```r
library(readr)
data_train <- read_csv("data-train.csv")
```

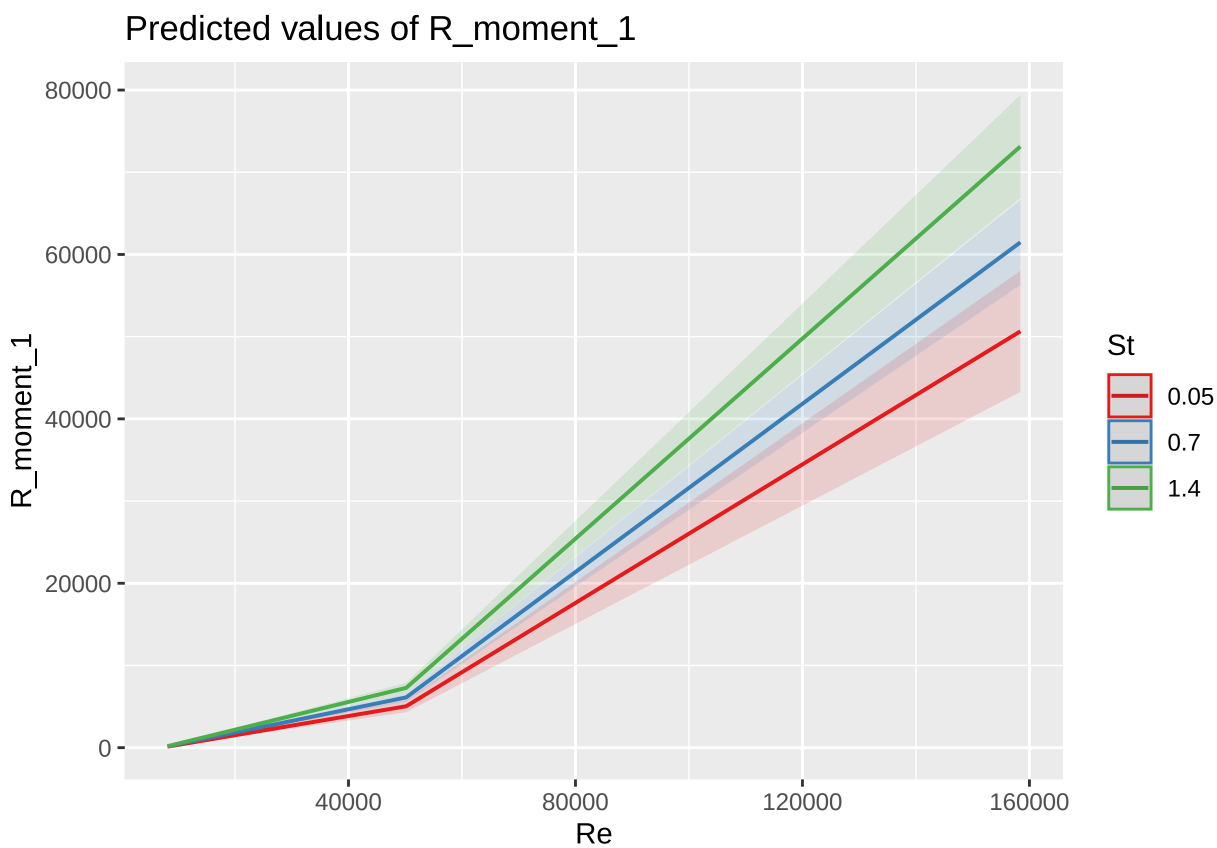
```
## Rows: 89 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (7): St, Re, Fr, R_moment_1, R_moment_2, R_moment_3, R_moment_4
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
attach(data_train)
data_train <- data_train%>% mutate(TFr = case_when(Fr>1~ .99999, Fr<1~Fr))
data_train<-data_train%>%mutate(TFr=logit(TFr))
```

```r
model_1<-lm(R_moment_1~I(Re^2)+St+Re+I(Re^2)*St+St:Re,data=data_train)
summary(model_1)
```

```
##
## Call:
## lm(formula = R_moment_1 ~ I(Re^2) + St + Re + I(Re^2) * St +
##     St:Re, data = data_train)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0274911 -0.0003305 -0.0000063  0.0001455  0.0298379
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.822e-01  7.757e-03  23.492  < 2e-16 ***
## I(Re^2)      1.993e-06  1.553e-07  12.838  < 2e-16 ***
## St           6.143e-02  6.487e-03   9.470 7.50e-15 ***
## Re          -1.250e-03  7.592e-05 -16.471  < 2e-16 ***
## I(Re^2):St   6.666e-07  1.364e-07   4.887 4.93e-06 ***
## St:Re       -4.195e-04  6.588e-05  -6.368 1.01e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01127 on 83 degrees of freedom
## Multiple R-squared:  0.9616, Adjusted R-squared:  0.9593
## F-statistic: 415.7 on 5 and 83 DF,  p-value: < 2.2e-16
```

```r
plot_model(model_1, type = "pred", terms = c("Re","St[.05,.7,1.4]"))
```



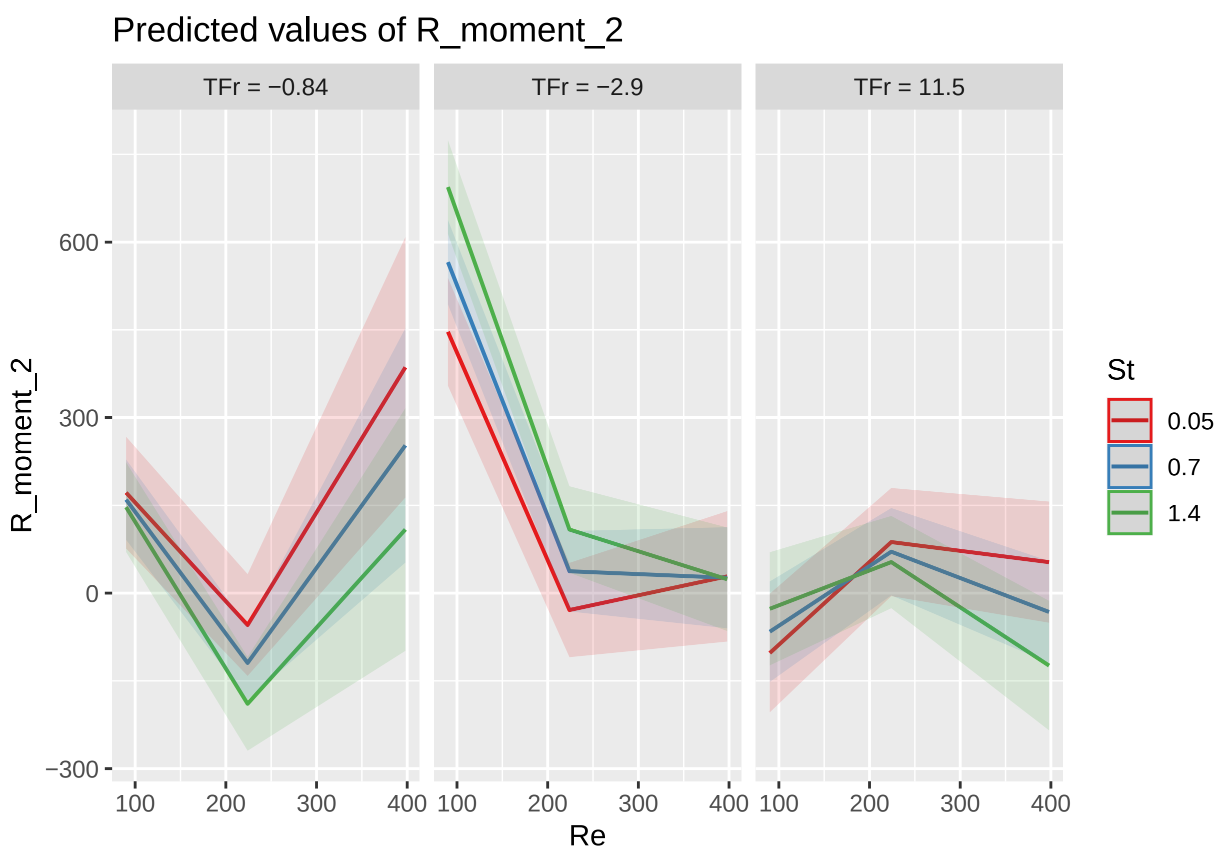
This plot of the effect of Re on the first moment using the nonlinear model with interactions sheds light on the relationship between Re and the first moment. First, as we determined with the simple linear model, St has a positive relationship with the first moment. The impact of Re on the first moment is higher at each level of Re if St increases. However, this plot shows that as Re grows beyond roughly 50000, its marginal influence on the first moment increases. Thus, at very high values R seems to have a nonlinear relationship to the average expected amount of clustering.

```
model_2<-lm(R_moment_2~Re*TFr*St+I(TFr^2):I(Re^2)+I(TFr^2):St +I(TFr^2)+I(Re^2)-Re:TFr:St,data=data_trai
summary(model_2)
```

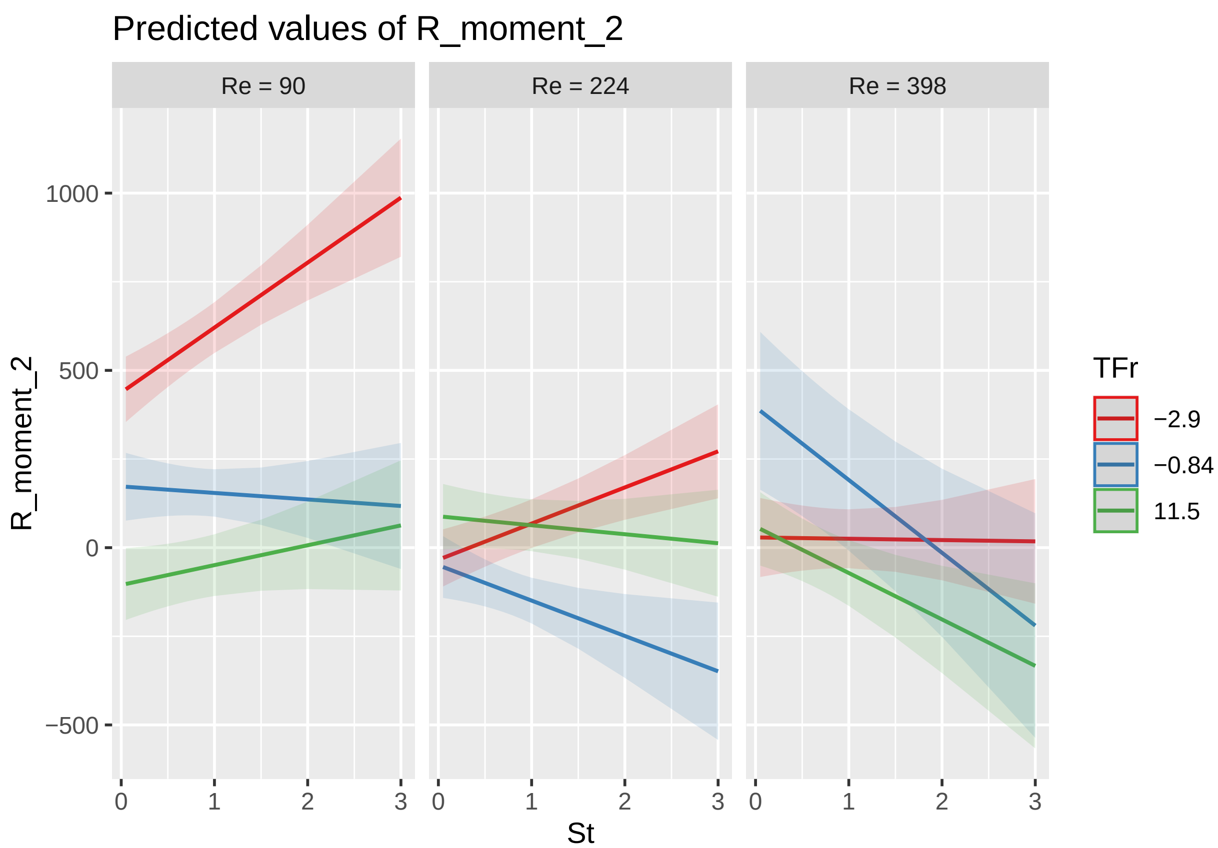
```
##
## Call:
## lm(formula = R_moment_2 ~ Re * TFr * St + I(TFr^2):I(Re^2) +
##     I(TFr^2):St + I(TFr^2) + I(Re^2) - Re:TFr:St, data = data_train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -446.53  -67.92  -17.88   79.33  283.59
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.519e+02  9.369e+01   4.824 6.84e-06 ***
## Re               -5.346e+00  7.367e-01  -7.256 2.55e-10 ***
## TFr              -1.670e+02  2.530e+01  -6.600 4.49e-09 ***
## St               -2.838e+01  5.586e+01  -0.508 0.612807
## I(TFr^2)          8.527e+00  2.235e+00   3.815 0.000271 ***
```

```
## I(Re^2)           1.381e-02  2.023e-03   6.825 1.70e-09 ***
## Re:TFr             7.336e-01  1.024e-01   7.165 3.81e-10 ***
## Re:St             -6.071e-01  1.722e-01  -3.525 0.000713 ***
## TFr:St            -7.090e+01  1.746e+01  -4.060 0.000116 ***
## I(TFr^2):I(Re^2) -1.440e-04  2.371e-05  -6.074 4.27e-08 ***
## St:I(TFr^2)        7.216e+00  1.910e+00   3.778 0.000307 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 134.5 on 78 degrees of freedom
## Multiple R-squared:  0.766,  Adjusted R-squared:  0.736
## F-statistic: 25.53 on 10 and 78 DF,  p-value: < 2.2e-16
```

```r
plot_model(model_2, type = "pred", terms = c("Re", "St[.05,.7,1.4]", "TFr[-.84,-2.9,11.5]"))
```




```r
plot_model(model_2, type = "pred", terms = c("St", "TFr[-.84,-2.9,11.5]", "Re[90,224,398]"))
```



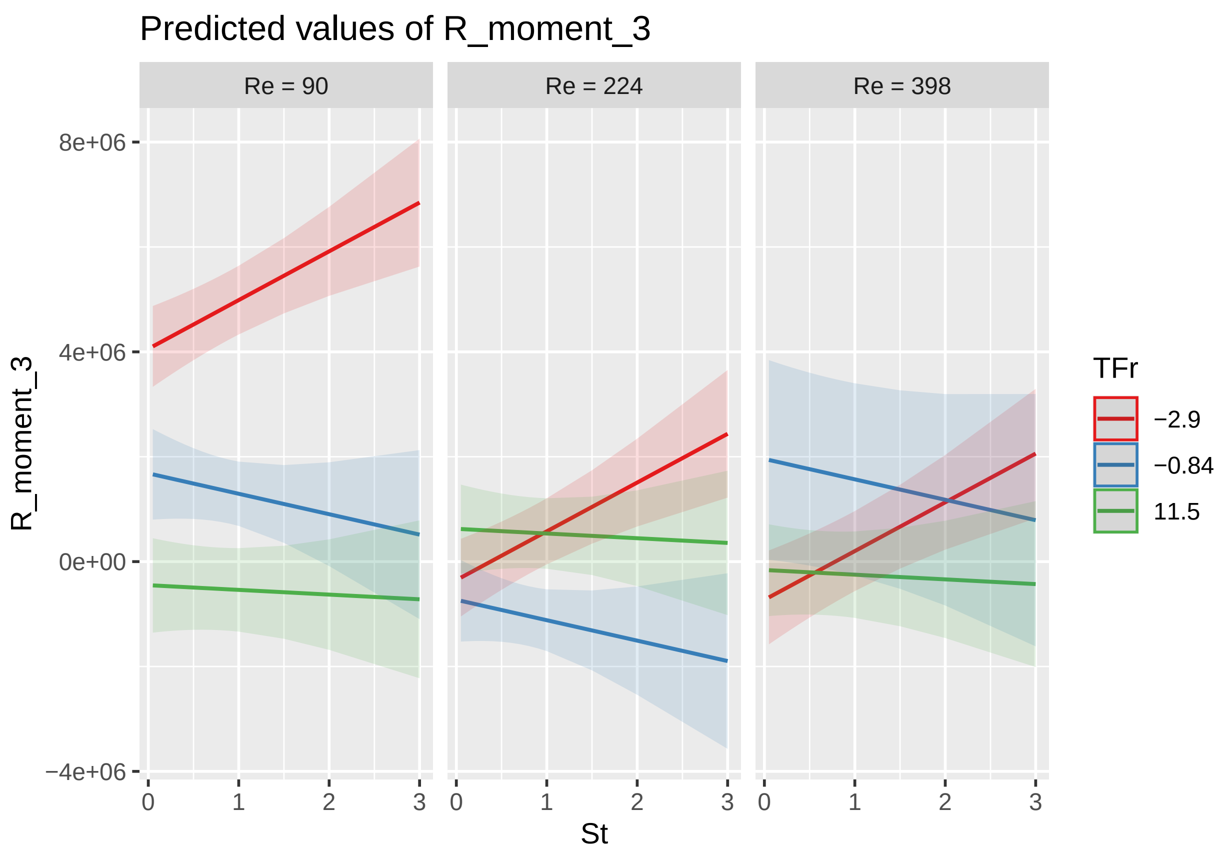
First plots: These plots show us a relationship that later holds true in the second and third moments as well. First, we can see by the blue and green lines that regardless of the level of Re and St, higher levels of Fr decrease the variance of clustering. We saw this relationship in the previous set of plots. The most interesting line to observe is the red one that shows what happens to variance as St changes at near-zero Fr. As Re increases, the effect of St converges to essentially nothing. Combined with what we learn about the first moment (that at extremely high levels of Re, average clustering increases), perhaps this means that Re becomes the most important determinant of clustering as it reaches high levels. At very high Re, we may see consistent clustering regardless of the other variables. Nonetheless, if Re is low (90 or 224), St contributes to increases in variance. Thus, at low Re and Fr, not only does St increase average clustering (as we observed in the linear models), it does so with a great deal of variance. This probably reflects the fact that particles colliding does increase in frequency as they grow in size, but collisions are very unpredictable and random events; sometimes they lead to clustering and other times they do not.

Second plots: From these plots of the second model, we can first notice a significant amount of overlap of error bounds in all three plots regardless of the level of St. Thus, when it comes to variance, the influence of the size of the particles is not particularly important. We can also see that the convergence of the fits tightens at TFr = 11.5 (Fr = inf). Because Fr = u/sqrt(gL), it makes sense that u (flow velocity) must be relatively large for Fr to be infinite (find any source). At high rates of flow may consistently break up the clusters and thus preventing the occasional examples of high levels of clustering. We also see that high Fr may limit the influence of Re. At near-zero Fr (TFr = -.84), high Re has a strong positive relationship with variance. Additionally, Re seems to exhibit some kind of thresholding behavior past Re = 224 where its relationship with variance reverses past this point. Re = 224 appears to be a special value worth studying more in the future.

```
model_3<-lm(R_moment_3~I(TFr^2)*I(Re^2)+St*TFr+Re*TFr+I(TFr^2):St,data=data_train)
summary(model_3)
```

```
##
## Call:
## lm(formula = R_moment_3 ~ I(TFr^2) * I(Re^2) + St * TFr + Re *
##     TFr + I(TFr^2):St, data = data_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4107519  -580683   194083   640231  3090996
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.236e+06  8.449e+05   5.013 3.21e-06 ***
## I(TFr^2)          7.720e+04  2.021e+04   3.820 0.000265 ***
## I(Re^2)           1.093e+02  1.856e+01   5.890 8.99e-08 ***
## St               -8.149e+05  4.841e+05  -1.683 0.096237 .
## TFr              -1.429e+06  2.300e+05  -6.214 2.28e-08 ***
## Re               -4.711e+04  6.672e+03  -7.062 5.69e-10 ***
## I(TFr^2):I(Re^2) -1.134e+00  2.175e-01  -5.216 1.43e-06 ***
## St:TFr           -4.678e+05  1.523e+05  -3.072 0.002917 **
## TFr:Re            5.905e+03  9.404e+02   6.280 1.72e-08 ***
## I(TFr^2):St       4.616e+04  1.667e+04   2.769 0.007001 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1236000 on 79 degrees of freedom
## Multiple R-squared:  0.7122, Adjusted R-squared:  0.6794
## F-statistic: 21.72 on 9 and 79 DF,  p-value: < 2.2e-16
```

```r
plot_model(model_3, type = "pred", terms = c("St", "TFr[-.84,-2.9,11.5]", "Re[90,224,398]"))
```



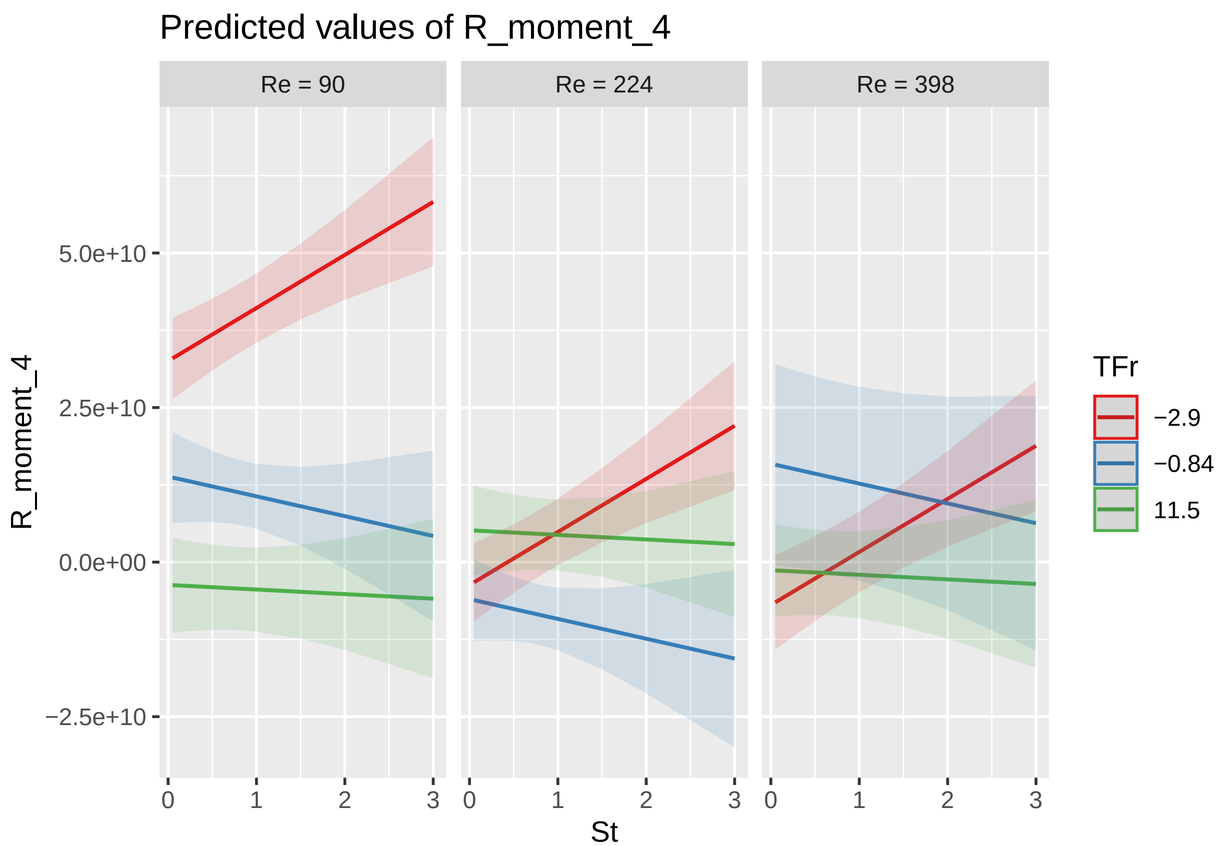
Once again, we see essentially the same relationship between St and the third moment as we do between St and the second moment. Thus, higher Fr and higher Re decreases skewness. However, the slope of the red lines is always positive. Thus, in this case, St appears to lengthen the right handed tail even as variance decreases due to Fr and Re. Thus, we can see that in a limited way, particle size increases the chances of rare high-clustering events within certain fixed conditions (where variance is limited overall by Fr and Re).

```
model_4<-lm(R_moment_4~I(TFr^2)*I(Re^2)+St*TFr+Re*TFr+I(TFr^2):St,data=data_train)
summary(model_4)
```

```
##
## Call:
## lm(formula = R_moment_4 ~ I(TFr^2) * I(Re^2) + St * TFr + Re *
##     TFr + I(TFr^2):St, data = data_train)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -3.298e+10 -4.906e+09  1.731e+09  5.650e+09  2.784e+10
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          3.504e+10  7.233e+09   4.845 6.20e-06 ***
## I(TFr^2)             6.062e+08  1.730e+08   3.503 0.000759 ***
## I(Re^2)              8.955e+05  1.589e+05   5.634 2.60e-07 ***
## St                  -7.011e+09  4.144e+09  -1.692 0.094617 .
## TFr                 -1.144e+10  1.969e+09  -5.811 1.25e-07 ***
## Re                  -3.864e+08  5.711e+07  -6.765 2.10e-09 ***
```

```
## I(TFr^2):I(Re^2) -9.301e+03  1.862e+03  -4.996 3.43e-06 ***
## St:TFr           -4.185e+09  1.304e+09  -3.211 0.001917 **
## TFr:Re            4.847e+07  8.050e+06   6.021 5.17e-08 ***
## I(TFr^2):St       4.113e+08  1.427e+08   2.882 0.005081 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.058e+10 on 79 degrees of freedom
## Multiple R-squared:  0.6992, Adjusted R-squared:  0.665
## F-statistic: 20.41 on 9 and 79 DF,  p-value: < 2.2e-16
```

```
plot_model(model_4, type = "pred", terms = c("St", "TFr[-.84,-2.9,11.5]", "Re[90,224,398]"))
```



The relationship between the predictors and kurtosis essentially seems identical to their relationship with skewness. Nonetheless, the Re = 398 plot is interesting because in this same plot for the second moment, we saw the effects of all variables converge to zero. However, St always has a positive effect on kurtosis as long as Fr is low. Thus, the size of the particles still makes the tails of the probability distribution of clustering heavier. Similar to our analysis of St's relationship to the third moment, it seems to increase the spread of the distribution within bounds set by the other parameters. Thus, as the other parameters make tail weight smaller overall, higher St makes them as small as possible within those bounds. I believe that this makes sense because Fr and Re seem to be more related to the environment than St, which has to do with the particles themselves. Given certain environmental parameters that limit the amount of variance overall, larger particle size will always increase the chance for rare events of high levels of clustering within the bounds of possibility.

#Conclusion: In conclusion, we have learned that St and the first moment have a mostly positive and linear relationship, although predictive accuracy can increase if we use a more complex model that includes interactions. Through the model with interactions, we saw that extremely high values of Re will increase average clustering. The other moments are best fitted with models that involve complex interactions and exponential terms due to their inherent nonlinearity. In general, high levels of Fr and Re decrease variance, skewness, and kurtosis. These parameters seem to make the results of the simulation more regular and consistent. On the other hand, St pushes for more positive skewness and kurtosis. I believe that this probably has to do with the inherently chaotic and unpredictable nature of collisions between particles. Lastly, with regards to variance, we saw Re = 224 exhibits some kind of thresholding behavior because its effect on variance switches at that point. In addition to this mystery, the specific relationship between Re and Fr is also worth further consideration. Intuitively, as Fr increases I would expect clustering behavior to become more regular because high flow probably cuts short any particularly unusual behavior of particles. However, Re when combined with Fr also seems to concentrate the probability of clustering around the mean. Theoretically, I would expect Re to both increase mean clustering and lead to less predictable behavior. Perhaps our results mean that high levels of Re in fact leads to more clustering, albeit in very consistent ways.