

Scatter plots and Histograms

23 January 2019

Objective: show relationship and distribution between inputs and outputs using scatter plots and histograms of process variables in SC2 for ~110 batches between Feb and May 2017

End goal: dimensionality reduction by finding tags that have a close to linear relationship with Parameters which can then be used to make a generalised model which will provide the basis for a generalised model from which causes can be inferred.

Problems: If it is possible to know which tag and parameter should have a close to linear relationship by domain experts, this pair can then be cross checked in the plots to know if the data is good or not.

Assumptions:

- *Only Parameters and Tags with close to normal distribution have been selected

- *5 minute frequency is sufficient

- *Mean is a good method of aggregation

Other methods that could be tried:

- *Aggregation by median, mode, SD and variance

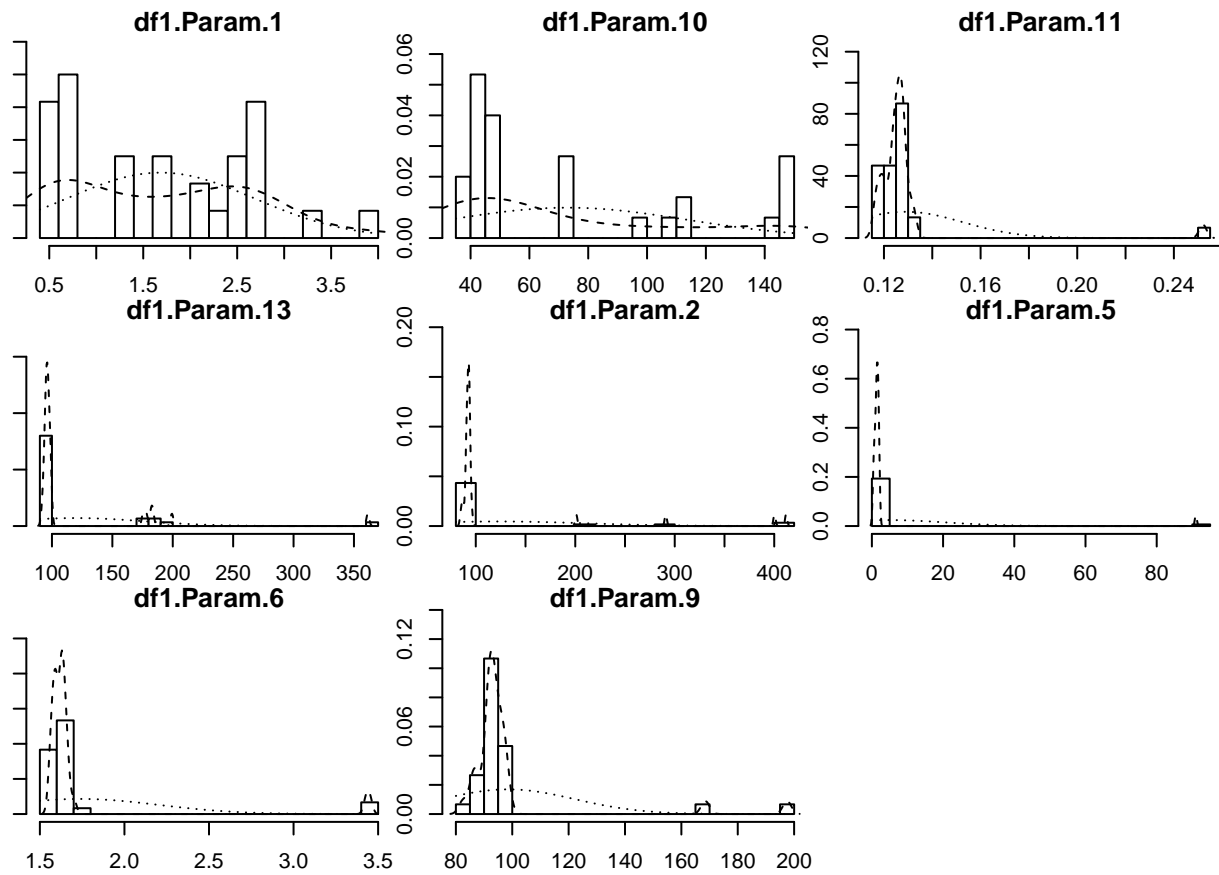
- *Transformation by log and square

- *Plot on and off spec batches together

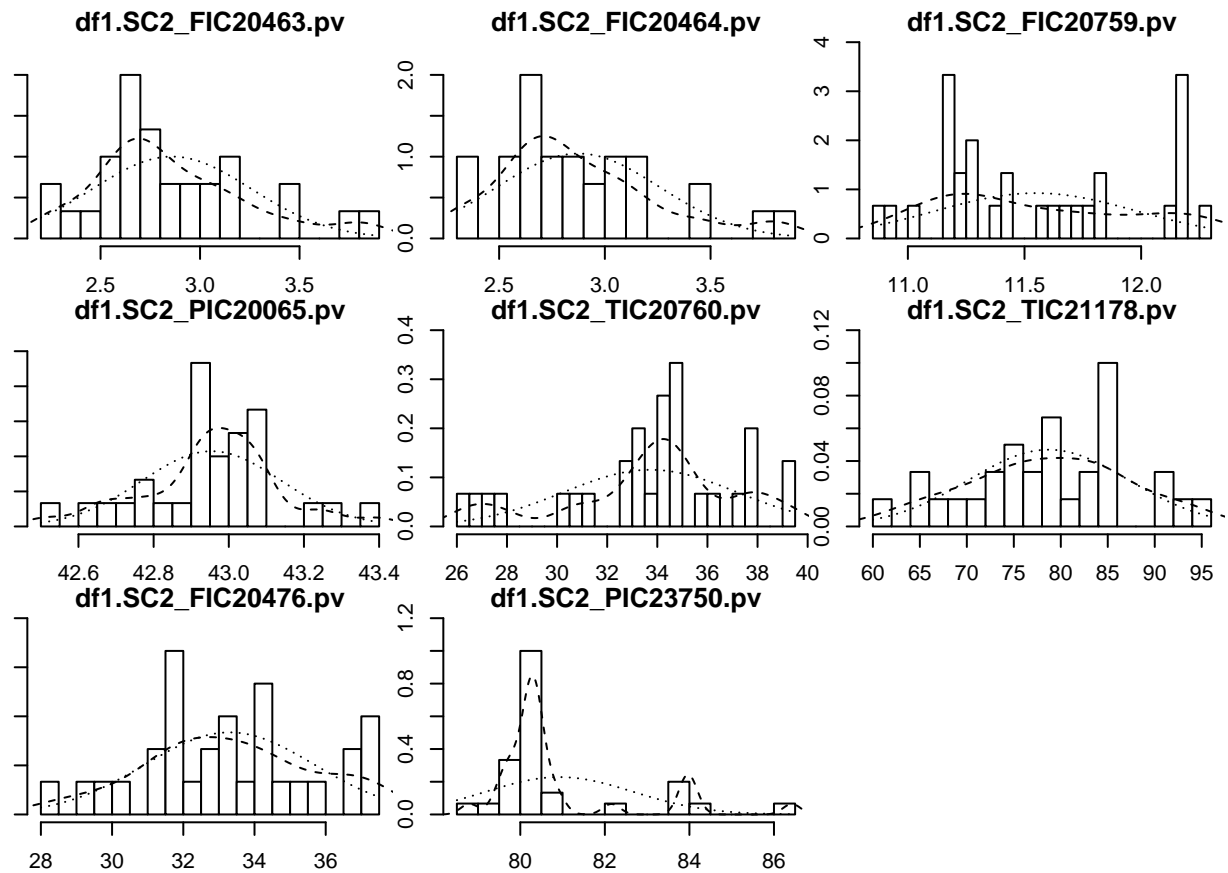
- *Run t-test between onspec and offspec batches

- *Find distance between important tags between onspec and offspec batches

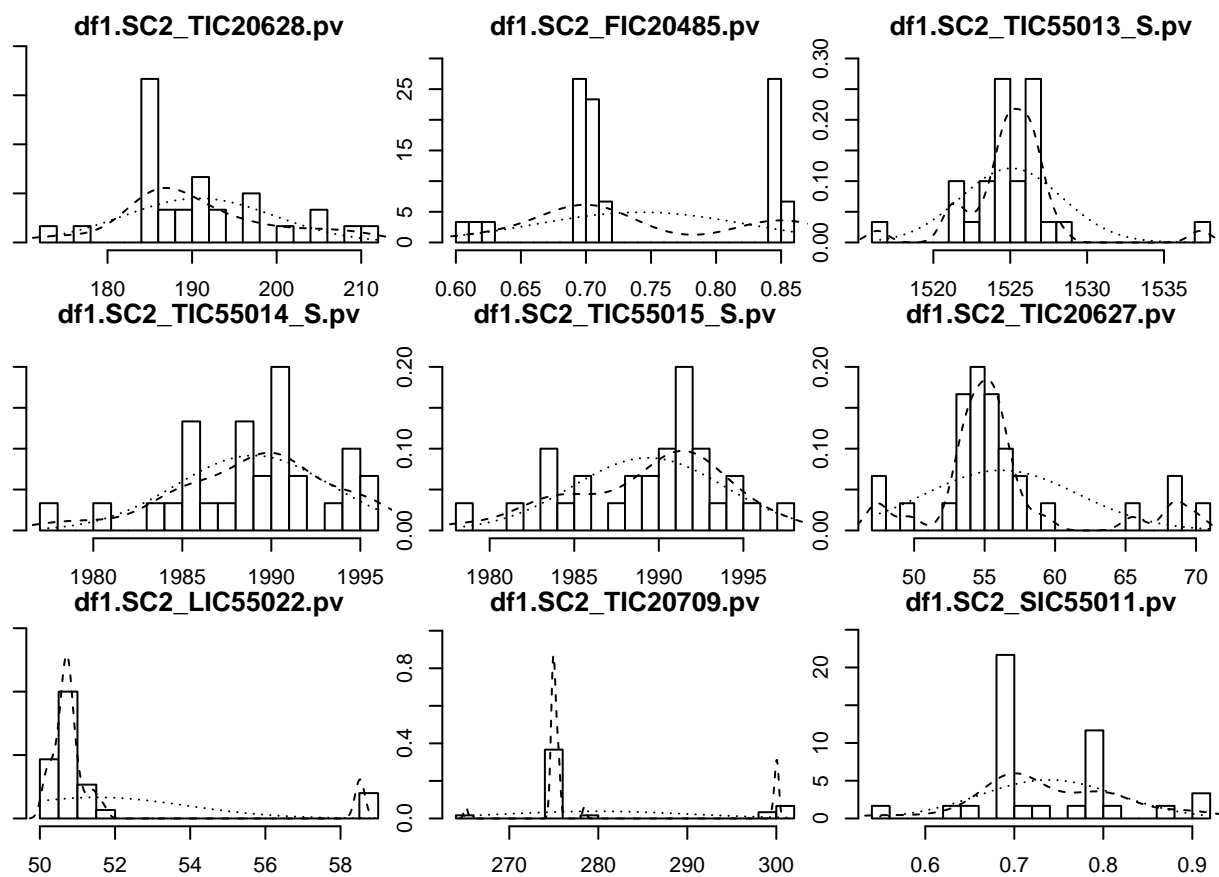
The below histograms and scatterplots are for OFFSPEC batches only



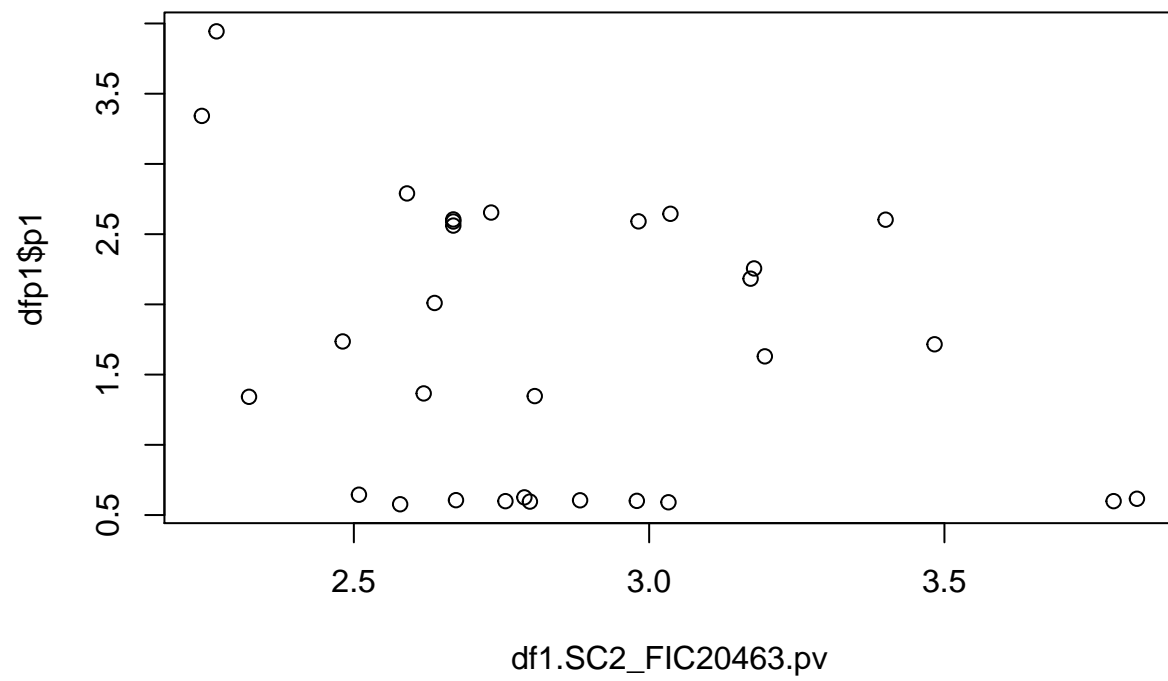
```
psych::multi.hist(input[,1:8])
```

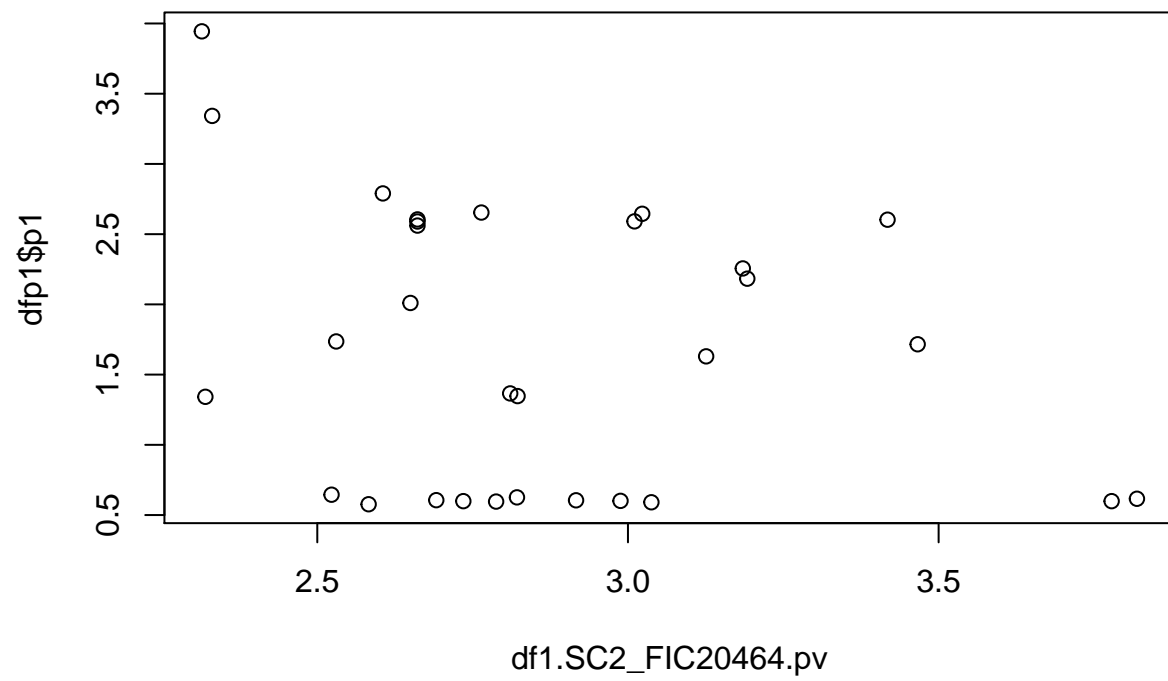


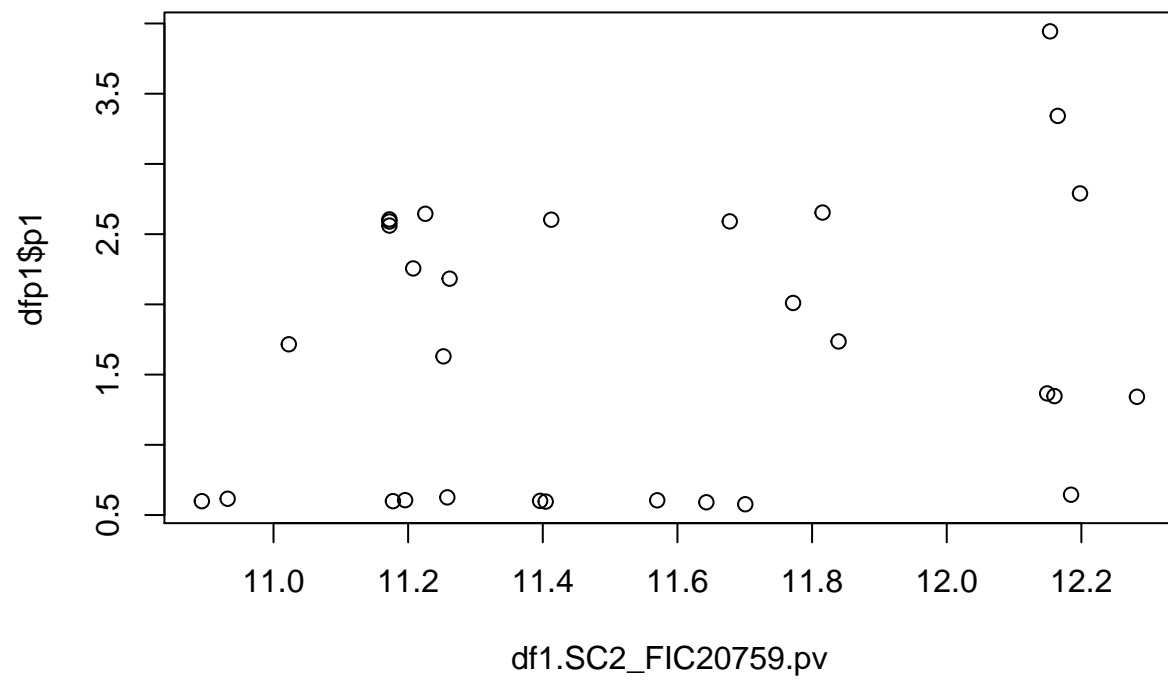
```
psych::multi.hist(input[,9:17])
```

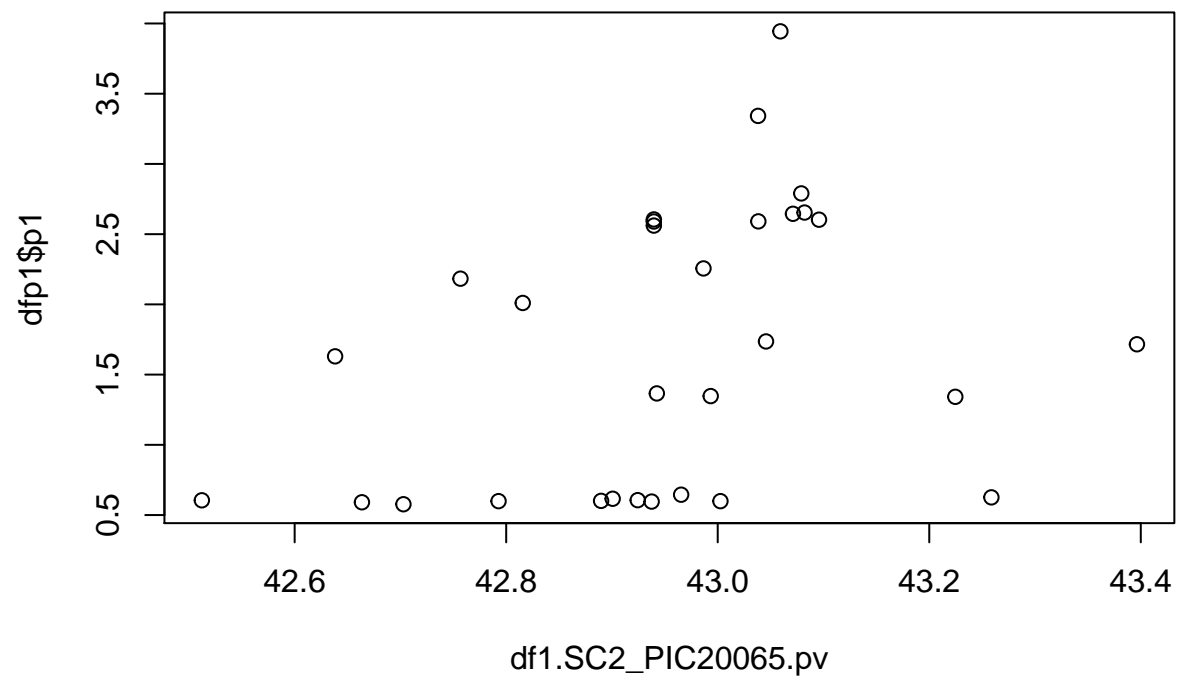


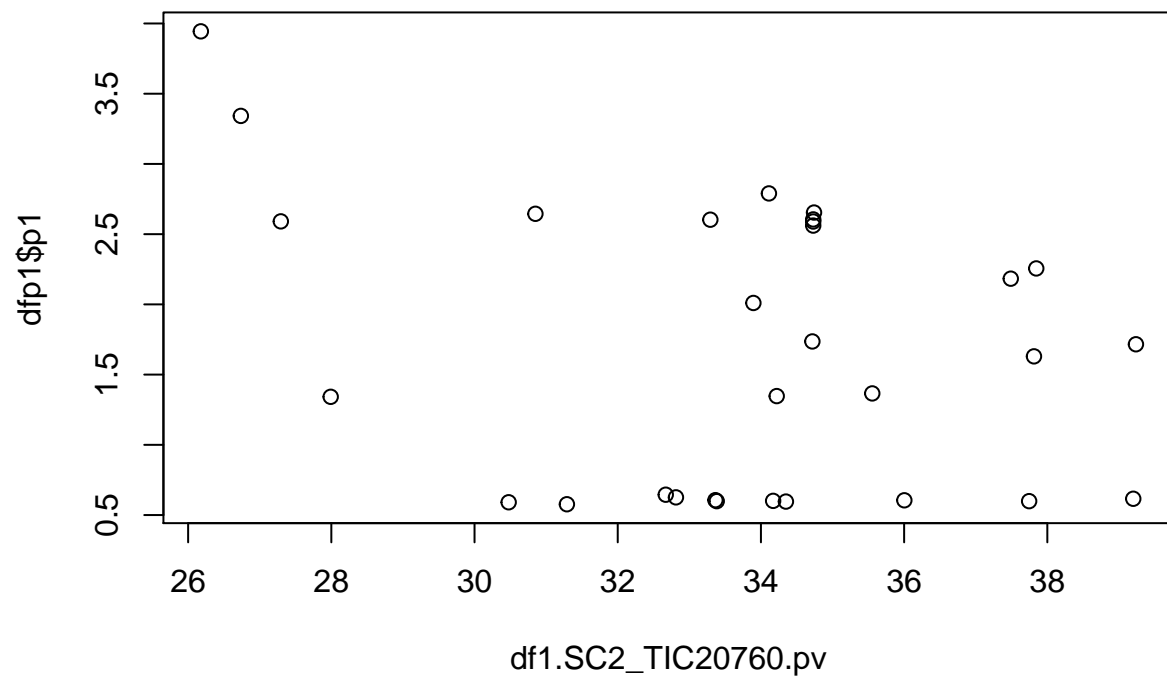
```
dfp1<-data.frame(p1,input)
plot(dfp1$p1~.,dfp1)
```

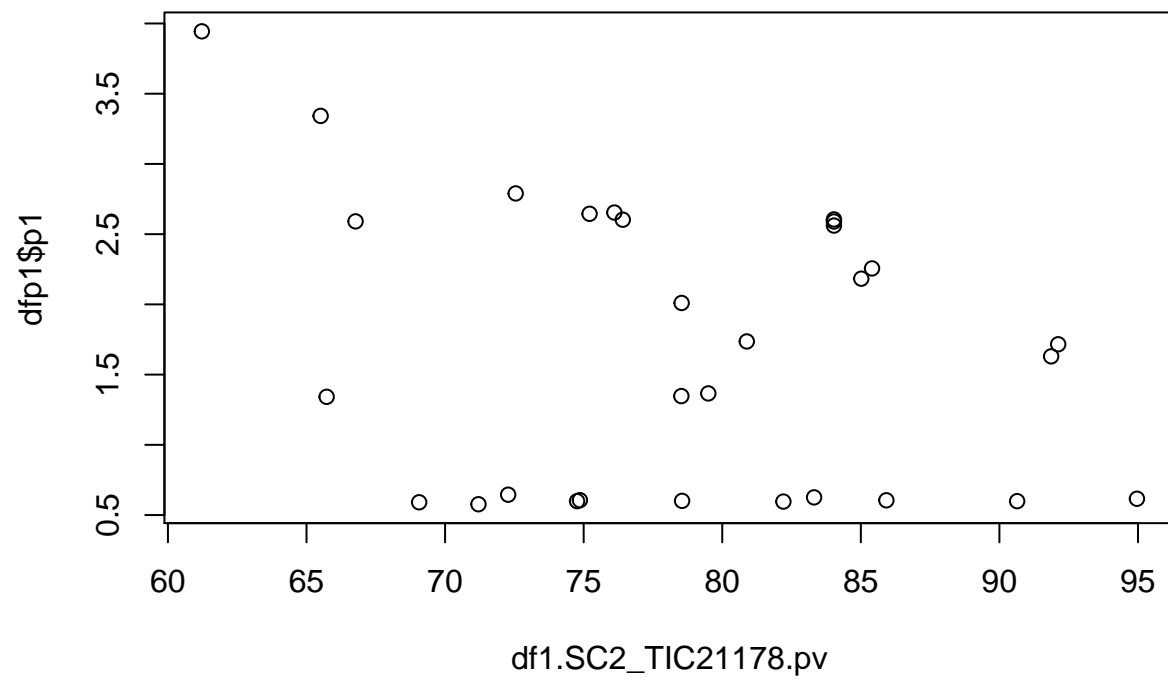


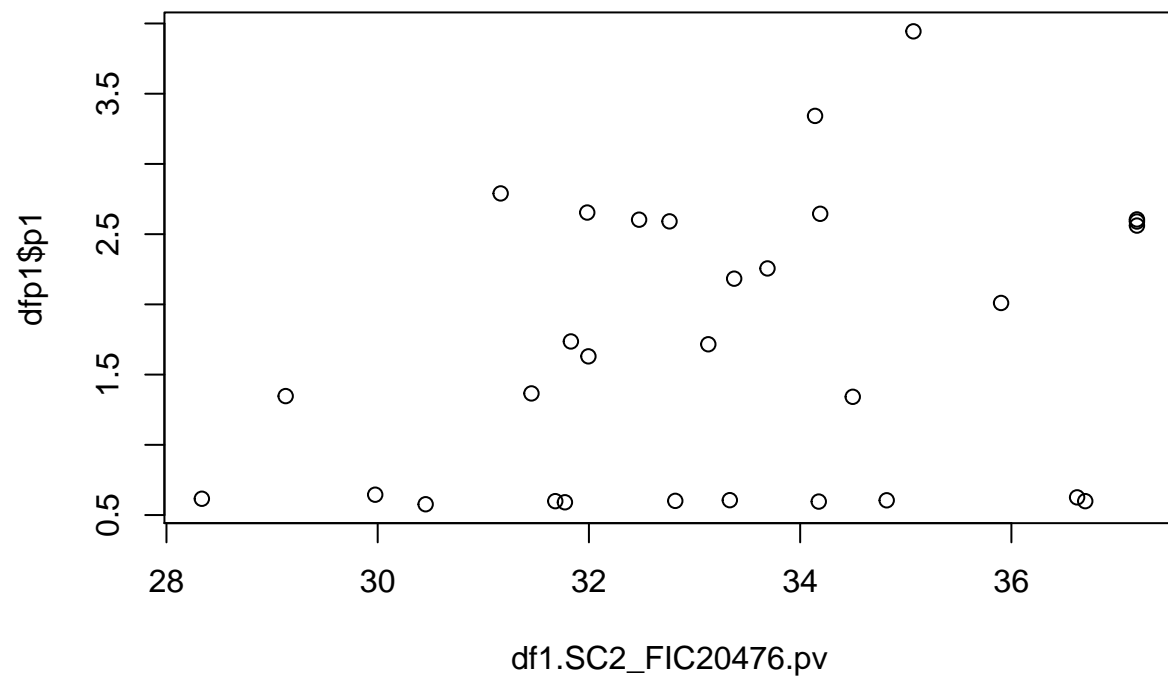


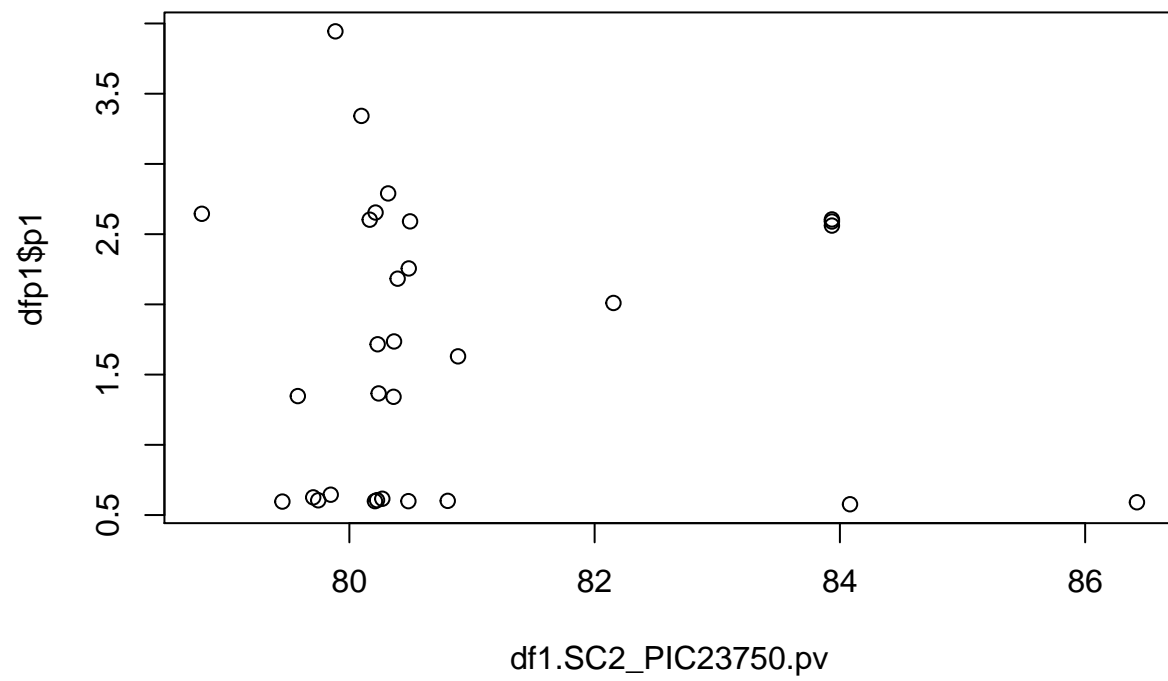


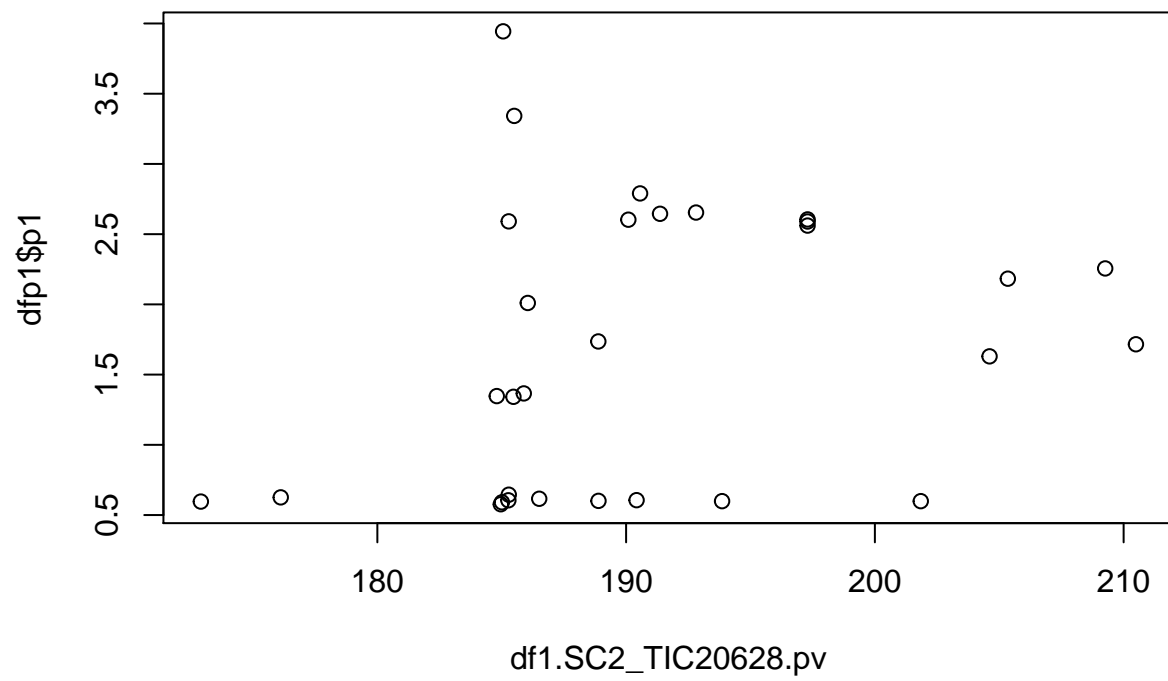


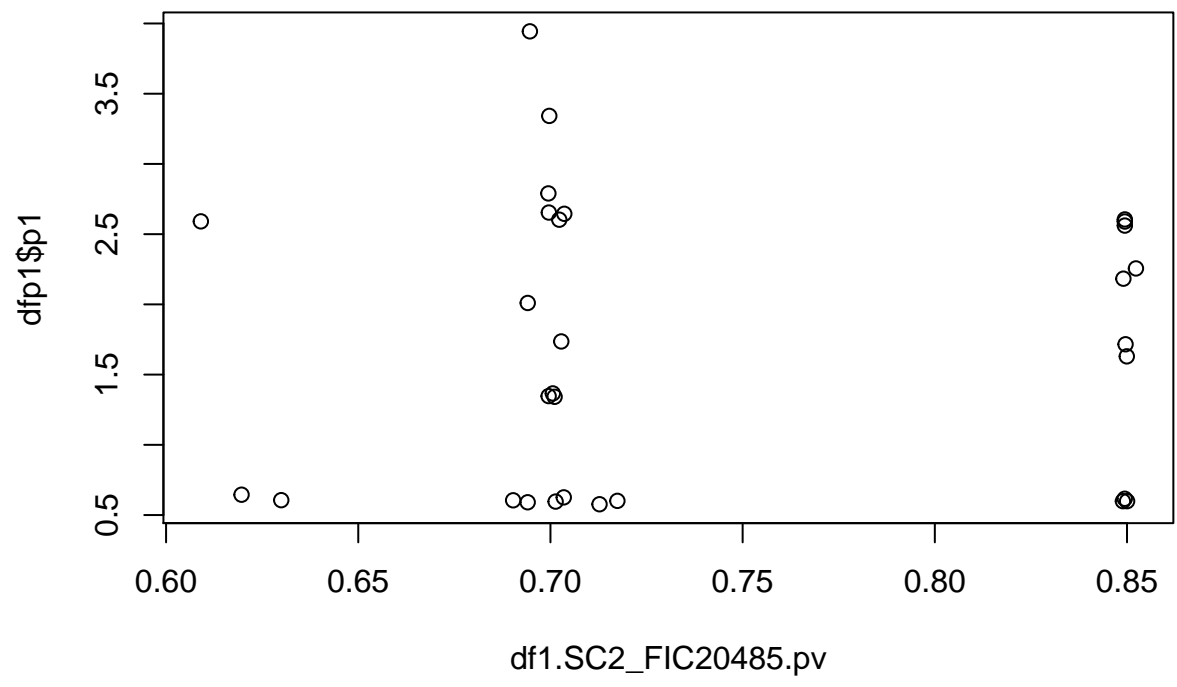


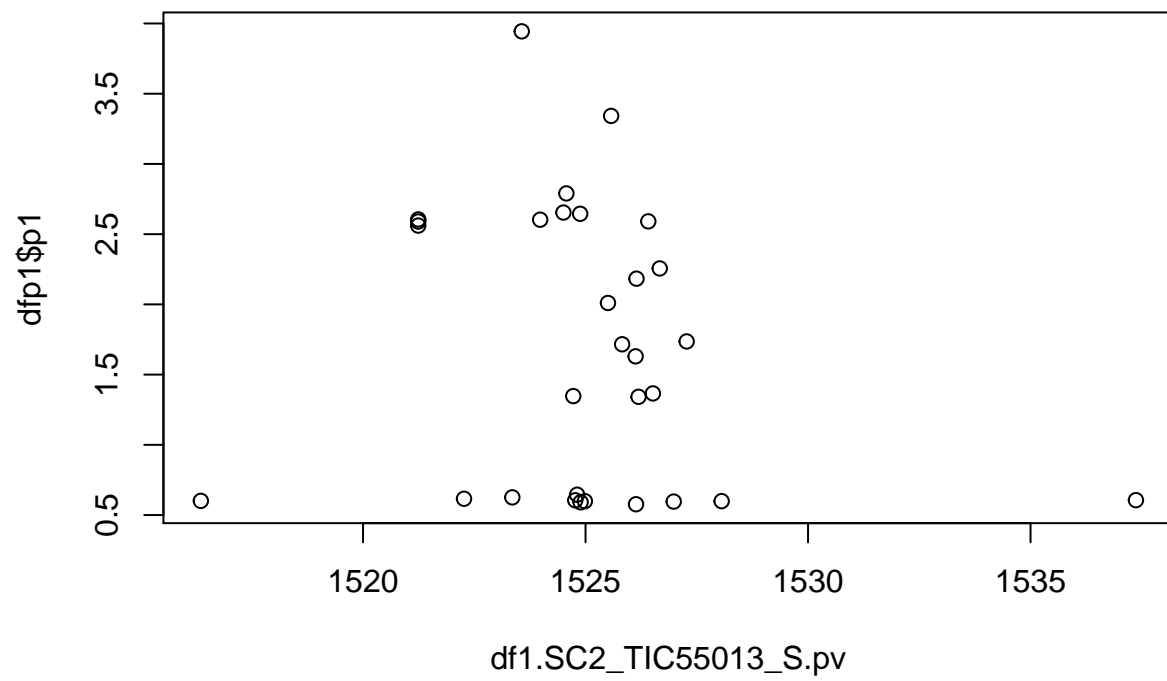


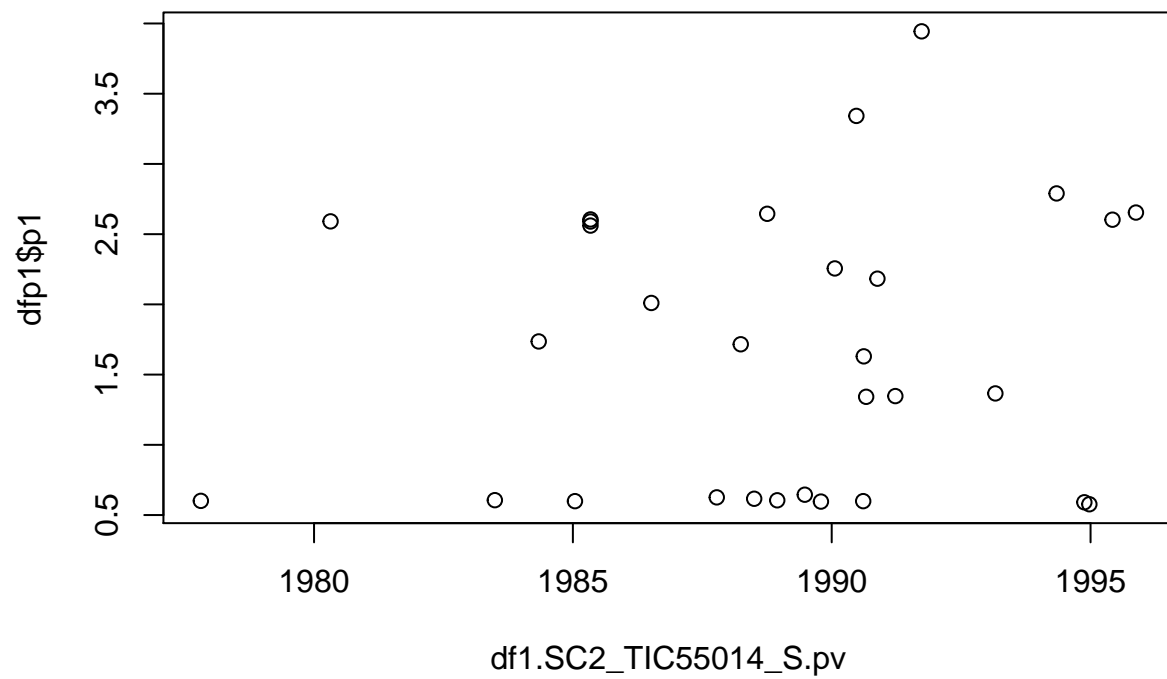


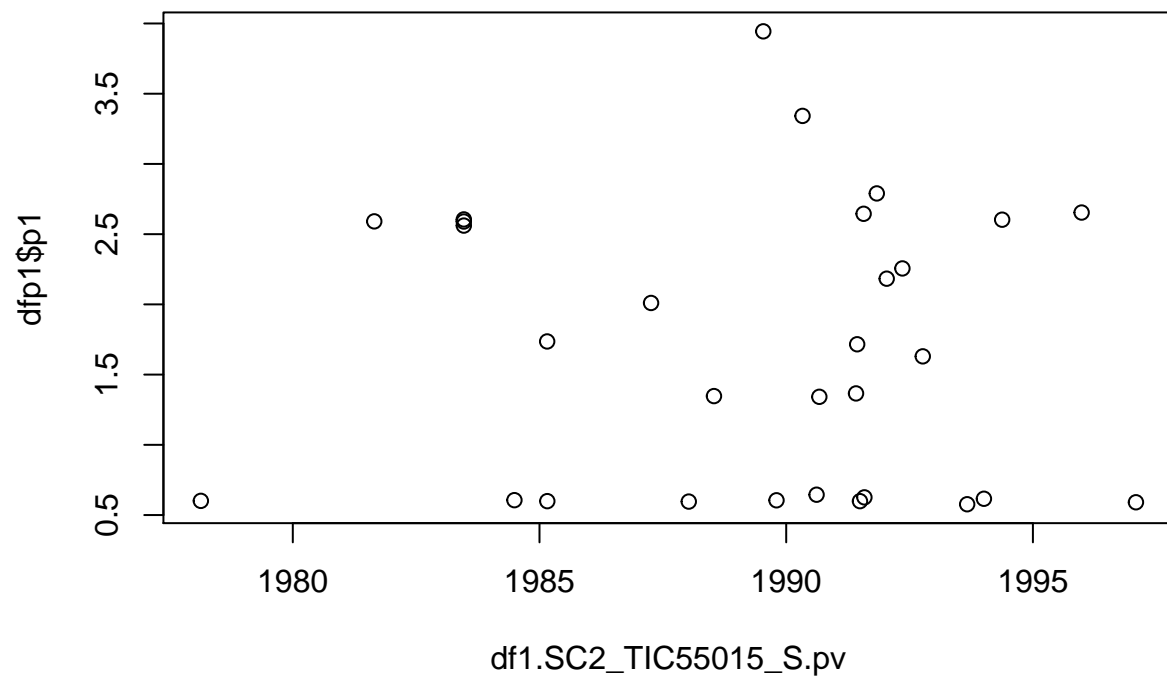


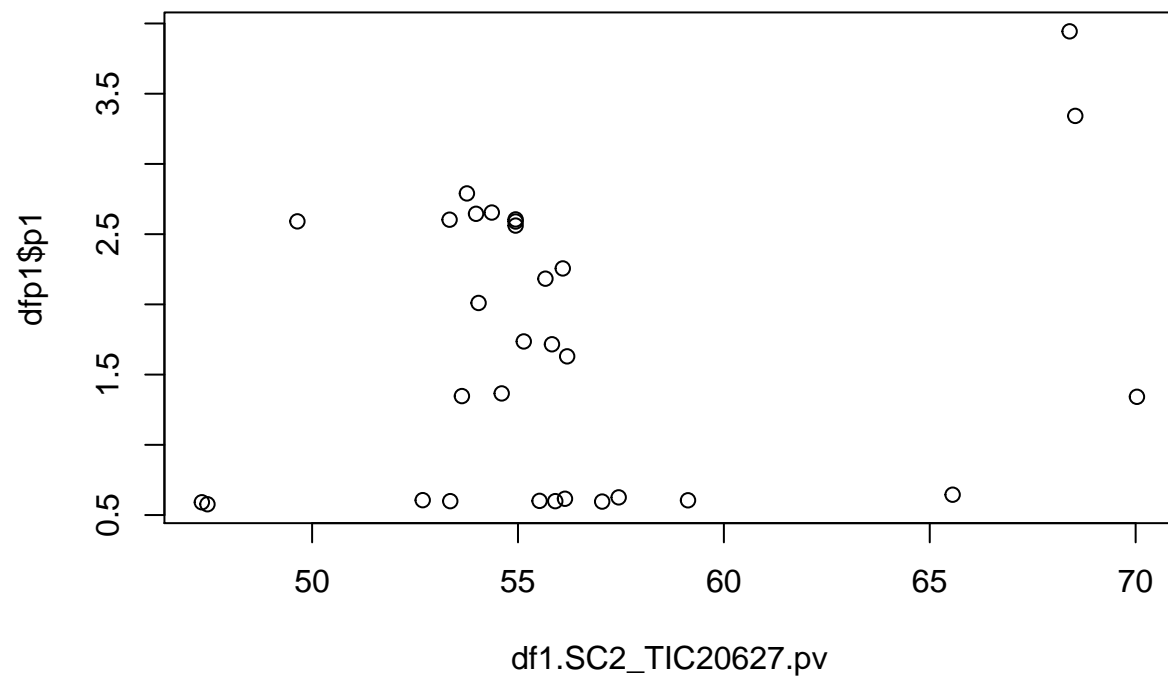


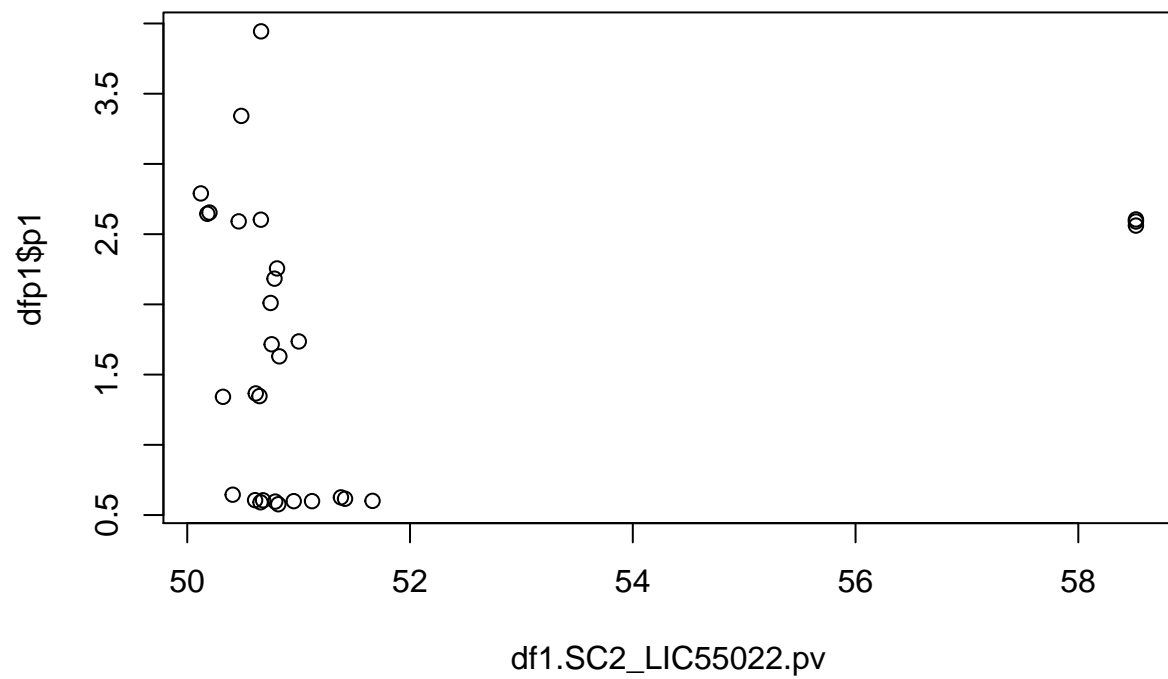


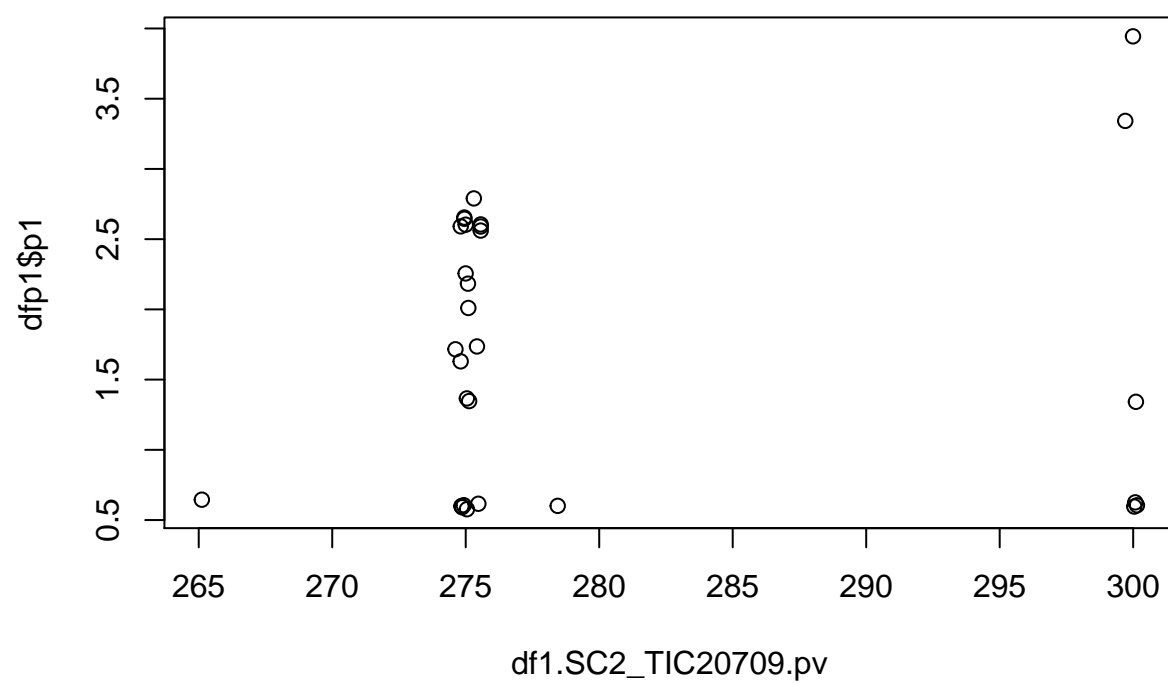


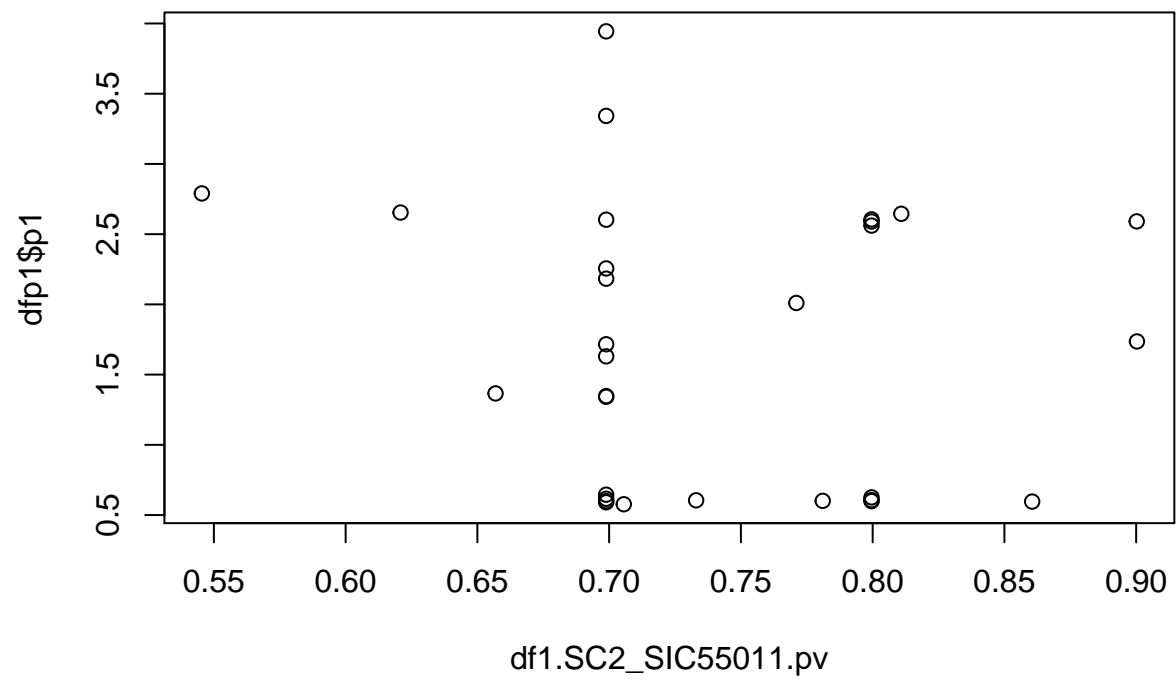




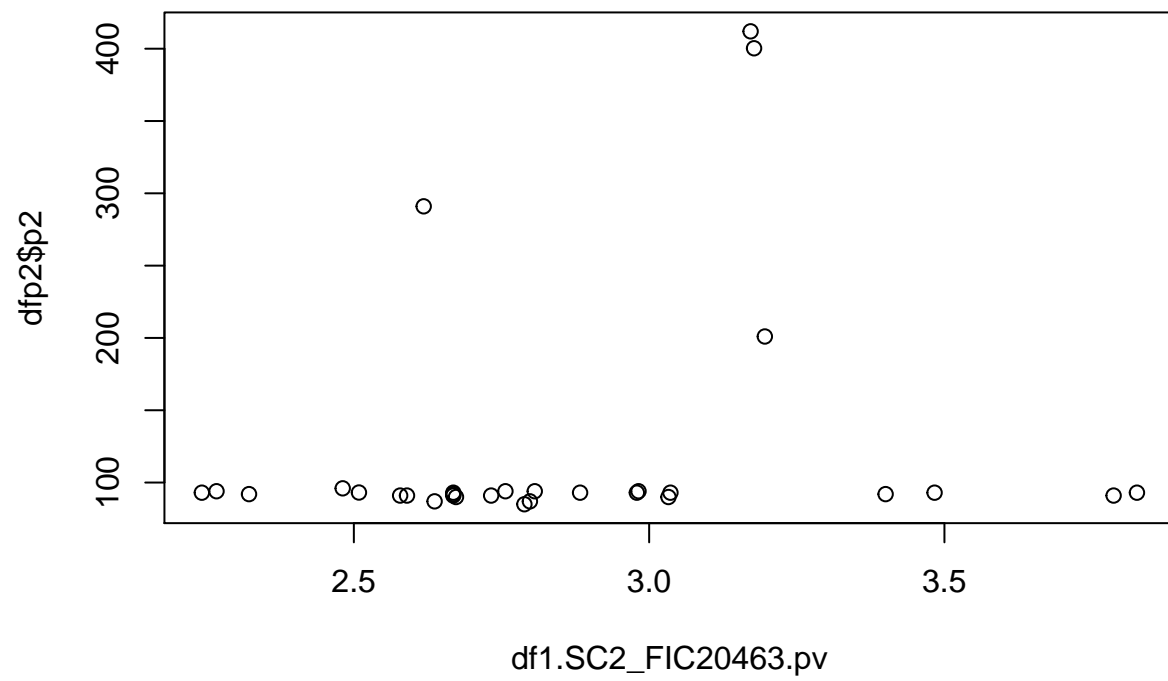


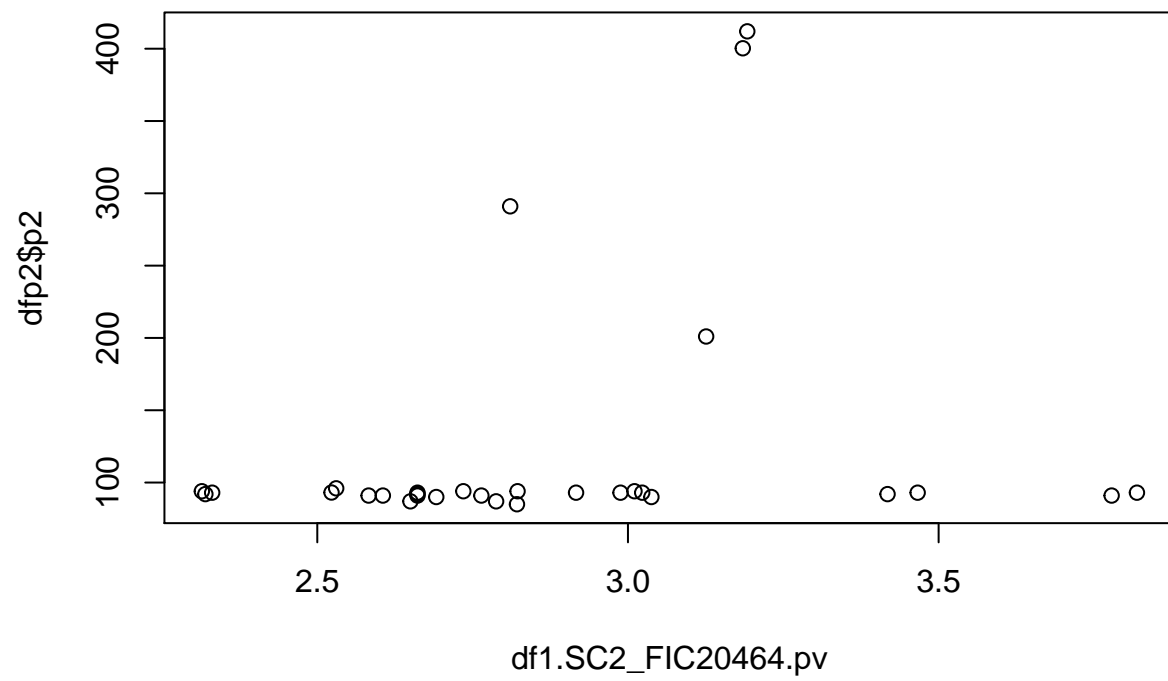


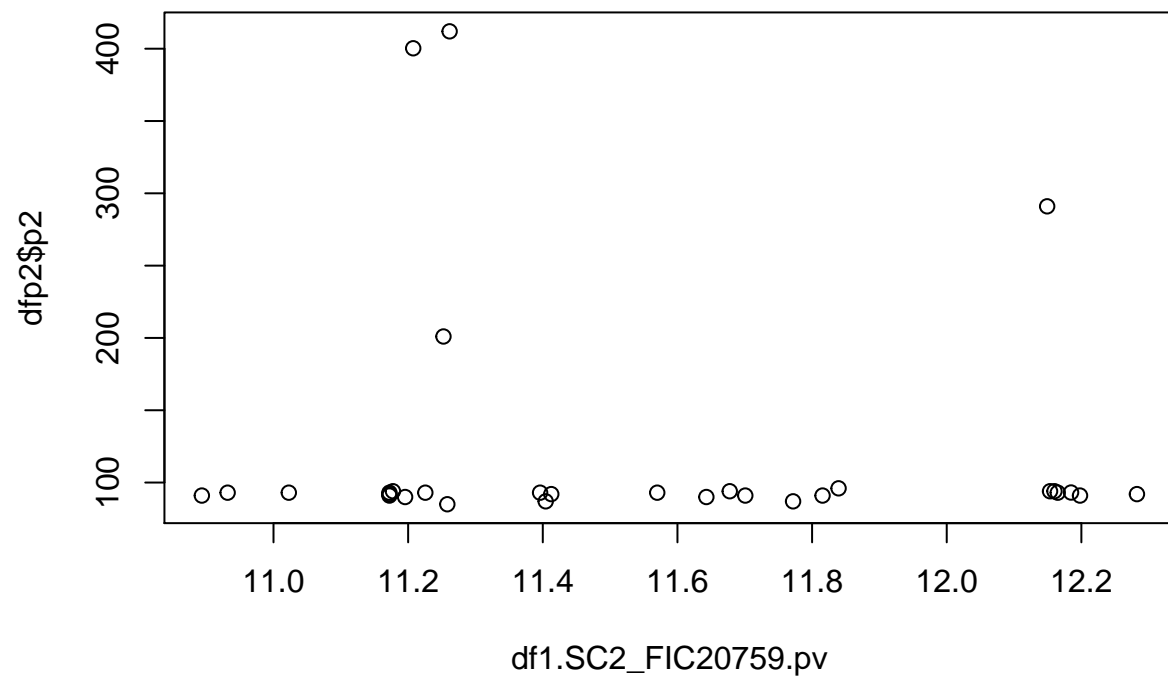


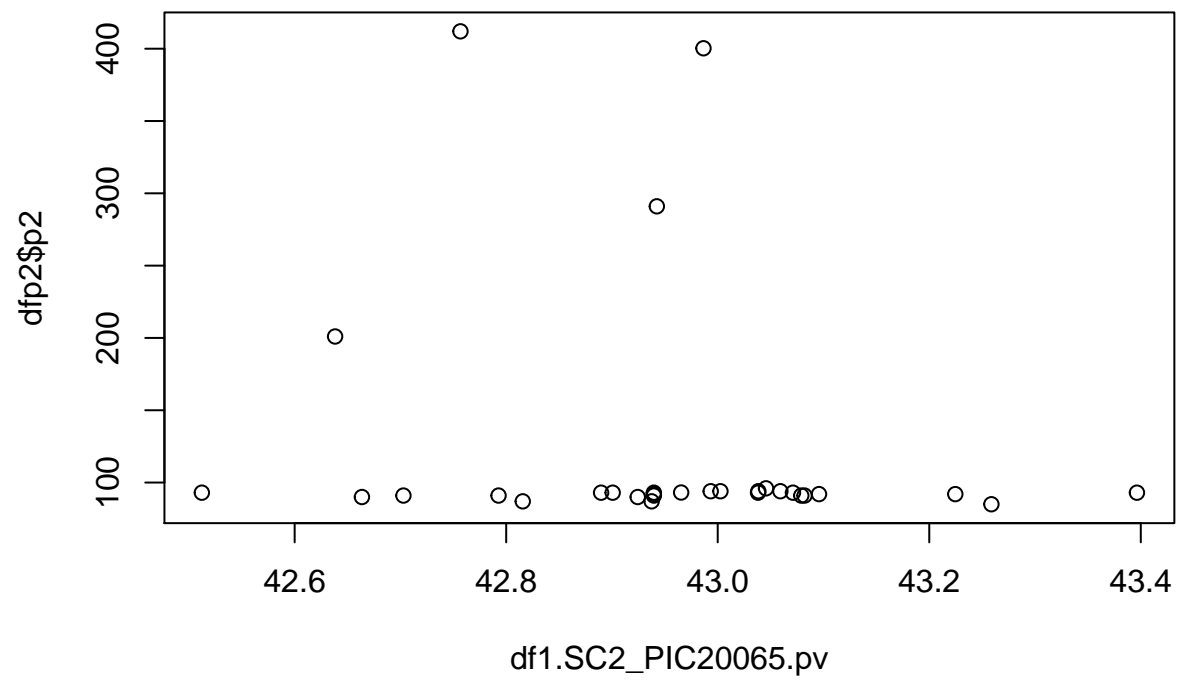


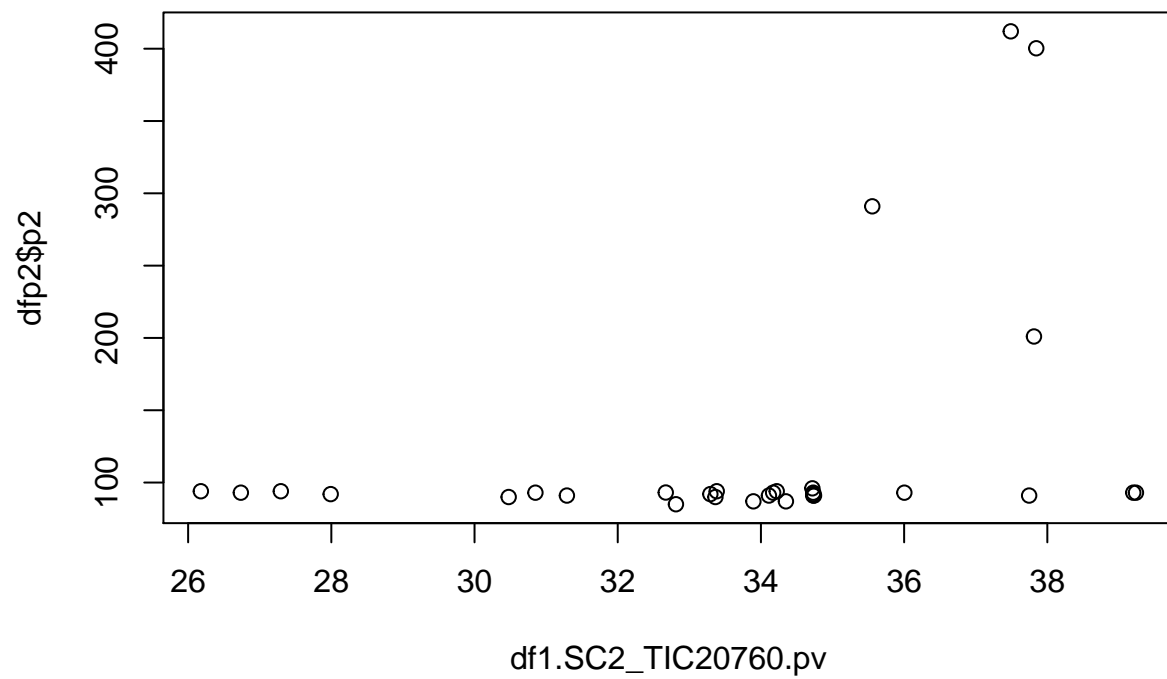
```
dfp2<-data.frame(p2,input)
plot(dfp2$p2~.,dfp2)
```

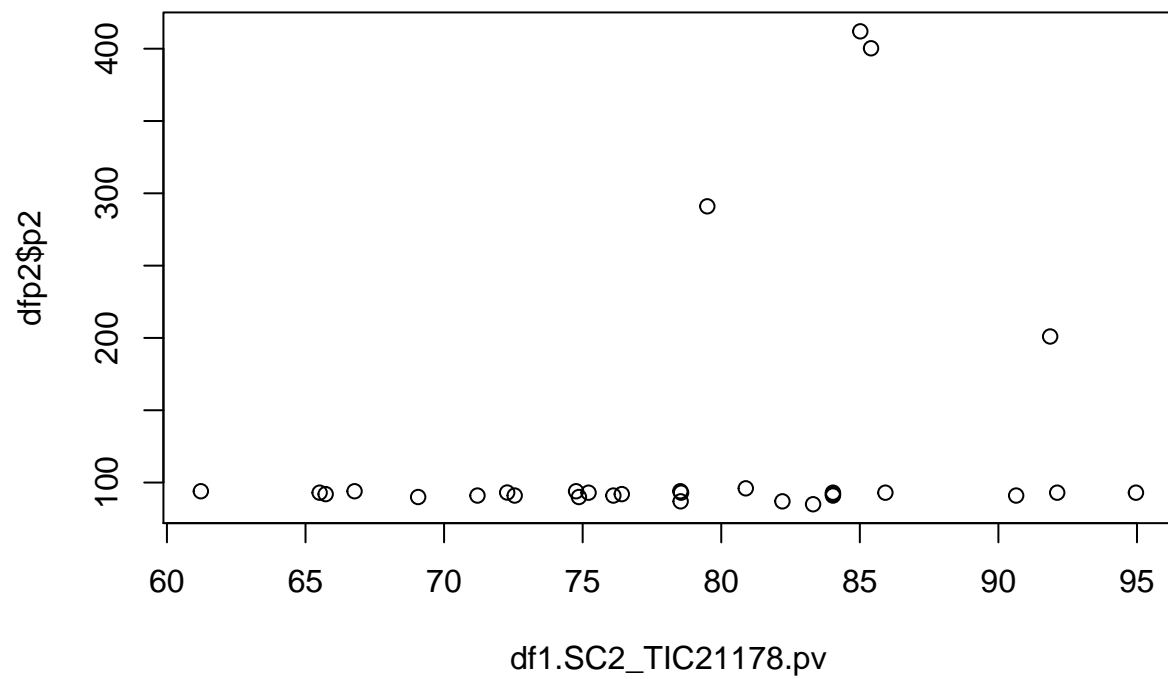


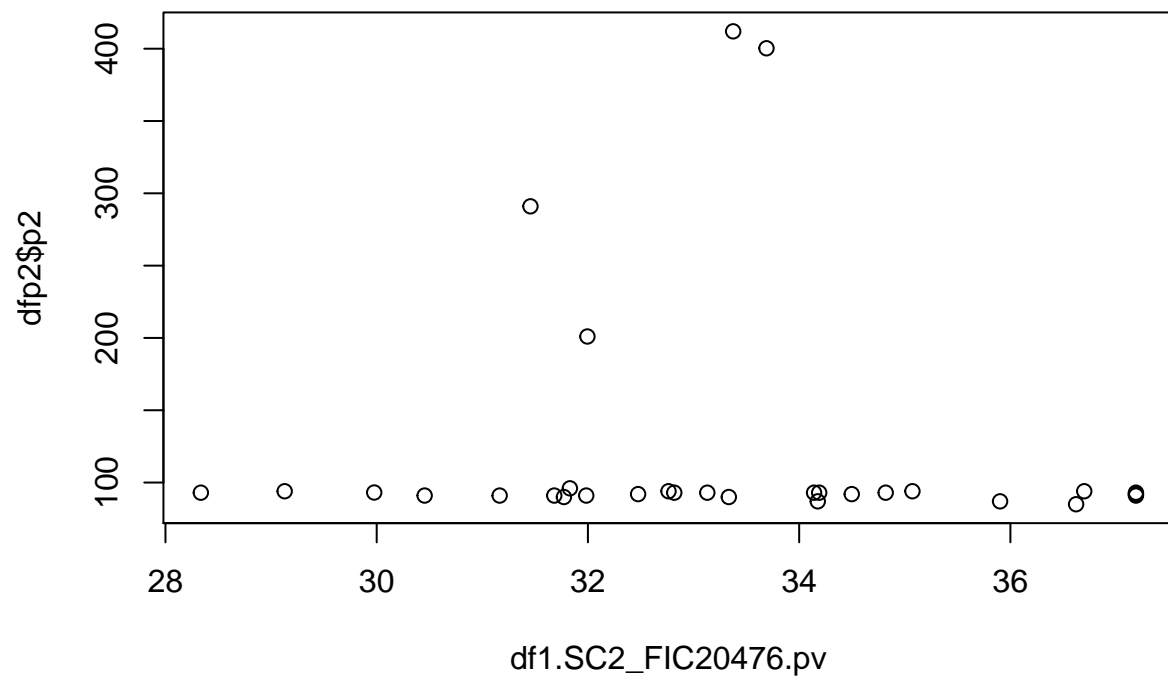


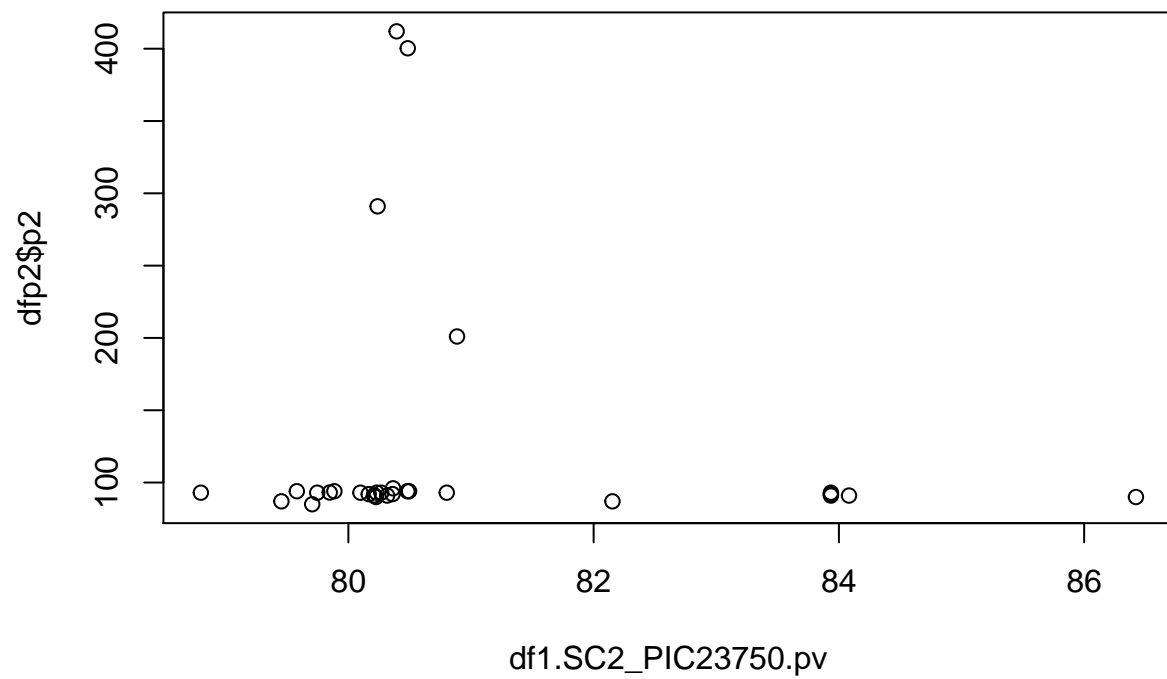


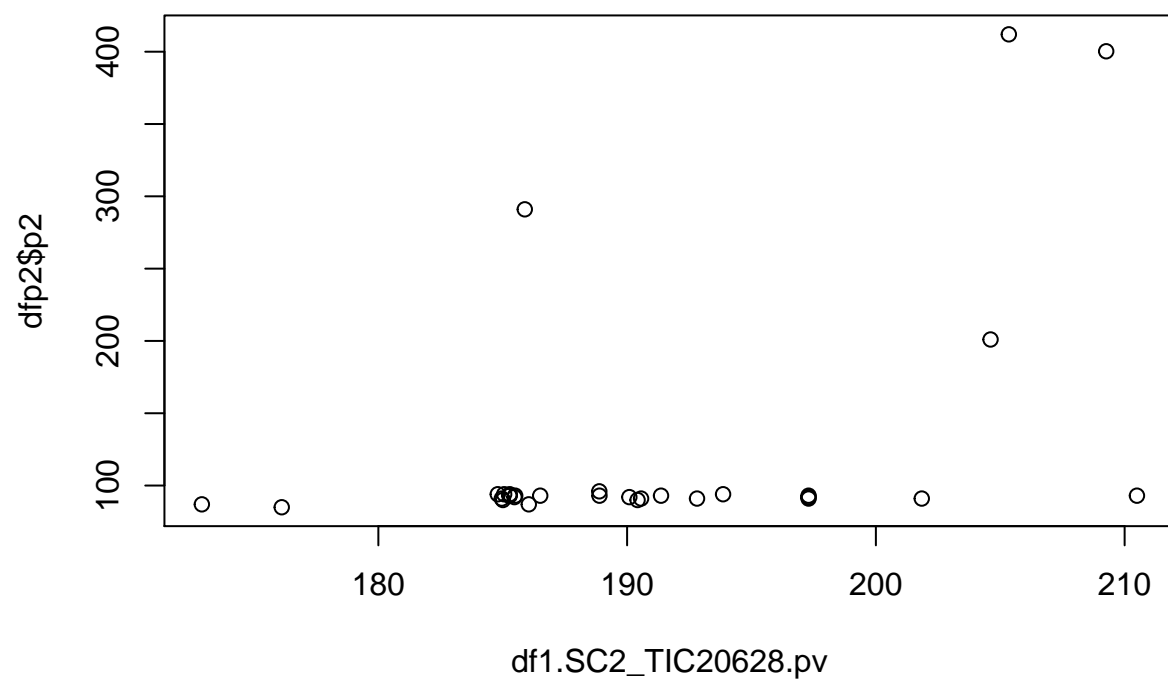


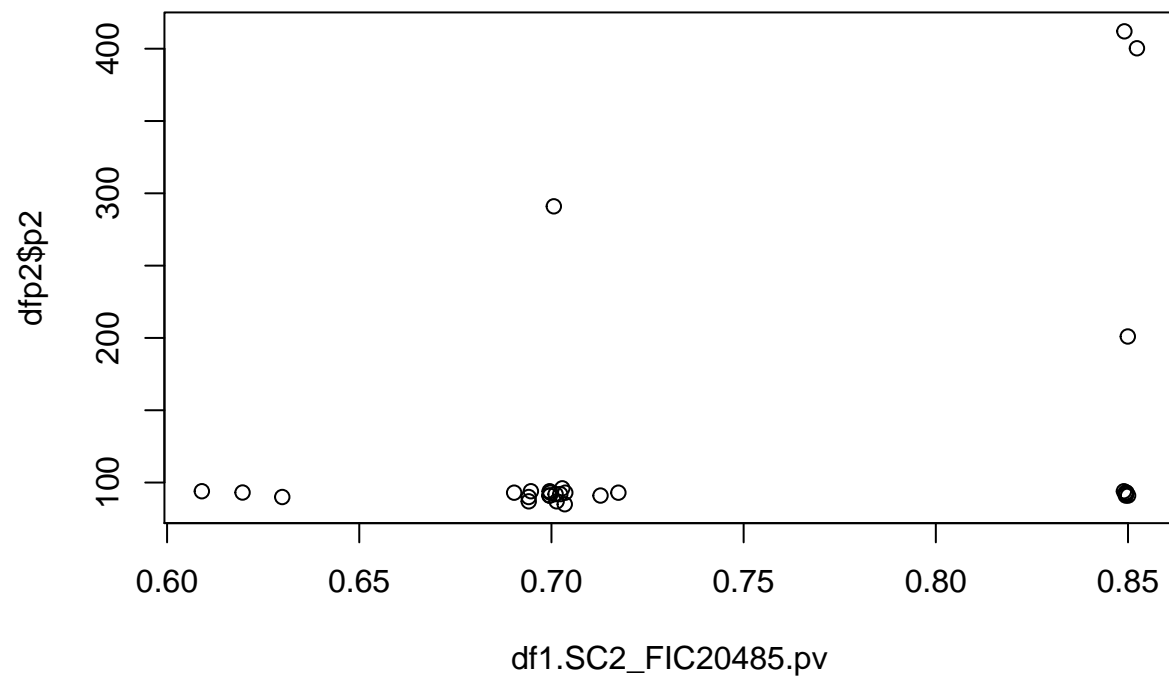


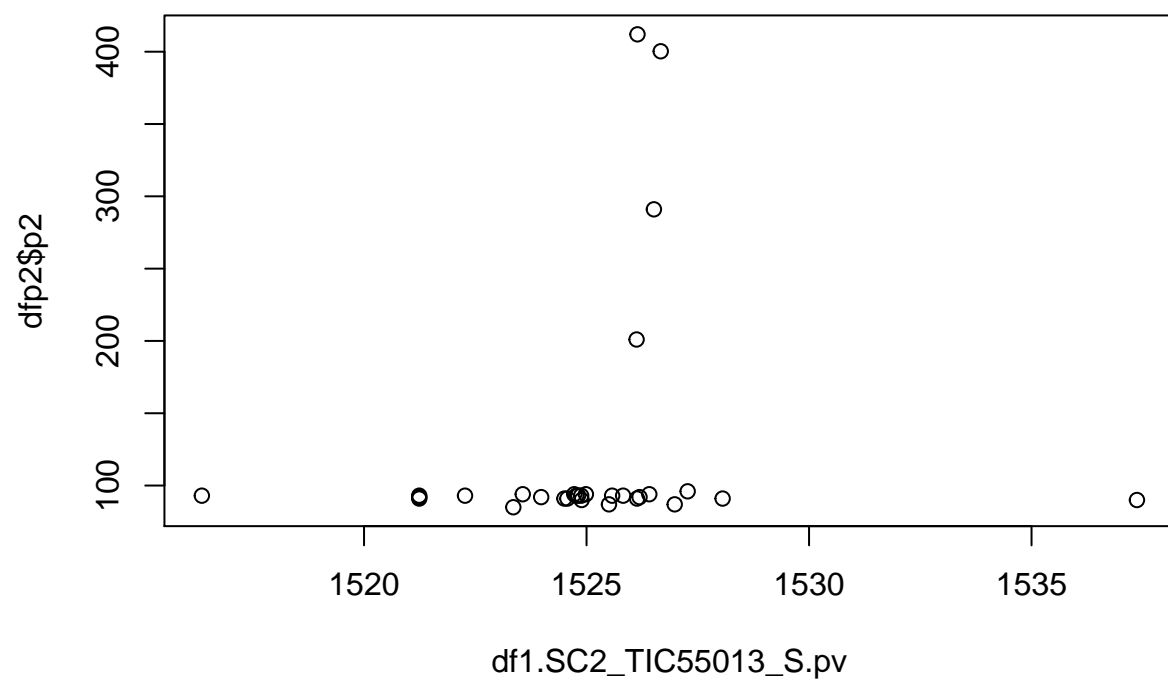


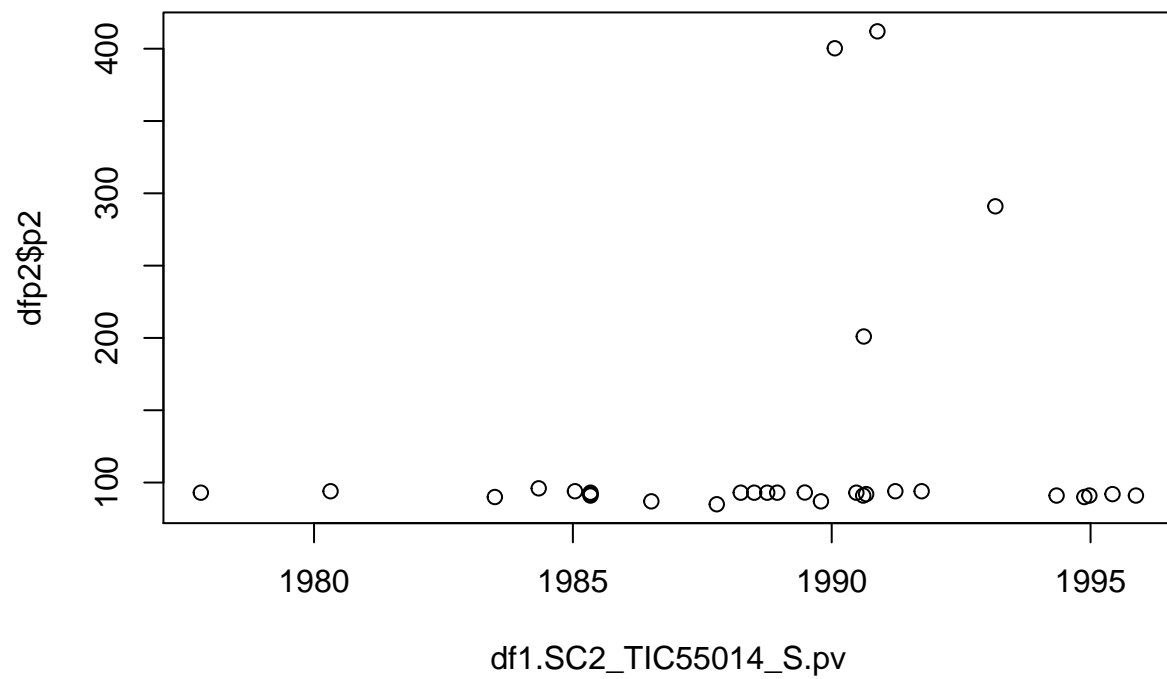


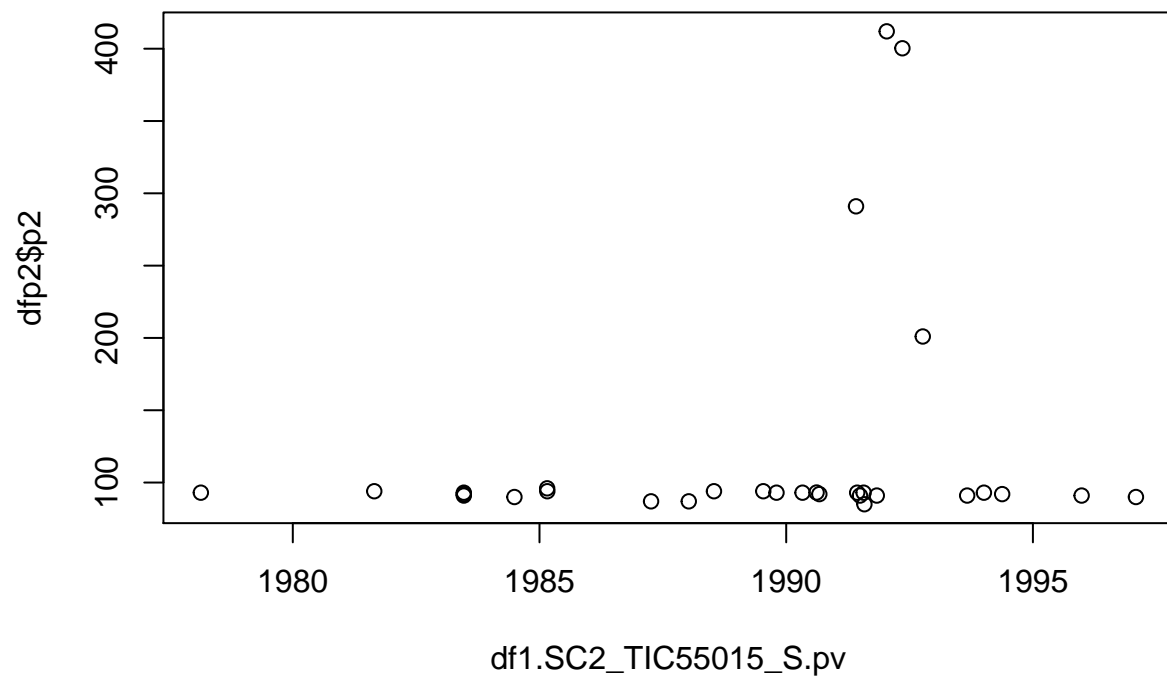


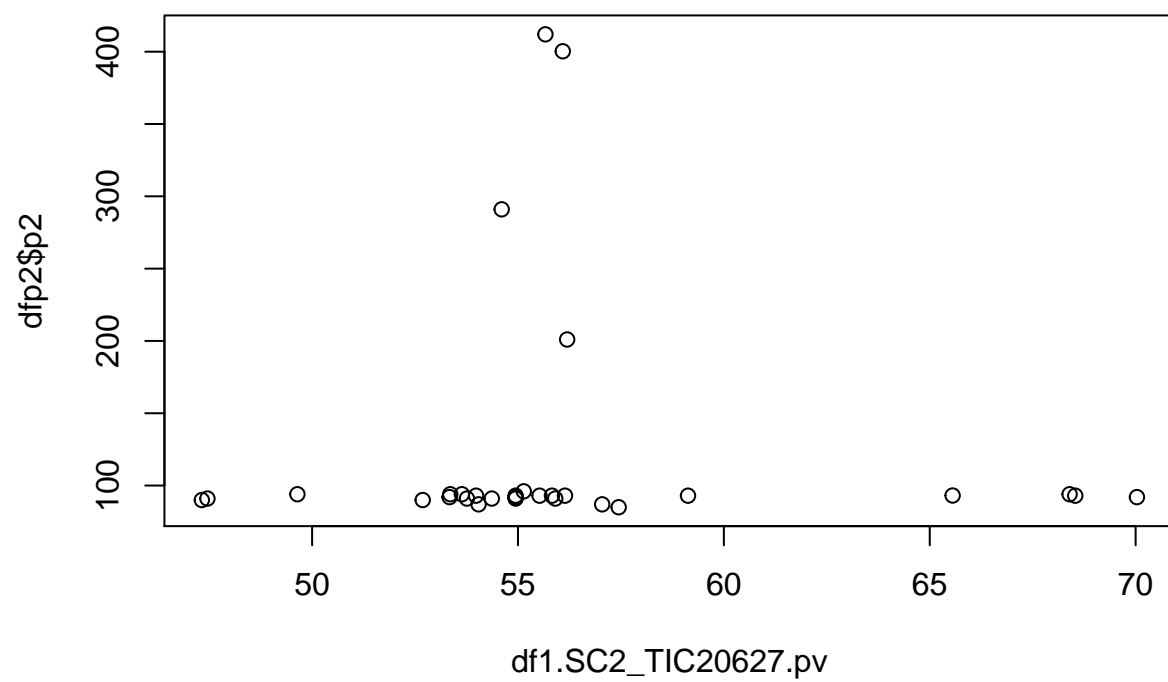


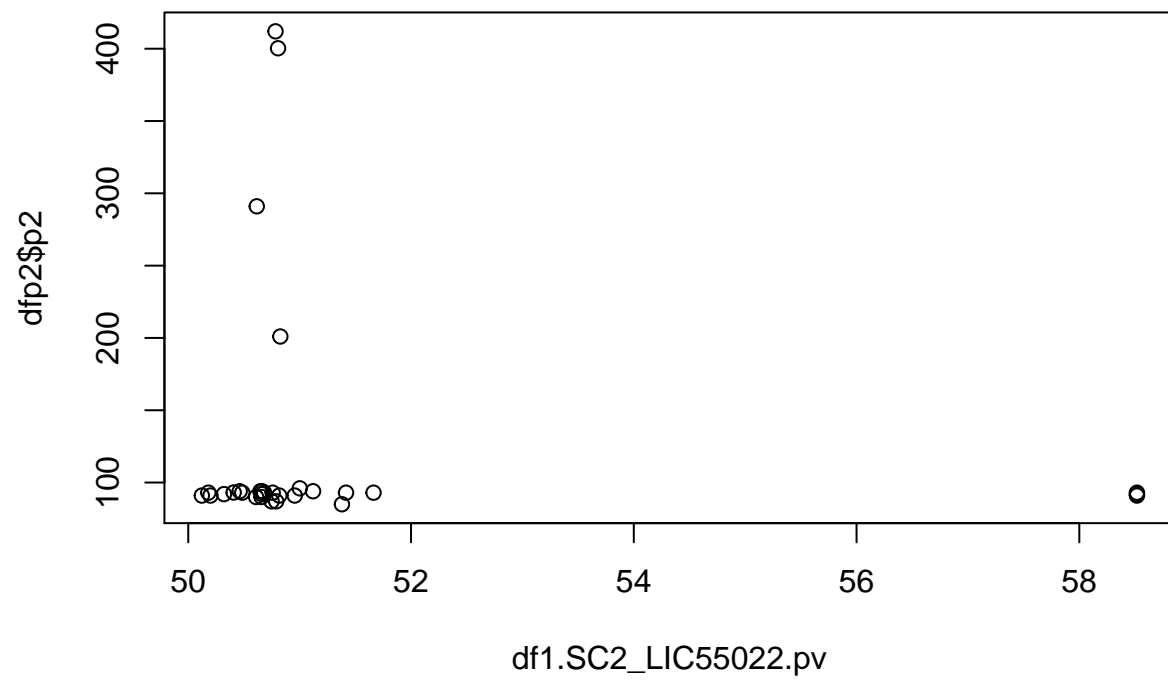


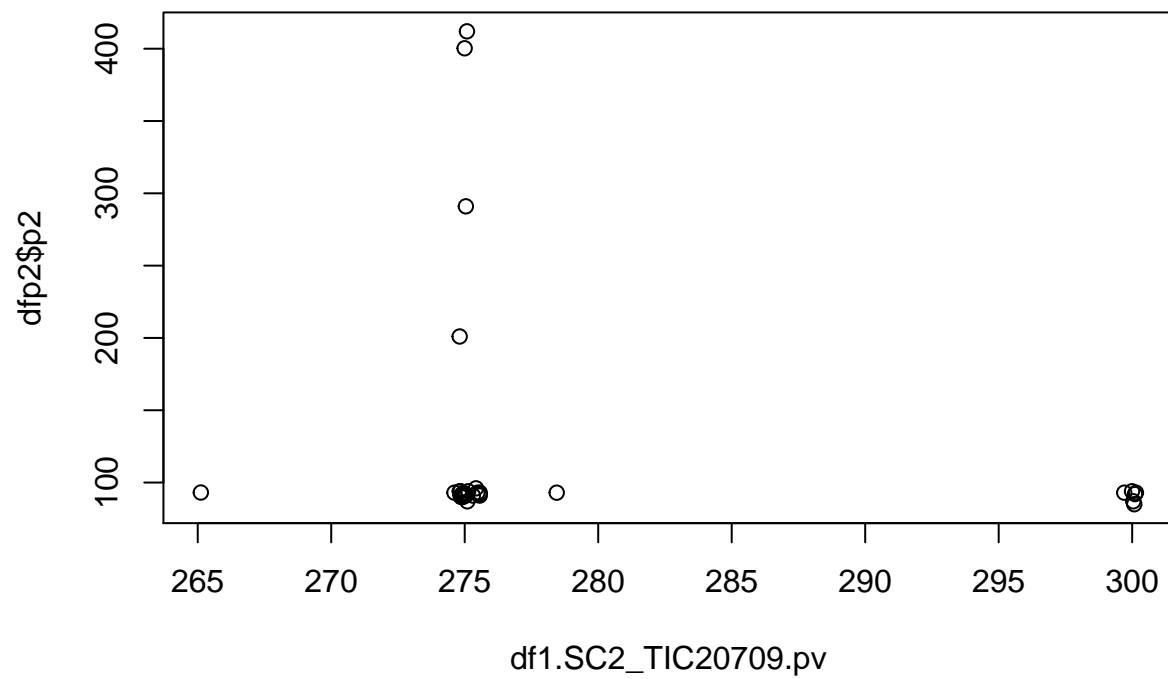


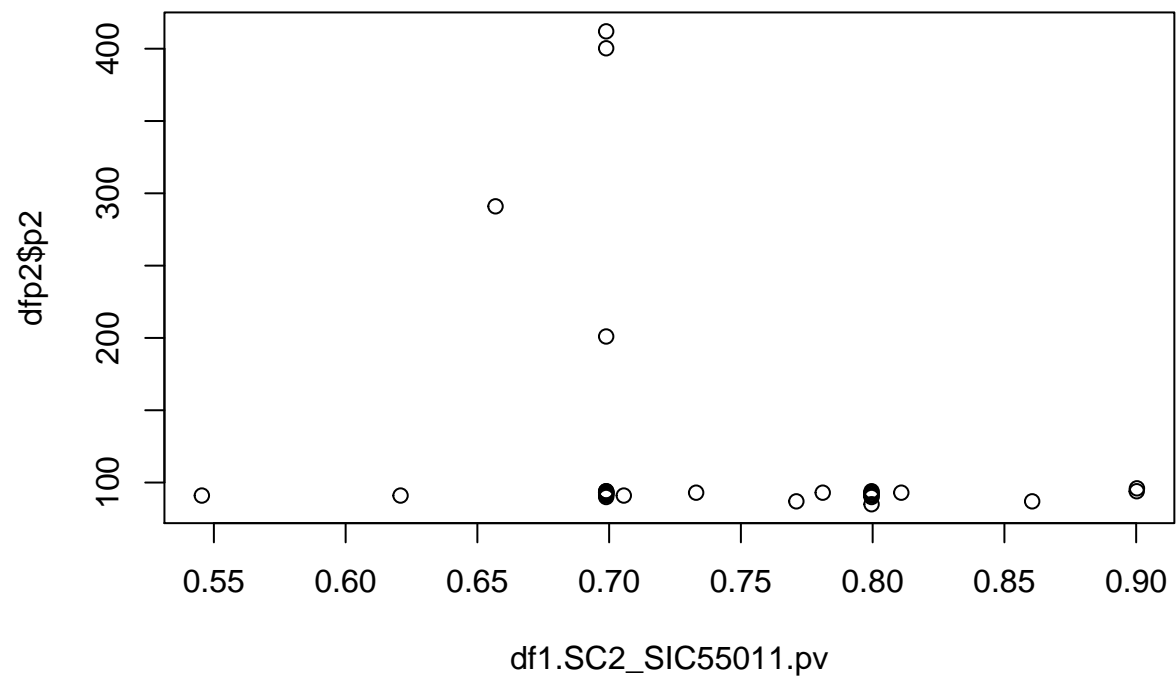




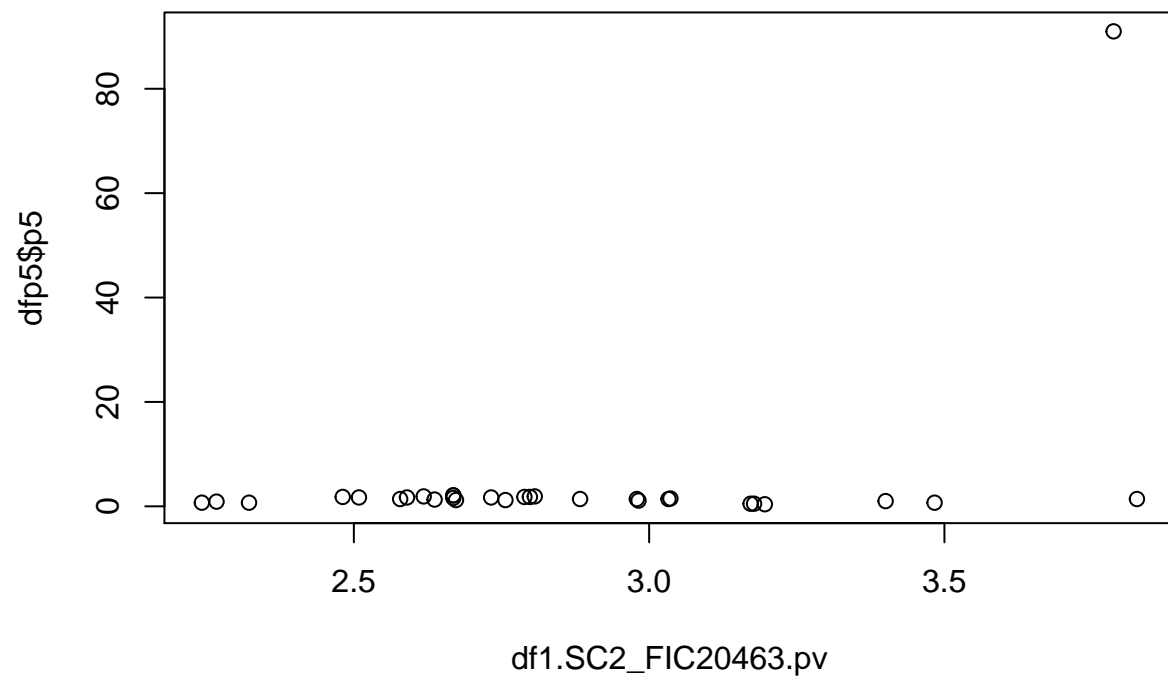


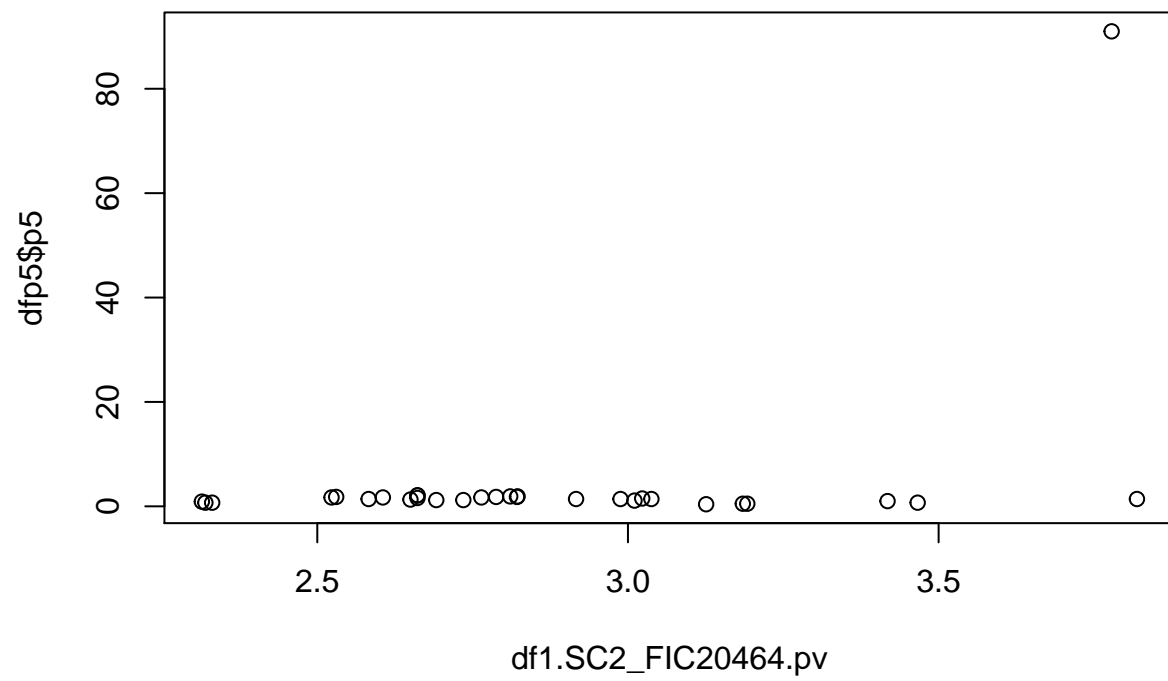


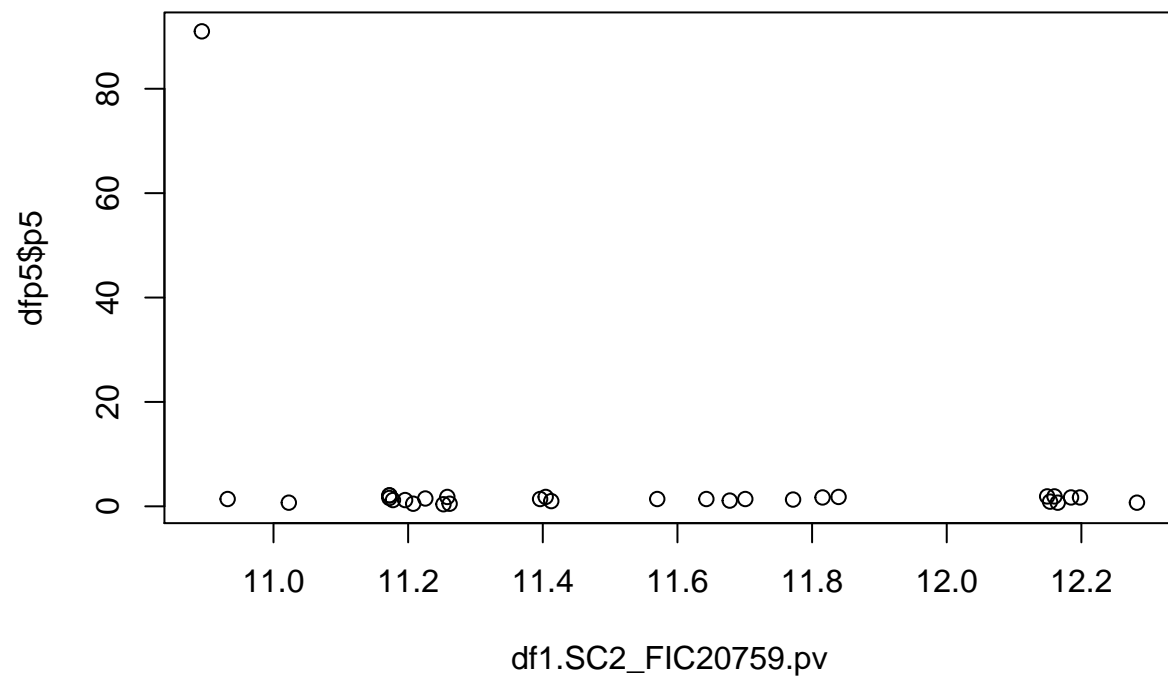


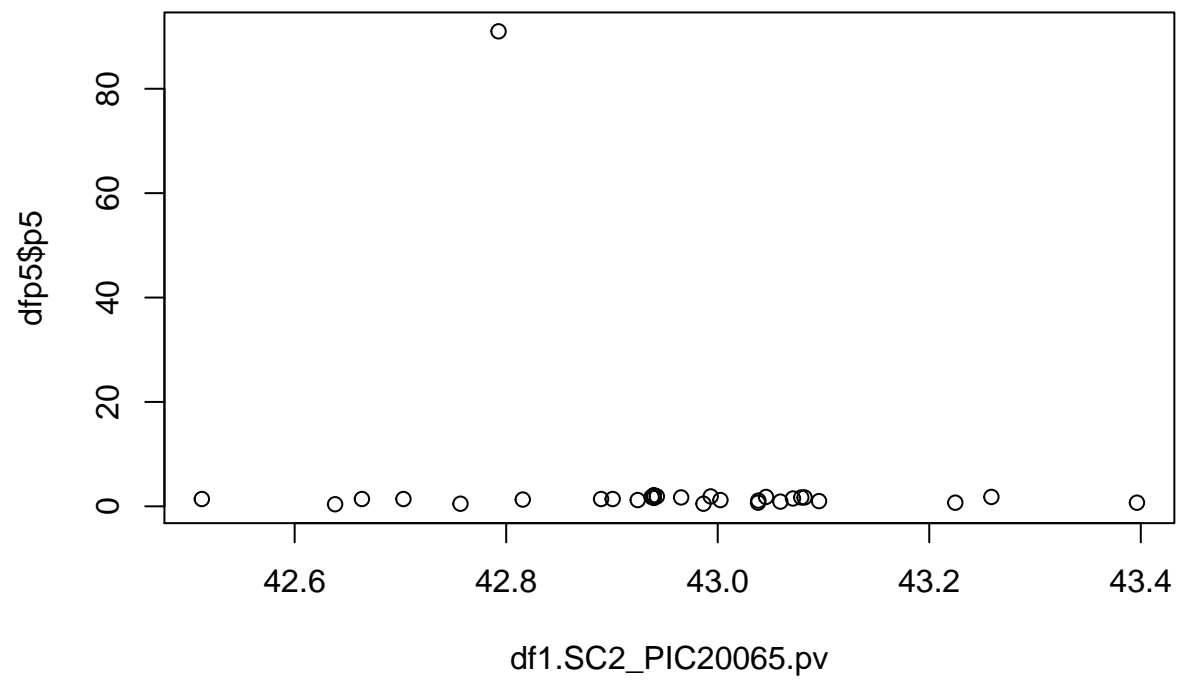


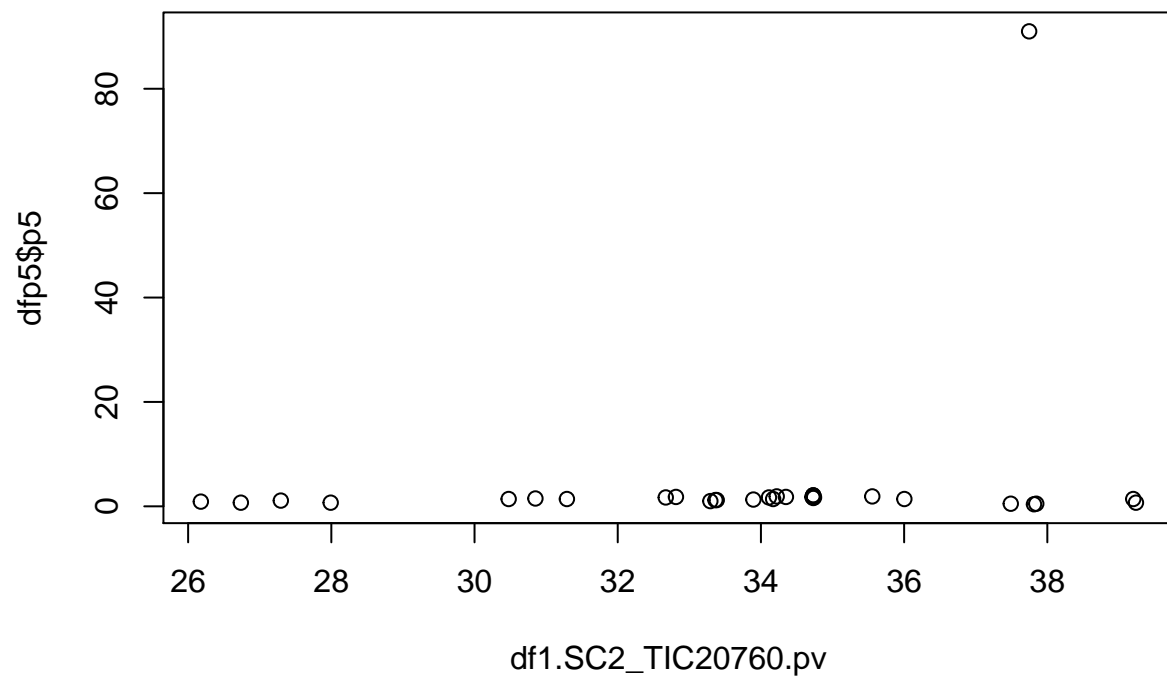
```
dfp5<-data.frame(p5,input)
plot(dfp5$p5~.,dfp5)
```

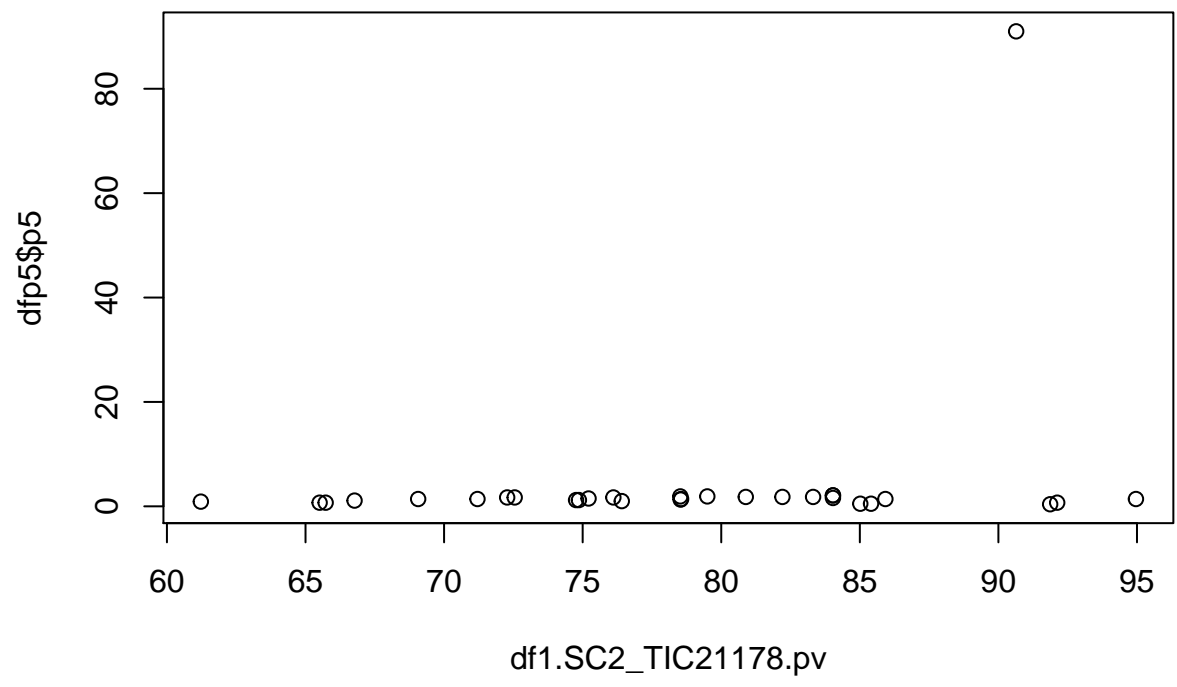


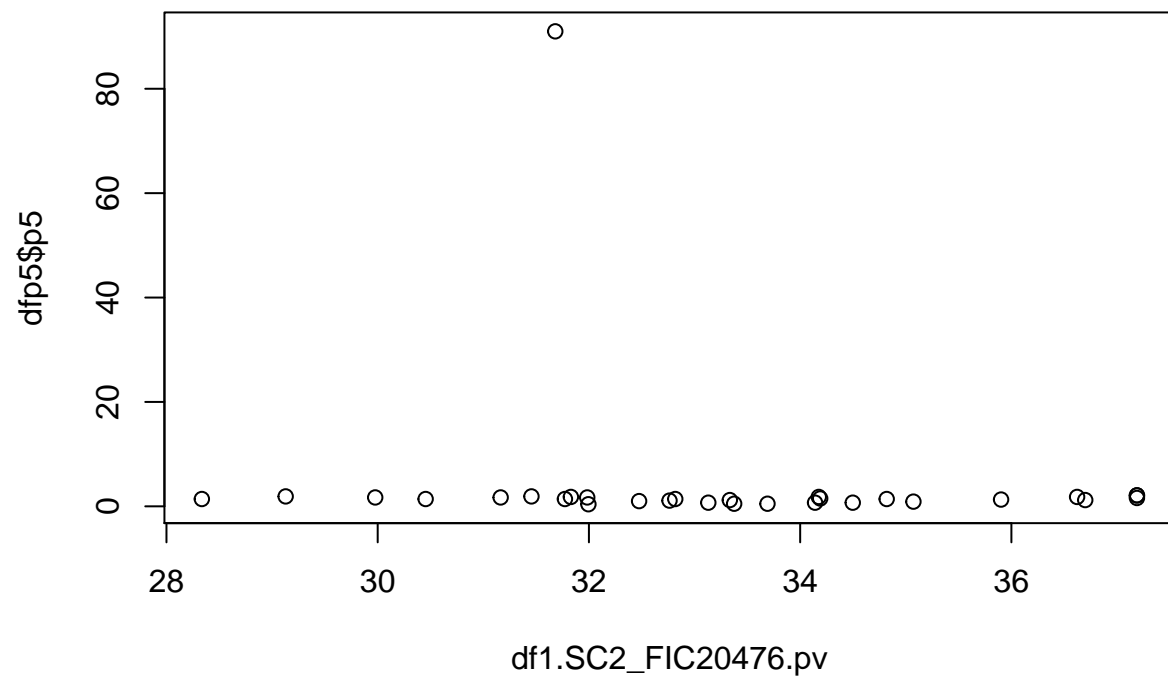


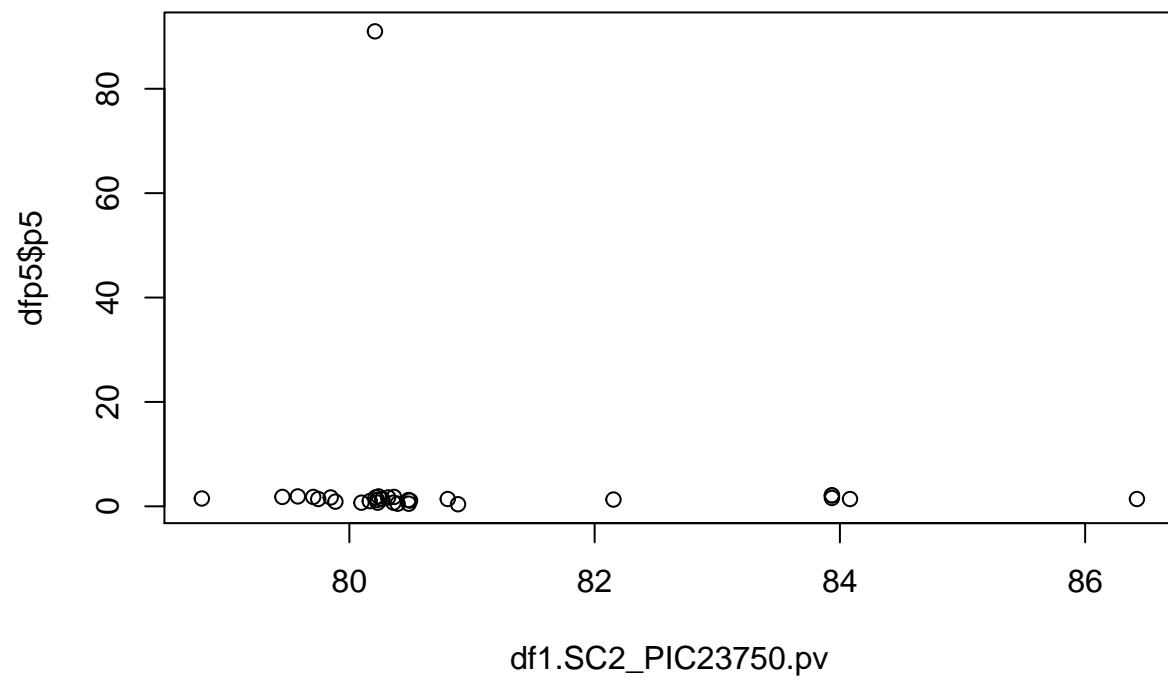


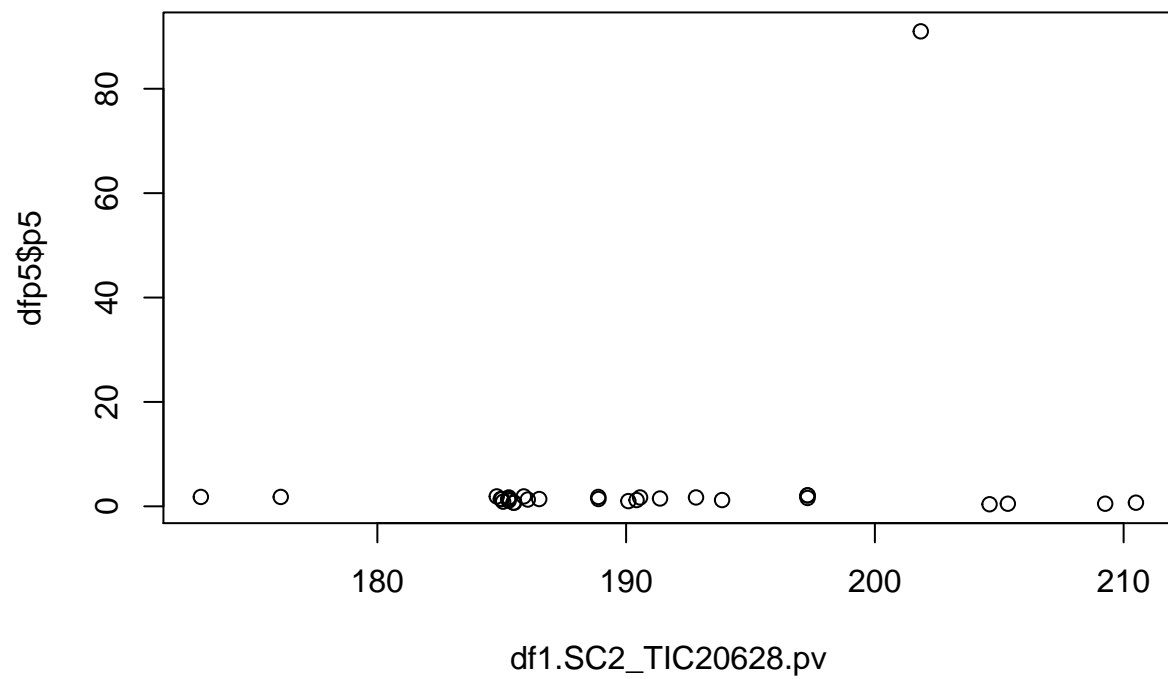


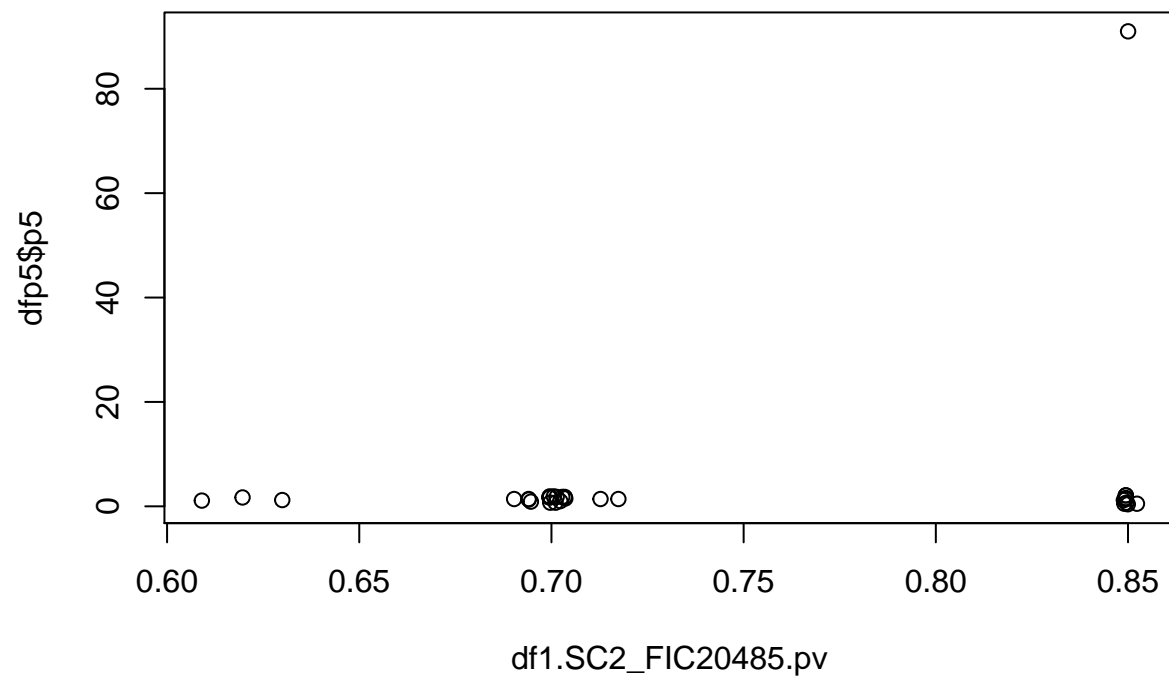


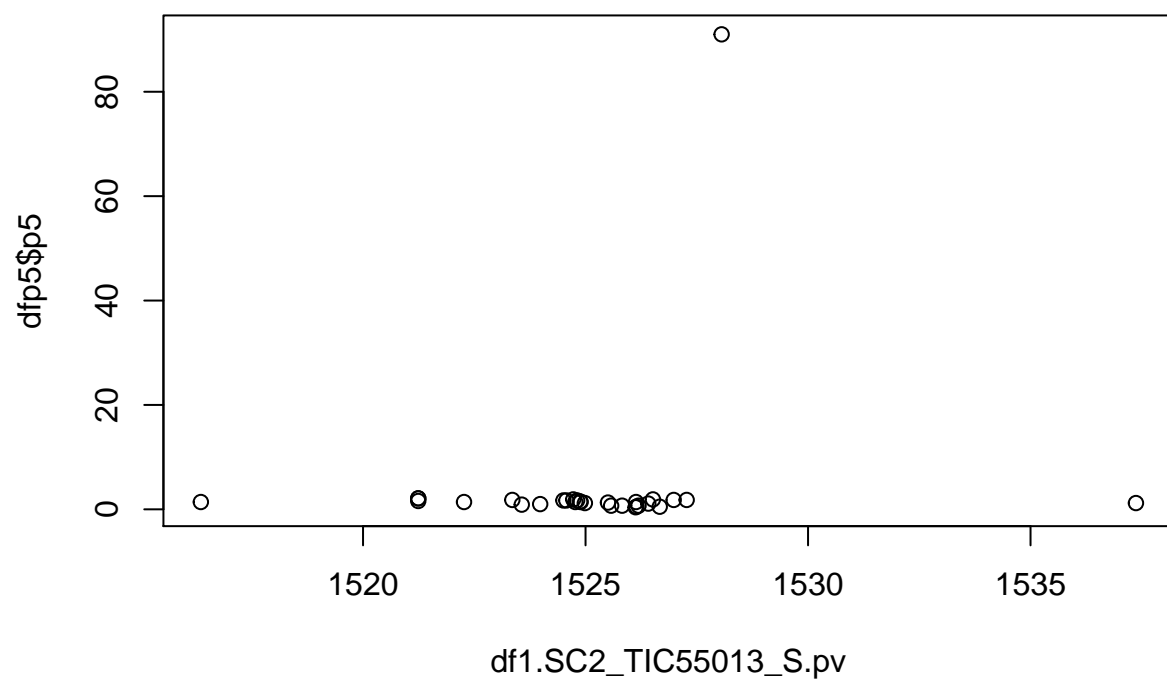


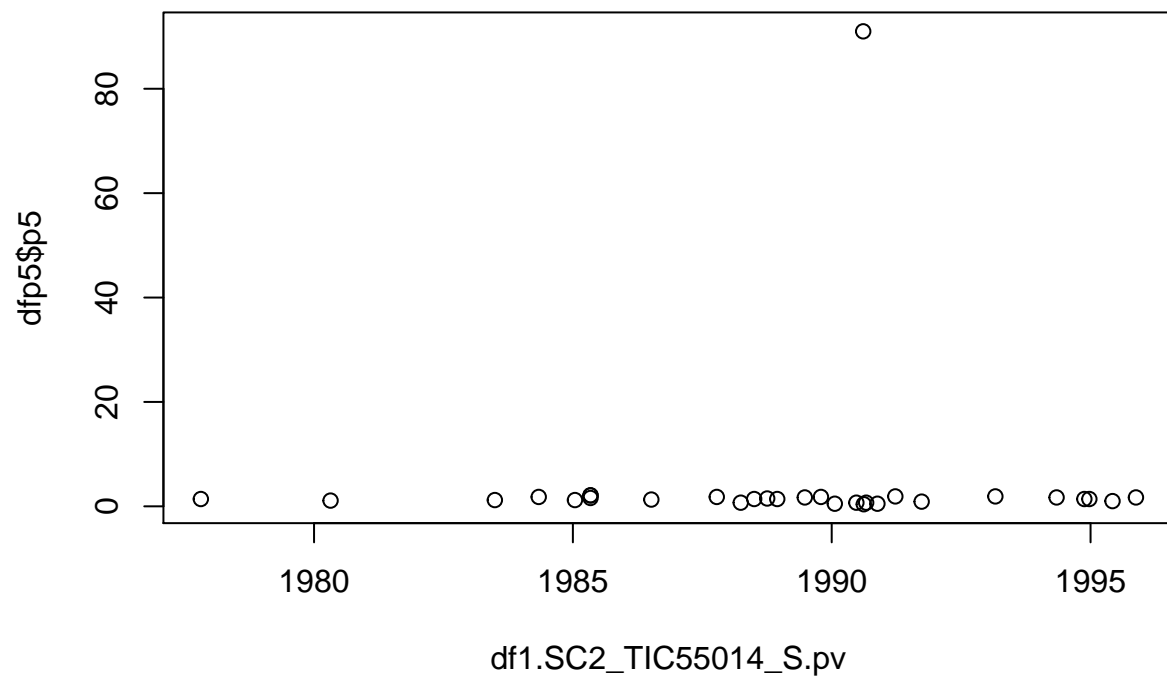


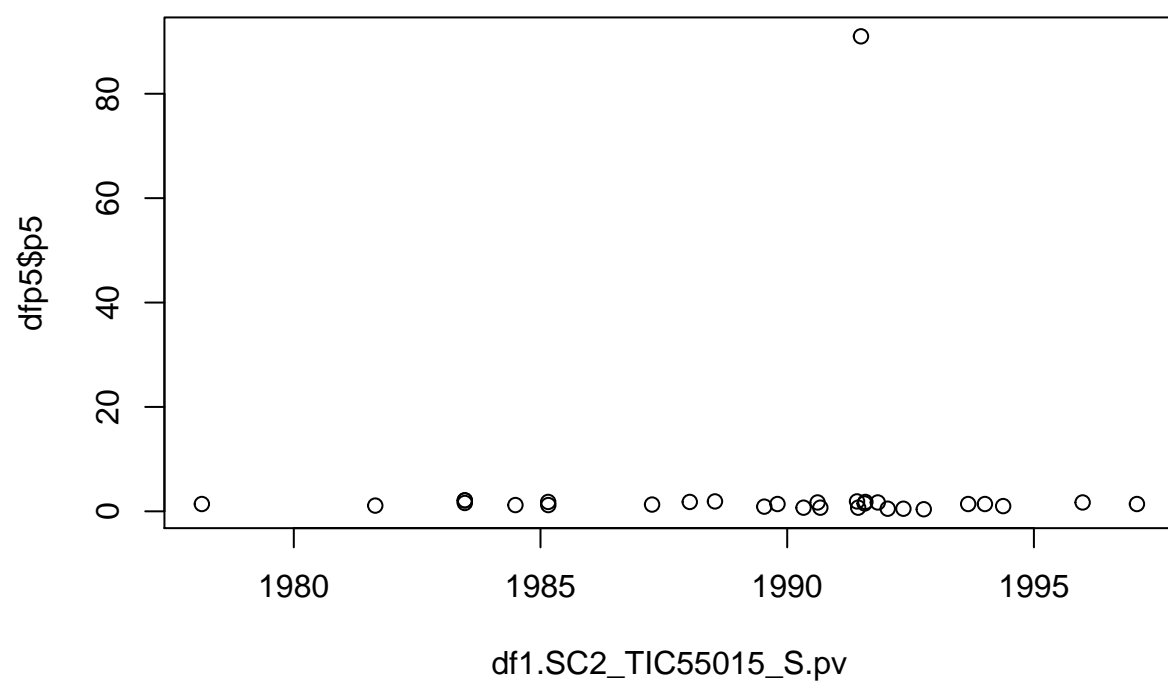


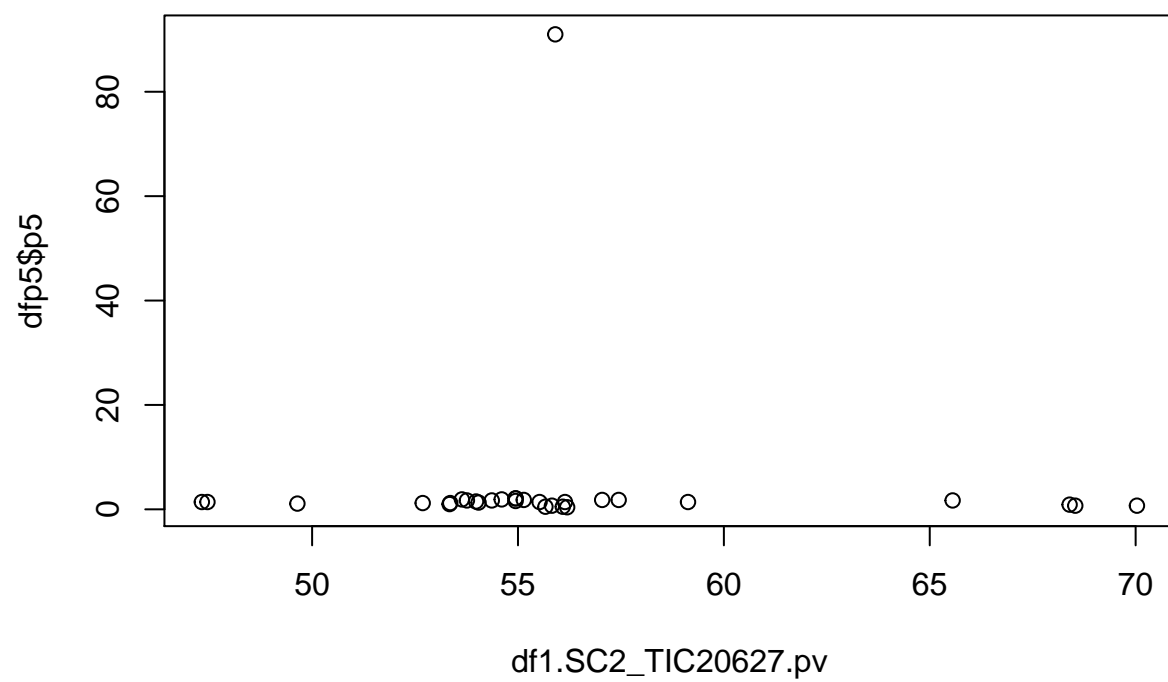


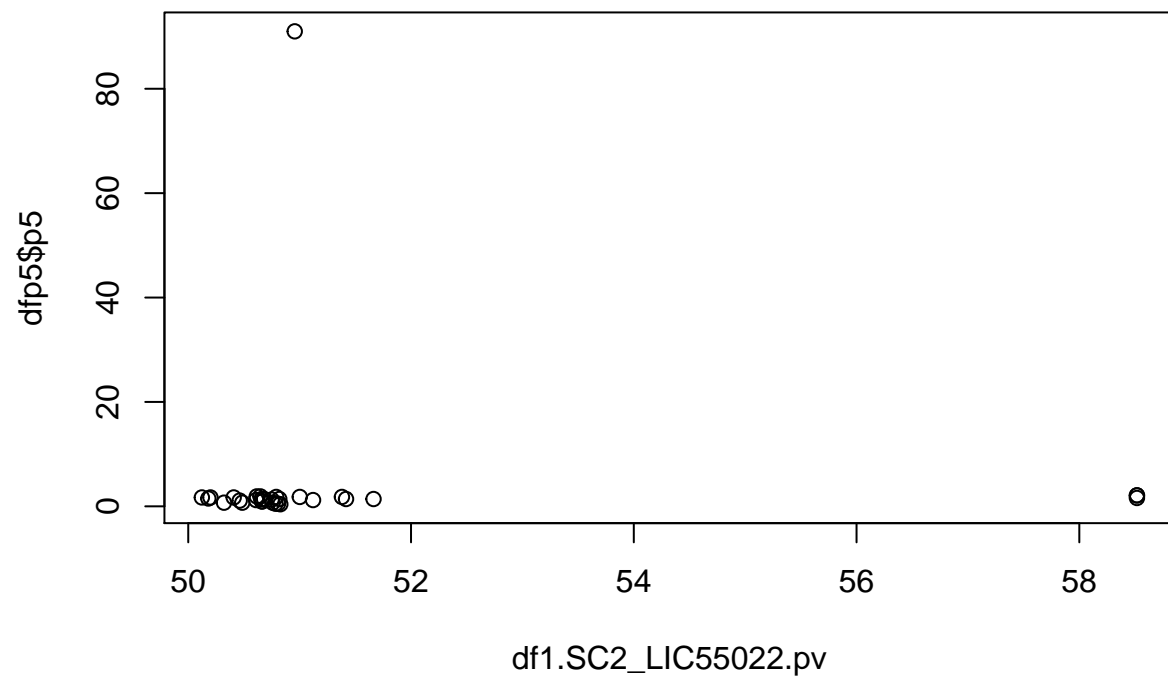


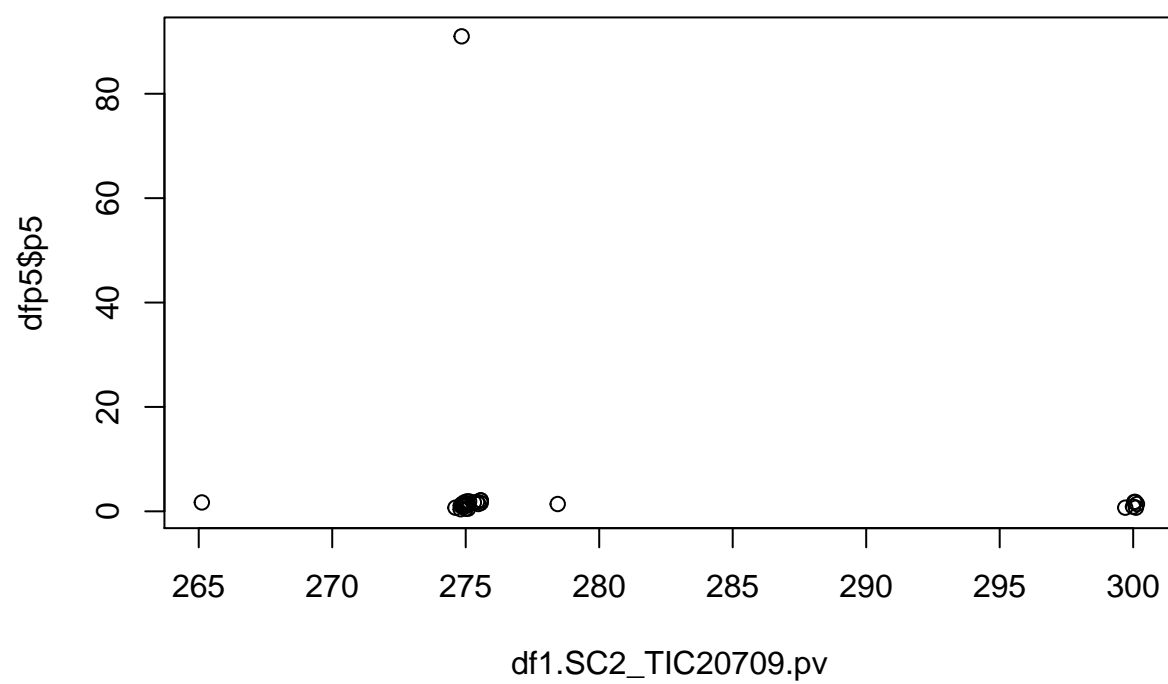


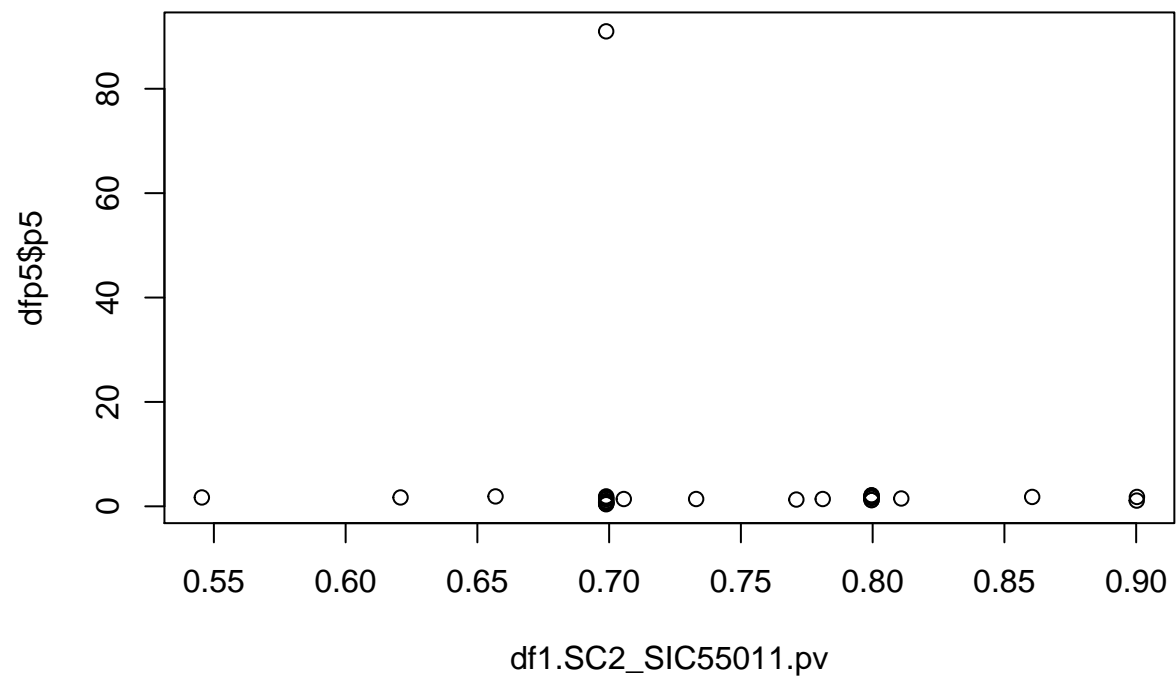




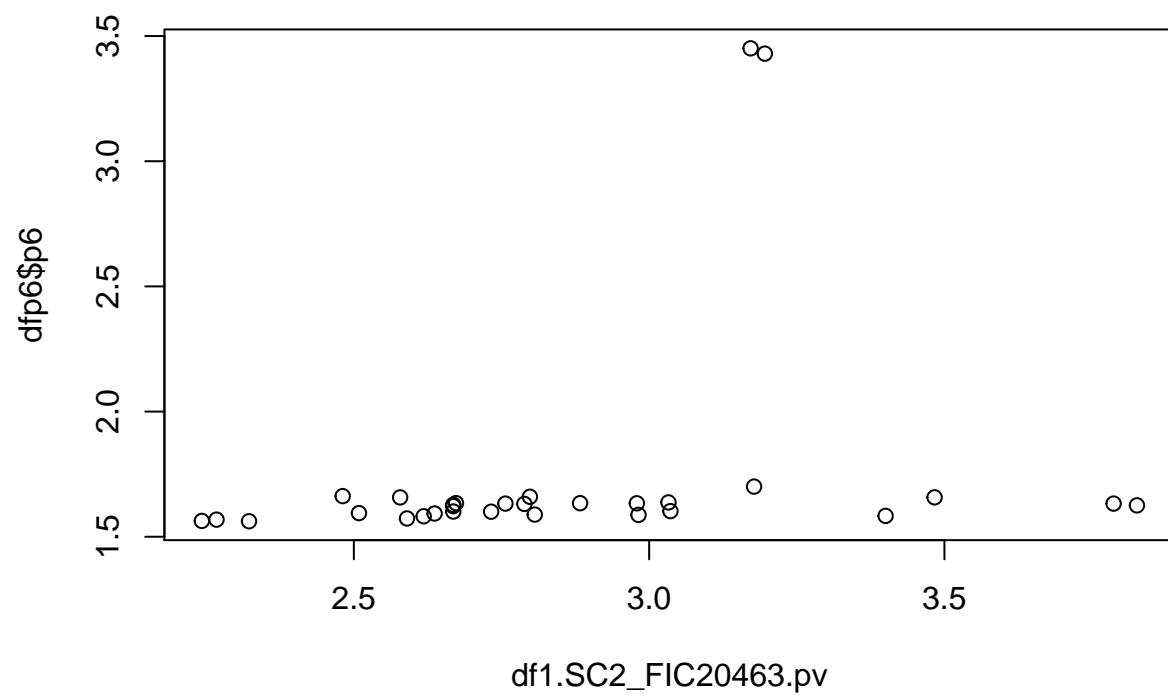


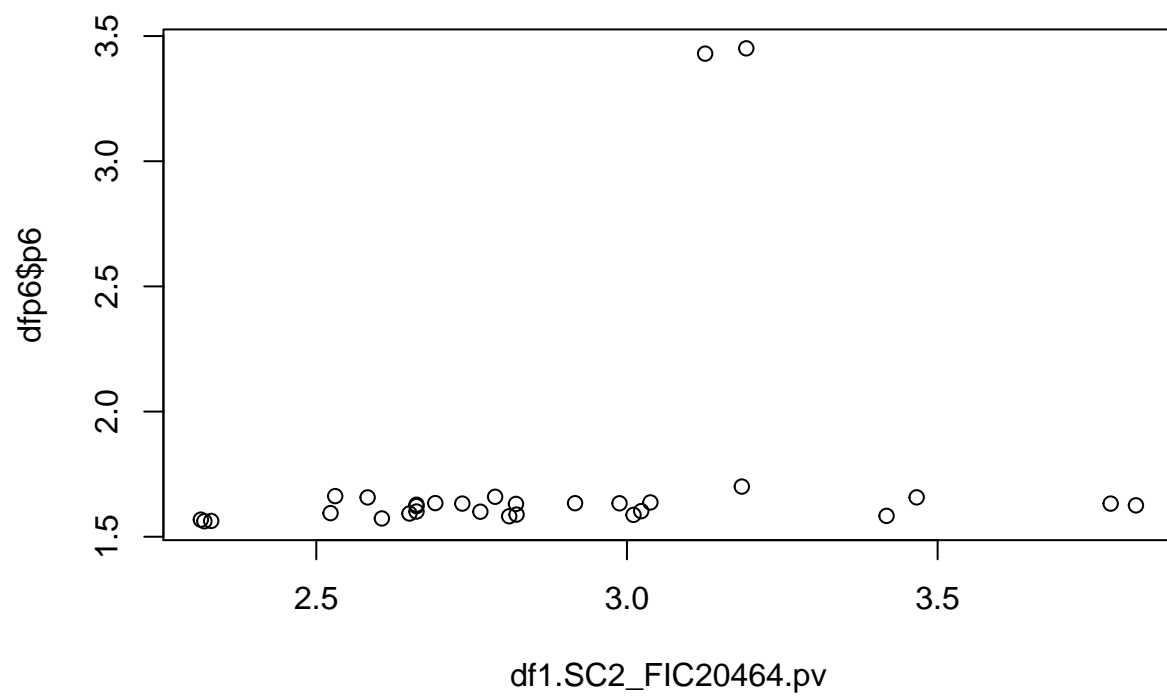


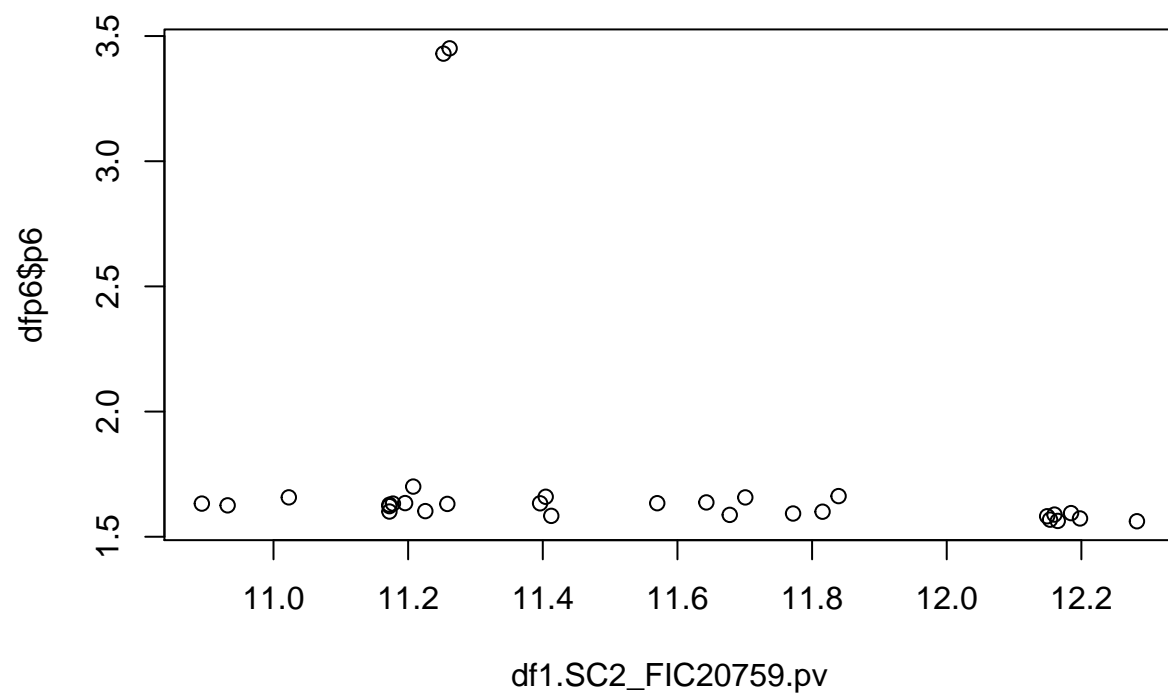


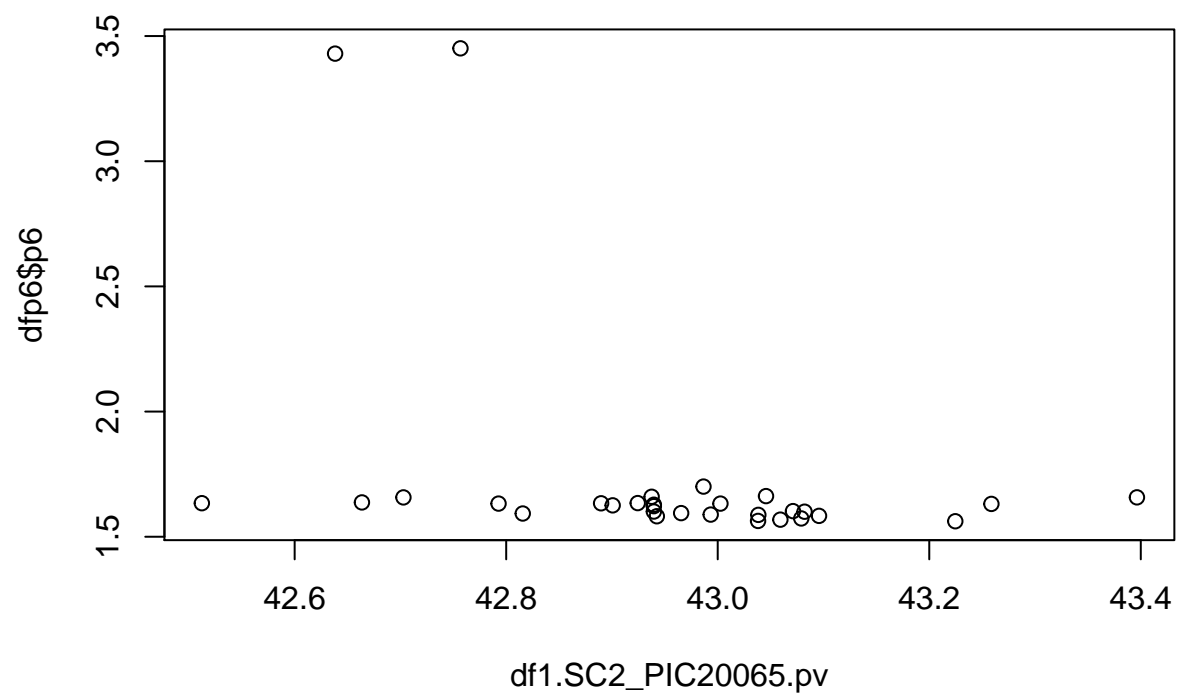


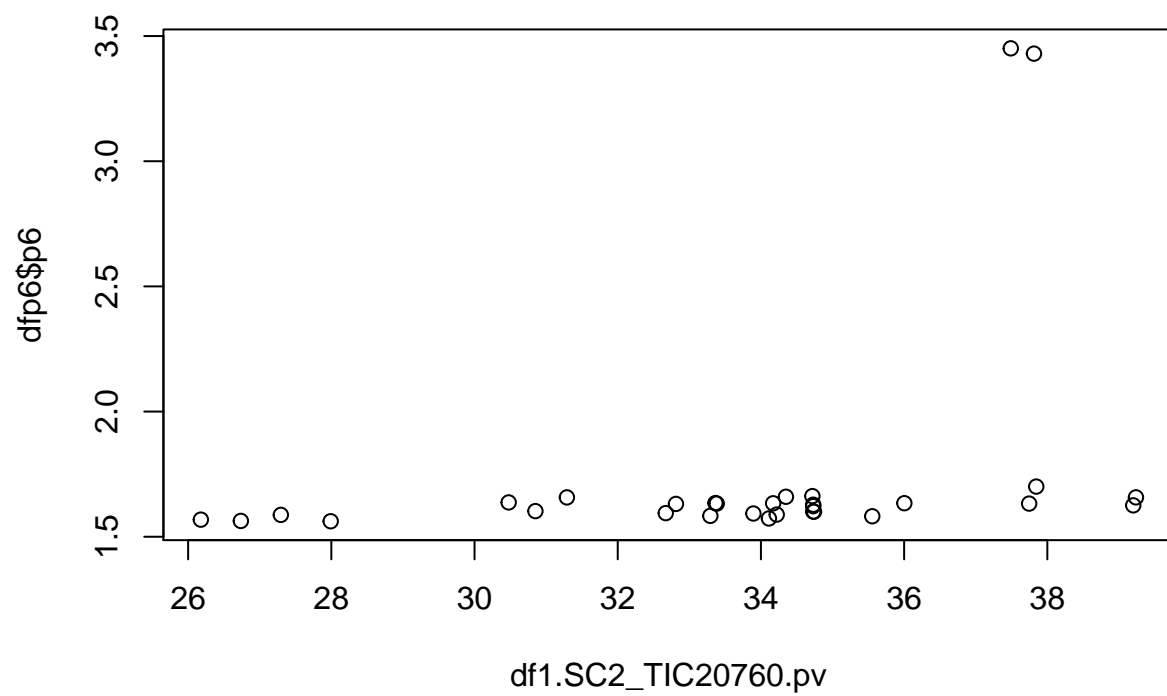
```
dfp6<-data.frame(p6,input)
plot(dfp6$p6~.,dfp6)
```

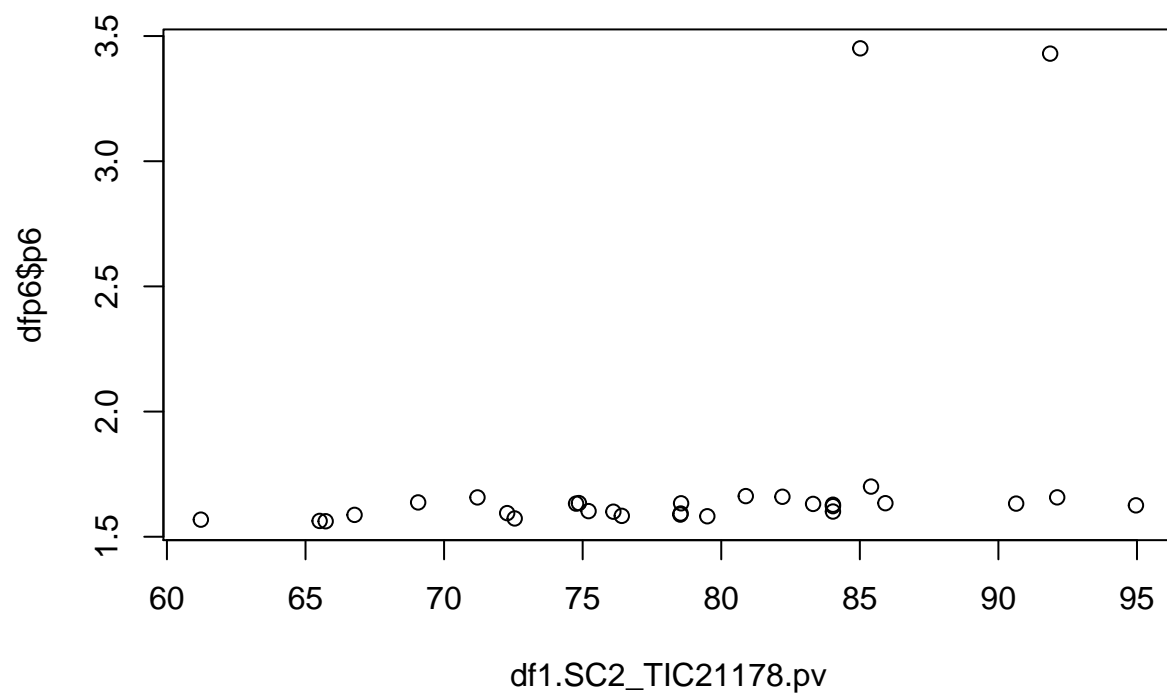


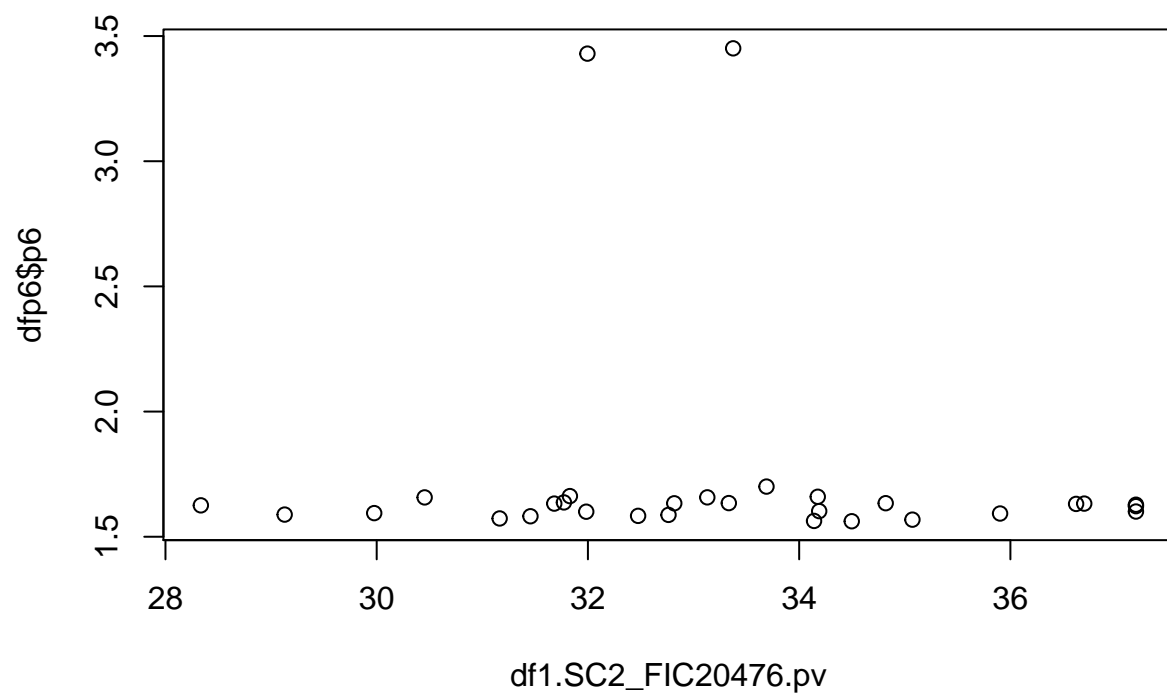


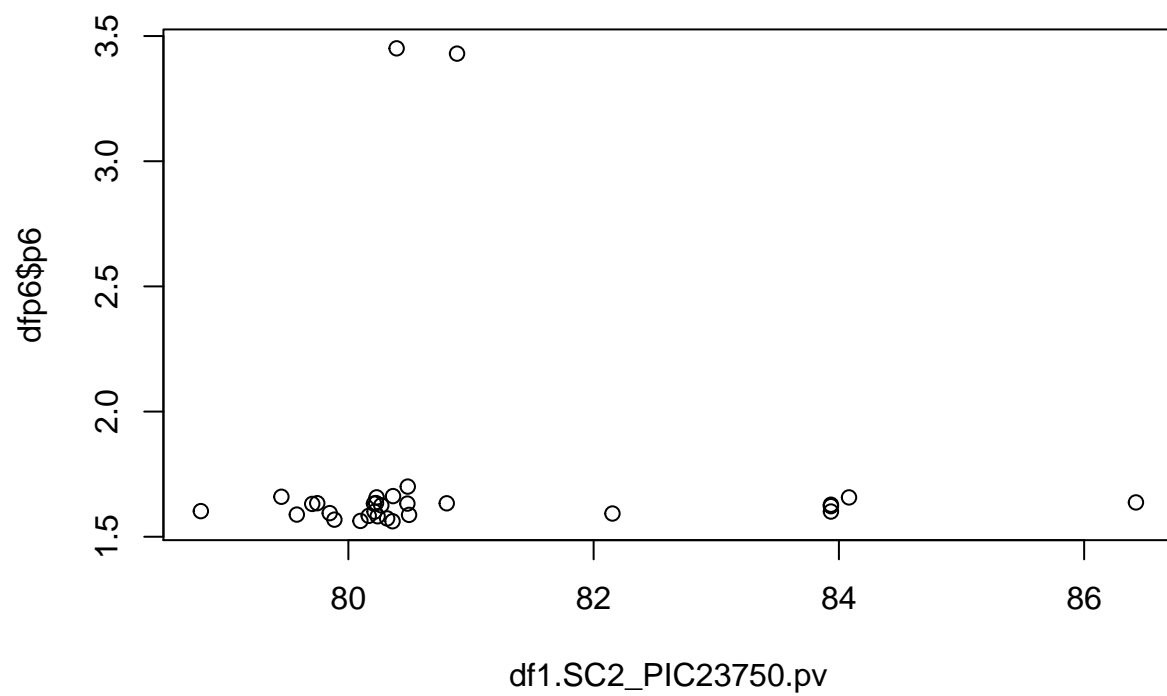


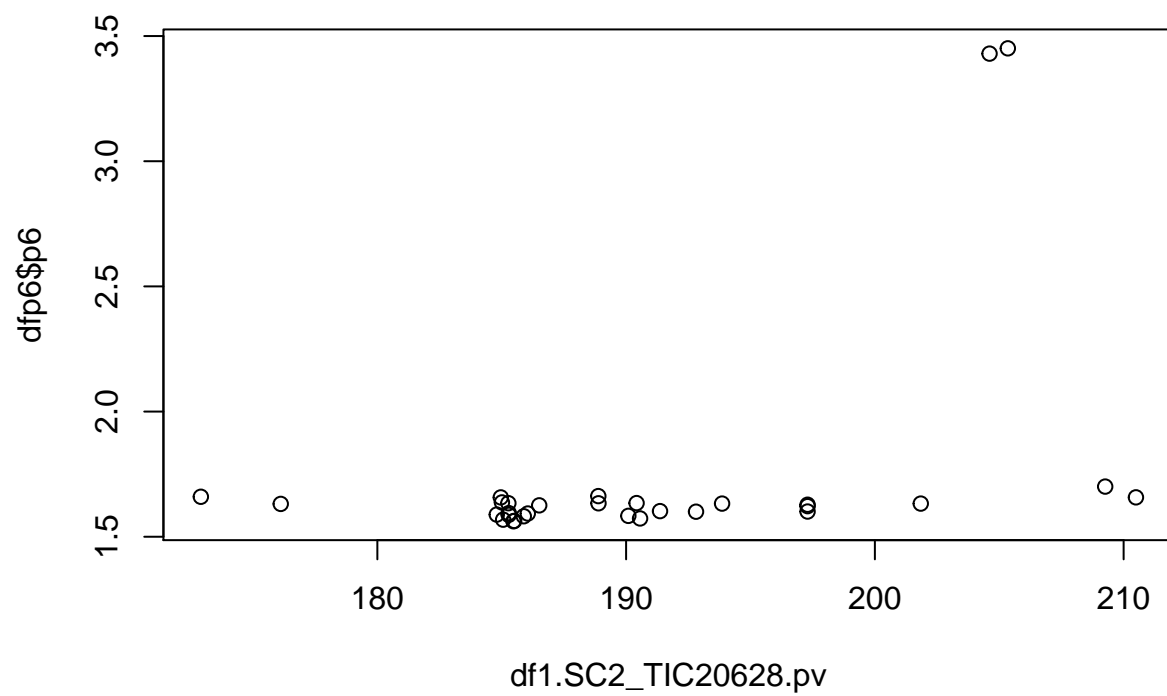


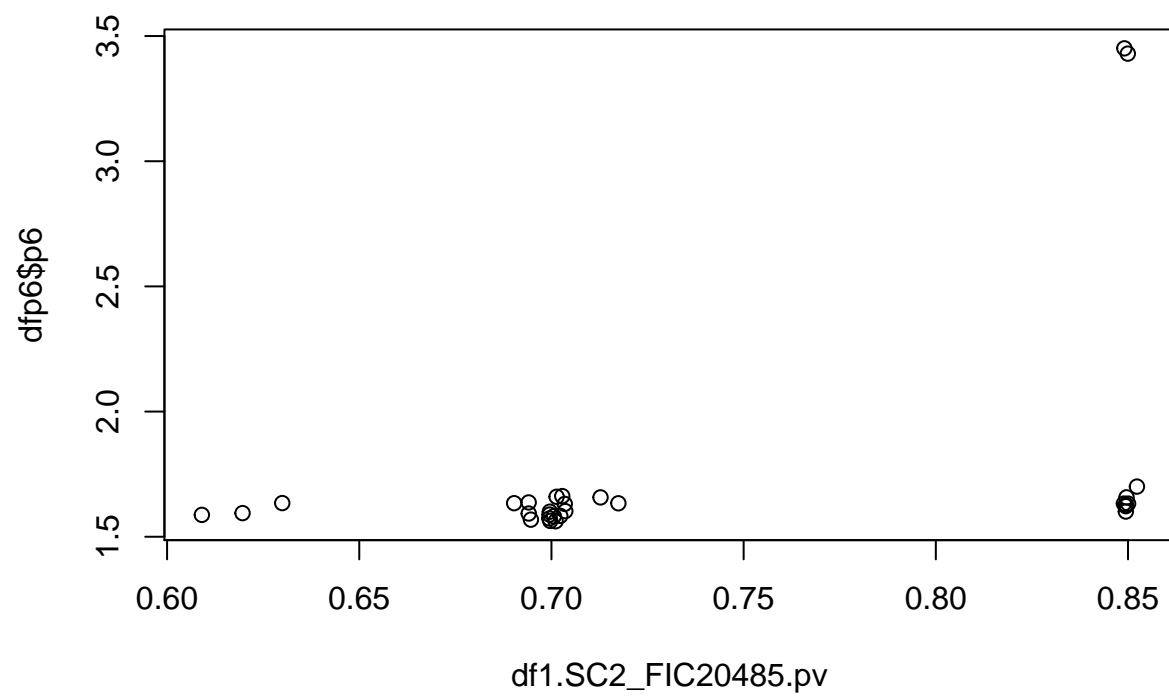




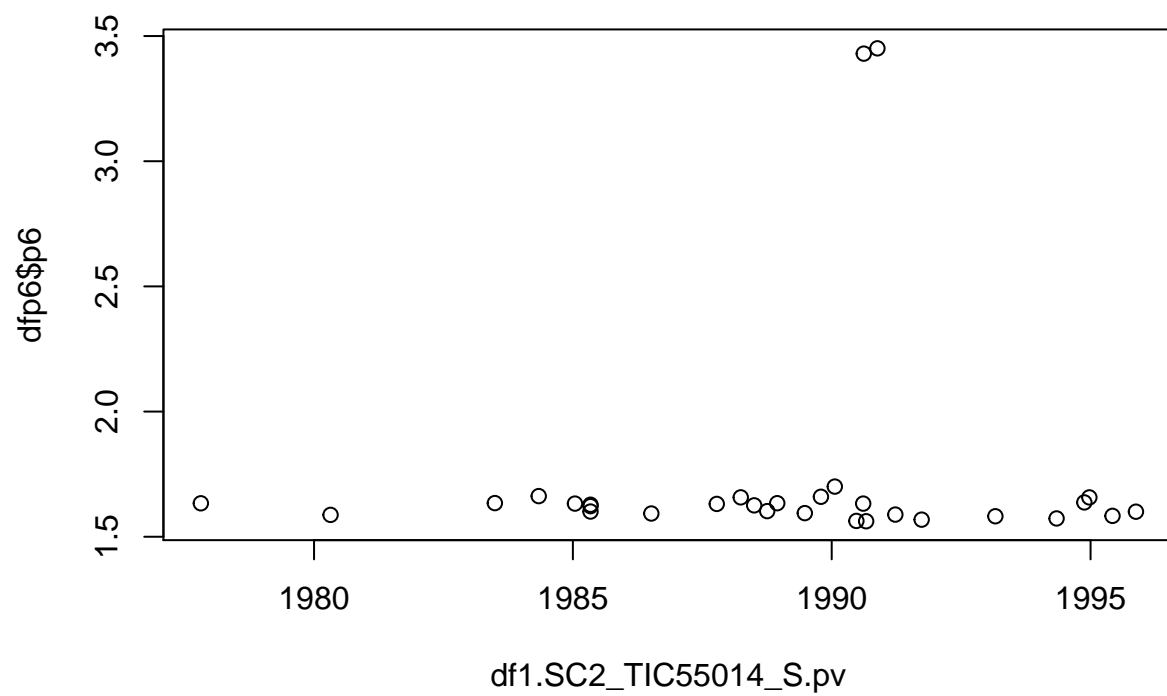


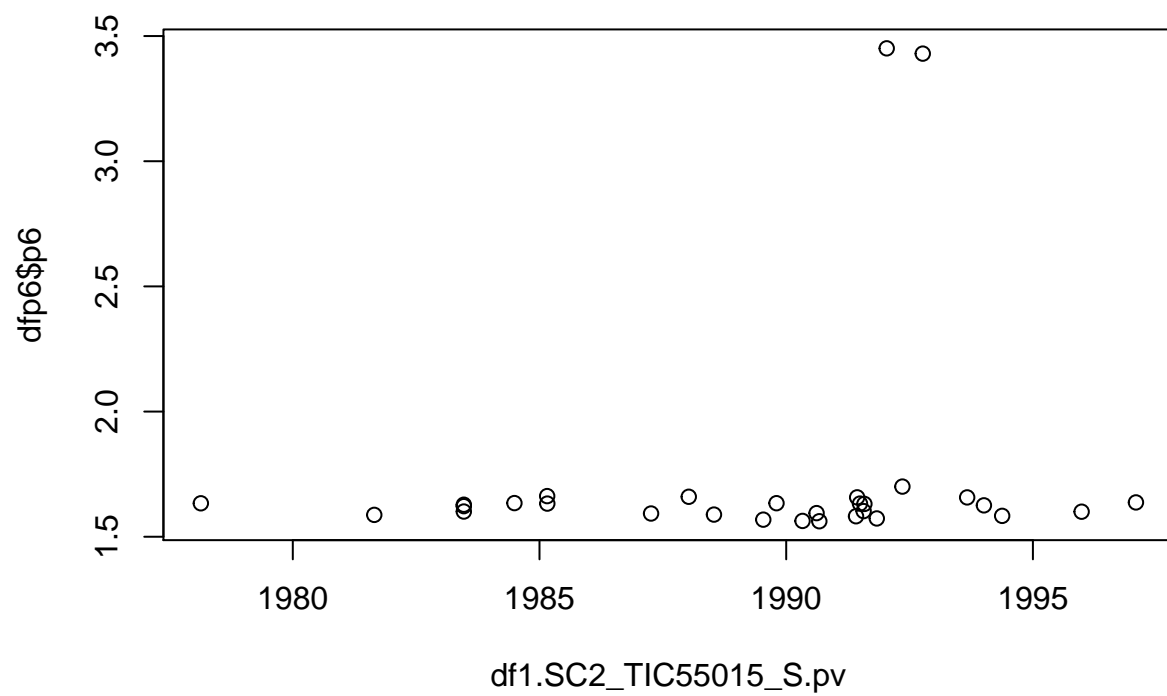




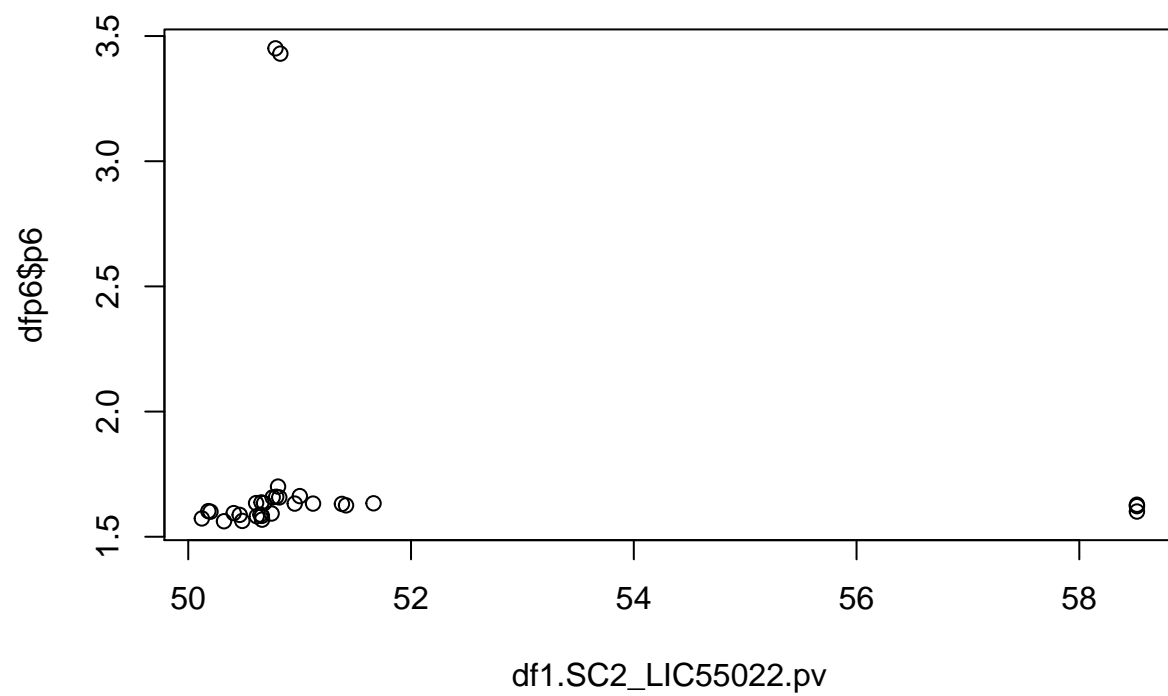


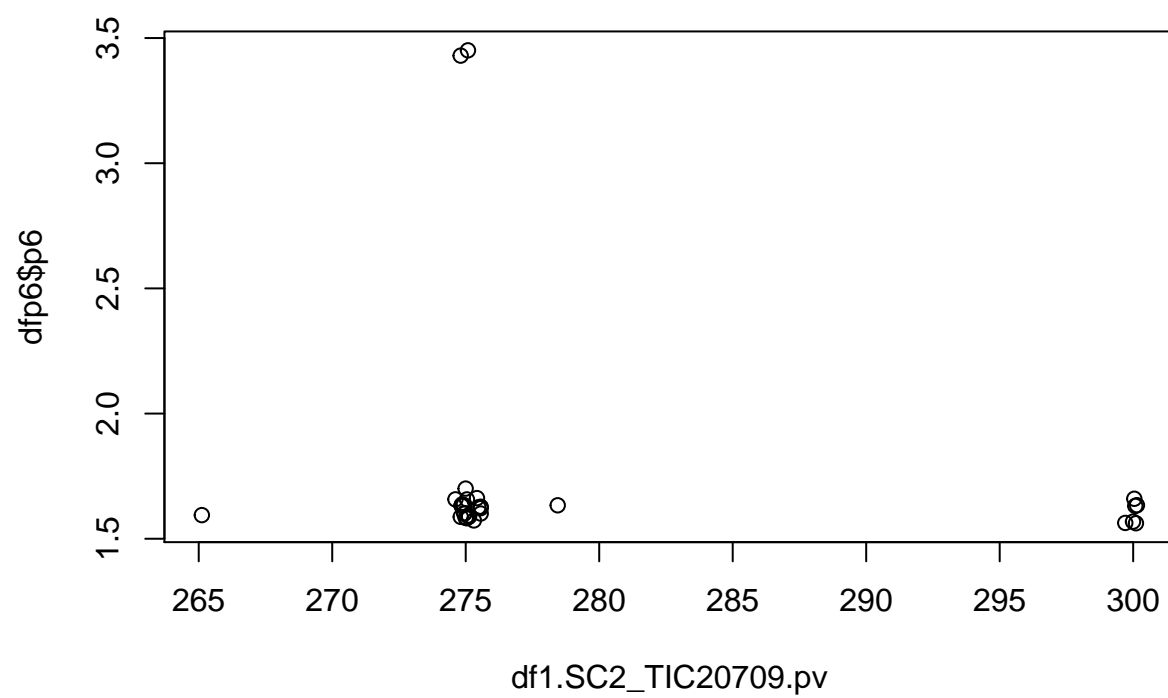


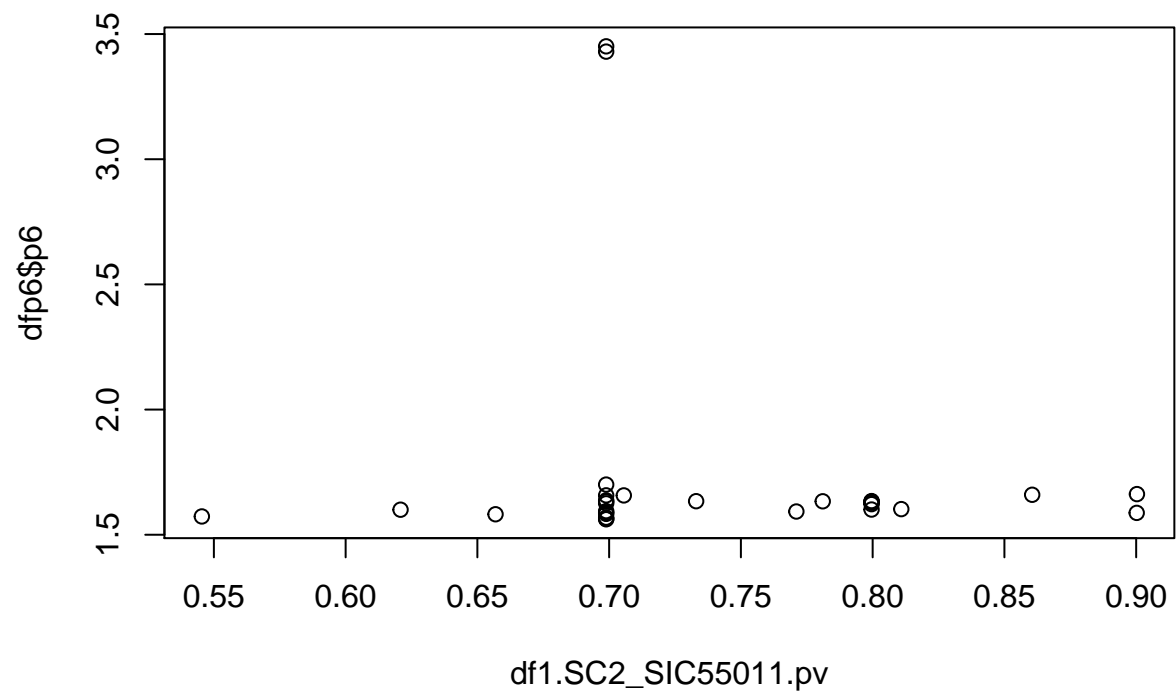




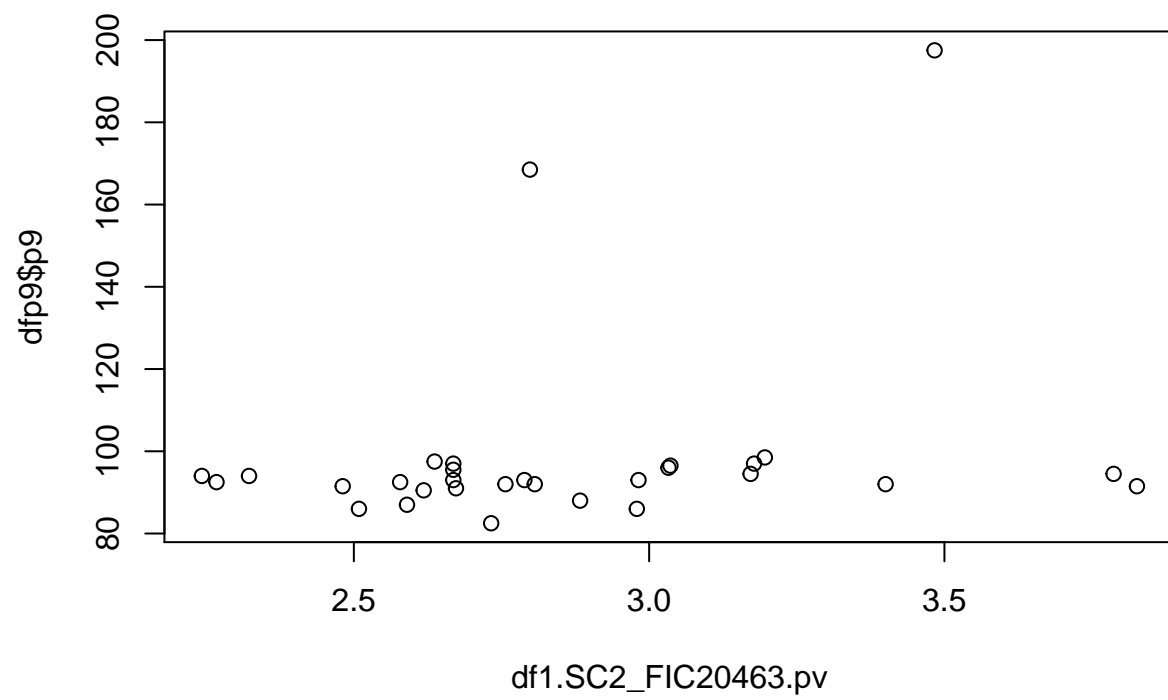


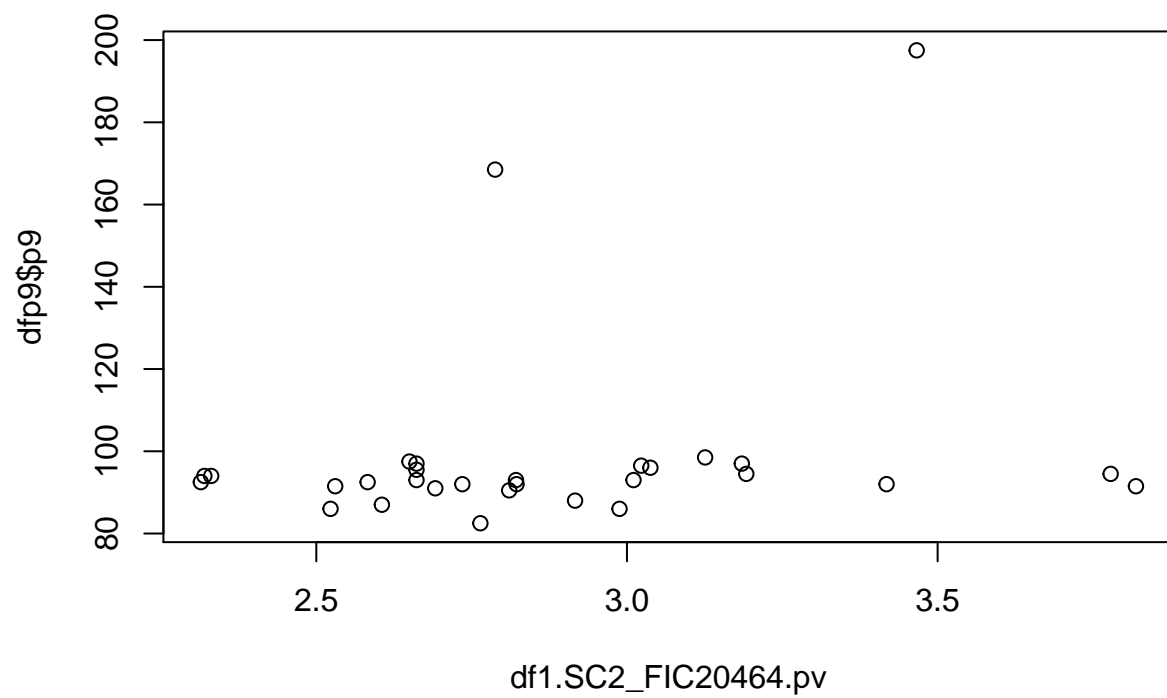


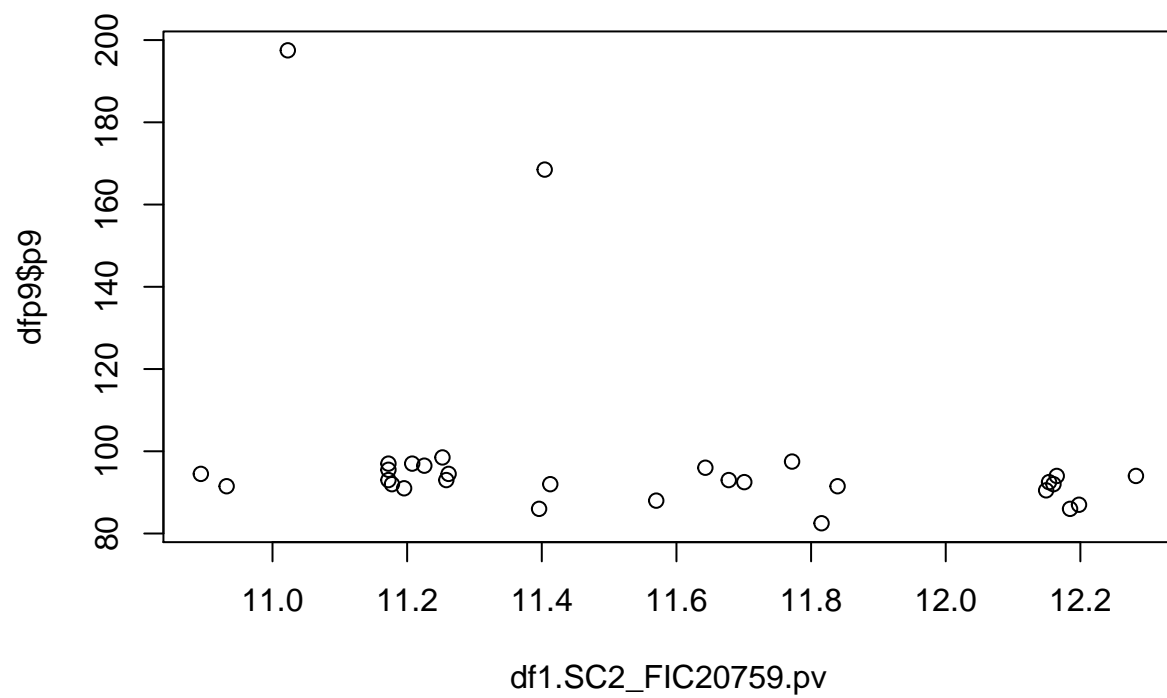


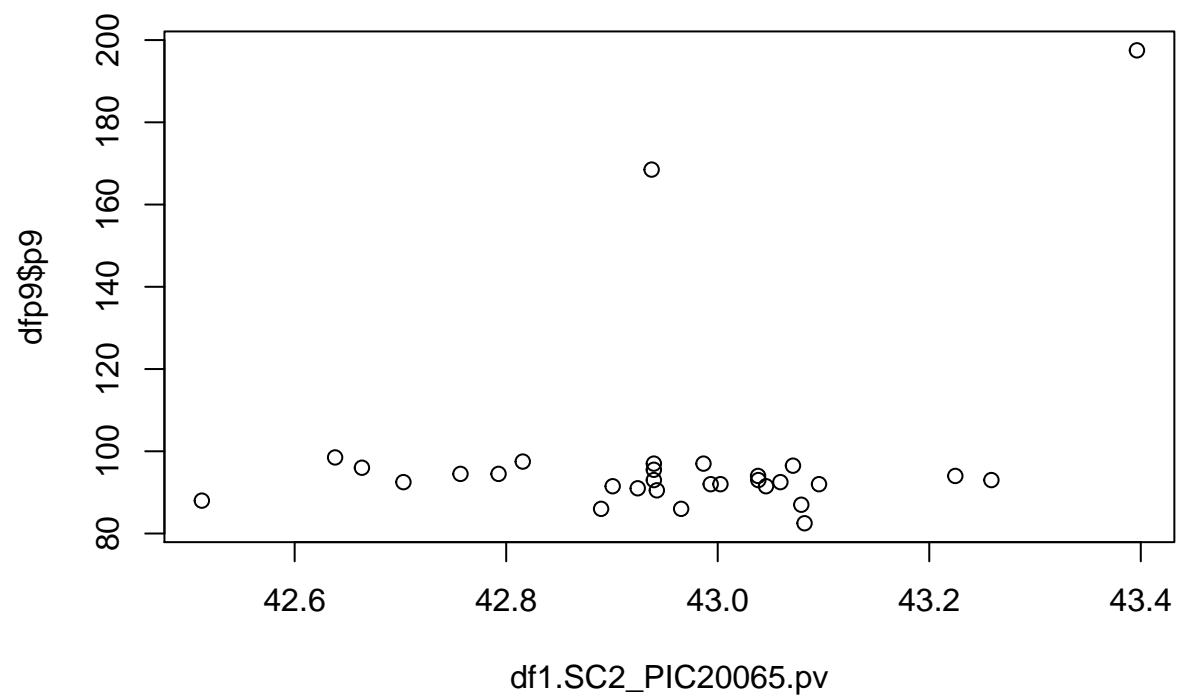


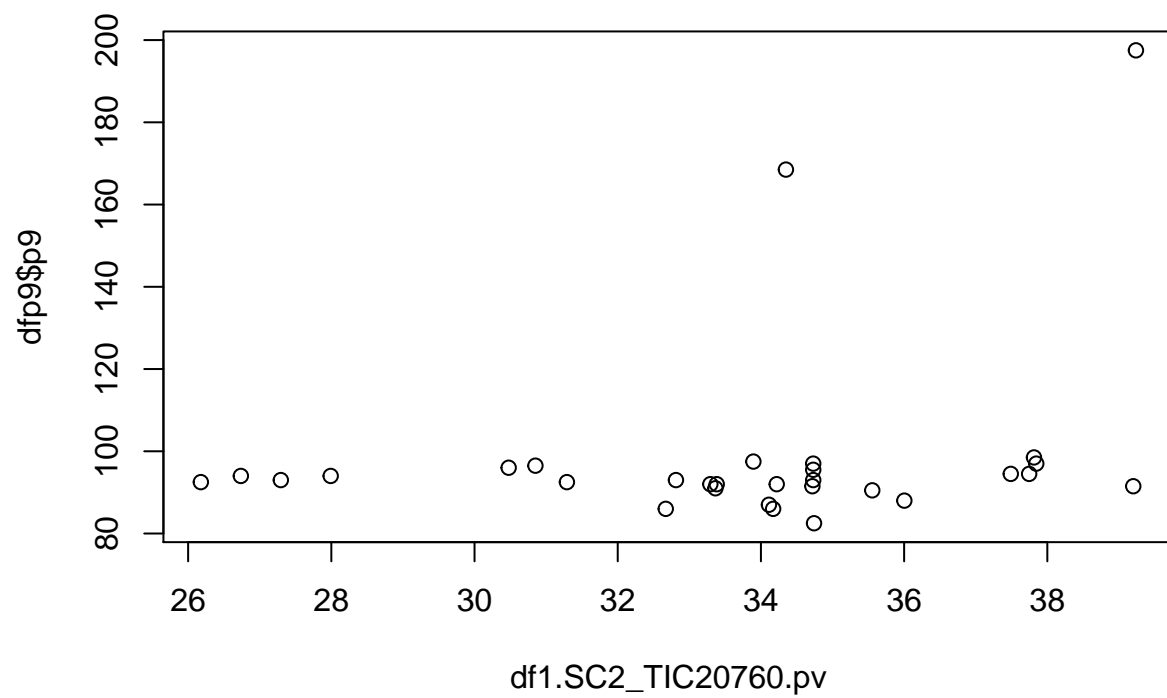
```
dfp9<-data.frame(p9,input)
plot(dfp9$p9~.,dfp9)
```

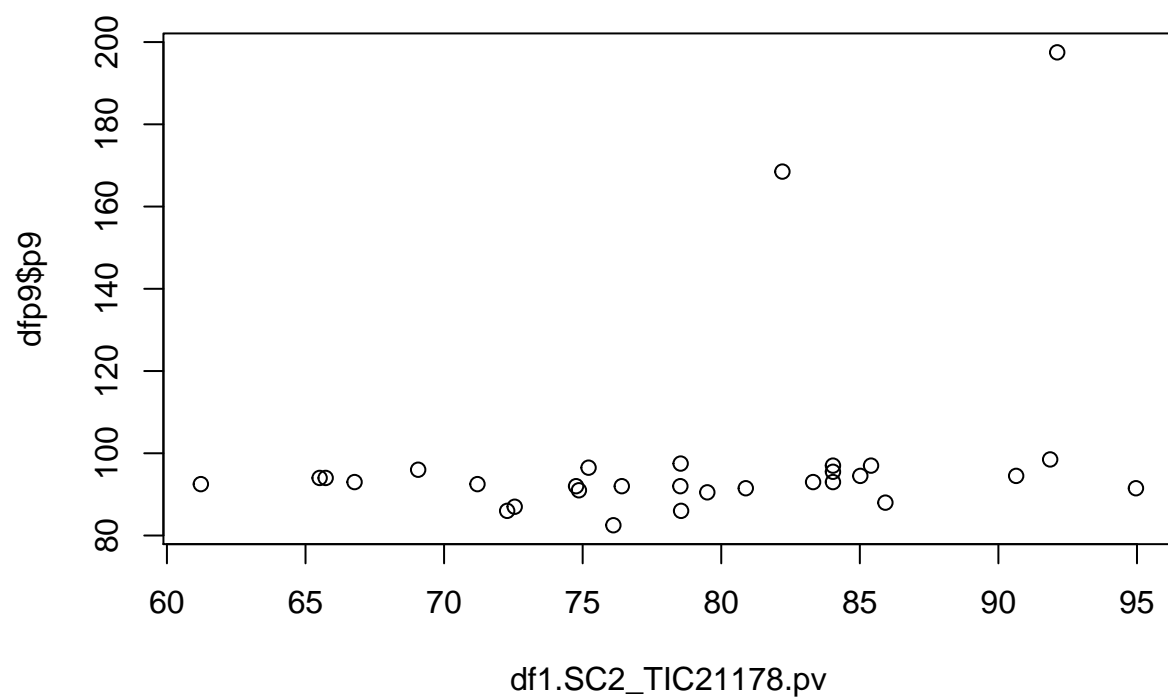



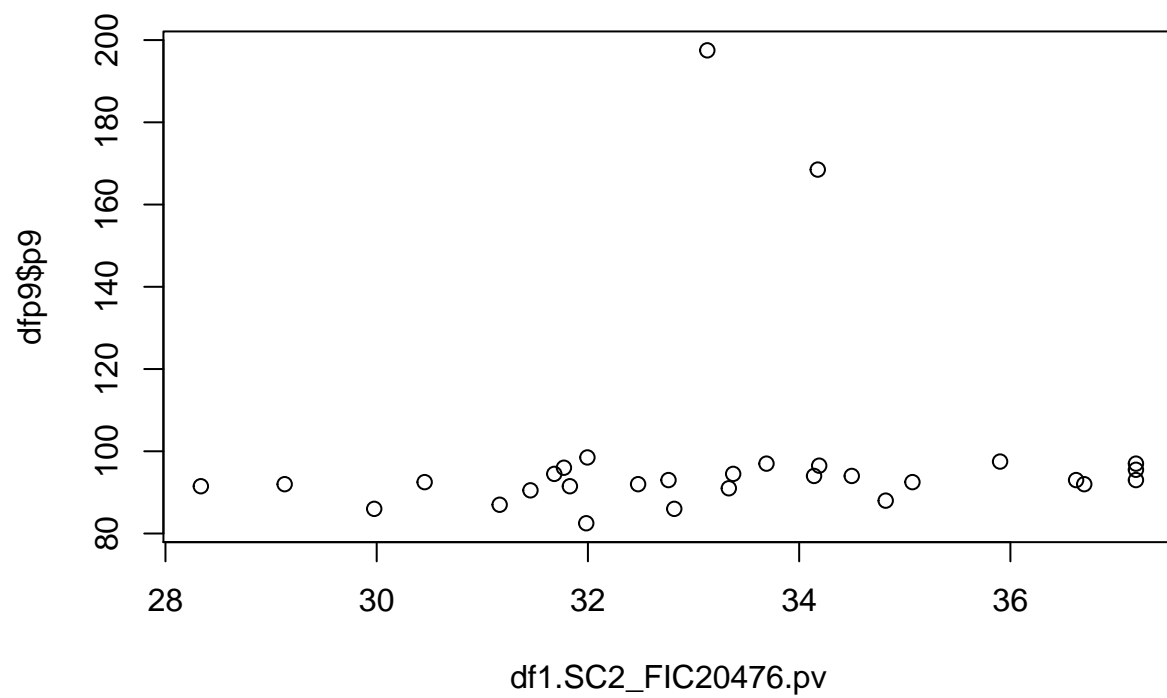


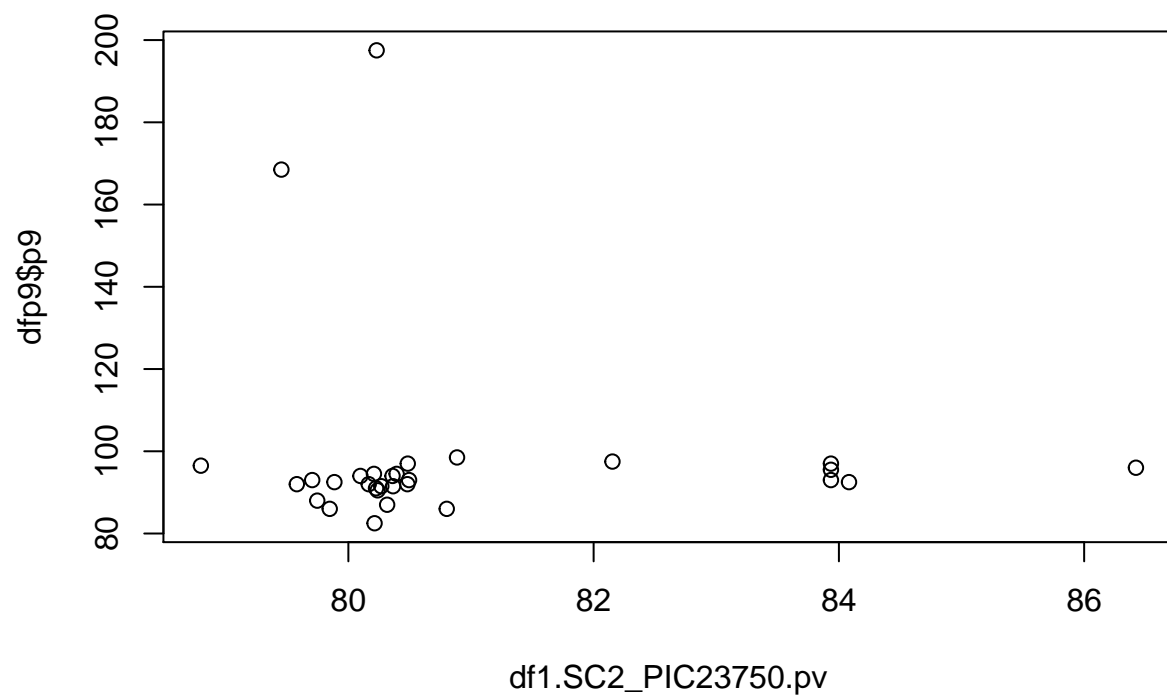


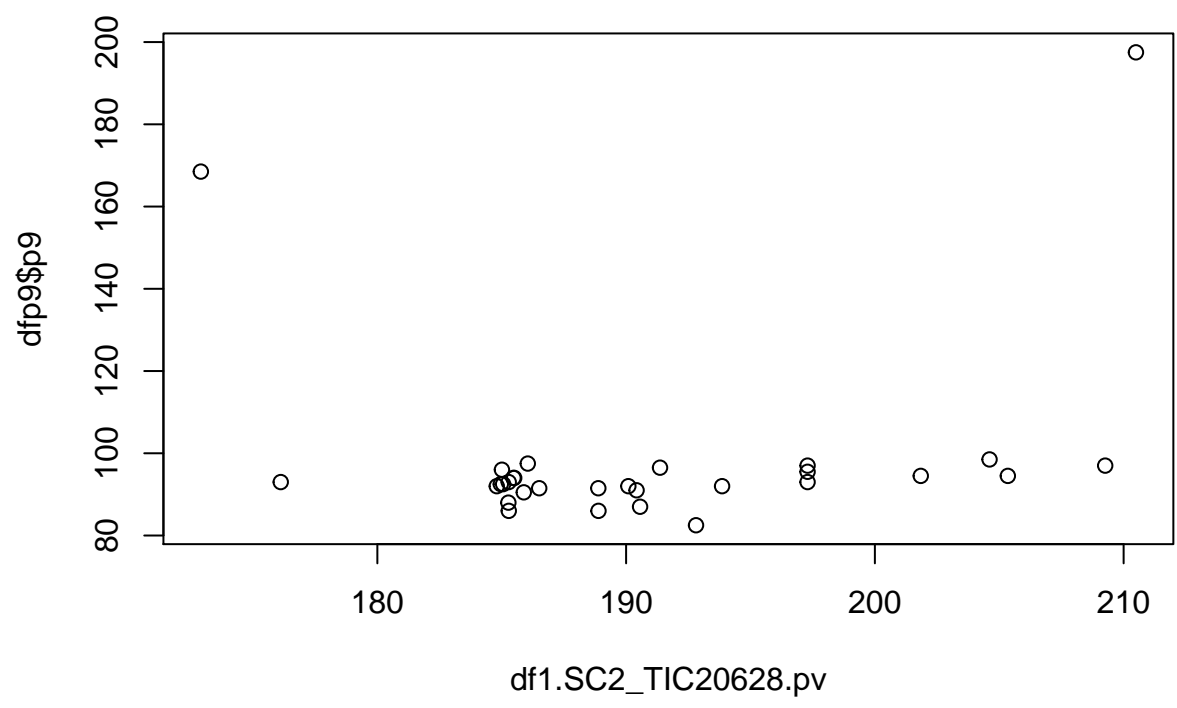


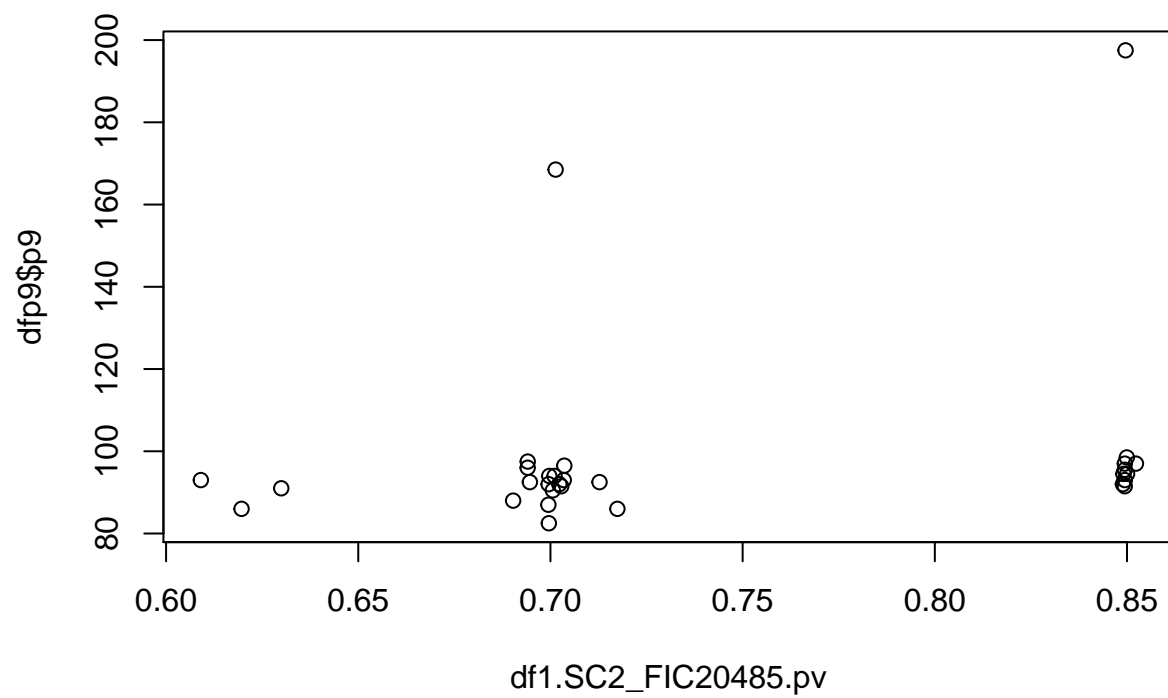


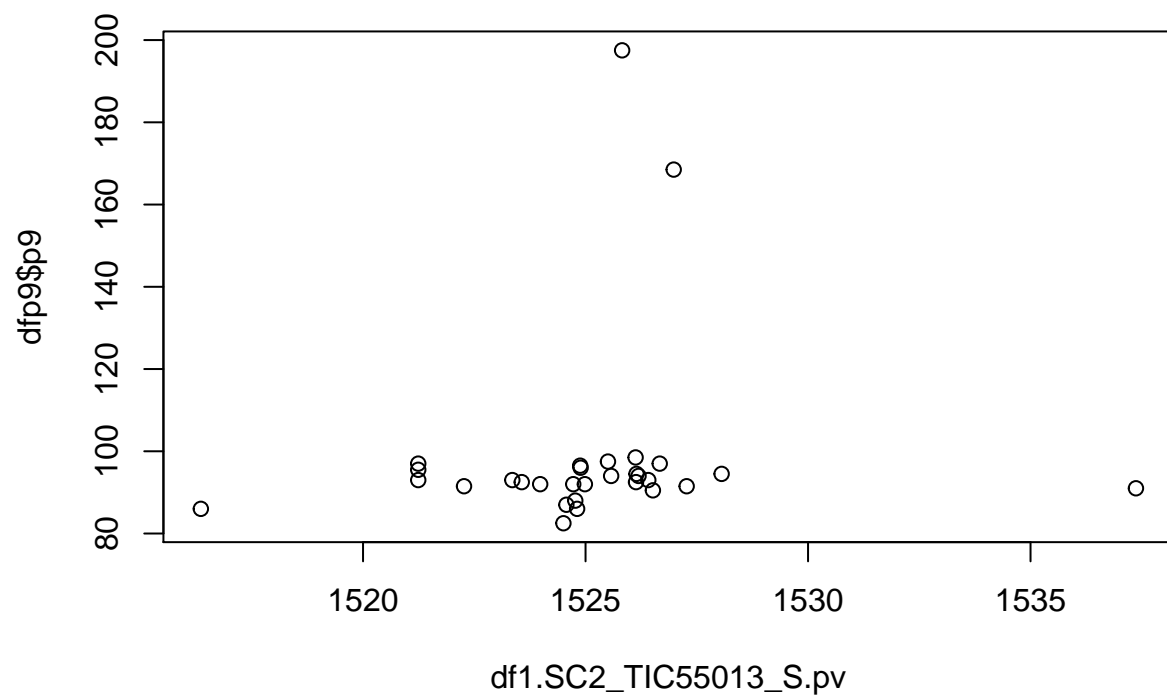


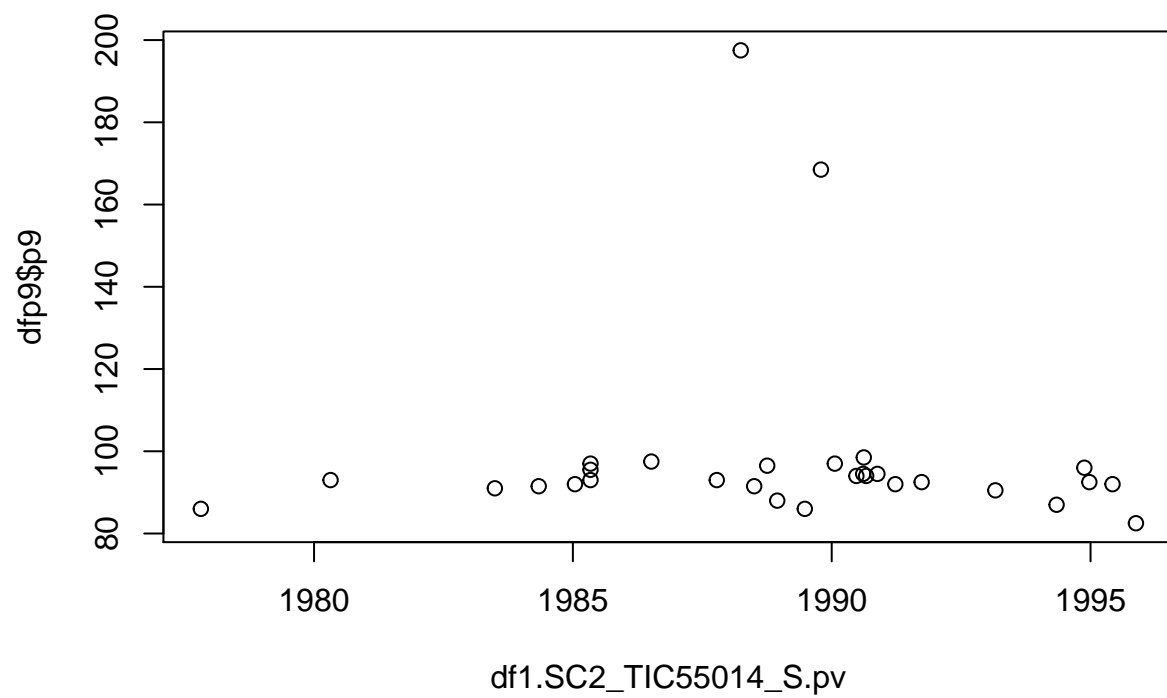


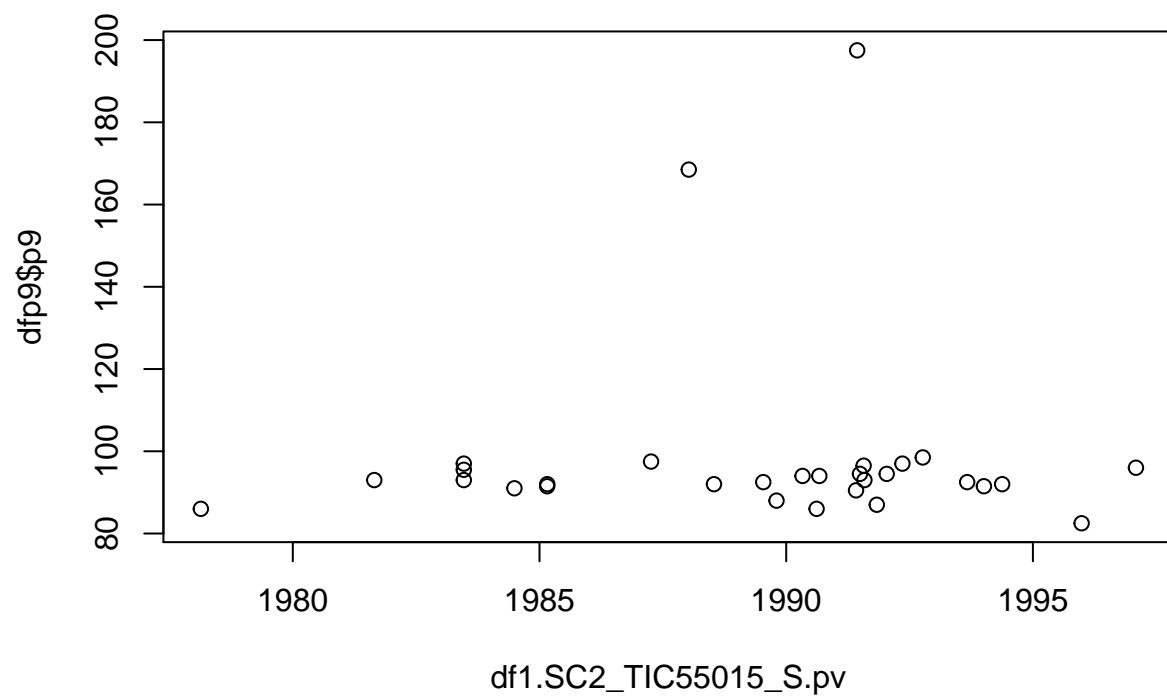


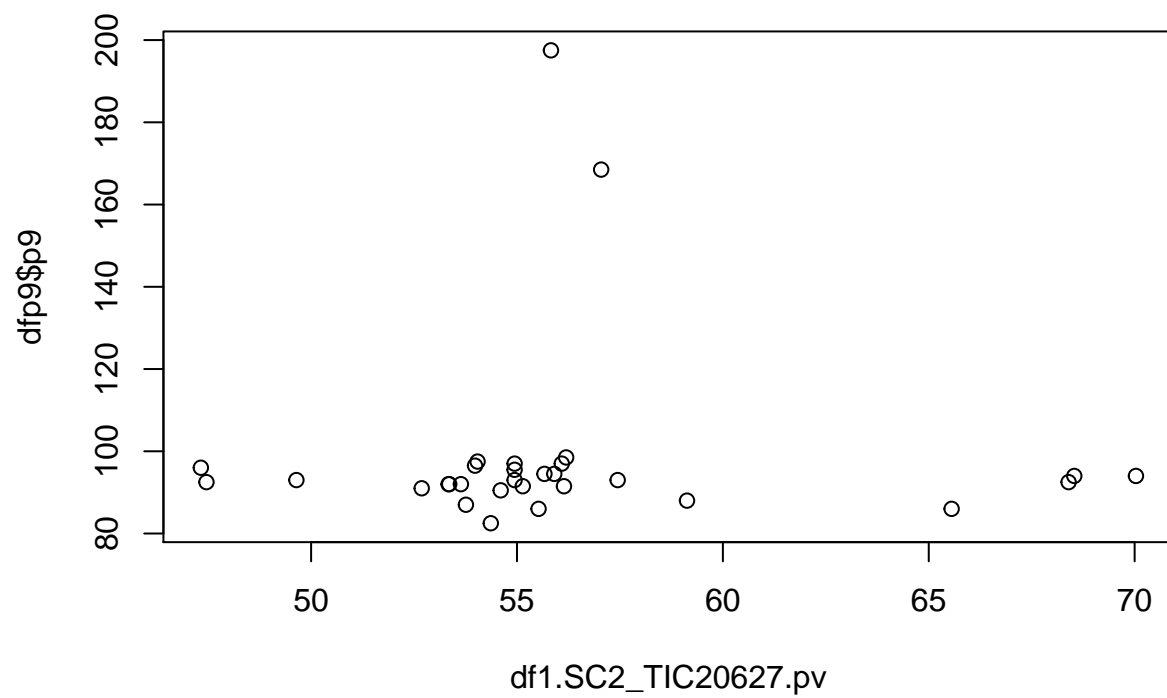


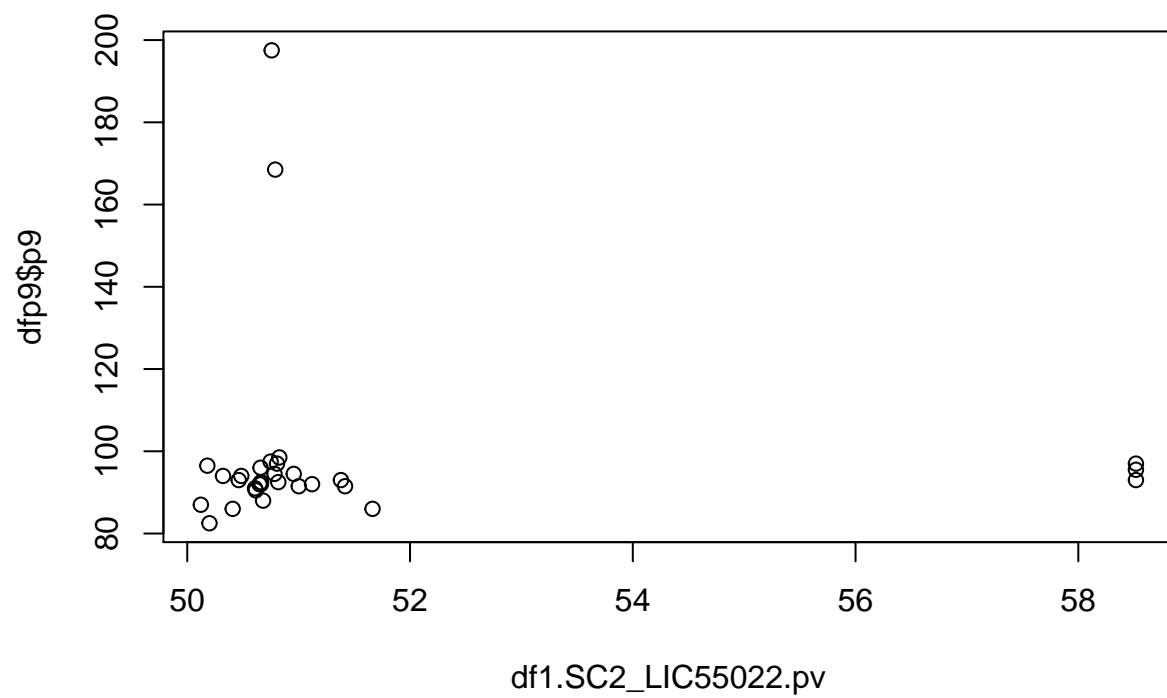


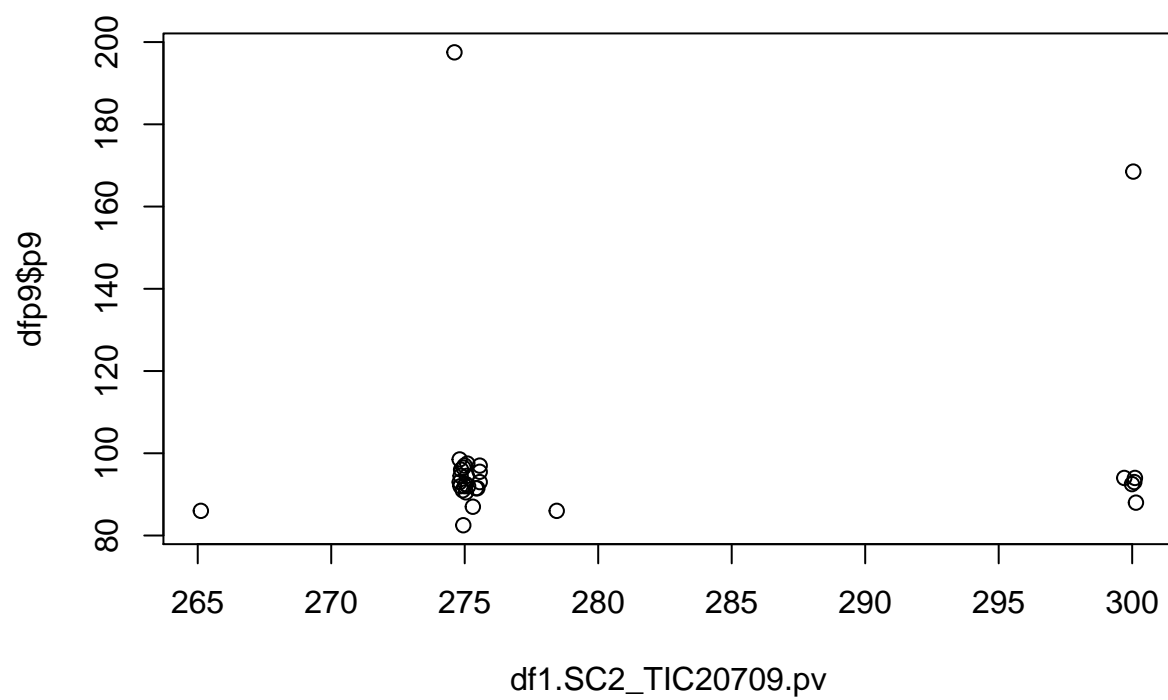


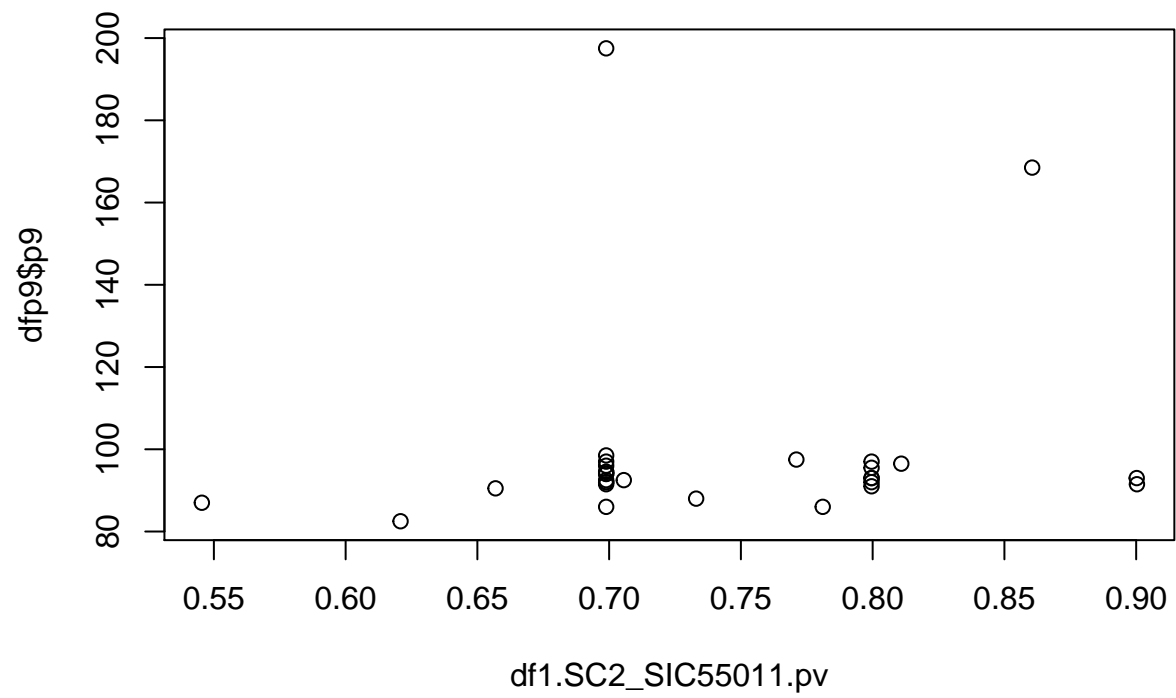




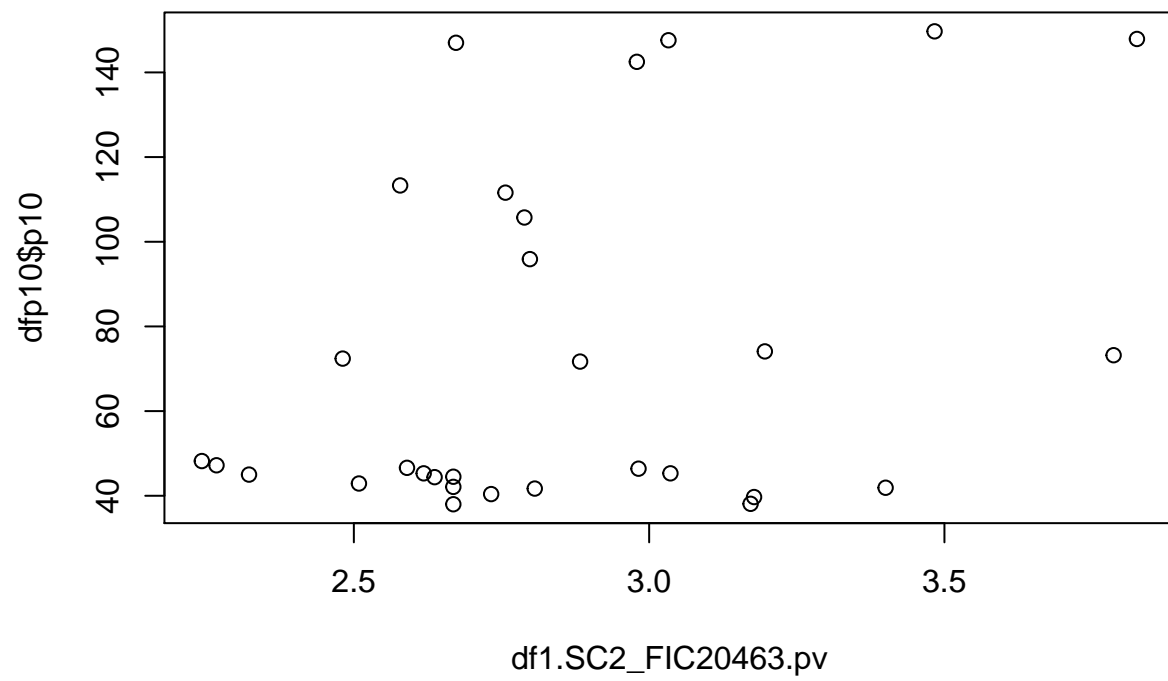


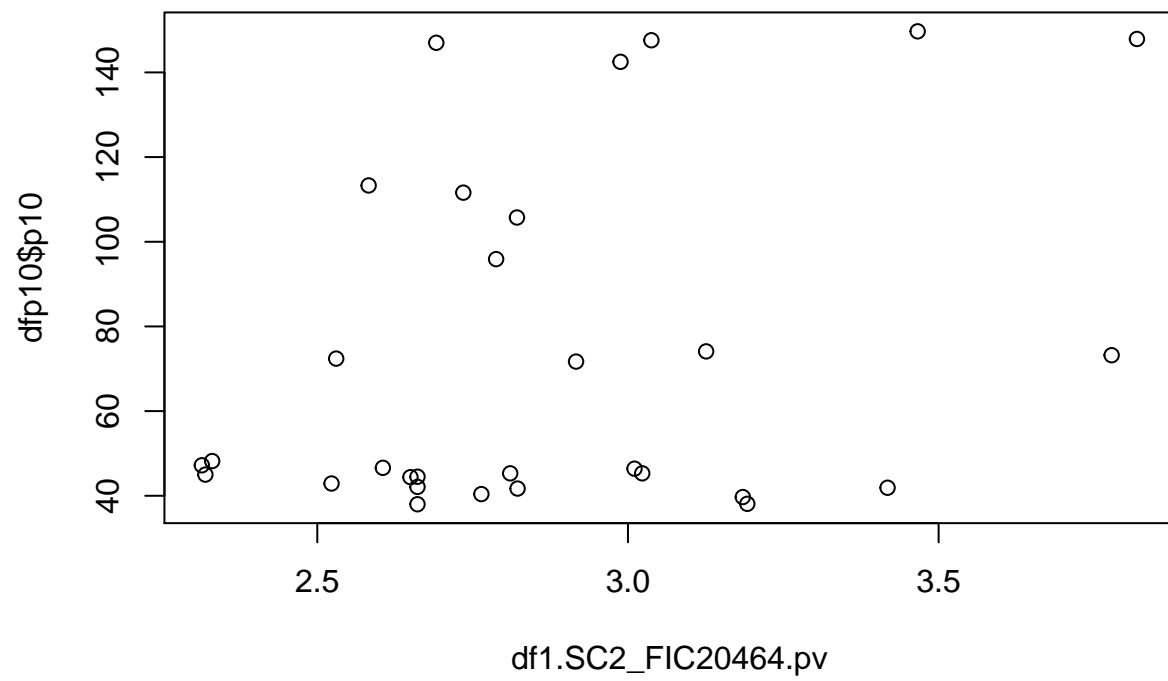


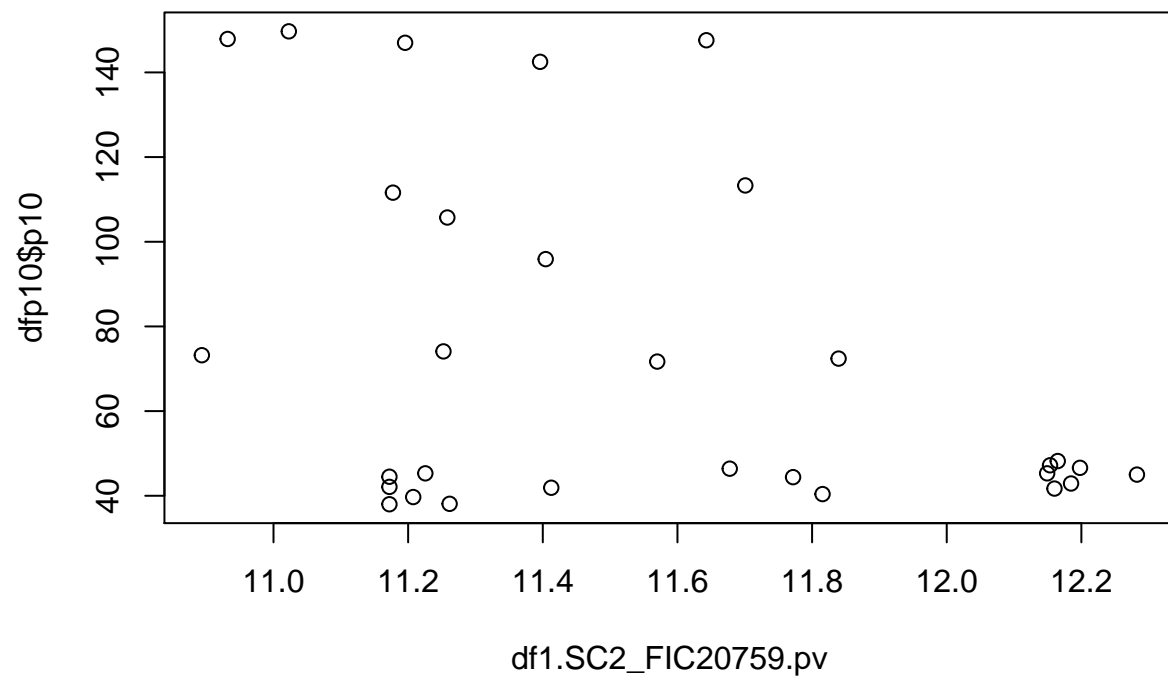


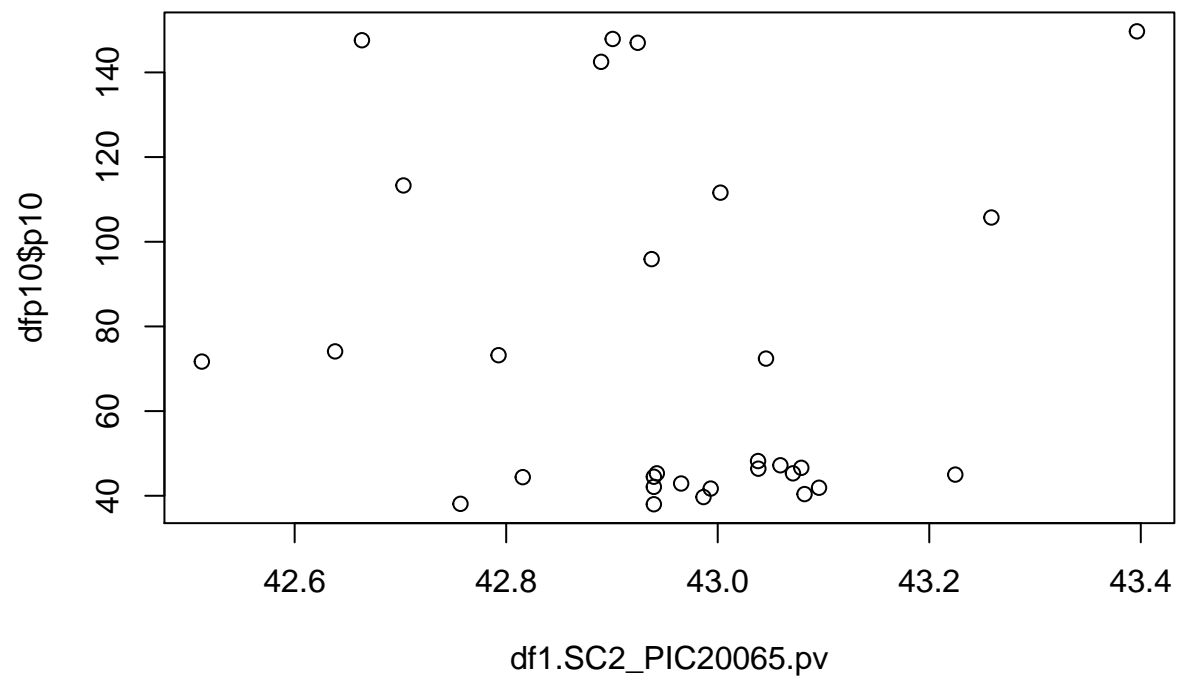


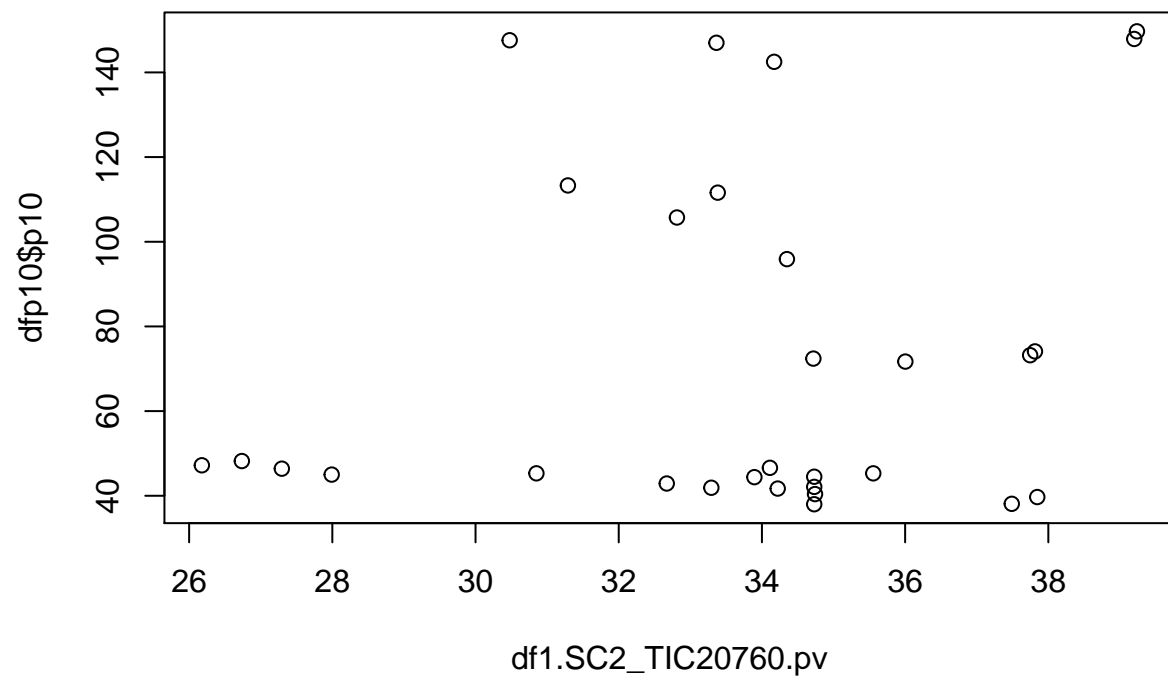
```
dfp10<-data.frame(p10,input)
plot(dfp10$p10~.,dfp10)
```

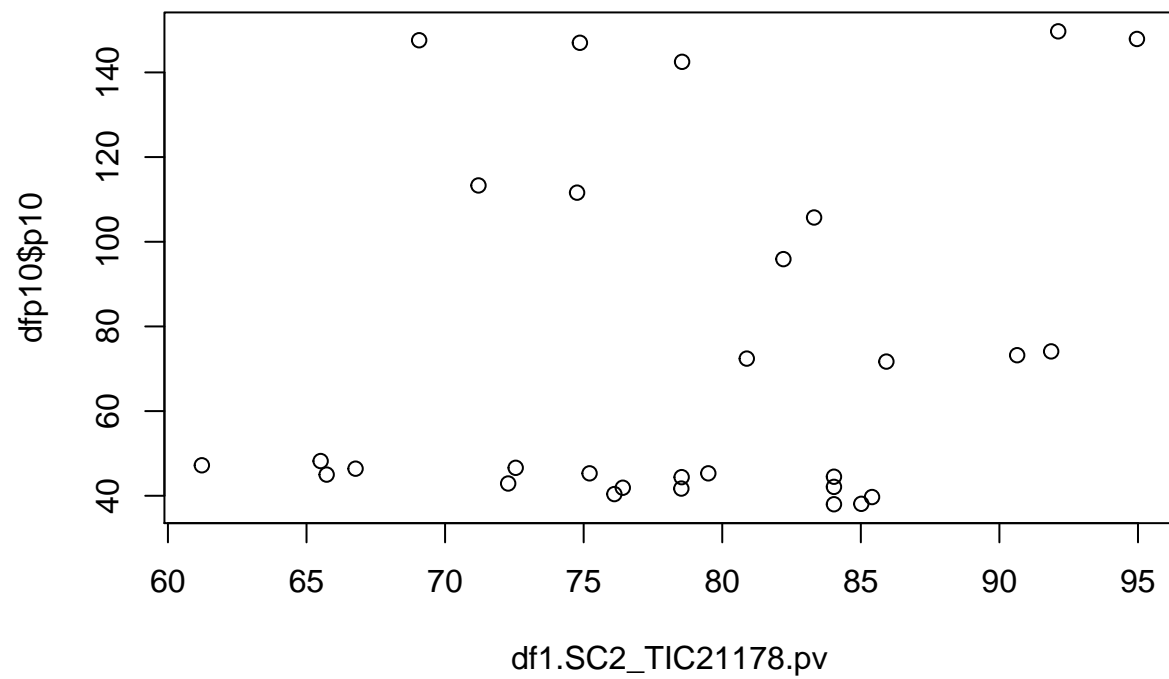


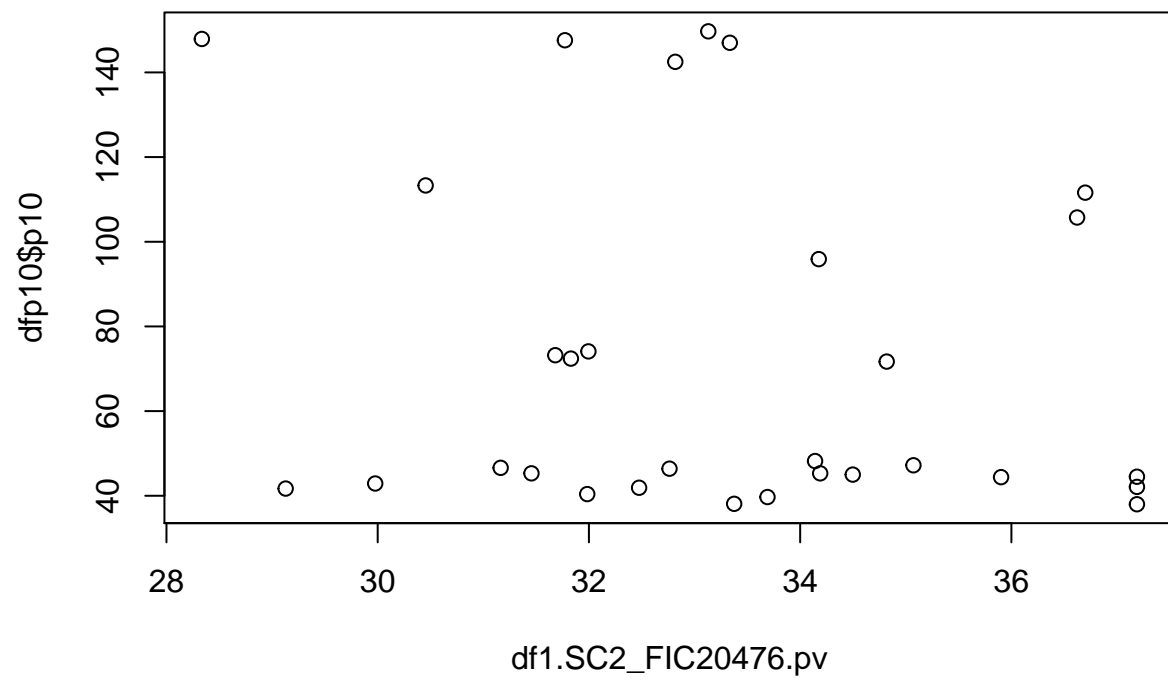


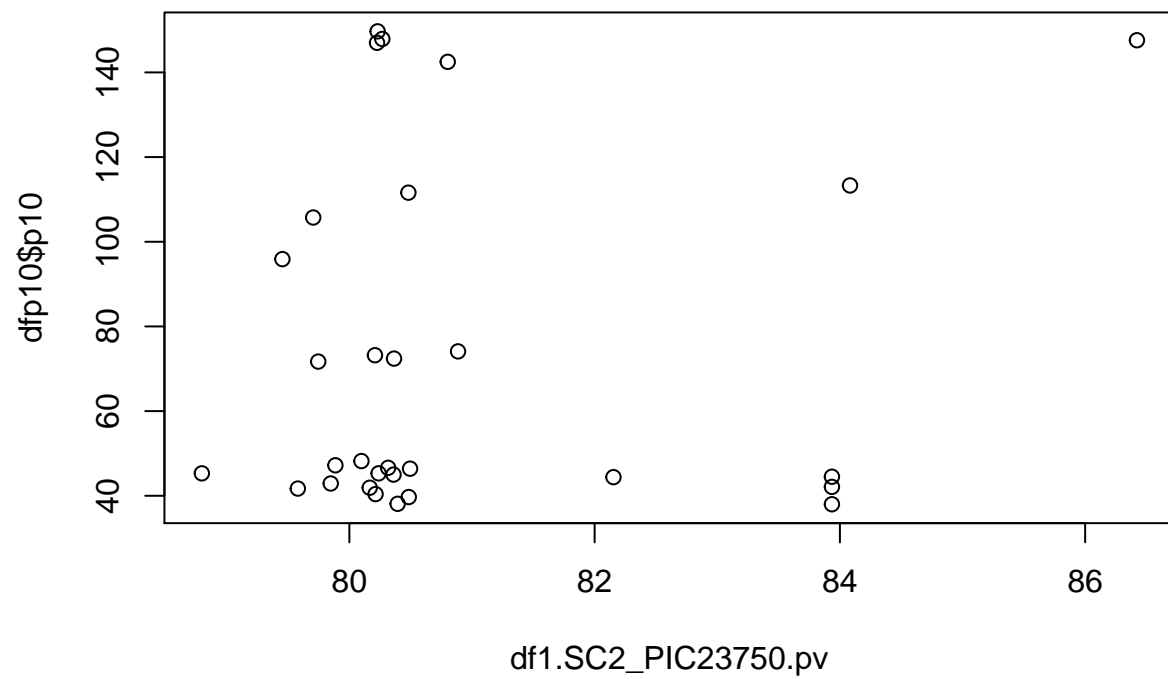


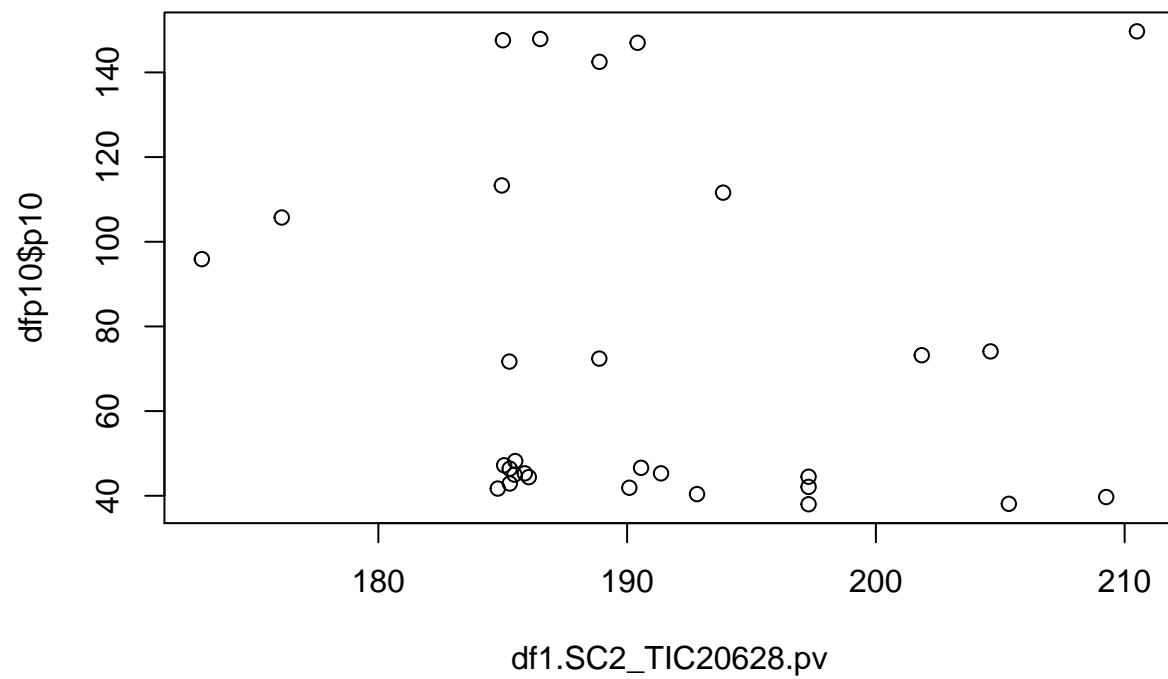


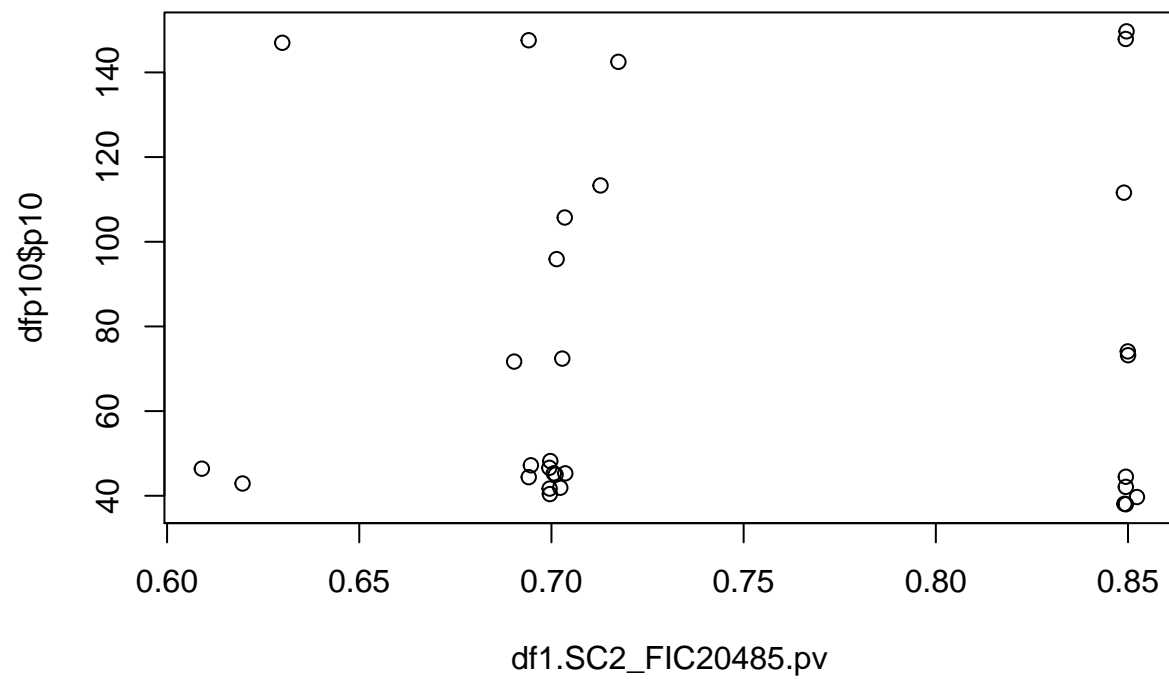


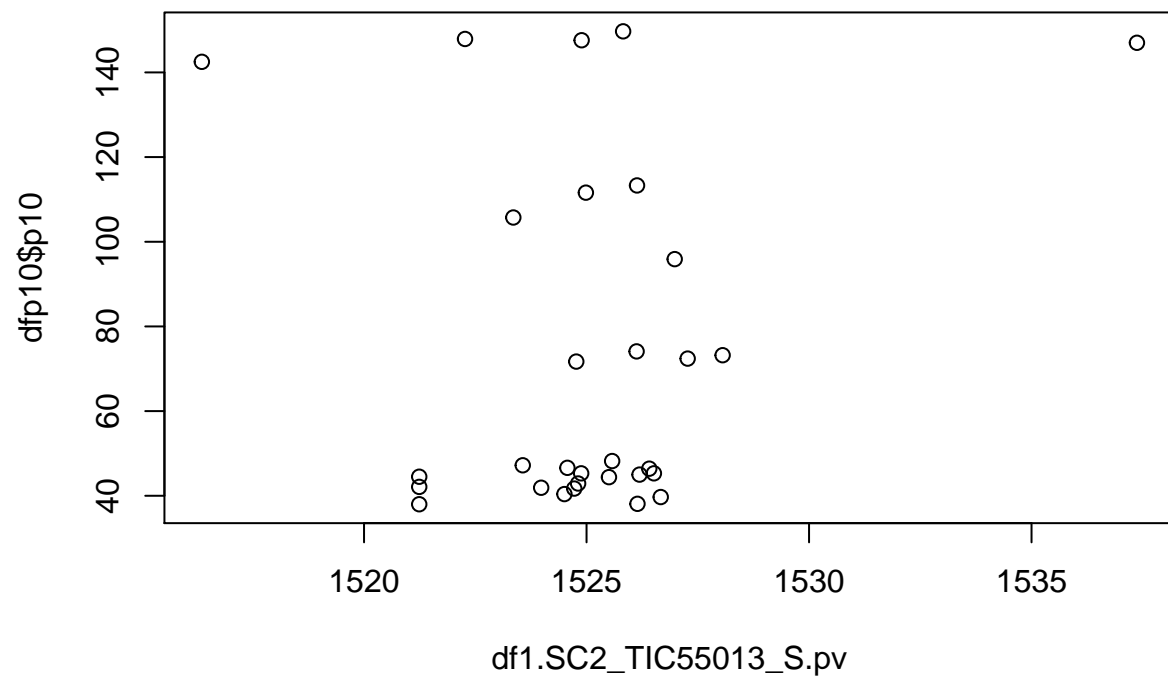


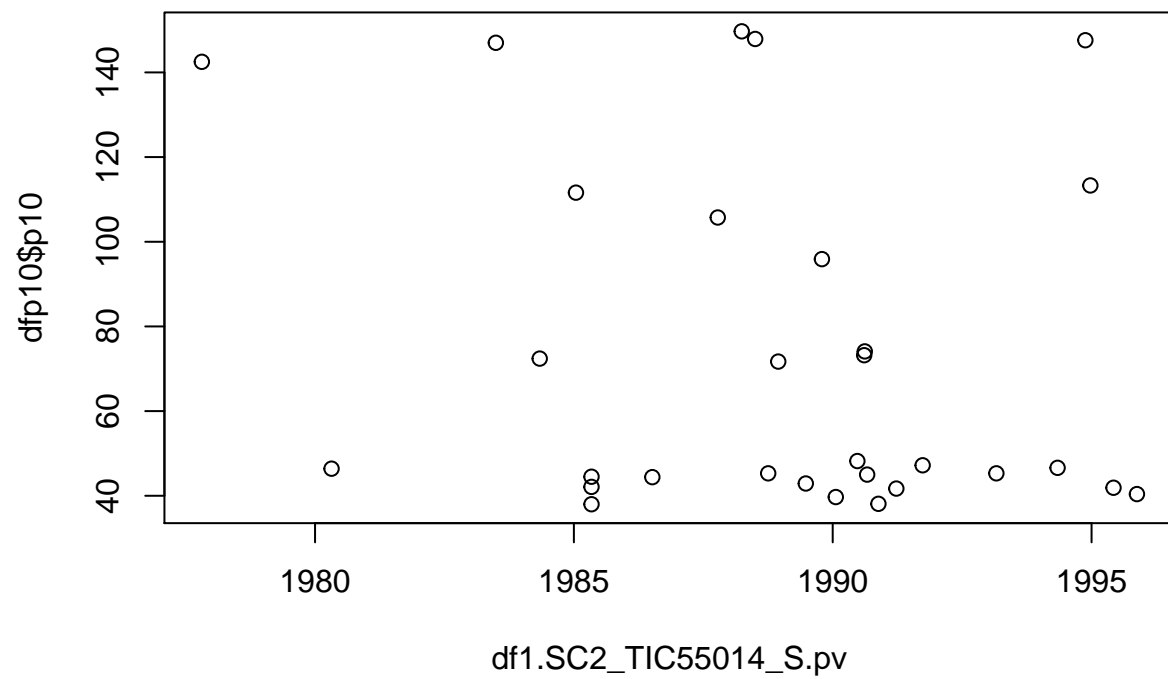


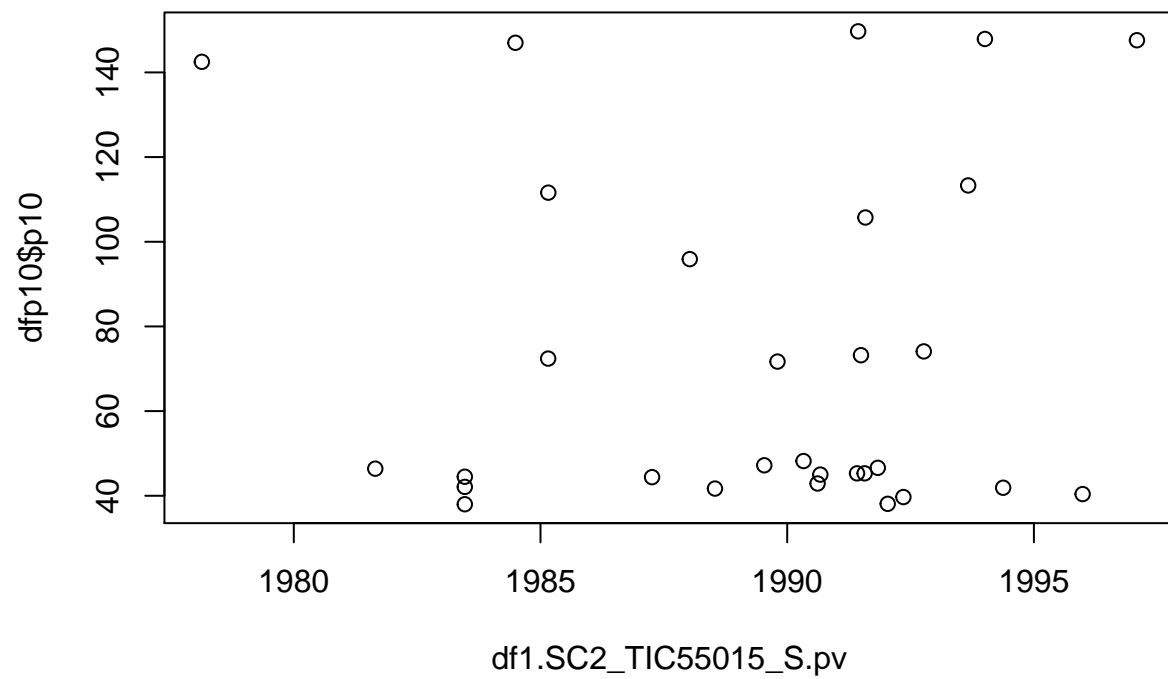


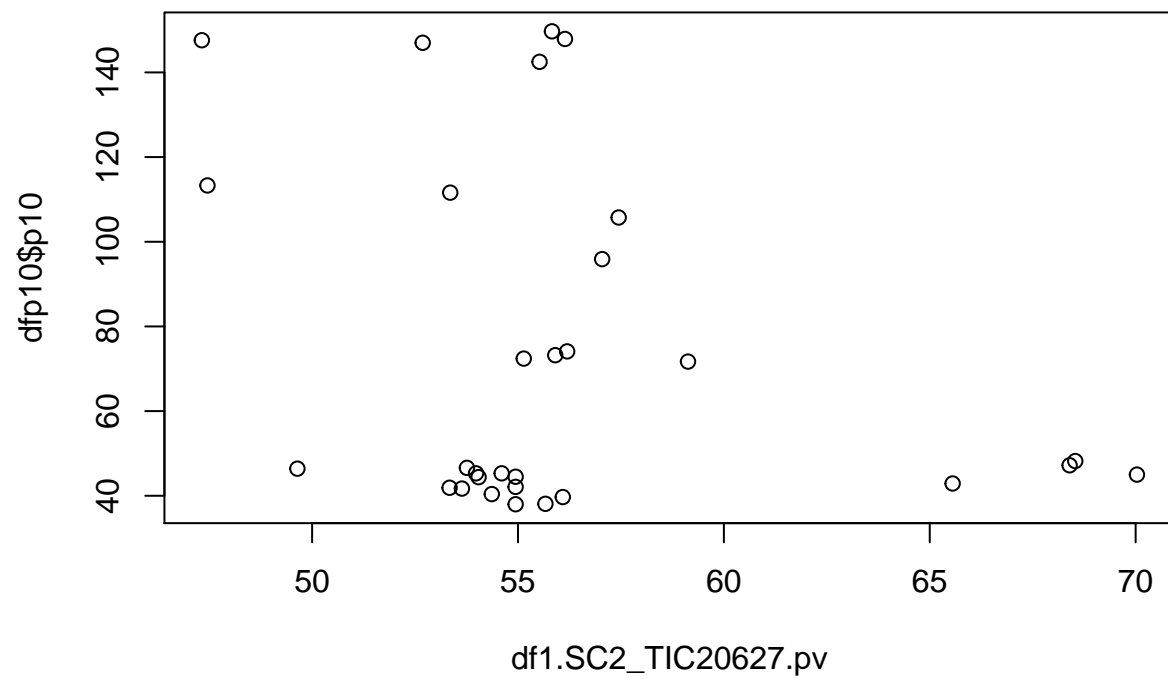


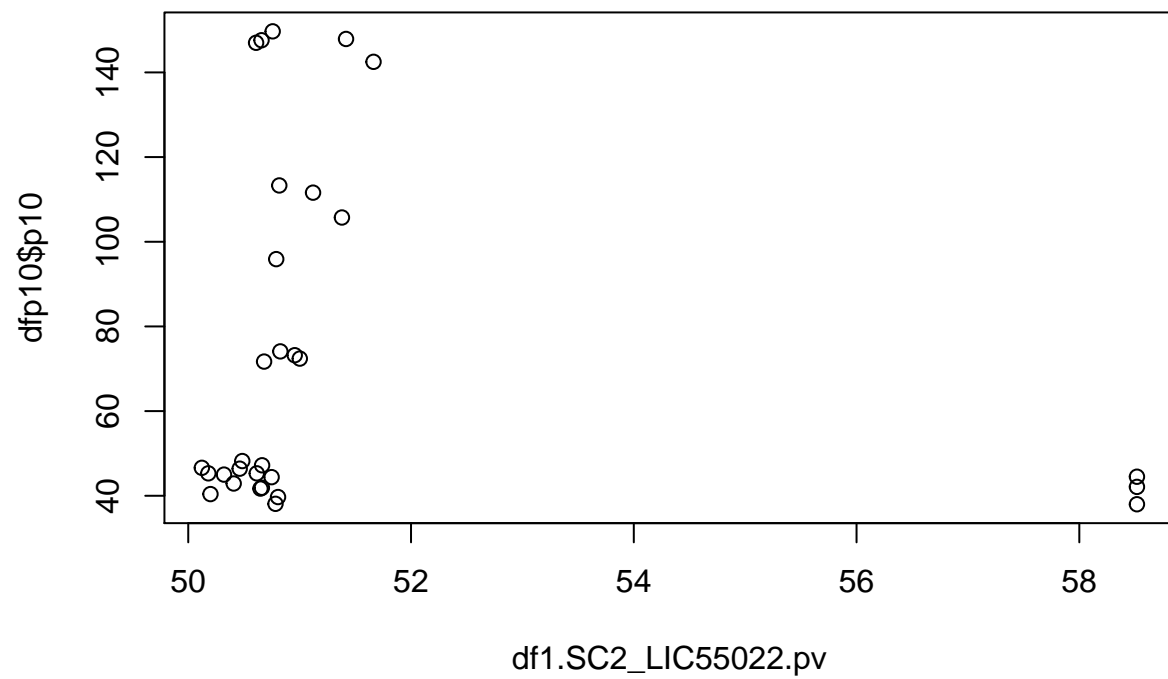


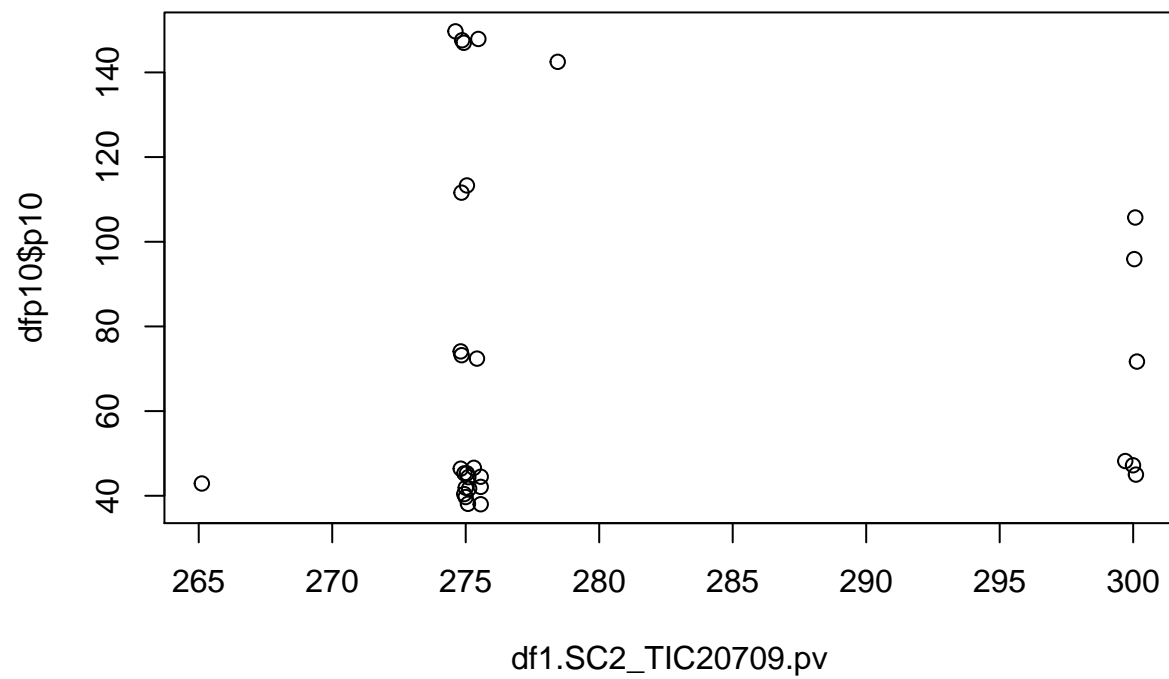


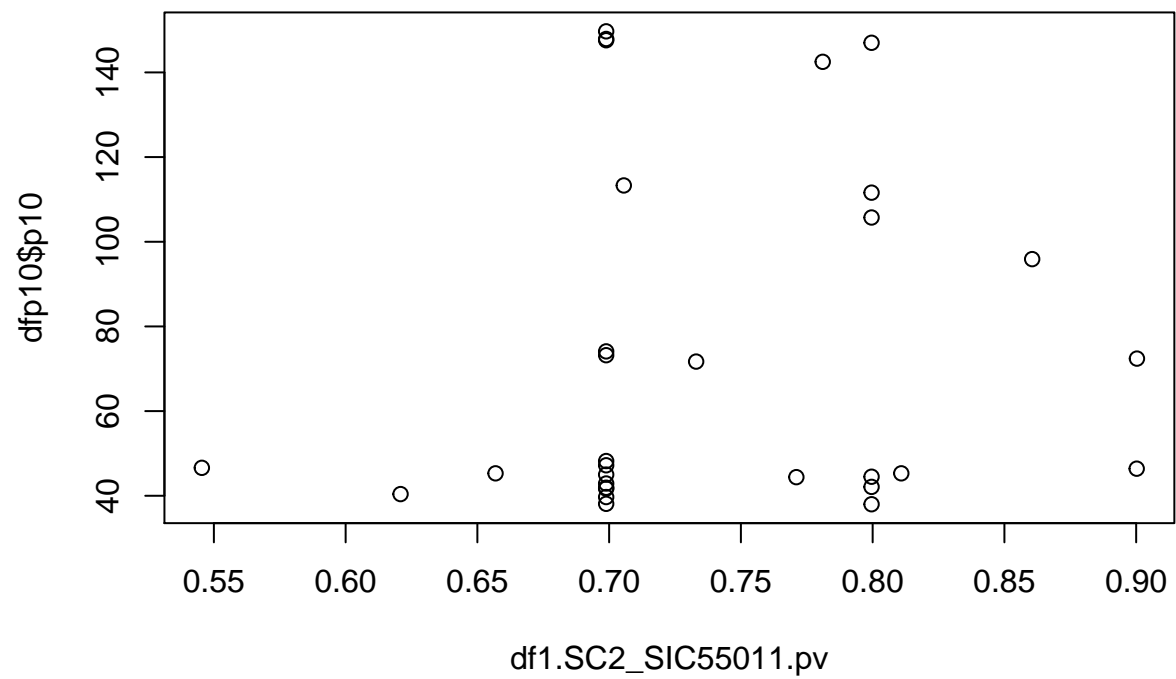




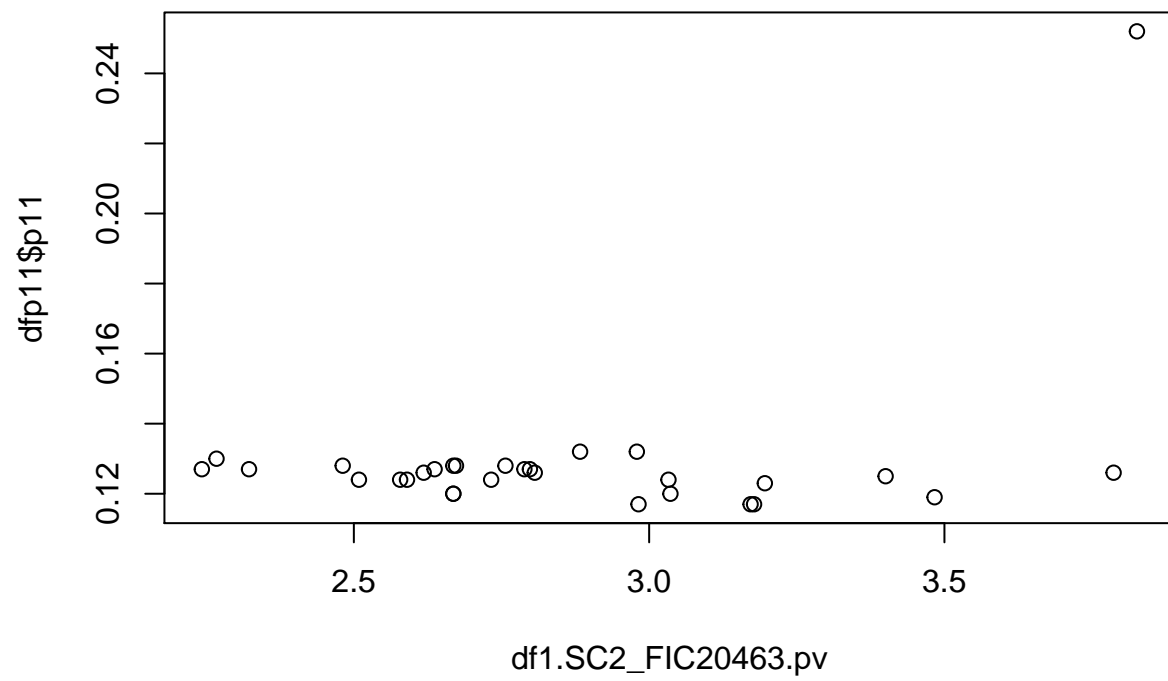


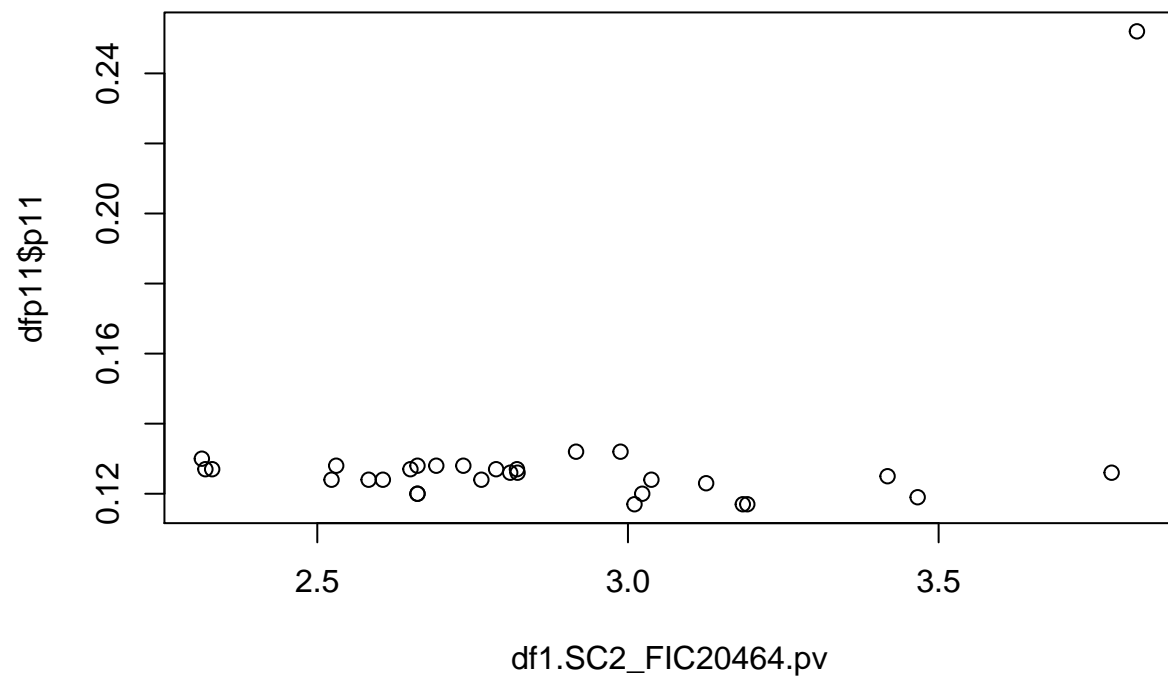




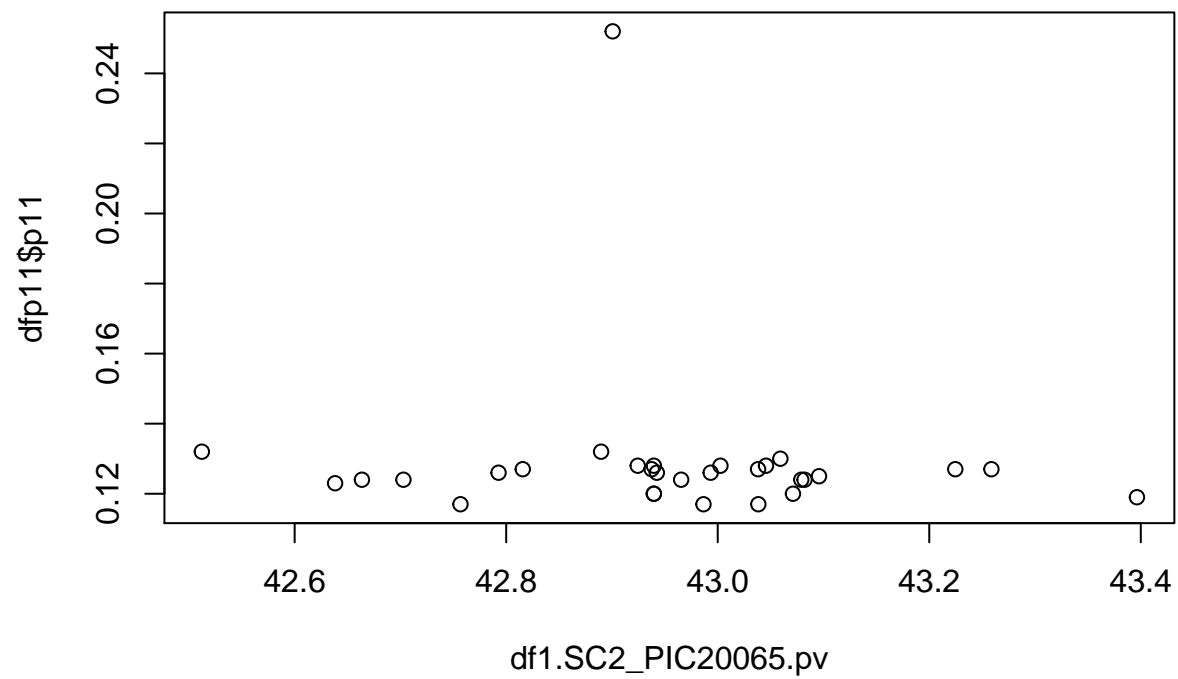


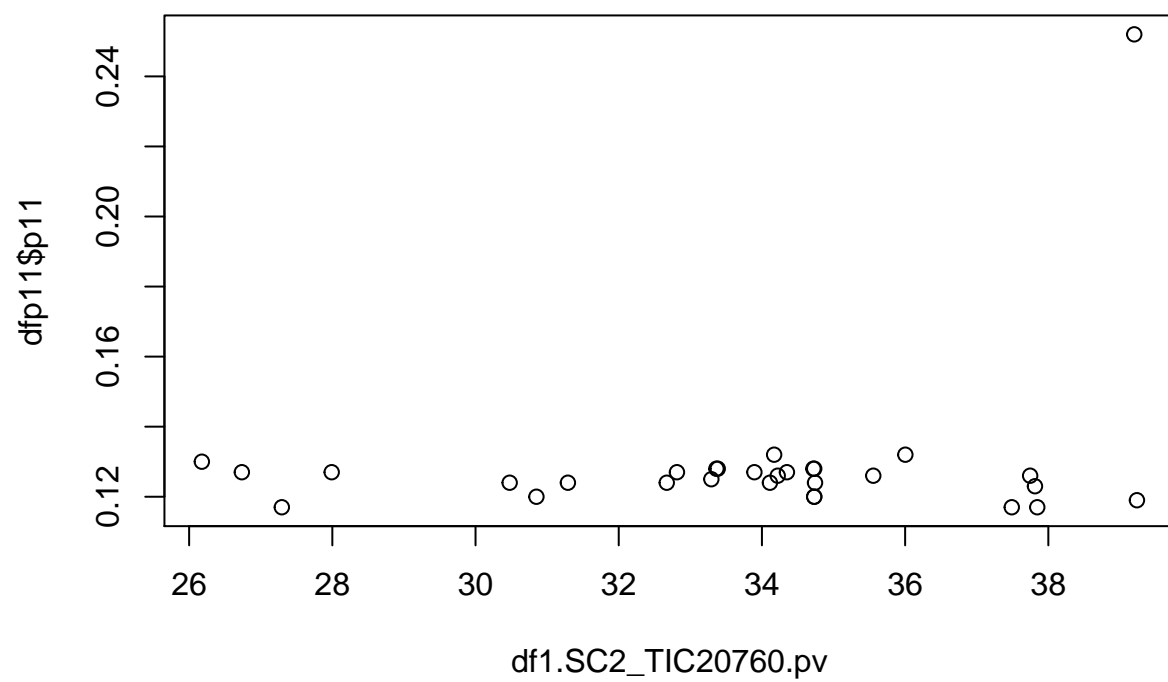
```
dfp11<-data.frame(p11,input)
plot(dfp11$p11~.,dfp11)
```

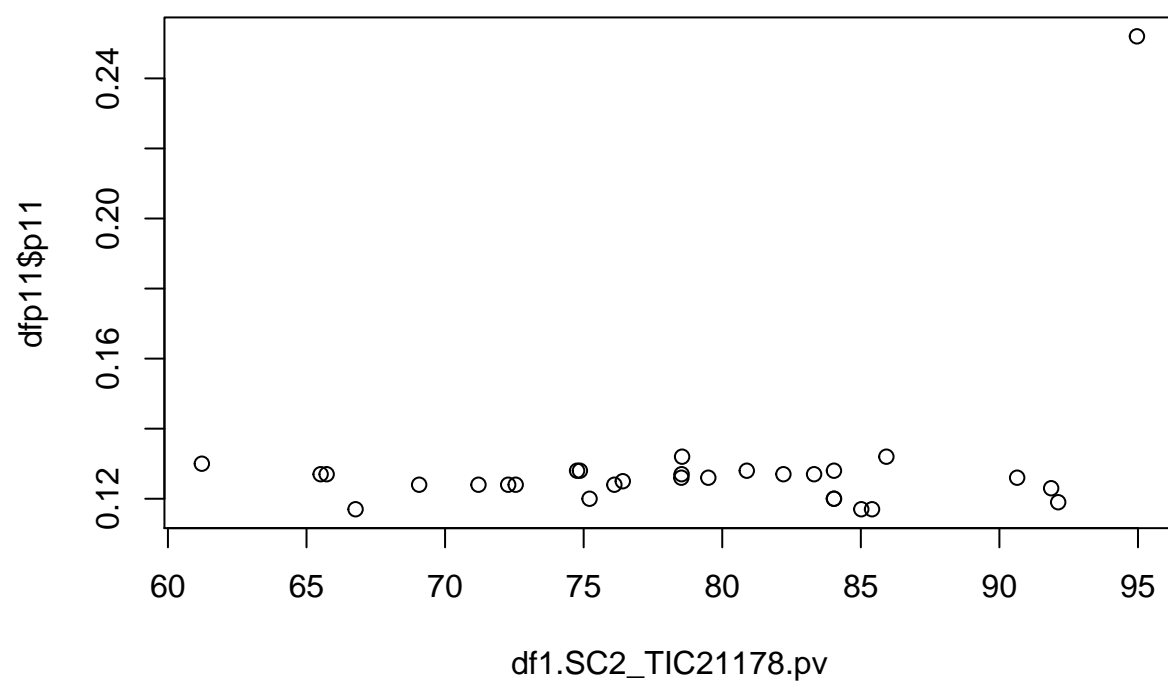


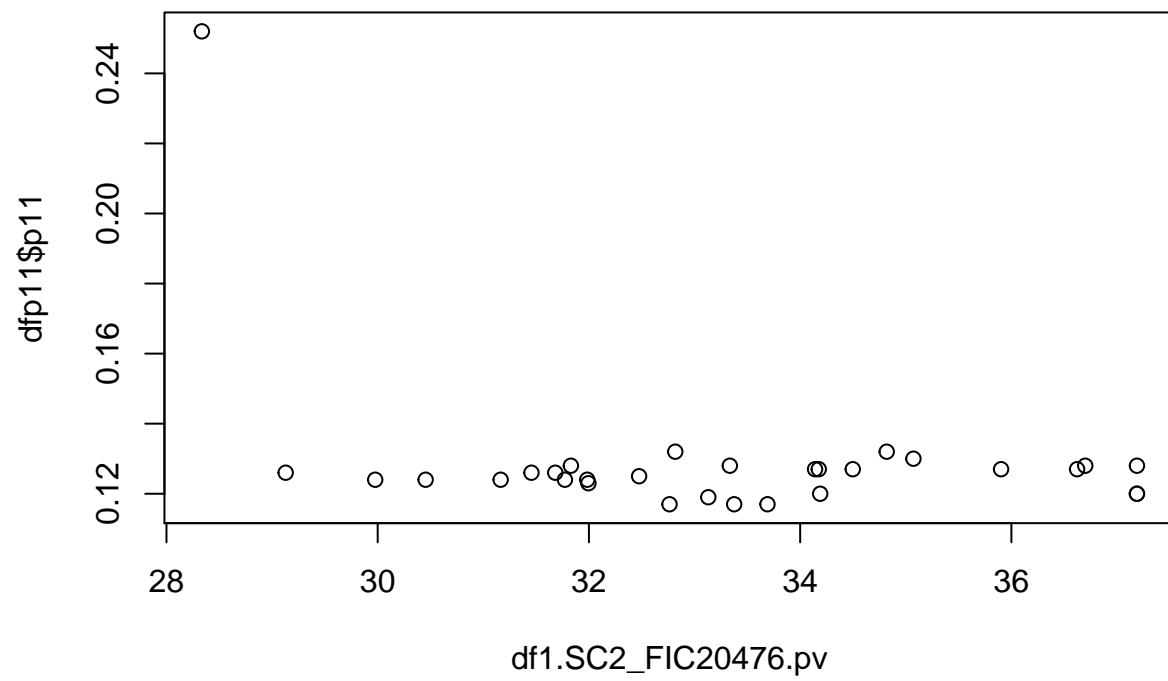


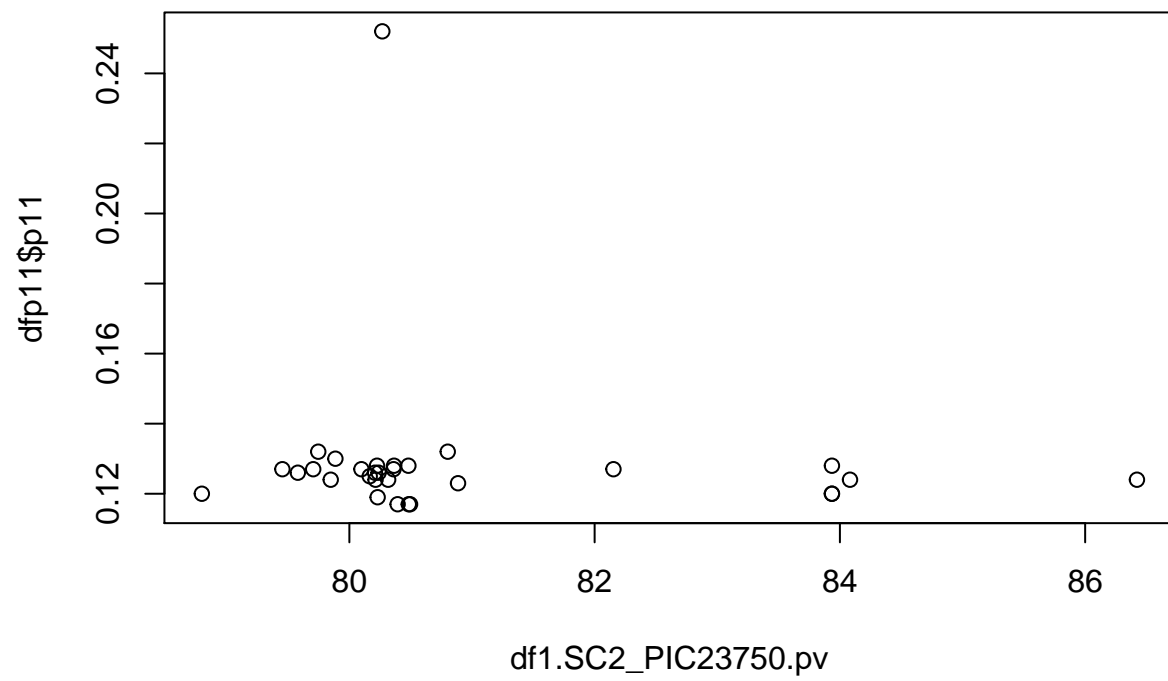


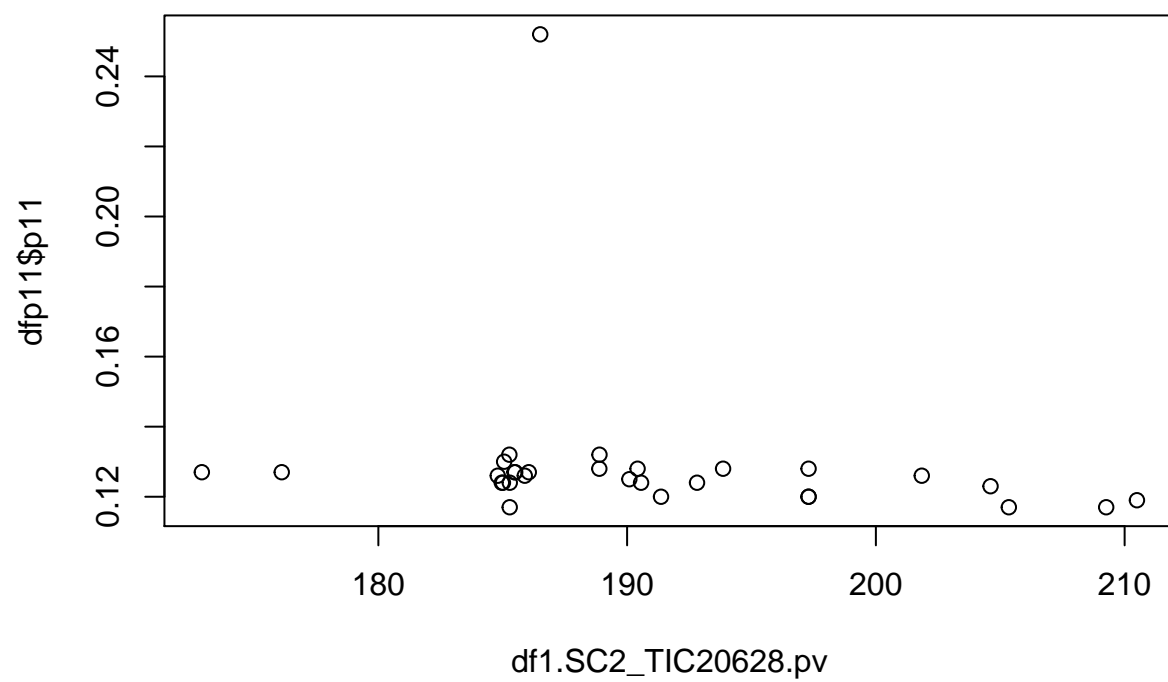


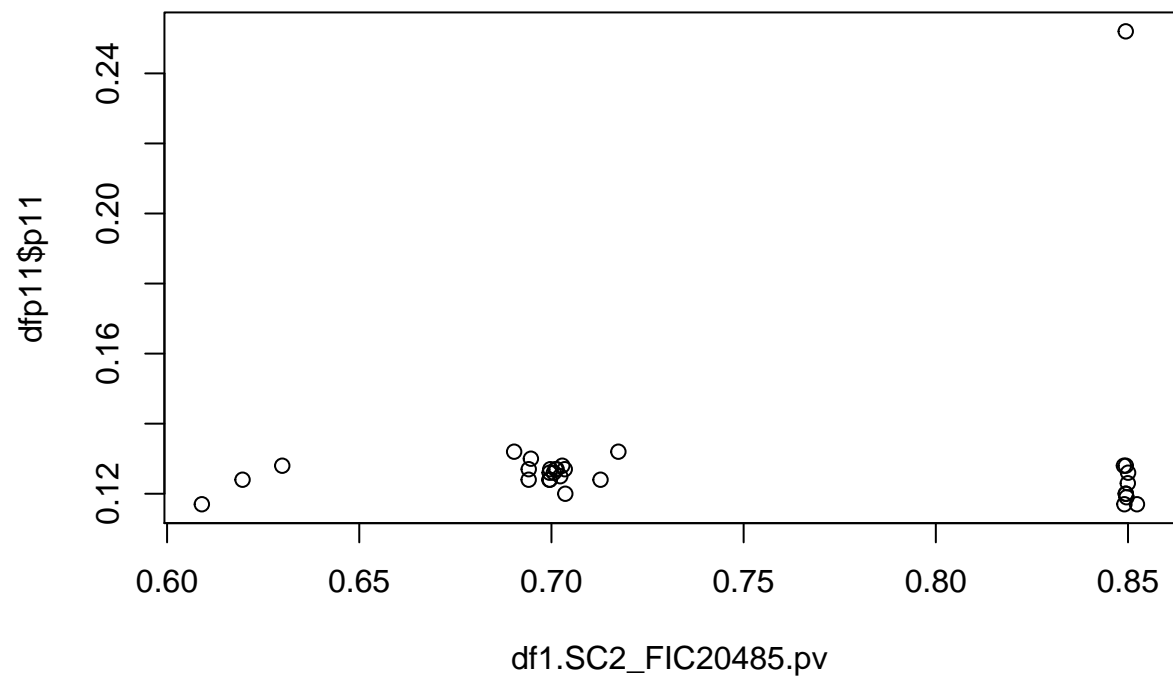


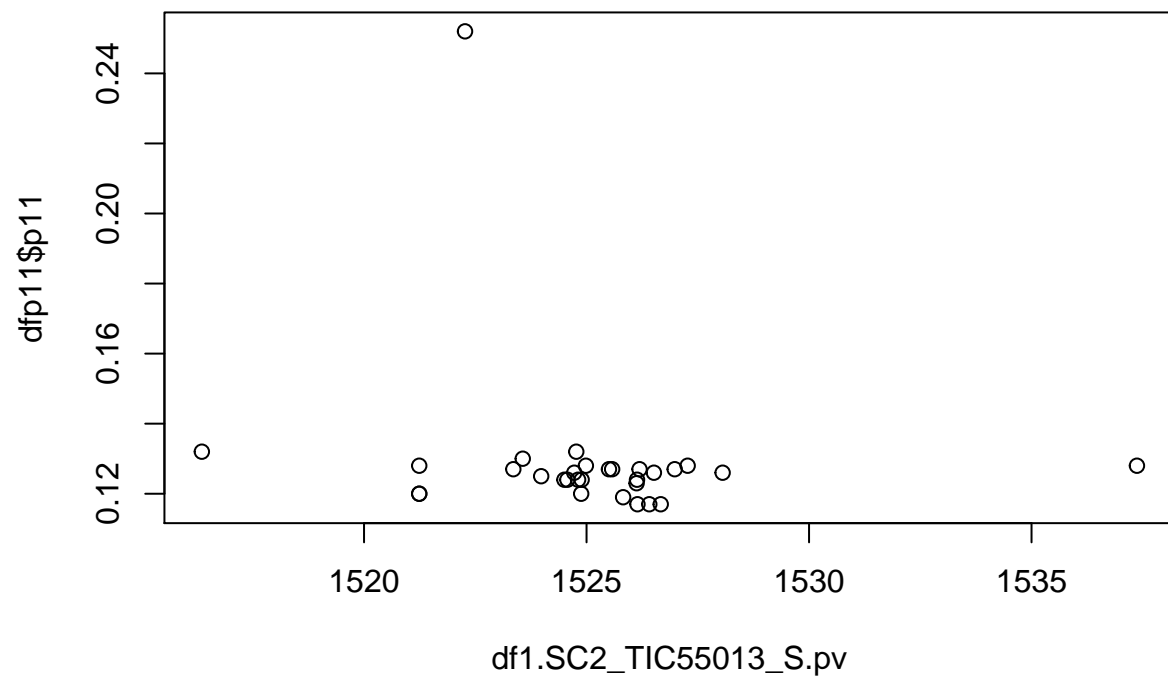


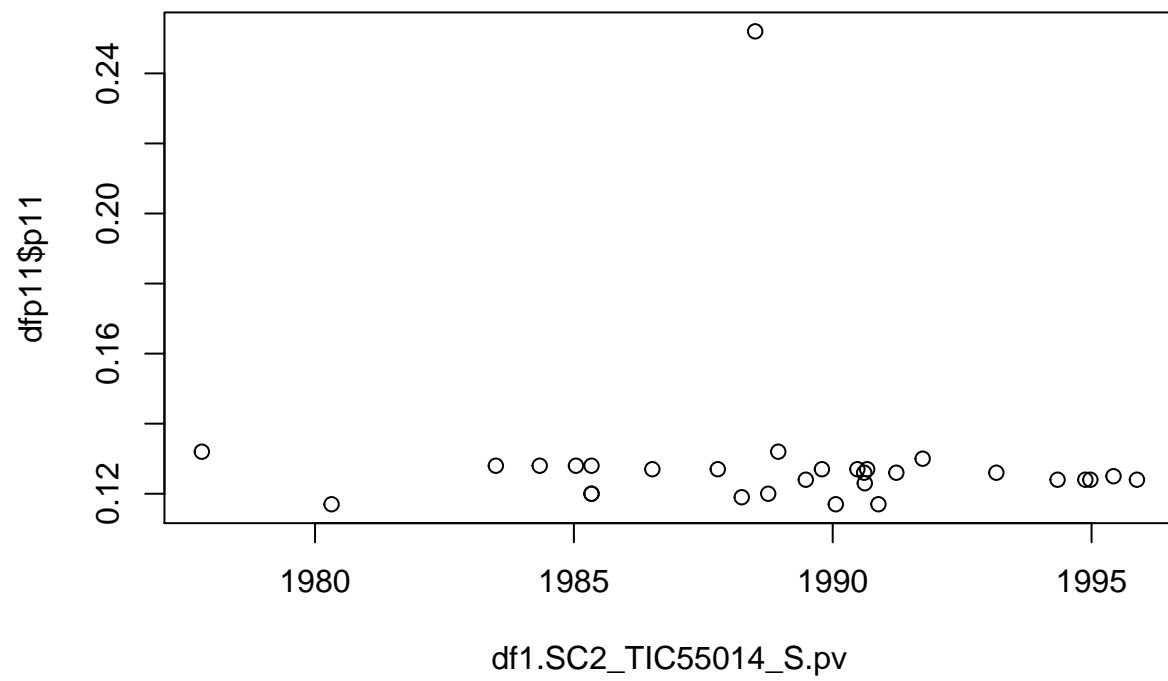


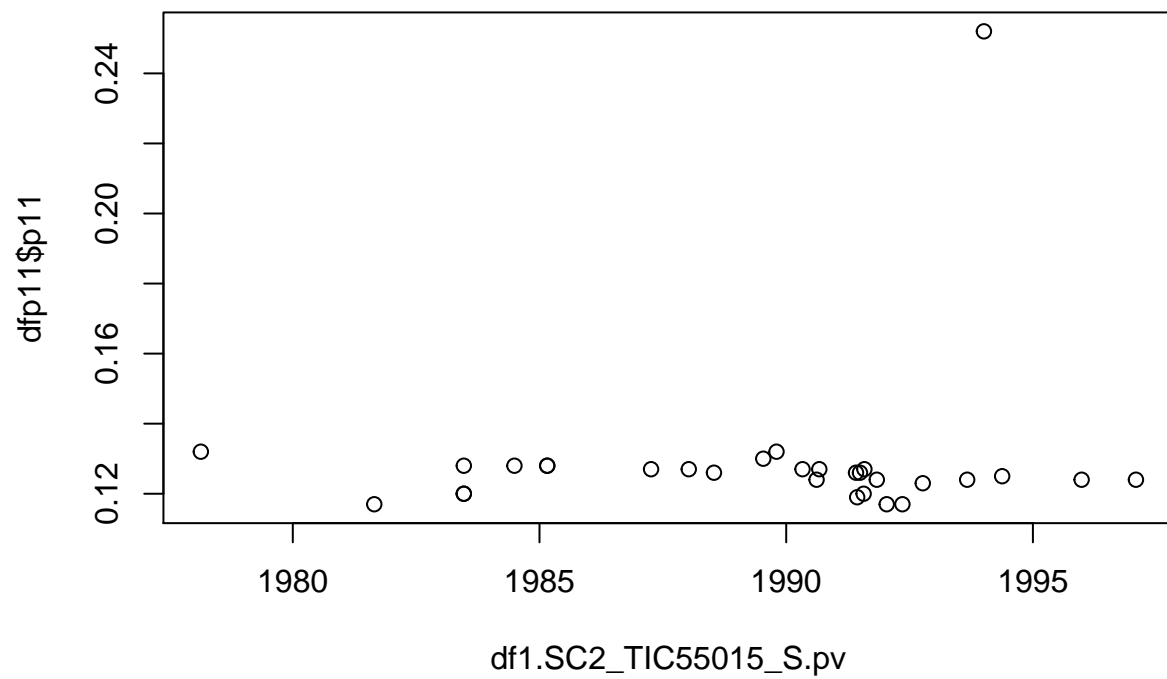


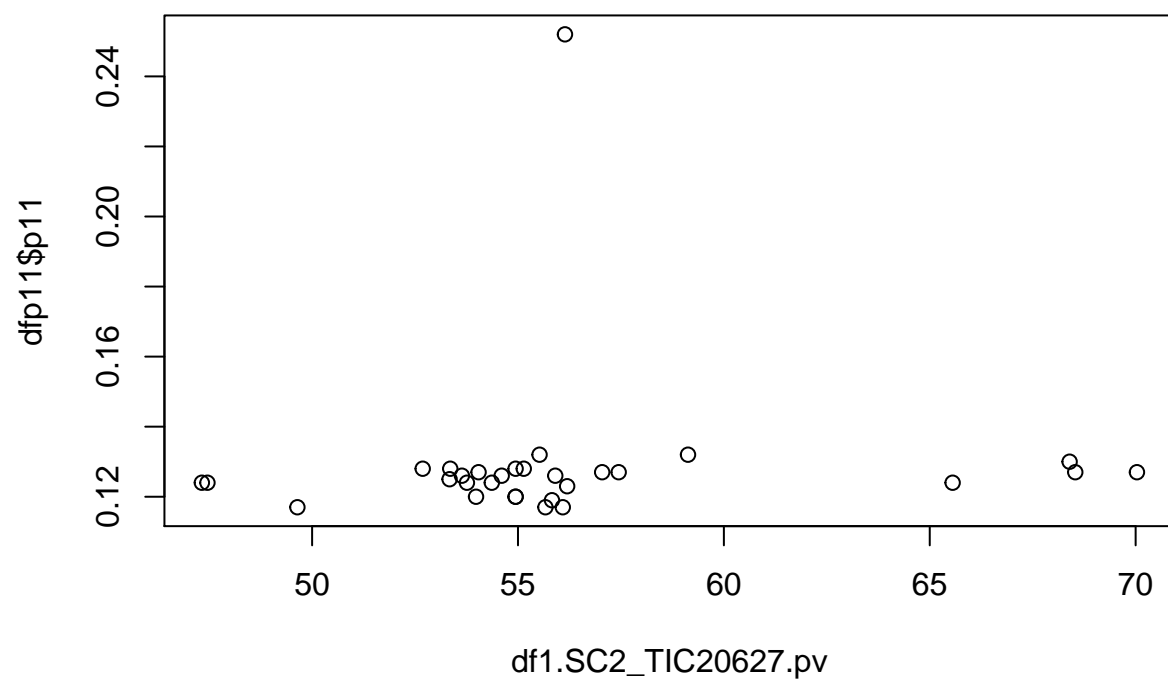


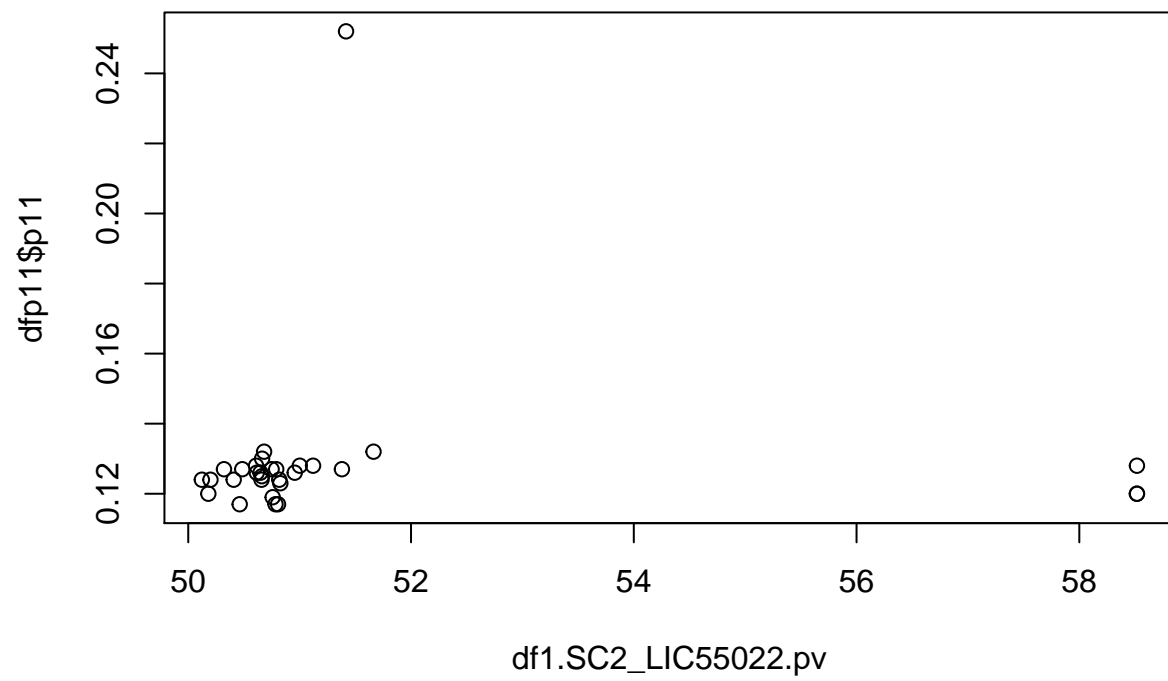




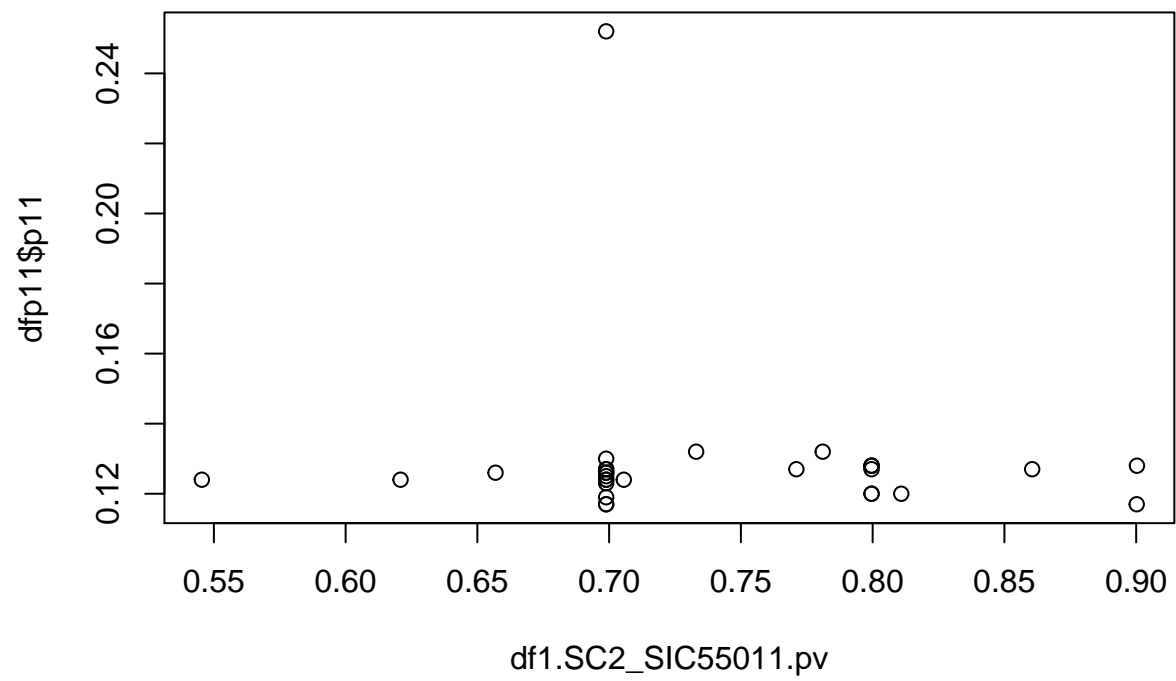












```
dfp13<-data.frame(p13,input)
plot(dfp13$p13~.,dfp13)
```

