Galaxiid Pilot Analyses

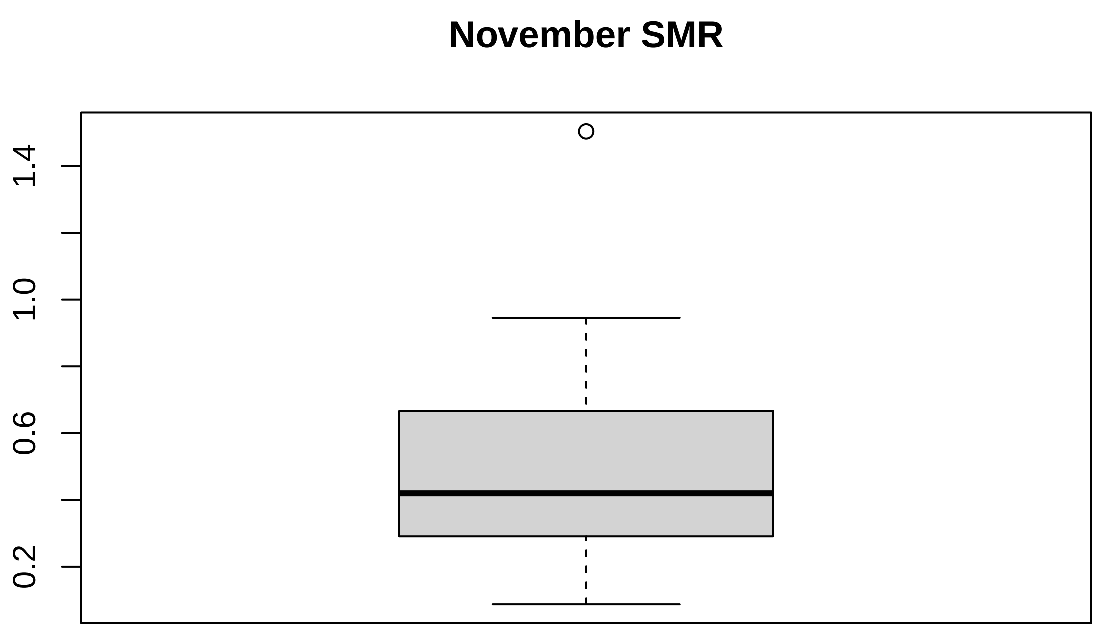
2024-02-06

Pilot Study Analysis

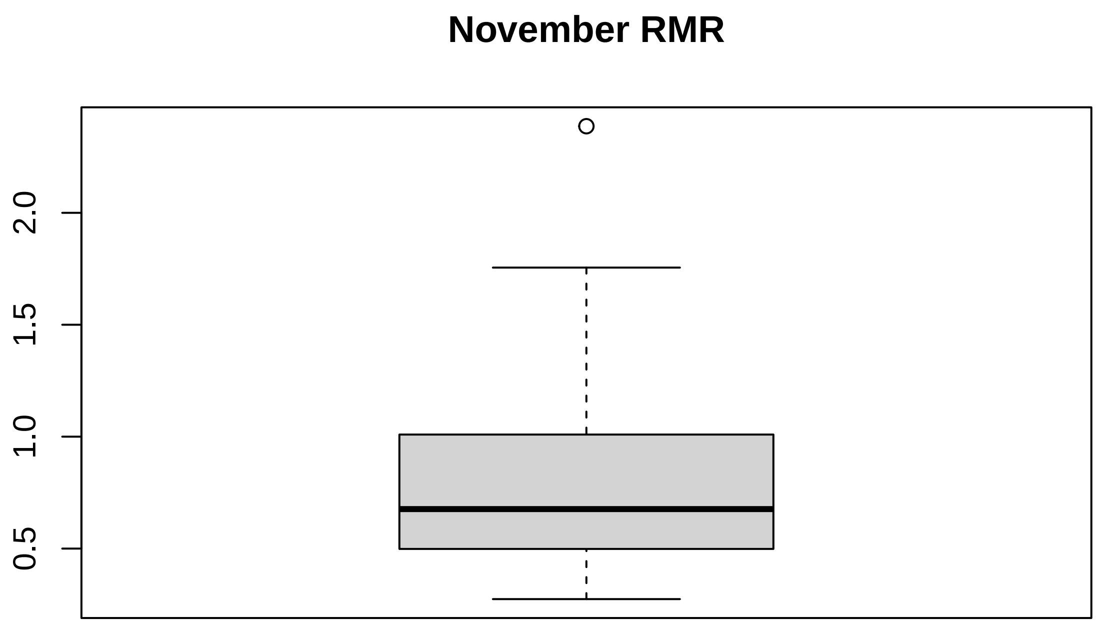
MR outlier analysis

Fish 23 appears to be an outlier in the Nov. SMR and RMR measurements. Let's test if it really is, and see how influential it might be for our analysis:

```
setwd("C:/Users/s222141733/OneDrive - Deakin University/Beth Hoots/Growth Performance Pilot/Results/Pro

library(ggplot2)
library(readxl)
library(ggpubr)

data <- read_excel("Respirometry Corrected BR2.xlsx",
                   sheet = "Growth, MR, Behavior")

boxplot(data$Nov_SMR, main="November SMR")
```



```r
boxplot(data$Nov_RMR, main="November RMR")
```



These boxplots suggest that we have one outlier in our November SMR and RMR data. We can validate that with a Bonferroni Outlier Test:

```r
library(car)
```

```
## Loading required package: carData
```

```r
model_nSMR <- lm(Nov_SMR ~ Nov, data = data)
outlierTest(model_nSMR)
```

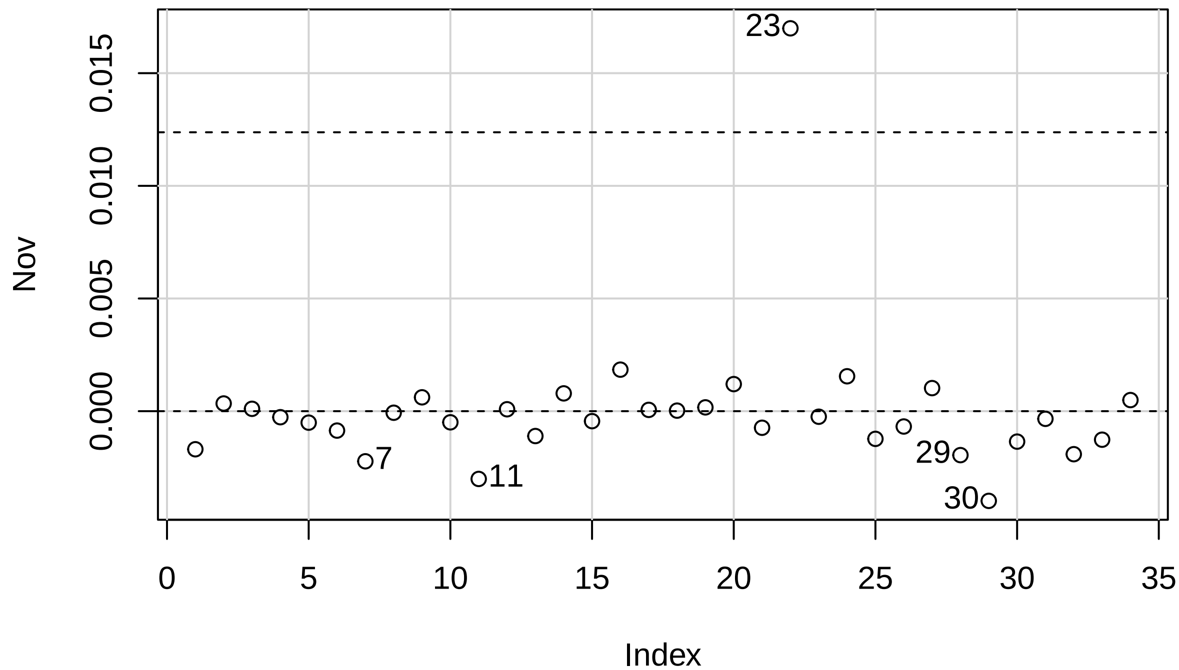
```
##    rstudent unadjusted p-value Bonferroni p
## 23 5.256632         1.0289e-05   0.00034983
```

```r
model_nRMR <- lm(Nov_RMR ~ Nov, data = data)
outlierTest(model_nRMR)
```

```
##    rstudent unadjusted p-value Bonferroni p
## 23 4.963529         2.3844e-05   0.00081068
```

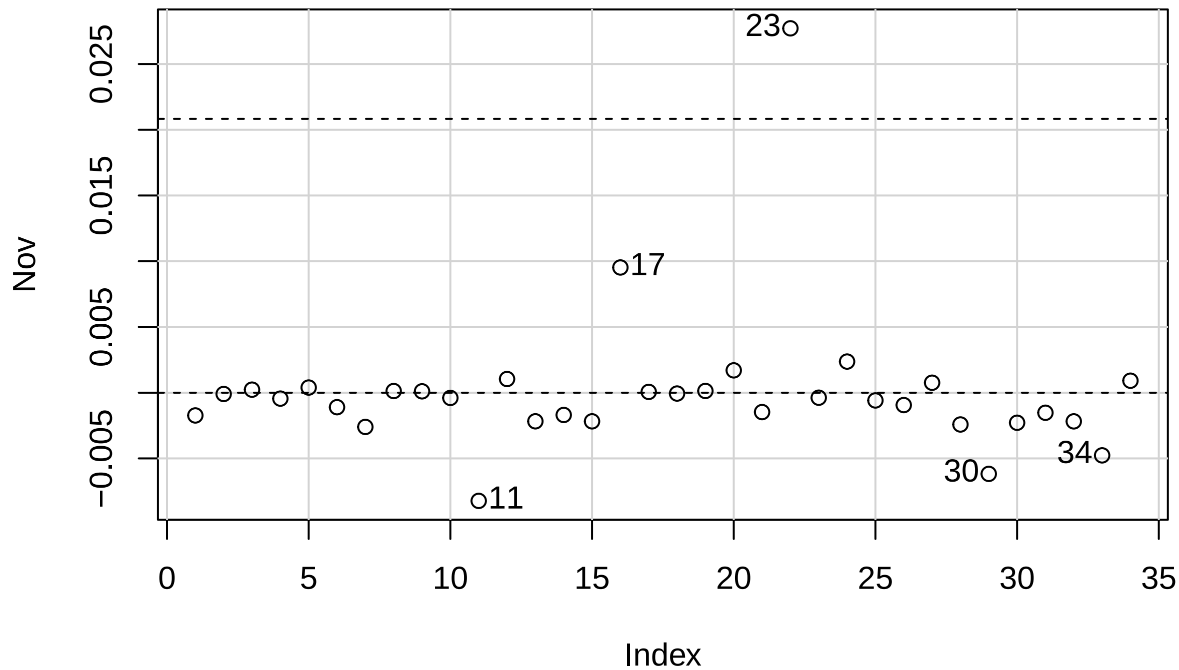
This is the point we suspected was an outlier, fish #23 (found in row 23, as highlighted in the above tables). What this test has done is look at the linear model between each of our two MR calculations for November (not mass standardized) and Nov (mass data), and then generate the Studentized residuals for each real value compared with a t-test to the model. It found that row 23 (fish #23) had a significantly high residual, with p < 0.0001 for both tests.

```r
plotdb <- dfbetaPlots(model_nSMR,id.n=5)
```



This is a Dfbeta plot for the SMR model, which shows the observations that are most influential on our regression model. The y-axis shows how much the regression coefficient changes when each individual observation is left out of the model. Again, we see point fish #23 far away from the rest of the data.

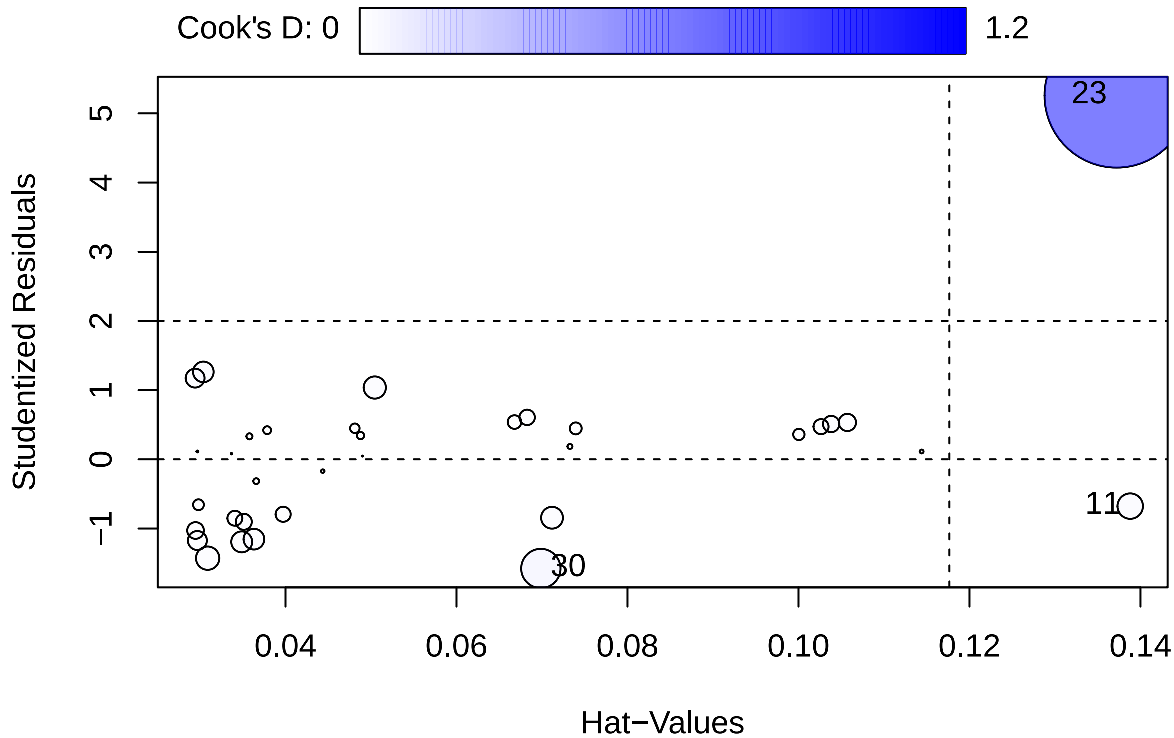
```
plotdb <- dfbetaPlots(model_nRMR,id.n=5)
```



Here again we see fish #23 as an outlier in our RMR model, but it's joined by fish #17 and #11 to a lesser extent. Fish #11 took a long time to settle in the chamber, based on the raw trace, and fish #17 really spiked when the lights came on. These look less concerning to me here, but we can look at how influential they all were.

Starting again with SMR, let's look at the influence plots:

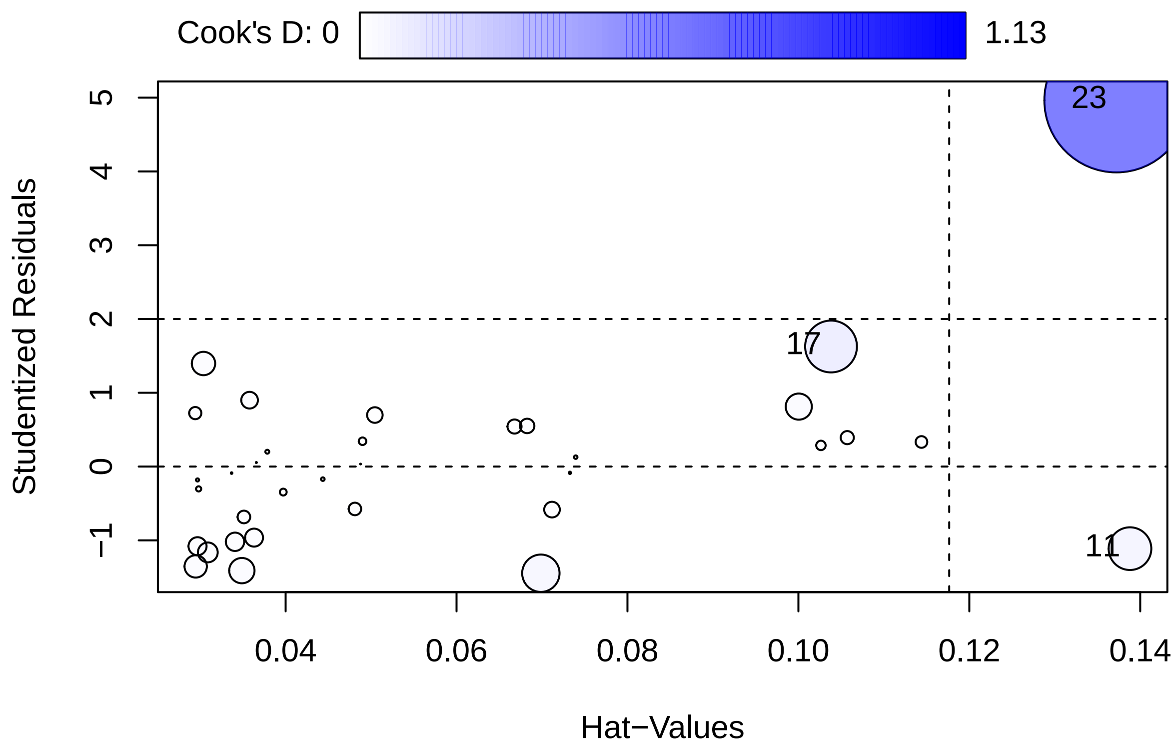
```
influencePlot(model_nSMR)
```




```
##       StudRes        Hat      CookD
## 11 -0.6759965 0.13879933 0.03746062
## 23  5.2566319 0.13720769 1.19914383
## 30 -1.5774326 0.06986601 0.08929950
```

This is interesting because it highlights 11 and 30 as additional points of concern. Still, 23 is definitely the biggest concern though, having far and away the highest Cook's distance. The rule of thumb for Cook's distance is any value more than 4/n, which is 0.11 for this study, is very influential. For Studentized Residuals, the rule is anything larger than 3 is an outlier. Taken together, these tests help us identify fish #23 as a highly influential outlier.

```
influencePlot(model_nRMR)
```



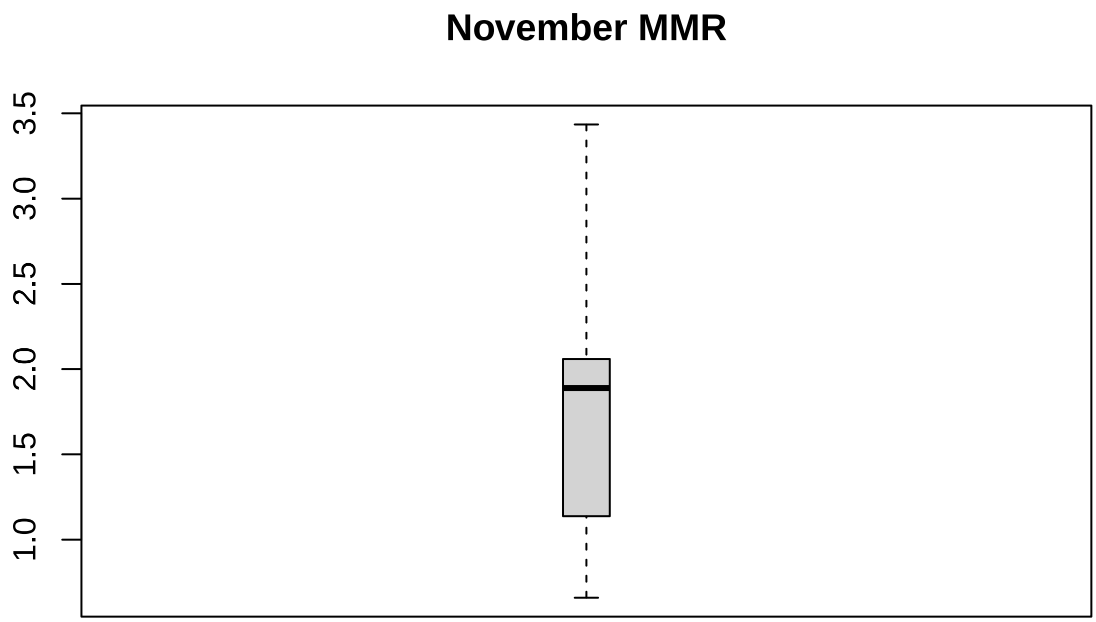

```
##      StudRes       Hat      CookD
## 11 -1.112658 0.1387993 0.09902799
## 17  1.627307 0.1038194 0.14587517
## 23  4.963529 0.1372077 1.12671109
```

Looking at our RMR model, we can quickly see that fish #23 is once again a highly influential outlier. At this point, with the evidence we've just seen, I am comfortable leaving this one point out of our analysis for the month of November.

This also flags fish #17 as a just-barely influential point, but it isn't considered an outlier by the Studentized Residuals metric. I will not exclude this point, but it will be good to keep this in mind as we continue analyses.

Let's also check through the other month and MR measurements:

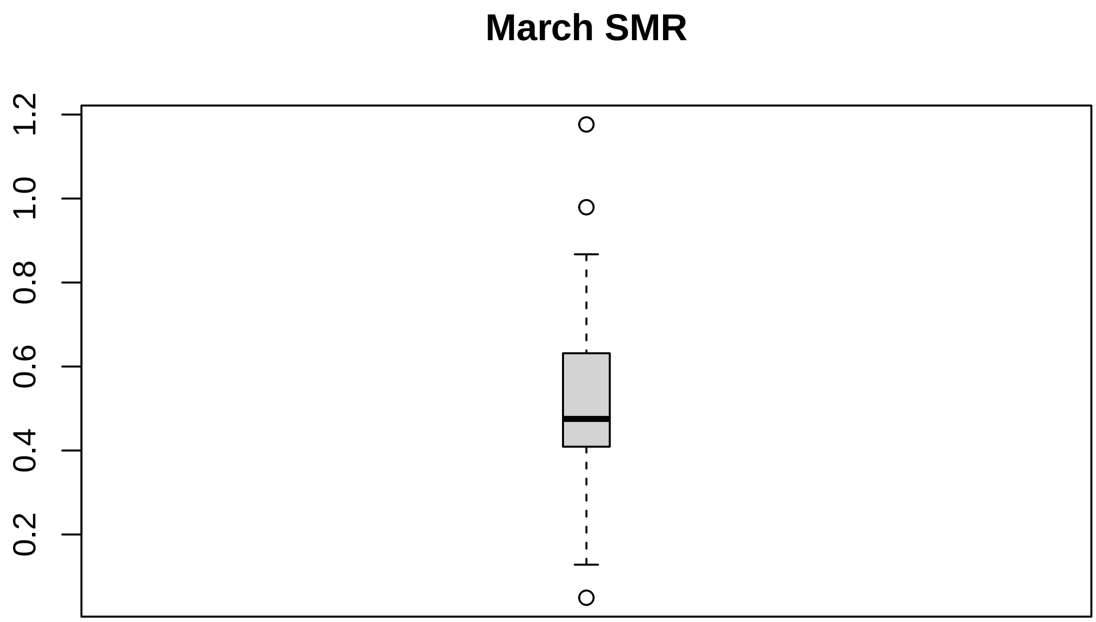
```
boxplot(data$Nov_MMR, main="November MMR", boxwex=0.1)
```




```r
model_nMMR <- lm(Nov_MMR ~ Nov, data = data)
outlierTest(model_nMMR)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferroni p
## 17 2.872993          0.0072767      0.24741
```

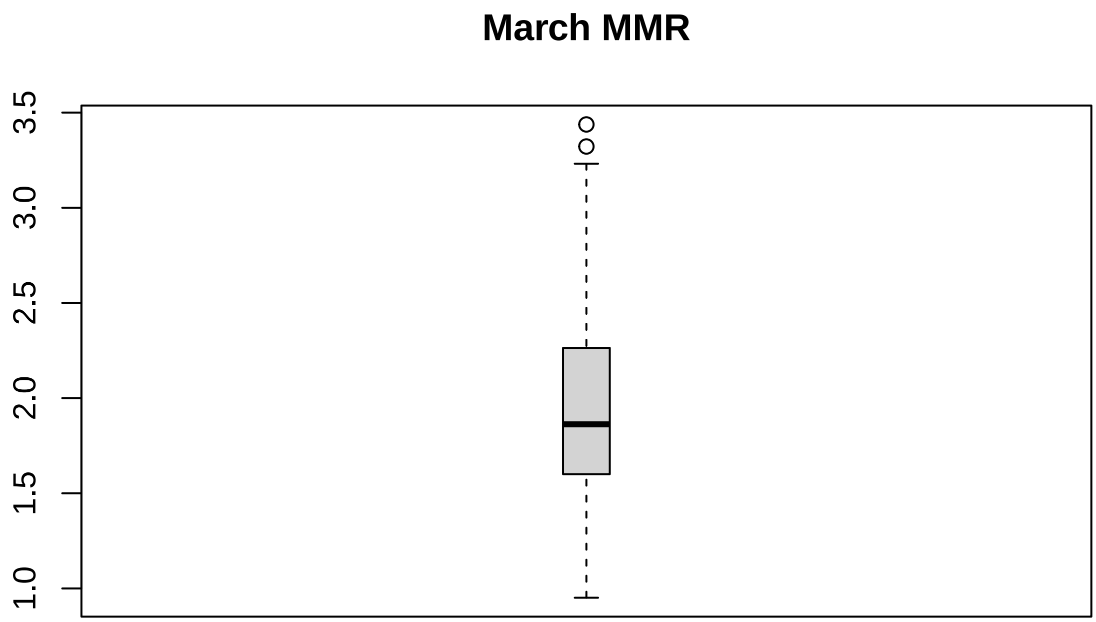
```r
boxplot(data$Mar_SMR, main="March SMR", boxwex=0.1)
```




```r
model_mSMR <- lm(Mar_SMR ~ Mar, data = data)
outlierTest(model_mSMR)
```

```
##    rstudent unadjusted p-value Bonferroni p
## 25 3.749471         0.00078631     0.025162
```

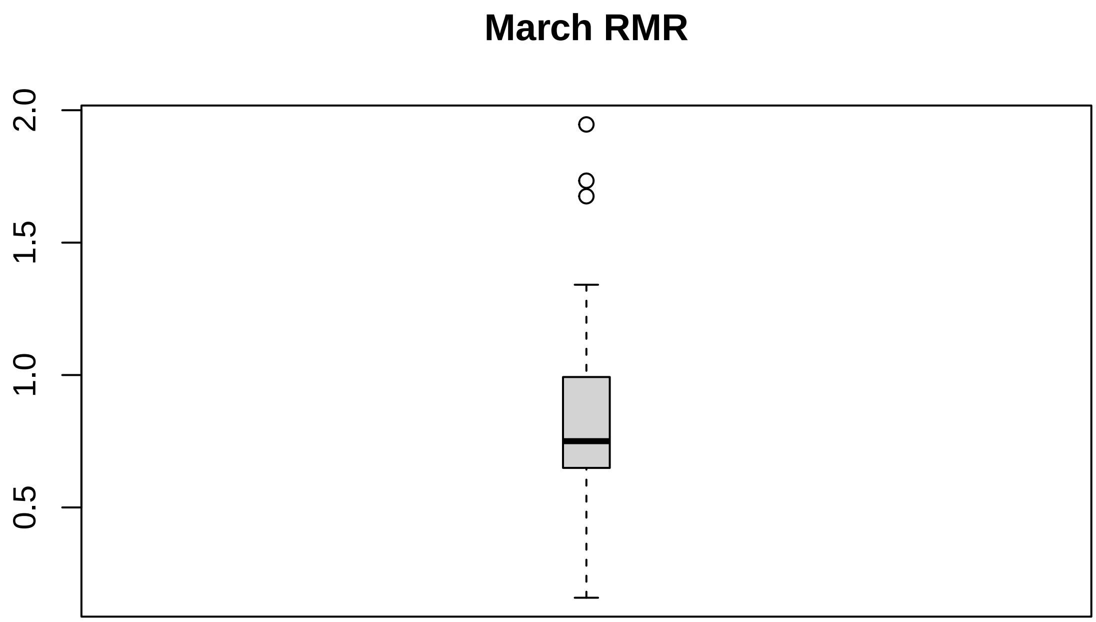
```r
boxplot(data$Mar_MMR, main="March MMR", boxwex=0.1)
```




```r
model_mMMR <- lm(Mar_MMR ~ Mar, data = data)
outlierTest(model_mMMR)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##   rstudent unadjusted p-value Bonferroni p
## 7 -2.08783           0.045707           NA
```

```r
boxplot(data$Mar_RMR, main="March RMR", boxwex=0.1)
```

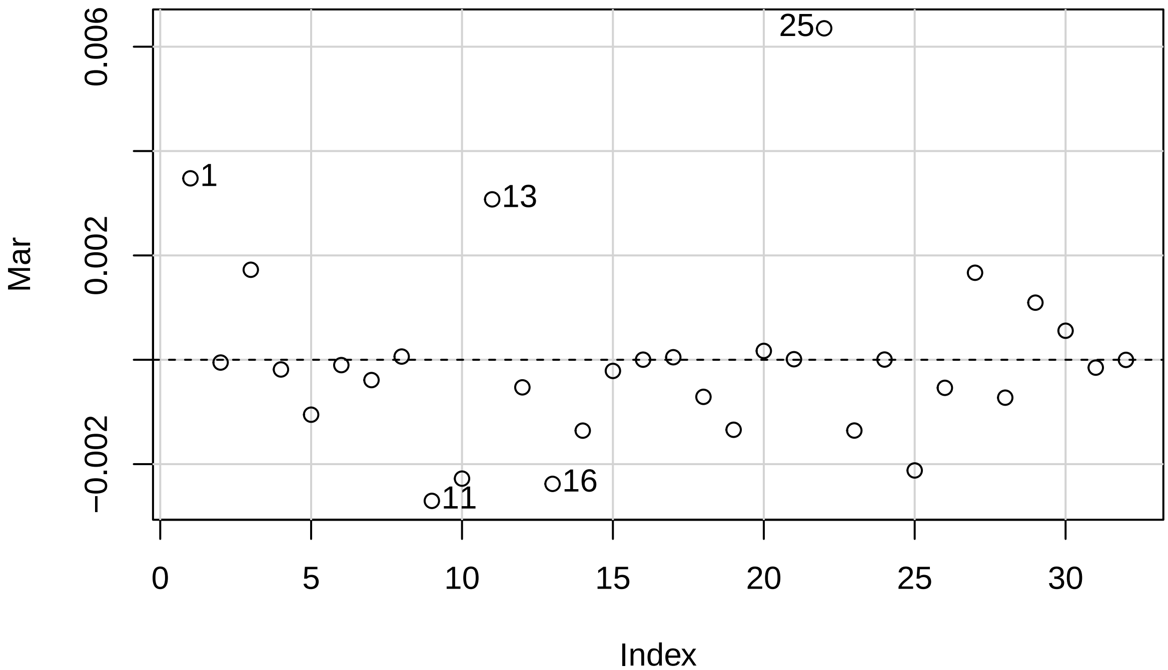



```
model_mRMR <- lm(Mar_RMR ~ Mar, data = data)
outlierTest(model_mRMR)
```

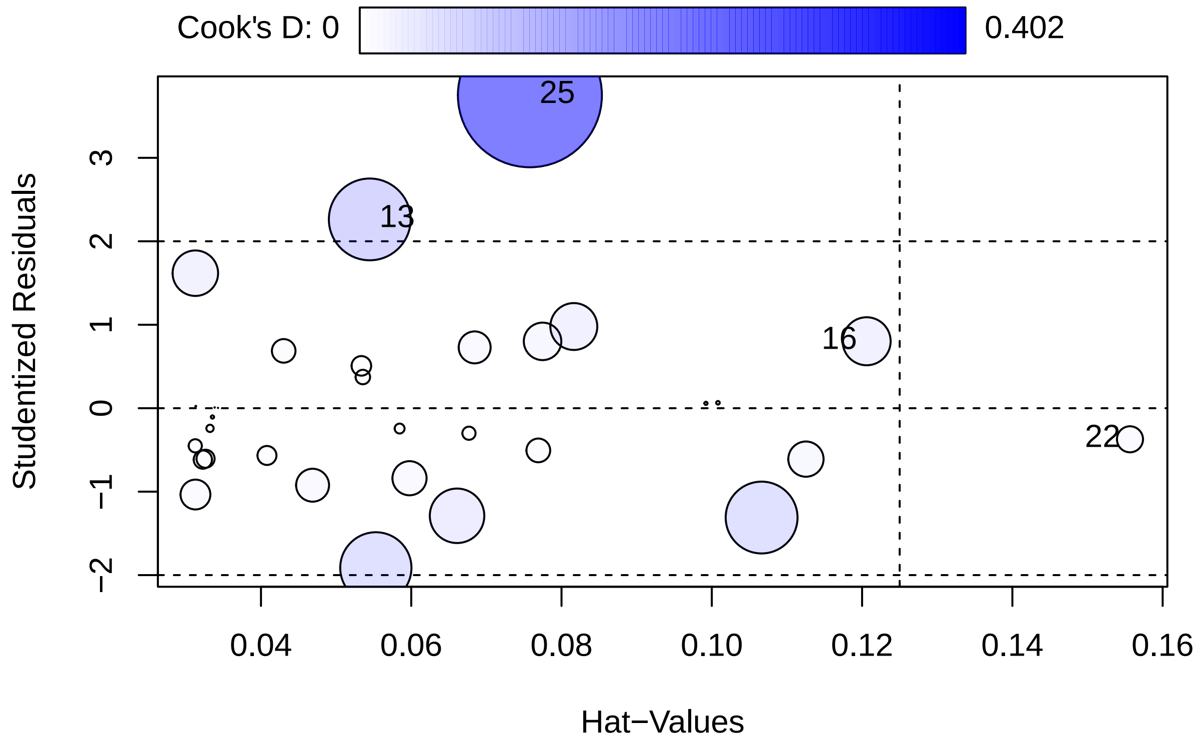
```
##    rstudent unadjusted p-value Bonferroni p
## 25 3.587772          0.0012094       0.0387
```

This looks like MMR models are outlier free, but our March SMR and RMR models might have a slight outlier in fish #25.

```
plotdb <- dfbetaPlots(model_mSMR,id.n=5)
```

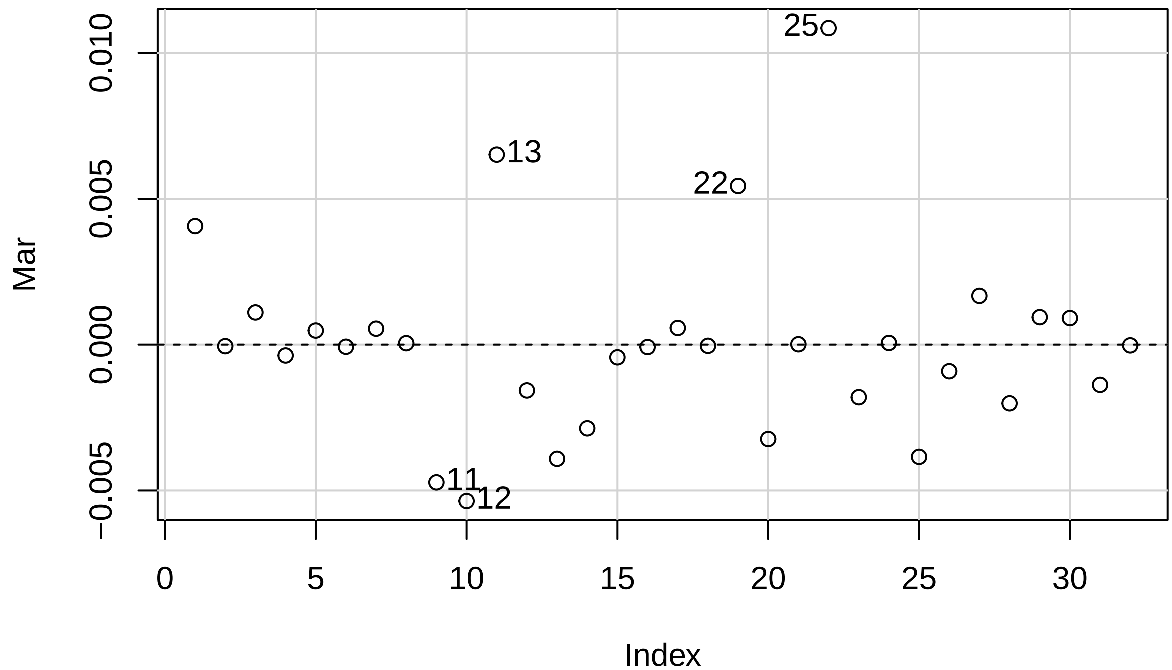



```r
influencePlot(model_mSMR)
```

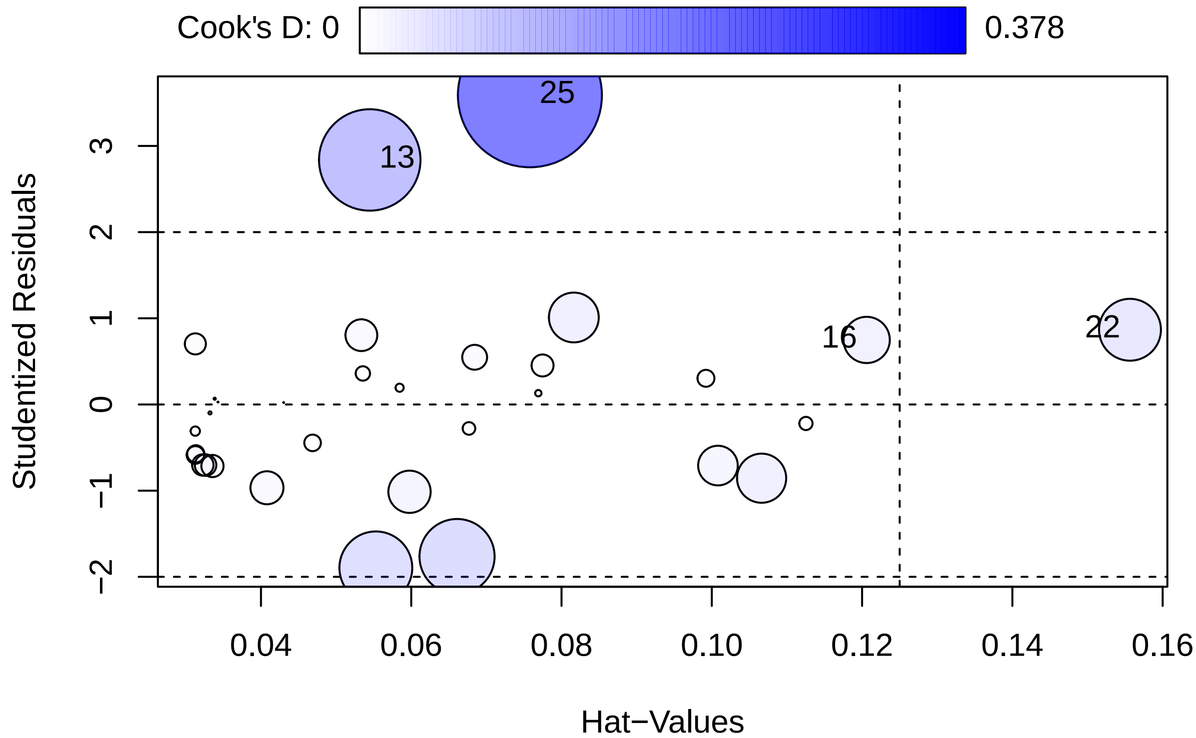



```
##        StudRes        Hat     CookD
## 13  2.2622365 0.05447803 0.12963944
## 16  0.8021290 0.12060686 0.04465195
## 22 -0.3725475 0.15565248 0.01317098
## 25  3.7494711 0.07579121 0.40162505
```

```
plotdb <- dfbetaPlots(model_mRMR,id.n=5)
```




```
influencePlot(model_mRMR)
```



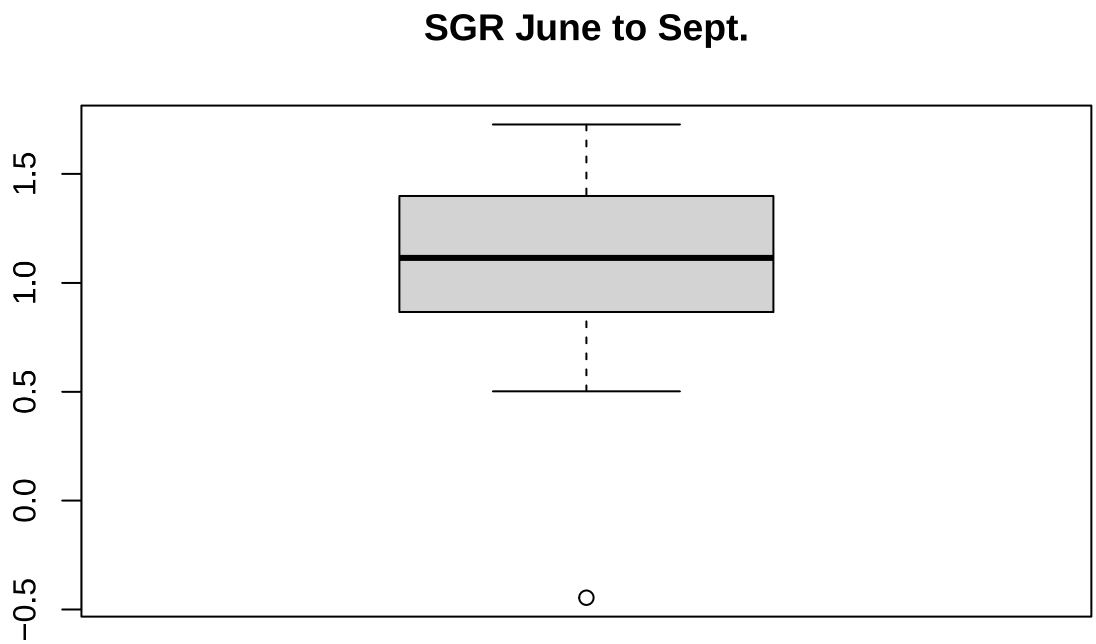

```
##       StudRes        Hat      CookD
## 13 2.8375370 0.05447803 0.18780925
## 16 0.7490504 0.12060686 0.03904649
## 22 0.8671179 0.15565248 0.06988238
## 25 3.5877724 0.07579121 0.37815073
```

Based on this plot, fish #25 is also a highly influential outlier for our purposes. It is not as influential as fish #23 was in the previous month, but it is still worth looking more closely at. From the raw slopes, it looks like fish #23 had a spike in the middle of the night, after initially settling down in the chamber. This may have prevented it from reaching a true SMR, and certainly would contribute to an increased RMR. I do not know if I would exclude it from the RMR analysis based just on this, as it might be a more active fish in the bins in general, but it will be worth keeping an eye on. This might be a good enought reason to exclude it from the SMR analysis.

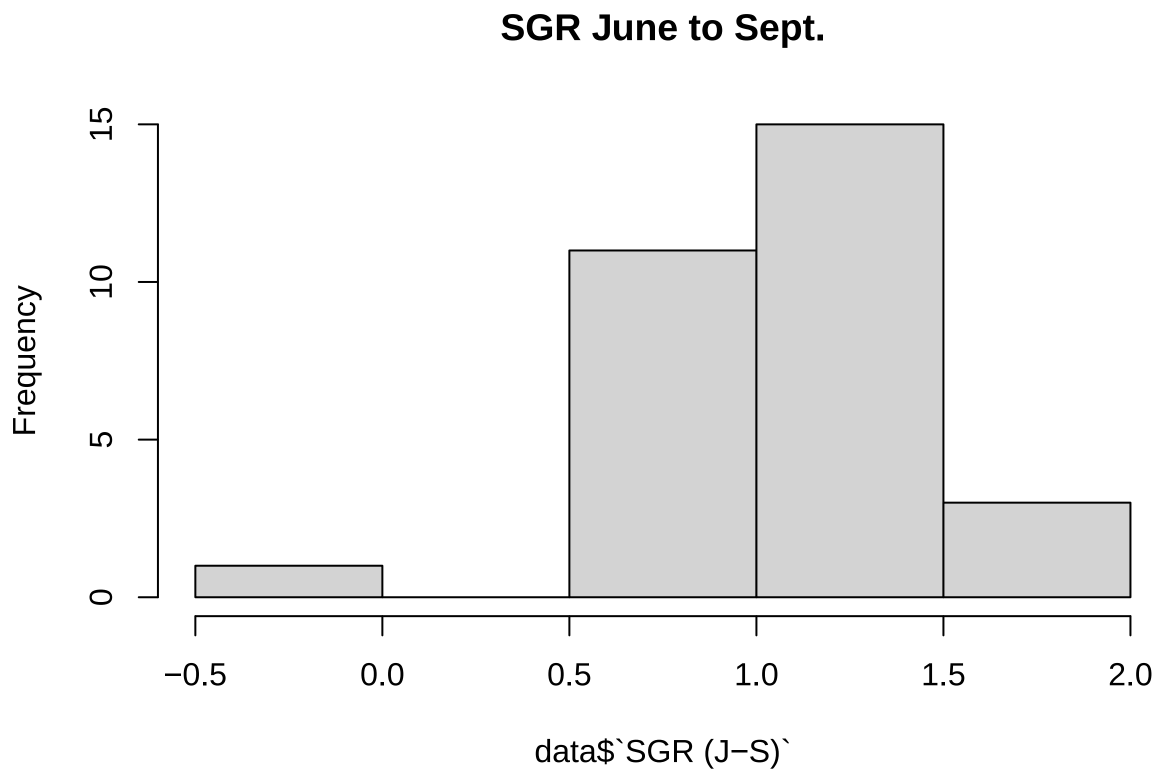
SGR Outlier Analysis

It looks like Fish 29 may be an influential outlier point as well, but for its November SGR calculation rather than its MO2 rates. Let's run the same analyses on our SGR calculations for each time step:

```
boxplot(data$`SGR (J-S)`, main="SGR June to Sept.")
```



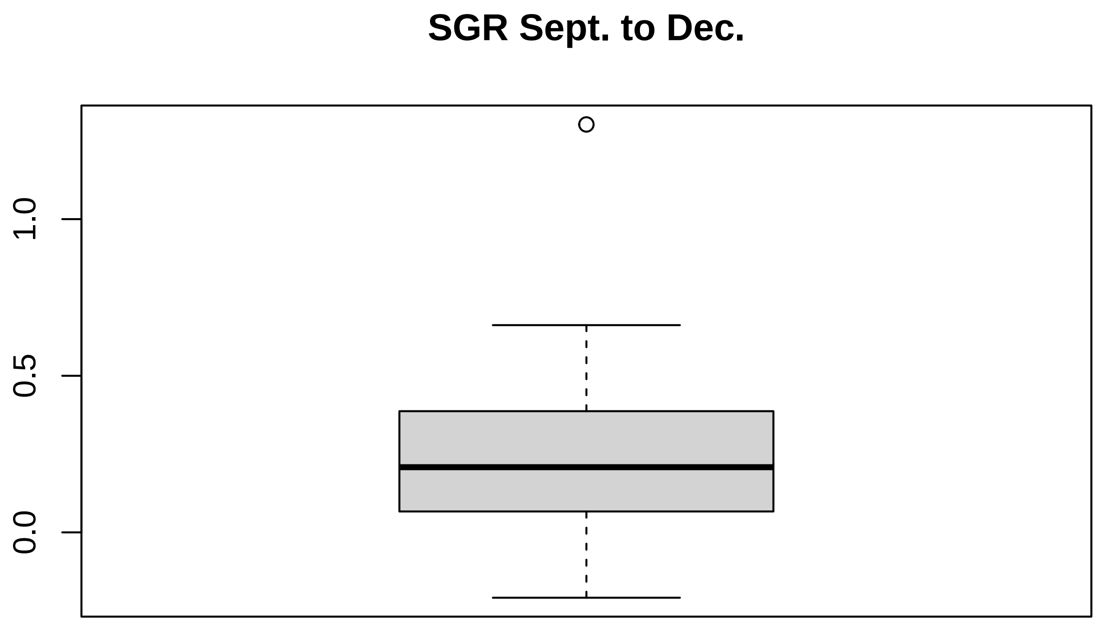

```r
hist(data$`SGR (J-S)`, breaks = sqrt(nrow(data)), main="SGR June to Sept.")
```



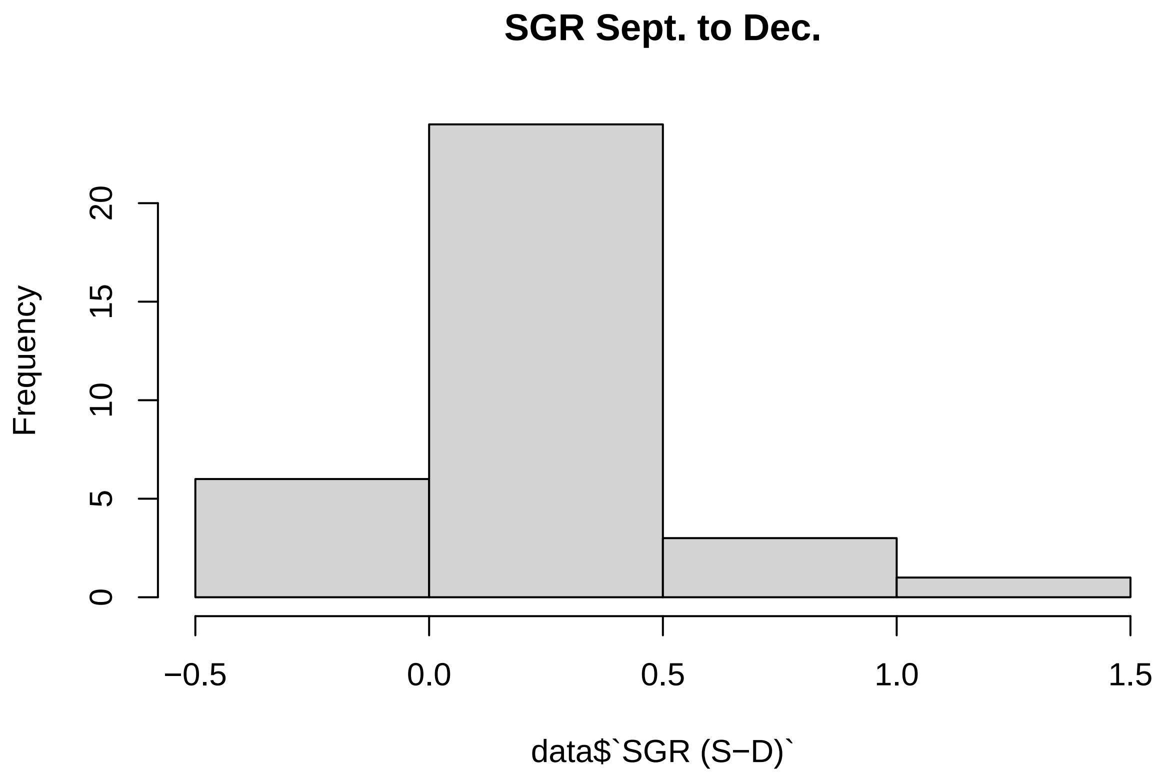

```r
boxplot.stats(data$`SGR (J-S)`)$out
```

```
## [1] -0.4458453
```

```r
boxplot(data$`SGR (S-D)`, main="SGR Sept. to Dec.")
```



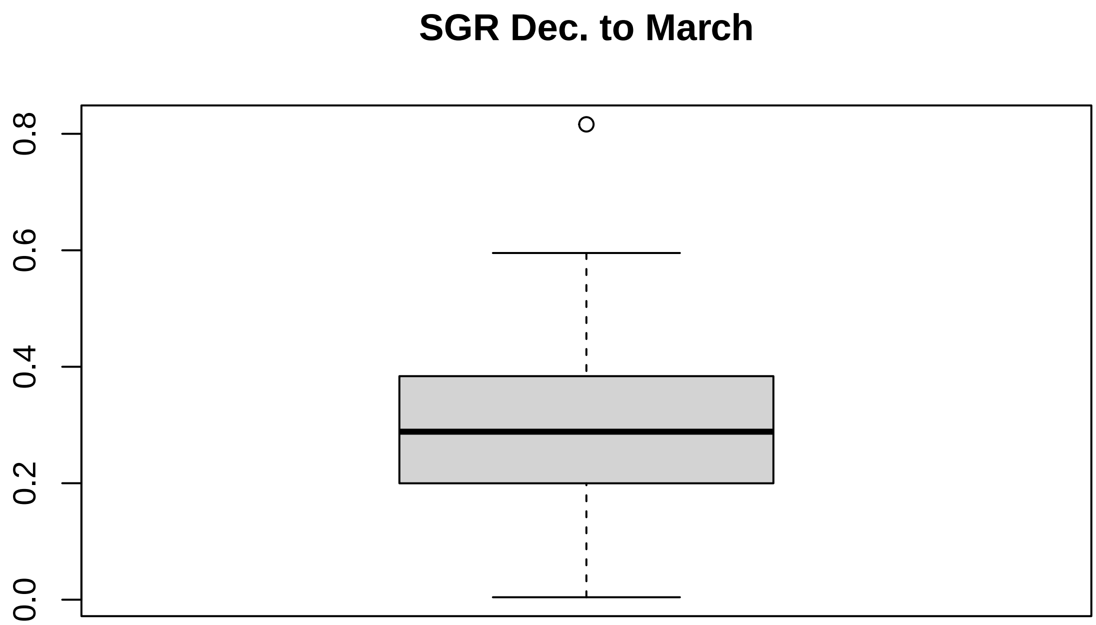

```r
hist(data$`SGR (S-D)`, breaks = sqrt(nrow(data)), main="SGR Sept. to Dec.")
```



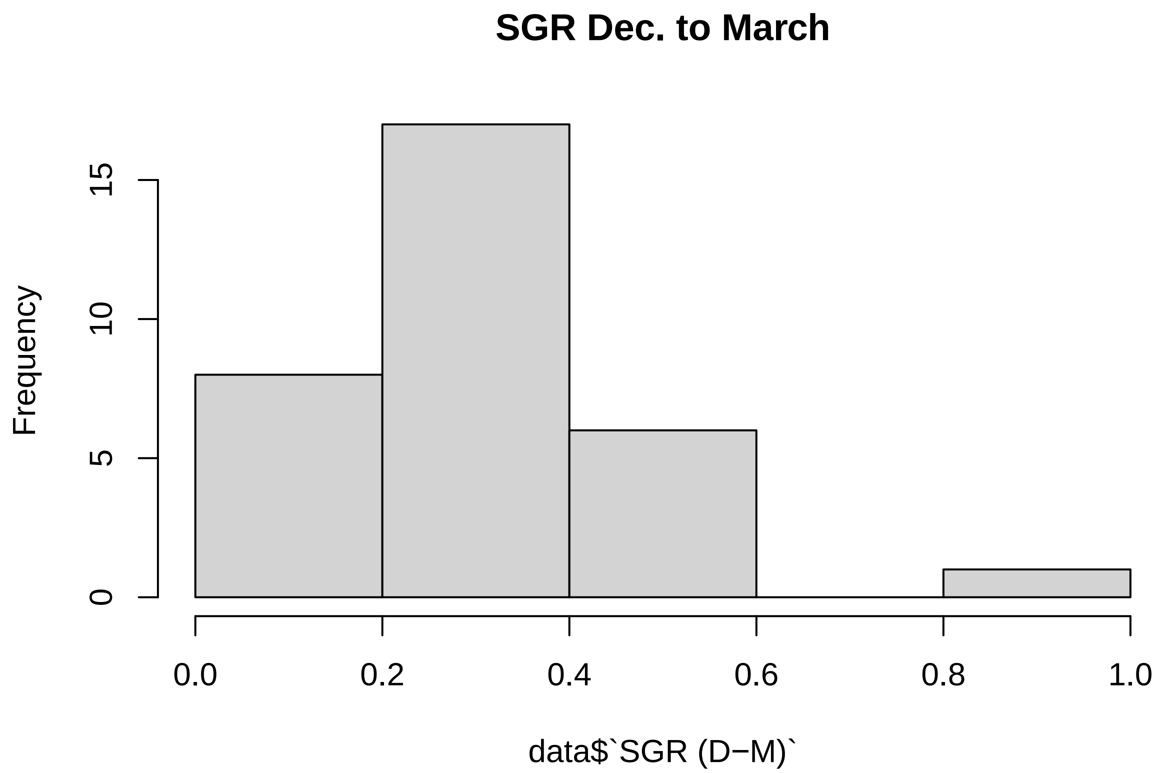

```
boxplot.stats(data$`SGR (S-D)`)$out
```

```
## [1] 1.302293
```

```
boxplot(data$`SGR (D-M)`, main="SGR Dec. to March")
```




```
hist(data$`SGR (D-M)`, breaks = sqrt(nrow(data)), main="SGR Dec. to March")
```

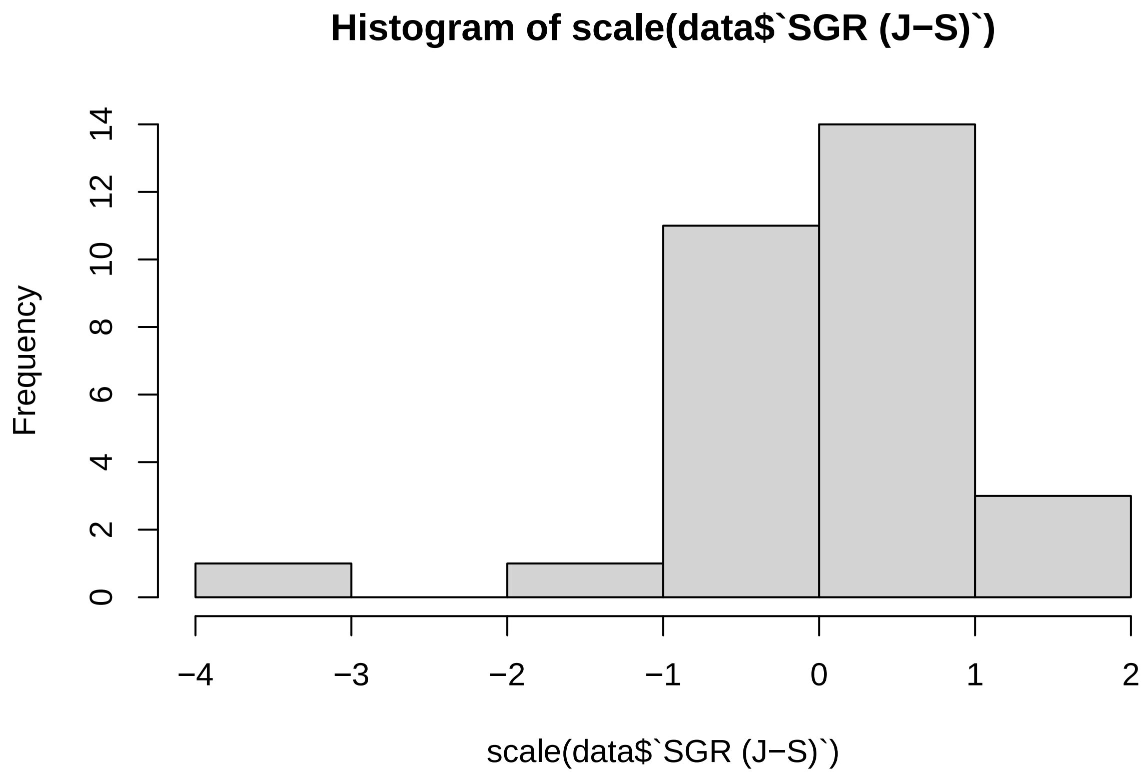



```
boxplot.stats(data$`SGR (D-M)`)$out
```

```
## [1] 0.8162421
```

Each of these time intervals has one SGR value that looks like it could be an outlier based on the IQR criterion. However, with this small dataset, the IQR might be too selective. Let's find their z-scores and work with those instead.

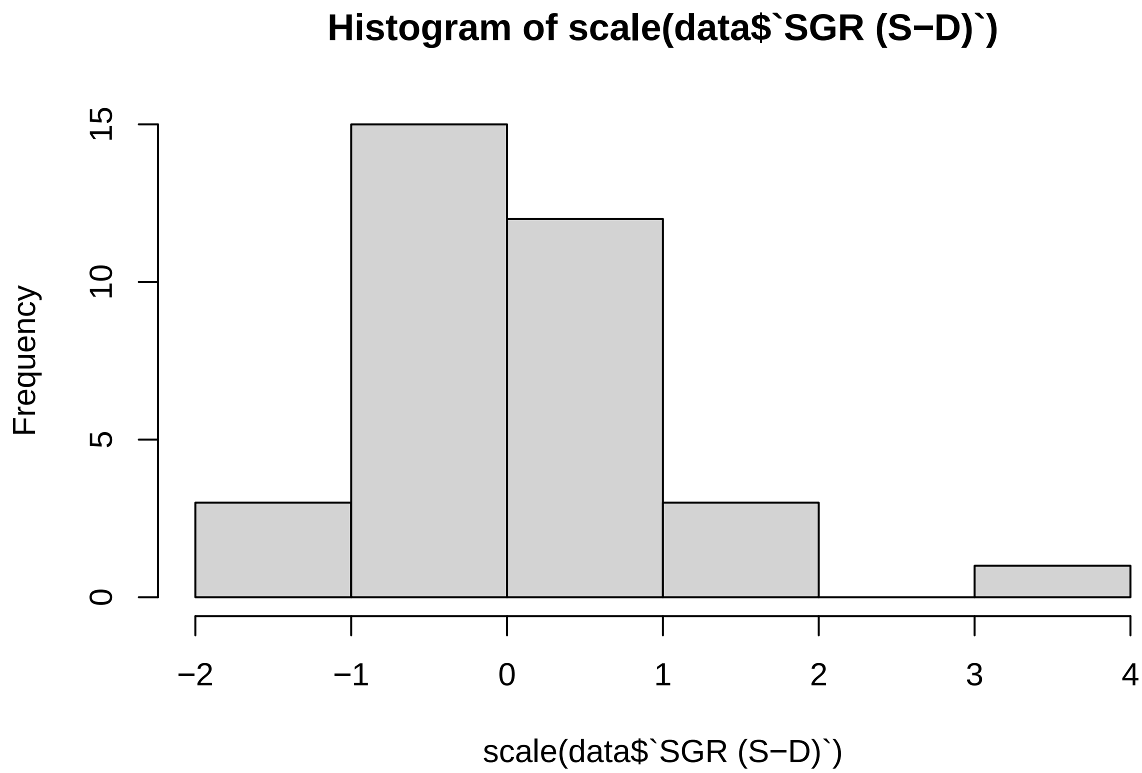
```
hist(scale(data$`SGR (J-S)`))
```




```
which(scale(data$`SGR (J-S)`)>3.29)
```

```
## integer(0)
```

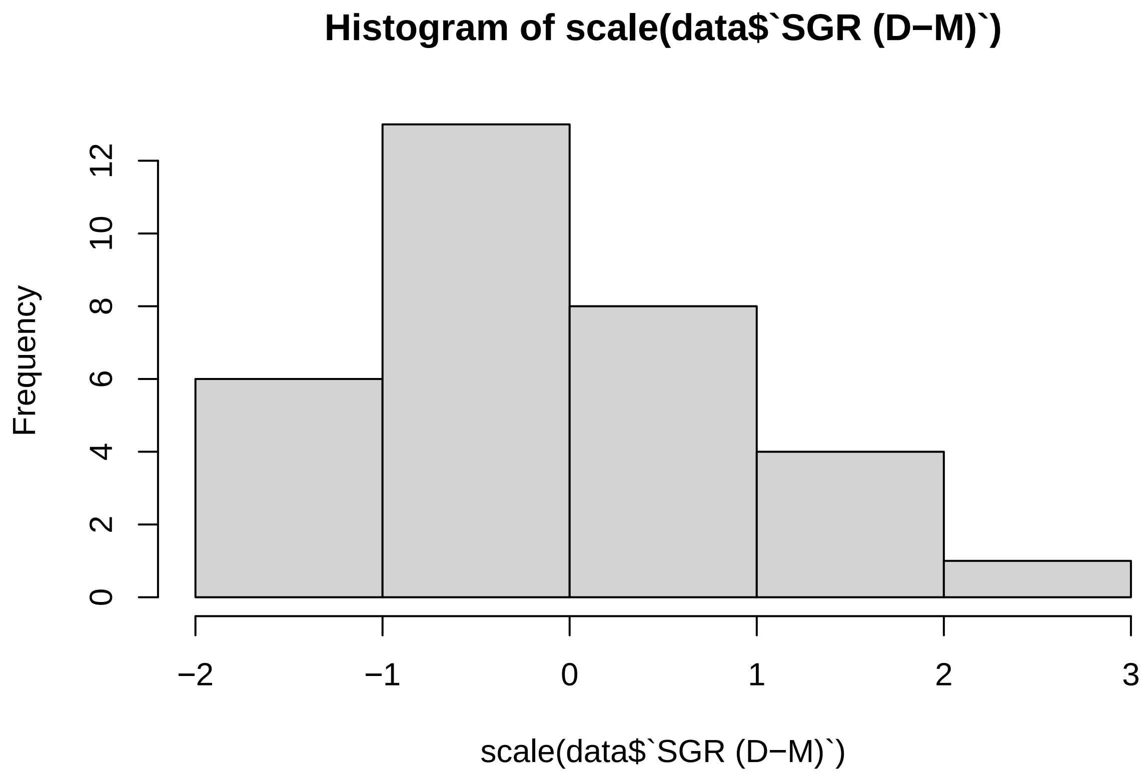
```
hist(scale(data$`SGR (S-D)`))
```




```r
which(scale(data$`SGR (S-D)`)>3.29)
```

```
## [1] 29
```

```r
hist(scale(data$`SGR (D-M)`))
```




```
which(scale(data$`SGR (D-M)`)>3.29)
```

```
## integer(0)
```

By convention, we would exclude any values with a z-score above |3.29| (this excludes 1 out of 1000 observations naturally, which would make it very uncommon that a point falls outside this range in my dataset). The only point that meets this criterion is in September to December. This point falling outside of this range is Fish #29, as suspected. I will remove this point from the linear model as well. No outliers meet the criteria for model exclusion from the June to September or December to March intervals.