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 infered.

Problems: If it is possible to know which tag and parameter shold 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

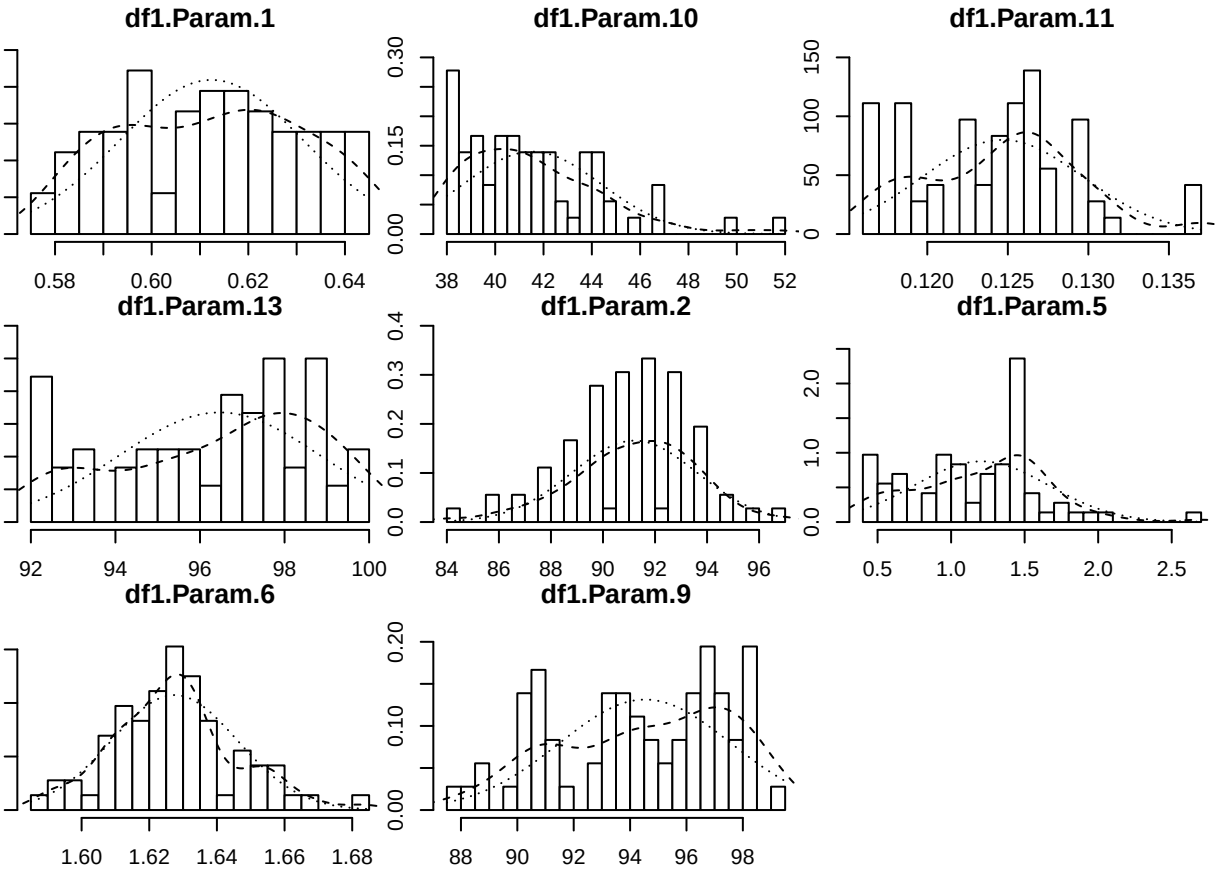
*Transformation by log and square

*Plot on and off spec batches together

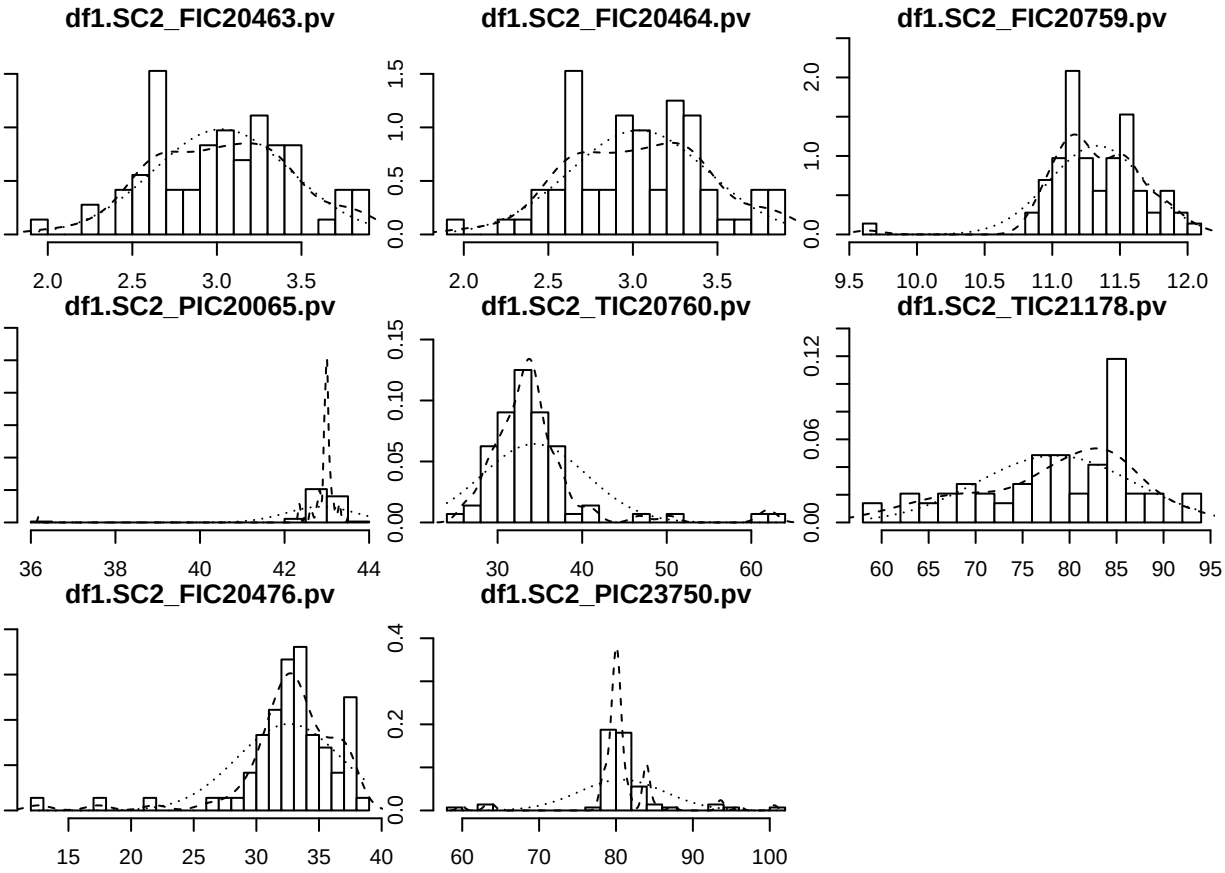
*Run t-test between onspec and offspec batches

*Find distance between important tags between onspec and offspec batches

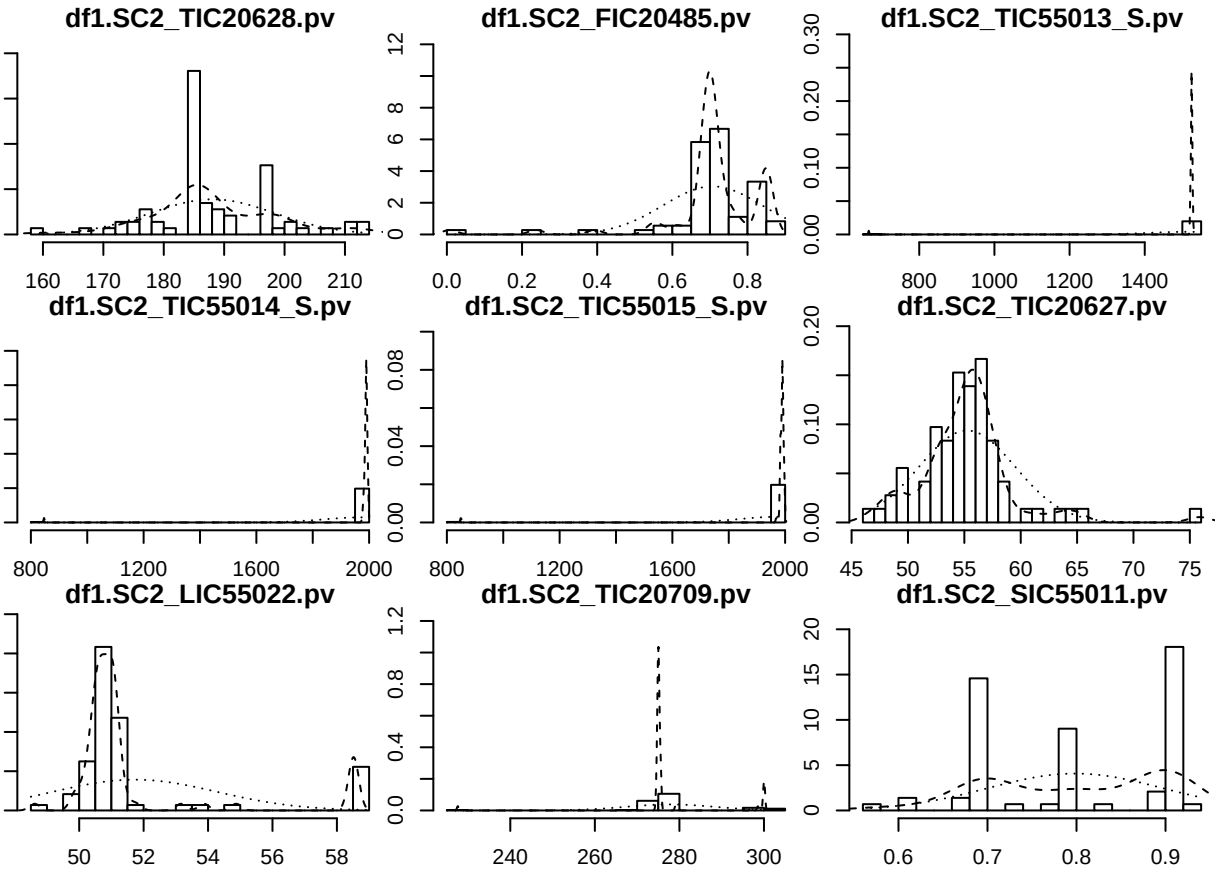
The below histograms and scatterplots are for ONSPEC batches only



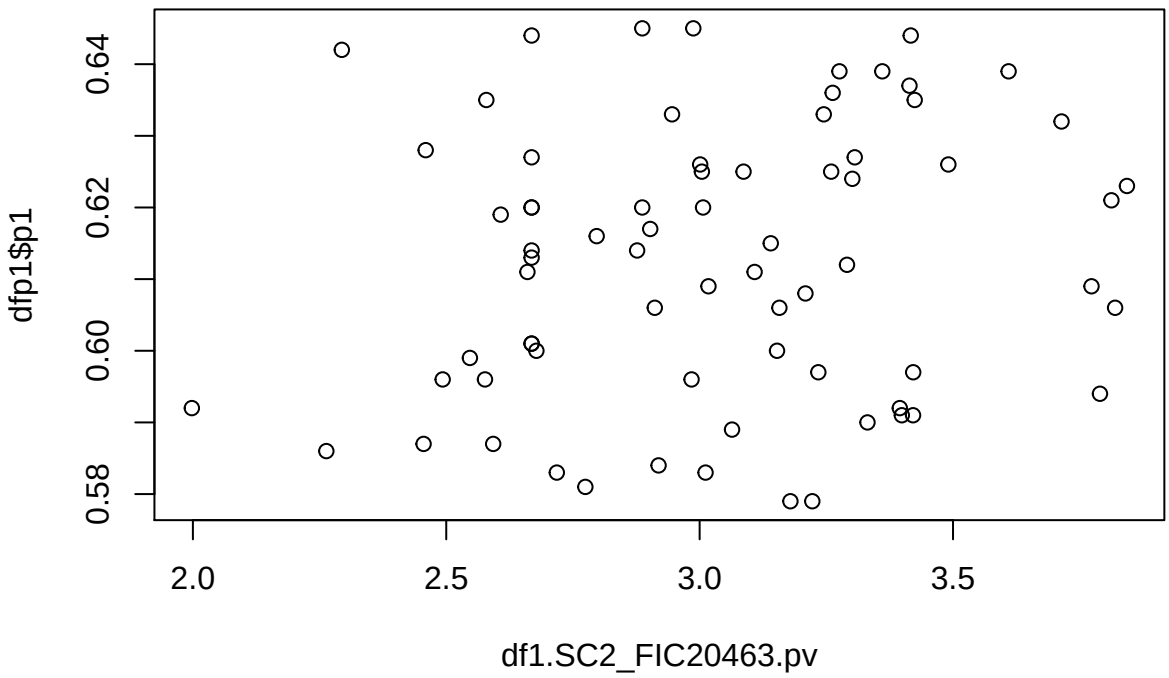

```
psych::multi.hist(input[,1:8])
```

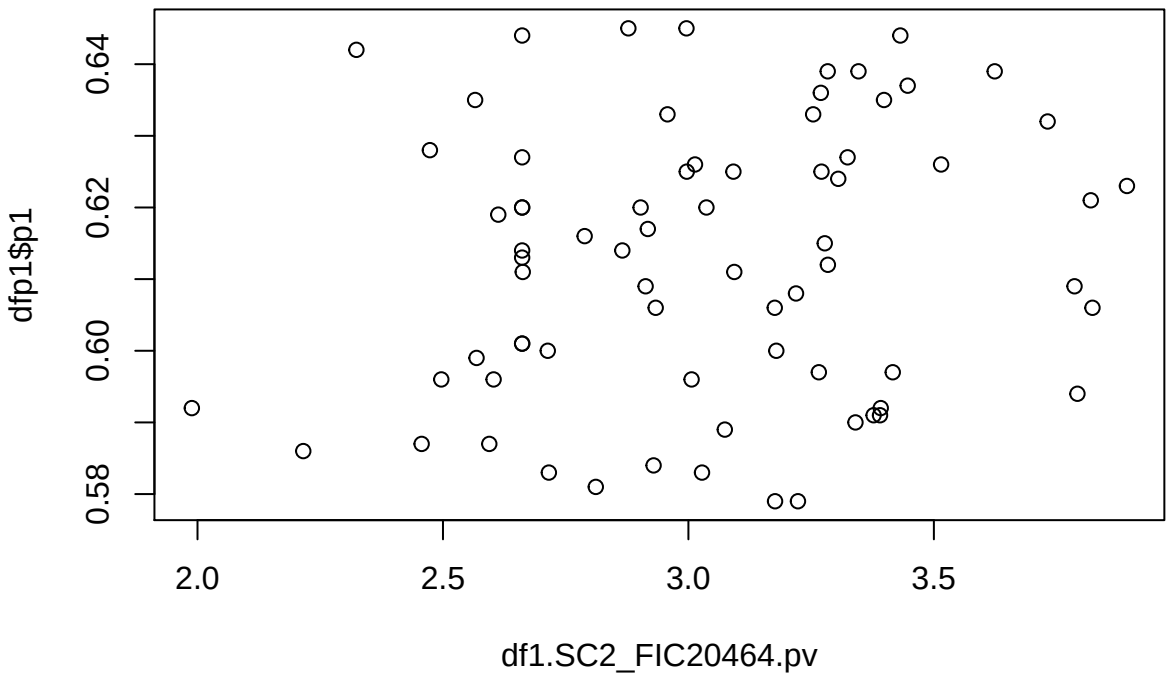


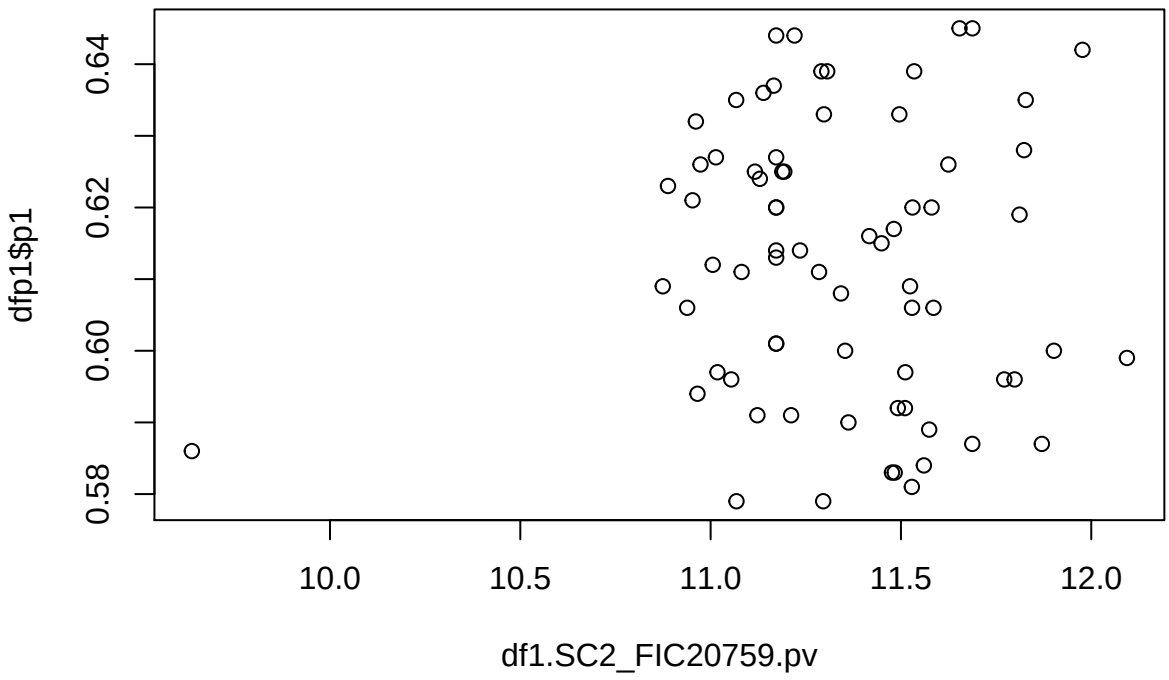

```
psych::multi.hist(input[,9:17])
```

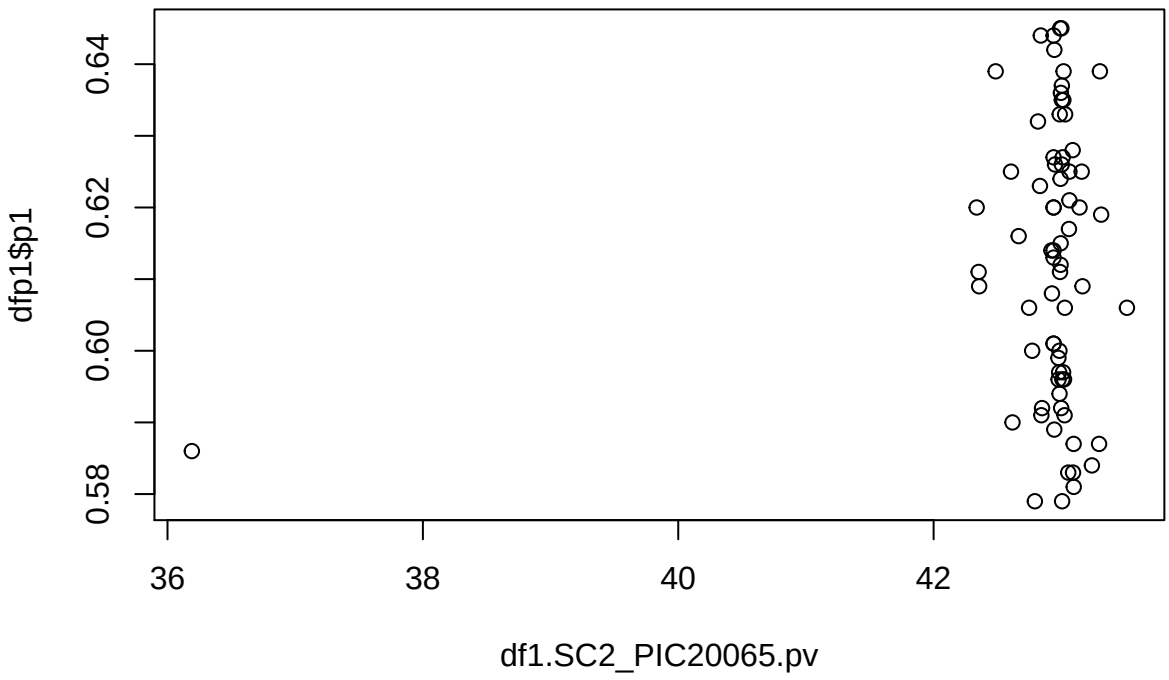


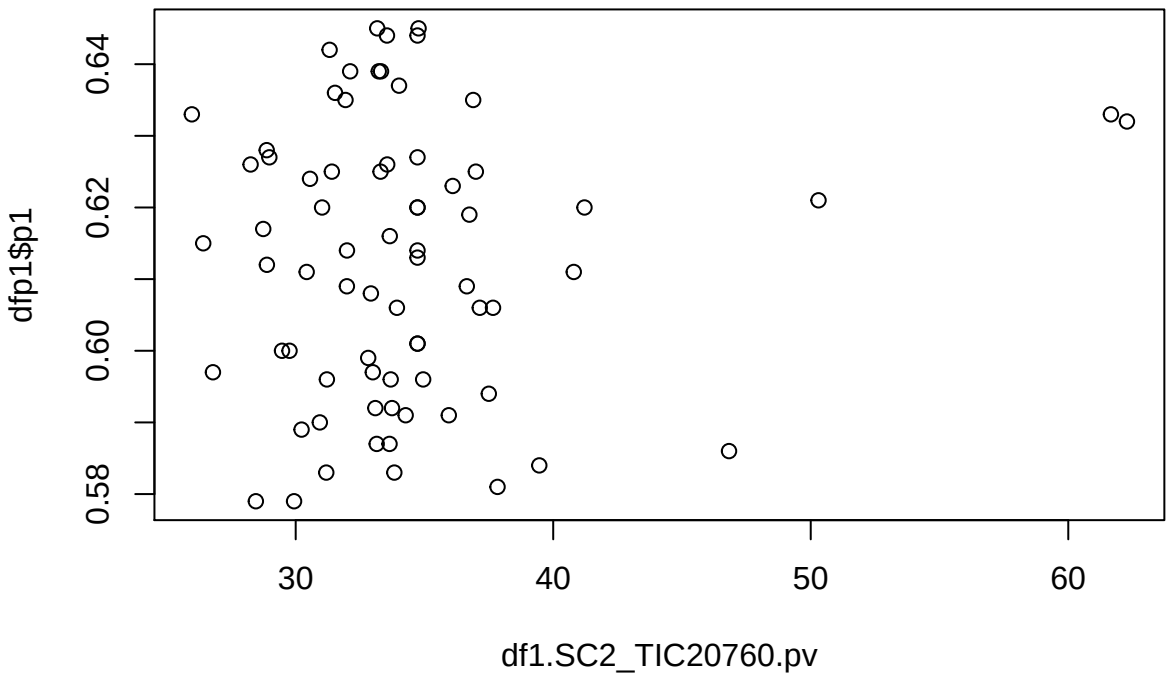

```
dfp1<-data.frame(p1,input)
plot(dfp1$p1~.,dfp1)
```

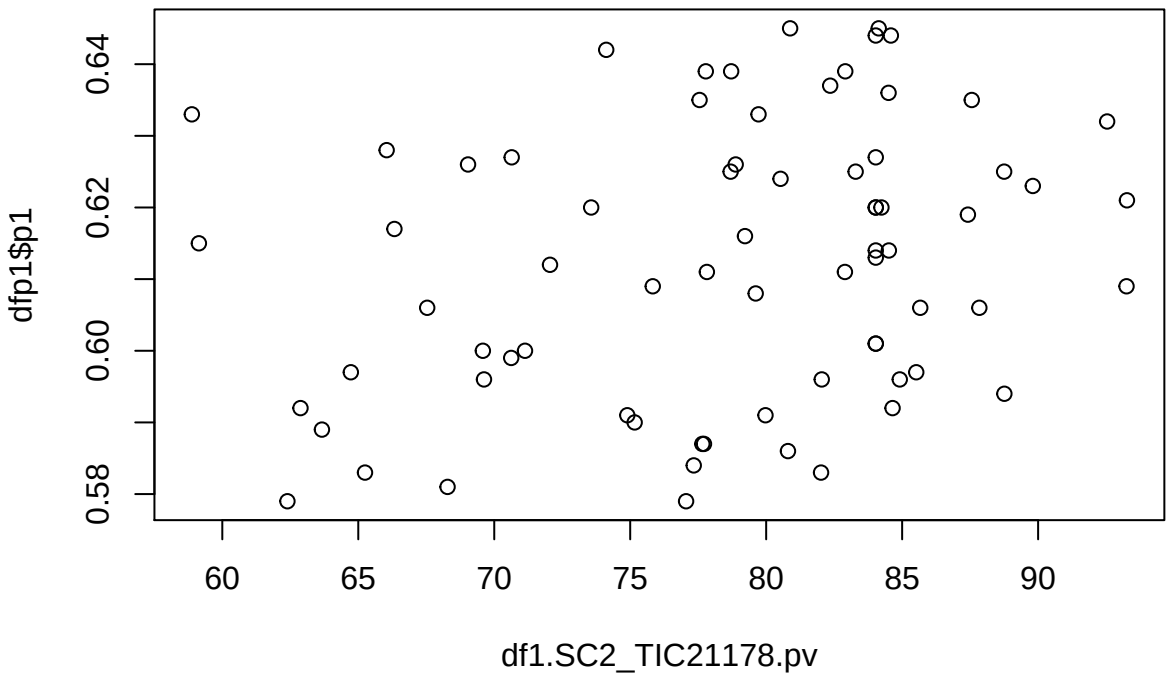

dfp1$p1
0.58
0.60
0.62
0.64
2.0
2.5
3.0
3.5
df1.SC2_FIC20463.pv

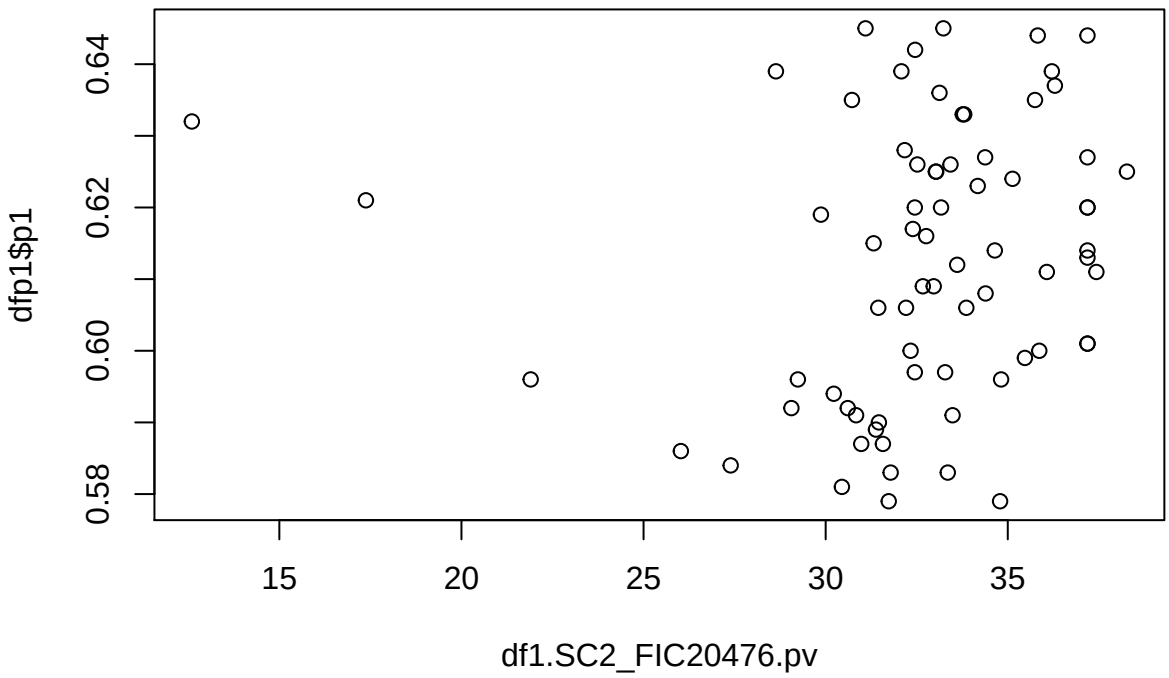

dfp1$p1
0.58
0.60
0.62
0.64
2.0
2.5
3.0
3.5
df1.SC2_FIC20464.pv

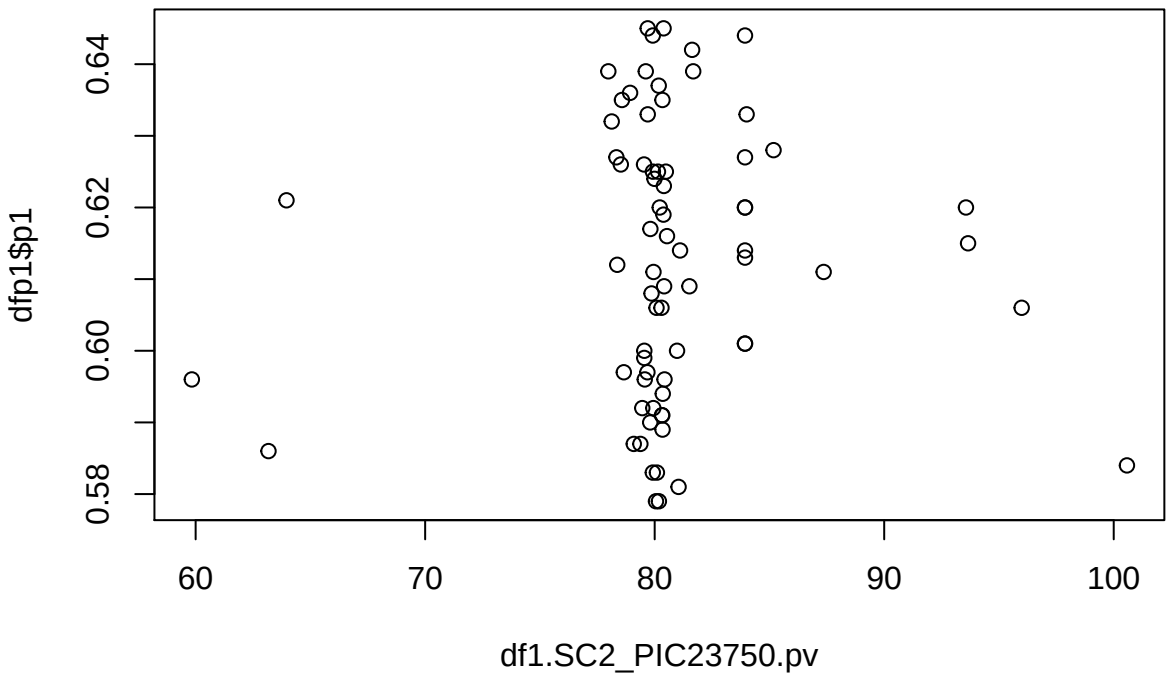

dfp1$p1
0.58
0.60
0.62
0.64
10.0
10.5
11.0
11.5
12.0
df1.SC2_FIC20759.pv

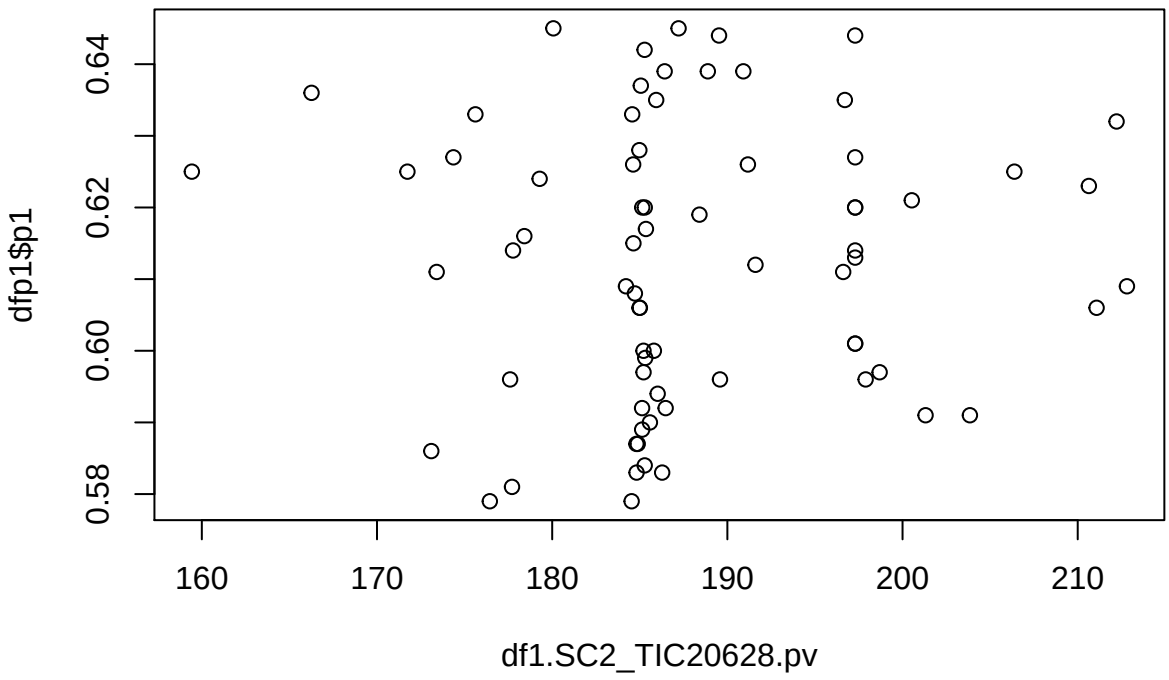

dfp1$p1
0.58
0.60
0.62
0.64
36
38
40
42
df1.SC2_PIC20065.pv

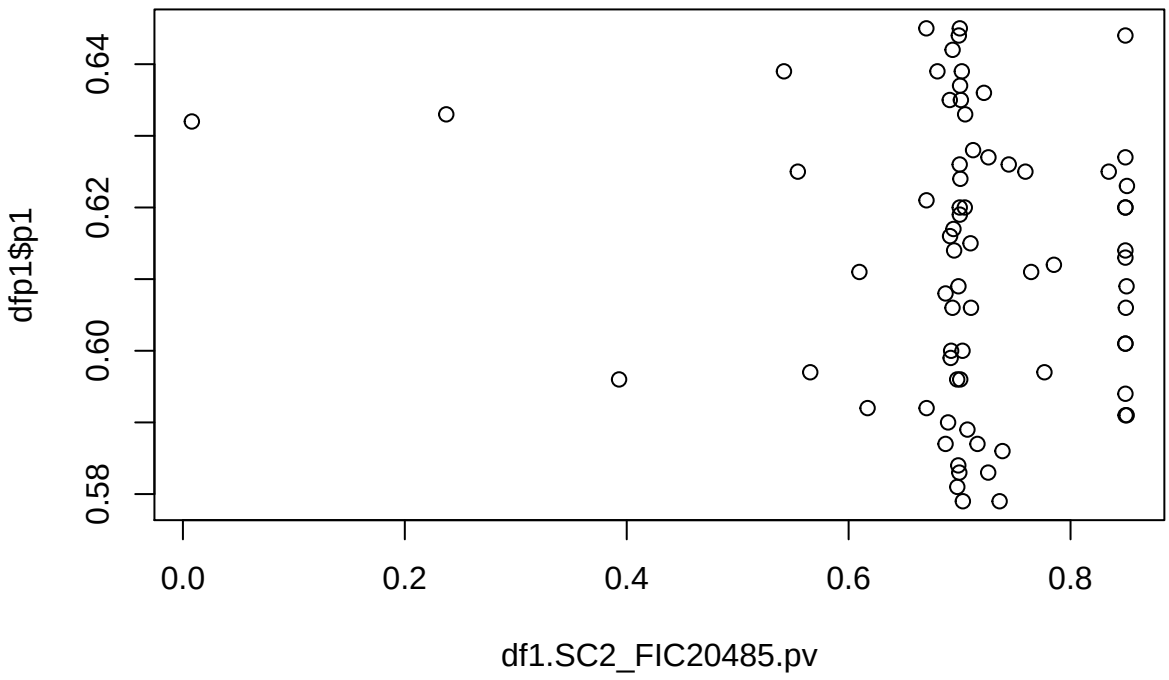

dfp1$p1
0.58
0.60
0.62
0.64
30
40
50
60
df1.SC2_TIC20760.pv

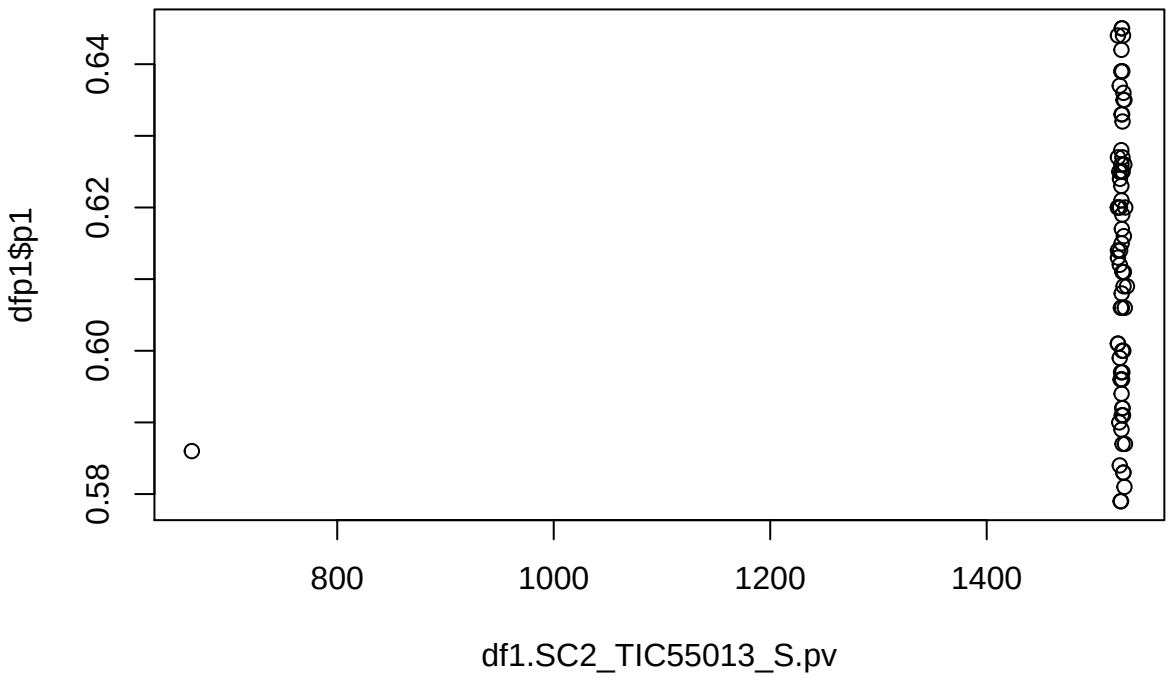

dfp1$p1
0.58
0.60
0.62
0.64
60
65
70
75
80
85
90
df1.SC2_TIC21178.pv

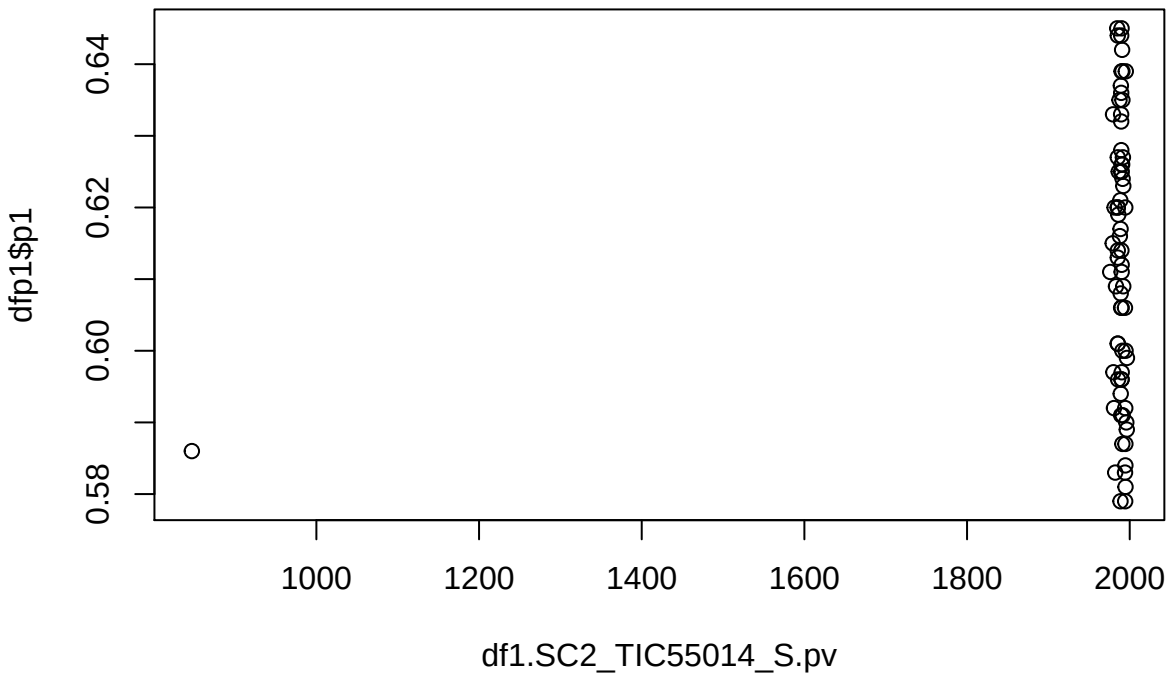

dfp1$p1
0.58
0.60
0.62
0.64
15
20
25
30
35
df1.SC2_FIC20476.pv

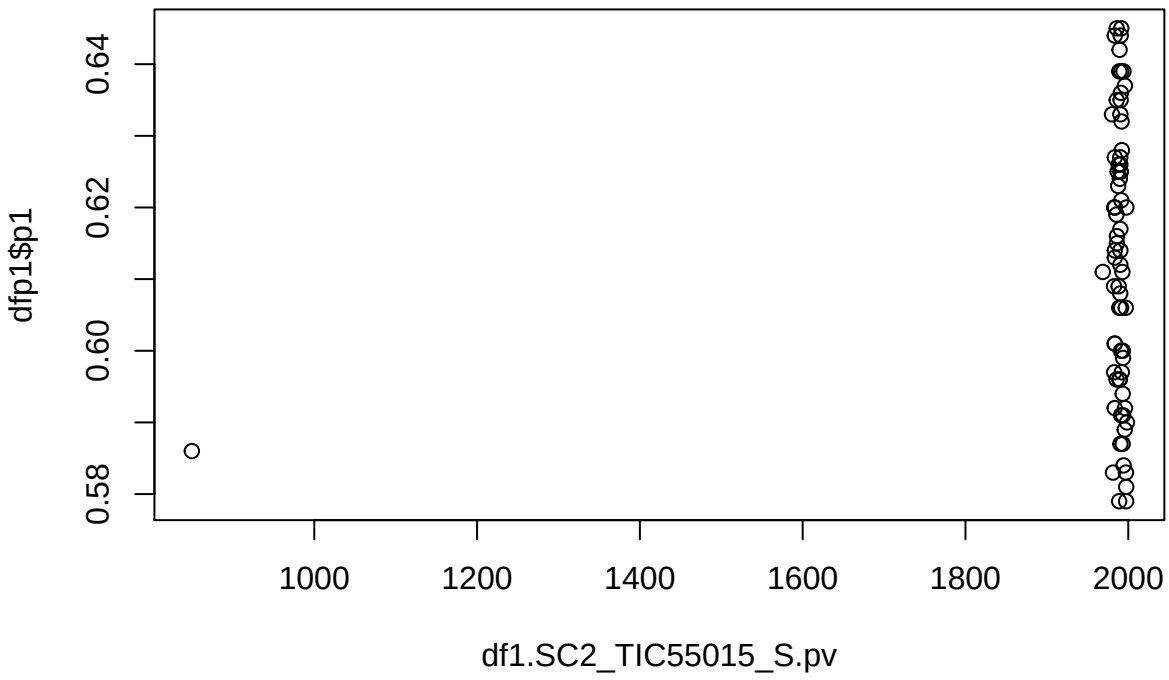

dfp1$p1
0.58
0.60
0.62
0.64
60
70
80
90
100
df1.SC2_PIC23750.pv

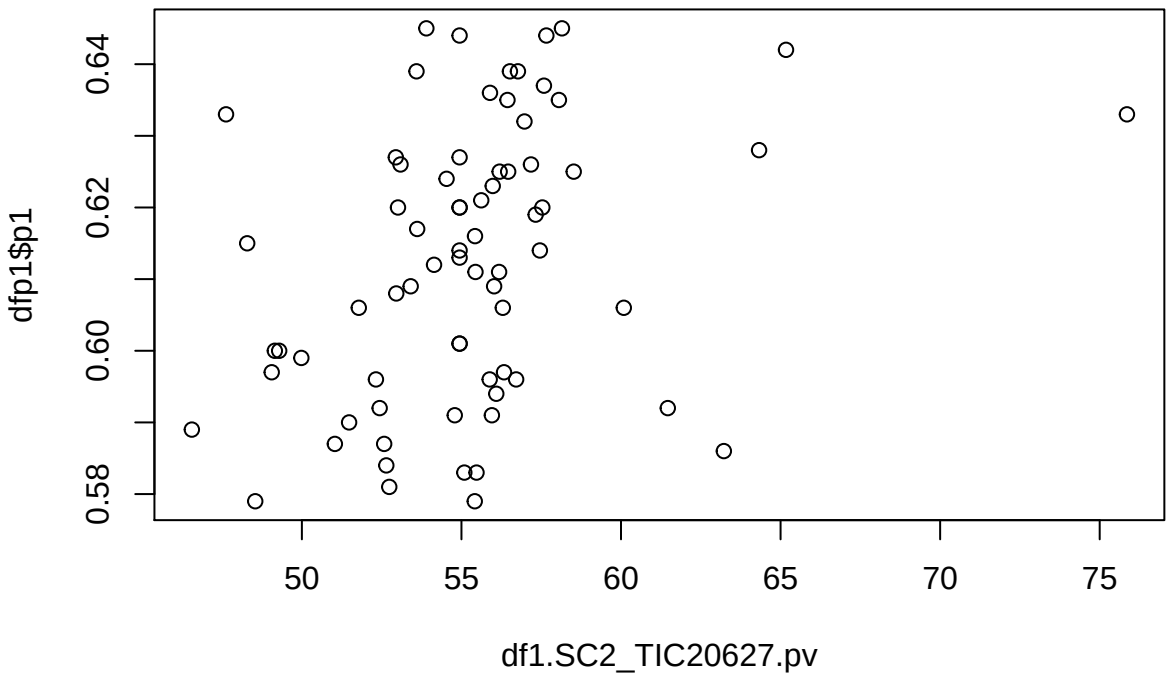

dfp1$p1
0.58
0.60
0.62
0.64
160
170
180
190
200
210
df1.SC2_TIC20628.pv

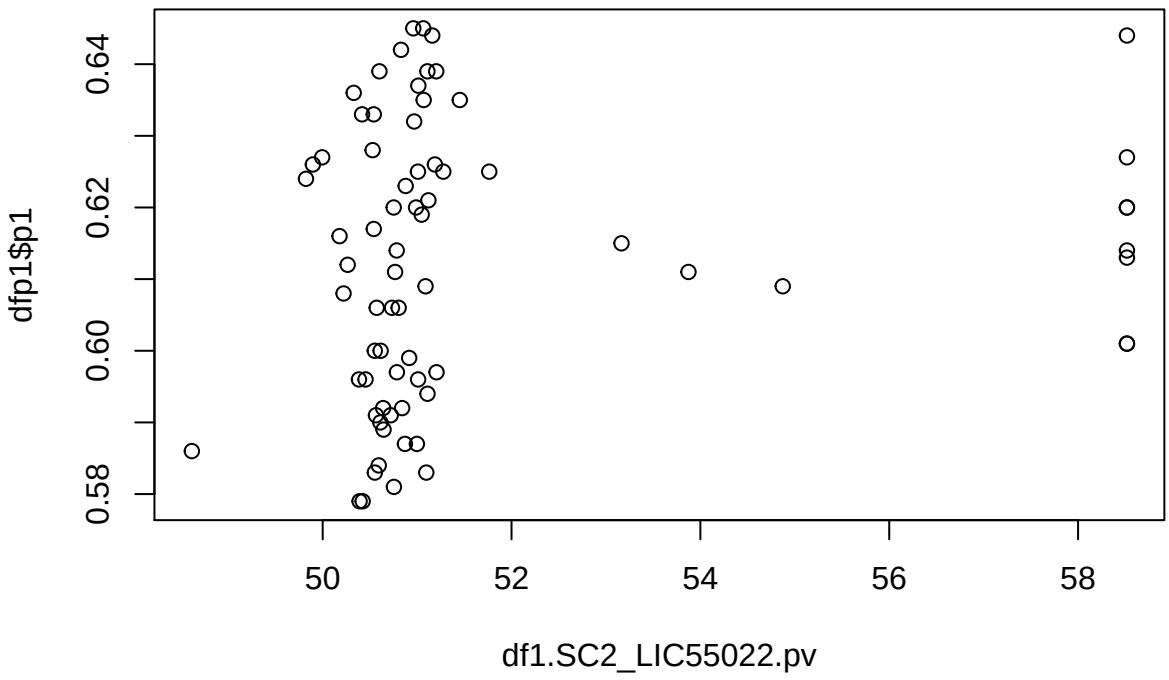

dfp1$p1
0.58
0.60
0.62
0.64
0.0
0.2
0.4
0.6
0.8
df1.SC2_FIC20485.pv

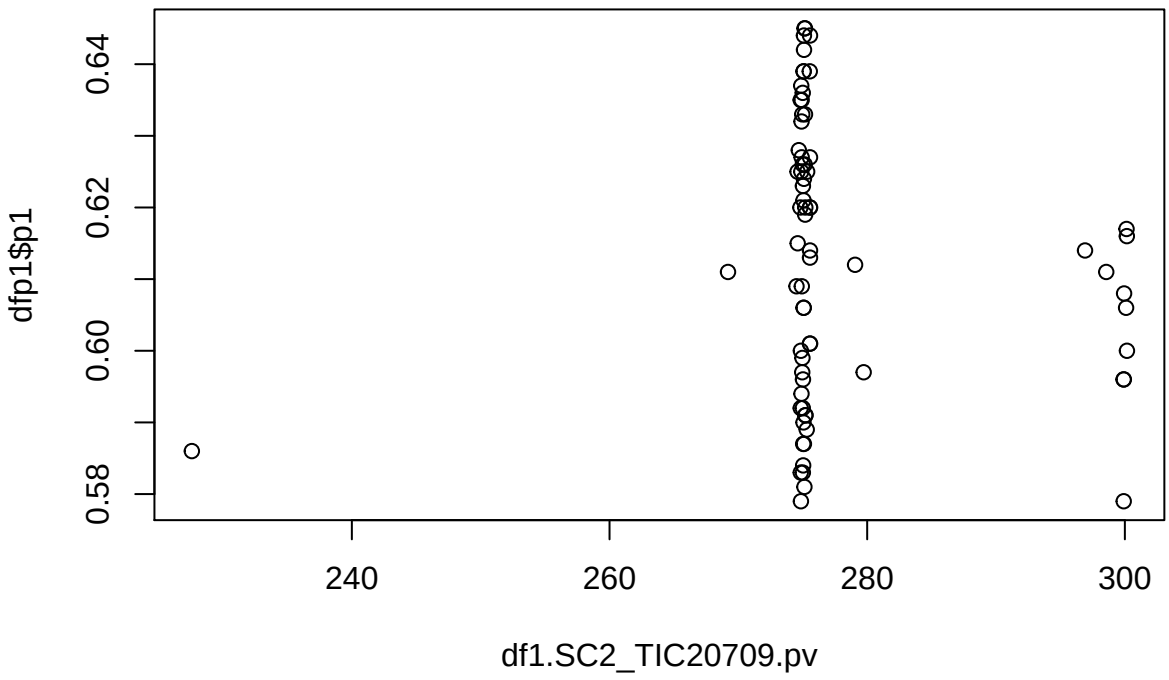

dfp1$p1
0.58
0.60
0.62
0.64
800
1000
1200
1400
df1.SC2_TIC55013_S.pv

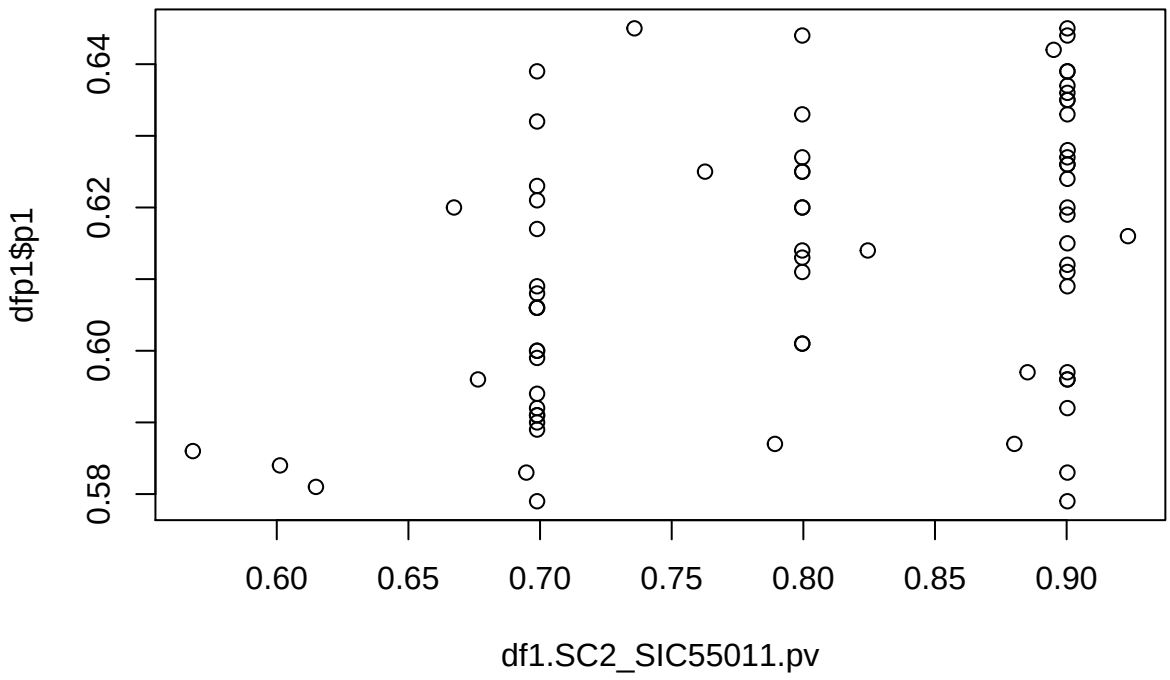

dfp1$p1
0.58
0.60
0.62
0.64
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55014_S.pv


dfp1$p1
0.58
0.60
0.62
0.64
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55015_S.pv

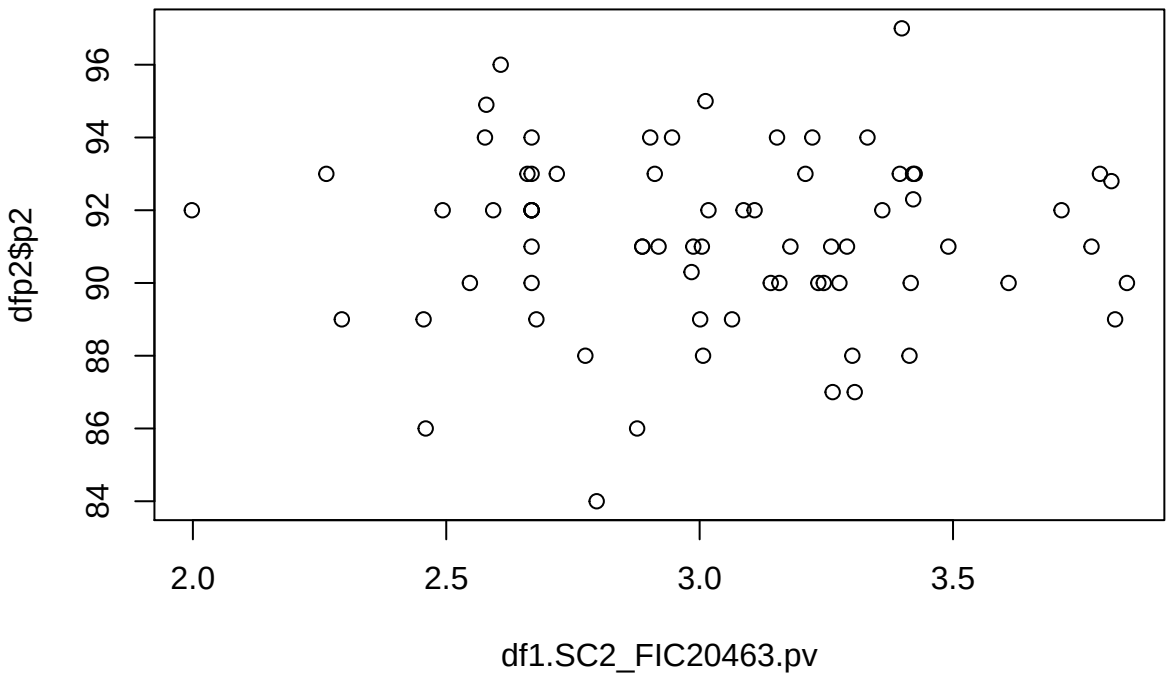

dfp1$p1
0.58
0.60
0.62
0.64
50
55
60
65
70
75
df1.SC2_TIC20627.pv

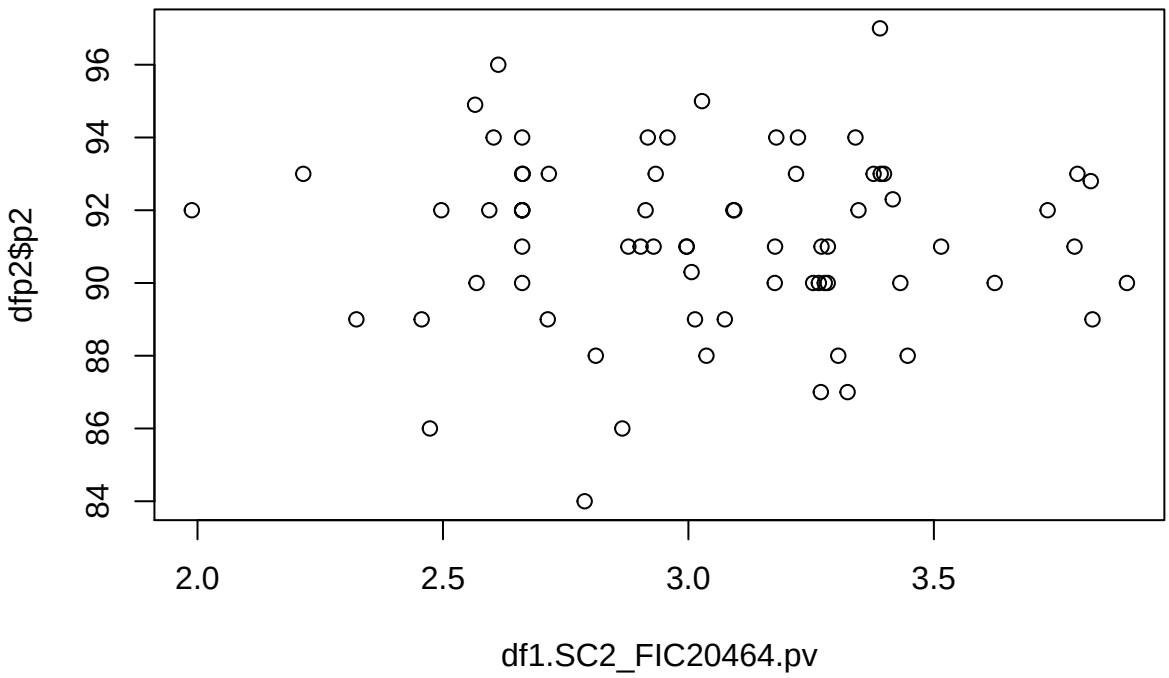

dfp1$p1
0.58
0.60
0.62
0.64
50
52
54
56
58
df1.SC2_LIC55022.pv

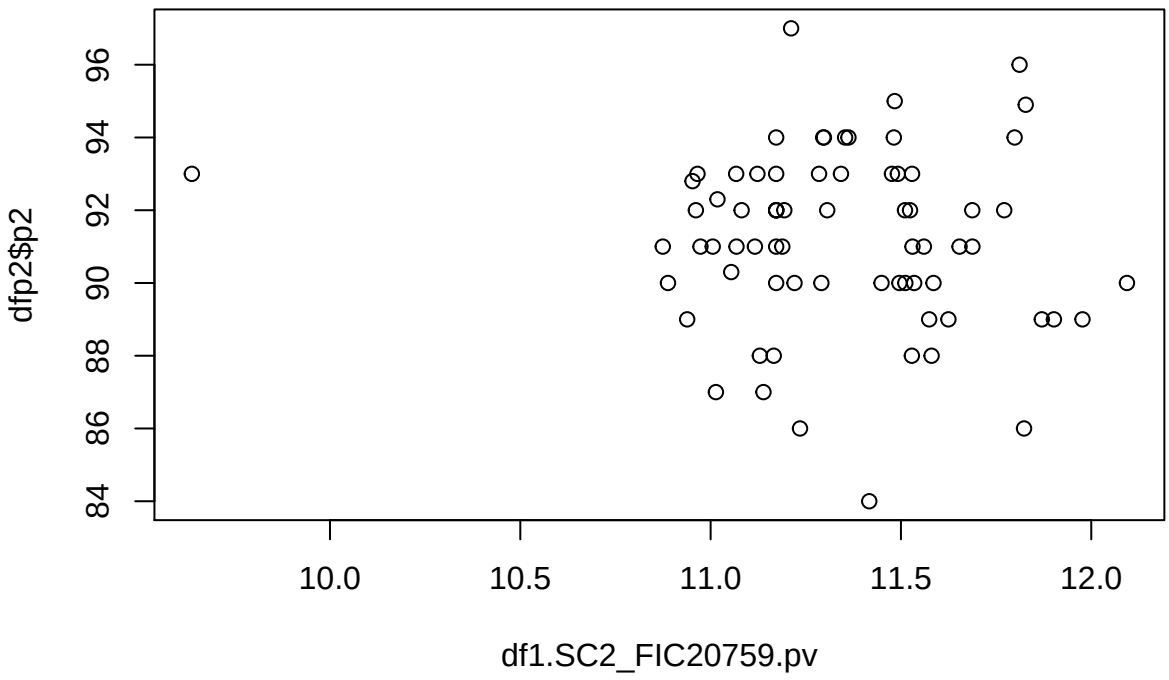

dfp1$p1
0.58
0.60
0.62
0.64
240
260
280
300
df1.SC2_TIC20709.pv

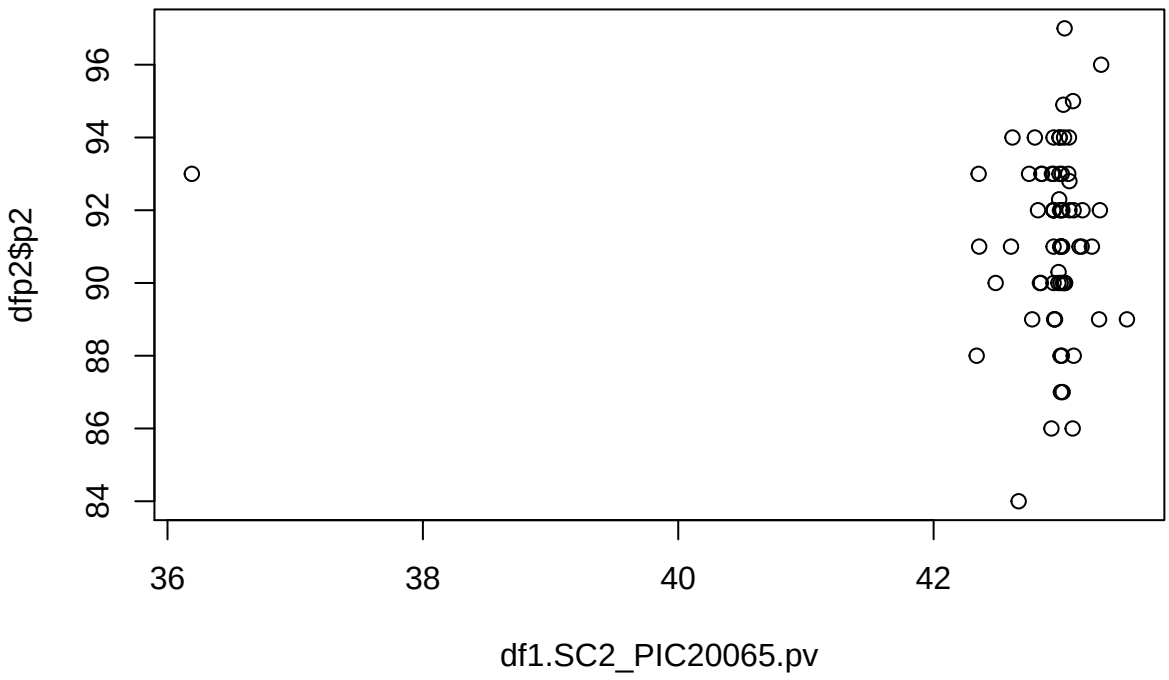


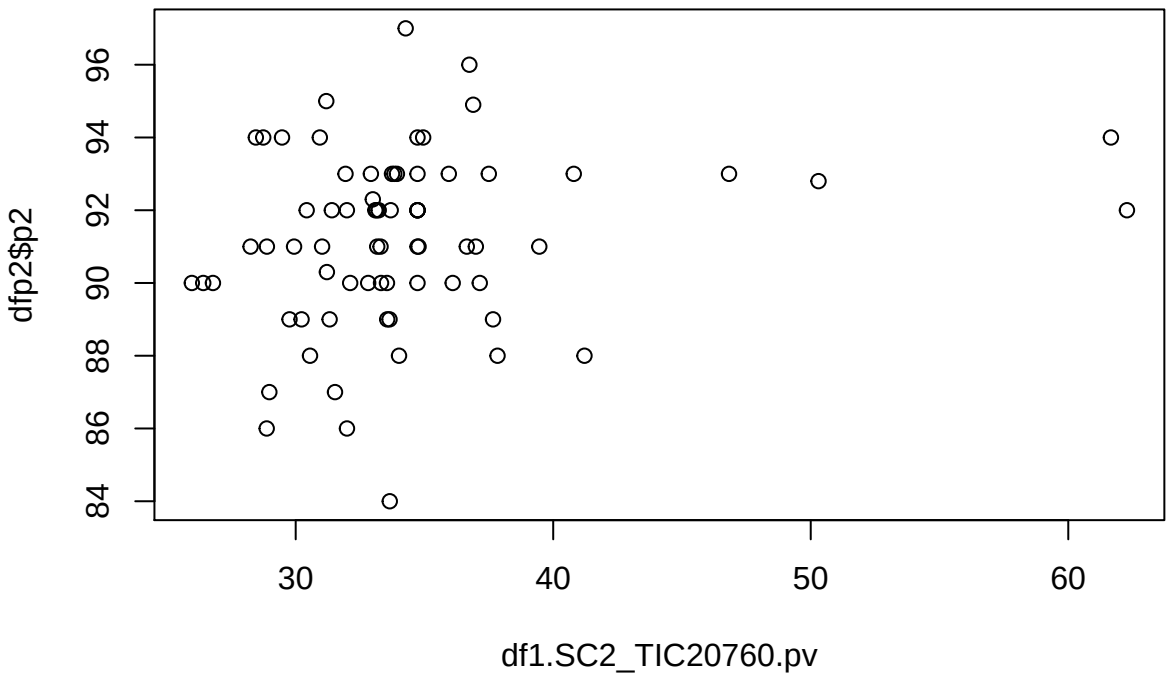

```
dfp2<-data.frame(p2,input)
plot(dfp2$p2~.,dfp2)
```

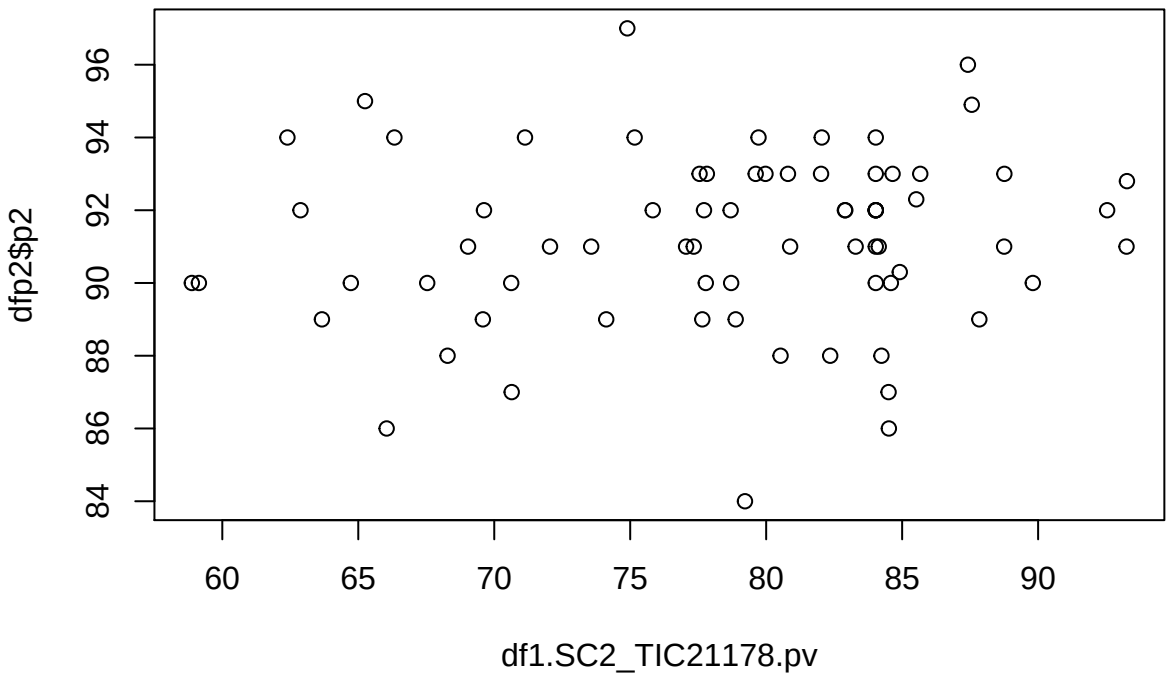

dfp2$p2
84
86
88
90
92
94
96
2.0
2.5
3.0
3.5
df1.SC2_FIC20463.pv

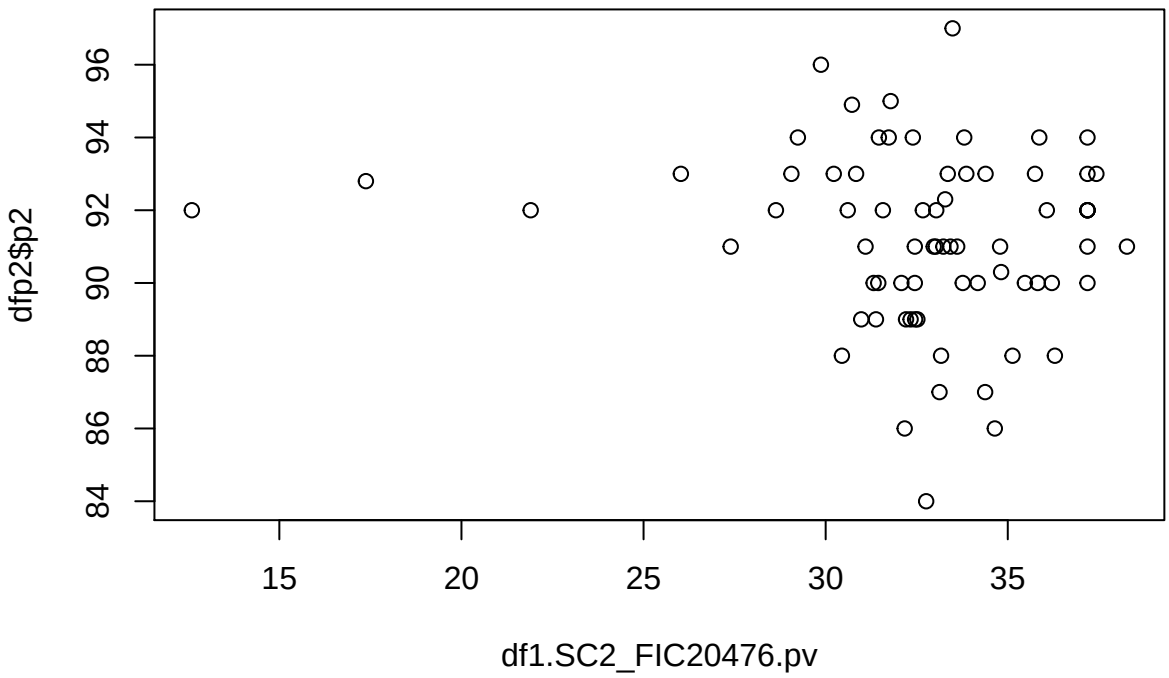

dfp2$p2
84
86
88
90
92
94
96
2.0
2.5
3.0
3.5
df1.SC2_FIC20464.pv

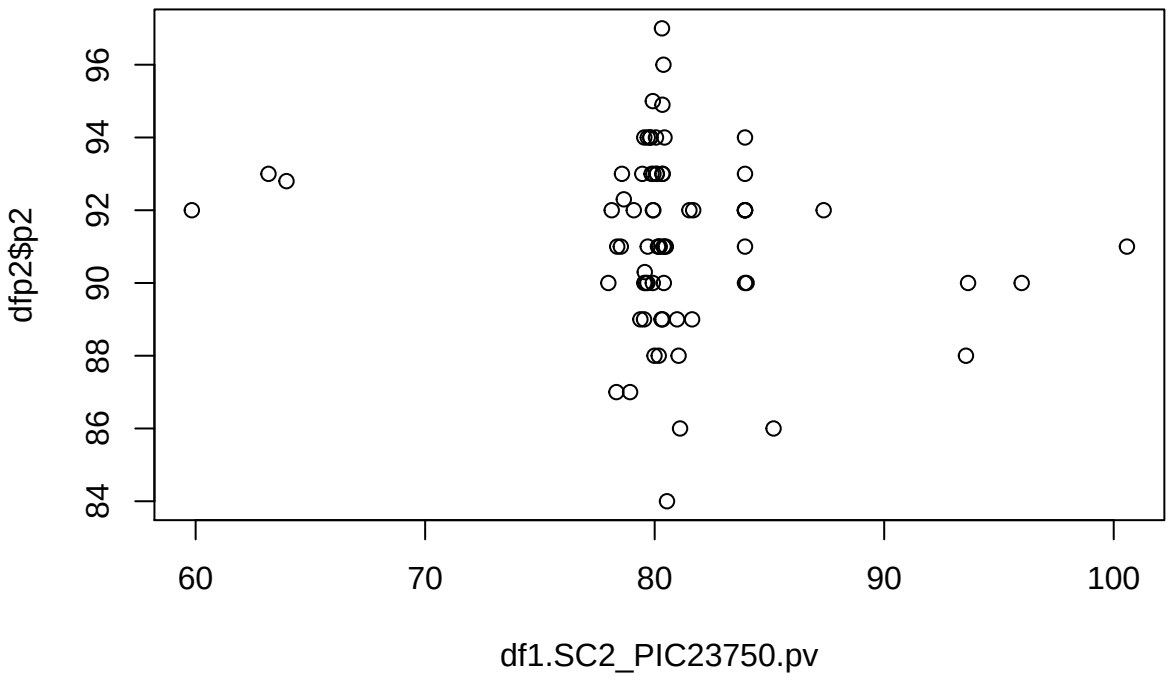

dfp2$p2
84
86
88
90
92
94
96
10.0
10.5
11.0
11.5
12.0
df1.SC2_FIC20759.pv

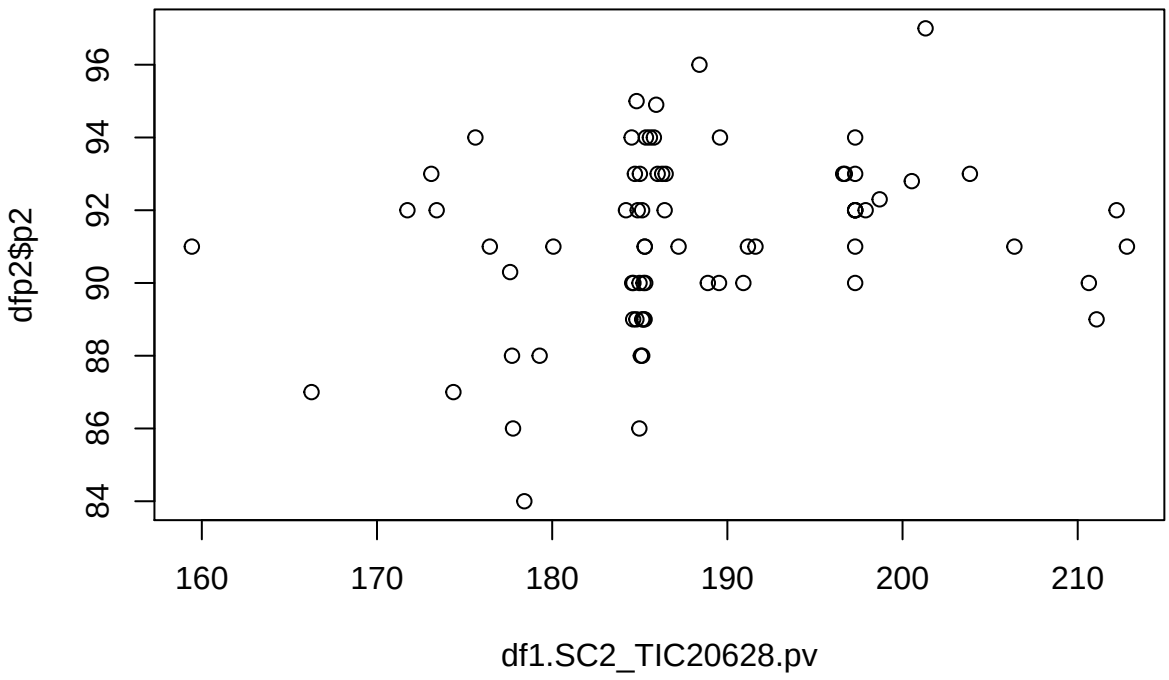

dfp2$p2
84
86
88
90
92
94
96
36
38
40
42
df1.SC2_PIC20065.pv

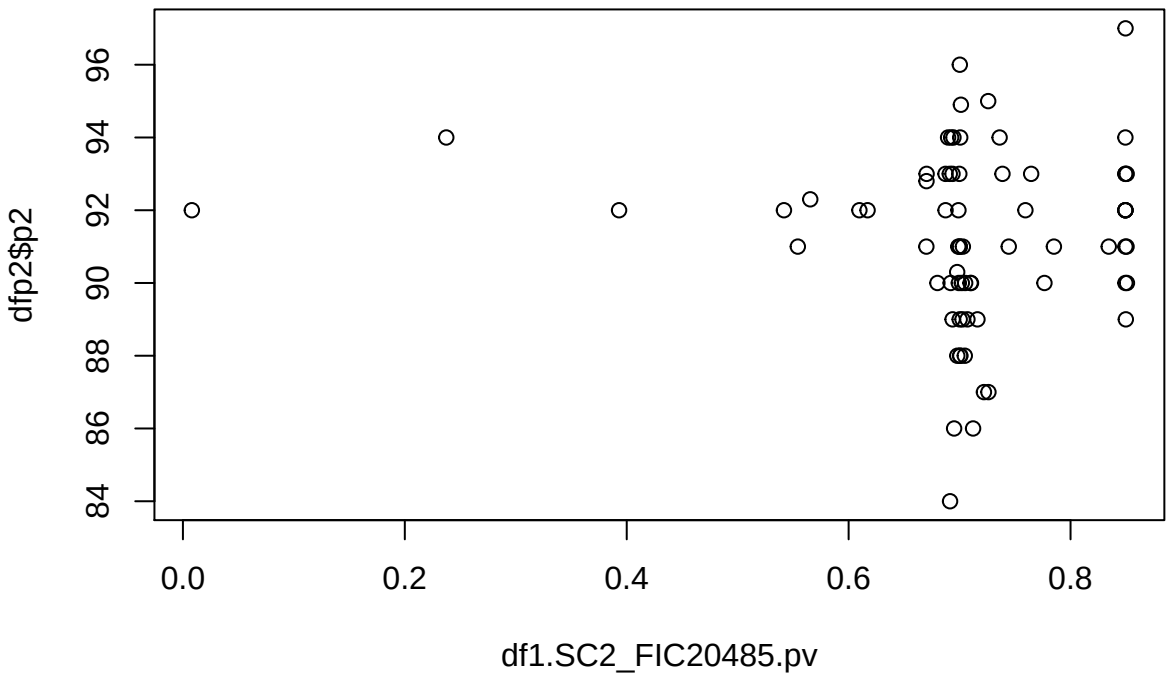

dfp2$p2
84
86
88
90
92
94
96
30
40
50
60
df1.SC2_TIC20760.pv

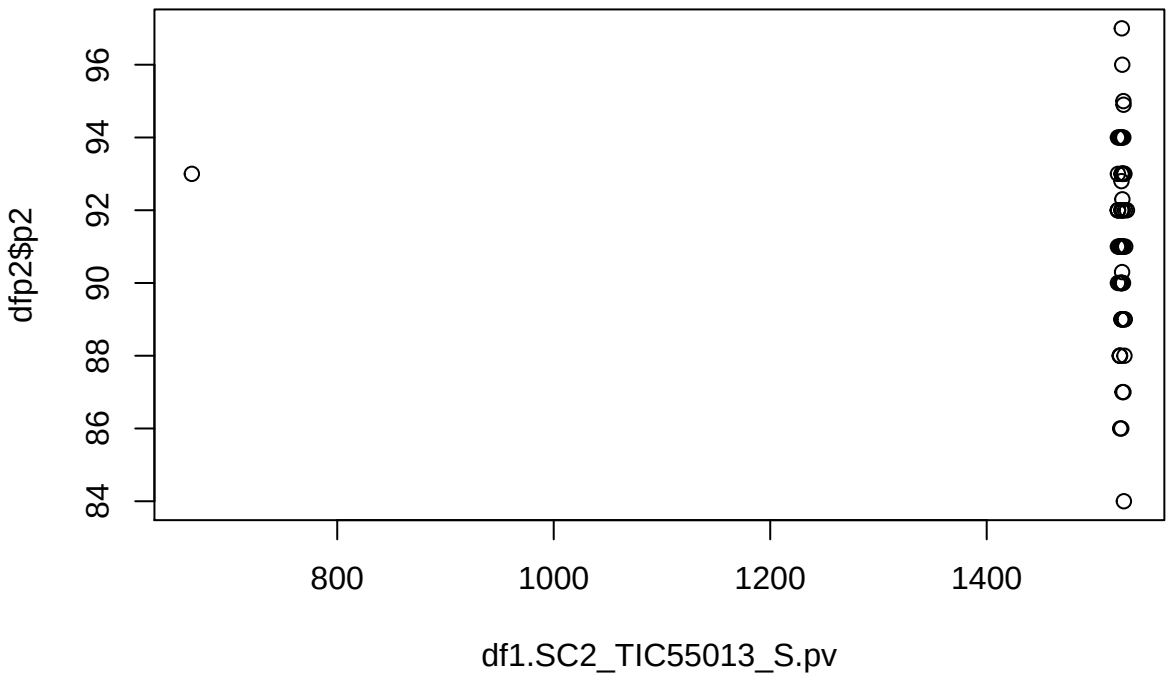

dfp2$p2
84
86
88
90
92
94
96
60
65
70
75
80
85
90
df1.SC2_TIC21178.pv

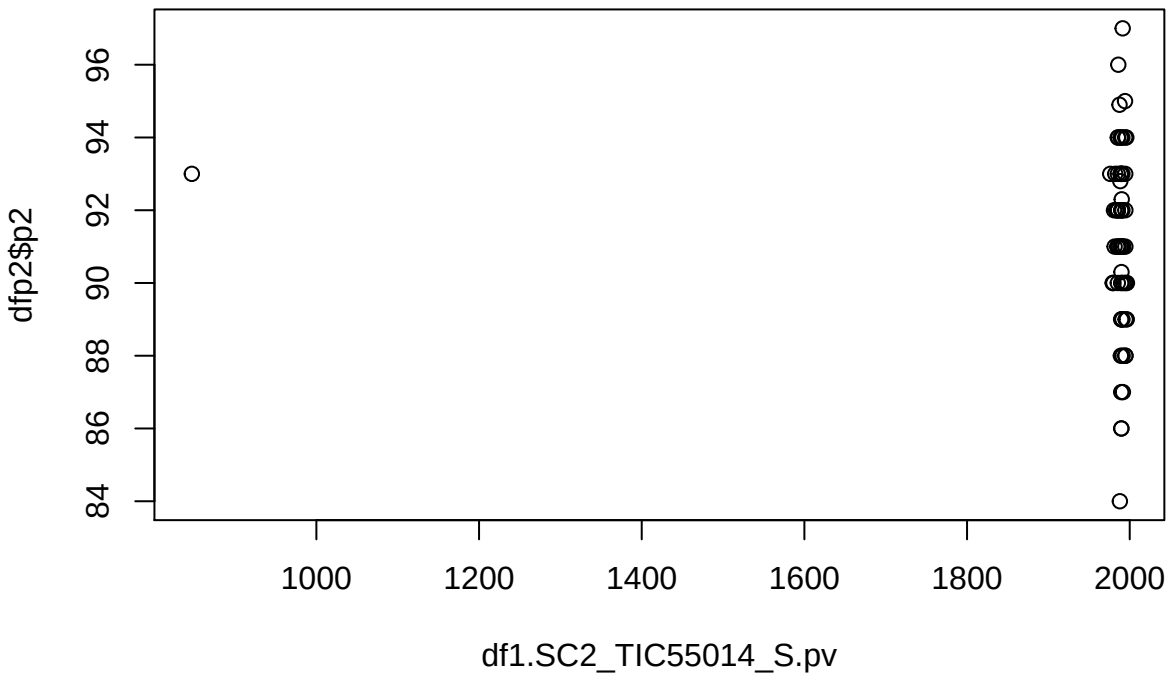

dfp2$p2
84
86
88
90
92
94
96
15
20
25
30
35
df1.SC2_FIC20476.pv

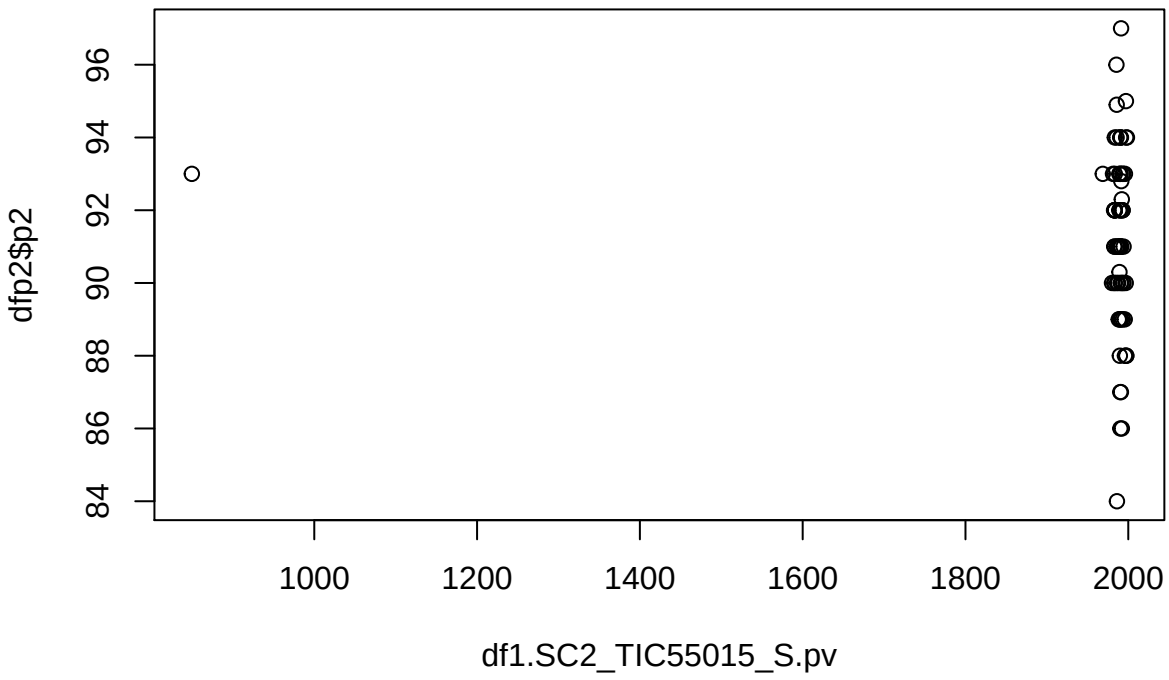

dfp2$p2
84
86
88
90
92
94
96
60
70
80
90
100
df1.SC2_PIC23750.pv

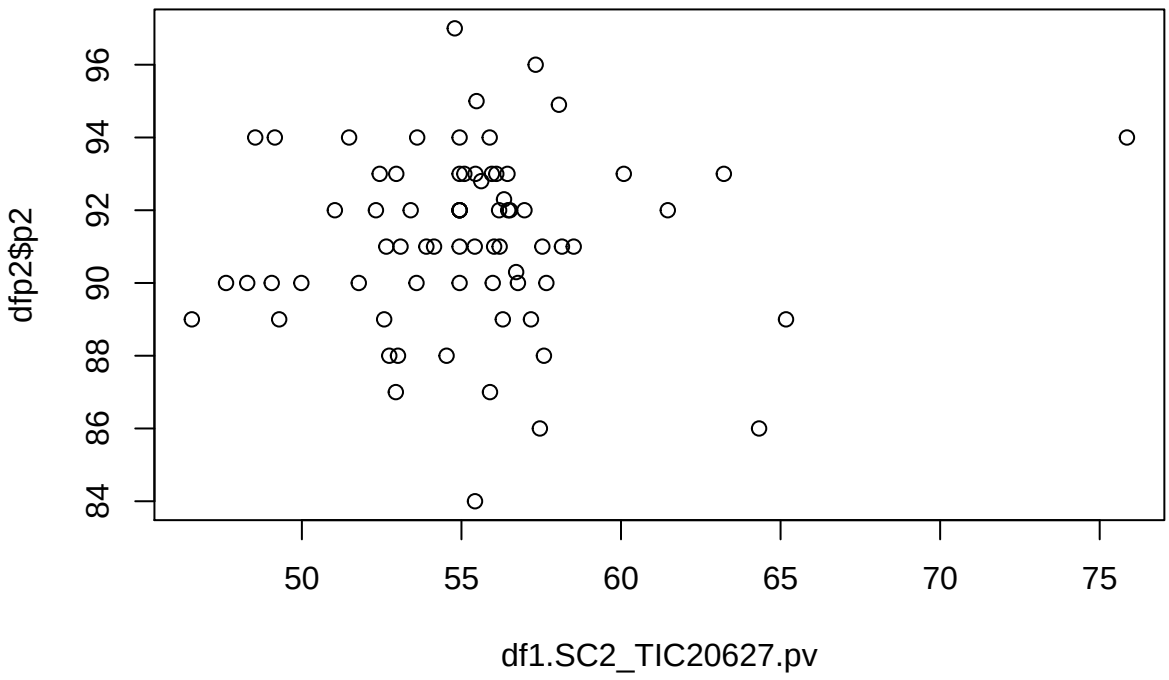

dfp2$p2
84
86
88
90
92
94
96
160
170
180
190
200
210
df1.SC2_TIC20628.pv

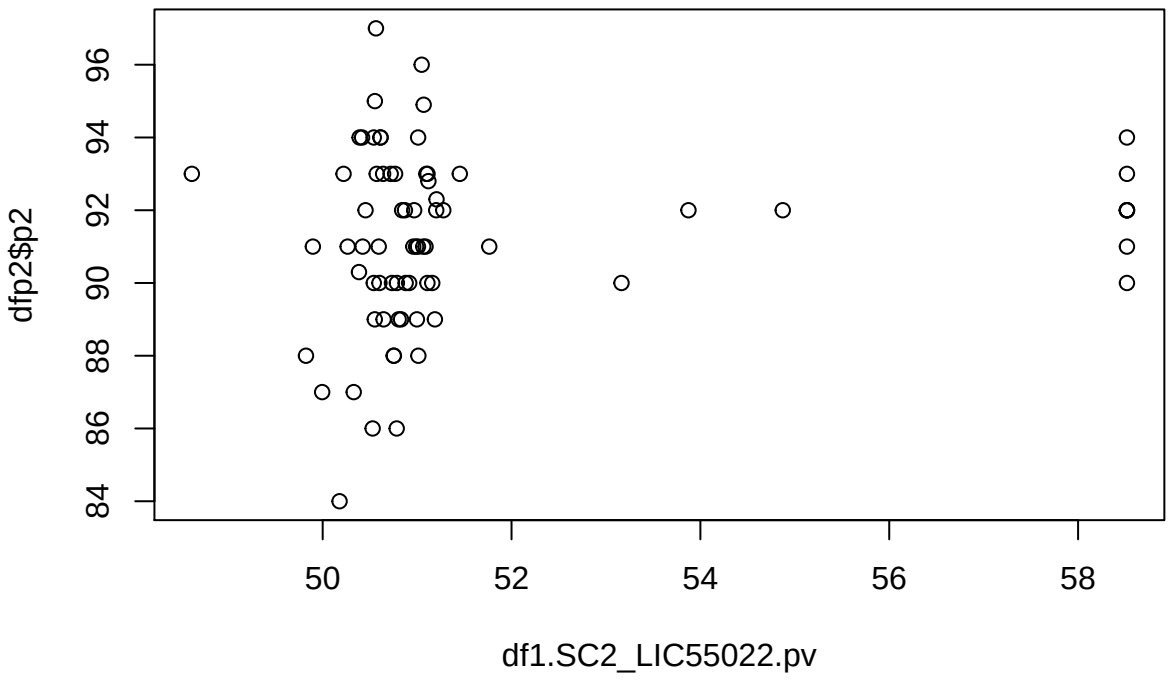

dfp2$p2
84
86
88
90
92
94
96
0.0
0.2
0.4
0.6
0.8
df1.SC2_FIC20485.pv

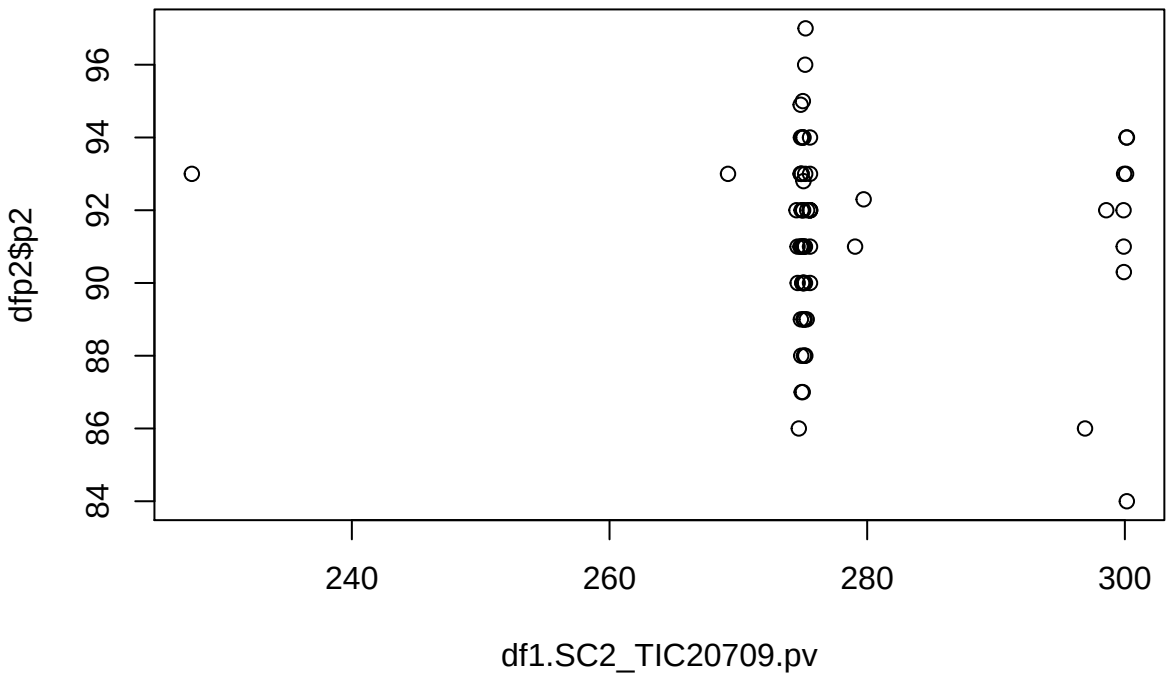

dfp2$p2
84
86
88
90
92
94
96
800
1000
1200
1400
df1.SC2_TIC55013_S.pv

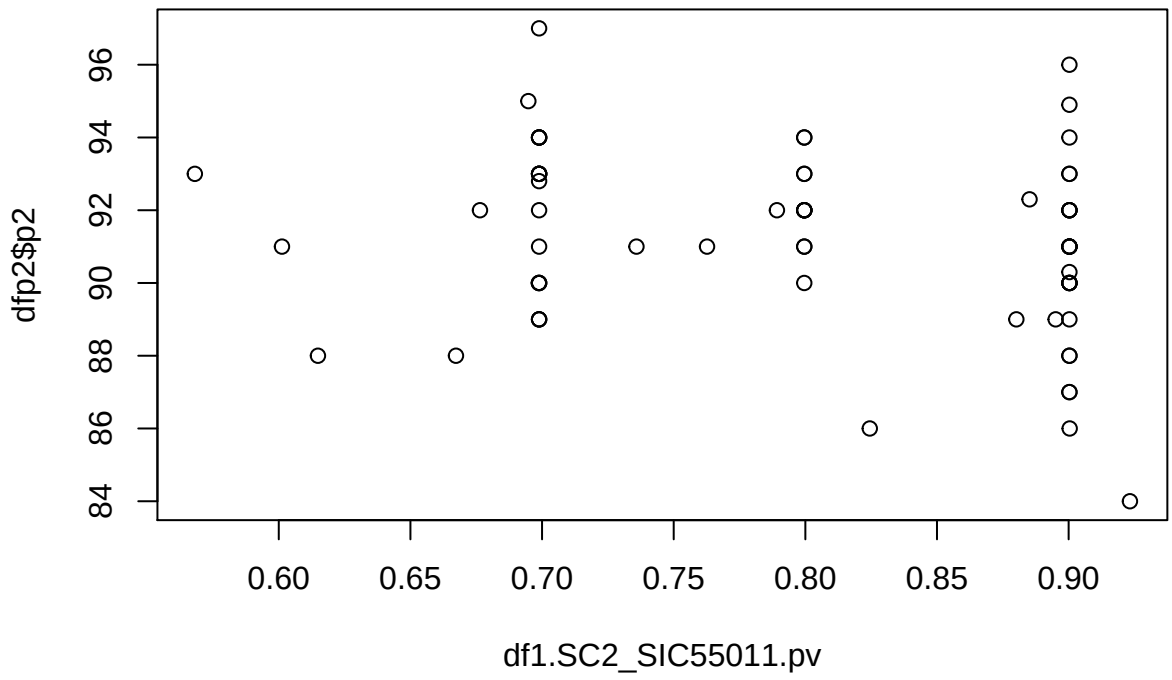

dfp2$p2
84
86
88
90
92
94
96
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55014_S.pv


dfp2$p2
84
86
88
90
92
94
96
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55015_S.pv

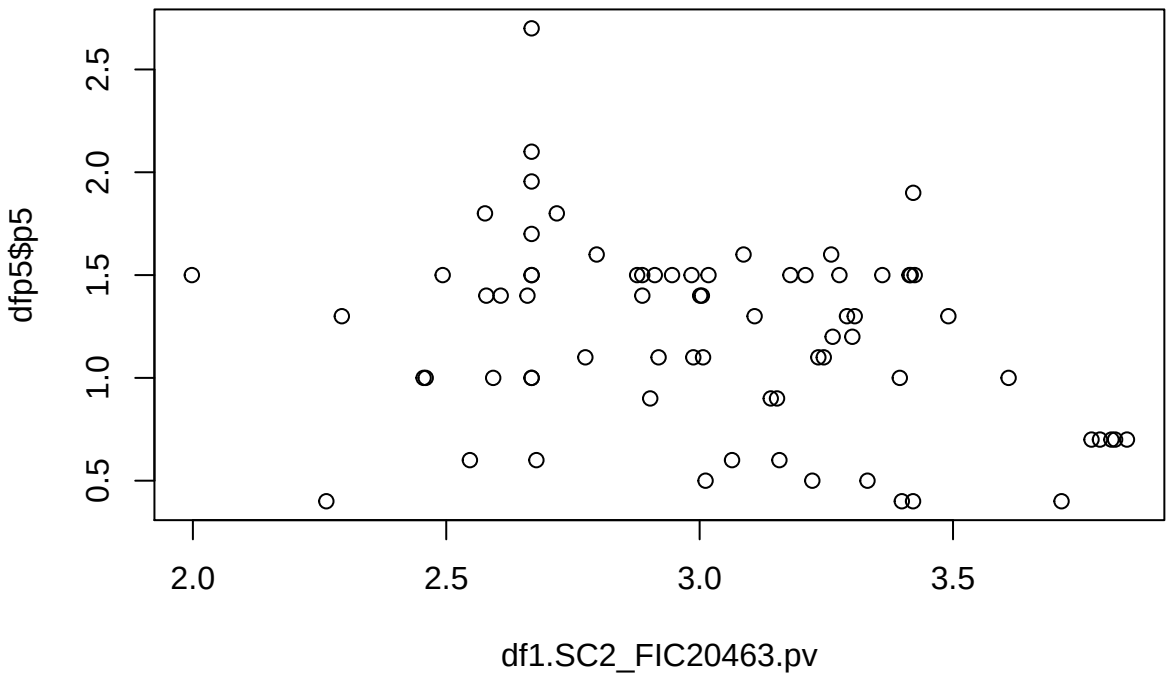

dfp2$p2
84
86
88
90
92
94
96
50
55
60
65
70
75
df1.SC2_TIC20627.pv

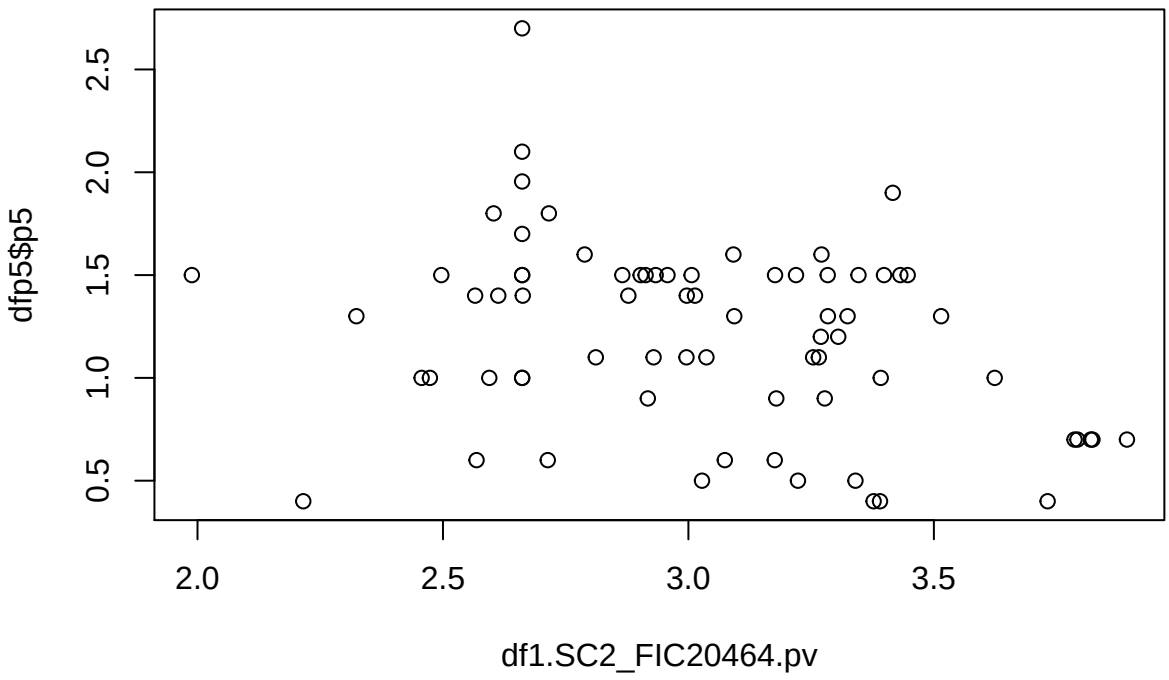

dfp2$p2
84
86
88
90
92
94
96
50
52
54
56
58
df1.SC2_LIC55022.pv

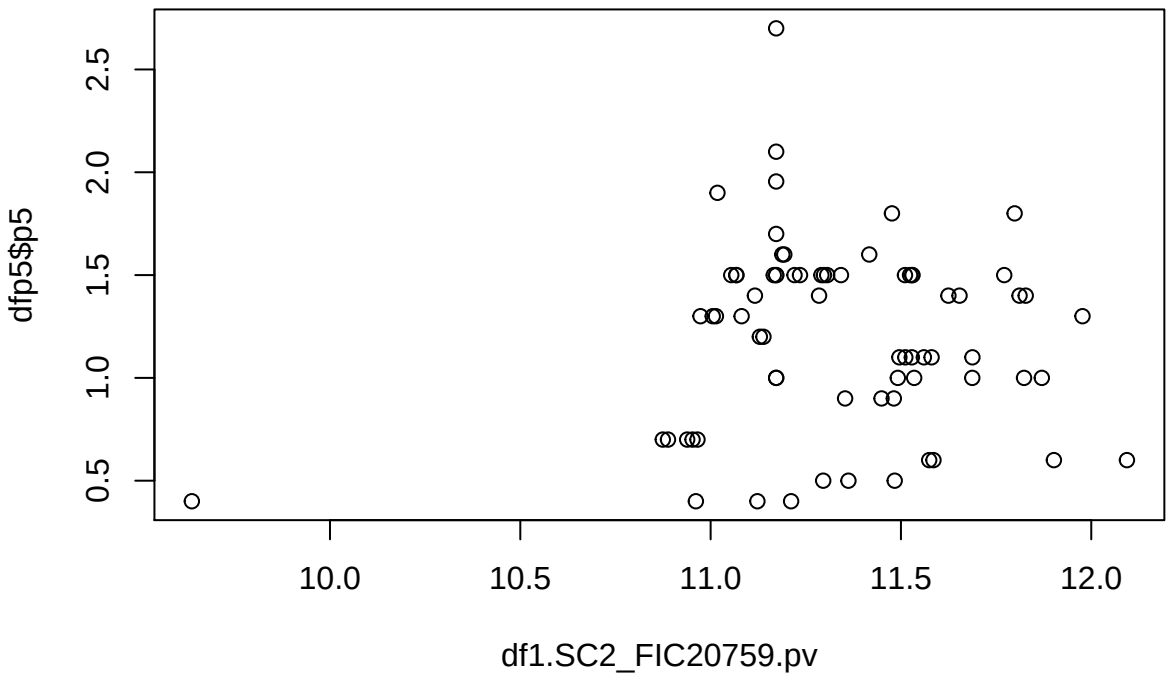

dfp2$p2
84
86
88
90
92
94
96
240
260
280
300
df1.SC2_TIC20709.pv

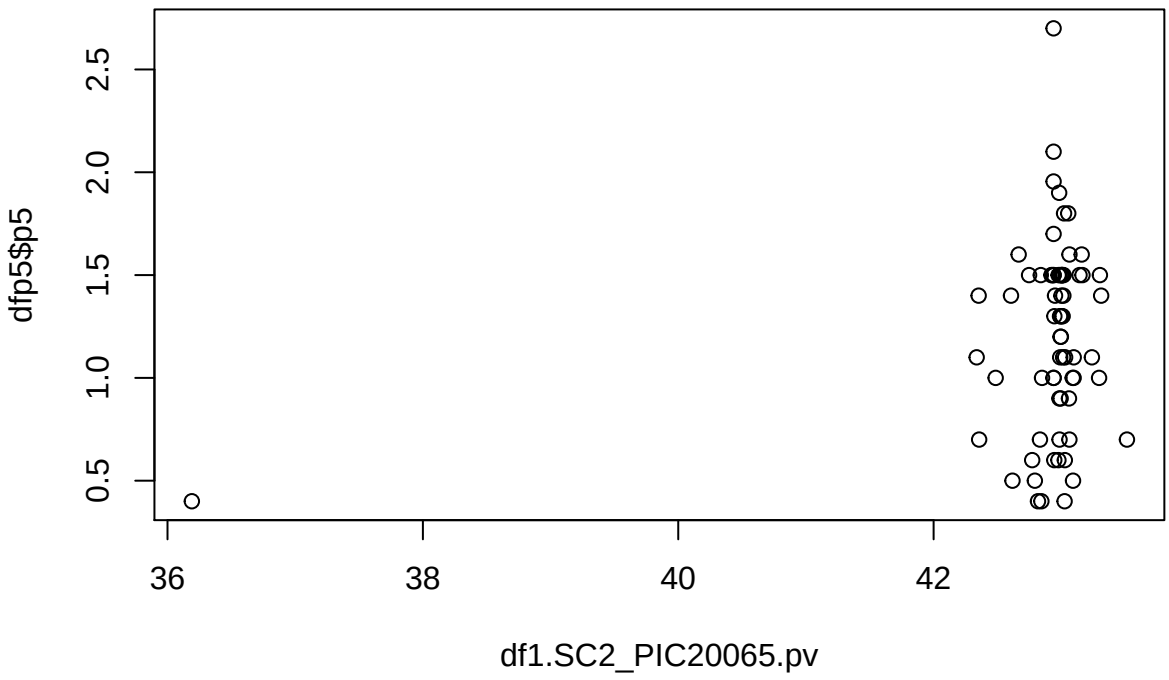


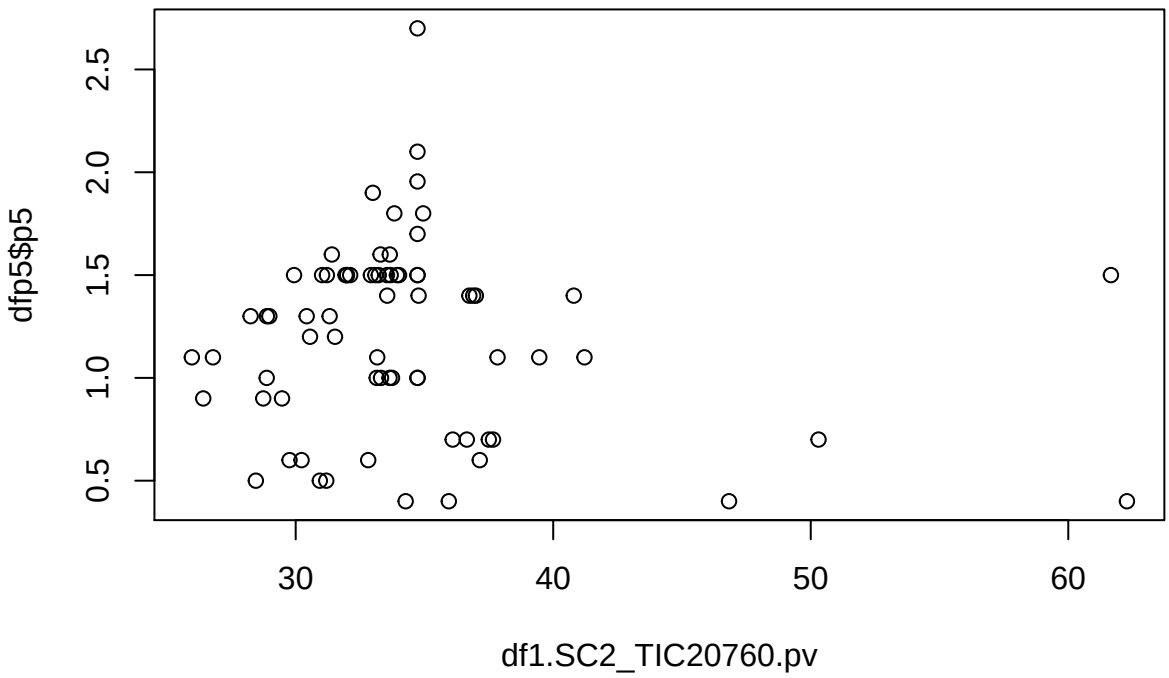

```
dfp5<-data.frame(p5,input)
plot(dfp5$p5~.,dfp5)
```

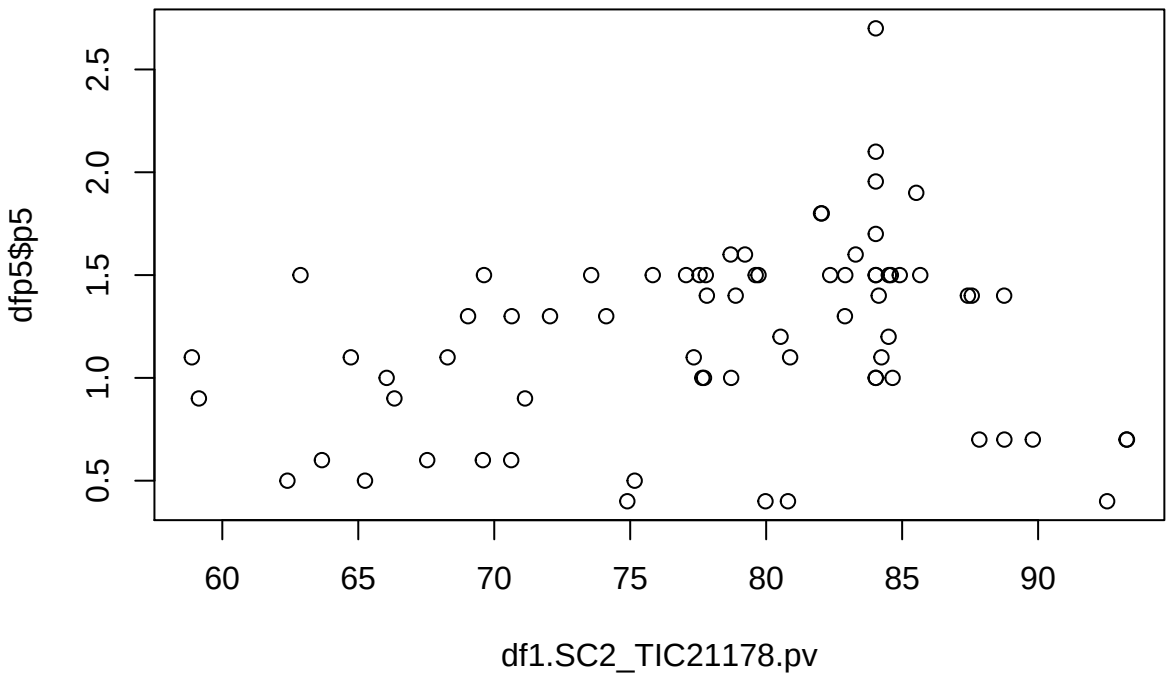

dfp5$p5
0.5
1.0
1.5
2.0
2.5
2.0
2.5
3.0
3.5
df1.SC2_FIC20463.pv

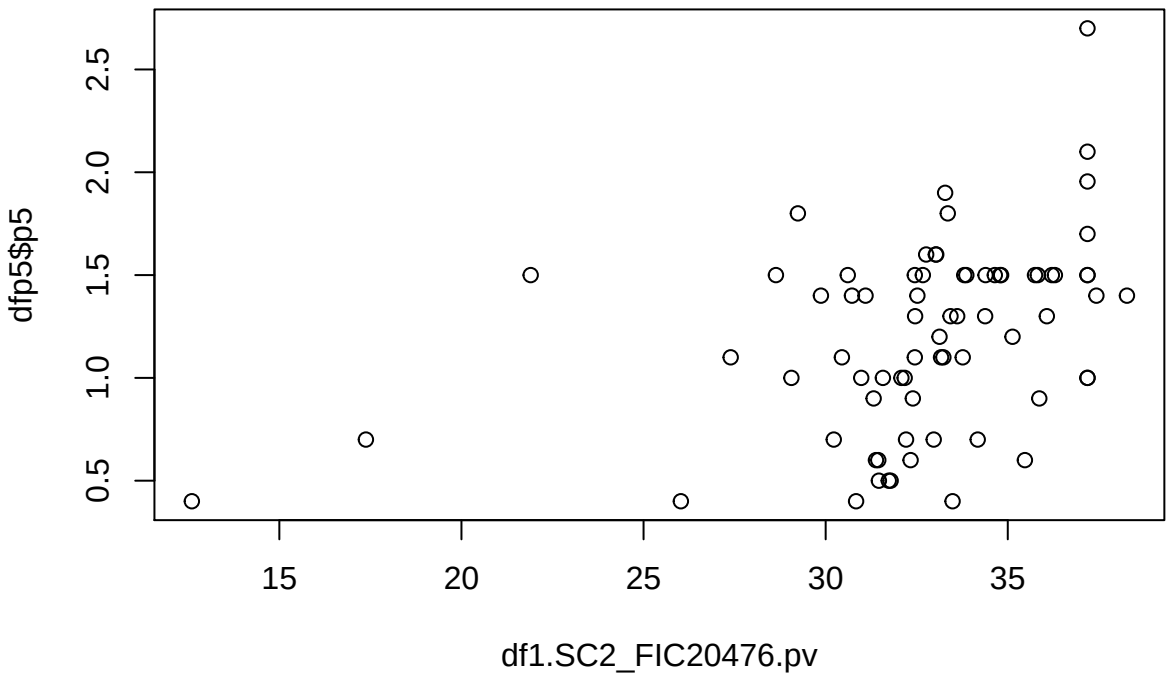

dfp5$p5
0.5
1.0
1.5
2.0
2.5
2.0
2.5
3.0
3.5
df1.SC2_FIC20464.pv

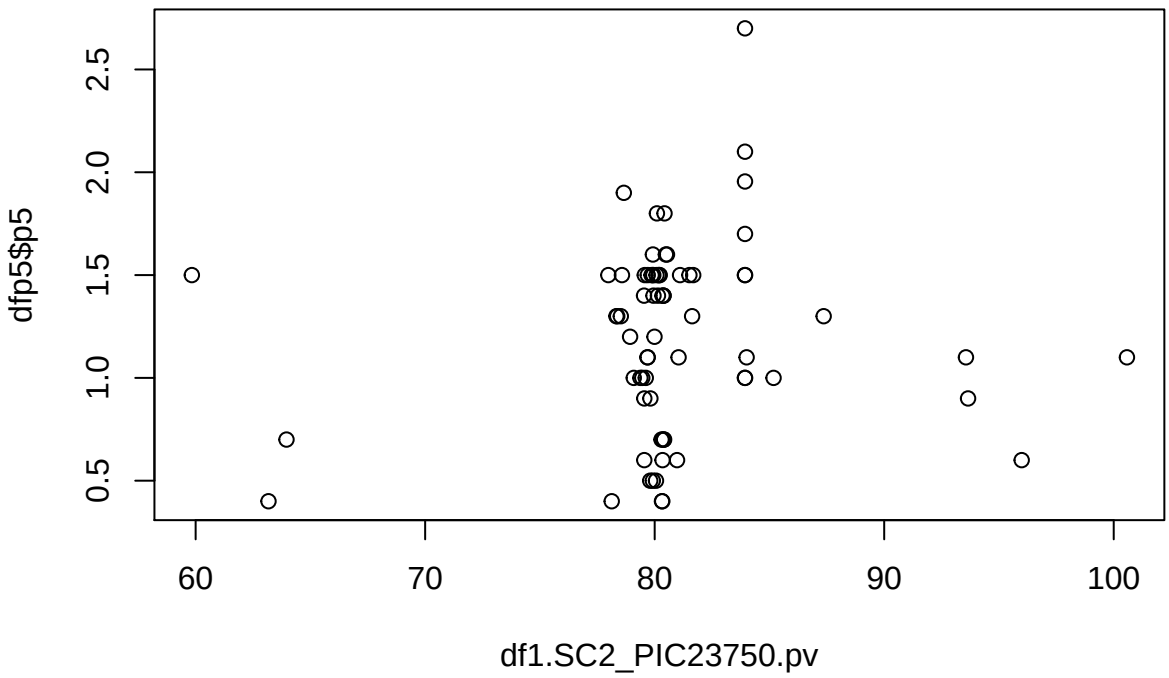

dfp5$p5
0.5
1.0
1.5
2.0
2.5
10.0
10.5
11.0
11.5
12.0
df1.SC2_FIC20759.pv

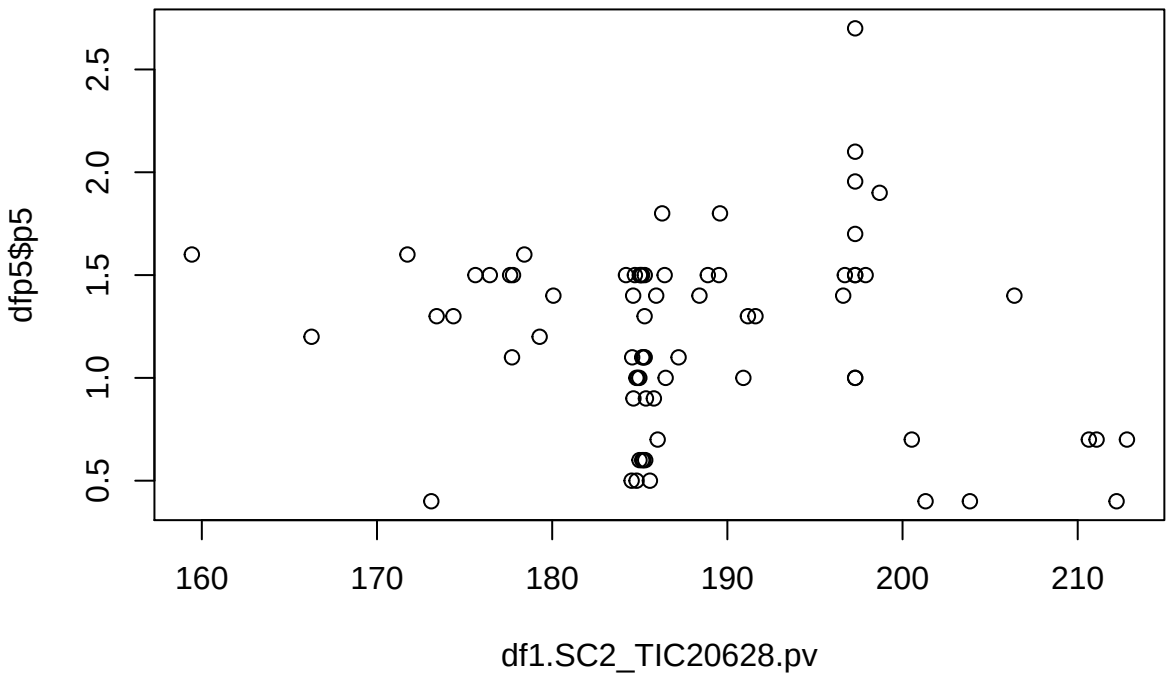

dfp5$p5
0.5
1.0
1.5
2.0
2.5
36
38
40
42
df1.SC2_PIC20065.pv

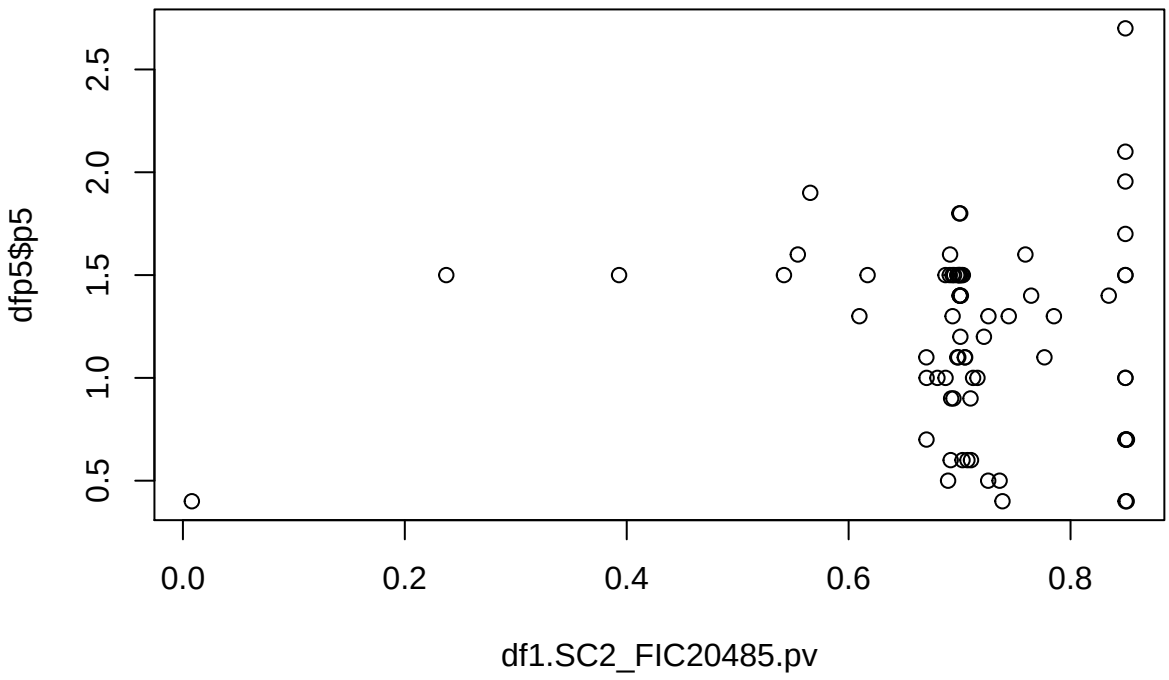

dfp5$p5
0.5
1.0
1.5
2.0
2.5
30
40
50
60
df1.SC2_TIC20760.pv

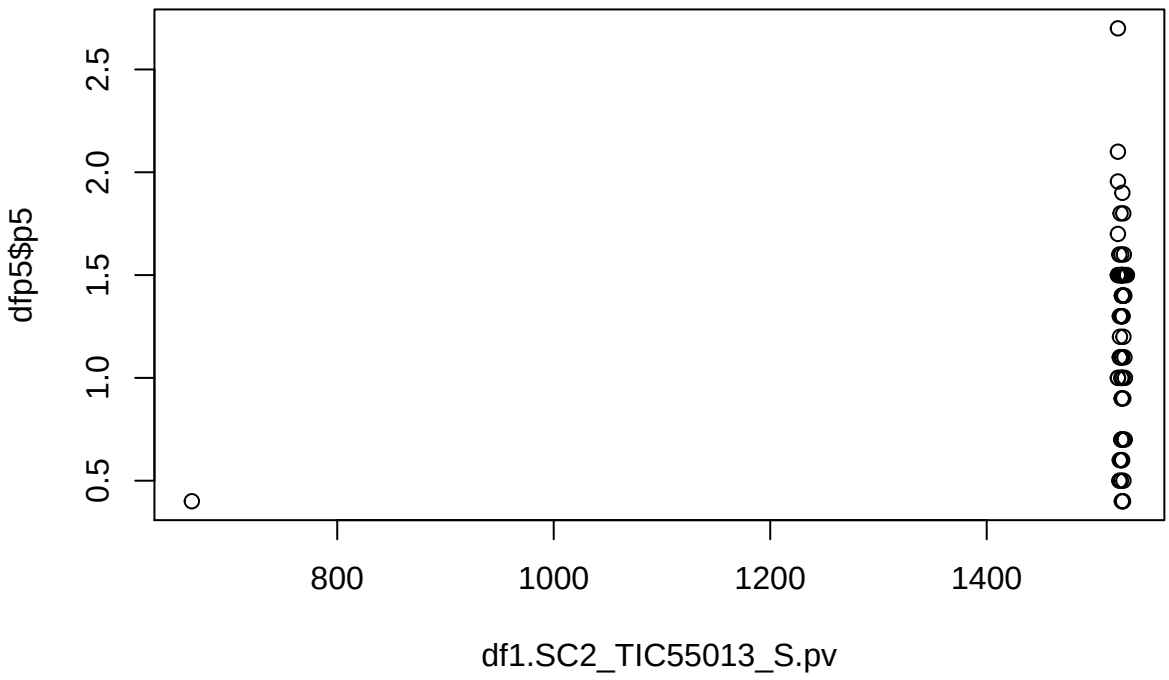

dfp5$p5
0.5
1.0
1.5
2.0
2.5
60
65
70
75
80
85
90
df1.SC2_TIC21178.pv

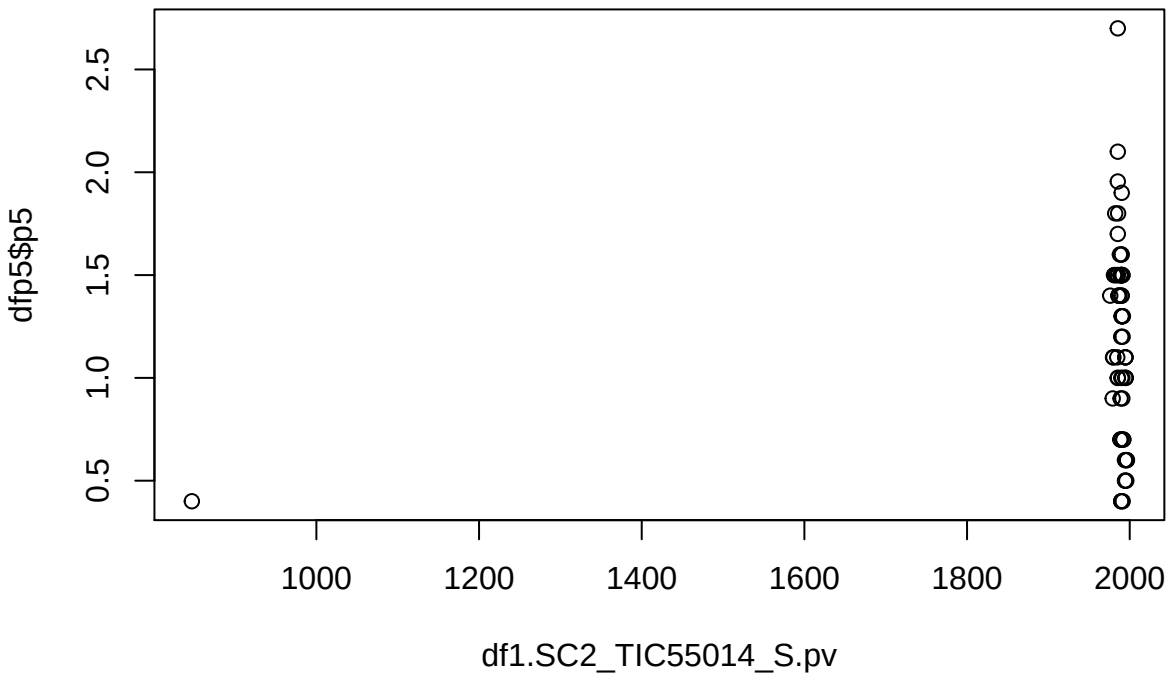

dfp5$p5
0.5
1.0
1.5
2.0
2.5
15
20
25
30
35
df1.SC2_FIC20476.pv

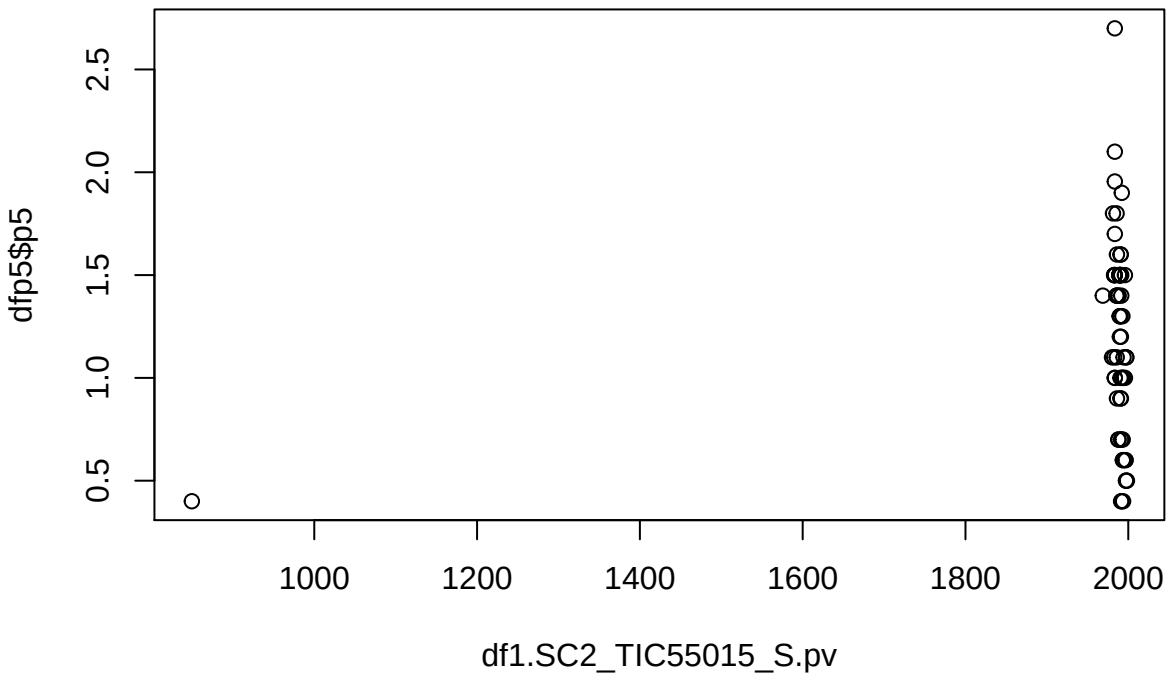

dfp5$p5
0.5
1.0
1.5
2.0
2.5
60
70
80
90
100
df1.SC2_PIC23750.pv

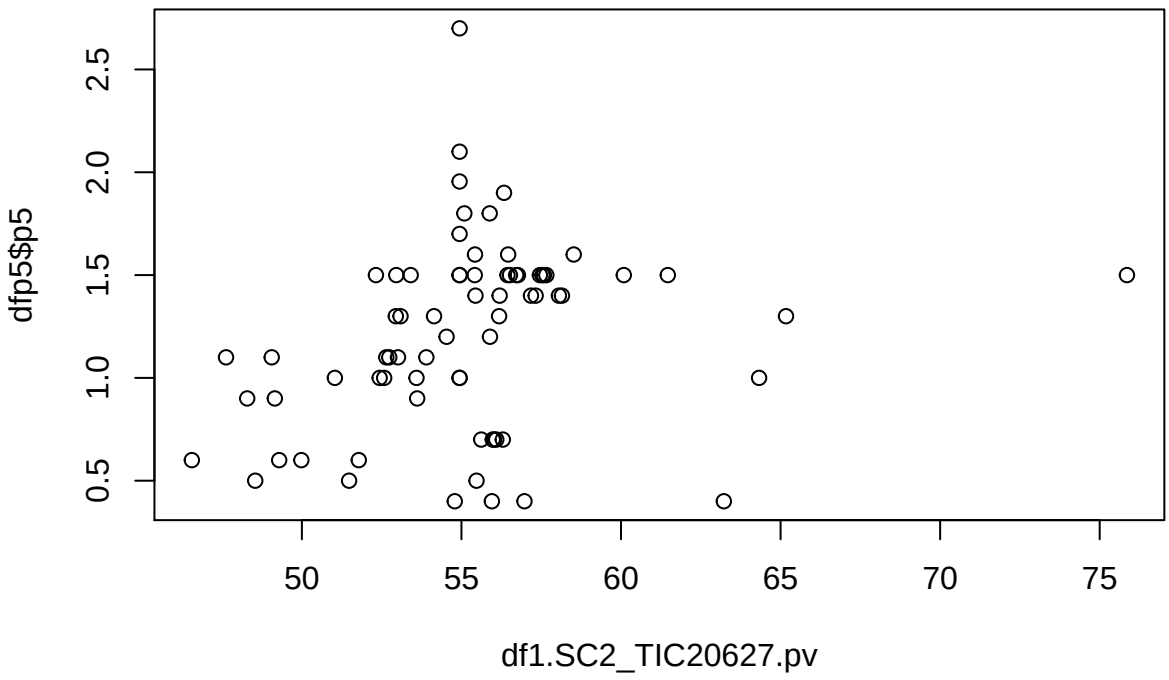

dfp5$p5
0.5
1.0
1.5
2.0
2.5
160
170
180
190
200
210
df1.SC2_TIC20628.pv

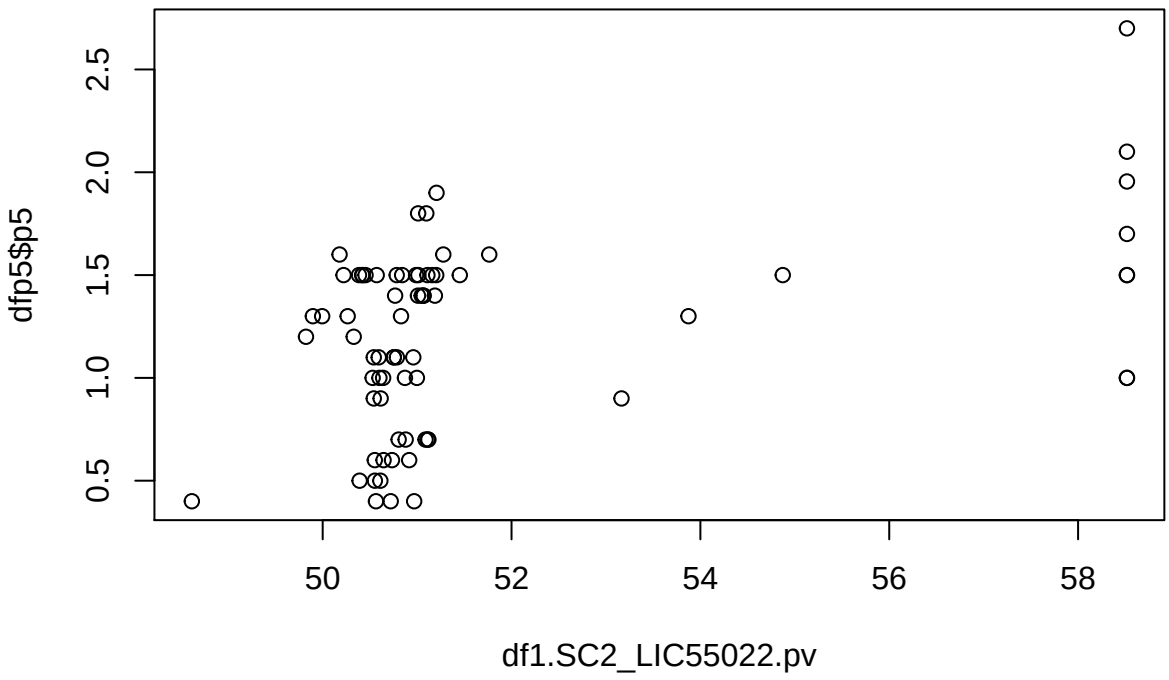

dfp5$p5
0.5
1.0
1.5
2.0
2.5
0.0
0.2
0.4
0.6
0.8
df1.SC2_FIC20485.pv

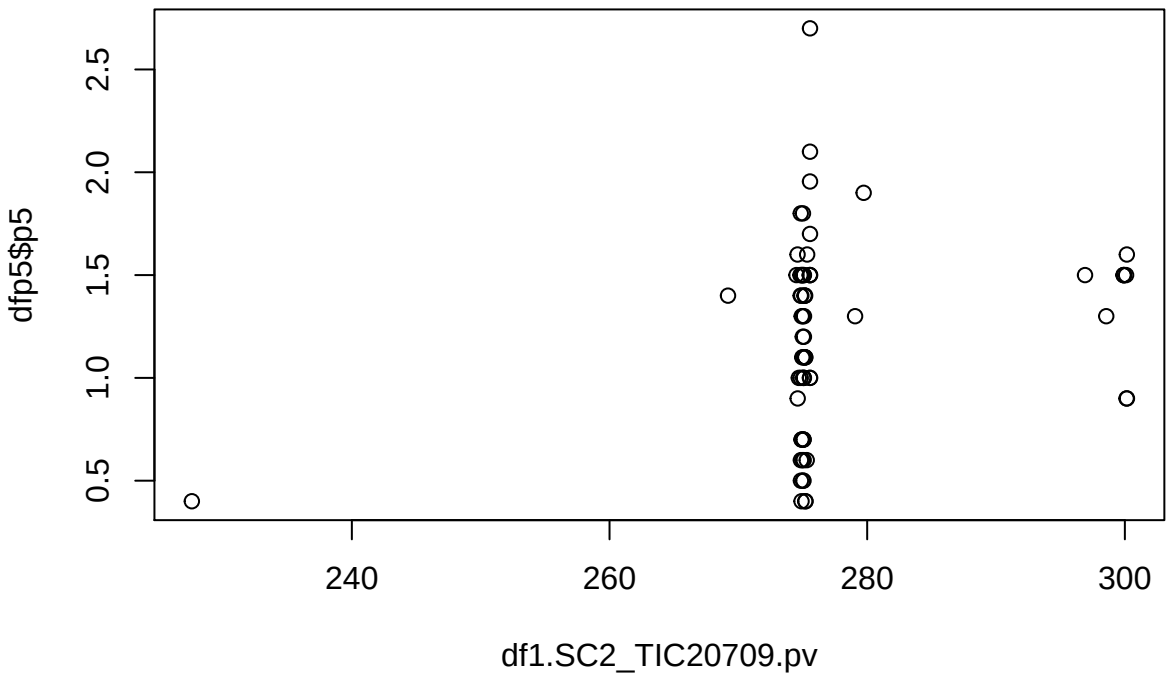

dfp5$p5
0.5
1.0
1.5
2.0
2.5
800
1000
1200
1400
df1.SC2_TIC55013_S.pv

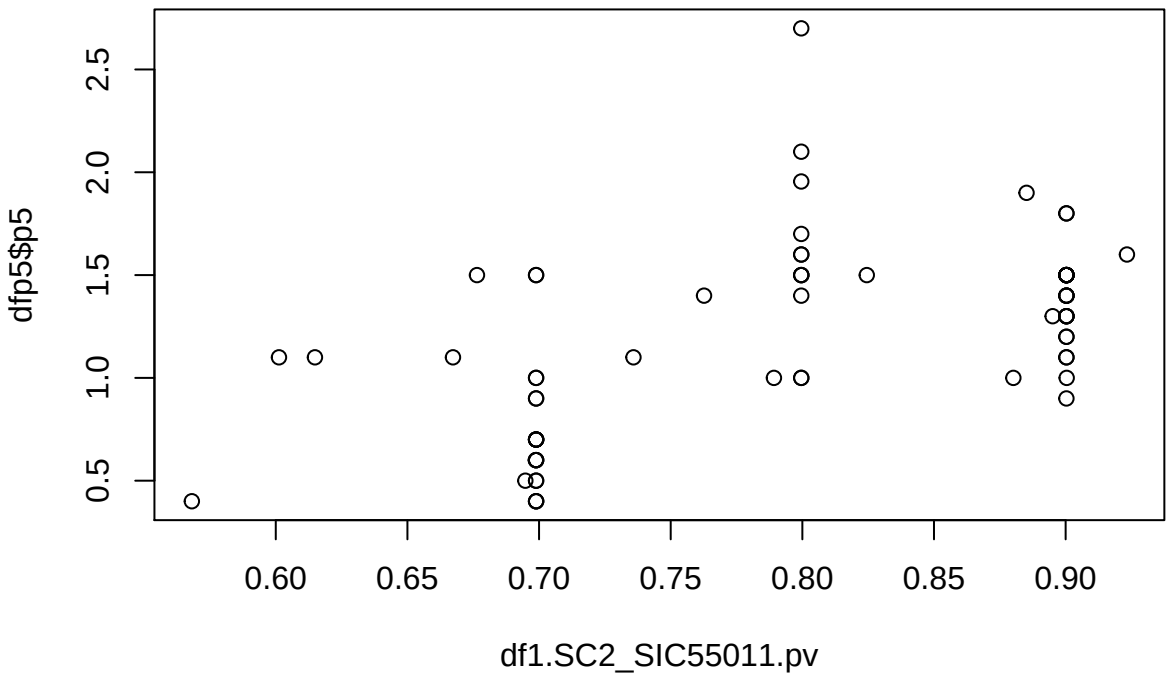

dfp5$p5
0.5
1.0
1.5
2.0
2.5
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55014_S.pv


dfp5$p5
0.5
1.0
1.5
2.0
2.5
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55015_S.pv

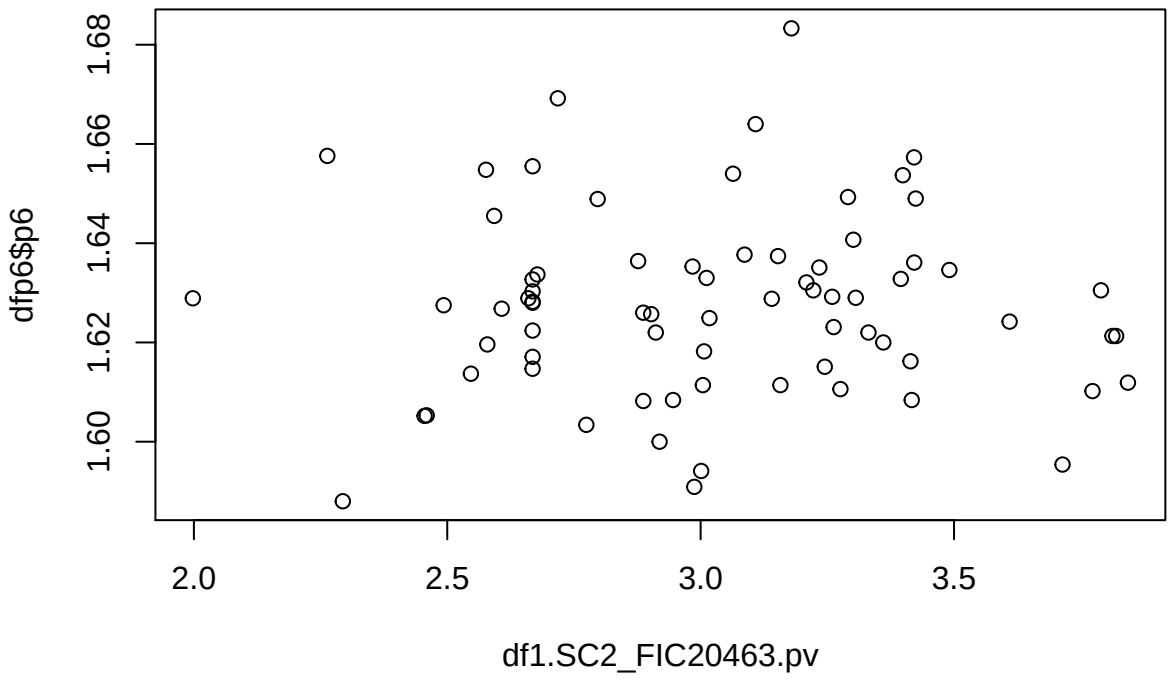

dfp5$p5
0.5
1.0
1.5
2.0
2.5
50
55
60
65
70
75
df1.SC2_TIC20627.pv

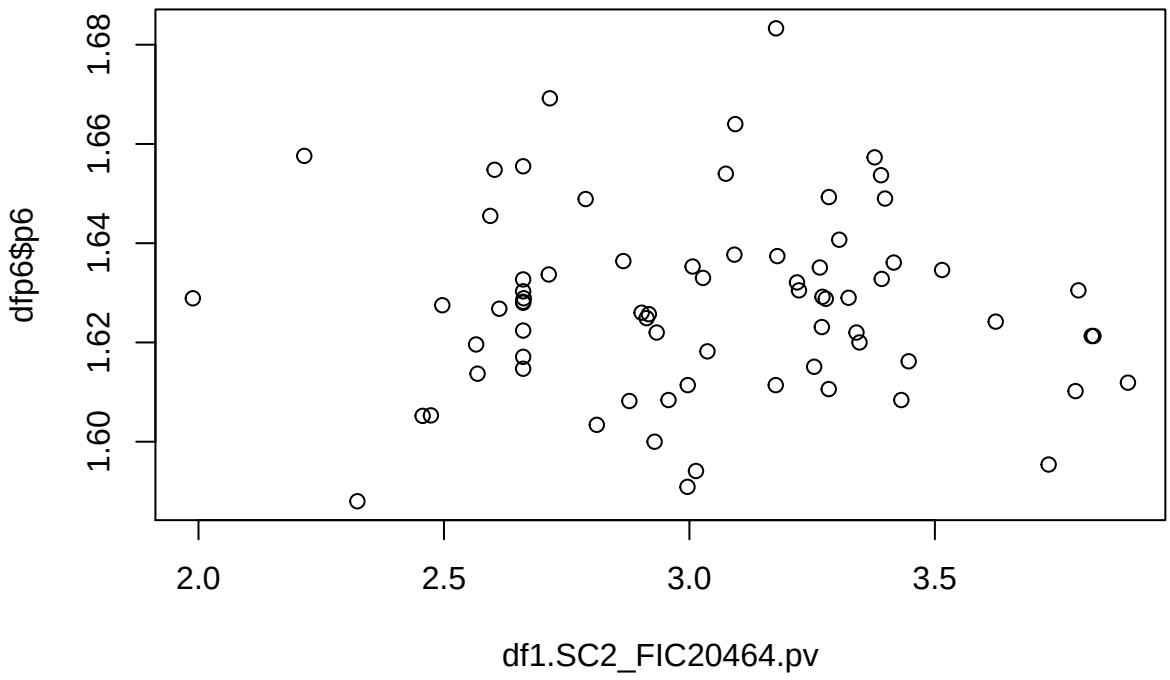

dfp5$p5
0.5
1.0
1.5
2.0
2.5
50
52
54
56
58
df1.SC2_LIC55022.pv

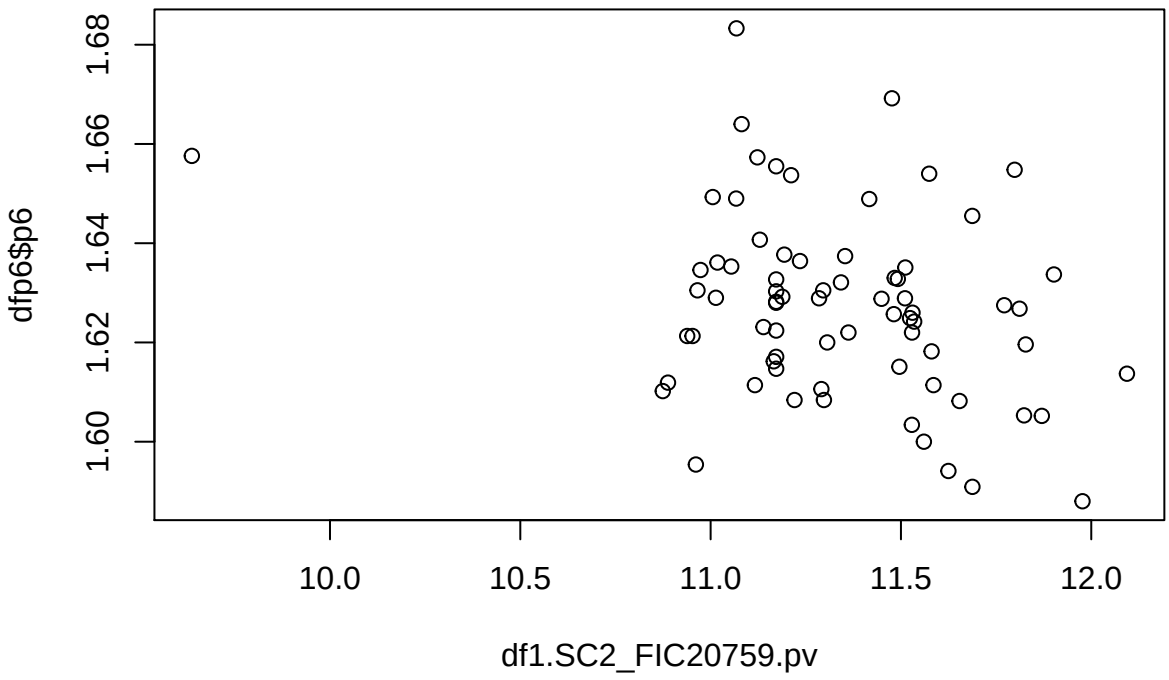

dfp5$p5
0.5
1.0
1.5
2.0
2.5
240
260
280
300
df1.SC2_TIC20709.pv

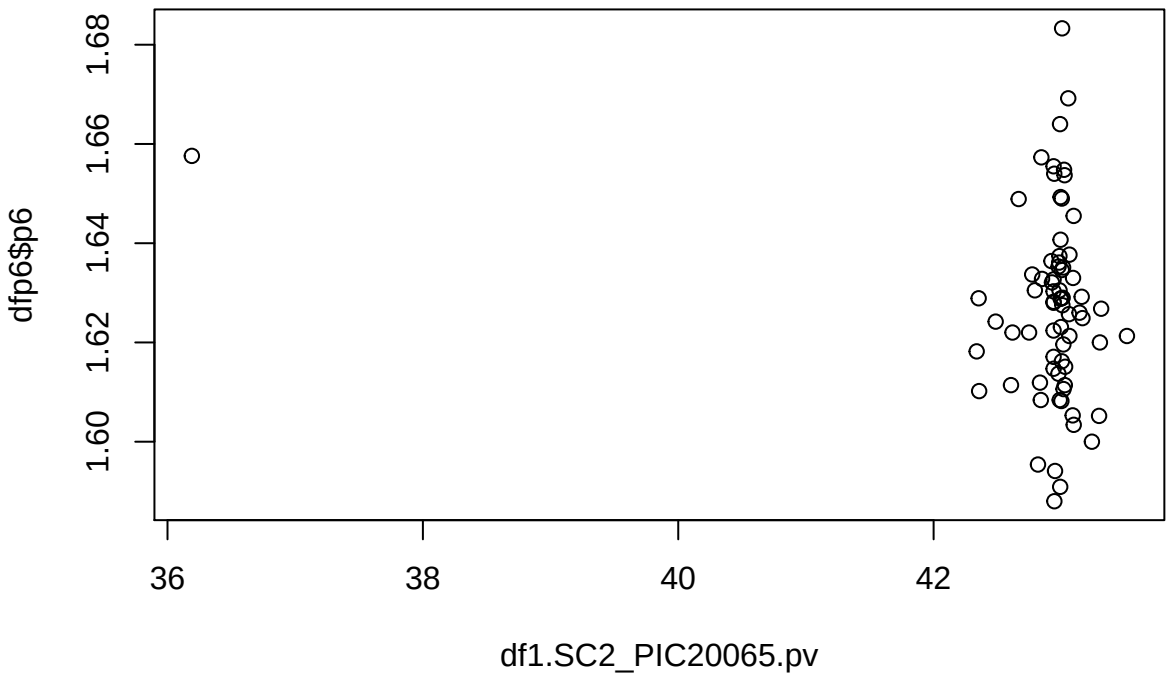


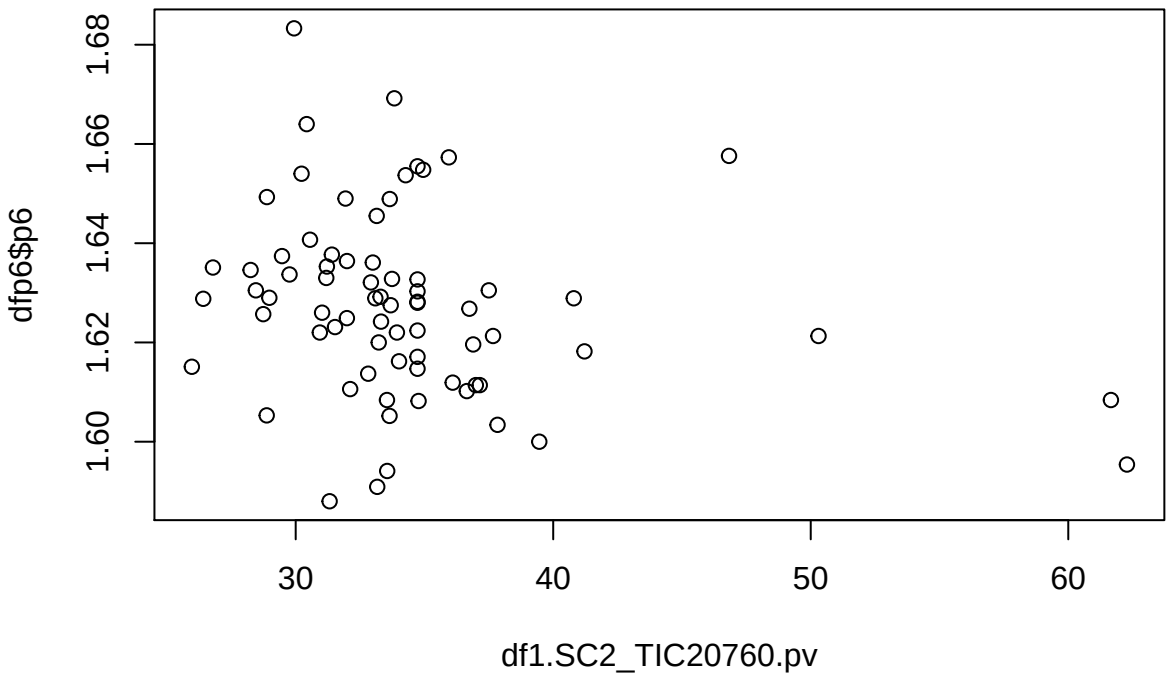

```
dfp6<-data.frame(p6,input)
plot(dfp6$p6~.,dfp6)
```

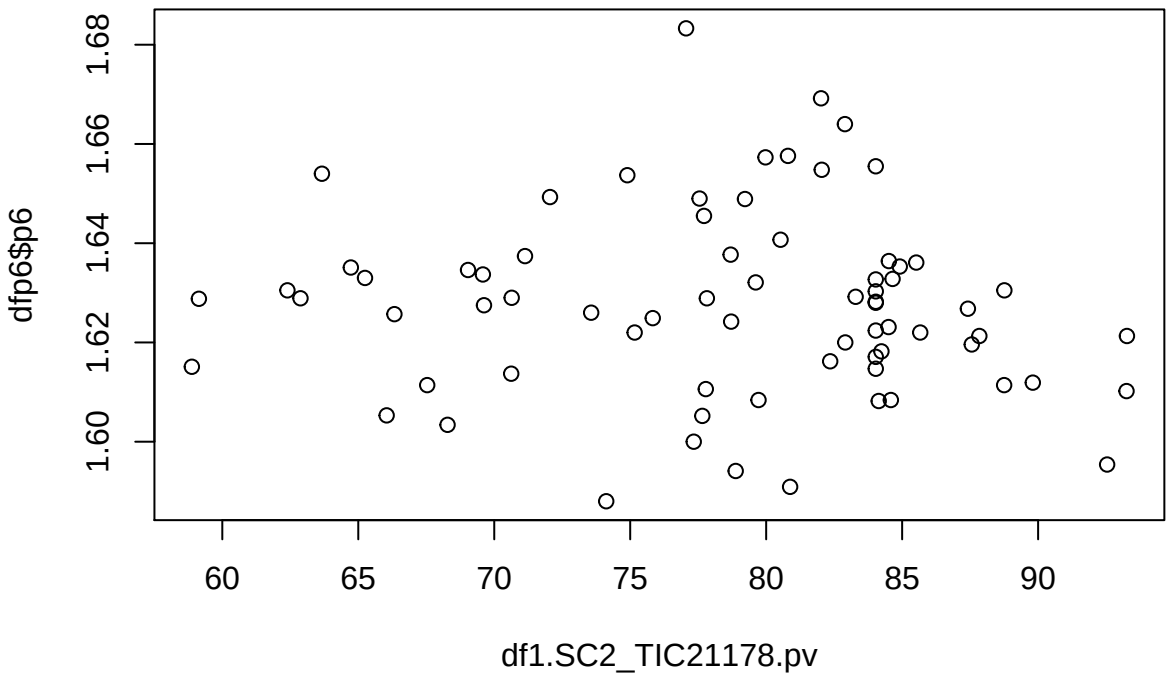

dfp6$p6
1.60
1.62
1.64
1.66
1.68
2.0
2.5
3.0
3.5
df1.SC2_FIC20463.pv

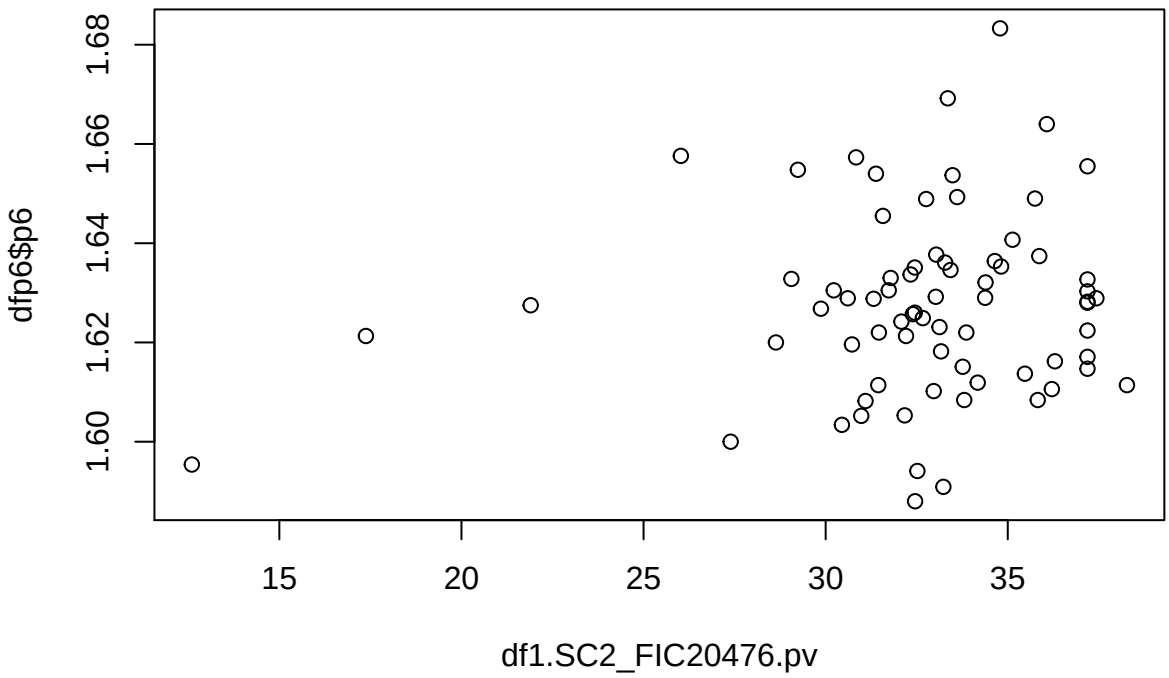

dfp6$p6
1.60
1.62
1.64
1.66
1.68
2.0
2.5
3.0
3.5
df1.SC2_FIC20464.pv

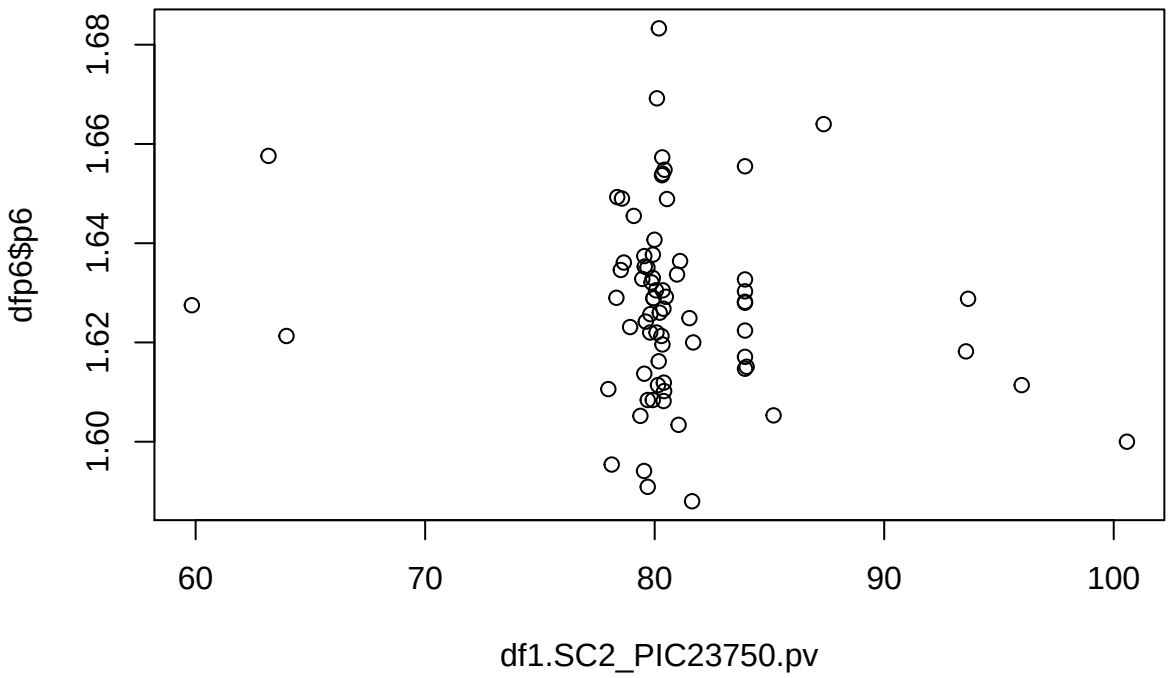

dfp6$p6
1.60
1.62
1.64
1.66
1.68
10.0
10.5
11.0
11.5
12.0
df1.SC2_FIC20759.pv

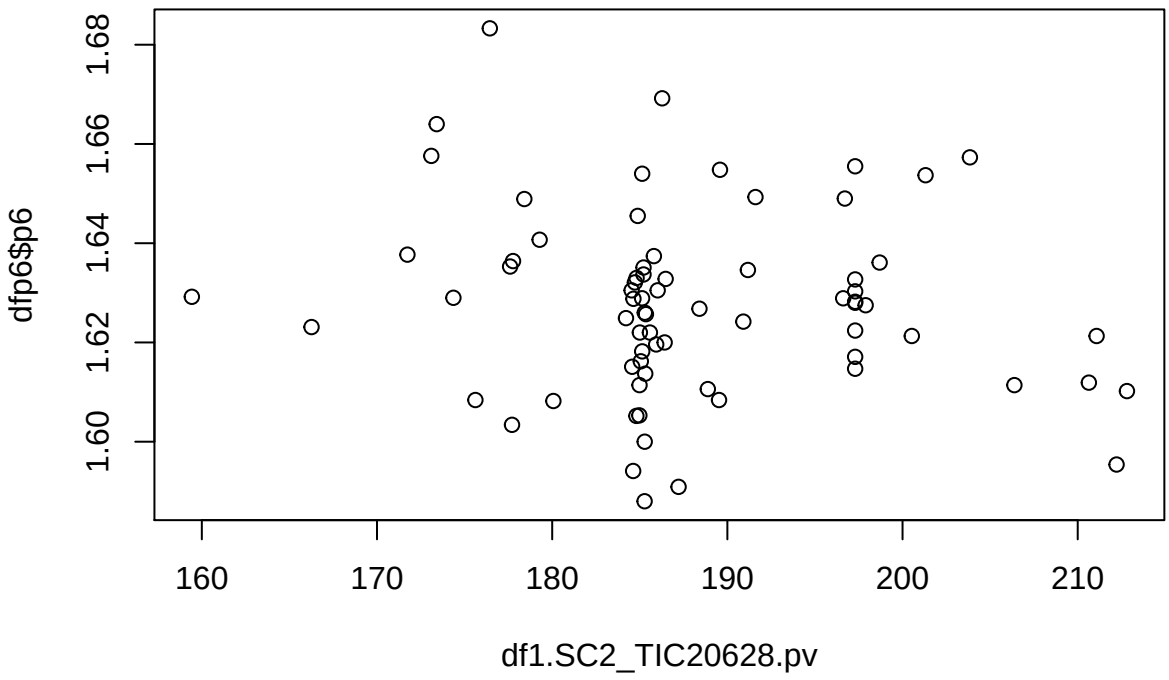

dfp6$p6
1.60
1.62
1.64
1.66
1.68
36
38
40
42
df1.SC2_PIC20065.pv

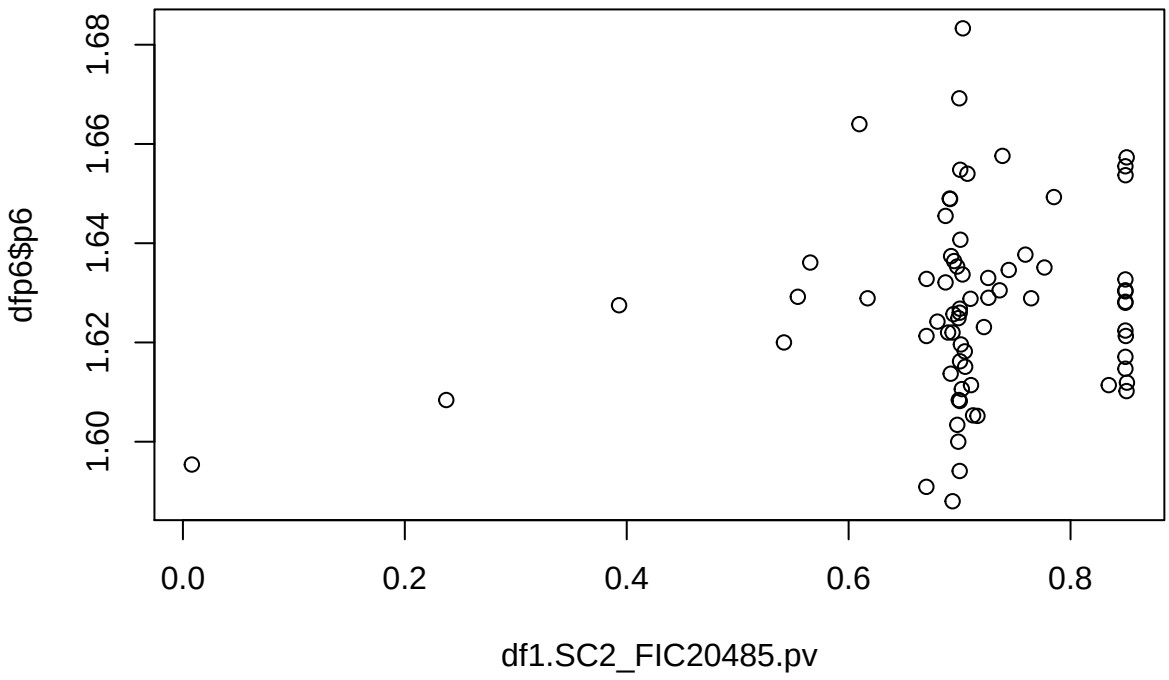

dfp6$p6
1.60
1.62
1.64
1.66
1.68
30
40
50
60
df1.SC2_TIC20760.pv

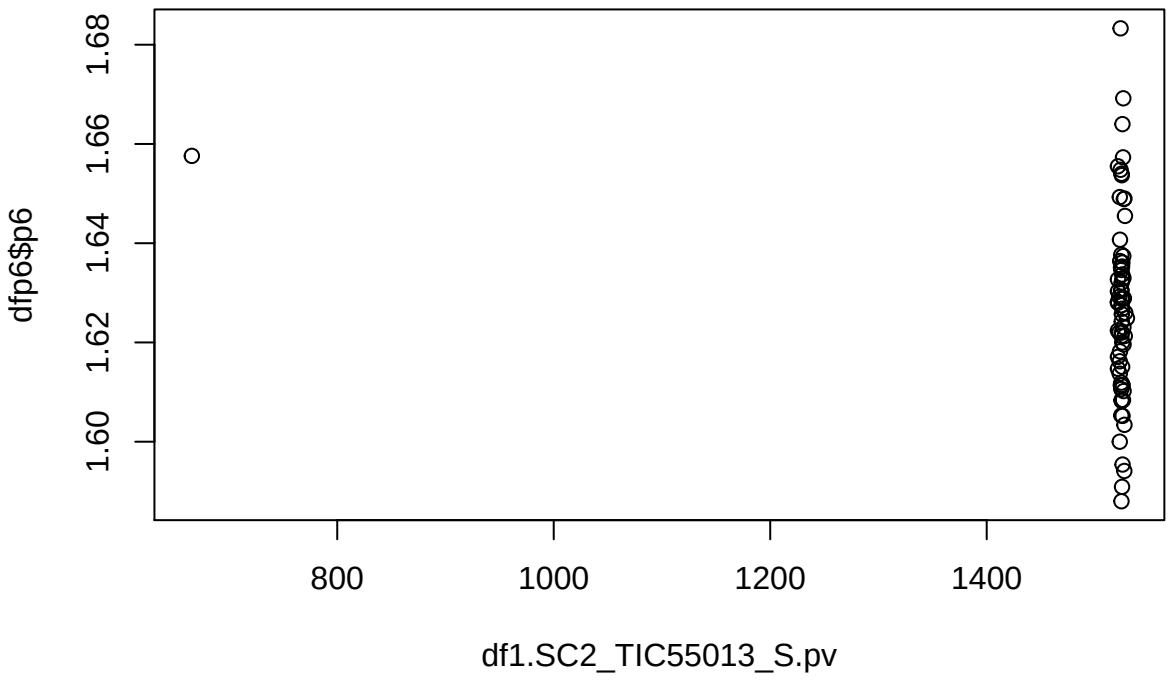

dfp6$p6
1.60
1.62
1.64
1.66
1.68
60
65
70
75
80
85
90
df1.SC2_TIC21178.pv

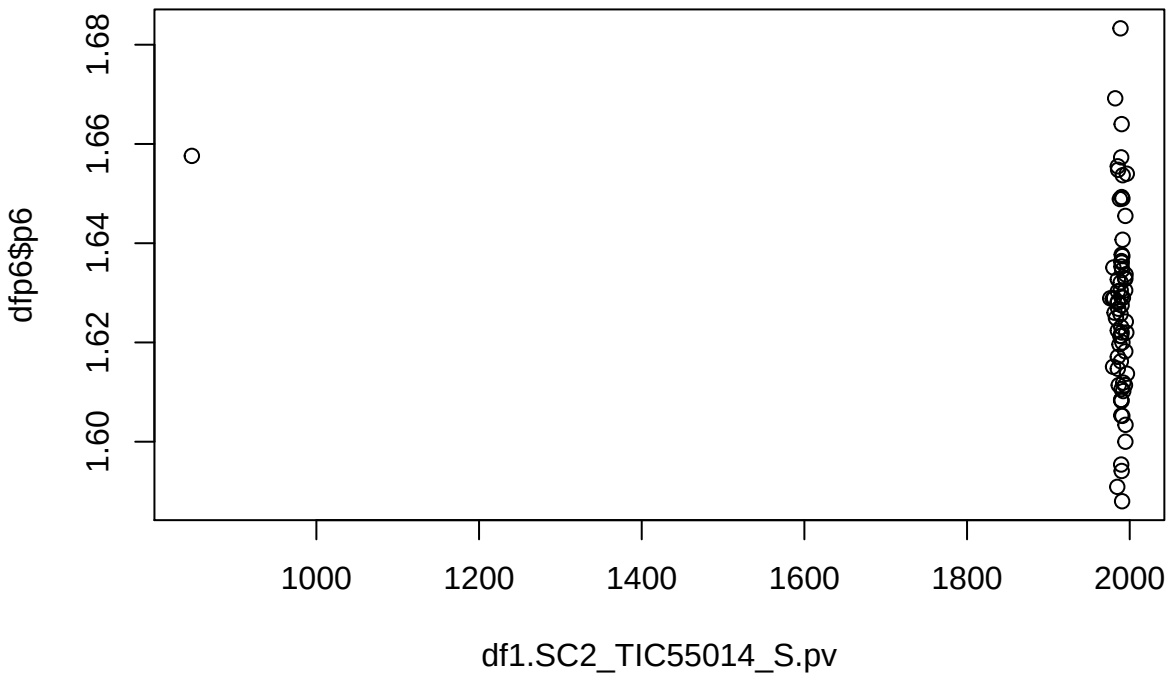

dfp6$p6
1.60
1.62
1.64
1.66
1.68
15
20
25
30
35
df1.SC2_FIC20476.pv

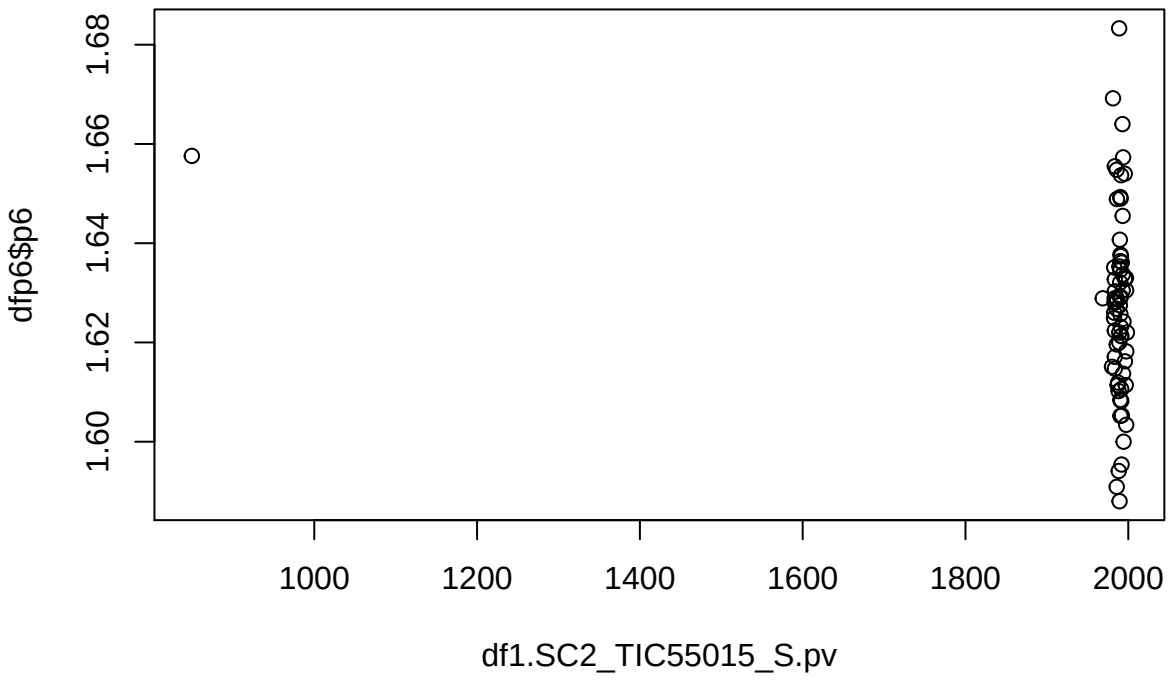

dfp6$p6
1.60
1.62
1.64
1.66
1.68
60
70
80
90
100
df1.SC2_PIC23750.pv

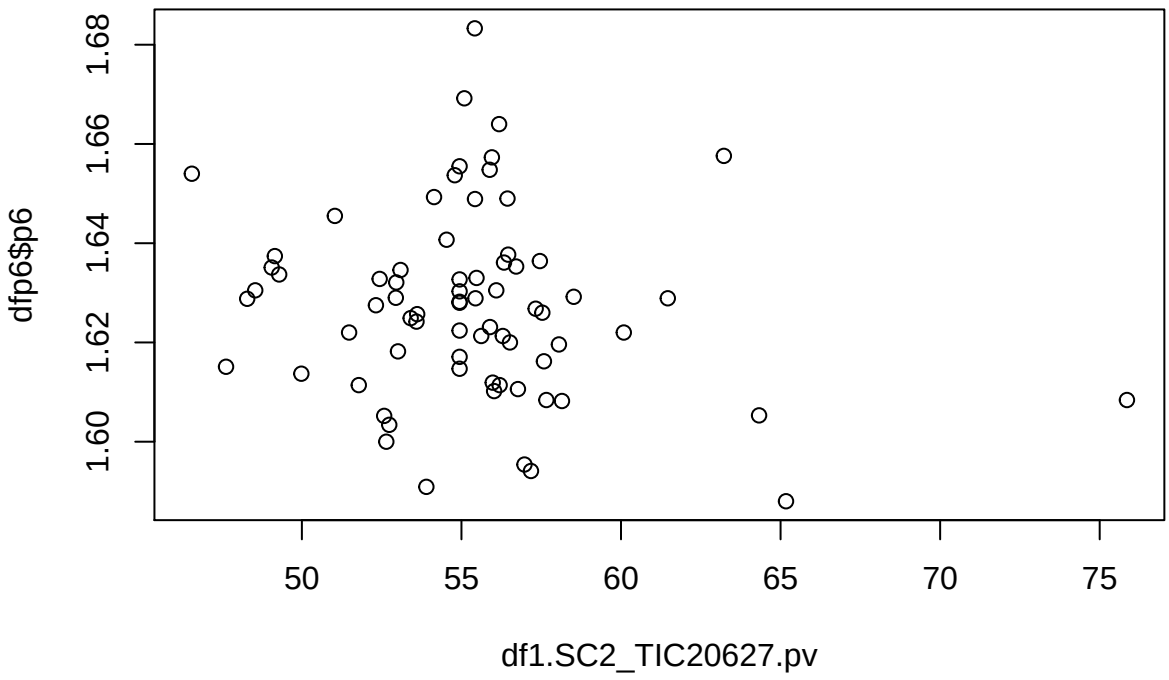

dfp6$p6
1.60
1.62
1.64
1.66
1.68
160
170
180
190
200
210
df1.SC2_TIC20628.pv

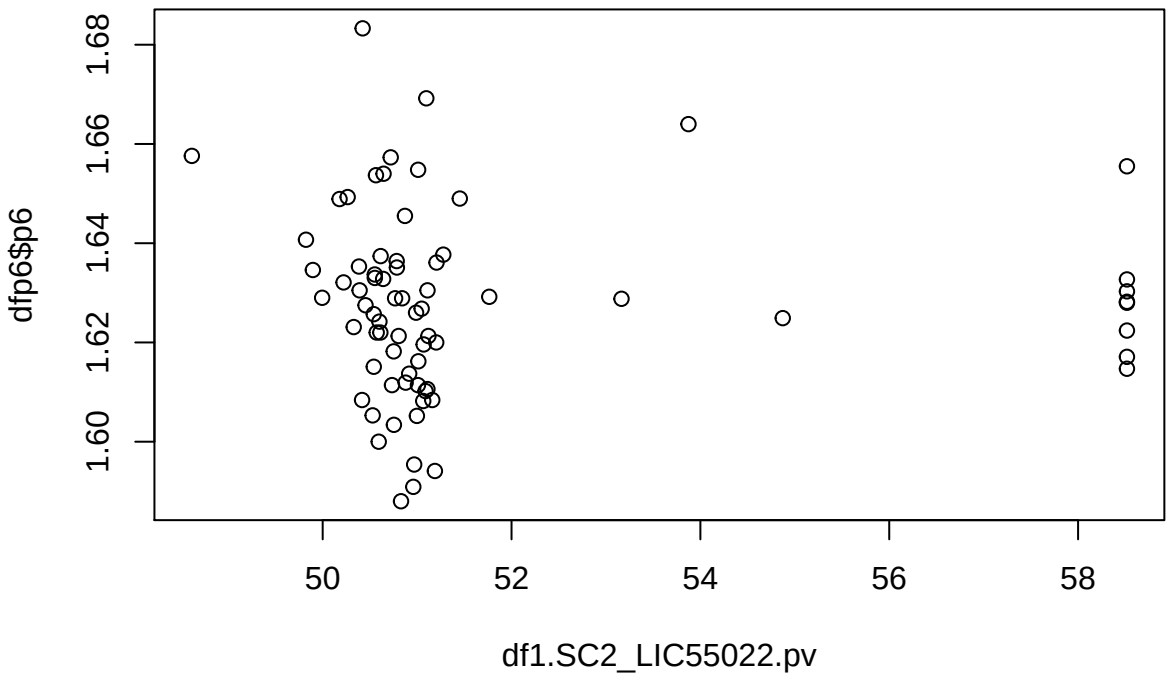

dfp6$p6
1.60
1.62
1.64
1.66
1.68
0.0
0.2
0.4
0.6
0.8
df1.SC2_FIC20485.pv

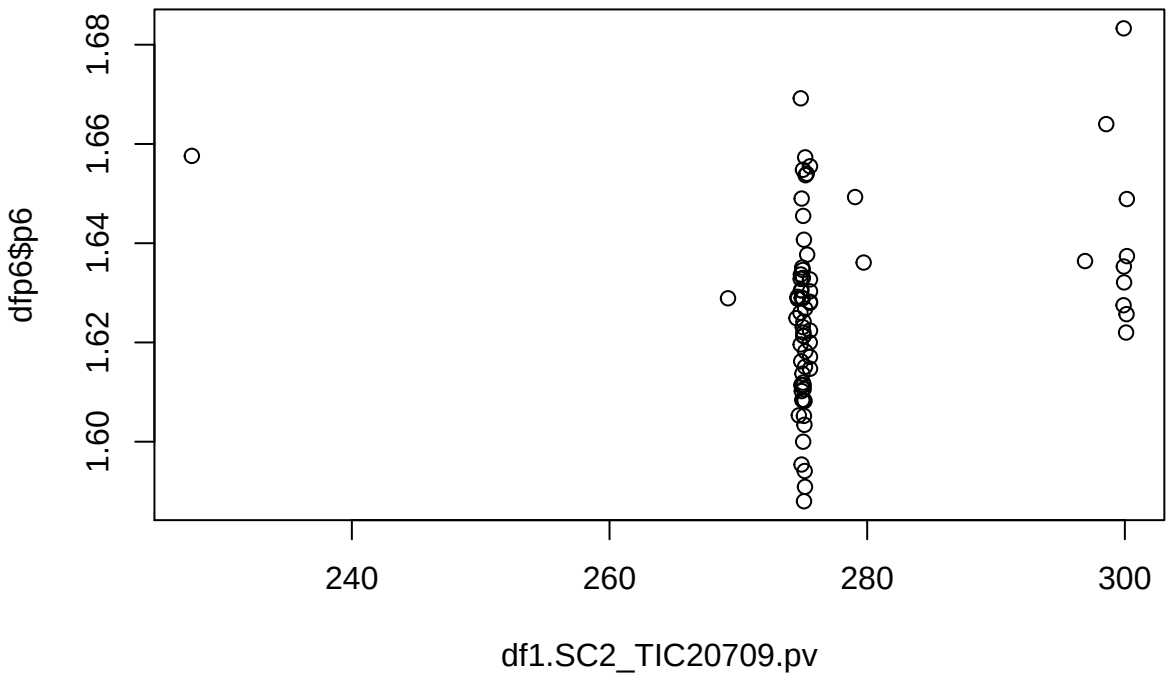

dfp6$p6
1.60
1.62
1.64
1.66
1.68
800
1000
1200
1400
df1.SC2_TIC55013_S.pv

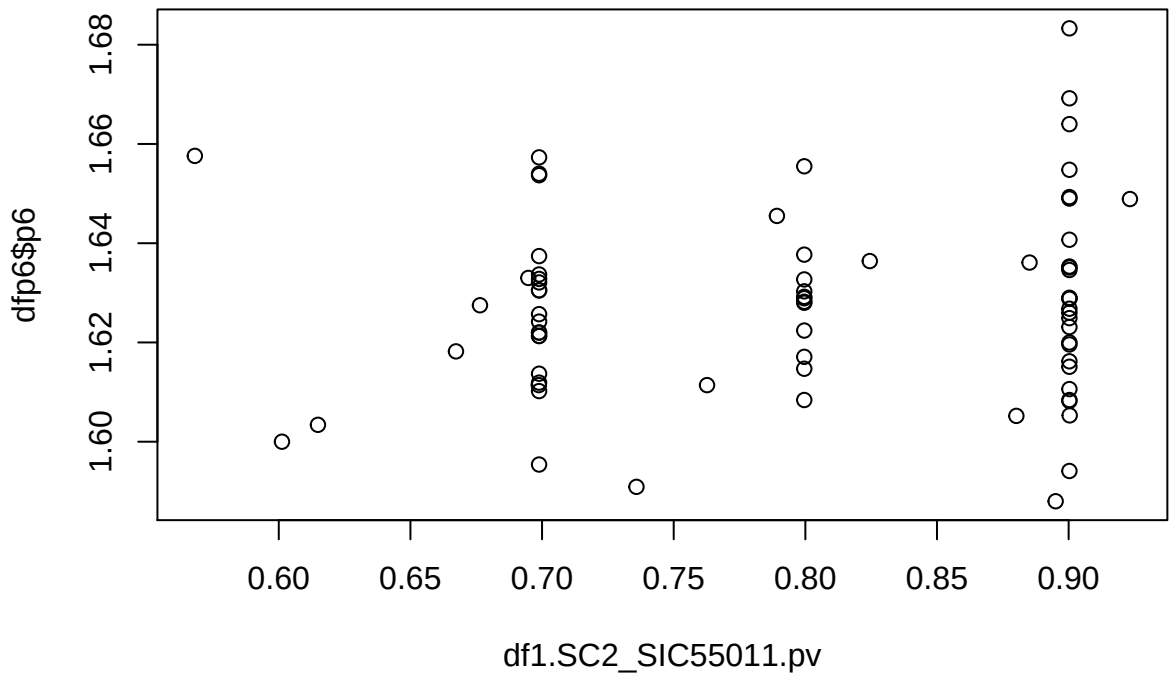

dfp6$p6
1.60
1.62
1.64
1.66
1.68
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55014_S.pv


dfp6$p6
1.60
1.62
1.64
1.66
1.68
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55015_S.pv

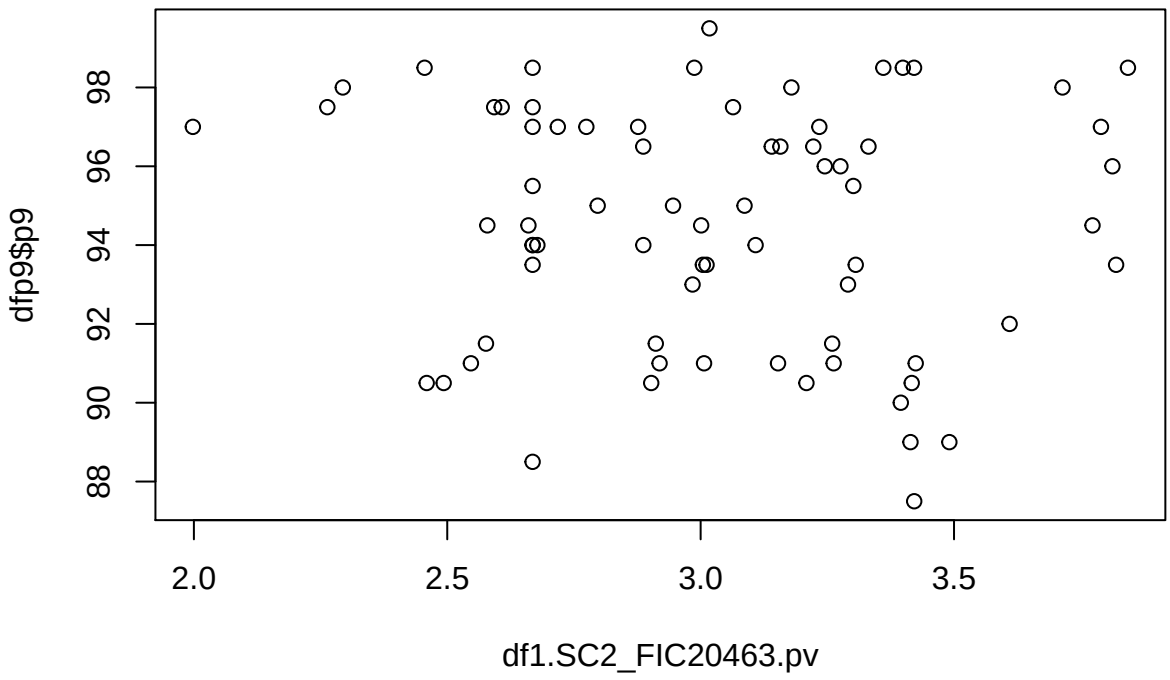

dfp6$p6
1.60
1.62
1.64
1.66
1.68
50
55
60
65
70
75
df1.SC2_TIC20627.pv

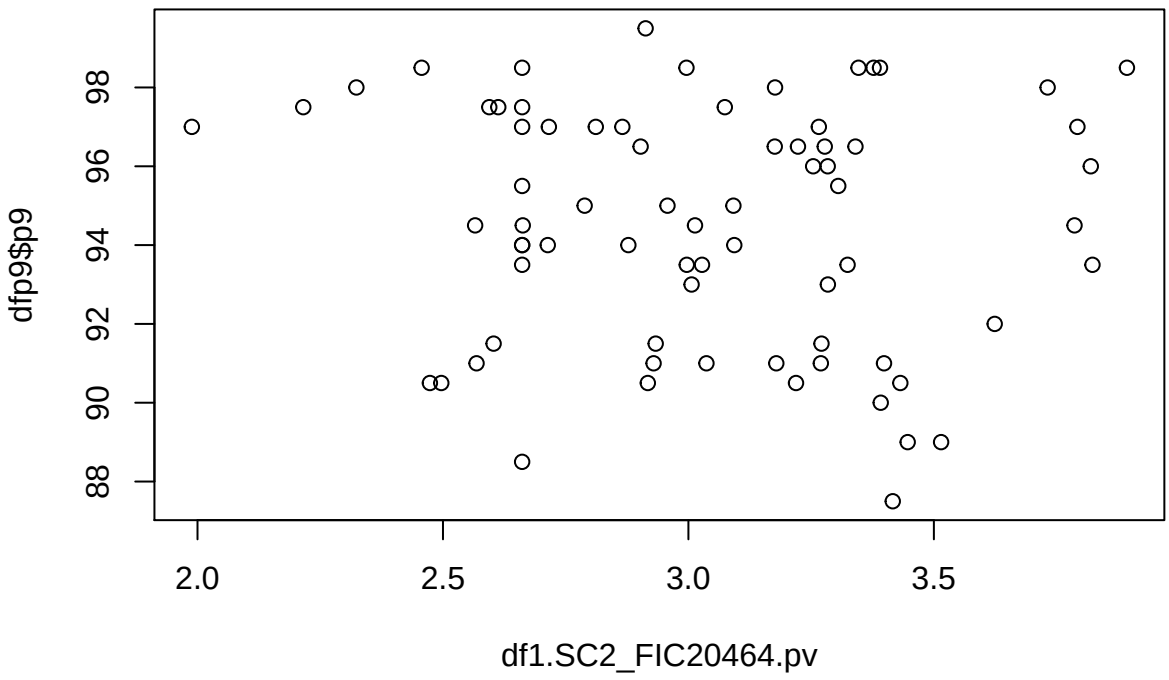

dfp6$p6
1.60
1.62
1.64
1.66
1.68
50
52
54
56
58
df1.SC2_LIC55022.pv

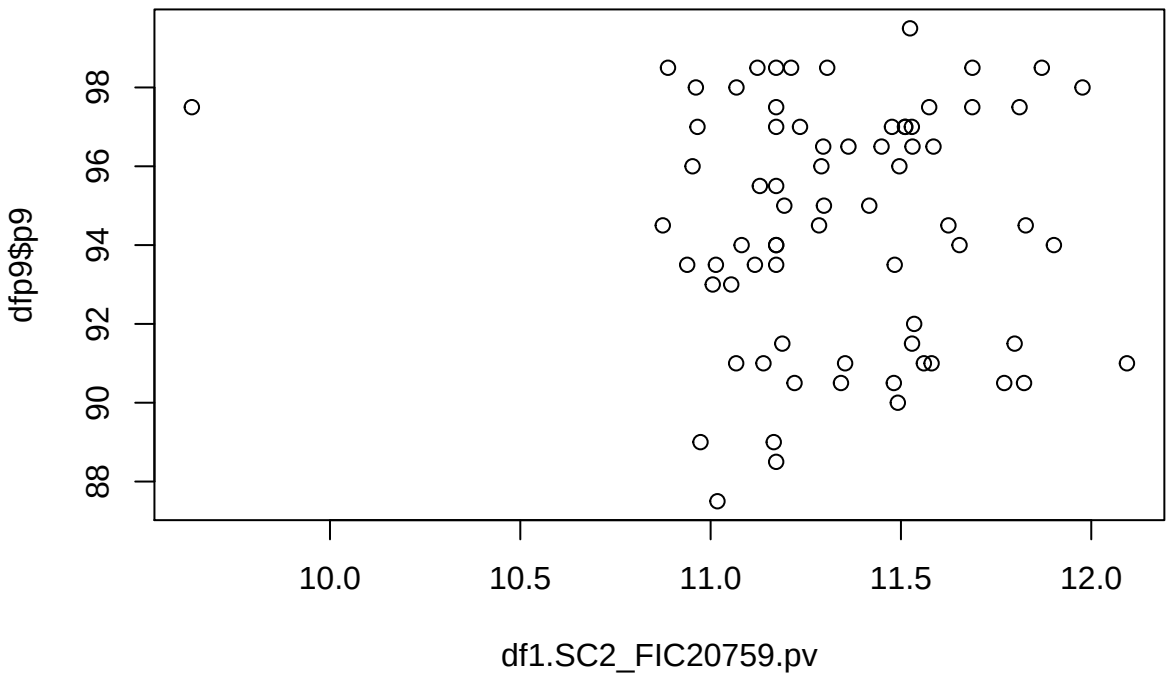

dfp6$p6
1.60
1.62
1.64
1.66
1.68
240
260
280
300
df1.SC2_TIC20709.pv

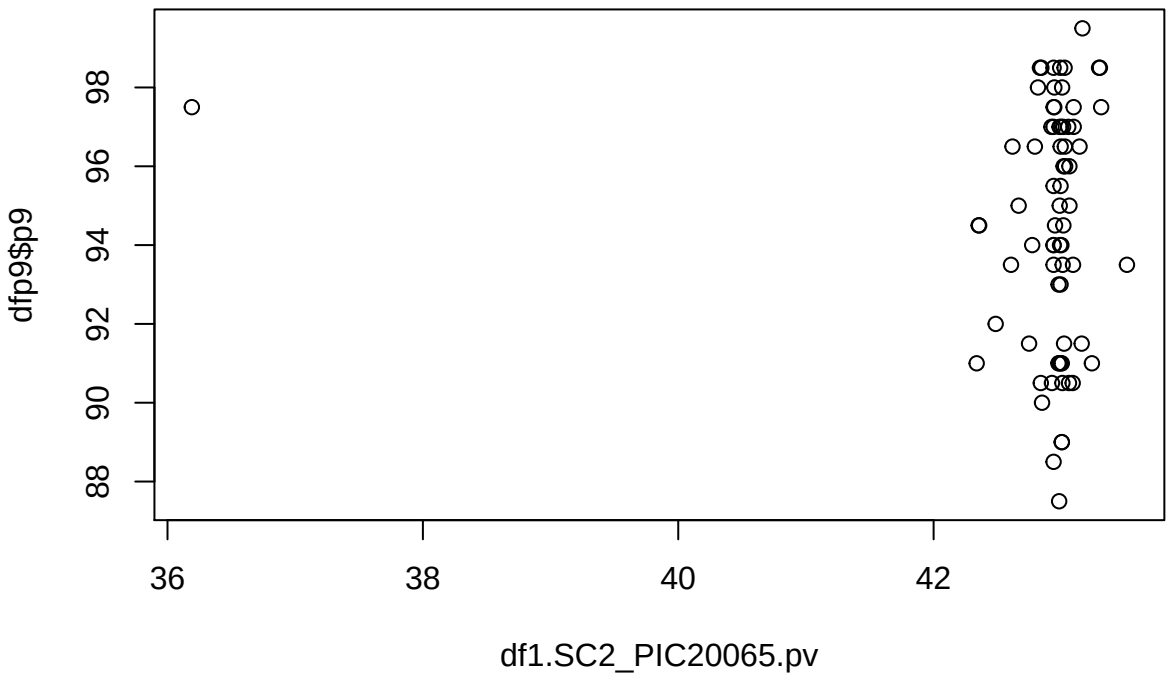


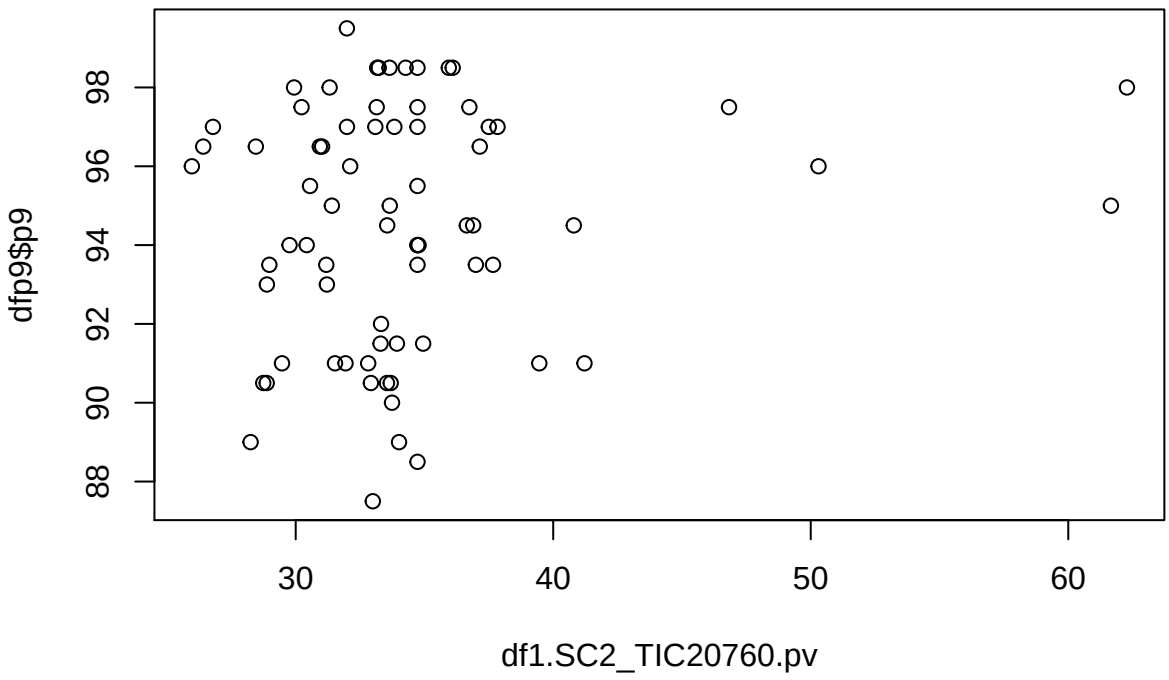

```
dfp9<-data.frame(p9,input)
plot(dfp9$p9~.,dfp9)
```

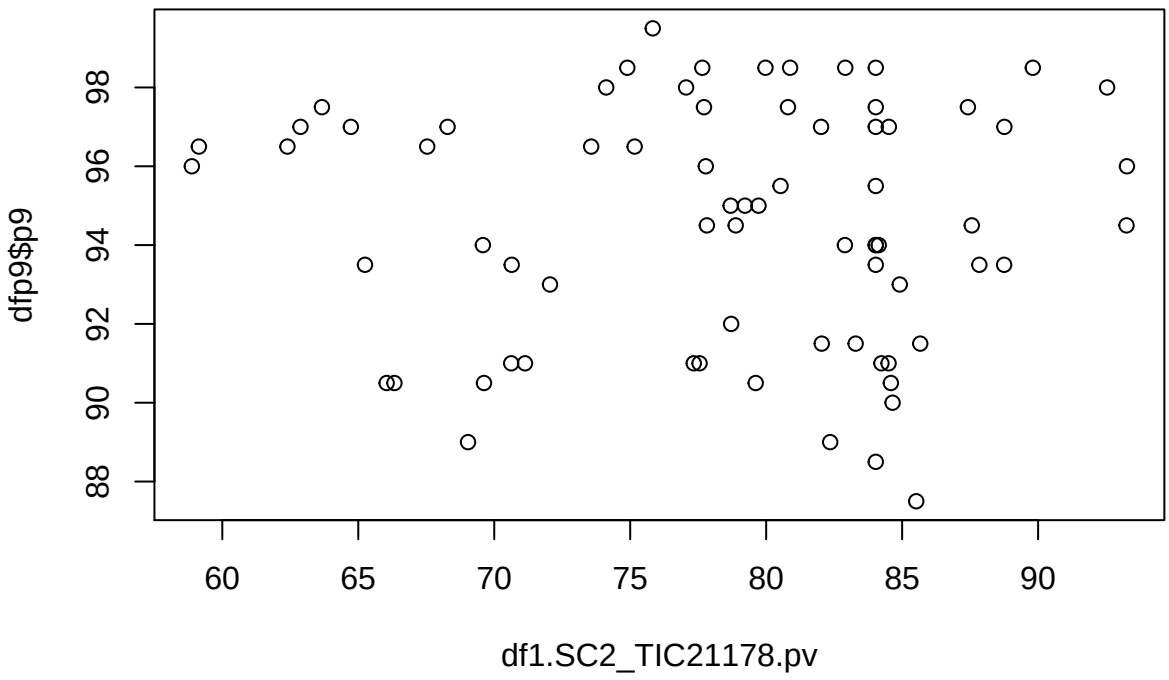

dfp9$p9
88
90
92
94
96
98
2.0
2.5
3.0
3.5
df1.SC2_FIC20463.pv

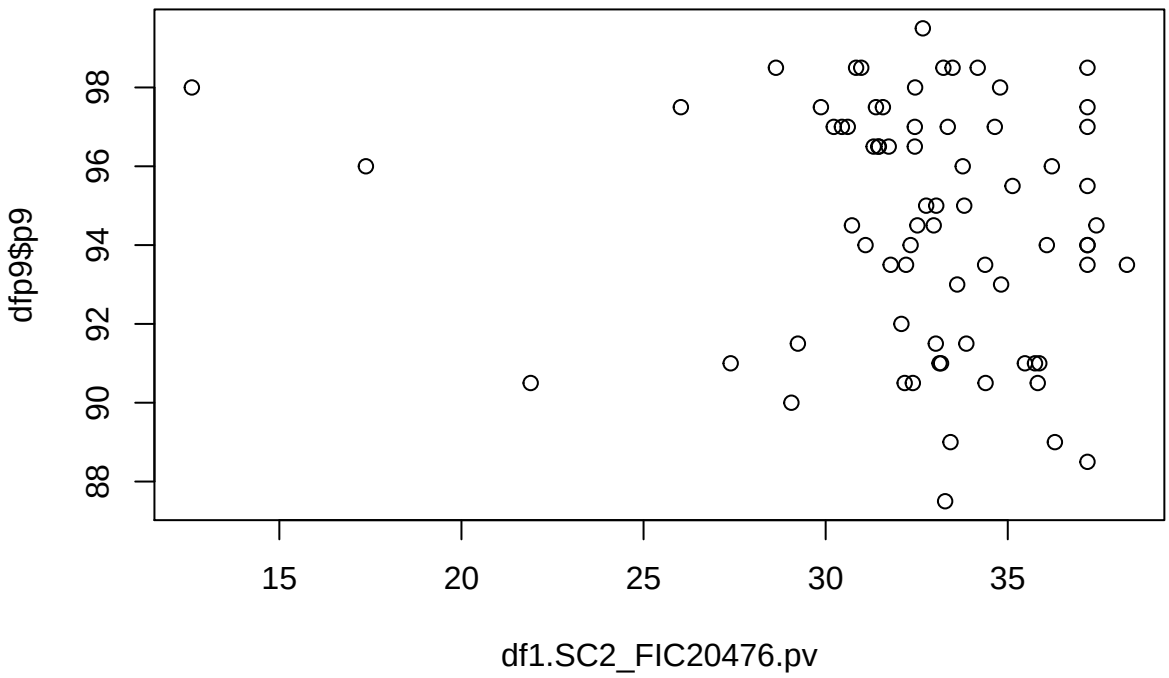

dfp9$p9
88
90
92
94
96
98
2.0
2.5
3.0
3.5
df1.SC2_FIC20464.pv

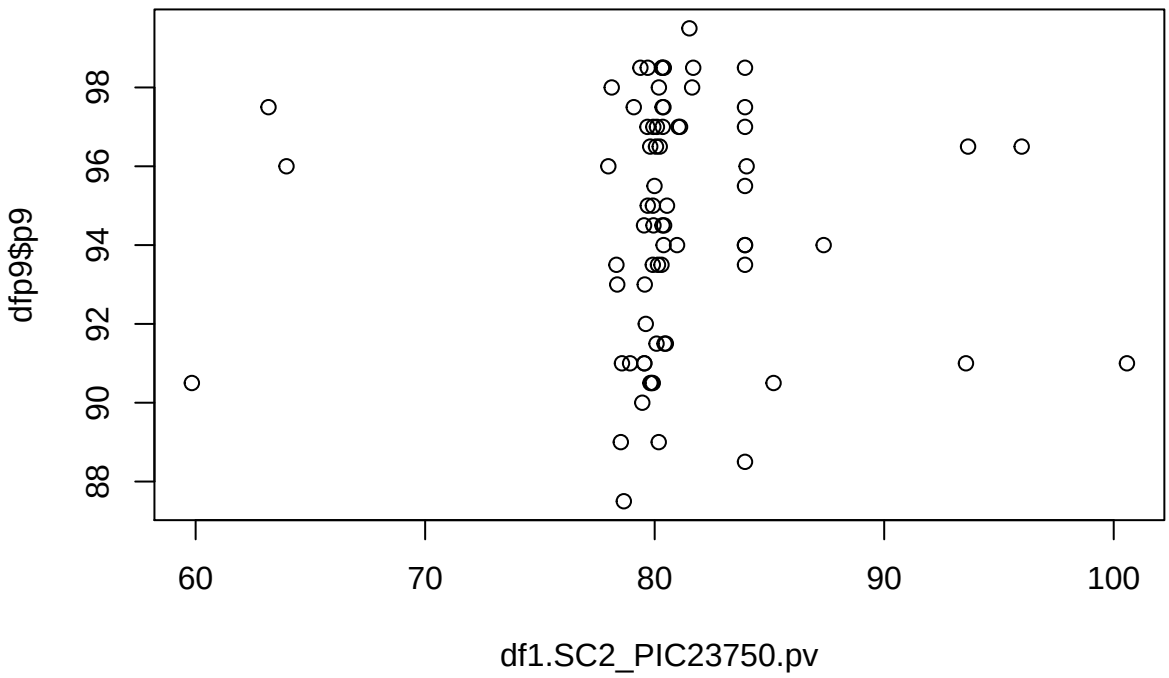

dfp9$p9
88
90
92
94
96
98
10.0
10.5
11.0
11.5
12.0
df1.SC2_FIC20759.pv

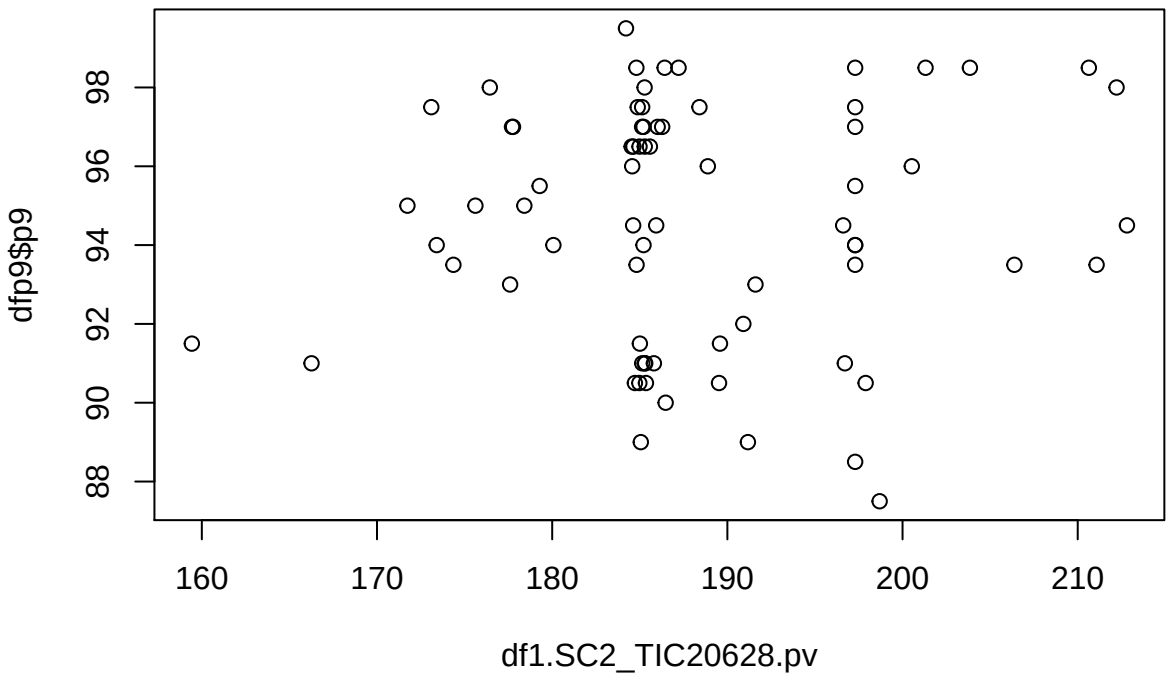

dfp9$p9
88
90
92
94
96
98
36
38
40
42
df1.SC2_PIC20065.pv

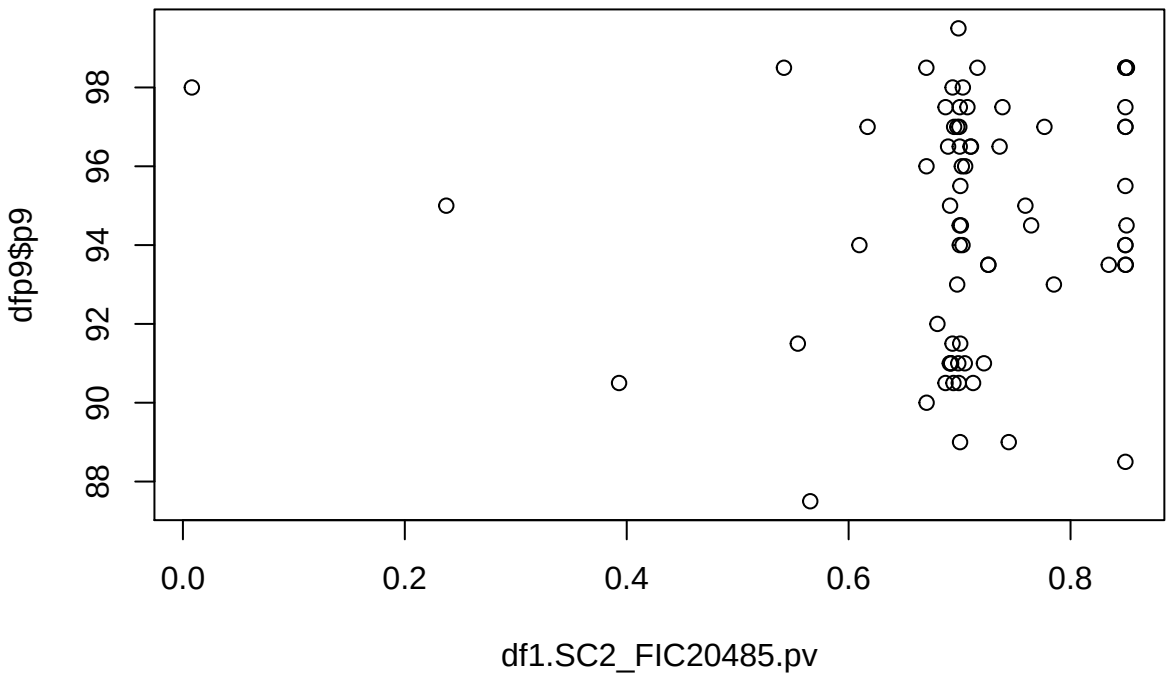

dfp9$p9
88
90
92
94
96
98
30
40
50
60
df1.SC2_TIC20760.pv

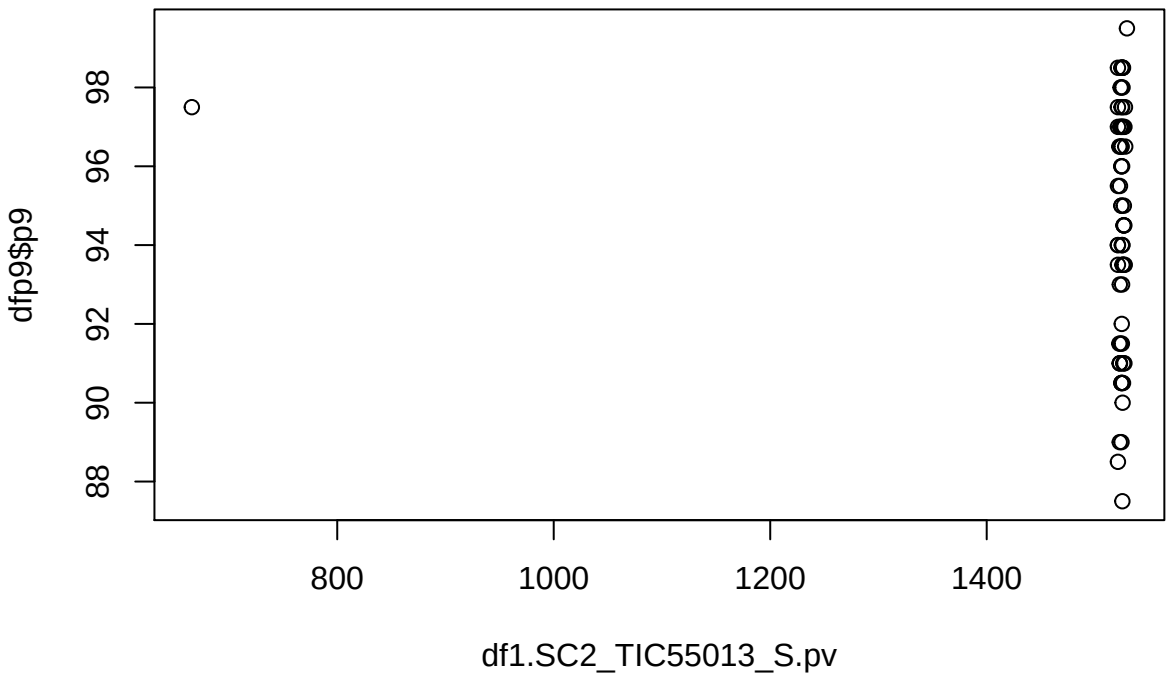

dfp9$p9
88
90
92
94
96
98
60
65
70
75
80
85
90
df1.SC2_TIC21178.pv

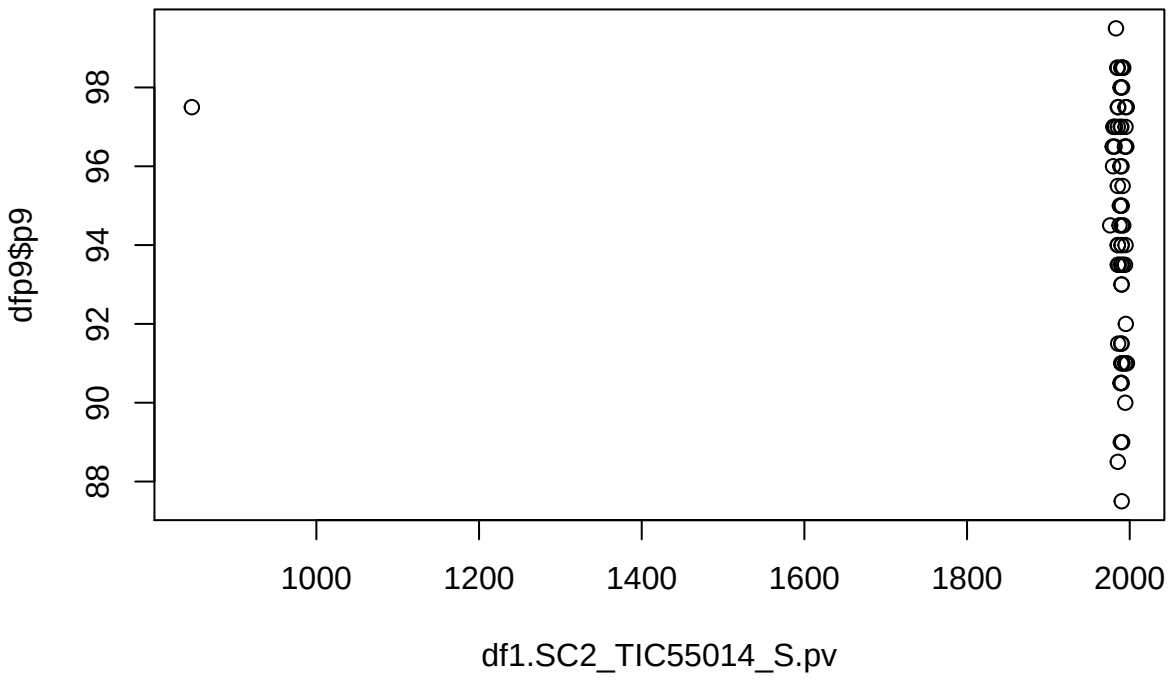

dfp9$p9
88
90
92
94
96
98
15
20
25
30
35
df1.SC2_FIC20476.pv

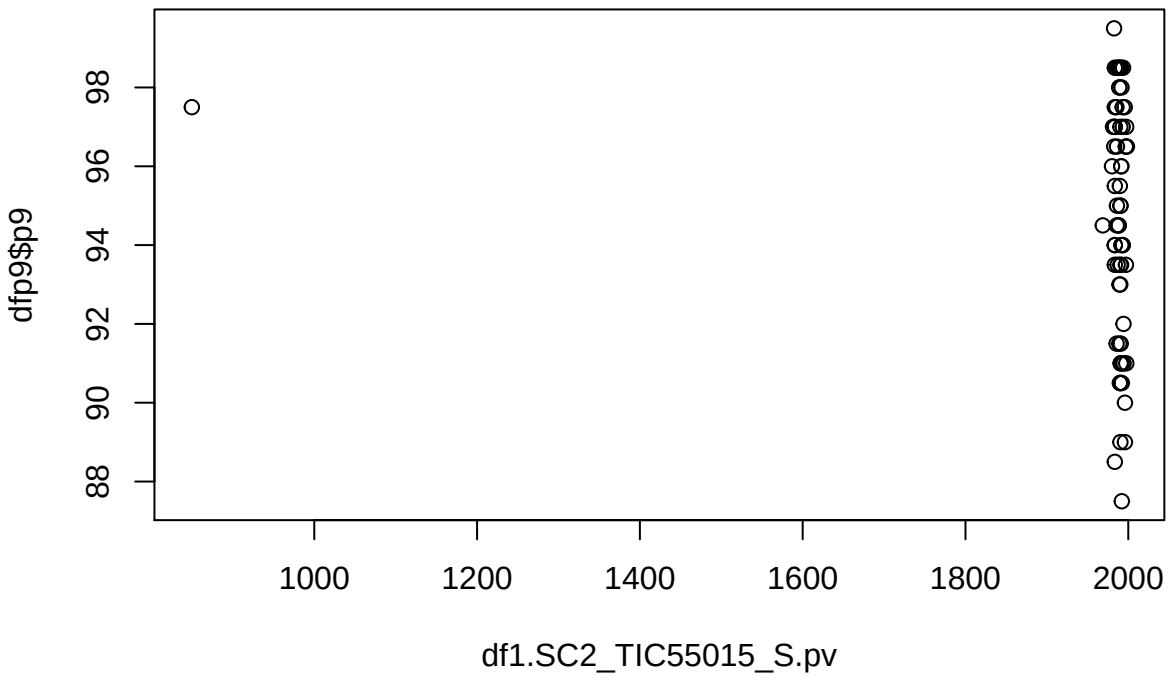

dfp9$p9
88
90
92
94
96
98
60
70
80
90
100
df1.SC2_PIC23750.pv

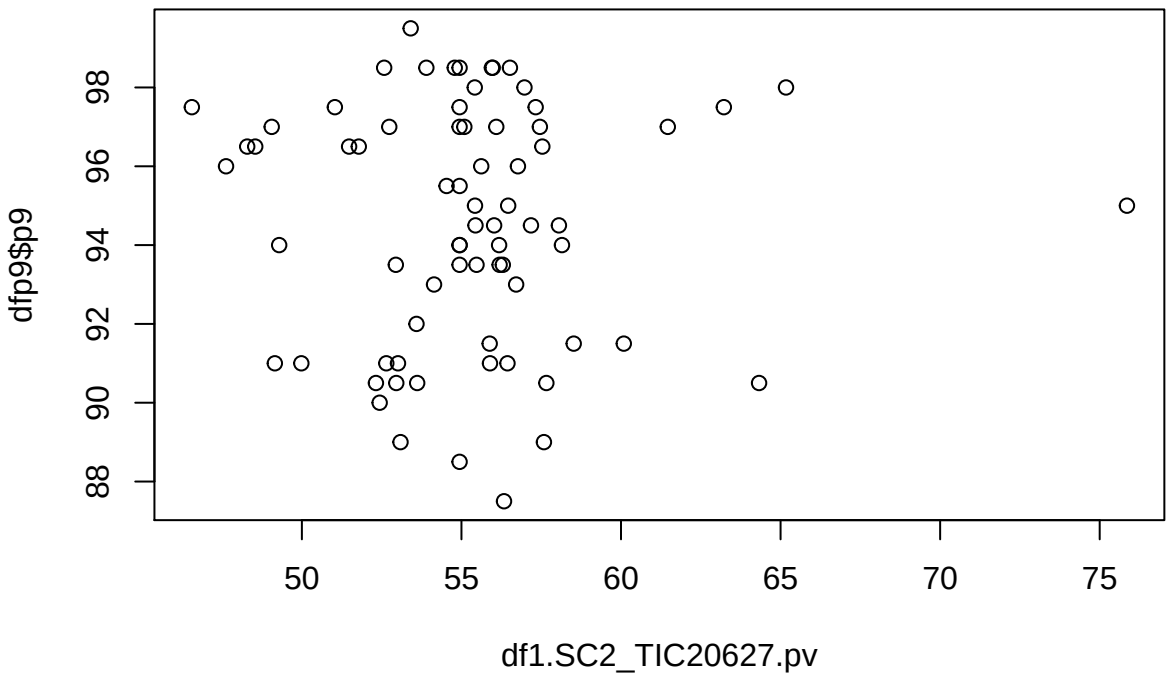

dfp9$p9
88
90
92
94
96
98
160
170
180
190
200
210
df1.SC2_TIC20628.pv

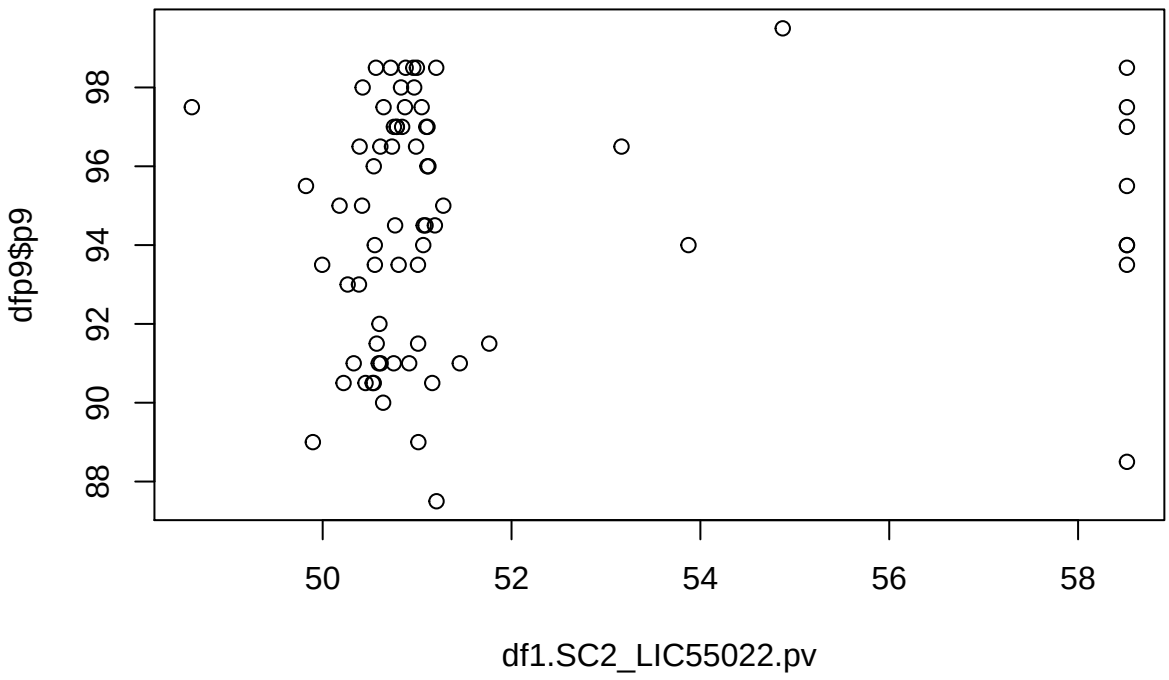

dfp9$p9
88
90
92
94
96
98
0.0
0.2
0.4
0.6
0.8
df1.SC2_FIC20485.pv

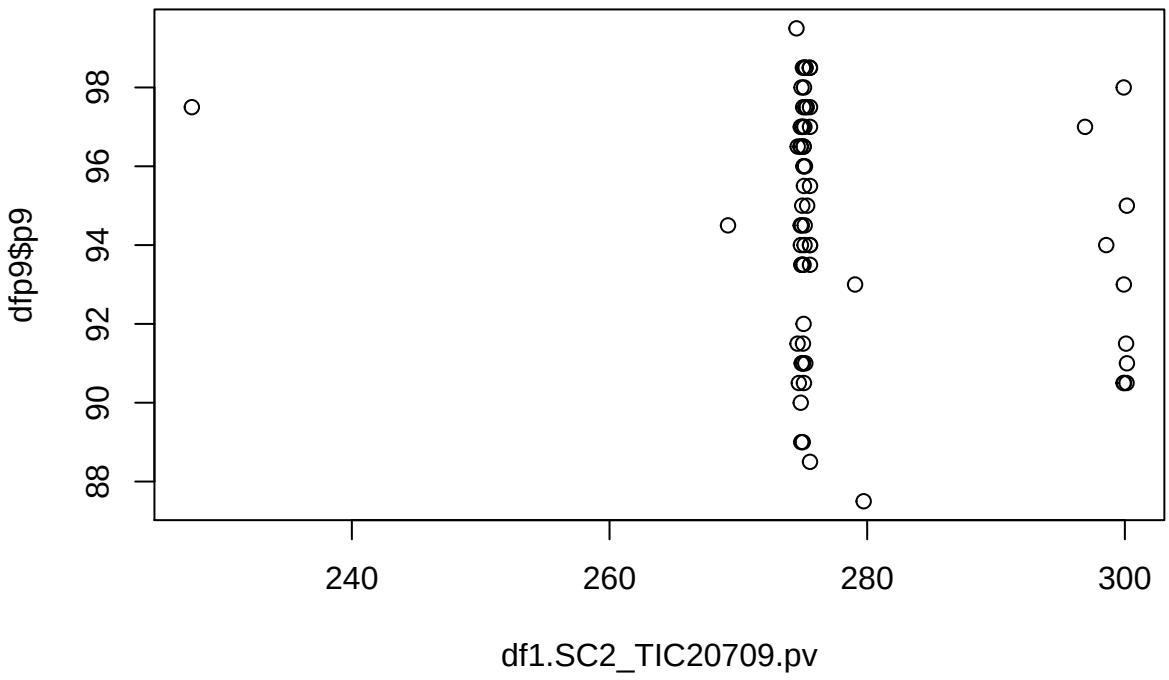

dfp9$p9
88
90
92
94
96
98
800
1000
1200
1400
df1.SC2_TIC55013_S.pv

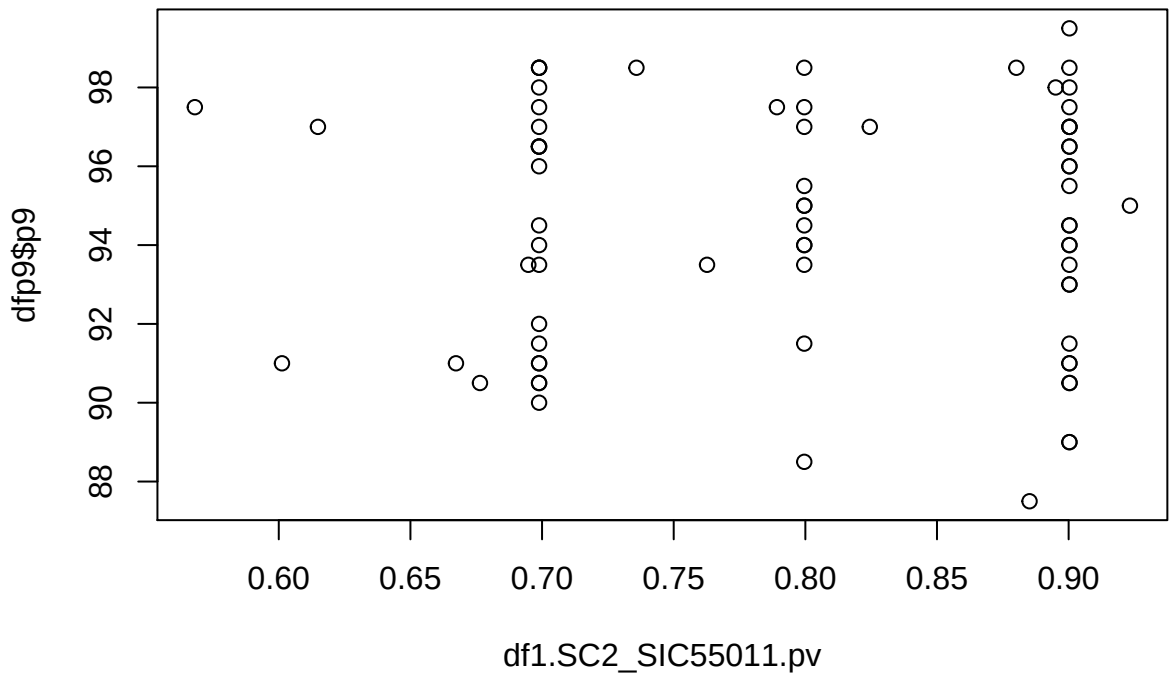

dfp9$p9
88
90
92
94
96
98
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55014_S.pv


dfp9$p9
88
90
92
94
96
98
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55015_S.pv

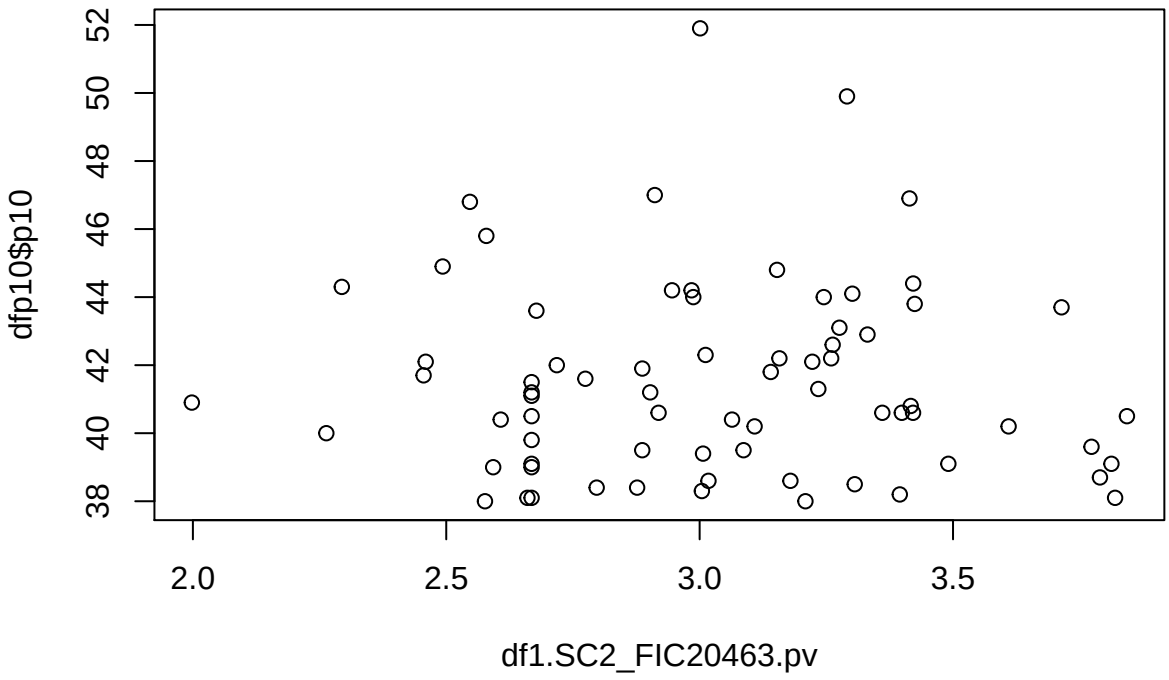

dfp9$p9
88
90
92
94
96
98
50
55
60
65
70
75
df1.SC2_TIC20627.pv

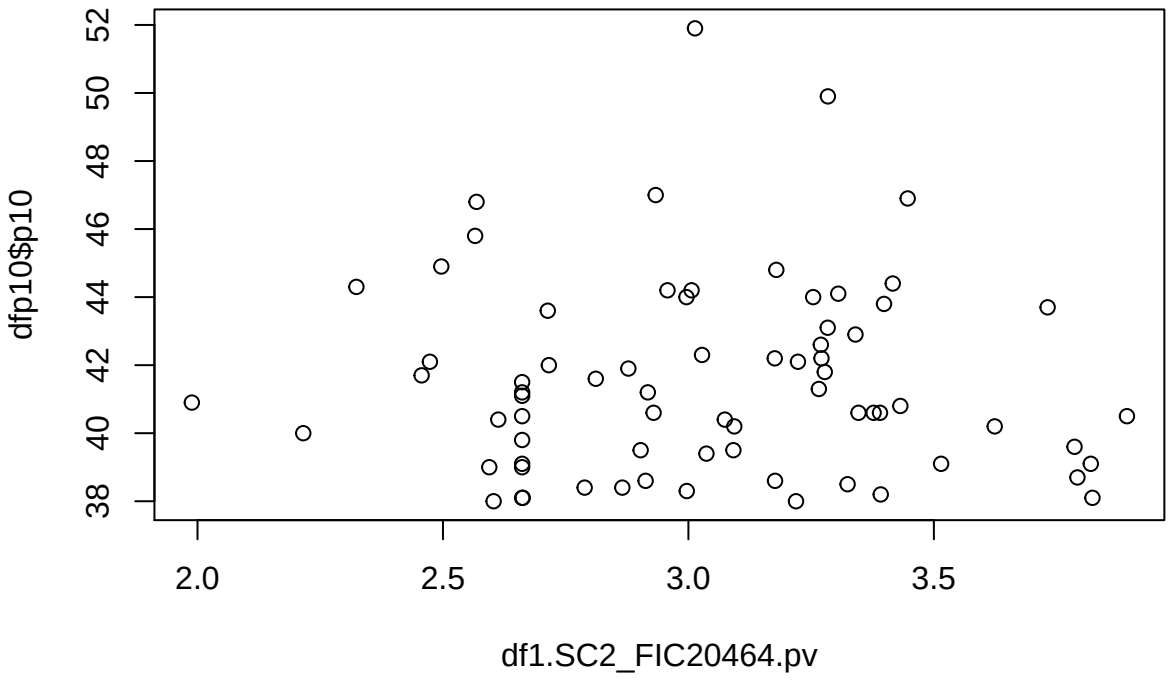

dfp9$p9
88
90
92
94
96
98
50
52
54
56
58
df1.SC2_LIC55022.pv

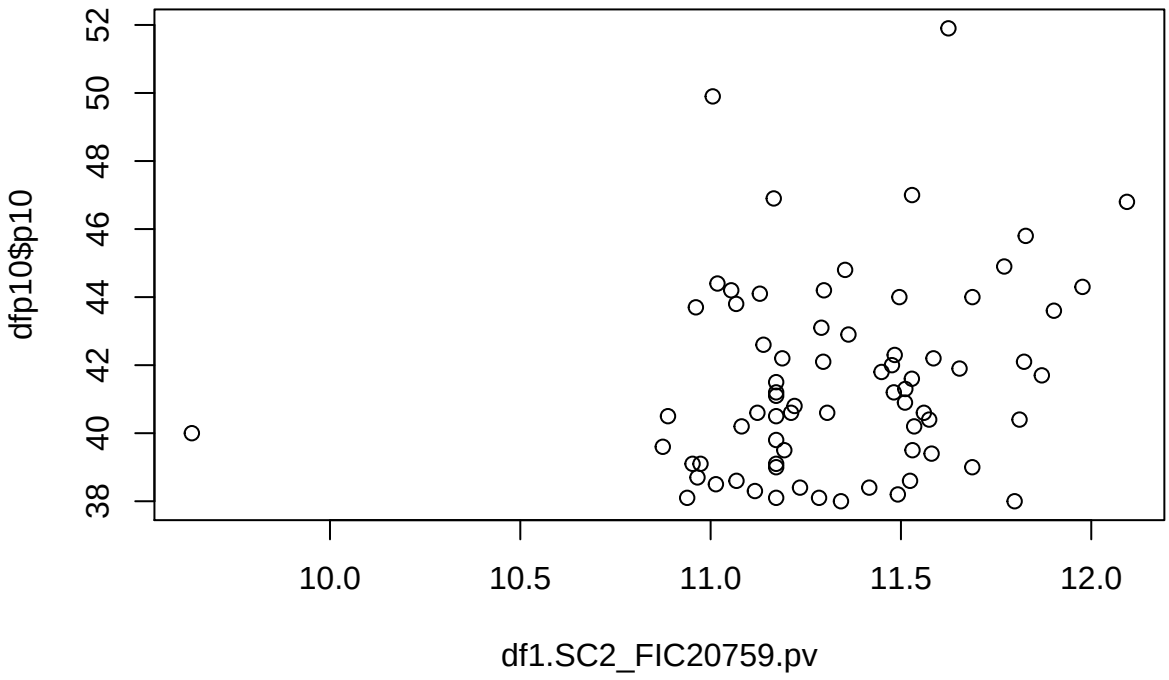

dfp9$p9
88
90
92
94
96
98
240
260
280
300
df1.SC2_TIC20709.pv

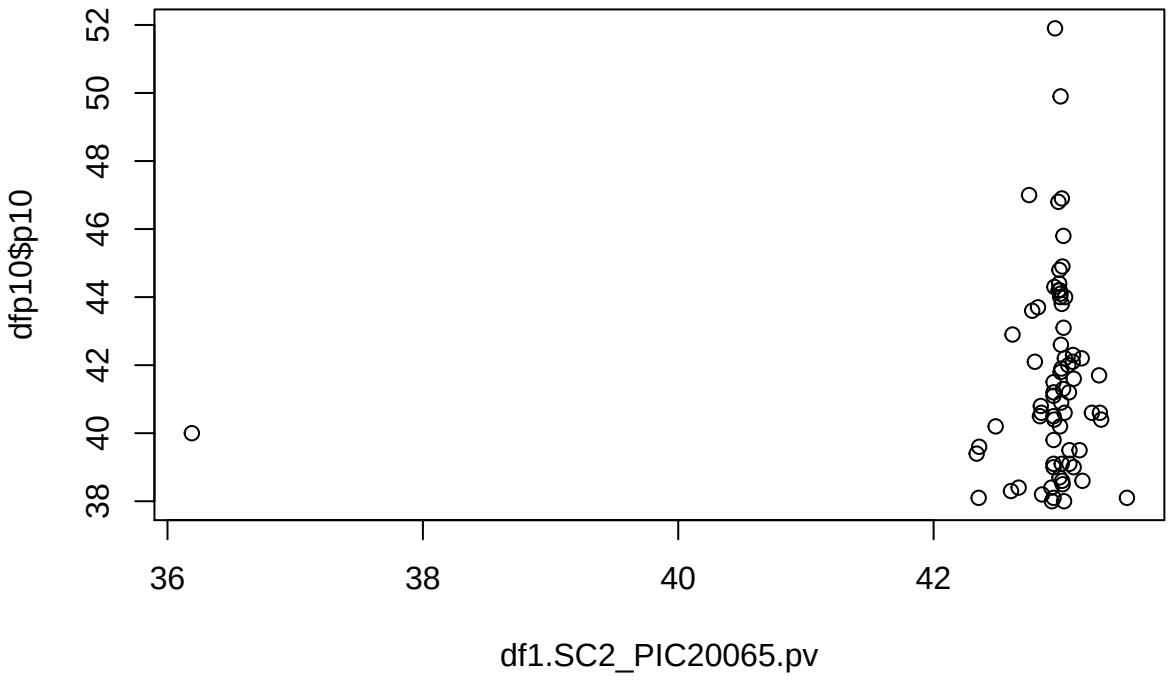


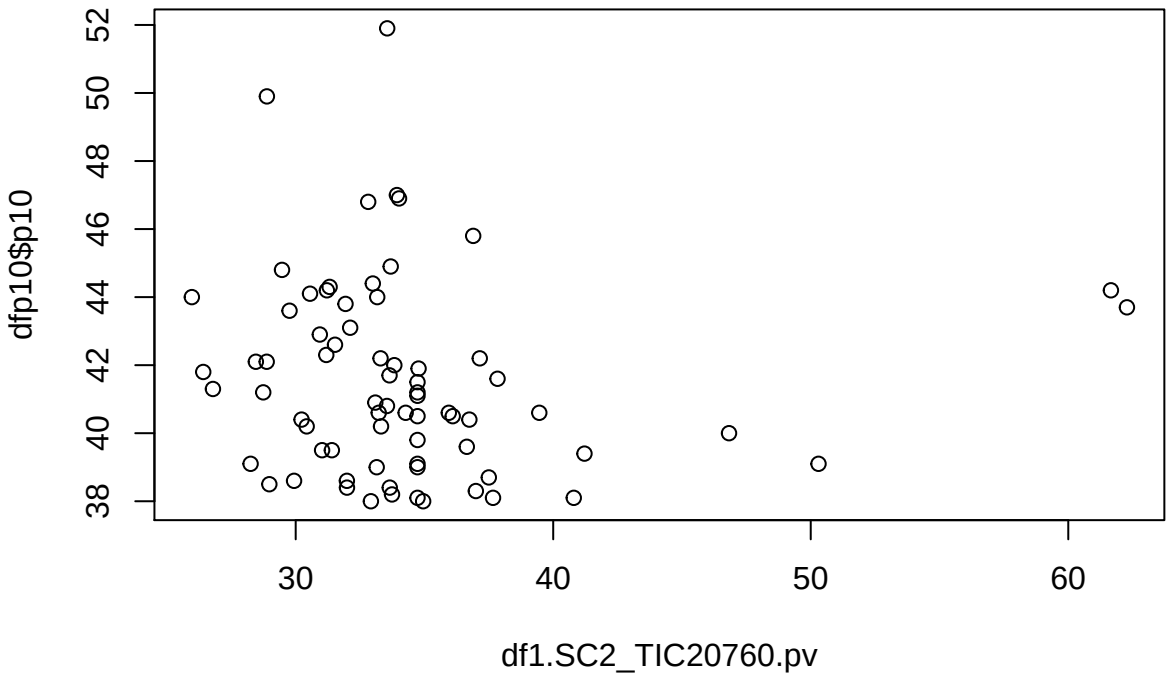

```
dfp10<-data.frame(p10,input)
plot(dfp10$p10~.,dfp10)
```

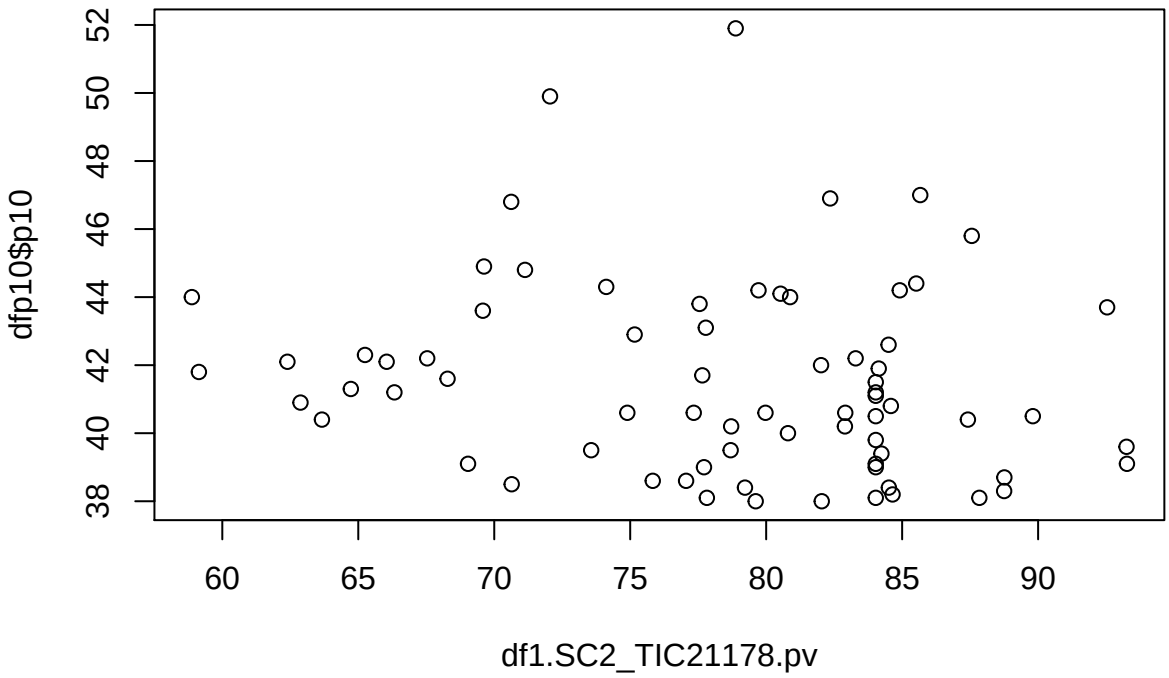

dfp10$p10
38
40
42
44
46
48
50
52
2.0
2.5
3.0
3.5
df1.SC2_FIC20463.pv

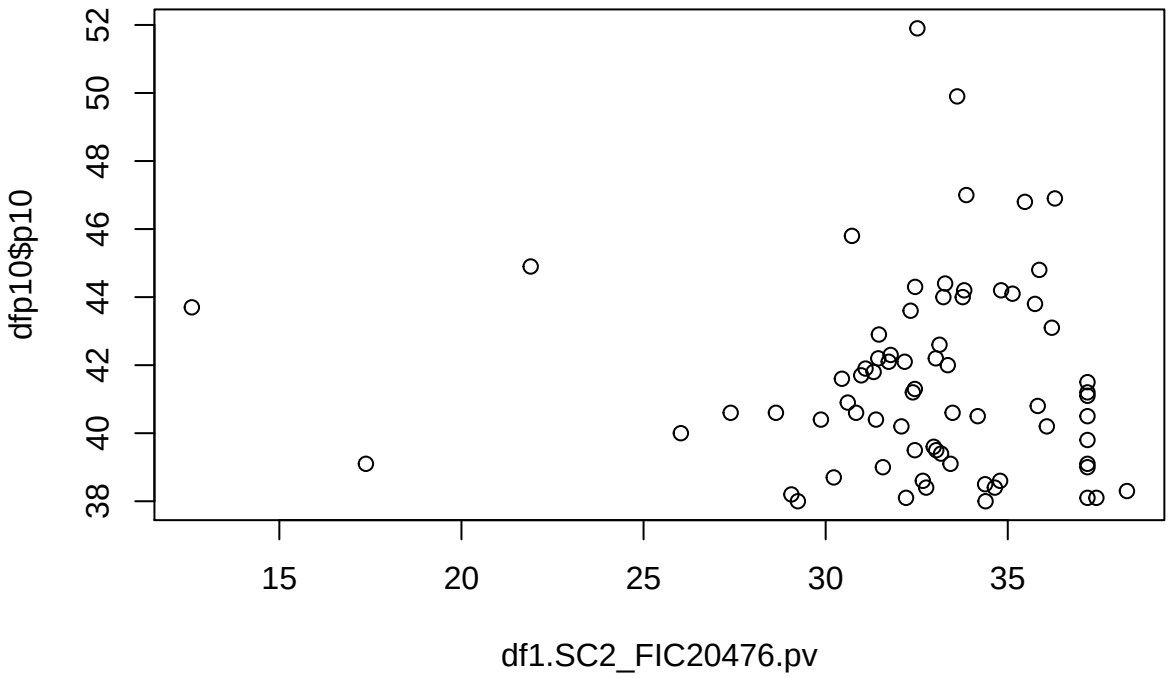

dfp10$p10
38
40
42
44
46
48
50
52
2.0
2.5
3.0
3.5
df1.SC2_FIC20464.pv

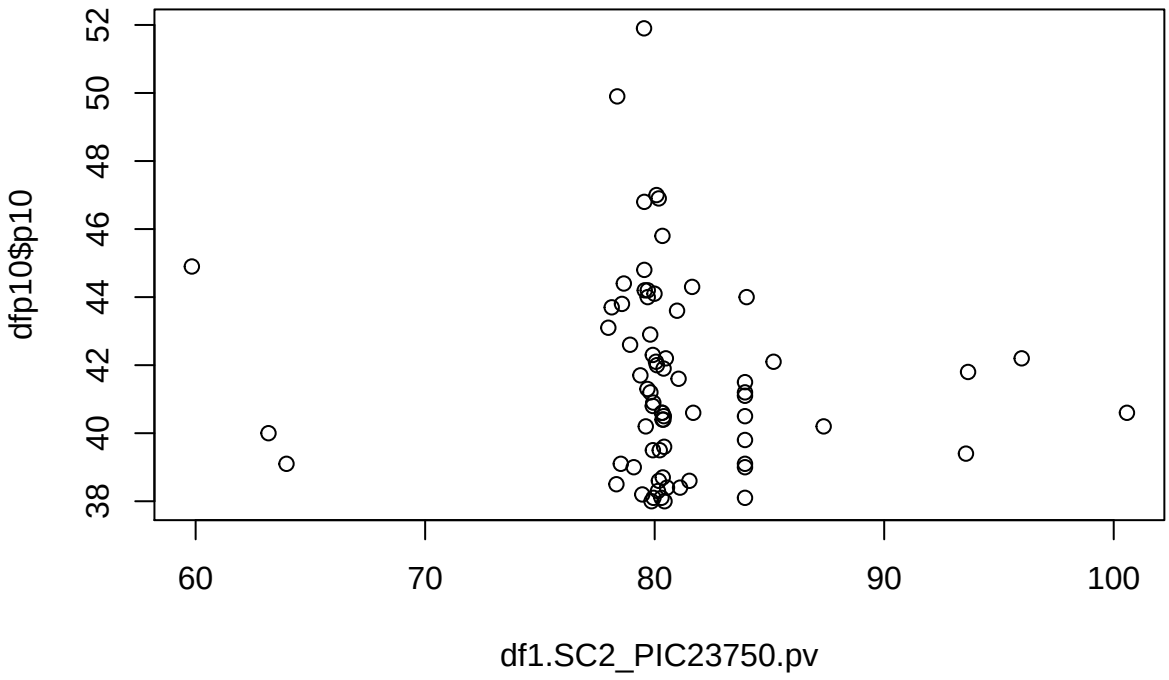

dfp10$p10
38
40
42
44
46
48
50
52
10.0
10.5
11.0
11.5
12.0
df1.SC2_FIC20759.pv

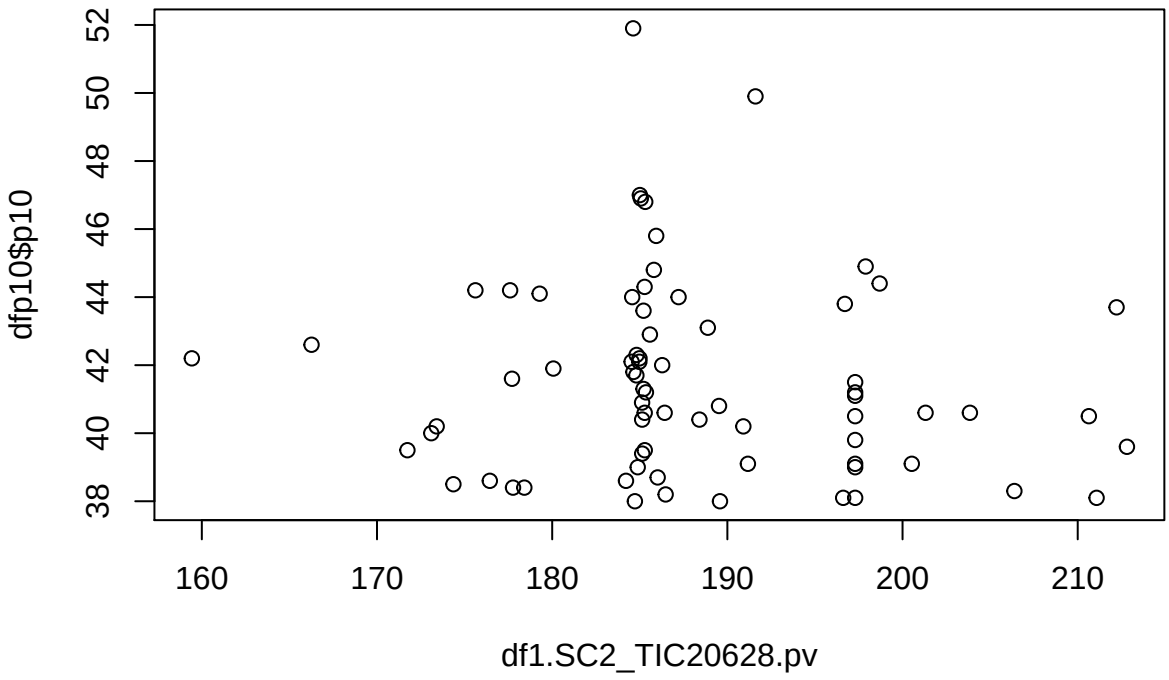

dfp10$p10
38
40
42
44
46
48
50
52
36
38
40
42
df1.SC2_PIC20065.pv

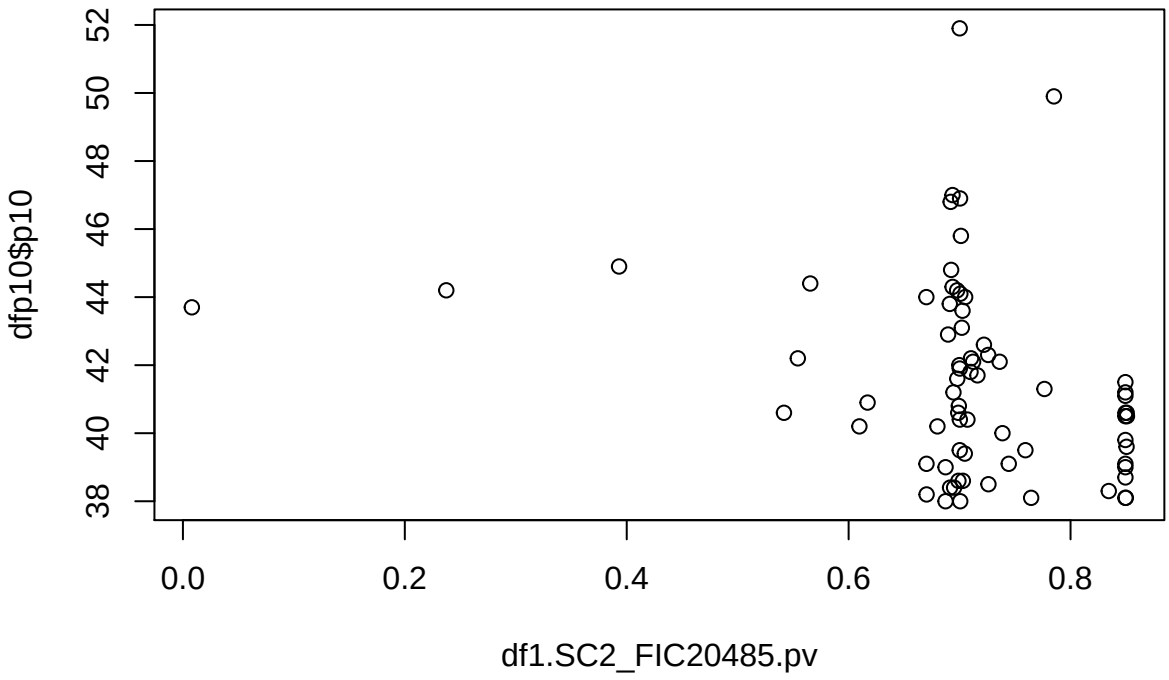

dfp10$p10
38
40
42
44
46
48
50
52
30
40
50
60
df1.SC2_TIC20760.pv

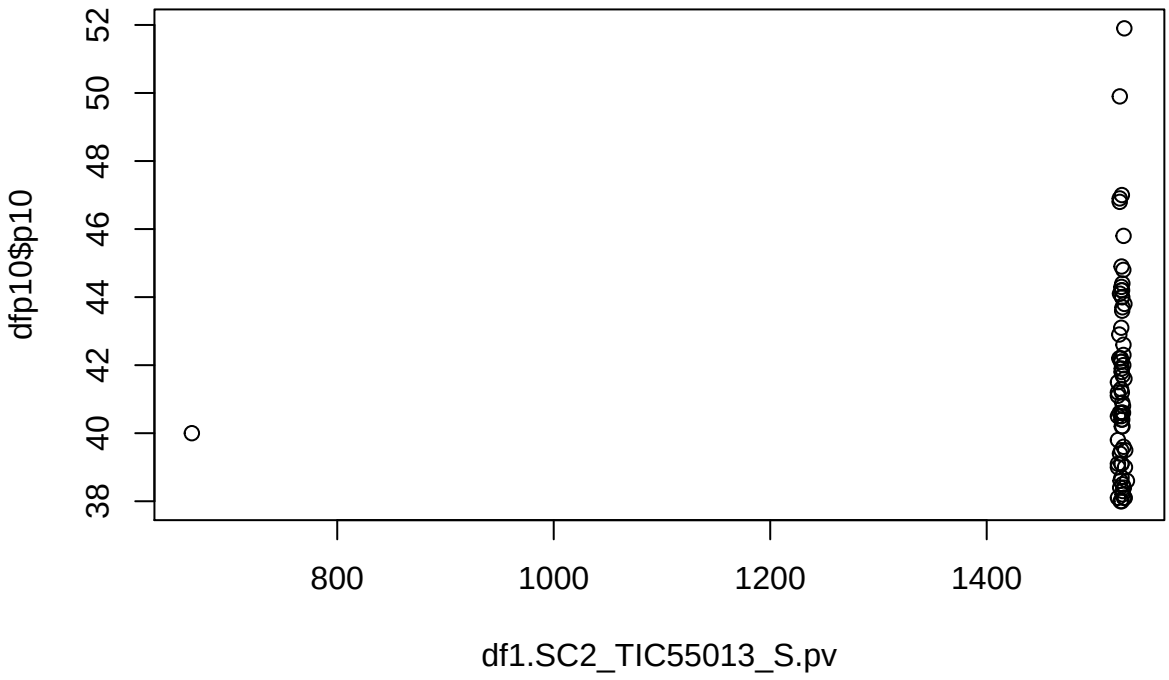

dfp10$p10
38
40
42
44
46
48
50
52
60
65
70
75
80
85
90
df1.SC2_TIC21178.pv

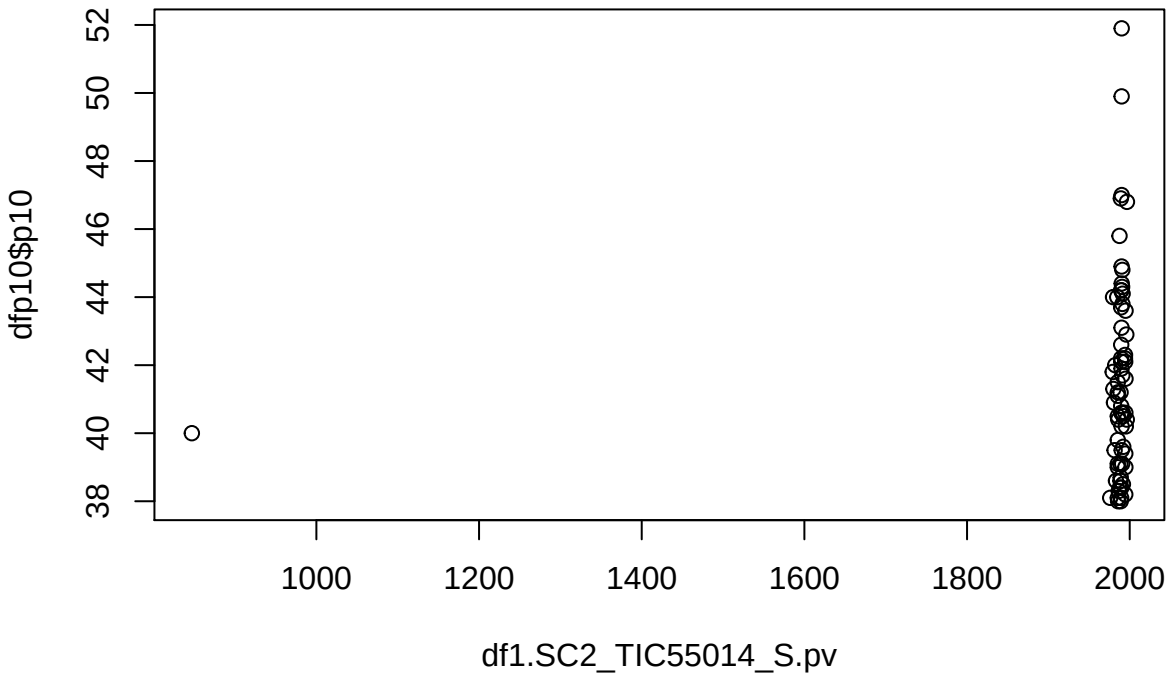

dfp10$p10
38
40
42
44
46
48
50
52
15
20
25
30
35
df1.SC2_FIC20476.pv

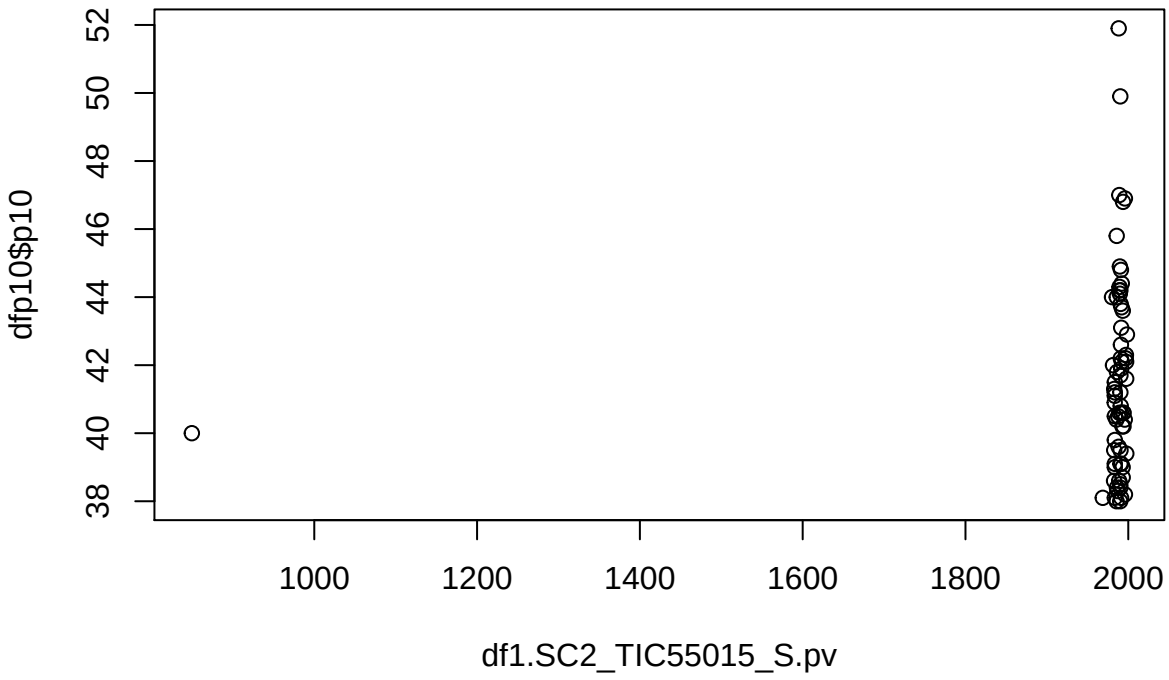

dfp10$p10
38
40
42
44
46
48
50
52
60
70
80
90
100
df1.SC2_PIC23750.pv

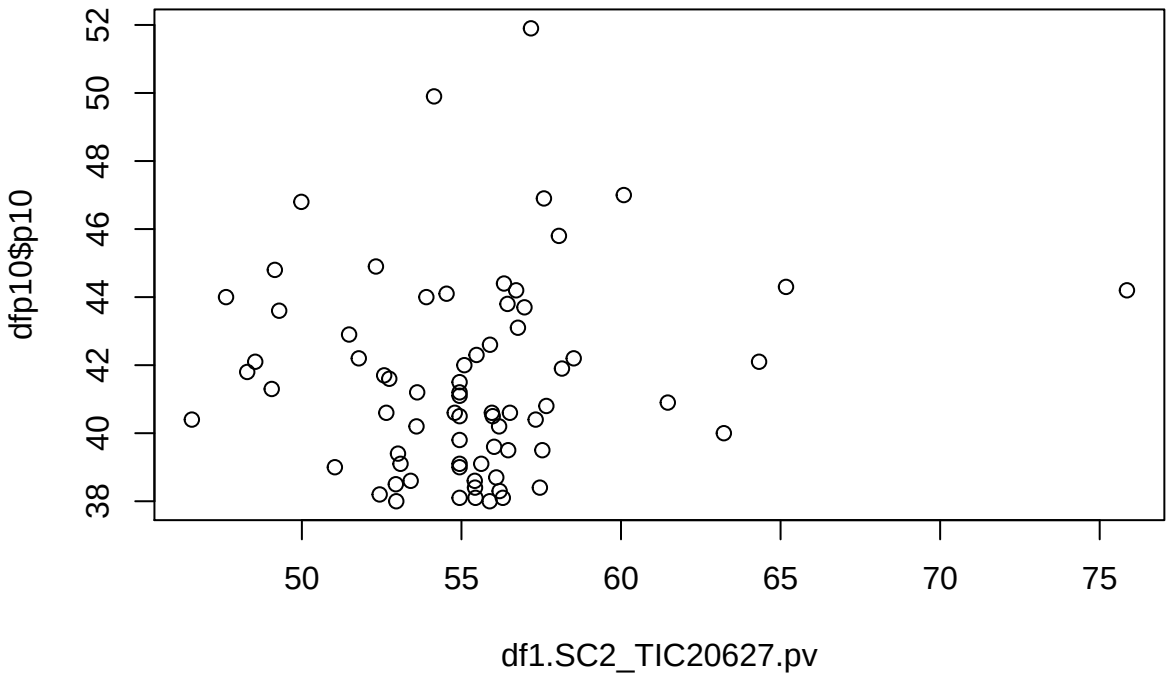

dfp10$p10
38
40
42
44
46
48
50
52
160
170
180
190
200
210
df1.SC2_TIC20628.pv

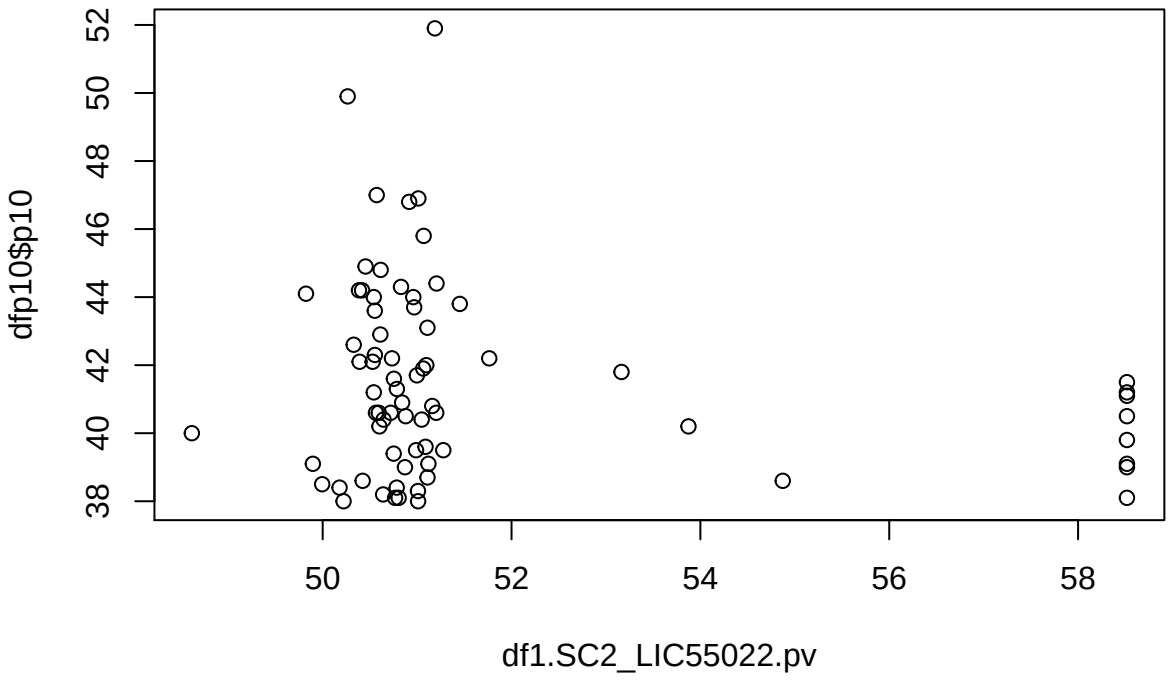

dfp10$p10
38
40
42
44
46
48
50
52
0.0
0.2
0.4
0.6
0.8
df1.SC2_FIC20485.pv

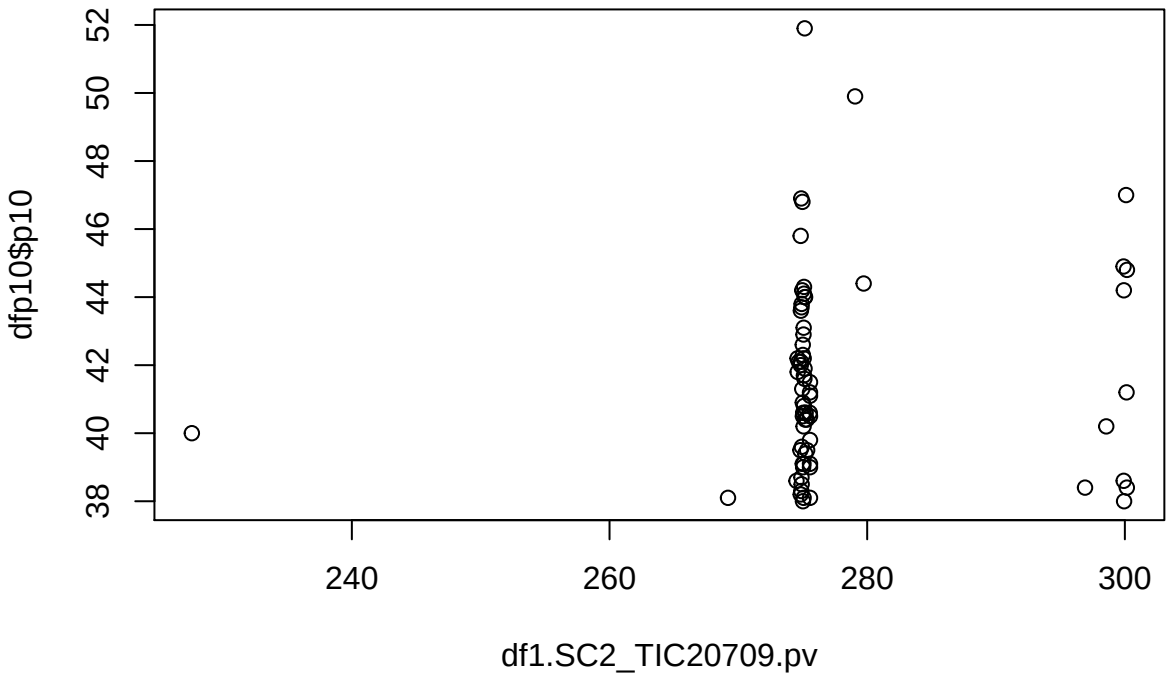

dfp10$p10
38
40
42
44
46
48
50
52
800
1000
1200
1400
df1.SC2_TIC55013_S.pv

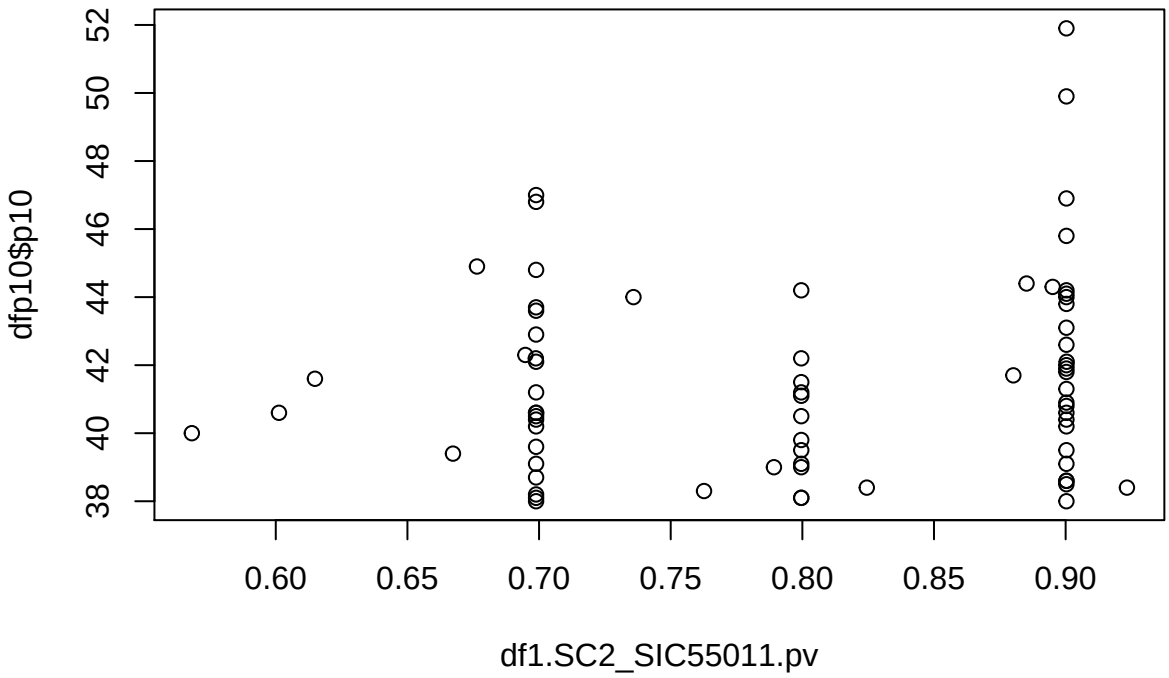

dfp10$p10
38
40
42
44
46
48
50
52
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55014_S.pv


dfp10$p10
38
40
42
44
46
48
50
52
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55015_S.pv

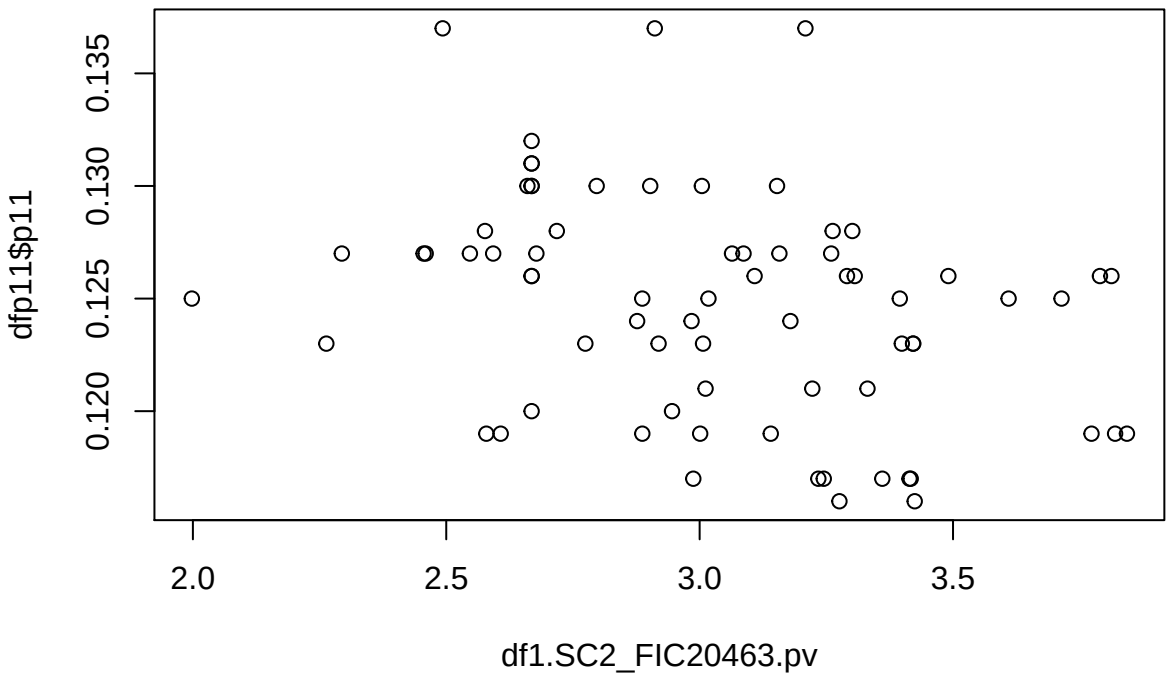

dfp10$p10
38
40
42
44
46
48
50
52
50
55
60
65
70
75
df1.SC2_TIC20627.pv

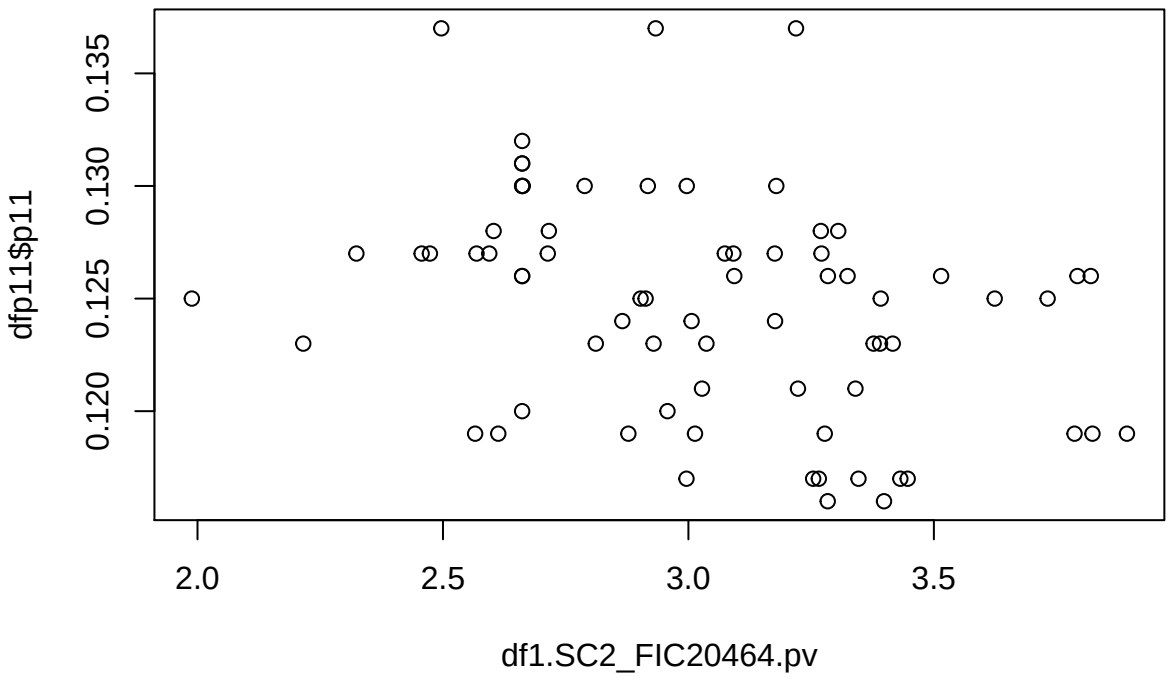

dfp10$p10
38
40
42
44
46
48
50
52
50
52
54
56
58
df1.SC2_LIC55022.pv

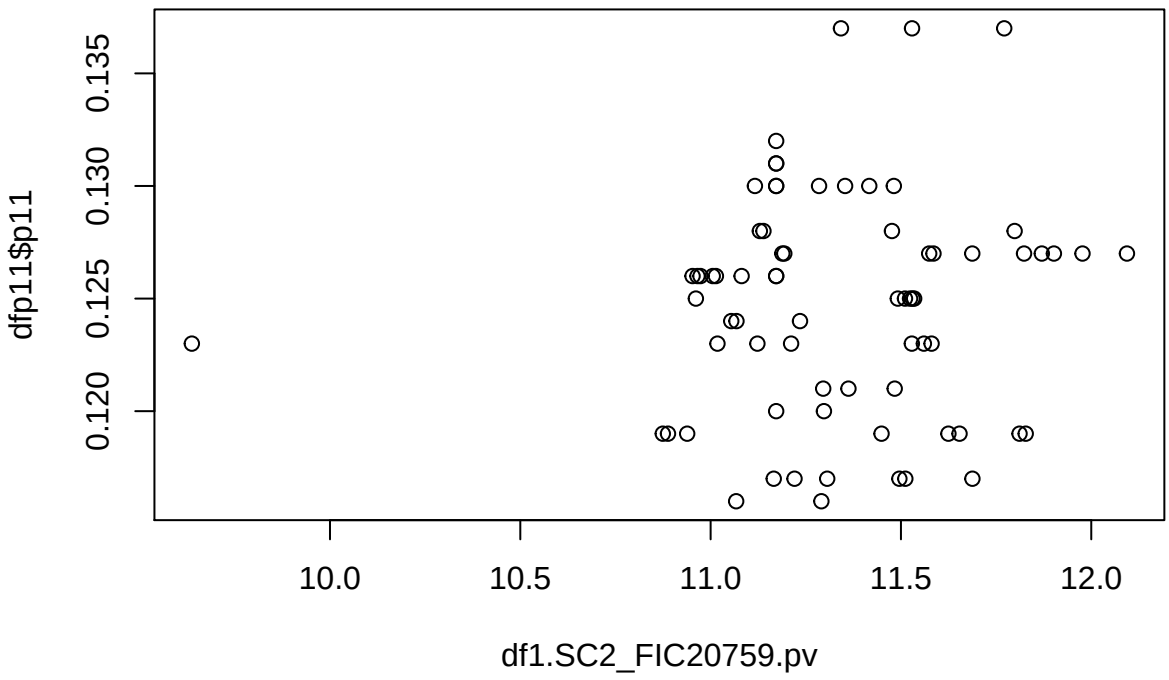

dfp10$p10
38
40
42
44
46
48
50
52
240
260
280
300
df1.SC2_TIC20709.pv

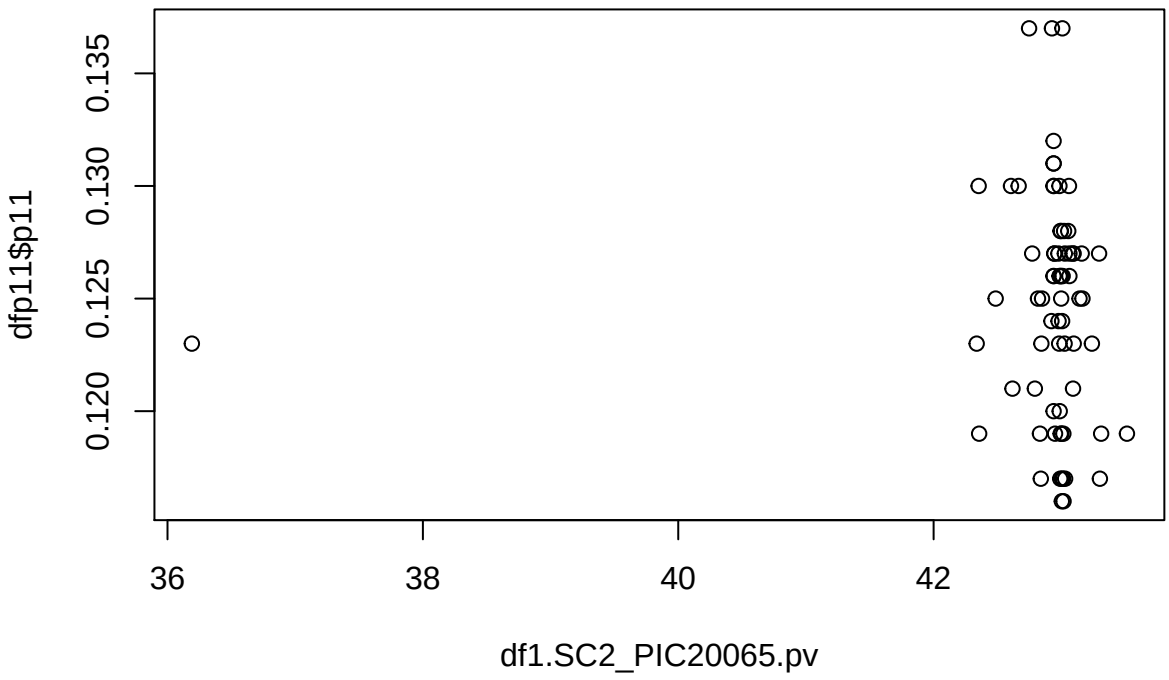


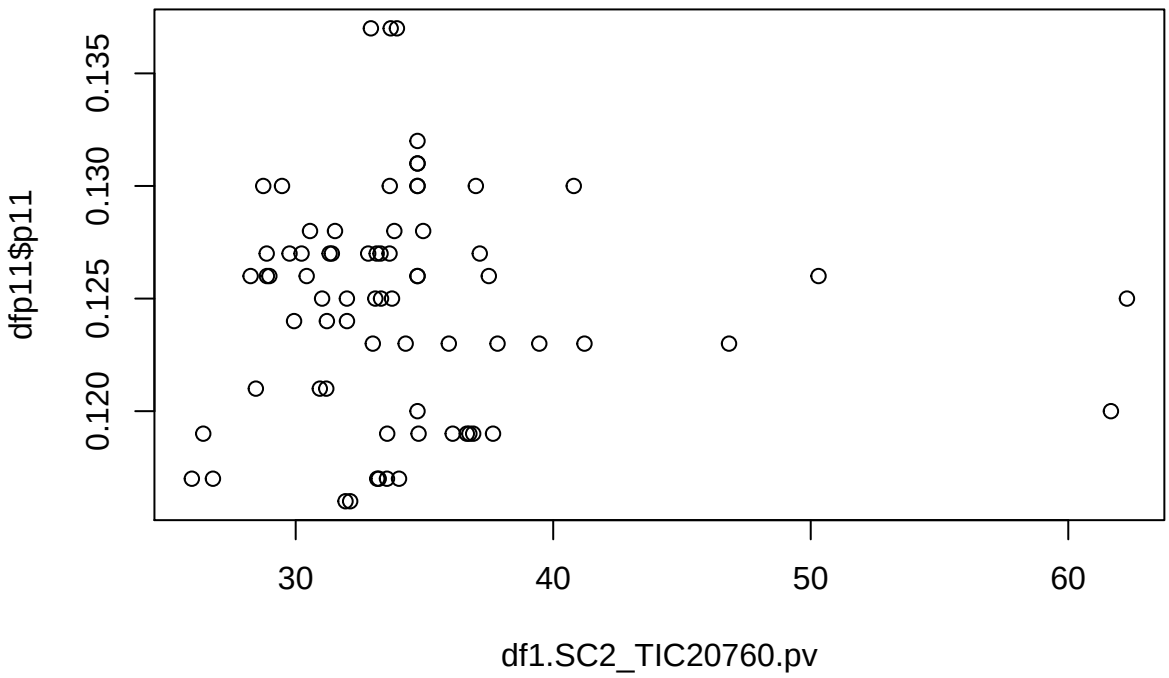

```
dfp11<-data.frame(p11,input)
plot(dfp11$p11~.,dfp11)
```

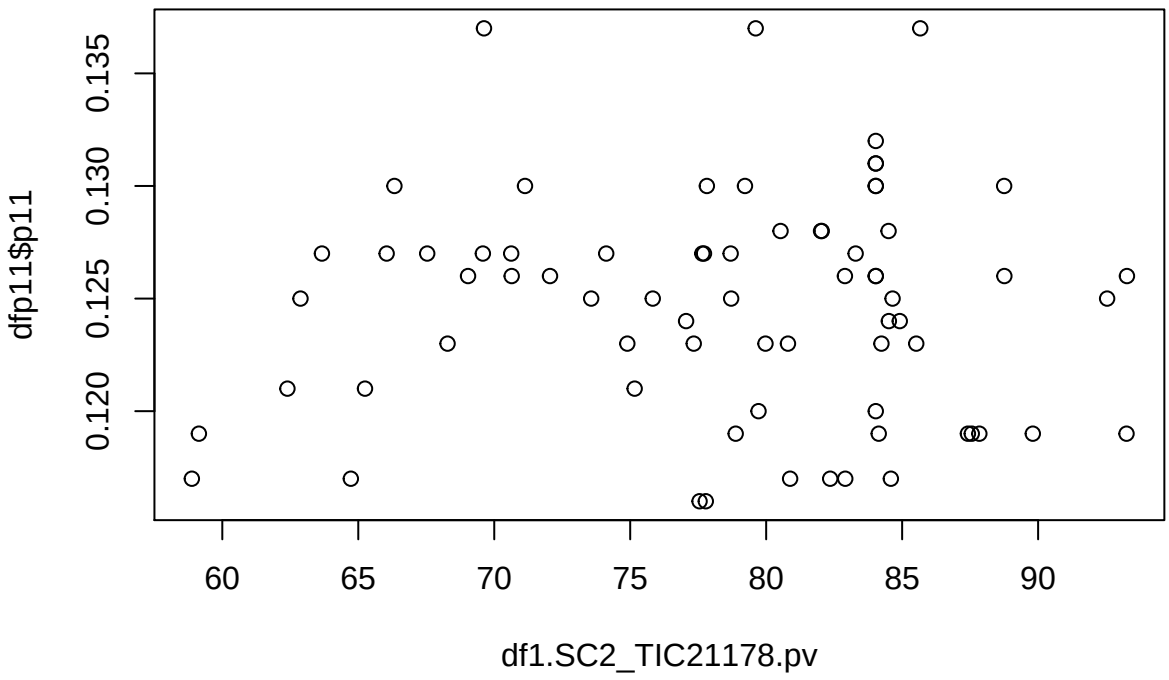

dfp11$p11
0.120
0.125
0.130
0.135
2.0
2.5
3.0
3.5
df1.SC2_FIC20463.pv

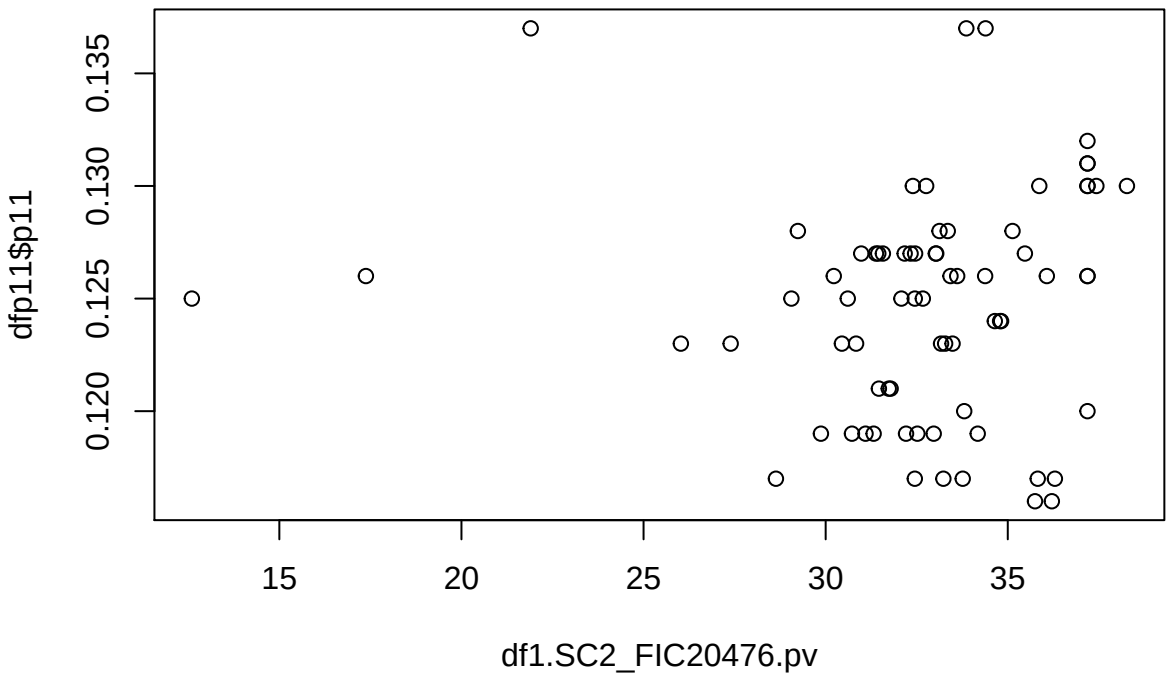

dfp11$p11
0.120
0.125
0.130
0.135
2.0
2.5
3.0
3.5
df1.SC2_FIC20464.pv

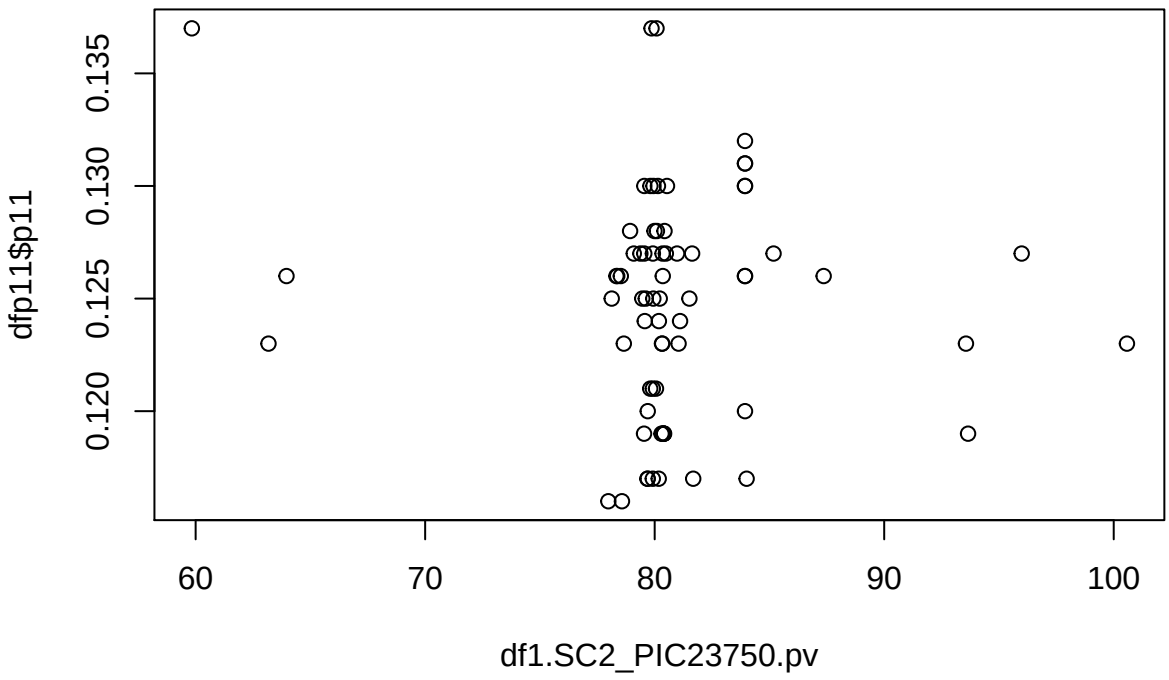

dfp11$p11
0.120
0.125
0.130
0.135
10.0
10.5
11.0
11.5
12.0
df1.SC2_FIC20759.pv

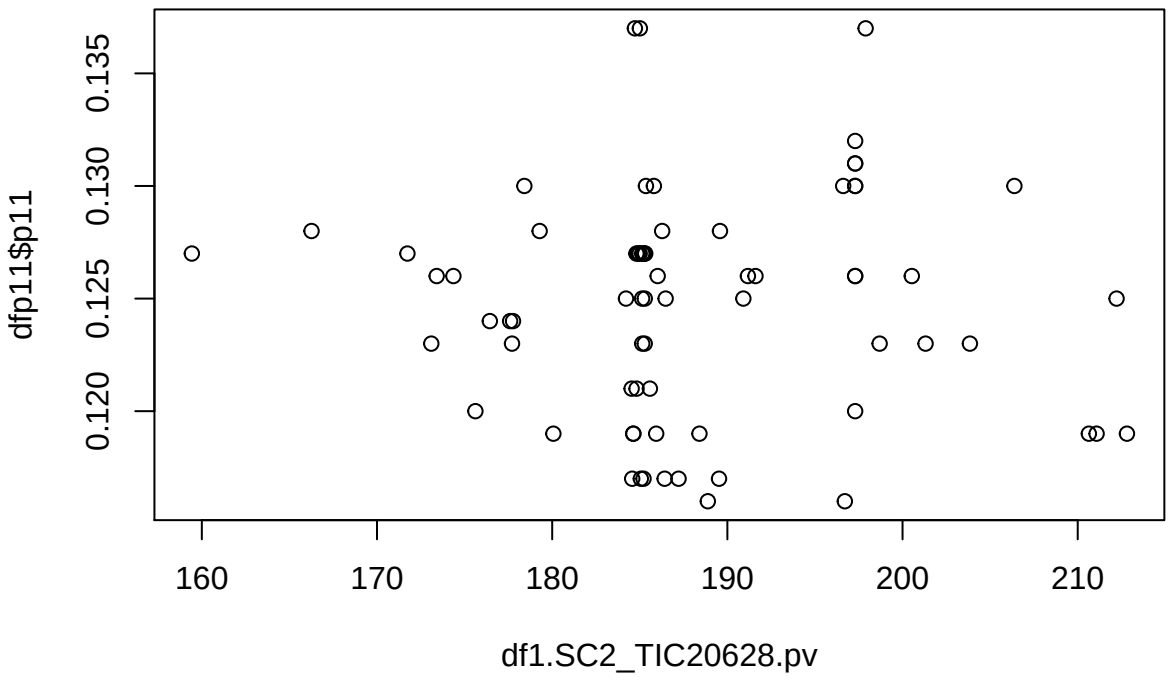

dfp11$p11
0.120
0.125
0.130
0.135
36
38
40
42
df1.SC2_PIC20065.pv

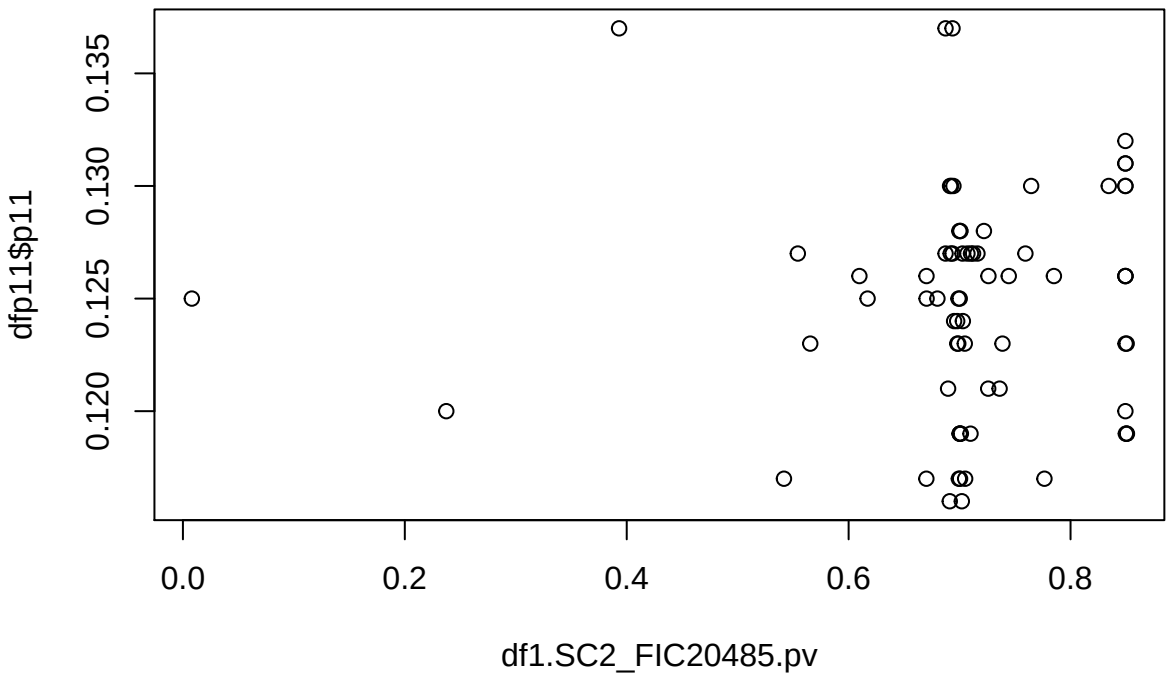

dfp11$p11
0.120
0.125
0.130
0.135
30
40
50
60
df1.SC2_TIC20760.pv

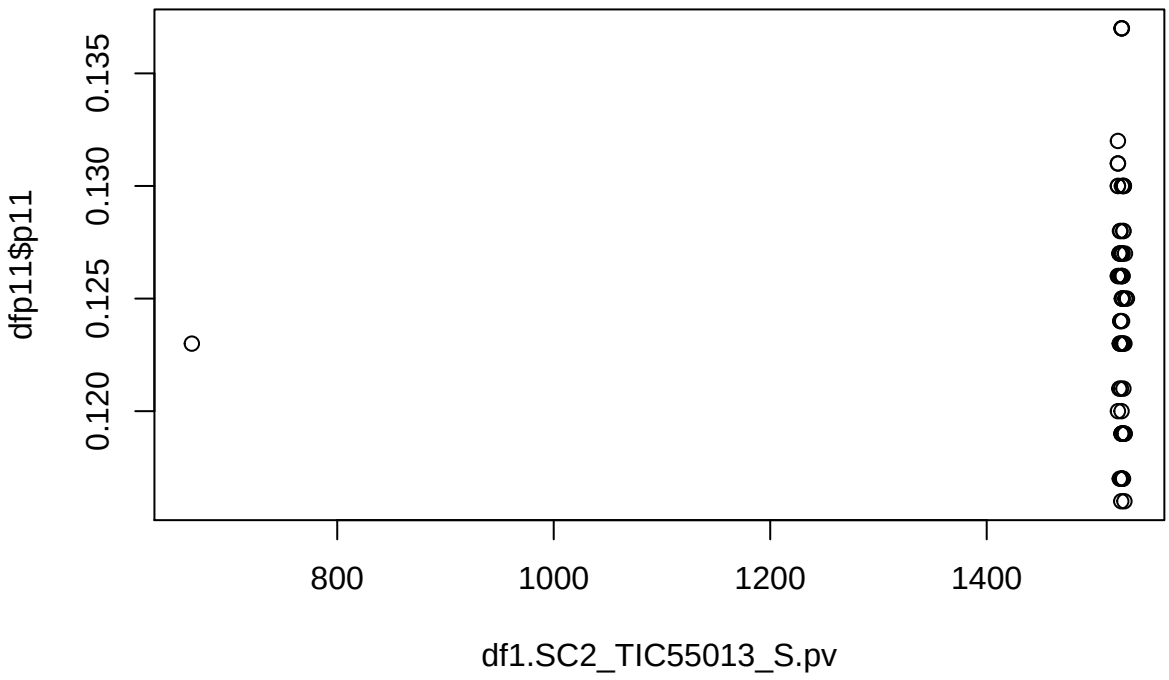

dfp11$p11
0.120
0.125
0.130
0.135
60
65
70
75
80
85
90
df1.SC2_TIC21178.pv

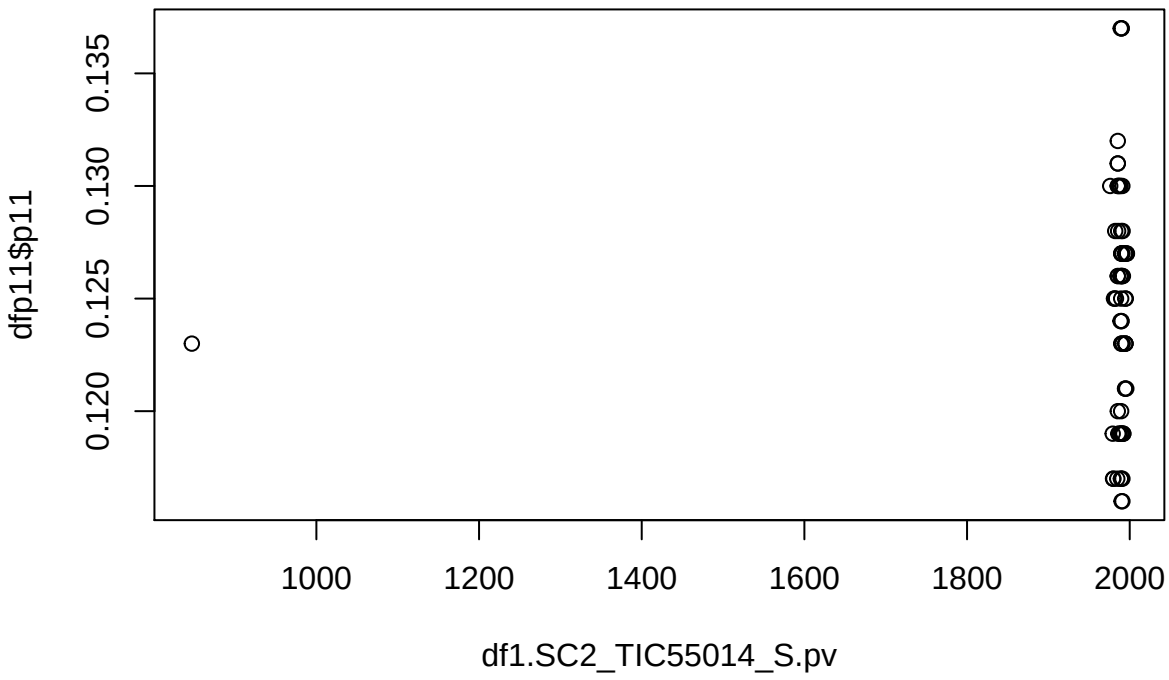

dfp11$p11
0.120
0.125
0.130
0.135
15
20
25
30
35
df1.SC2_FIC20476.pv

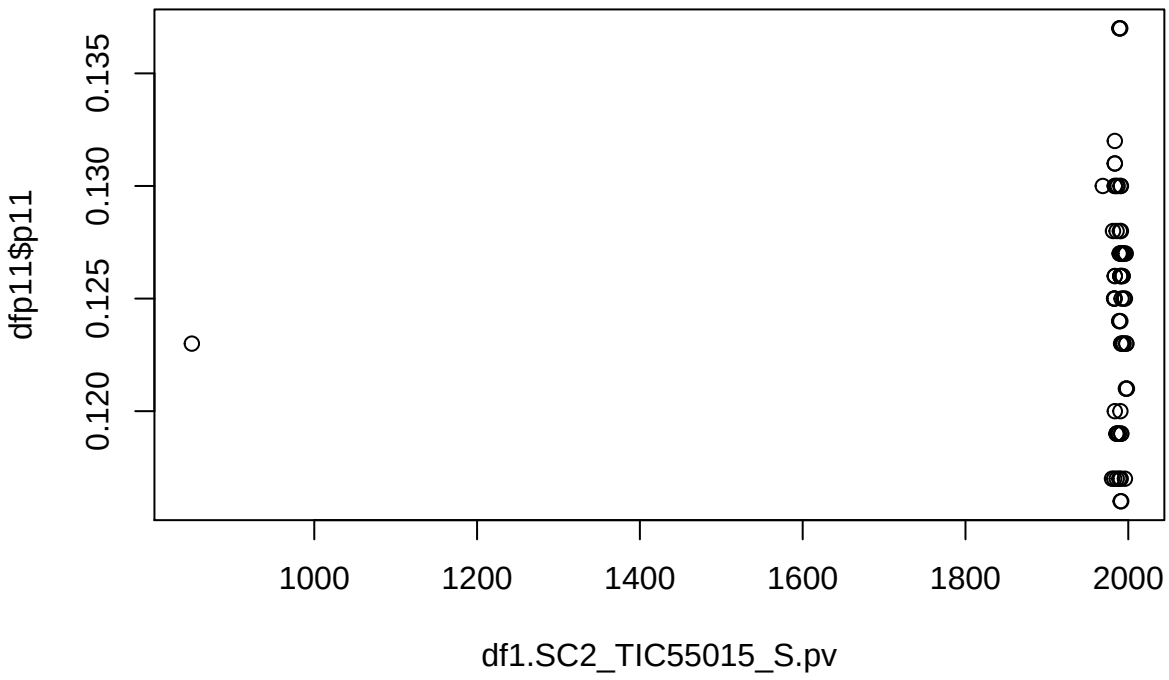

dfp11$p11
0.120
0.125
0.130
0.135
60
70
80
90
100
df1.SC2_PIC23750.pv

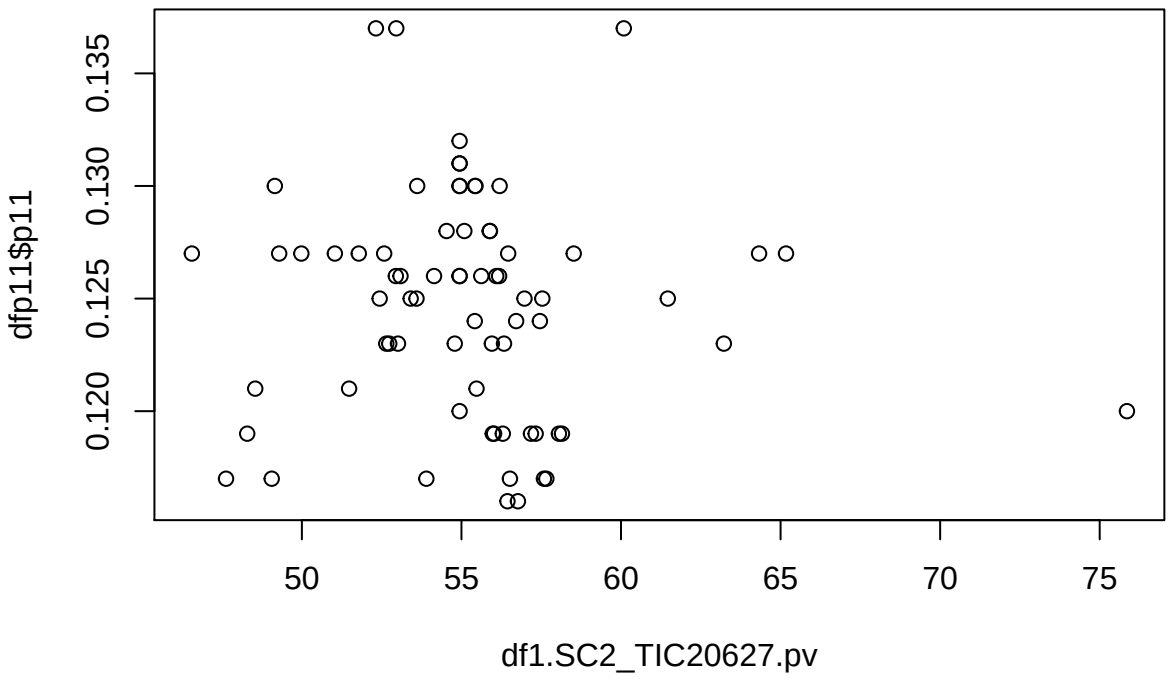

dfp11$p11
0.120
0.125
0.130
0.135
160
170
180
190
200
210
df1.SC2_TIC20628.pv

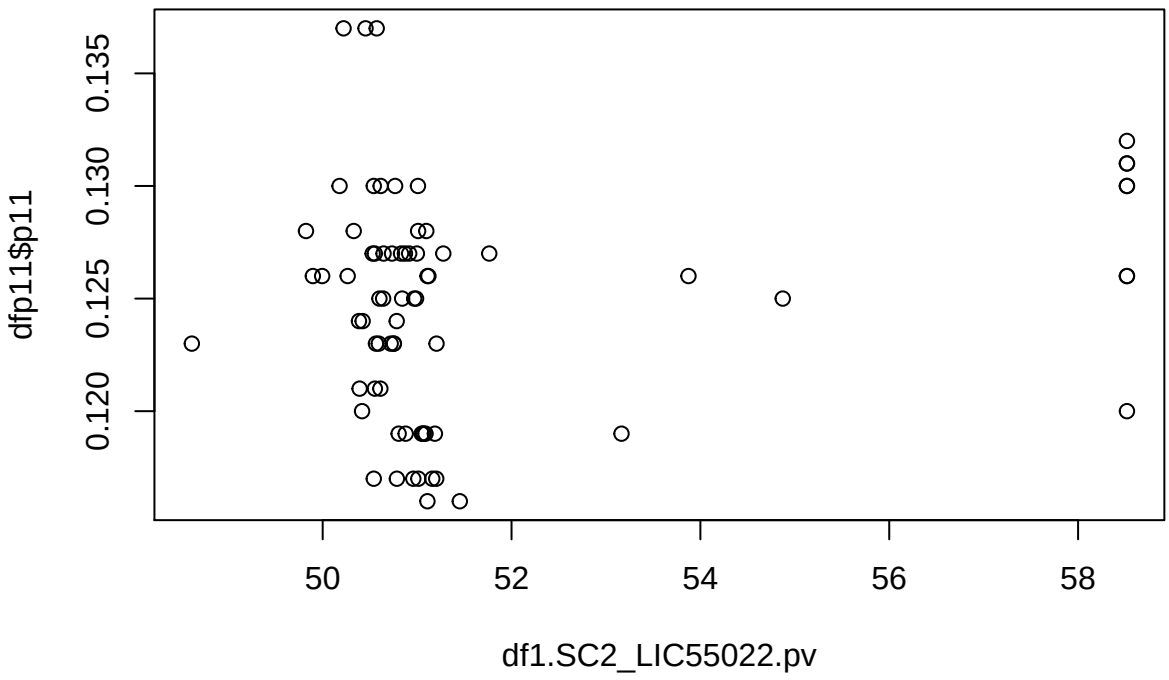

dfp11$p11
0.120
0.125
0.130
0.135
0.0
0.2
0.4
0.6
0.8
df1.SC2_FIC20485.pv

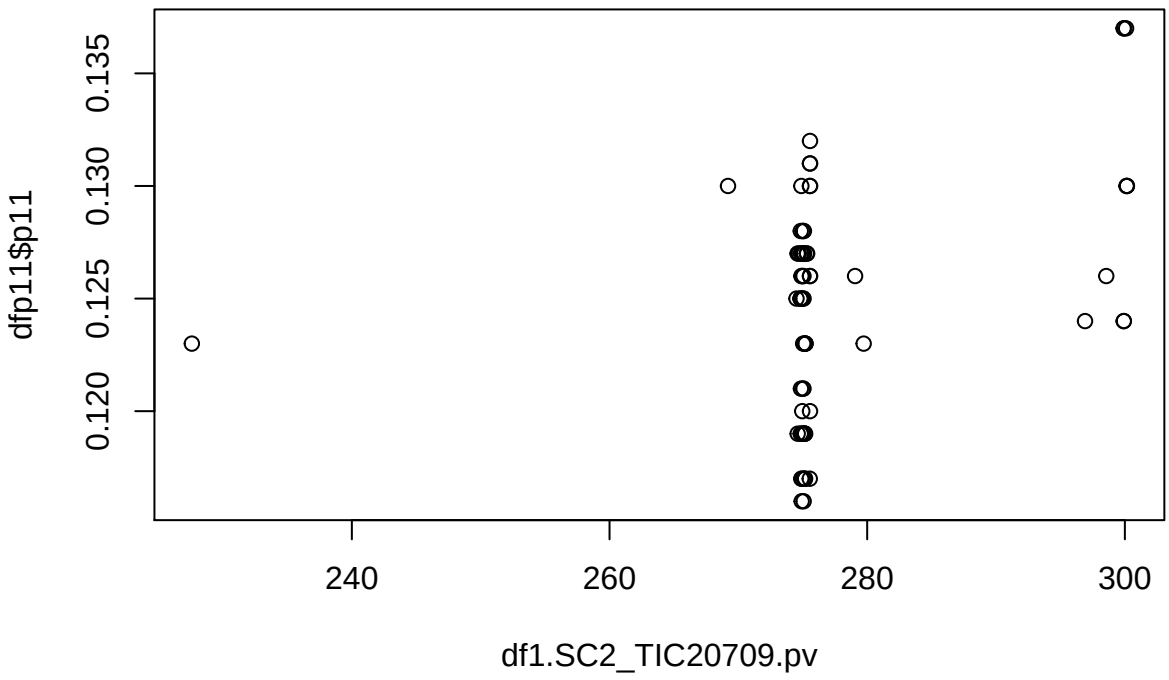

dfp11$p11
0.120
0.125
0.130
0.135
800
1000
1200
1400
df1.SC2_TIC55013_S.pv

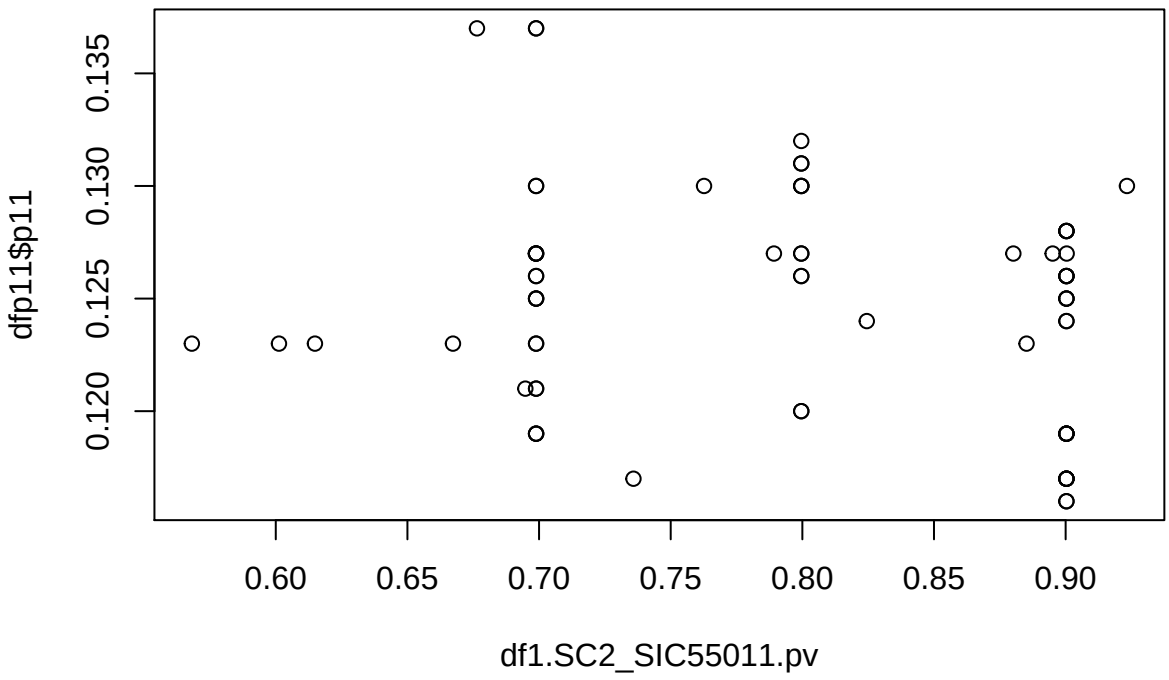

dfp11$p11
0.120
0.125
0.130
0.135
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55014_S.pv


dfp11$p11
0.120
0.125
0.130
0.135
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55015_S.pv

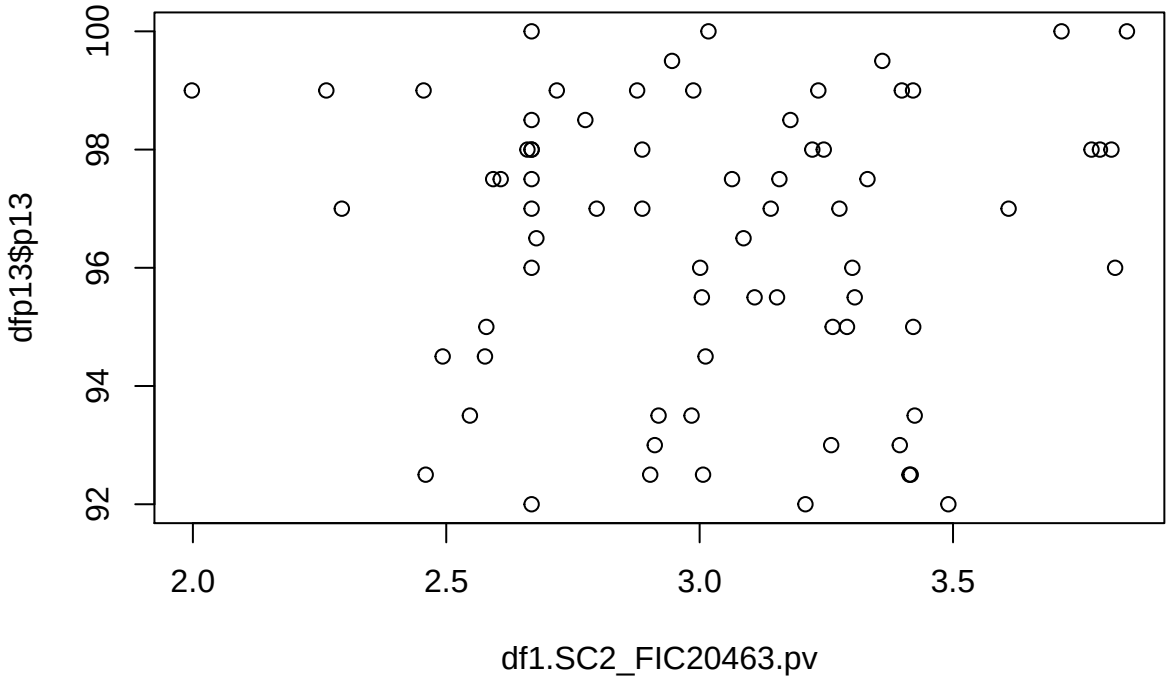

dfp11$p11
0.120
0.125
0.130
0.135
50
55
60
65
70
75
df1.SC2_TIC20627.pv

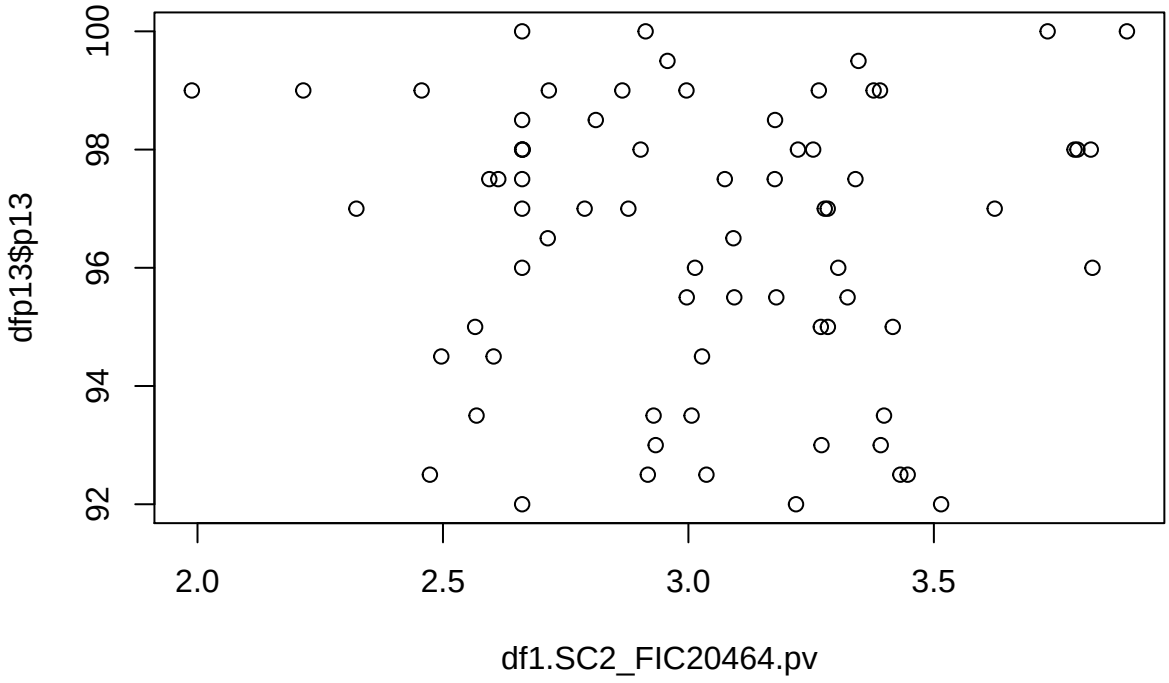

dfp11$p11
0.120
0.125
0.130
0.135
50
52
54
56
58
df1.SC2_LIC55022.pv

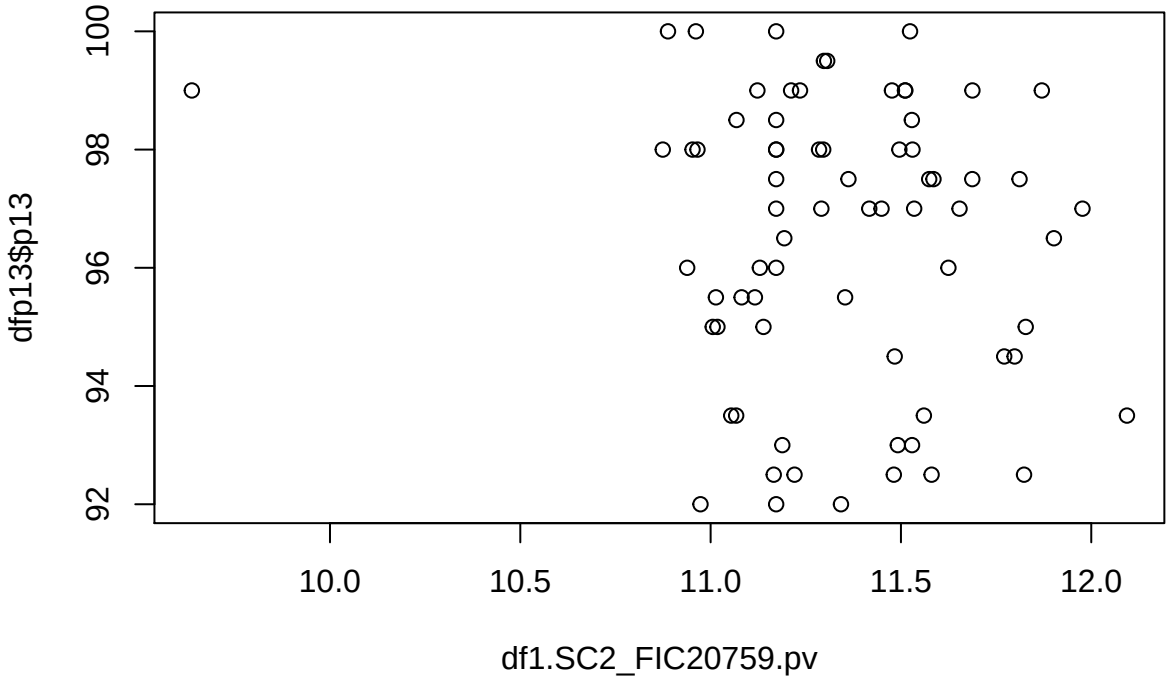

dfp11$p11
0.120
0.125
0.130
0.135
240
260
280
300
df1.SC2_TIC20709.pv

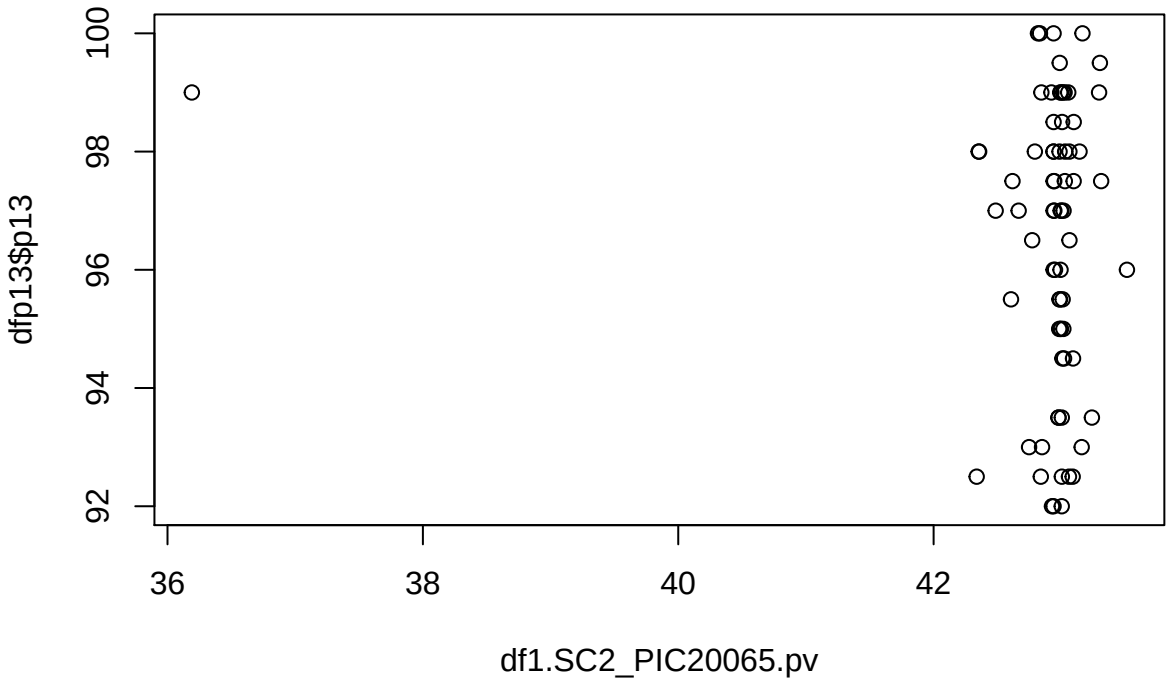


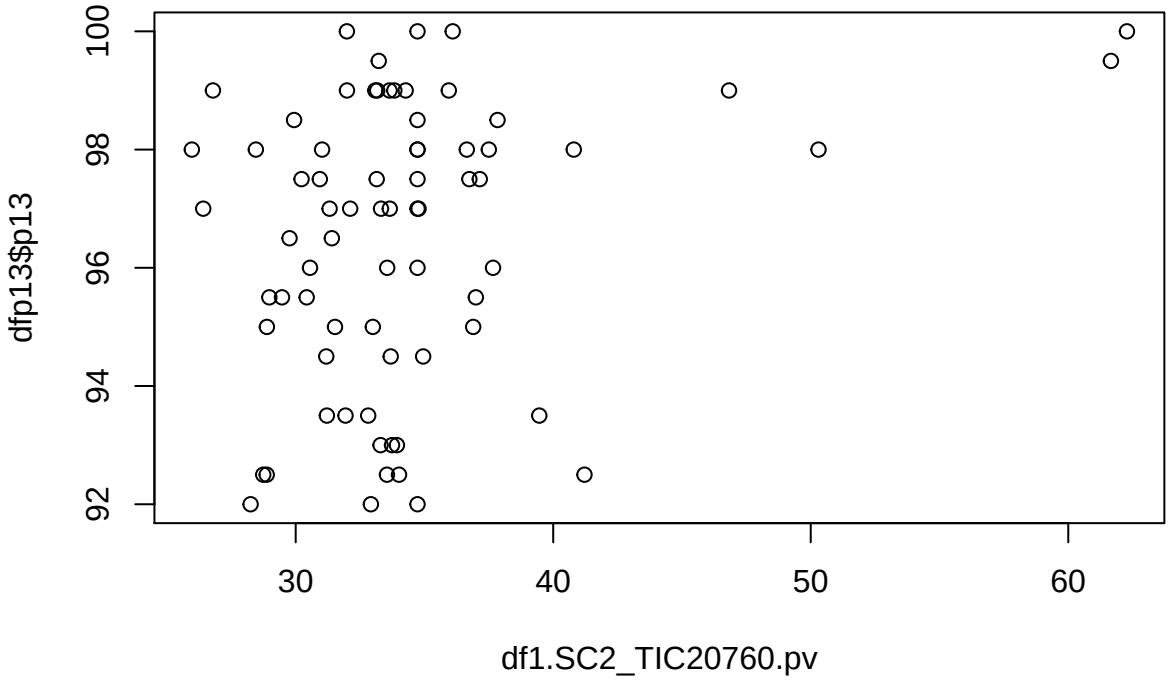

```
dfp13<-data.frame(p13,input)
plot(dfp13$p13~.,dfp13)
```

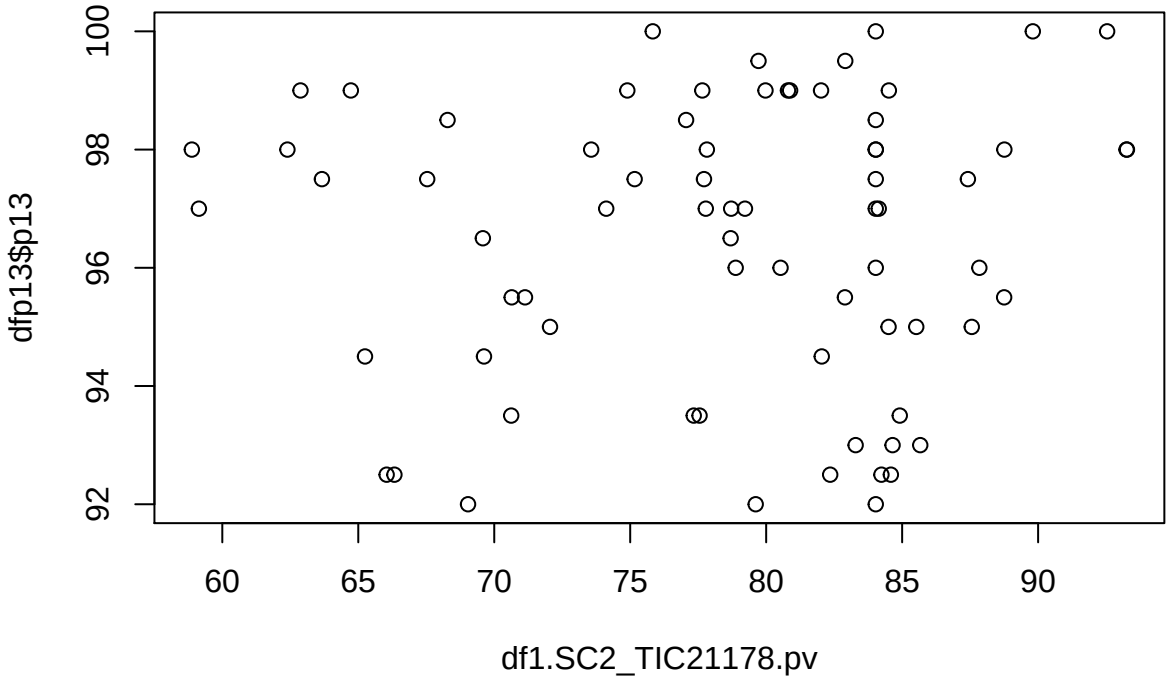

dfp13$p13
100
98
96
94
92
2.0
2.5
3.0
3.5
df1.SC2_FIC20463.pv

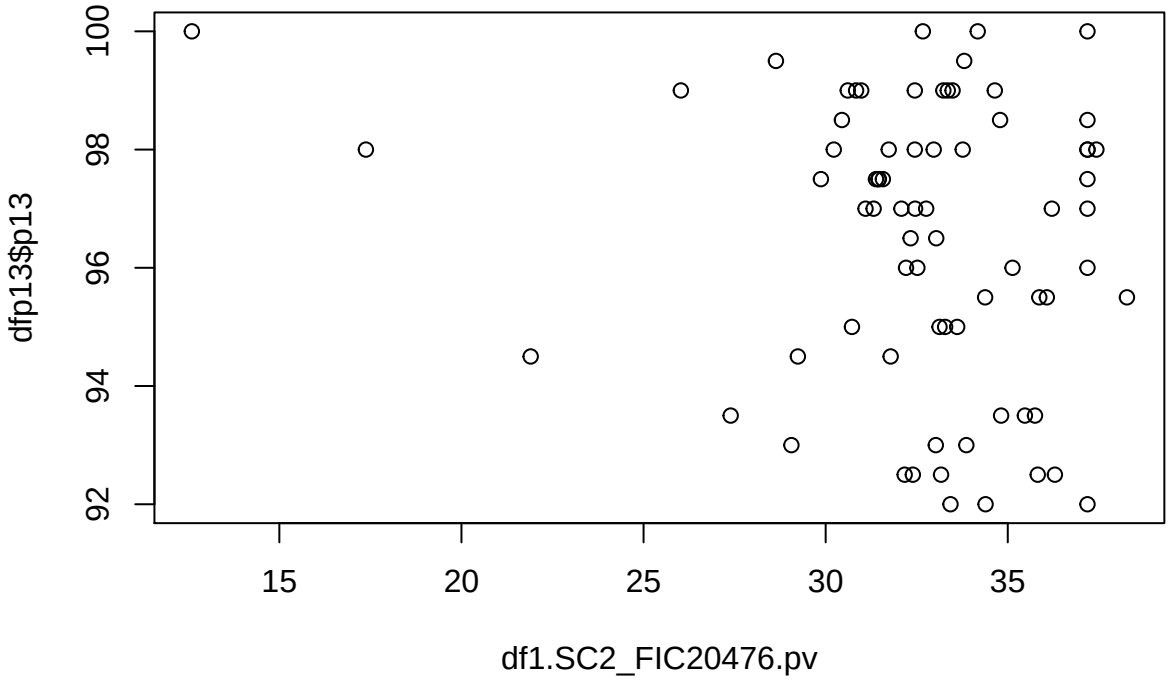

dfp13$p13
100
98
96
94
92
2.0
2.5
3.0
3.5
df1.SC2_FIC20464.pv

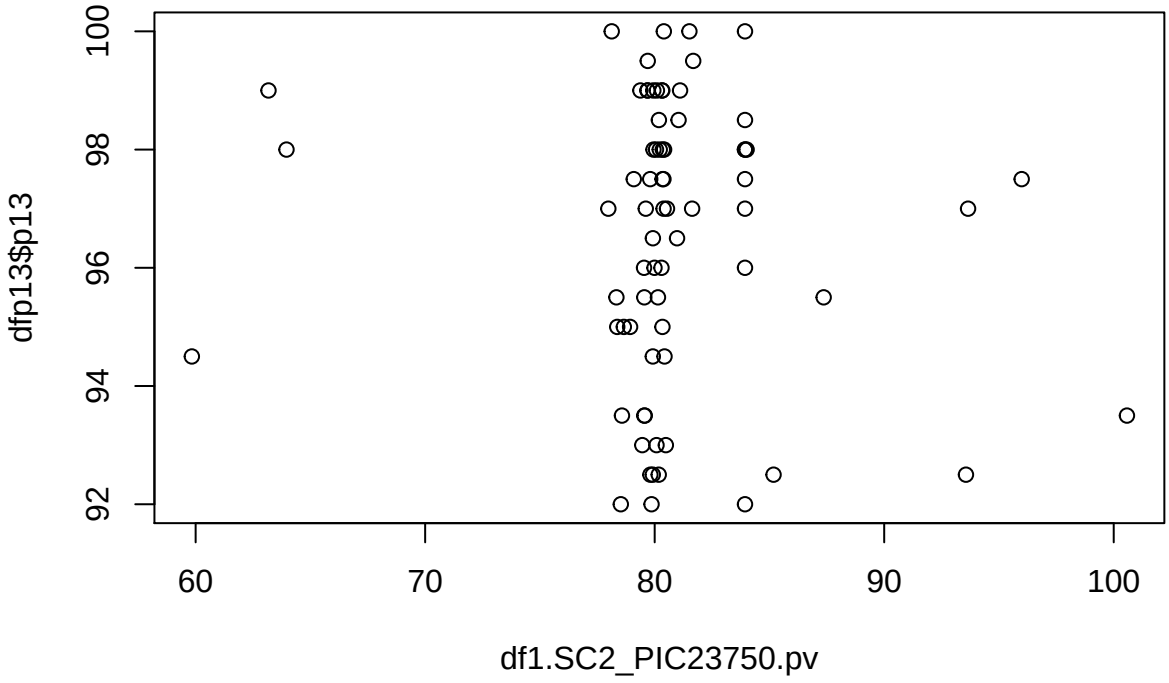

dfp13$p13
100
98
96
94
92
10.0
10.5
11.0
11.5
12.0
df1.SC2_FIC20759.pv

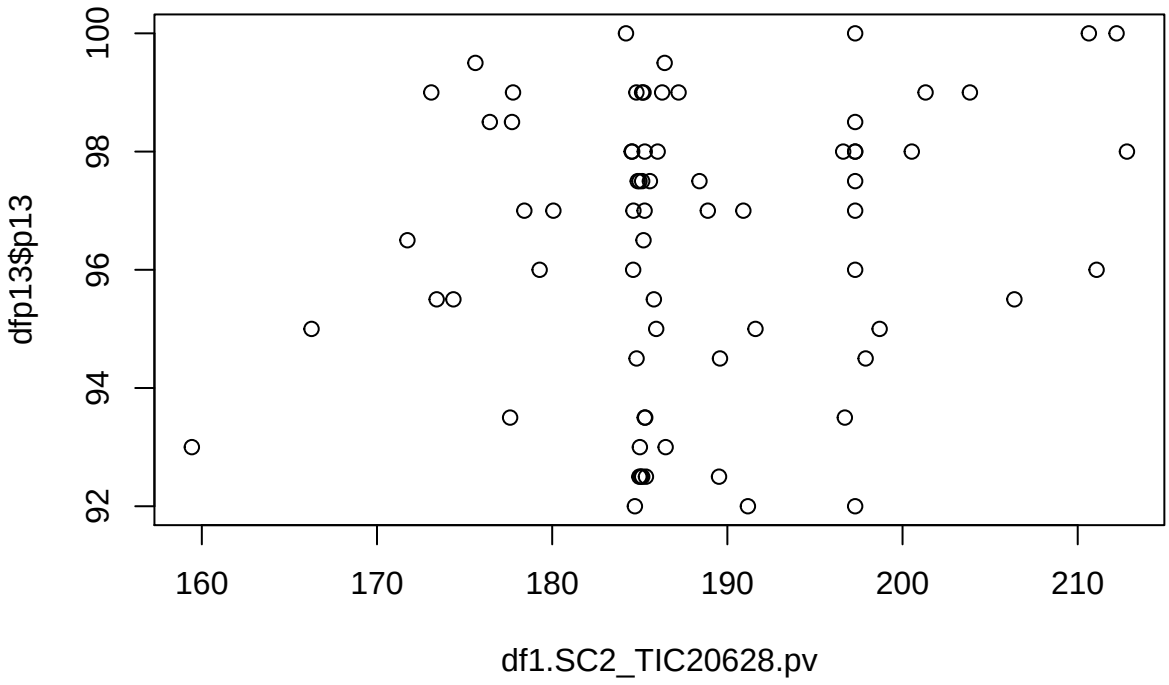

dfp13$p13
100
98
96
94
92
36
38
40
42
df1.SC2_PIC20065.pv

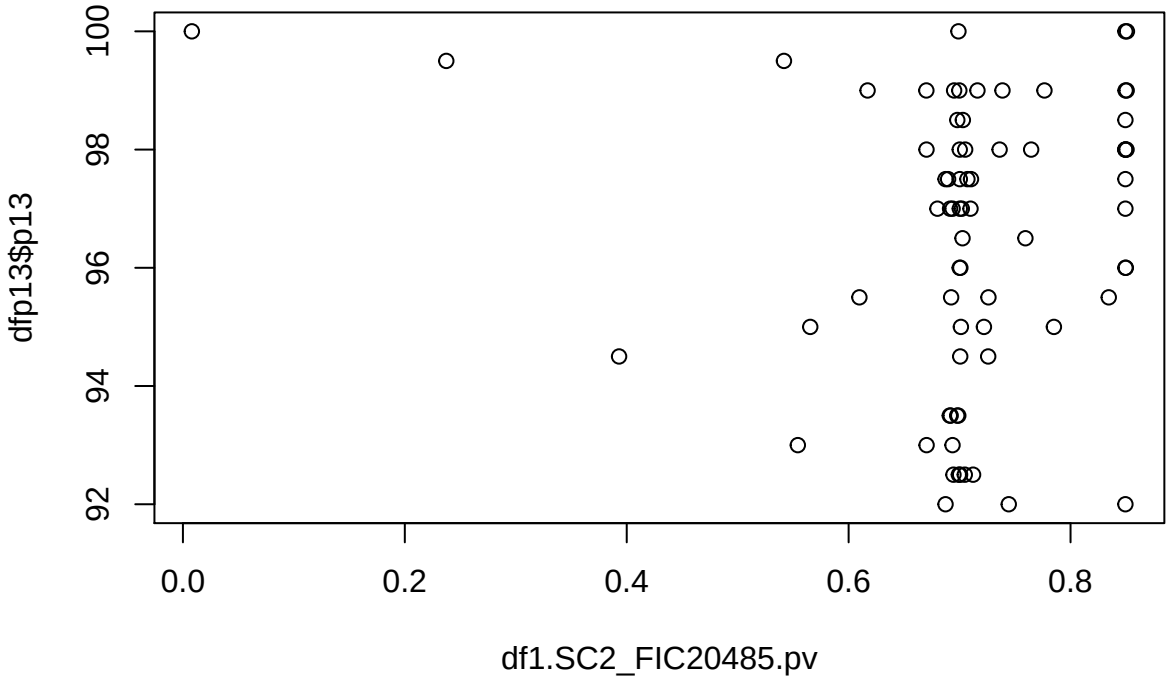

dfp13$p13
100
98
96
94
92
30
40
50
60
df1.SC2_TIC20760.pv

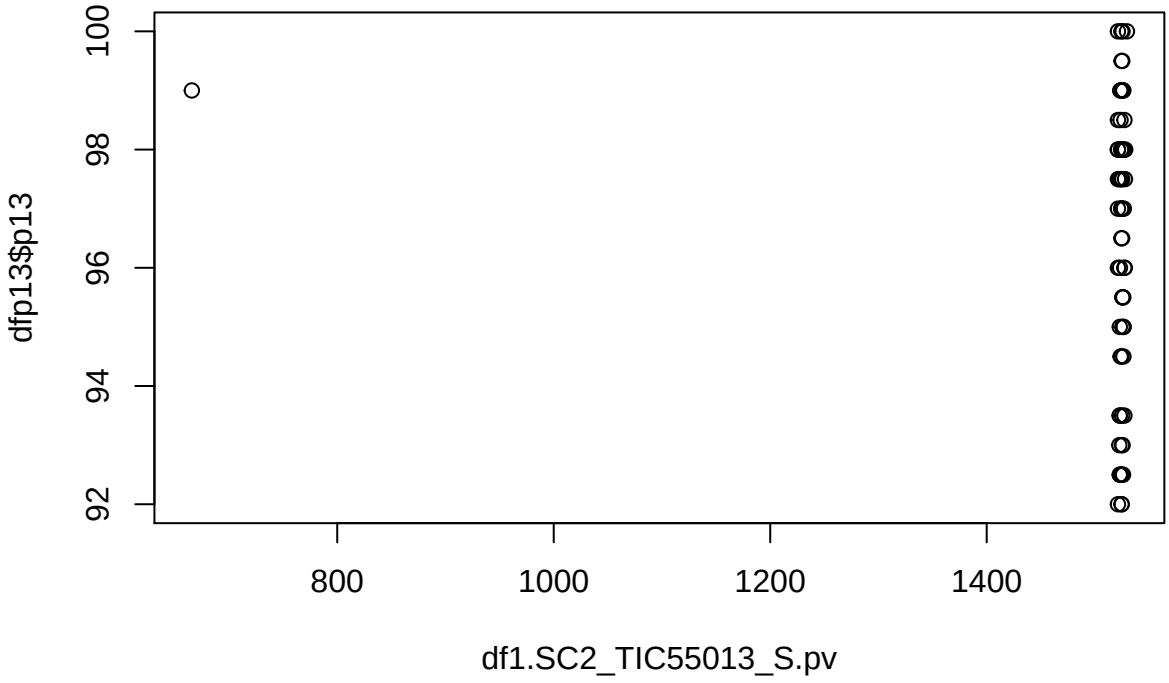

dfp13$p13
100
98
96
94
92
60
65
70
75
80
85
90
df1.SC2_TIC21178.pv

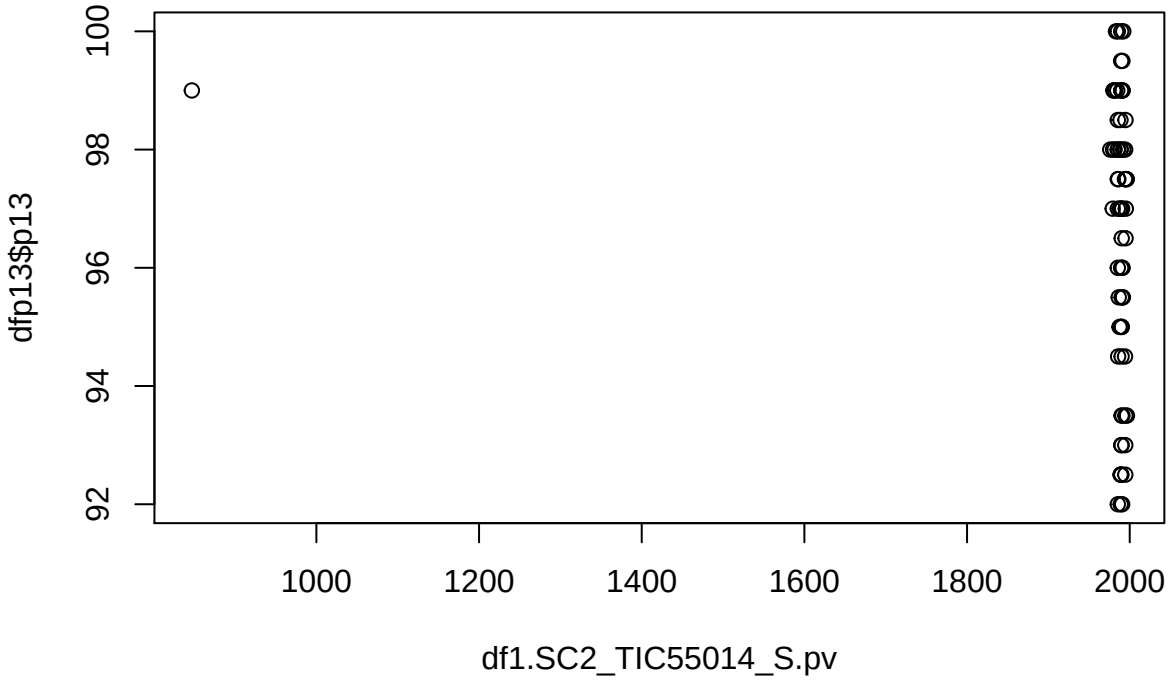

dfp13$p13
100
98
96
94
92
15
20
25
30
35
df1.SC2_FIC20476.pv

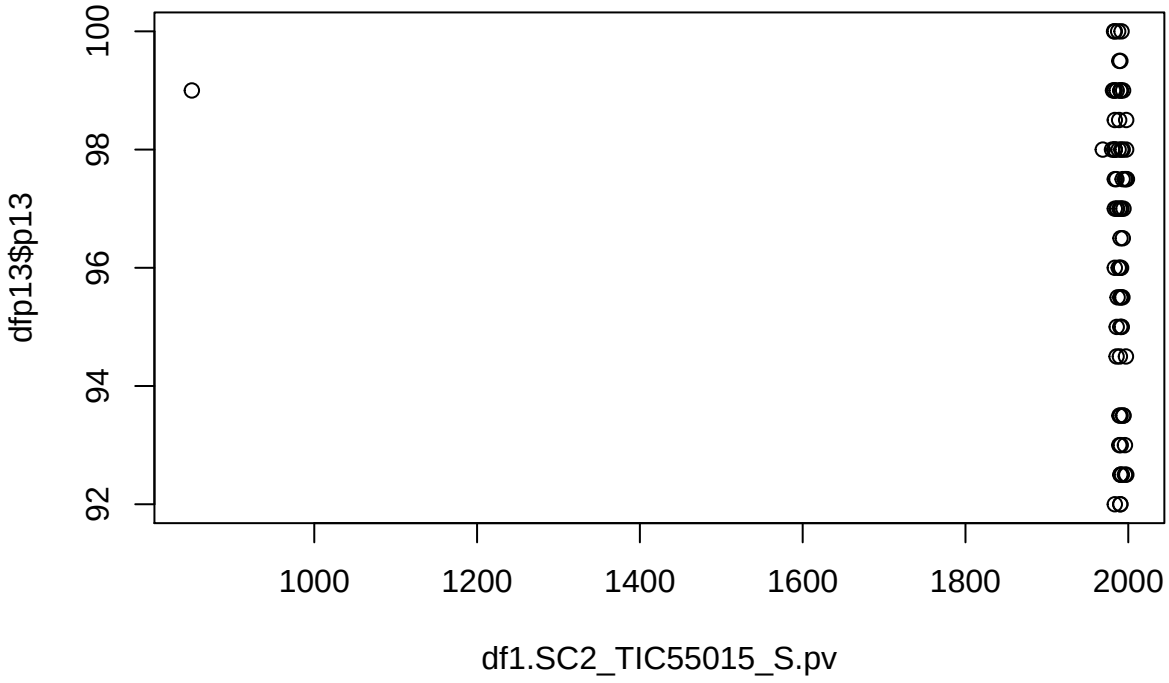

dfp13$p13
100
98
96
94
92
60
70
80
90
100
df1.SC2_PIC23750.pv

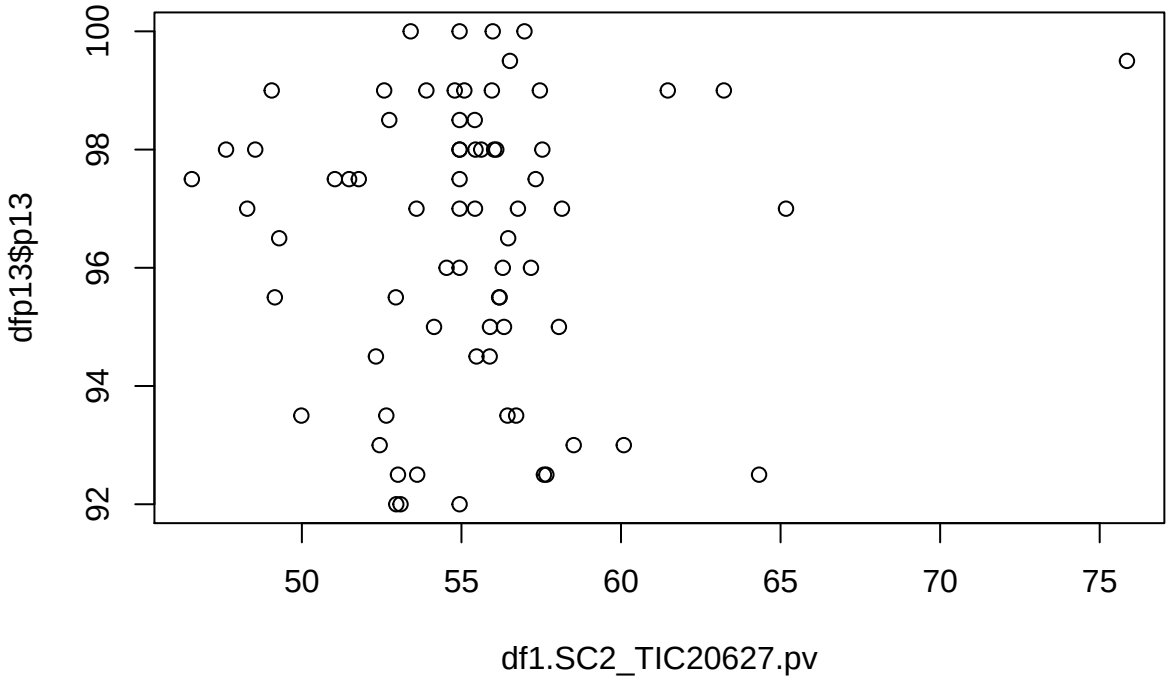

dfp13$p13
100
98
96
94
92
160
170
180
190
200
210
df1.SC2_TIC20628.pv

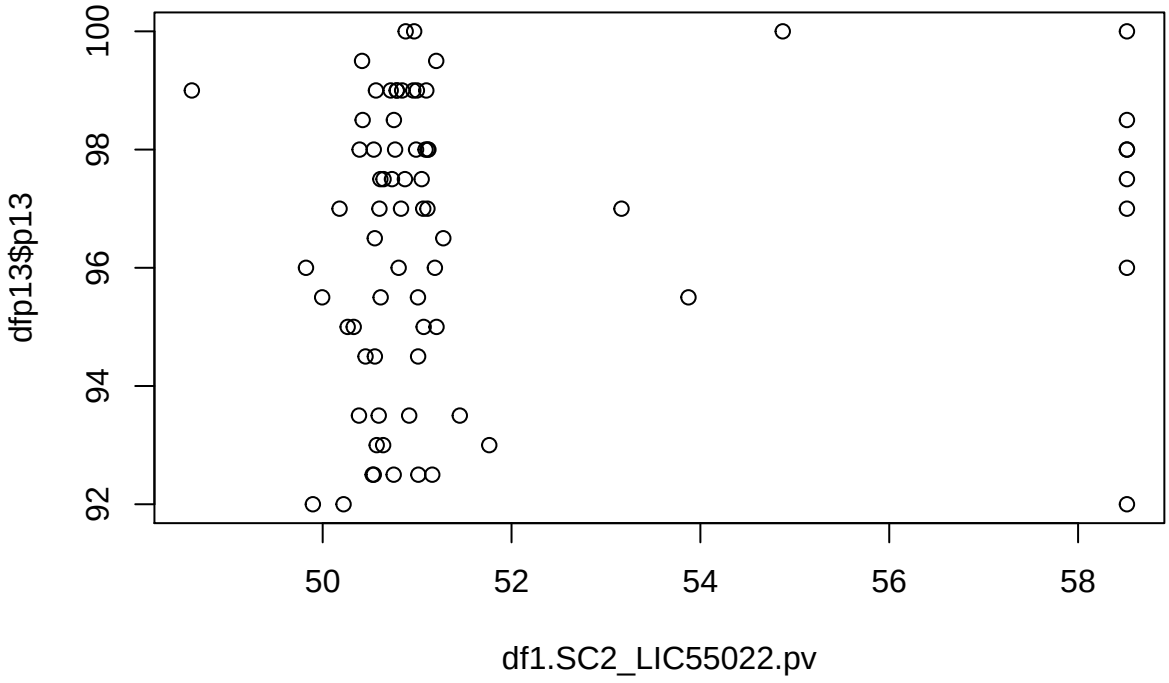

dfp13$p13
100
98
96
94
92
0.0
0.2
0.4
0.6
0.8
df1.SC2_FIC20485.pv

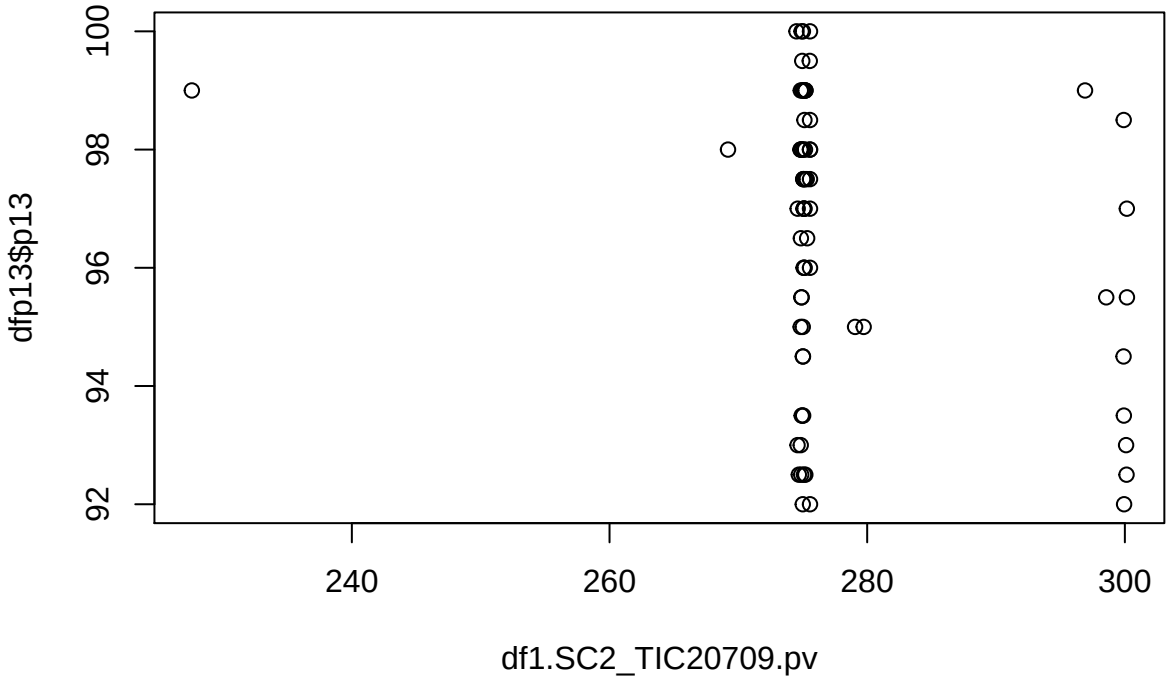

dfp13$p13
100
98
96
94
92
800
1000
1200
1400
df1.SC2_TIC55013_S.pv

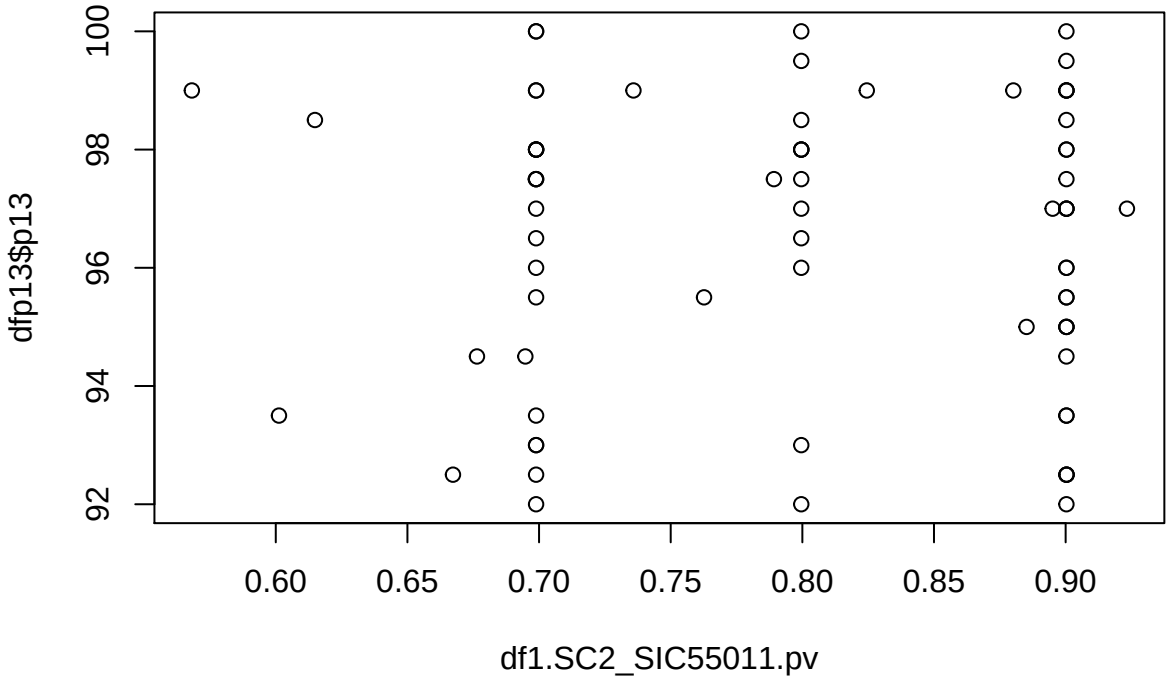

dfp13$p13
100
98
96
94
92
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55014_S.pv


dfp13$p13
100
98
96
94
92
1000
1200
1400
1600
1800
2000
df1.SC2_TIC55015_S.pv


dfp13$p13
100
98
96
94
92
50
55
60
65
70
75
df1.SC2_TIC20627.pv


dfp13$p13
100
98
96
94
92
50
52
54
56
58
df1.SC2_LIC55022.pv


dfp13$p13
100
98
96
94
92
240
260
280
300
df1.SC2_TIC20709.pv


dfp13$p13
100
98
96
94
92
0.60
0.65
0.70
0.75
0.80
0.85
0.90
df1.SC2_SIC55011.pv