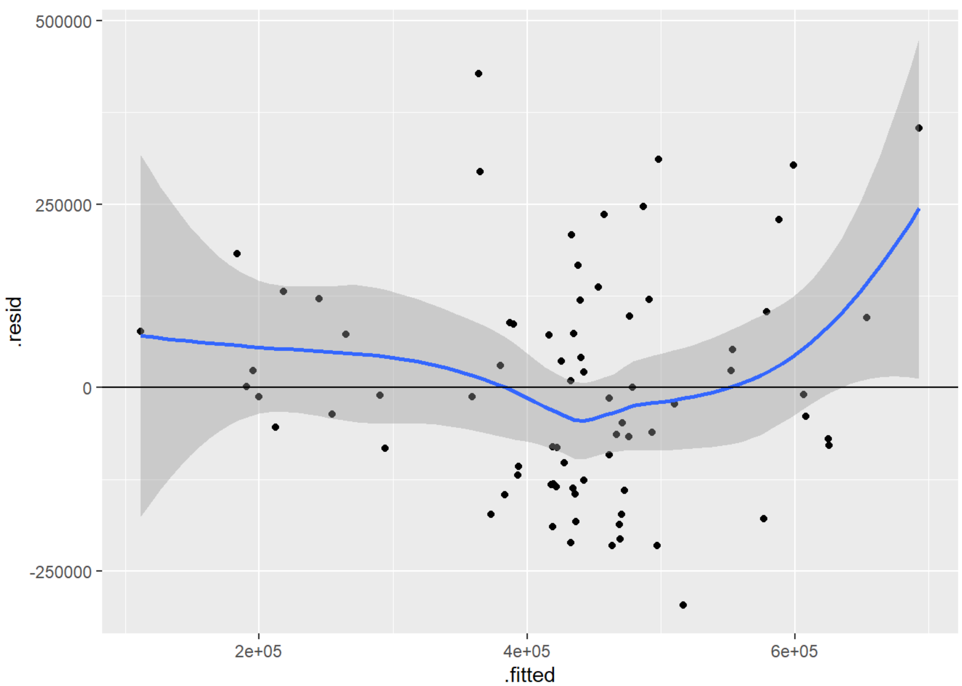

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + I(TVD_FT * ScaleInhibitor_LBS^2) +
##     ProppantIntensity_LBSPerFT * ProppantLoading_LBSPerGAL, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -359229 -119032  -22312   97914  414749
##
## Coefficients:
##                                                       Estimate Std. Error
## (Intercept)                                         -4.705e+05  3.827e+05
## FluidIntensity_BBLPerFT                              2.110e+04  2.942e+04
## FrictionReducer_LBS                                  4.562e+01  1.515e+01
## I(TVD_FT * ScaleInhibitor_LBS^2)                    -4.326e-07  1.788e-07
## ProppantIntensity_LBSPerFT                           3.294e+02  4.276e+02
## ProppantLoading_LBSPerGAL                            4.522e+05  2.582e+05
## FluidIntensity_BBLPerFT:FrictionReducer_LBS         -1.404e+00  9.571e-01
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL -2.876e+02  1.176e+02
##                                                     t value Pr(>|t|)
## (Intercept)                                          -1.229  0.22313
## FluidIntensity_BBLPerFT                               0.717  0.47574
## FrictionReducer_LBS                                   3.011  0.00365 **
## I(TVD_FT * ScaleInhibitor_LBS^2)                     -2.420  0.01822 *
## ProppantIntensity_LBSPerFT                            0.770  0.44378
## ProppantLoading_LBSPerGAL                             1.752  0.08436 .
## FluidIntensity_BBLPerFT:FrictionReducer_LBS          -1.467  0.14706
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL  -2.445  0.01707 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 164000 on 68 degrees of freedom
##   (167 observations deleted due to missingness)
## Multiple R-squared:  0.3328, Adjusted R-squared:  0.2641
## F-statistic: 4.846 on 7 and 68 DF,  p-value: 0.0001795
```

```
# quadratic model 9
# The interactive term in this model is that of TVD and scale inhibitor (the reason why was mentioned before). The values th
at are returned for each parameter indicate that it is a feasible model since the R adjusted square and pvalue are good whil
e the RMSE and F test could be improved upon.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(TVD_FT*ScaleInhibitor_LBS^1)+ProppantInte
nsity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



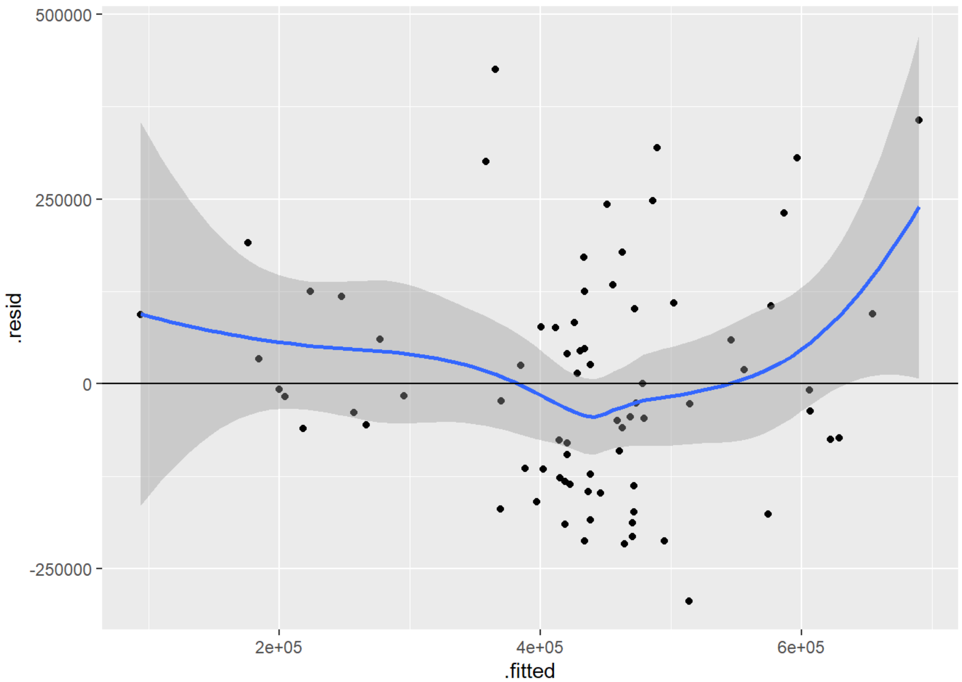

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + I(TVD_FT * ScaleInhibitor_LBS^1) +
##     ProppantIntensity_LBSPerFT * ProppantLoading_LBSPerGAL, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -296818 -121320  -12556   89741  427277
##
## Coefficients:
##                                                      Estimate Std. Error
## (Intercept)                                        -5.944e+05  3.760e+05
## FluidIntensity_BBLPerFT                             3.846e+04  2.966e+04
## FrictionReducer_LBS                                 5.820e+01  1.522e+01
## I(TVD_FT * ScaleInhibitor_LBS^1)                   -5.493e-03  1.699e-03
## ProppantIntensity_LBSPerFT                          7.130e+01  4.298e+02
## ProppantLoading_LBSPerGAL                           5.353e+05  2.532e+05
## FluidIntensity_BBLPerFT:FrictionReducer_LBS        -2.133e+00  9.396e-01
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL -2.884e+02  1.141e+02
##                                                    t value Pr(>|t|)
## (Intercept)                                         -1.581 0.118555
## FluidIntensity_BBLPerFT                              1.297 0.199159
## FrictionReducer_LBS                                  3.825 0.000287 ***
## I(TVD_FT * ScaleInhibitor_LBS^1)                    -3.234 0.001888 **
## ProppantIntensity_LBSPerFT                           0.166 0.868733
## ProppantLoading_LBSPerGAL                            2.114 0.038173 *
## FluidIntensity_BBLPerFT:FrictionReducer_LBS         -2.270 0.026362 *
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL  -2.527 0.013826 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 159100 on 68 degrees of freedom
##   (167 observations deleted due to missingness)
## Multiple R-squared:  0.3719, Adjusted R-squared:  0.3073
## F-statistic: 5.753 on 7 and 68 DF,  p-value: 2.967e-05
```

```
# quadratic model 10
# The interactive term in this model is that of TVD and scale inhibitor (the reason why was mentioned before). The values th
at are returned for each parameter indicate that it is a feasible model since the R adjusted square and pvalue are good (imp
roved upon from the previous model) while the RMSE and F test could be improved upon.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(MD_FT*ScaleInhibitor_LBS^1)+ProppantInten
sity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



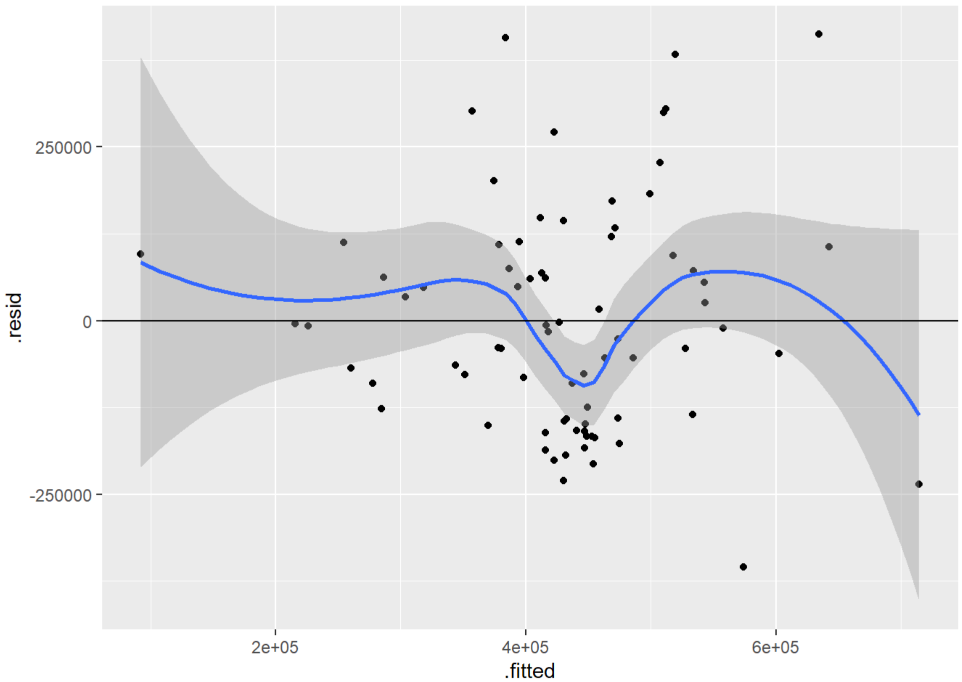

```r
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + I(MD_FT * ScaleInhibitor_LBS^1) + ProppantIntensity_LBSPerFT *
##     ProppantLoading_LBSPerGAL, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -294538 -117579  -20357   93793  425638
##
## Coefficients:
##                                                      Estimate Std. Error
## (Intercept)                                        -6.288e+05  3.786e+05
## FluidIntensity_BBLPerFT                             3.846e+04  2.967e+04
## FrictionReducer_LBS                                 5.903e+01  1.529e+01
## I(MD_FT * ScaleInhibitor_LBS^1)                    -2.547e-03  7.893e-04
## ProppantIntensity_LBSPerFT                          9.404e+01  4.279e+02
## ProppantLoading_LBSPerGAL                           5.546e+05  2.544e+05
## FluidIntensity_BBLPerFT:FrictionReducer_LBS        -2.186e+00  9.428e-01
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL -3.019e+02  1.142e+02
##                                                    t value Pr(>|t|)
## (Intercept)                                         -1.661 0.101352
## FluidIntensity_BBLPerFT                              1.296 0.199303
## FrictionReducer_LBS                                  3.861 0.000254 ***
## I(MD_FT * ScaleInhibitor_LBS^1)                     -3.227 0.001924 **
## ProppantIntensity_LBSPerFT                           0.220 0.826723
## ProppantLoading_LBSPerGAL                            2.180 0.032711 *
## FluidIntensity_BBLPerFT:FrictionReducer_LBS         -2.318 0.023465 *
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL  -2.642 0.010205 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 159100 on 68 degrees of freedom
##   (167 observations deleted due to missingness)
## Multiple R-squared:  0.3716, Adjusted R-squared:  0.3069
## F-statistic: 5.745 on 7 and 68 DF,  p-value: 3.014e-05
```

```r
# quadratic model 11
# The interactive term in this model is that of MD and scale inhibitor (the reason why was mentioned before). The values tha
t are returned for each parameter indicate that it is a feasible model since the R adjusted square and pvalue are good while
the RMSE and F test could be improved upon.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(MD_FT*ScaleInhibitor_LBS^2)+ProppantInten
sity_LBSPerFT, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



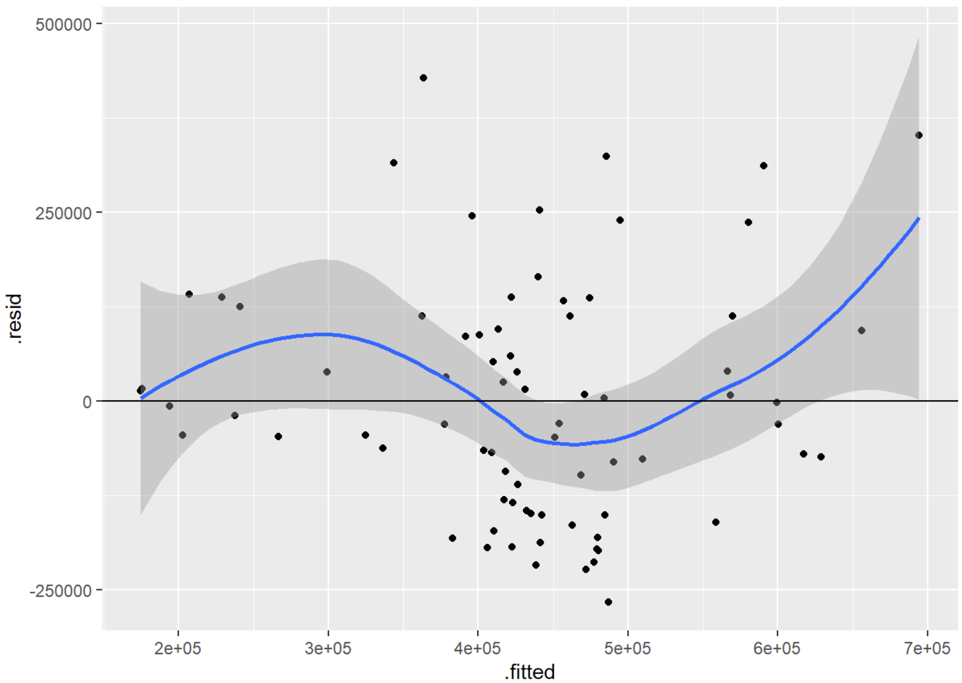

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + I(MD_FT * ScaleInhibitor_LBS^2) + ProppantIntensity_LBSPerFT,
##     data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -354485 -141032  -13075   98333  412273
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                 1.061e+05  1.145e+05   0.926
## FluidIntensity_BBLPerFT                     8.085e+02  1.132e+04   0.071
## FrictionReducer_LBS                         3.621e+01  1.148e+01   3.156
## I(MD_FT * ScaleInhibitor_LBS^2)            -1.913e-07  8.090e-08  -2.365
## ProppantIntensity_LBSPerFT                  3.298e+02  1.145e+02   2.881
## FluidIntensity_BBLPerFT:FrictionReducer_LBS -9.051e-01  6.184e-01  -1.464
##                                            Pr(>|t|)
## (Intercept)                                 0.35755
## FluidIntensity_BBLPerFT                     0.94329
## FrictionReducer_LBS                         0.00236 **
## I(MD_FT * ScaleInhibitor_LBS^2)             0.02081 *
## ProppantIntensity_LBSPerFT                  0.00526 **
## FluidIntensity_BBLPerFT:FrictionReducer_LBS  0.14779
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 168700 on 70 degrees of freedom
##   (167 observations deleted due to missingness)
## Multiple R-squared:  0.273,  Adjusted R-squared:  0.2211
## F-statistic: 5.258 on 5 and 70 DF,  p-value: 0.0003713
```

```
# quadratic model 12
# The interactive term in this model is that ofMD and scale inhibitor (the reason why was mentioned before). The values that are returned for each parameter indicate that it is a feasible model since the R adjusted square and pvalue are good but not as good as the previous model while the RMSE and F test could be improved upon.All variables that were entered were done so
 with the specific aim to improve the 36 month production.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(ElevationKB_FT*ScaleInhibitor_LBS^1)+ProppantIntensity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



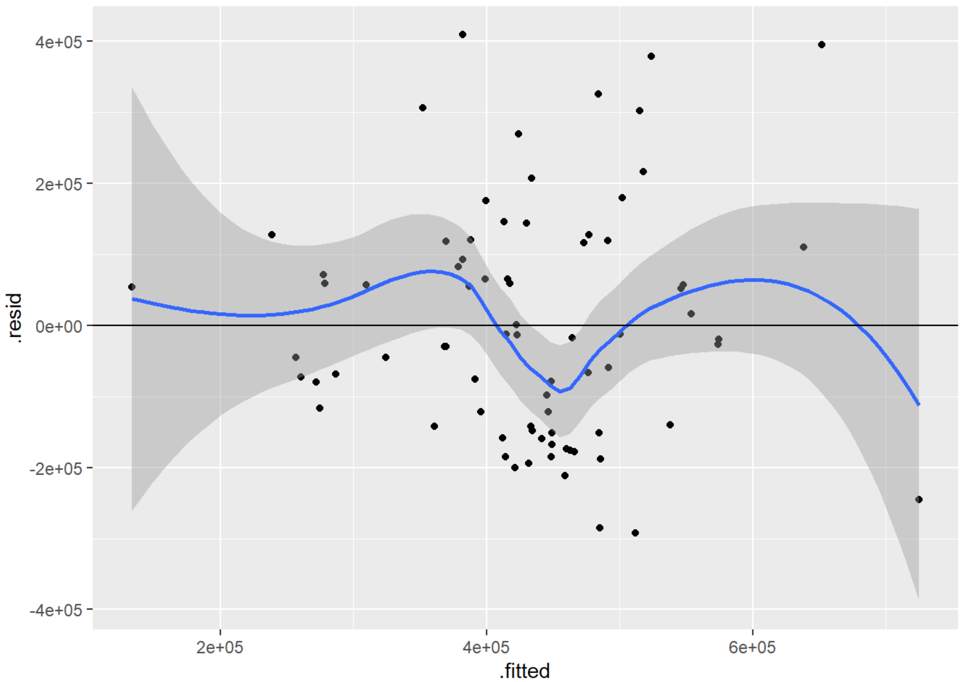

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + I(ElevationKB_FT * ScaleInhibitor_LBS^1) +
##     ProppantIntensity_LBSPerFT * ProppantLoading_LBSPerGAL, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -267165 -131735  -13061   99530  427893
##
## Coefficients:
##                                                       Estimate Std. Error
## (Intercept)                                         -6.073e+05  3.940e+05
## FluidIntensity_BBLPerFT                              3.364e+04  3.084e+04
## FrictionReducer_LBS                                  5.823e+01  1.598e+01
## I(ElevationKB_FT * ScaleInhibitor_LBS^1)            -1.047e-02  4.241e-03
## ProppantIntensity_LBSPerFT                           1.610e+02  4.423e+02
## ProppantLoading_LBSPerGAL                            5.322e+05  2.627e+05
## FluidIntensity_BBLPerFT:FrictionReducer_LBS         -1.991e+00  9.648e-01
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL -3.015e+02  1.176e+02
##                                                     t value Pr(>|t|)
## (Intercept)                                          -1.541 0.127859
## FluidIntensity_BBLPerFT                               1.091 0.279228
## FrictionReducer_LBS                                   3.643 0.000521 ***
## I(ElevationKB_FT * ScaleInhibitor_LBS^1)             -2.470 0.016045 *
## ProppantIntensity_LBSPerFT                            0.364 0.716984
## ProppantLoading_LBSPerGAL                             2.026 0.046694 *
## FluidIntensity_BBLPerFT:FrictionReducer_LBS          -2.063 0.042894 *
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL  -2.564 0.012564 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 163700 on 68 degrees of freedom
##   (167 observations deleted due to missingness)
## Multiple R-squared:  0.335,  Adjusted R-squared:  0.2666
## F-statistic: 4.894 on 7 and 68 DF,  p-value: 0.0001629
```

```
# quadratic model 13
# The interactive terms in this model are that of Elevation and scale inhibitor. Scale inhibitor prevents the build up of pa
rticulates on the pipe therefore preventing flow from slowing down while elevation determines the amount of pressure.  The v
alues that are returned for each parameter indicate that it is a feasible model since the R adjusted square and pvalue are g
ood while the RMSE and F test could be improved upon.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(ElevationKB_FT*ScaleInhibitor_LBS^2)+Prop
pantLoading_LBSPerGAL, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



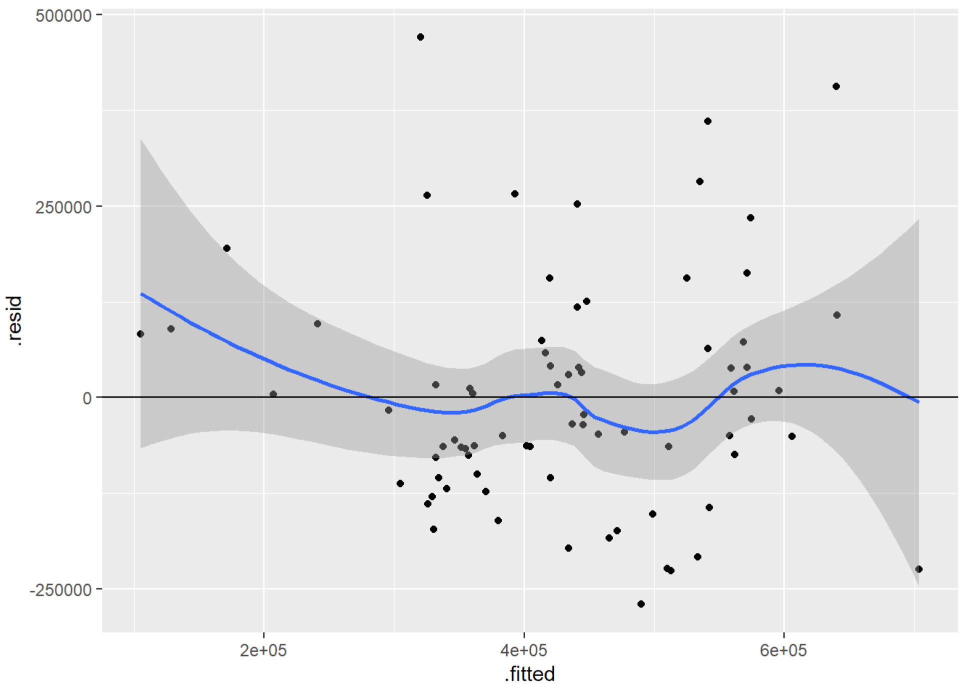

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + I(ElevationKB_FT * ScaleInhibitor_LBS^2) +
##     ProppantLoading_LBSPerGAL, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -291951 -142030  -18593  112056  409379
##
## Coefficients:
##                                             Estimate Std. Error t value
## (Intercept)                               -8.997e+04  1.408e+05  -0.639
## FluidIntensity_BBLPerFT                    2.030e+04  1.008e+04   2.014
## FrictionReducer_LBS                        4.199e+01  1.191e+01   3.525
## I(ElevationKB_FT * ScaleInhibitor_LBS^2)  -1.205e-06  5.069e-07  -2.377
## ProppantLoading_LBSPerGAL                  1.206e+05  4.495e+04   2.682
## FluidIntensity_BBLPerFT:FrictionReducer_LBS -1.104e+00  6.746e-01  -1.637
##                                           Pr(>|t|)
## (Intercept)                               0.524811
## FluidIntensity_BBLPerFT                   0.047827 *
## FrictionReducer_LBS                       0.000752 ***
## I(ElevationKB_FT * ScaleInhibitor_LBS^2)  0.020168 *
## ProppantLoading_LBSPerGAL                 0.009113 **
## FluidIntensity_BBLPerFT:FrictionReducer_LBS 0.106099
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 170200 on 70 degrees of freedom
##   (167 observations deleted due to missingness)
## Multiple R-squared:  0.2601, Adjusted R-squared:  0.2072
## F-statistic: 4.921 on 5 and 70 DF,  p-value: 0.0006456
```

```
# quadratic model 14
# The interactive terms in this model are that of Elevation and scale inhibitor. Scale inhibitor prevents the build up of pa
rticulates on the pipe therefore preventing flow from slowing down while elevation determines the amount of pressure.  The v
alues that are returned for each parameter indicate that it is a feasible model since the R adjusted square and pvalue are g
ood while the RMSE and F test could be improved upon. The only difference between this model and the one from before is that
in the previous model the scale inhibitor was not quared here it is and that affects all the parameters as can be seen makin
g this model slightly less accurate.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(ScaleInhibitor_LBS^2)+ProppantIntensity_LBSPerFT*ElevationGL_FT, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



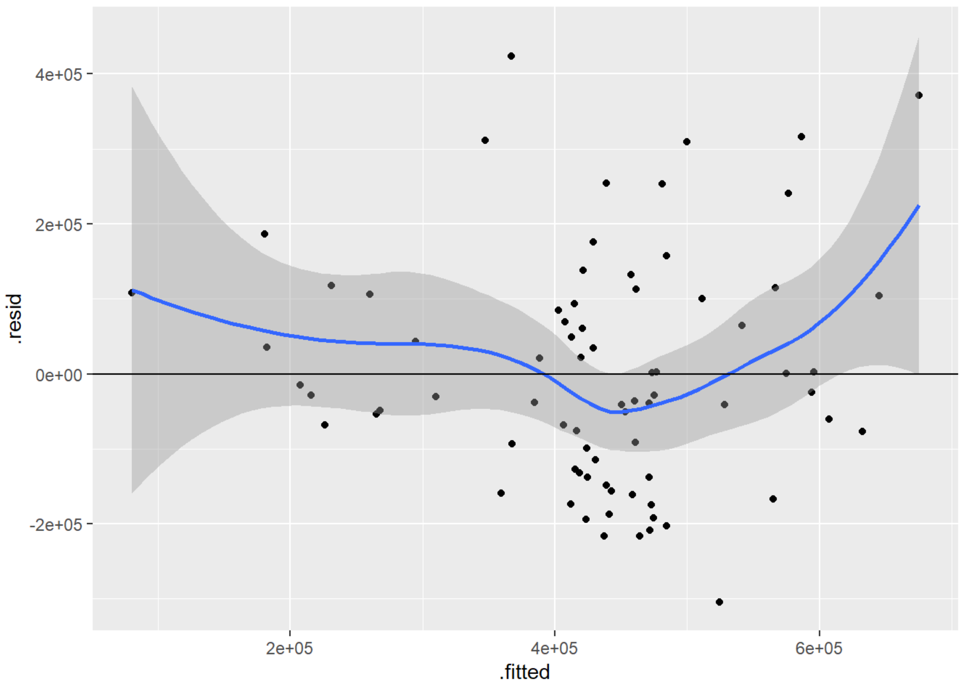

```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + I(ScaleInhibitor_LBS^2) + ProppantIntensity_LBSPerFT *
##     ElevationGL_FT, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -270487 -101534  -31199   73026  470882
##
## Coefficients:
##                                           Estimate Std. Error t value
## (Intercept)                             -7.626e+05  4.149e+05  -1.838
## FluidIntensity_BBLPerFT                 -5.897e+02  1.085e+04  -0.054
## FrictionReducer_LBS                      2.972e+01  1.097e+01   2.708
## I(ScaleInhibitor_LBS^2)                 -3.001e-03  1.312e-03  -2.286
## ProppantIntensity_LBSPerFT               1.337e+03  7.445e+02   1.796
## ElevationGL_FT                           3.457e+02  1.473e+02   2.347
## FluidIntensity_BBLPerFT:FrictionReducer_LBS -6.100e-01  6.195e-01  -0.985
## ProppantIntensity_LBSPerFT:ElevationGL_FT   -3.820e-01  2.689e-01  -1.421
##                                           Pr(>|t|)
## (Intercept)                                0.07040 .
## FluidIntensity_BBLPerFT                    0.95680
## FrictionReducer_LBS                        0.00856 **
## I(ScaleInhibitor_LBS^2)                    0.02535 *
## ProppantIntensity_LBSPerFT                 0.07700 .
## ElevationGL_FT                             0.02185 *
## FluidIntensity_BBLPerFT:FrictionReducer_LBS  0.32829
## ProppantIntensity_LBSPerFT:ElevationGL_FT    0.15995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 158500 on 68 degrees of freedom
##   (167 observations deleted due to missingness)
## Multiple R-squared:  0.3763, Adjusted R-squared:  0.312
## F-statistic:  5.86 on 7 and 68 DF,  p-value: 2.411e-05
```

```
# quadratic model 15
# The interactive term in this model is that of scale inhibitor. Scale inhibitor prevents the build up of particulates on th
e pipe therefore preventing flow from slowing down.  The values that are returned for each parameter indicate that it is a f
easible model since the R adjusted square and pvalue are good while the RMSE and F test could be improved upon.The variables
that are chosen to go with this model are specifically meant to fit with the interactive term.
```

```
secondmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(PerfInterval_FT*ScaleInhibitor_LBS^1)+Pro
ppantIntensity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(secondmodel, aes(x=.fitted, y=.resid)) +geom_point() +geom_smooth()+geom_hline(yintercept = 0)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
summary(secondmodel)
```

```
##
## Call:
## lm(formula = First36MonthProd_BOE ~ FluidIntensity_BBLPerFT *
##     FrictionReducer_LBS + I(PerfInterval_FT * ScaleInhibitor_LBS^1) +
##     ProppantIntensity_LBSPerFT * ProppantLoading_LBSPerGAL, data = WELLS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -304753 -117779  -28301  100938  424098
##
## Coefficients:
##                                                       Estimate Std. Error
## (Intercept)                                         -6.308e+05  3.858e+05
## FluidIntensity_BBLPerFT                              3.504e+04  2.998e+04
## FrictionReducer_LBS                                  5.757e+01  1.546e+01
## I(PerfInterval_FT * ScaleInhibitor_LBS^1)           -4.335e-03  1.489e-03
## ProppantIntensity_LBSPerFT                           1.568e+02  4.306e+02
## ProppantLoading_LBSPerGAL                            5.583e+05  2.589e+05
## FluidIntensity_BBLPerFT:FrictionReducer_LBS         -2.147e+00  9.555e-01
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL -3.142e+02  1.160e+02
##                                                     t value Pr(>|t|)
## (Intercept)                                          -1.635  0.10670
## FluidIntensity_BBLPerFT                               1.169  0.24664
## FrictionReducer_LBS                                   3.725  0.00040 ***
## I(PerfInterval_FT * ScaleInhibitor_LBS^1)            -2.911  0.00487 **
## ProppantIntensity_LBSPerFT                            0.364  0.71685
## ProppantLoading_LBSPerGAL                             2.157  0.03457 *
## FluidIntensity_BBLPerFT:FrictionReducer_LBS          -2.247  0.02789 *
## ProppantIntensity_LBSPerFT:ProppantLoading_LBSPerGAL  -2.709  0.00854 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 161100 on 68 degrees of freedom
##   (167 observations deleted due to missingness)
## Multiple R-squared:  0.3557, Adjusted R-squared:  0.2893
## F-statistic: 5.362 on 7 and 68 DF,  p-value: 6.387e-05
```

```
# quadratic model 16
# The interactive terms in this model are that of Perf interval and scale inhibitor. Scale inhibitor prevents the build up o
f particulates on the pipe therefore preventing flow from slowing down while Perf Interval determines the flow rate as well
 as it is the section of the wellbore that has been prepared for production.The values that are returned for each parameter
 indicate that it is a feasible model since the R adjusted square and pvalue are good while the RMSE and F test could be imp
roved upon.
```

```
fullmodel<-lm(ProppantLoading_LBSPerGAL~Surfactant_LBS+I(FrictionReducer_LBS*First36MonthProd_BOE^8)+FluidIntensity_BBLPerFT
+PerfInterval_FT, data=WELLS)
options(error=function() dump.frames(to.file=TRUE))
backmodel=ols_step_backward_p(fullmodel, prem=0.001 , details=TRUE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . Surfactant_LBS
## 2 . I(FrictionReducer_LBS * First36MonthProd_BOE^8)
## 3 . FluidIntensity_BBLPerFT
## 4 . PerfInterval_FT
##
## We are eliminating variables based on p value...
##
## - I(FrictionReducer_LBS * First36MonthProd_BOE^8)
##
## Backward Elimination: Step 1
##
##  Variable I(FrictionReducer_LBS * First36MonthProd_BOE^8) Removed
##
##                         Model Summary
## --------------------------------------------------------------
## R                       0.507       RMSE                0.473
## R-Squared               0.257       Coef. Var          33.506
## Adj. R-Squared          0.245       MSE                 0.224
## Pred R-Squared          0.227       MAE                 0.343
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     14.156          3          4.719    21.074    0.0000
## Residual       40.976        183          0.224
## Total          55.132        186
## --------------------------------------------------------------------
##
##                                         Parameter Estimates
## ---------------------------------------------------------------------------------------------------
##                   model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ---------------------------------------------------------------------------------------------------
##             (Intercept)     2.541         0.191                 13.326    0.000     2.165     2.918
##          Surfactant_LBS     0.000         0.000       -0.280    -4.194    0.000     0.000     0.000
## FluidIntensity_BBLPerFT    -0.040         0.009       -0.314    -4.688    0.000    -0.057    -0.023
##         PerfInterval_FT     0.000         0.000       -0.175    -2.736    0.007     0.000     0.000
## ---------------------------------------------------------------------------------------------------
##
##
## - PerfInterval_FT
##
## Backward Elimination: Step 2
##
##  Variable PerfInterval_FT Removed
##
##                         Model Summary
## --------------------------------------------------------------
## R                       0.476       RMSE                0.481
## R-Squared               0.226       Coef. Var          34.092
## Adj. R-Squared          0.218       MSE                 0.232
## Pred R-Squared          0.203       MAE                 0.355
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     12.481          2          6.240    26.921    0.0000
## Residual       42.652        184          0.232
```

```
## Total          55.132        186
## -------------------------------------------------------------------
##
##                                        Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##                   model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------------
##             (Intercept)     2.099         0.103                 20.391    0.000     1.896     2.302
##          Surfactant_LBS     0.000         0.000       -0.293    -4.310    0.000     0.000     0.000
## FluidIntensity_BBLPerFT    -0.038         0.009       -0.299    -4.399    0.000    -0.055    -0.021
## ----------------------------------------------------------------------------------------------------
##
##
##
## No more variables satisfy the condition of p value = 0.001
##
##
## Variables Removed:
##
## - I(FrictionReducer_LBS * First36MonthProd_BOE^8)
## - PerfInterval_FT
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## -------------------------------------------------------------
## R                       0.476       RMSE               0.481
## R-Squared               0.226       Coef. Var         34.092
## Adj. R-Squared          0.218       MSE                0.232
## Pred R-Squared          0.203       MAE                0.355
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square      F         Sig.
## -------------------------------------------------------------------
## Regression     12.481          2          6.240    26.921    0.0000
## Residual       42.652        184          0.232
## Total          55.132        186
## -------------------------------------------------------------------
##
##                                        Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##                   model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------------
##             (Intercept)     2.099         0.103                 20.391    0.000     1.896     2.302
##          Surfactant_LBS     0.000         0.000       -0.293    -4.310    0.000     0.000     0.000
## FluidIntensity_BBLPerFT    -0.038         0.009       -0.299    -4.399    0.000    -0.055    -0.021
## ----------------------------------------------------------------------------------------------------
```

```
fullmodel<-lm(ProppantLoading_LBSPerGAL~Surfactant_LBS+I(FrictionReducer_LBS*First36MonthProd_BOE^8)+FluidIntensity_BBLPerFT
+PerfInterval_FT, data=WELLS)
formodel=ols_step_forward_p(fullmodel,penter = 0.001, details=TRUE)
```

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. Surfactant_LBS
## 2. I(FrictionReducer_LBS * First36MonthProd_BOE^8)
## 3. FluidIntensity_BBLPerFT
## 4. PerfInterval_FT
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - FluidIntensity_BBLPerFT
##
##                         Model Summary
## --------------------------------------------------------------
## R                       0.377       RMSE               0.513
## R-Squared               0.142       Coef. Var         36.389
## Adj. R-Squared          0.139       MSE                0.263
## Pred R-Squared          0.129       MAE                0.384
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     10.216          1         10.216    38.797    0.0000
## Residual       61.615        234          0.263
## Total          71.831        235
## --------------------------------------------------------------------
##
##                                         Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##                   model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------------
##             (Intercept)     1.932         0.090                 21.415    0.000     1.754     2.110
## FluidIntensity_BBLPerFT    -0.050         0.008       -0.377    -6.229    0.000    -0.066    -0.034
## ----------------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 2
##
## - Surfactant_LBS
##
##                         Model Summary
## --------------------------------------------------------------
## R                       0.476       RMSE               0.481
## R-Squared               0.226       Coef. Var         34.092
## Adj. R-Squared          0.218       MSE                0.232
## Pred R-Squared          0.203       MAE                0.355
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     12.481          2          6.240    26.921    0.0000
## Residual       42.652        184          0.232
## Total          55.132        186
## --------------------------------------------------------------------
##
##                                         Parameter Estimates
```

```
## ---------------------------------------------------------------------------------------------------
##                  model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ---------------------------------------------------------------------------------------------------
##            (Intercept)     2.099         0.103                 20.391    0.000     1.896     2.302
## FluidIntensity_BBLPerFT    -0.038         0.009       -0.299    -4.399    0.000    -0.055    -0.021
##         Surfactant_LBS     0.000         0.000       -0.293    -4.310    0.000     0.000     0.000
## ---------------------------------------------------------------------------------------------------
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + FluidIntensity_BBLPerFT
## + Surfactant_LBS
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## --------------------------------------------------------------
## R                       0.476       RMSE               0.481
## R-Squared               0.226       Coef. Var         34.092
## Adj. R-Squared          0.218       MSE                0.232
## Pred R-Squared          0.203       MAE                0.355
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     12.481         2          6.240    26.921    0.0000
## Residual       42.652       184          0.232
## Total          55.132       186
## --------------------------------------------------------------------
##
##                                         Parameter Estimates
## ---------------------------------------------------------------------------------------------------
##                  model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ---------------------------------------------------------------------------------------------------
##            (Intercept)     2.099         0.103                 20.391    0.000     1.896     2.302
## FluidIntensity_BBLPerFT    -0.038         0.009       -0.299    -4.399    0.000    -0.055    -0.021
##         Surfactant_LBS     0.000         0.000       -0.293    -4.310    0.000     0.000     0.000
## ---------------------------------------------------------------------------------------------------
```

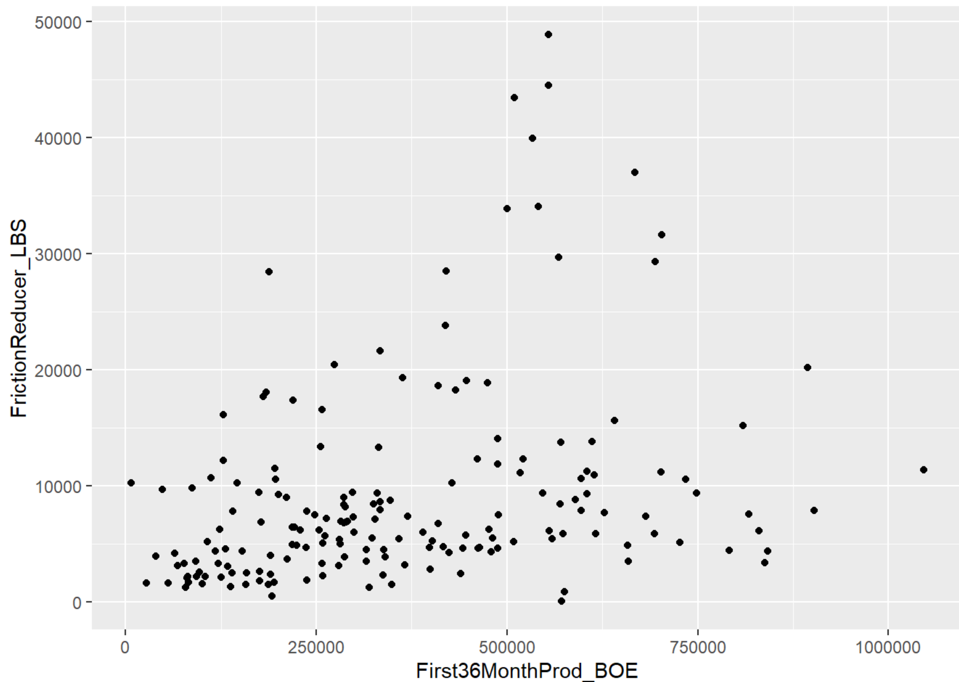
```
# We are only discussing models based on the Y value being 36 month production. This model was for determining how the backw
ard and forward elimination method would run if we changed our Y value. Here the Y value is Propant Loading and we can see t
hat the process runs more smoothly for the backward model and the forward model. In the forward model more terms/variables a
re kept like fluid intensity and surfactant. Even though the model seems to be more accurate it is not due to the Y value no
t being the main term in our model.
```

```
# we will do binomial dependent next
fullmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(TVD_FT*ScaleInhibitor_LBS^1)+ProppantIntens
ity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(data = WELLS, mapping = aes(x = First36MonthProd_BOE, y = FrictionReducer_LBS))+geom_point()+
 stat_smooth(method="glm",method.args=list(family="binomial"),se=FALSE)
```

```
## Warning: Removed 62 rows containing non-finite values (stat_smooth).
```

```
## Warning: Computation failed in `stat_smooth()`:
## y values must be 0 <= y <= 1
```

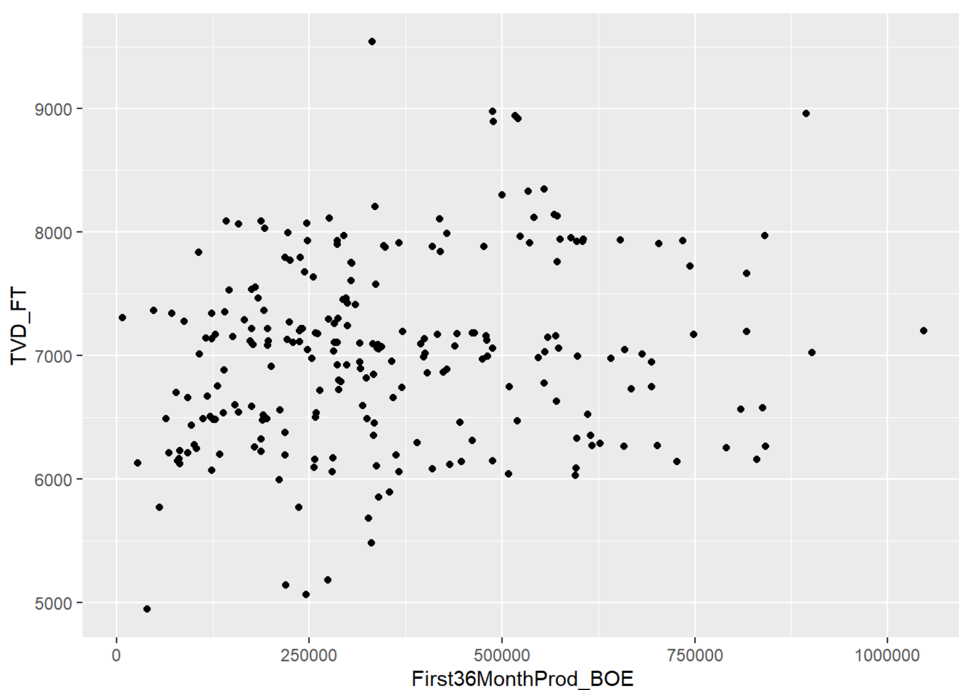
```
## Warning: Removed 62 rows containing missing values (geom_point).
```




```
# THe binary dependent model shows that the as production begins to increase along the x axis the variations between the friction reducer and 36 month productio increase. When production is below 250000 then the friction reducer is much less however after this poin it dramatically increases.
```

```
fullmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(TVD_FT*ScaleInhibitor_LBS^2)+ProppantIntensity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(data = WELLS, mapping = aes(x = First36MonthProd_BOE, y = TVD_FT ))+geom_point()+
 stat_smooth(method="glm",method.args=list(family="binomial"),se=FALSE)
```

```
## Warning: Computation failed in `stat_smooth()`:
## y values must be 0 <= y <= 1
```



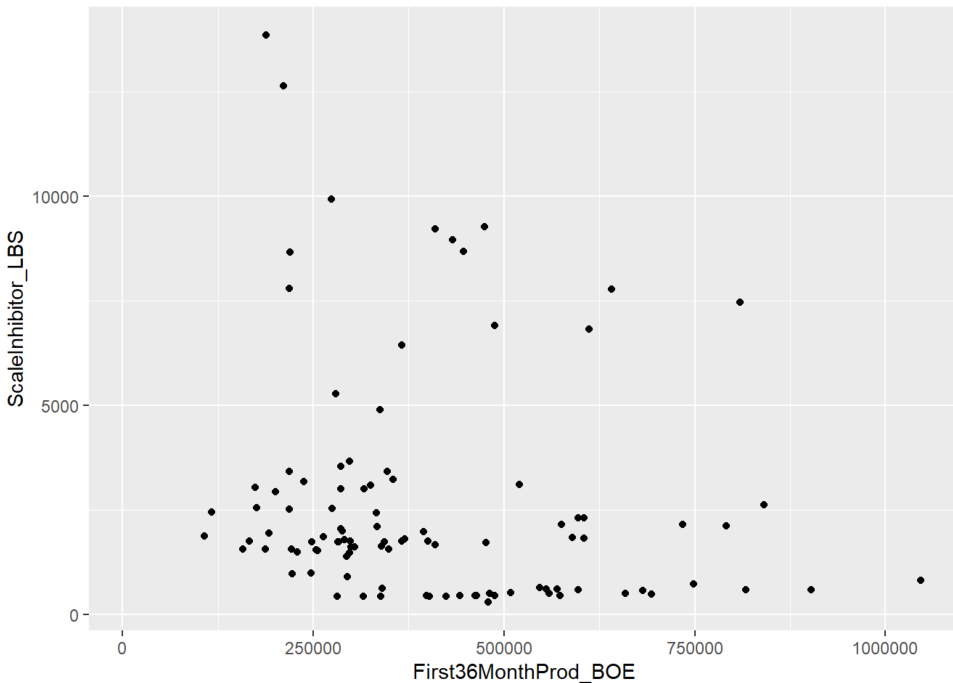

```
# As we can see from this binary dependent model indicates that when production is below 260000 the values are much closer together indicating that the variance is low however between production values of 275000 and approximately 510000 there seems to be more vairance. Overall though the values are evenly spread out and there is limited variance.
```

```
fullmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(TVD_FT*PerfInterval_FT^2)+ProppantIntensity
_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(data = WELLS, mapping = aes(x = First36MonthProd_BOE, y =ScaleInhibitor_LBS  ))+geom_point()+
 stat_smooth(method="glm",method.args=list(family="binomial"),se=FALSE)
```

```
## Warning: Removed 141 rows containing non-finite values (stat_smooth).
```

```
## Warning: Computation failed in `stat_smooth()`:
## y values must be 0 <= y <= 1
```

```
## Warning: Removed 141 rows containing missing values (geom_point).
```



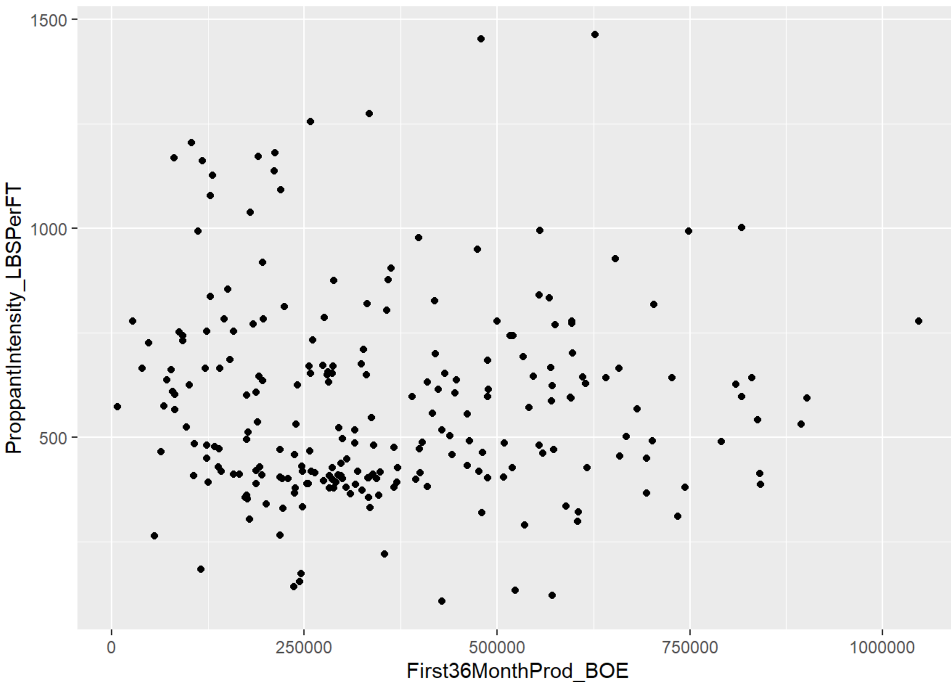

```
#In this binary dependent model between the 36 month production and scale inhibitor there are many outliers of data. As can
 be seen from the plot most of the points are below the 5000 scale inhibitor value and there are outliers above this thresho
ld. We also note that after 250000 production value the outliers begin to decrease in value.
```

```
fullmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(TVD_FT*ScaleInhibitor_LBS^1)+ProppantIntens
ity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(data = WELLS, mapping = aes(x = First36MonthProd_BOE, y = ProppantIntensity_LBSPerFT ))+geom_point()+
 stat_smooth(method="glm",method.args=list(family="binomial"),se=FALSE)
```

```
## Warning: Removed 3 rows containing non-finite values (stat_smooth).
```

```
## Warning: Computation failed in `stat_smooth()`:
## y values must be 0 <= y <= 1
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```



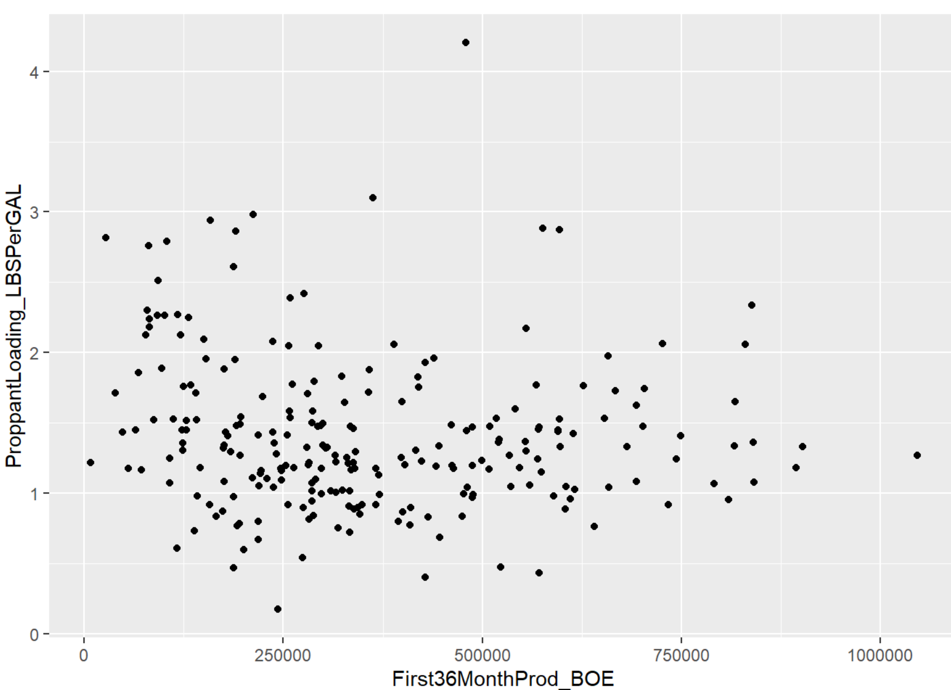

```
# From this binary dependent model we can see that most of the values lies between 250 and 1000 propant intensity while the
 rest would be considered outliers. We also notice that there is a slight increase to values as production increases. There
 are also regions in the plot where data points are clumped like around the 260000 prodcuion value. Finally we notice that t
he total amount variance stays the same.
```

```
fullmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*I(TVD_FT*ScaleInhibitor_LBS^1)+ProppantIntensity_LBSPerFT*Proppan
tLoading_LBSPerGAL, data=WELLS)
ggplot(data = WELLS, mapping = aes(x = First36MonthProd_BOE, y = ProppantLoading_LBSPerGAL ))+geom_point()+
 stat_smooth(method="glm",method.args=list(family="binomial"),se=FALSE)
```

```
## Warning: Removed 7 rows containing non-finite values (stat_smooth).
```

```
## Warning: Computation failed in `stat_smooth()`:
## y values must be 0 <= y <= 1
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```



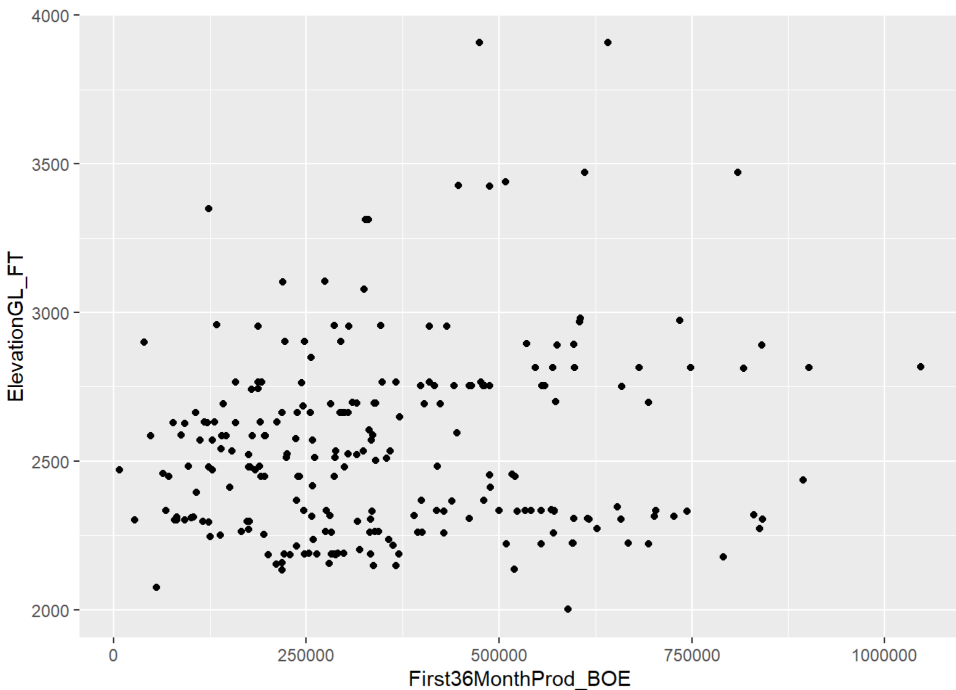
```
# Similar to the Propant Intensity plot the Propant Loading has values that are between 0.5 and 2 for propant loading while
 the rest can be considered outliers. Just as we had clumps in the previouos model around 250000 and 275000 there is a large
clump of data that indicates that this regions has the most events occuring. The variance in the data is also low between 0.
5 and 2 for propant loading.
```

```
fullmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(TVD_FT*ScaleInhibitor_LBS^1)+ProppantIntens
ity_LBSPerFT, data=WELLS)
ggplot(data = WELLS, mapping = aes(x = First36MonthProd_BOE, y = ElevationGL_FT ))+geom_point()+
 stat_smooth(method="glm",method.args=list(family="binomial"),se=FALSE)
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Computation failed in `stat_smooth()`:
## y values must be 0 <= y <= 1
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



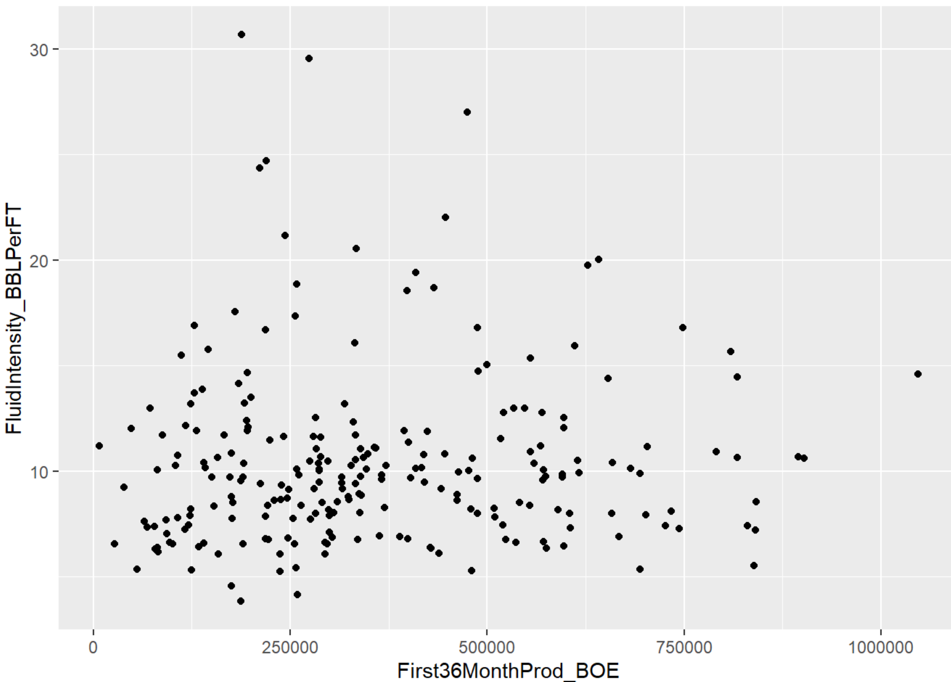

```
# From this binary dependent plot between elevation and production we notice that most of the points are between 2000 and 27
50 ft. This means that most of the production that is occuring will happen in this area of elevation. The points after 2750
 can be considered outliers. We also take into account that the variance in this region is relatively low.
```

```
fullmodel<-lm(First36MonthProd_BOE~FluidIntensity_BBLPerFT*FrictionReducer_LBS+I(TVD_FT*ScaleInhibitor_LBS^1)+ProppantIntens
ity_LBSPerFT*ProppantLoading_LBSPerGAL+PerfInterval_FT, data=WELLS)
ggplot(data = WELLS, mapping = aes(x = First36MonthProd_BOE, y = FluidIntensity_BBLPerFT ))+geom_point()+
 stat_smooth(method="glm",method.args=list(family="binomial"),se=FALSE)
```

```
## Warning: Removed 7 rows containing non-finite values (stat_smooth).
```

```
## Warning: Computation failed in `stat_smooth()`:
## y values must be 0 <= y <= 1
```

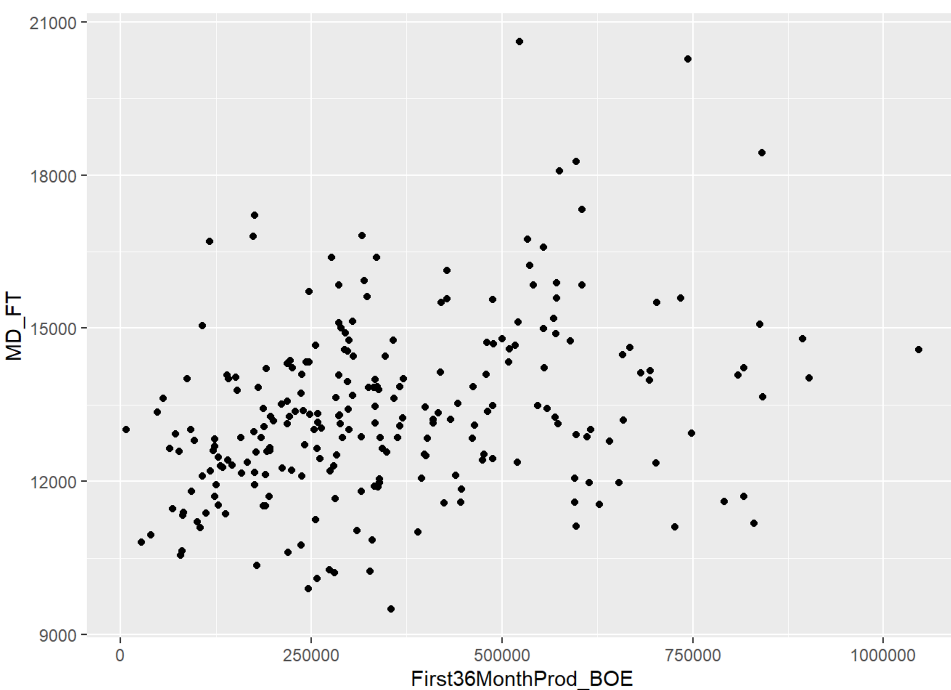
```
## Warning: Removed 7 rows containing missing values (geom_point).
```




```
# As can be seen from this binary dependent model most of the values lie between 5 and 15 for Fluid intensity and here seems to be little variance. Points that are considered after 15 can be considered outliers. From this model we can determine that most of the production occurs when fluid intensity is between 5 and 15.
```

```
fullmodel<-lm(First36MonthProd_BOE~FrictionReducer_LBS+I(TVD_FT*ScaleInhibitor_LBS^1)+ProppantIntensity_LBSPerFT*ProppantLoading_LBSPerGAL, data=WELLS)
ggplot(data = WELLS, mapping = aes(x = First36MonthProd_BOE, y = MD_FT ))+geom_point()+
 stat_smooth(method="glm",method.args=list(family="binomial"),se=FALSE)
```

```
## Warning: Computation failed in `stat_smooth()`:
## y values must be 0 <= y <= 1
```




```
# As can be seen from the binary dependent model we notice that most of the data points are present in the early part of the plot. Most of the data points are clustered between 0 and 375000 for production and 9000 and 150000 for MD meaning that most of the operations are within this region of the plot. We also take inot account that the variance in this regions is lower.
```