Turbulence Analysis

Tingnan Hu, Peter Liu, Islina Shan, Ken Ye

2023-10-22

Load Data & Libraries

```
train <- read.csv("data-train.csv")
test <- read.csv("data-test.csv")
library(ggplot2)
```

```
head(train)
```

```
##     St  Re    Fr R_moment_1 R_moment_2 R_moment_3 R_moment_4
## 1 0.10 224 0.052 0.00215700  0.1303500   14.37400  1586.5000
## 2 3.00 224 0.052 0.00379030  0.4704200   69.94000 10404.0000
## 3 0.70 224   Inf 0.00290540  0.0434990    0.82200    15.5510
## 4 0.05  90   Inf 0.06352800  0.0906530    0.46746     3.2696
## 5 0.70 398   Inf 0.00036945  0.0062242    0.12649     2.5714
## 6 2.00  90 0.300 0.14780000  2.0068000   36.24900   671.6700
```

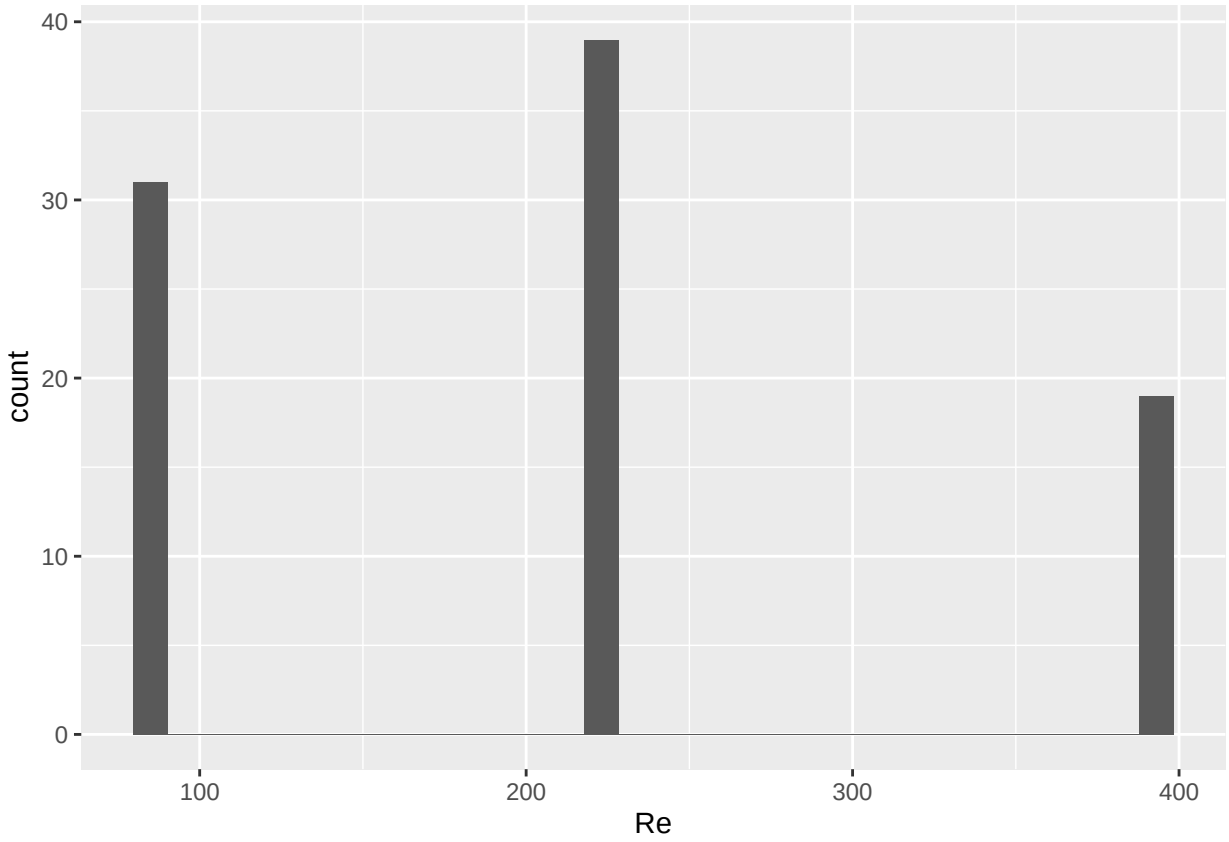
EDA

```
summary(train)
```

```
##        St               Re              Fr          R_moment_1
##  Min.   :0.0500   Min.   : 90.0   Min.   :0.052   Min.   :0.000222
##  1st Qu.:0.3000   1st Qu.: 90.0   1st Qu.:0.052   1st Qu.:0.002157
##  Median :0.7000   Median :224.0   Median :0.300   Median :0.002958
##  Mean   :0.8596   Mean   :214.5   Mean   :  Inf   Mean   :0.040394
##  3rd Qu.:1.0000   3rd Qu.:224.0   3rd Qu.:  Inf   3rd Qu.:0.087868
##  Max.   :3.0000   Max.   :398.0   Max.   :  Inf   Max.   :0.172340
##    R_moment_2          R_moment_3        R_moment_4
##  Min.   :   0.0001   Min.   :      0   Min.   :0.000e+00
##  1st Qu.:   0.0245   1st Qu.:      0   1st Qu.:3.000e+00
##  Median :   0.0808   Median :      1   Median :2.100e+01
##  Mean   :  92.4902   Mean   : 753370   Mean   :6.194e+09
##  3rd Qu.:   0.5345   3rd Qu.:     40   3rd Qu.:5.345e+03
##  Max.   :1044.3000   Max.   :9140000   Max.   :8.000e+10
```

```r
# Create histograms for the predictor variables
ggplot(train, aes(x = Re)) + geom_histogram()
```

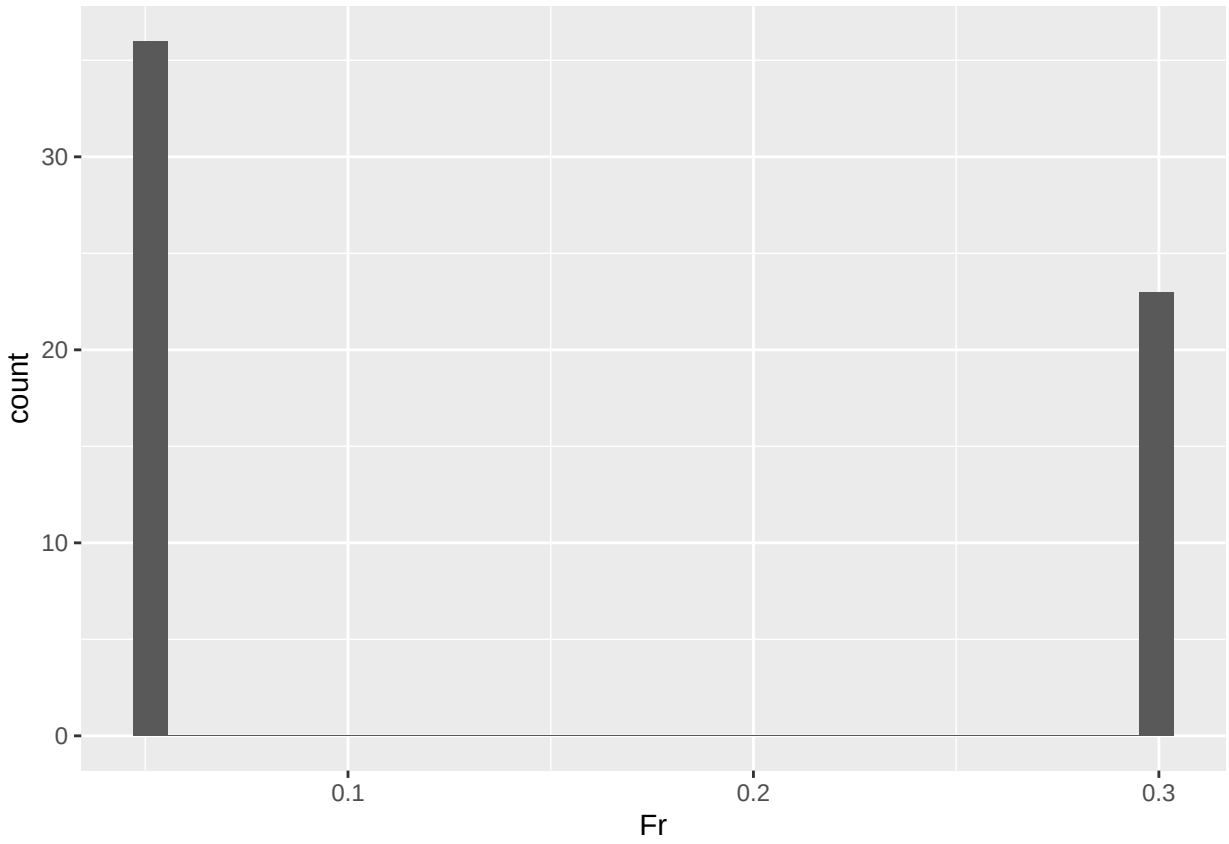
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```r
ggplot(train, aes(x = Fr)) + geom_histogram()
```

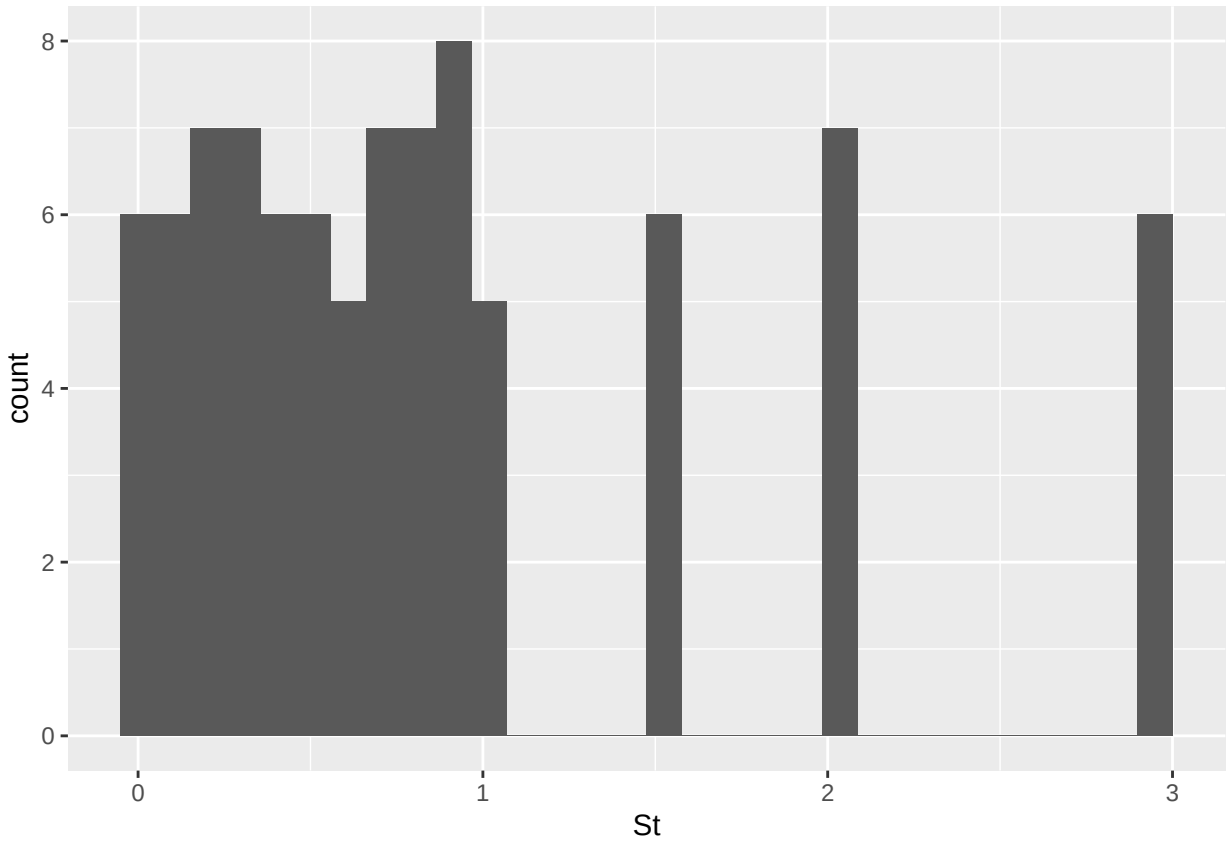
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 30 rows containing non-finite values (`stat_bin()`).
```

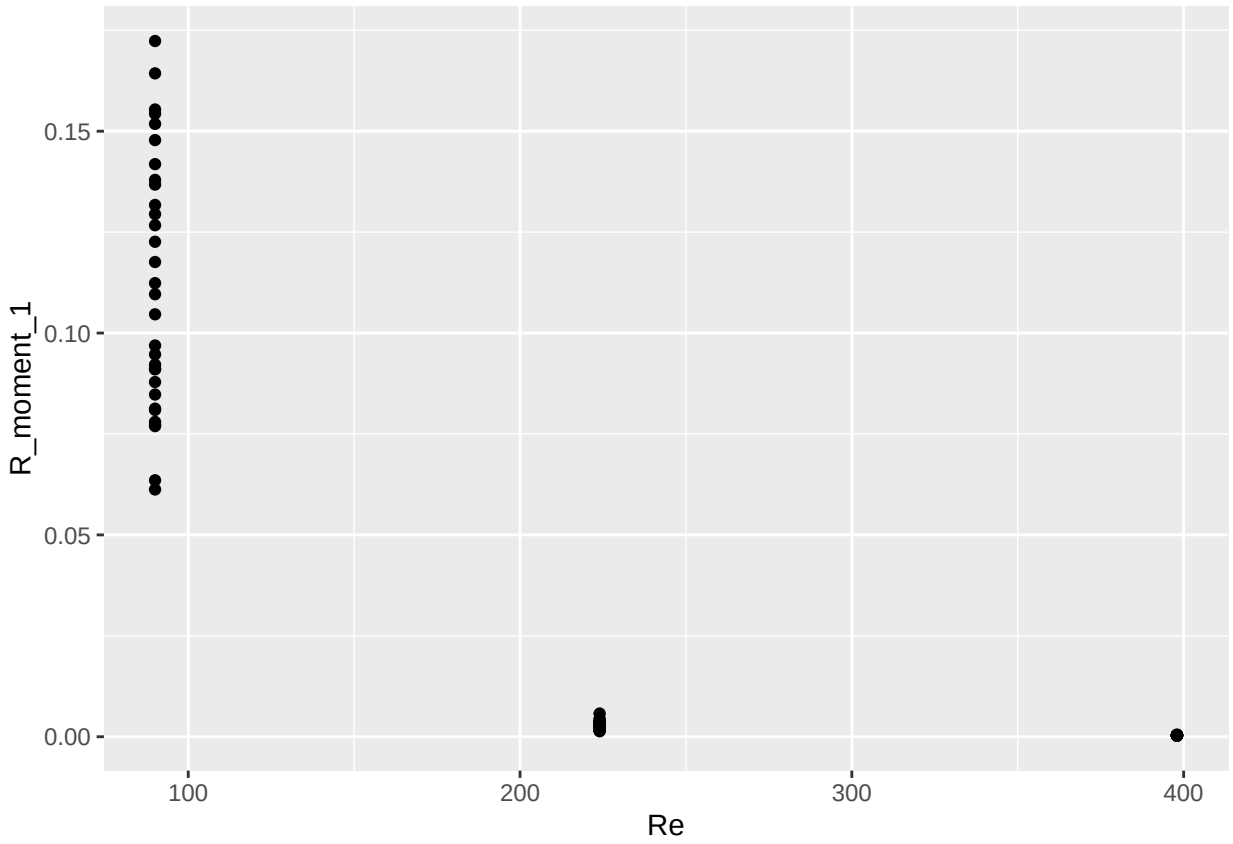



```
ggplot(train, aes(x = St)) + geom_histogram()
```

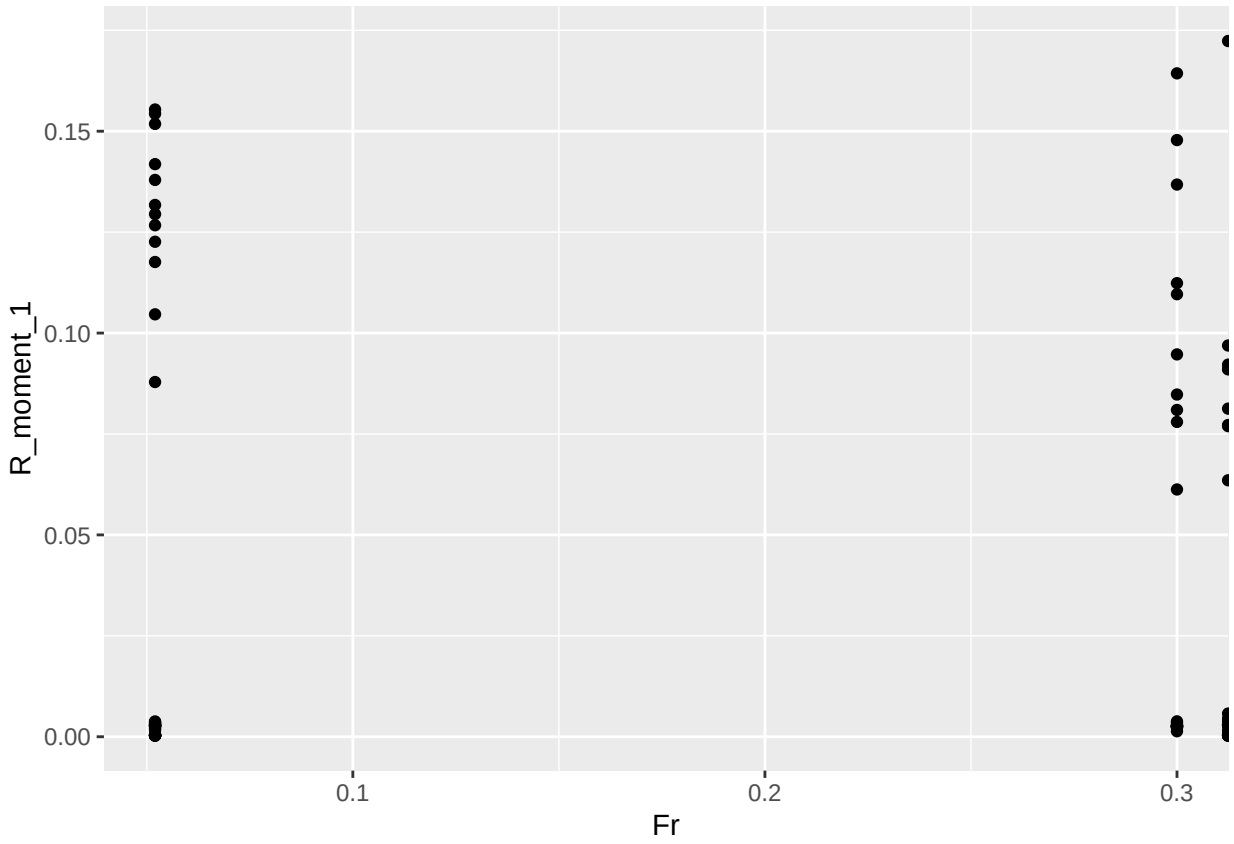
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



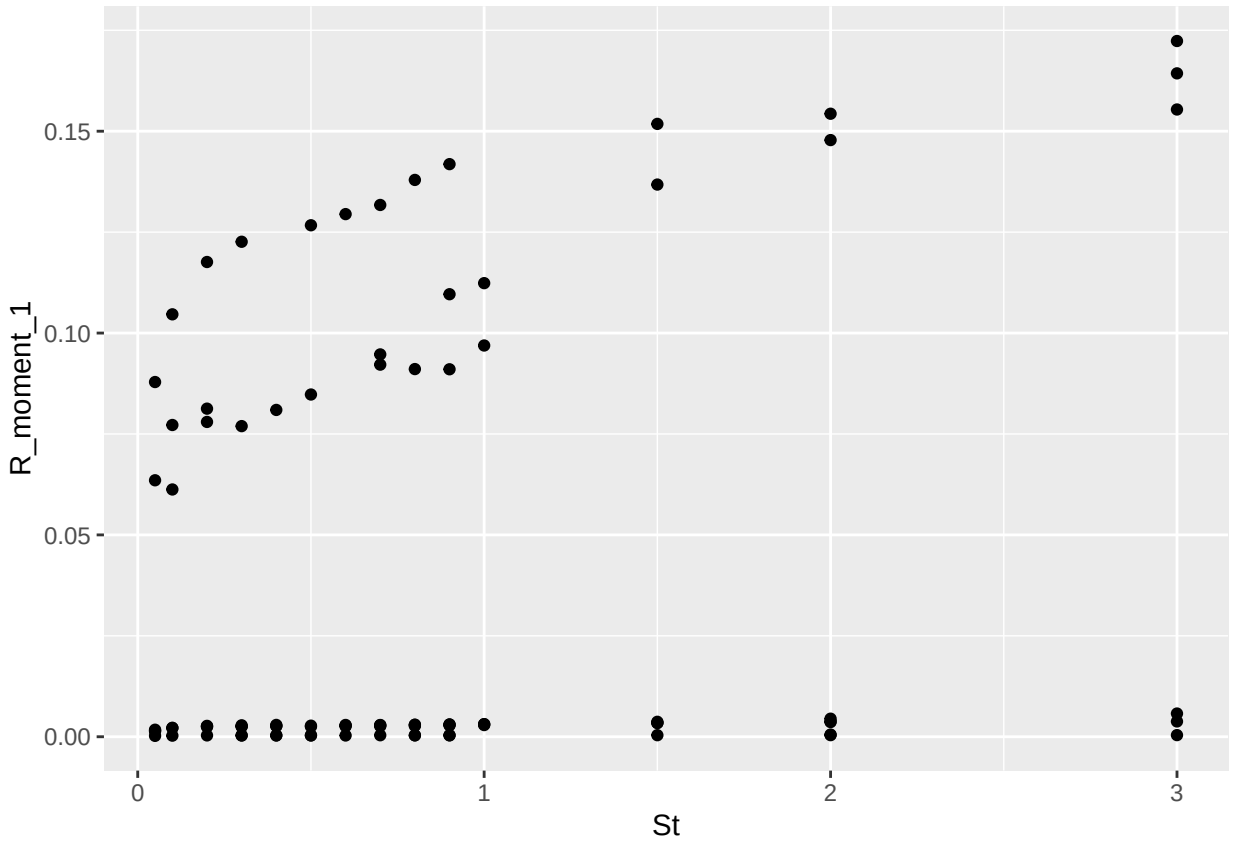

```
# Create scatterplots to explore relationships between predictor variables with R_moment_1
ggplot(train, aes(x = Re, y = R_moment_1)) + geom_point()
```



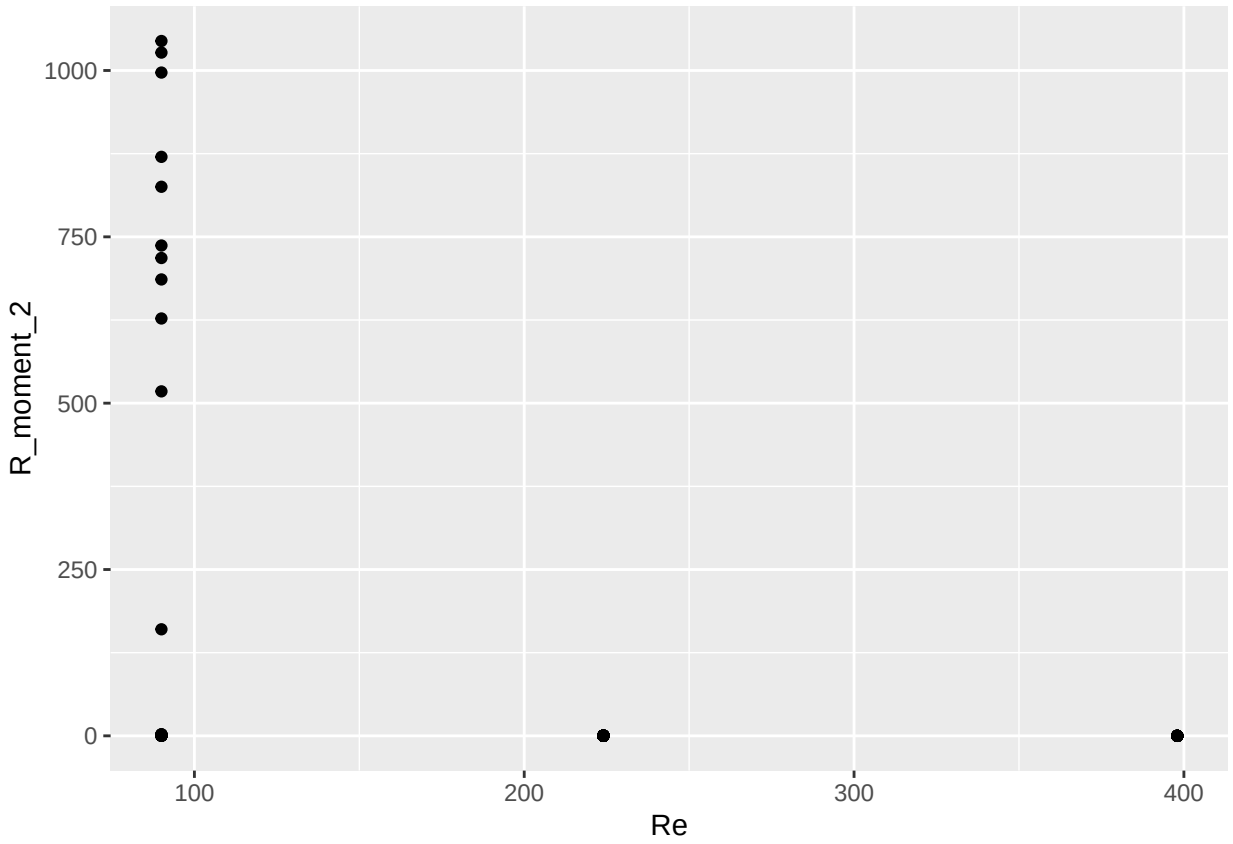

```
ggplot(train, aes(x = Fr, y = R_moment_1)) + geom_point()
```



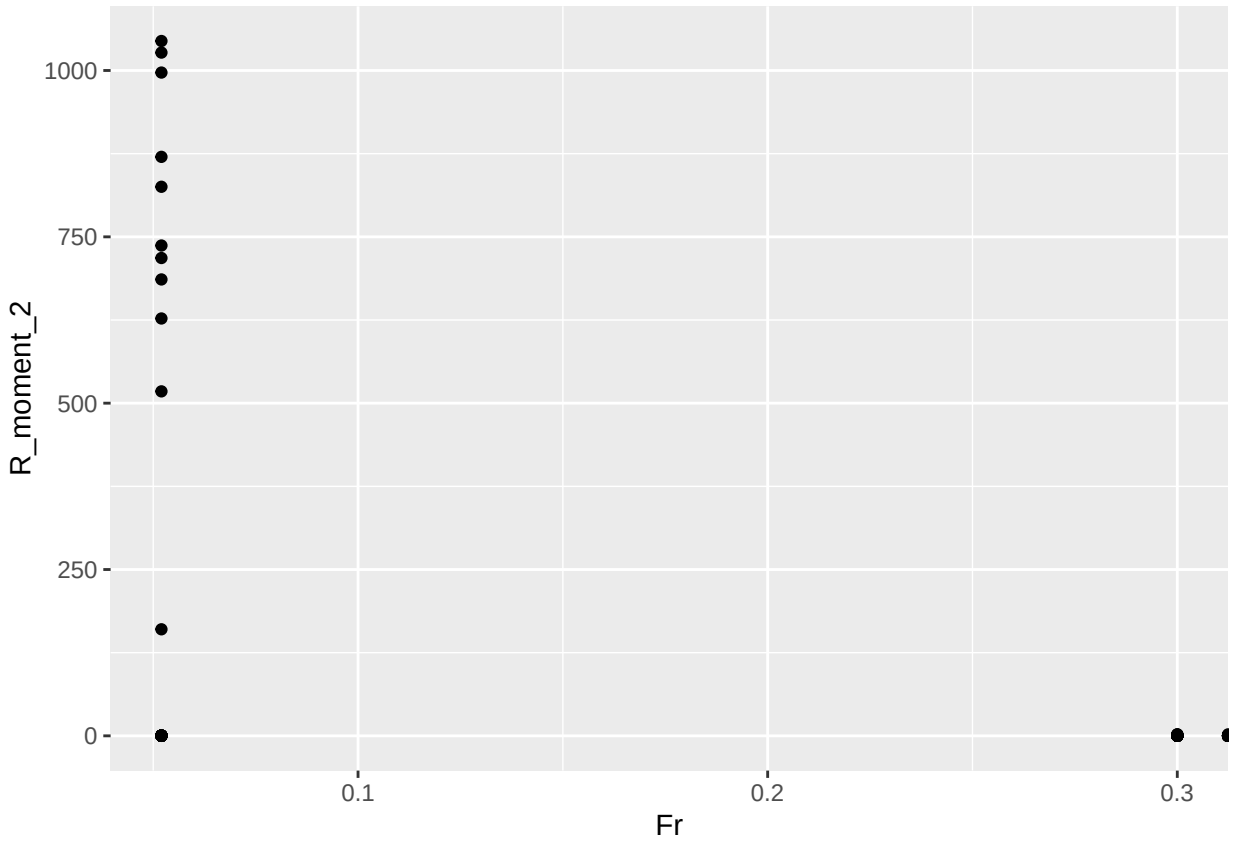

```
ggplot(train, aes(x = St, y = R_moment_1)) + geom_point()
```



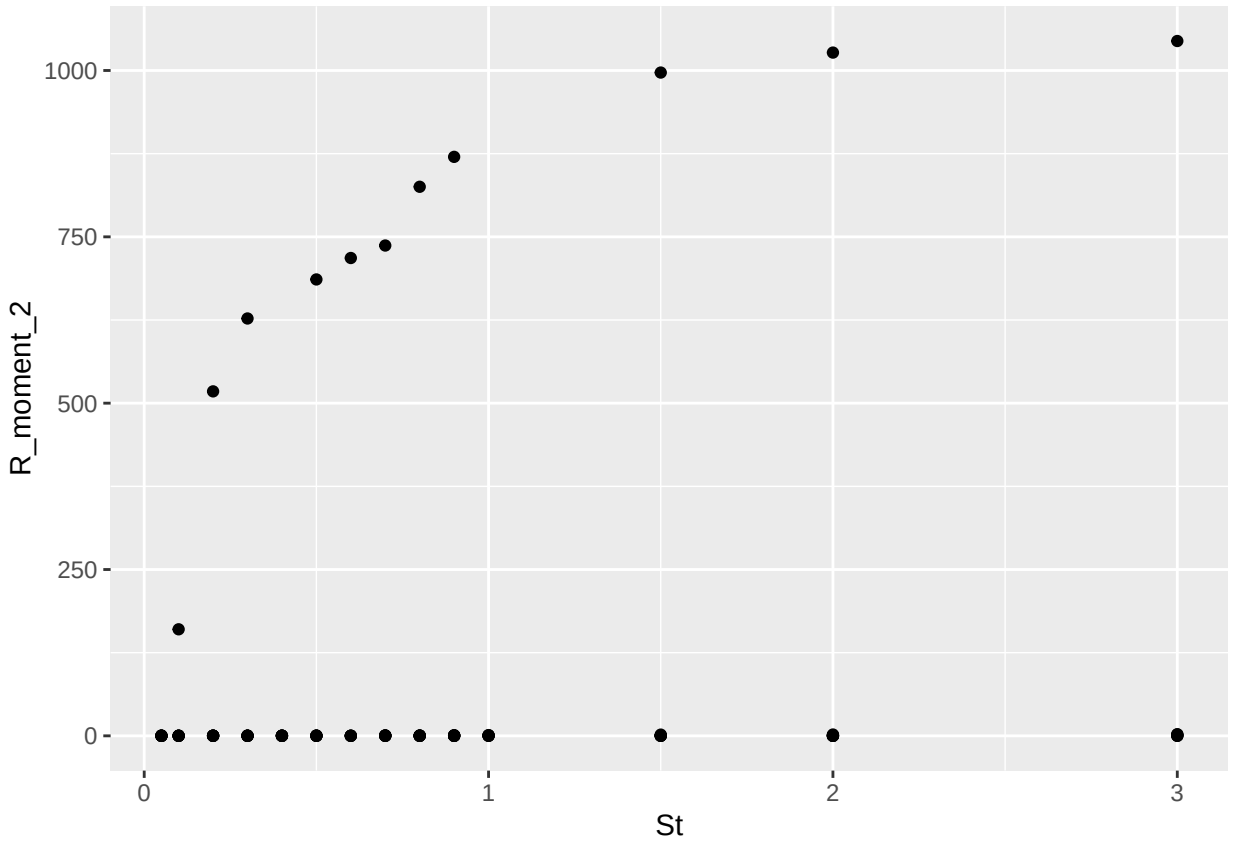

```
# Create scatterplots to explore relationships between predictor variables with R_moment_2
ggplot(train, aes(x = Re, y = R_moment_2)) + geom_point()
```



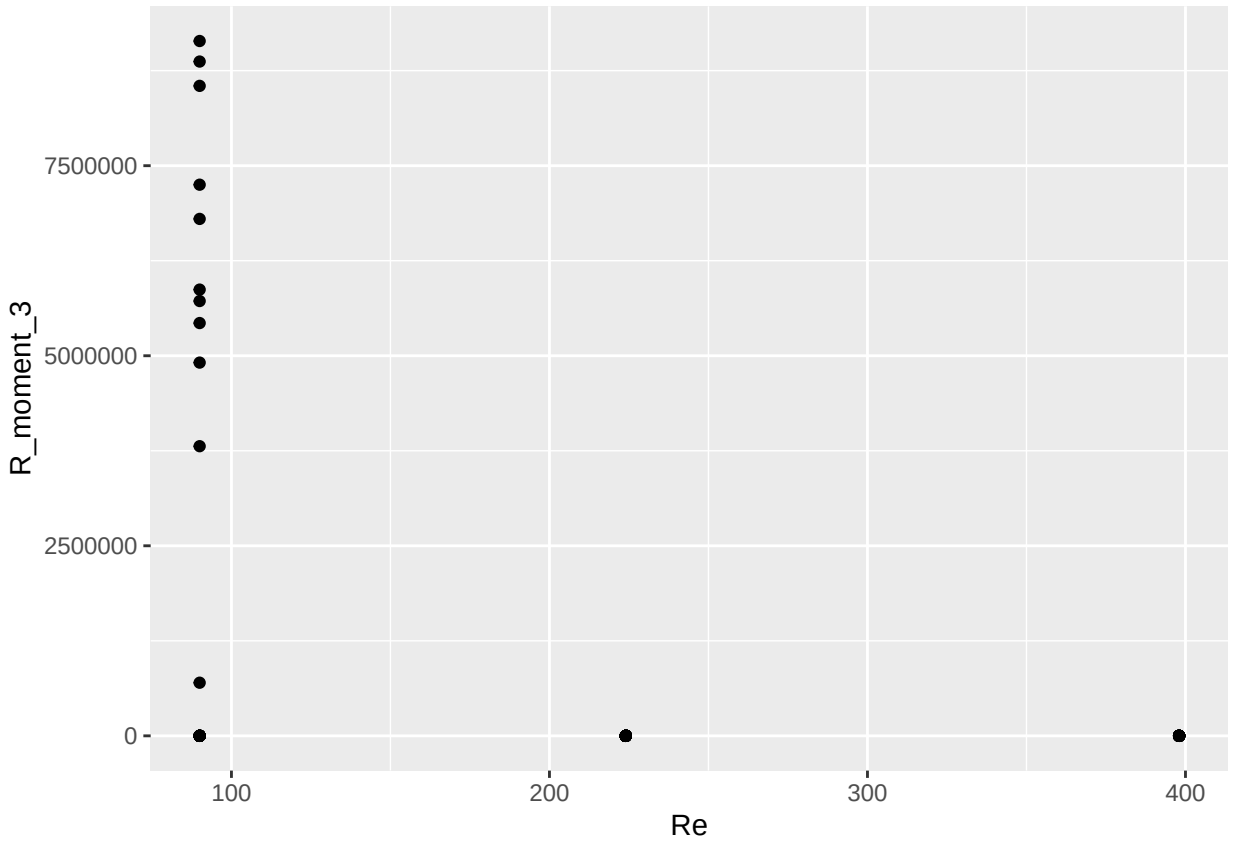

```
ggplot(train, aes(x = Fr, y = R_moment_2)) + geom_point()
```



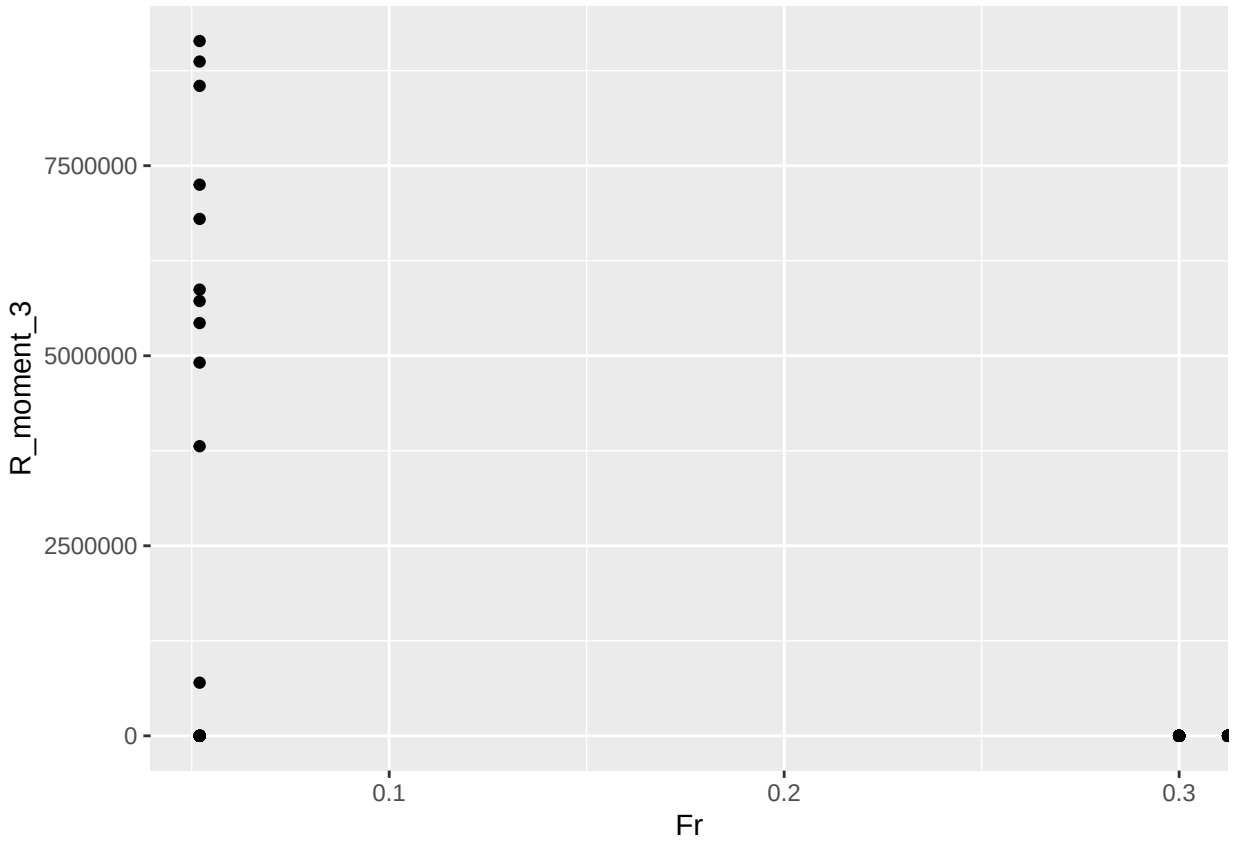

```
ggplot(train, aes(x = St, y = R_moment_2)) + geom_point()
```



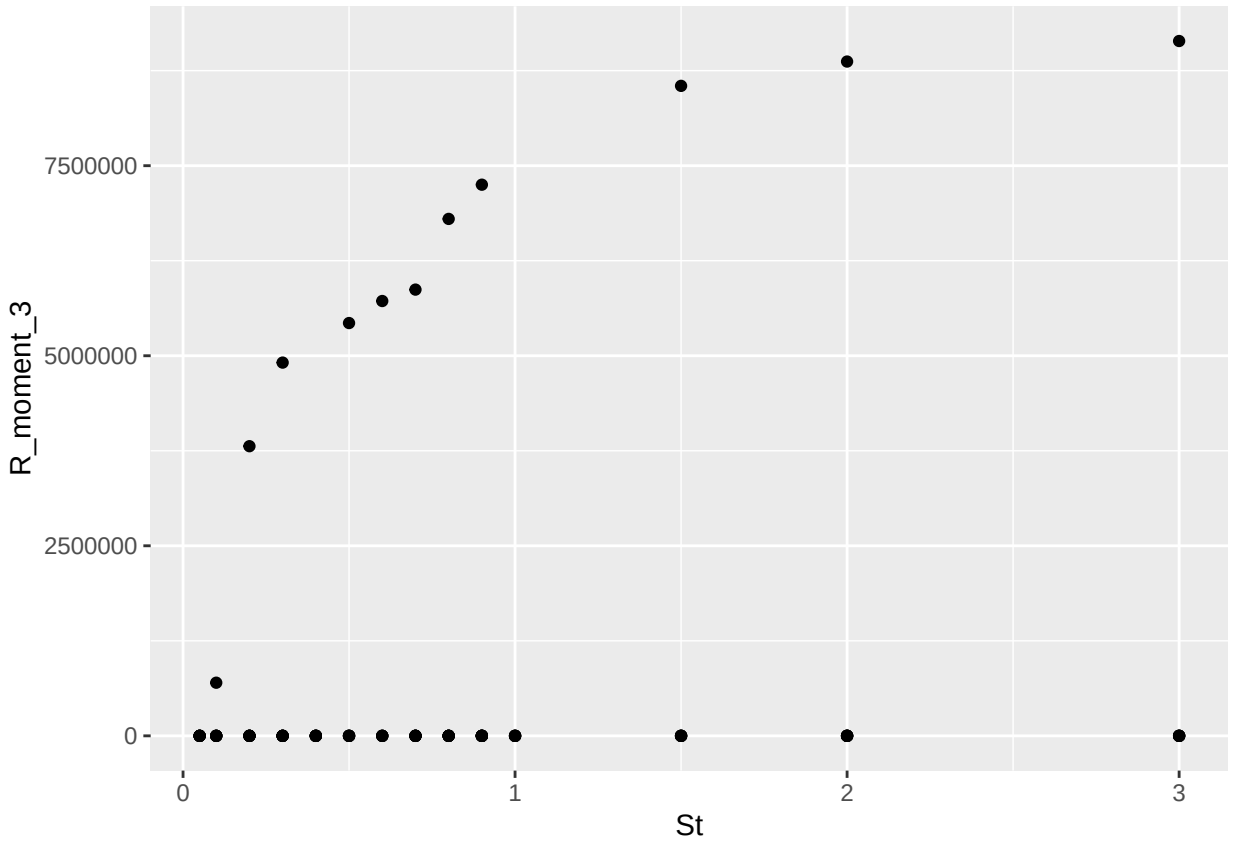

```
# Create scatterplots to explore relationships between predictor variables with R_moment_3
ggplot(train, aes(x = Re, y = R_moment_3)) + geom_point()
```



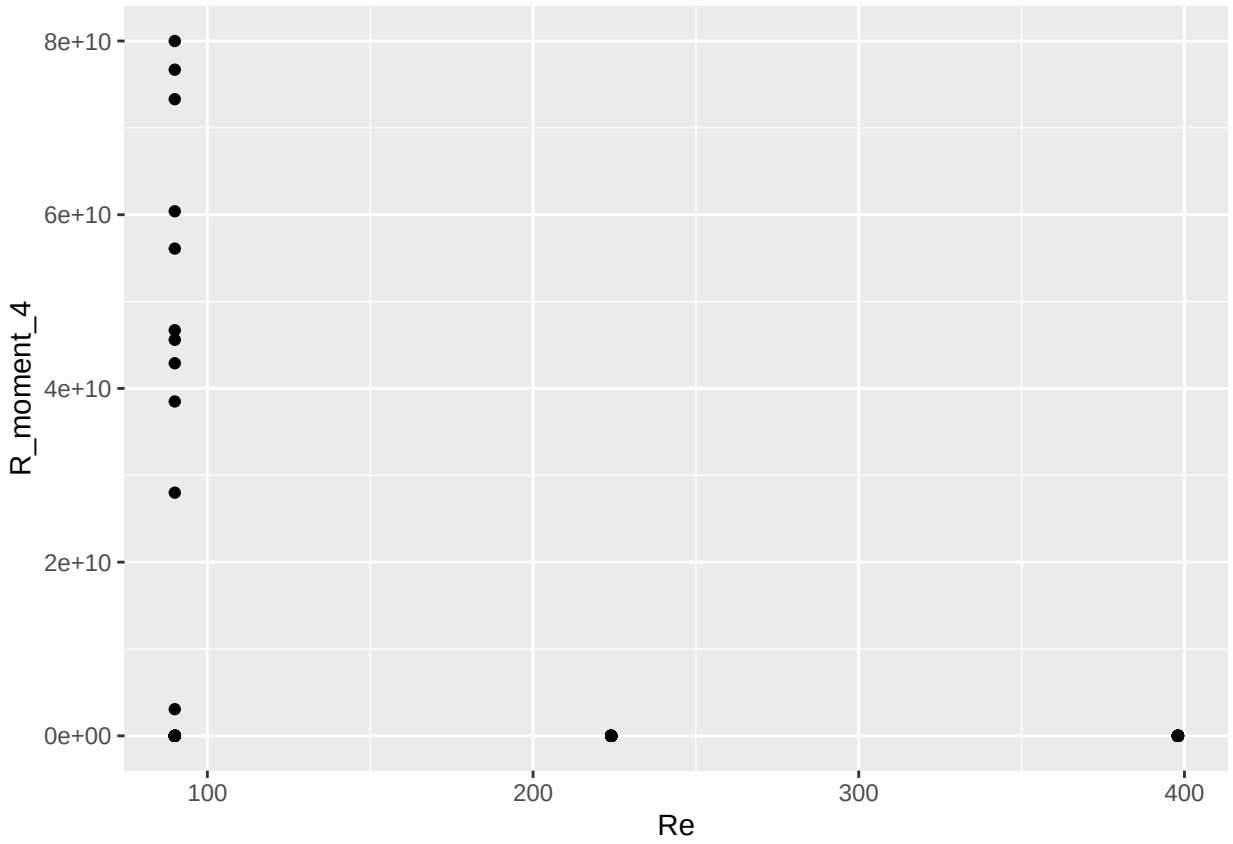

```
ggplot(train, aes(x = Fr, y = R_moment_3)) + geom_point()
```



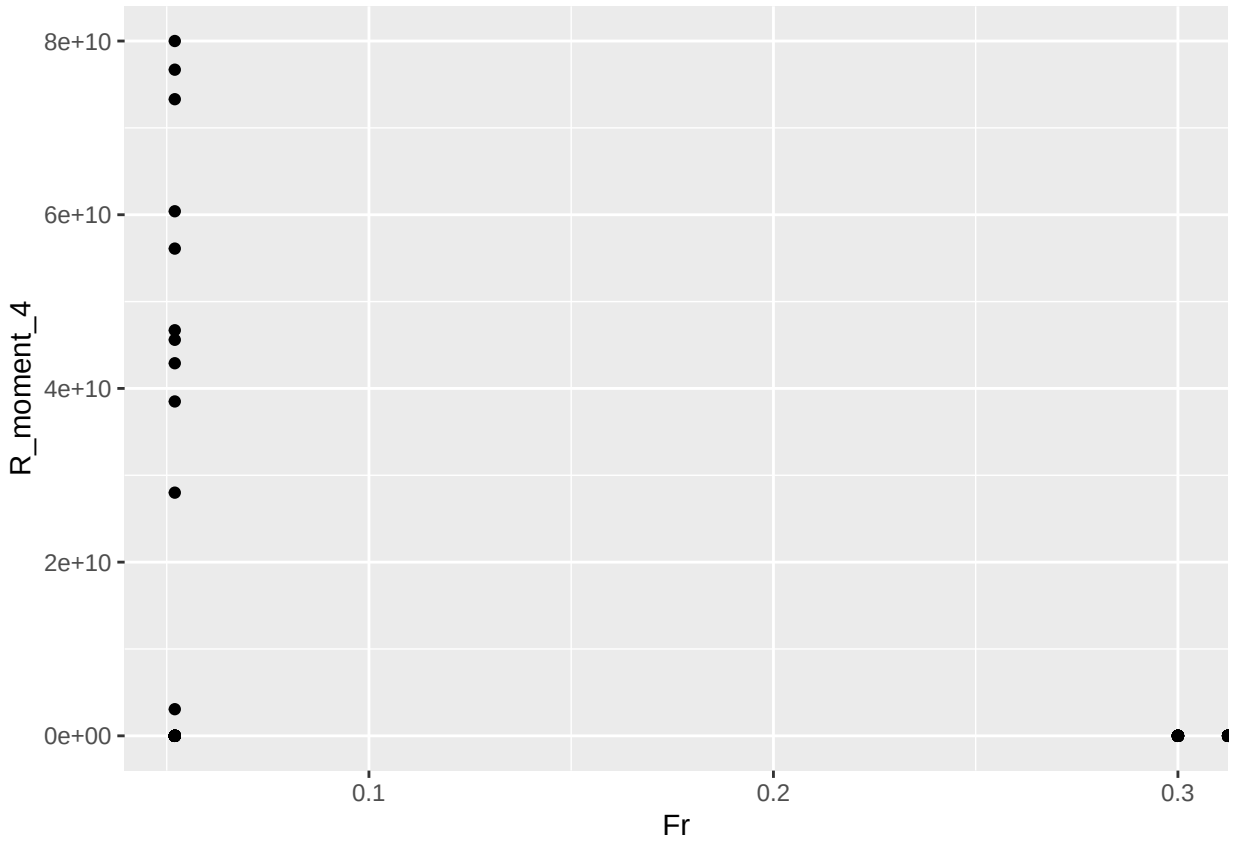

```
ggplot(train, aes(x = St, y = R_moment_3)) + geom_point()
```



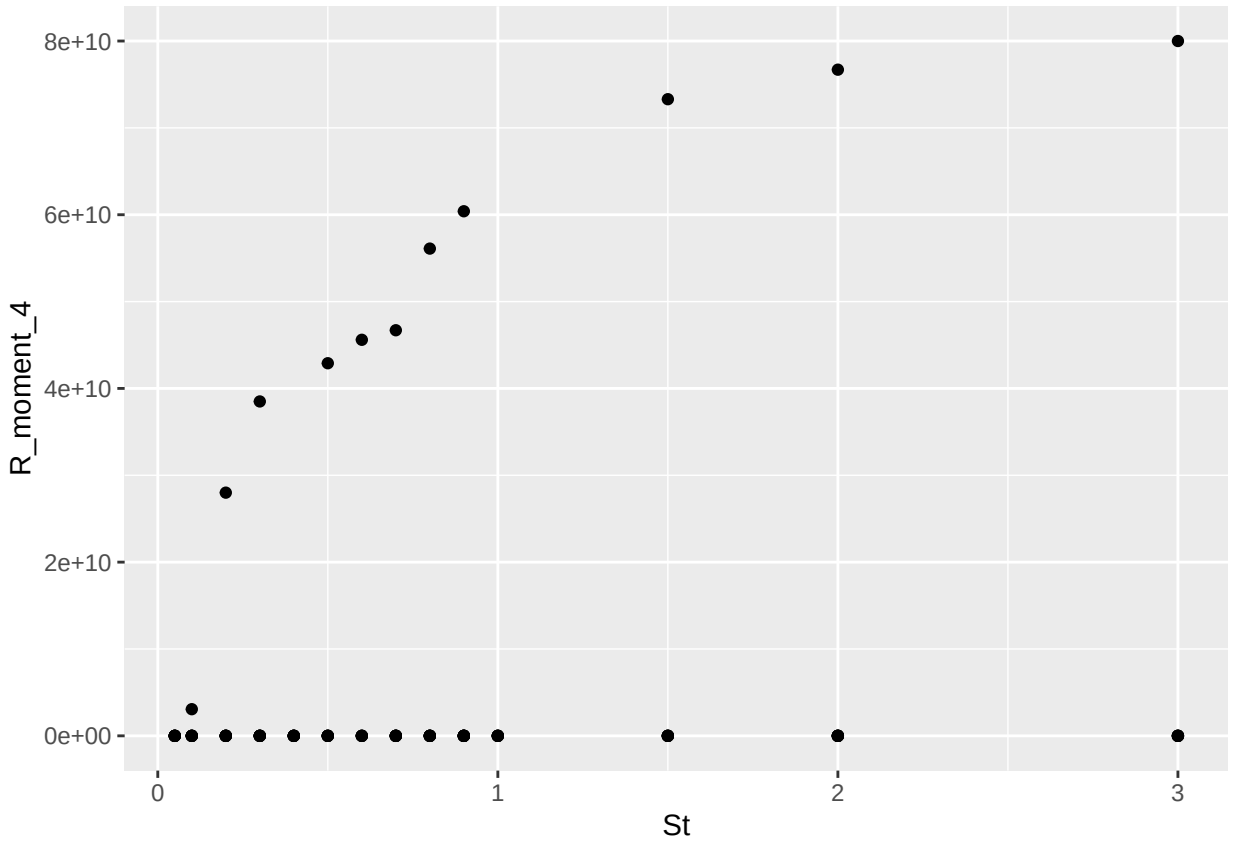

```
# Create scatterplots to explore relationships between predictor variables with R_moment_4
ggplot(train, aes(x = Re, y = R_moment_4)) + geom_point()
```




```
ggplot(train, aes(x = Fr, y = R_moment_4)) + geom_point()
```




```
ggplot(train, aes(x = St, y = R_moment_4)) + geom_point()
```



Simple Linear Regression