Project.Final

Leslie Speight Youtsey and Liz Weatherup

11/13/2020

Title : Fluxes in the ocean

Abstract:

Introduction:

Many environmental and anthropogenic activities can affect the amount of nutrient runoff into water ecosystems (1). The amount of nitrogen (N) and phosphorus (P) entering an ecosystem can affect their flux as well as many other environmental impacts (1). In a study done in Erhai Lake, it showed that nitrogen and phosphorus diffusion flux can affect the biomass and population of autotrophs in this lake that can result in algal blooms (2). We can expect a similar trend to happen in the ocean, but due to the open oceans variability, it is much easier to study these trends in a lake system. Nitrogen and phosphorus fluxes are influenced by internal and external factors such as: nutritive salt concentration, DO, pH, algal biomass, microbial activity, chlorophyll a, etc. (2). It is important to look at these controlling factors and how they will affect the fluxes and the ecosystem as a whole. For example, Erhai Lake is in a transition phase from mesotrophic to eutrophic bodies of water due to the increase of nutritive salts affecting the N and P diffusive fluxes (2).

Carbon also enters the ocean and causes stress to organisms as it increases ocean acidity, but it enters through the earth's atmosphere (3). The amount of carbon in the atmosphere has been steadily increasing overtime due to the large input of fossil fuels into the environment (3). Again, carbon input, like nitrogen and phosphorus is caused by anthropogenic stress.

Nitrogen, phosphorus, and carbon are all connected in that they relate to the biological pump and the solubility pump. They all interact with marine organisms that use them for growth and energy, and they are eventually respired or excreted out of these organisms to sink to depth (4). Ocean physics and the solubility pump also helps to aid these to sink to depth, especially in areas of deep, cold water mixing, like the Southern Ocean (4). The sinking of this nitrogen, phosphorus, and carbon is flux (4). These fluxes are important, especially when it comes to carbon sequestration, which has often been a hot topic as a way to decrease the amount of carbon in the atmosphere. Typically, these fluxes tend to decrease with depth due to the carbon, phosphorus, and nitrogen being used by organisms along the way down to the benthos (4).

The best way to measure these fluxes is using sediment traps. Sediments traps are a method for collecting and determining the sinking fluxes of particulate matter, carbon, and nitrogen in the sea. These values are expressed as mg m^-2 day^-1. Total particulate mass flux is defined as the amount of sinking particulate matter passing through a depth level as (mg dry weight m^-2 day^-2). Total particulate carbon flux is defined as the amount of sinking particulate organic carbon passing through a depth level (mg carbon m^-2 day^-1). Total nitrogen mass flux is defined as the amount of sinking particulate organic nitrogen passing through a depth level (mg nitrogen m^-2 day^-1) (5). In our study we looked at fluxes of nitrogen, phosphorus, carbon,

and mass in the Bermuda Atlantic Time-series through The Joint Global Ocean Flux Study. The Joint Global Ocean FLux Study (JGOFS) is a multi-disciplinary and international study that aims to understand the ocean's role in global carbon and nutrient cycles. One of the objectives of the U.S JGOFS- sponsored Bermuda Atlantic Time-series Study (BATS) is to "observe and interpret the annual and interannual variability in the rates of particle flux and the apparent rates of particle remineralization over the entire water column". BATS began in October 1988 and is through December 2016.(6)

Toolbox discussion:

Description:

Implementation:

See Appendix for all code

Reading csv data in

Here is what our data looks like

	cr	dep	yymmdd1	yymmdd2	Lat2	Lat2.1	Long1	Long2	M_avg
	<int>	<int>	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
605	605	10205	150	20051010	20051014	31.596	31.616	64.164	64.161	64.99
606	606	10205	200	20051010	20051014	31.596	31.616	64.164	64.161	43.24
607	607	10205	300	20051010	20051014	31.596	31.616	64.164	64.161	47.75
608	608	10206	150	20051124	20051126	31.578	31.479	64.180	64.157	48.27
609	609	10206	200	20051124	20051126	31.578	31.479	64.180	64.157	33.98
610	610	10206	300	20051124	20051126	31.578	31.479	64.180	64.157	32.07

6 rows | 1-10 of 13 columns

Subsetted each flux out seperatly

Subsetted each depth and Renaming Columns

Bar plots to access relationship between fluxes and depths

n values for C-flux at each depth

```
## [1] 314
```

```
## [1] 304
```

```
## [1] 305
```

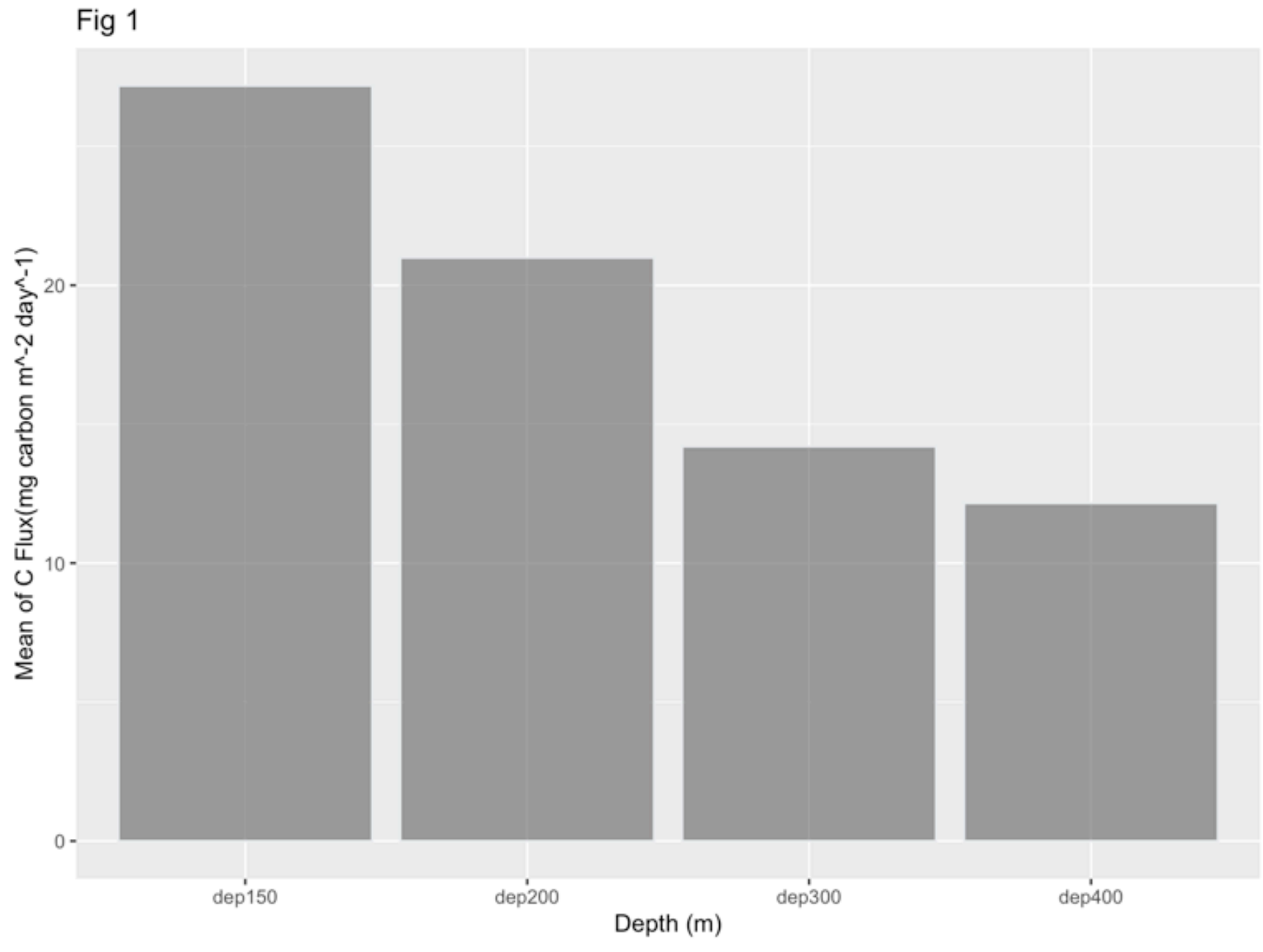
```
## [1] 18
```

Mean and SE of each depth of C-flux

Depth <fctr>	**C_flux.Mean** <dbl>	**C_flux.SE** <dbl>
dep150	27.17930	1.0048090
dep200	20.99128	1.2497363
dep300	14.19941	0.5975955
dep400	12.15778	1.1502888

4 rows

Graph of C-flux



C-flux decreasing with depth

n values of P-flux at each depth

```
## [1] 109
```

```
## [1] 106
```

```
## [1] 105
```

```
## [1] 0
```

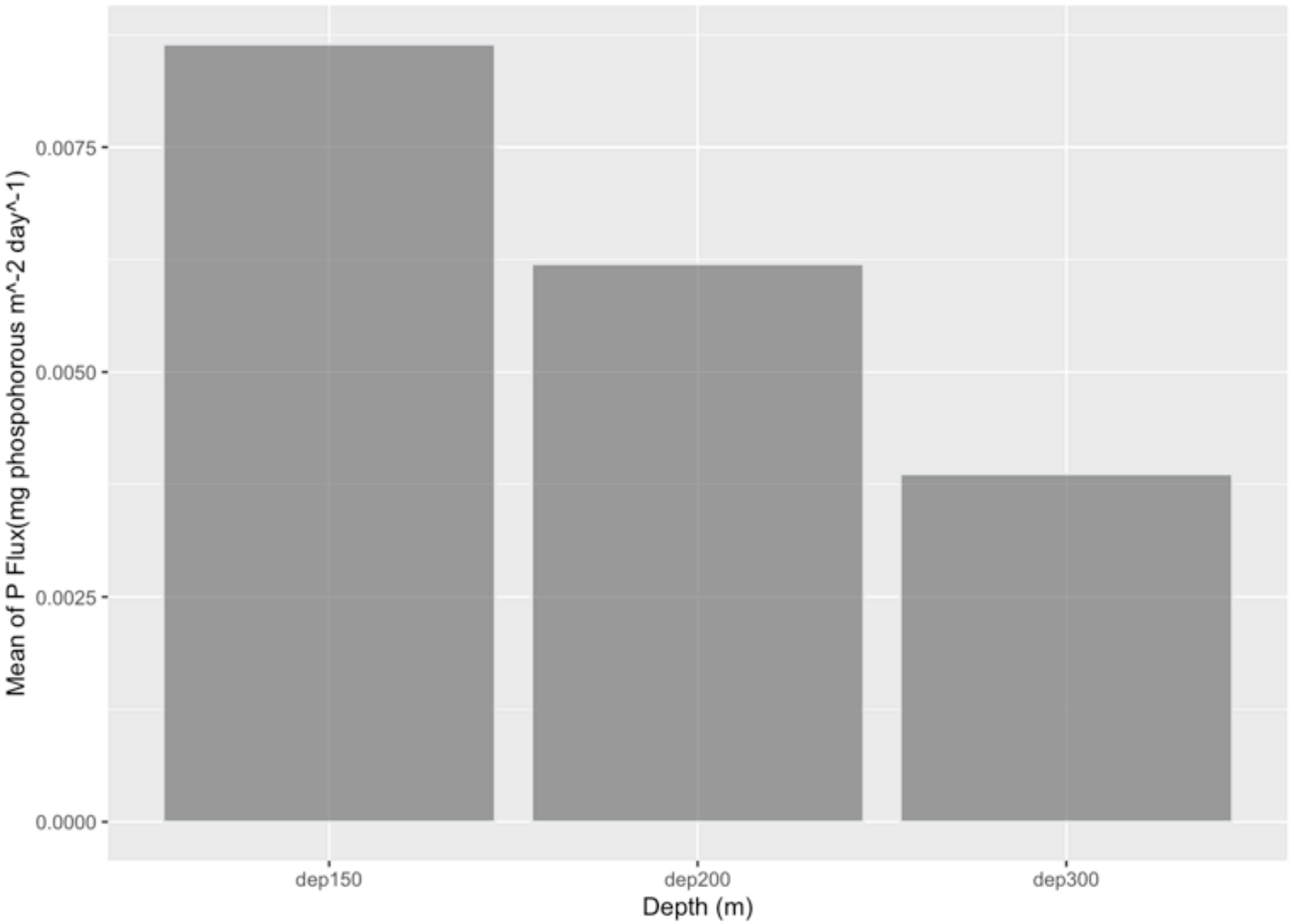
Mean and SE of each depth of P-flux

Depth <fctr>	**P_flux.Mean** <dbl>	**P_flux.SE** <dbl>
dep150	0.008644037	0.001411858
dep200	0.006201887	0.001151720
dep300	0.003866667	0.000852678

3 rows

Graph of P-flux

Fig 2



P-flux decreaing with depth

n value of N-flux at each depth

```
## [1] 312
```

```
## [1] 302
```

```
## [1] 303
```

```
## [1] 18
```

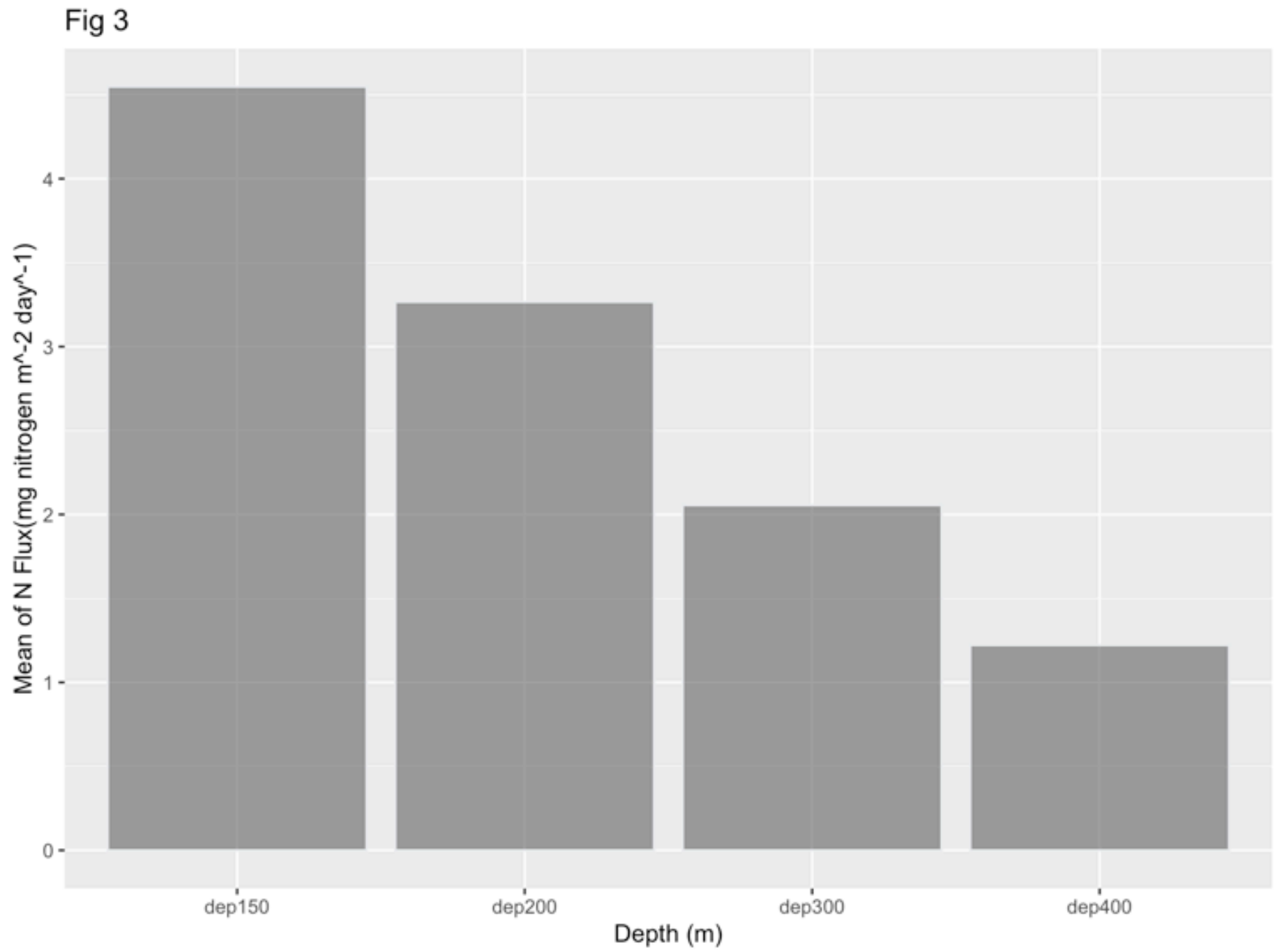
Mean and SE of each depth of N-flux

Depth	**N_flux.Mean**	**N_flux.SE**
<fctr>	<dbl>	<dbl>

dep150	4.548077	0.16714222
dep200	3.265960	0.15298129
dep300	2.055512	0.09043535
dep400	1.221667	0.19282938

4 rows

Graph of N-flux



N-flux decreasing with depth

n value for M-flux at each depth

```
## [1] 314
```

```
## [1] 305
```

```
## [1] 305
```

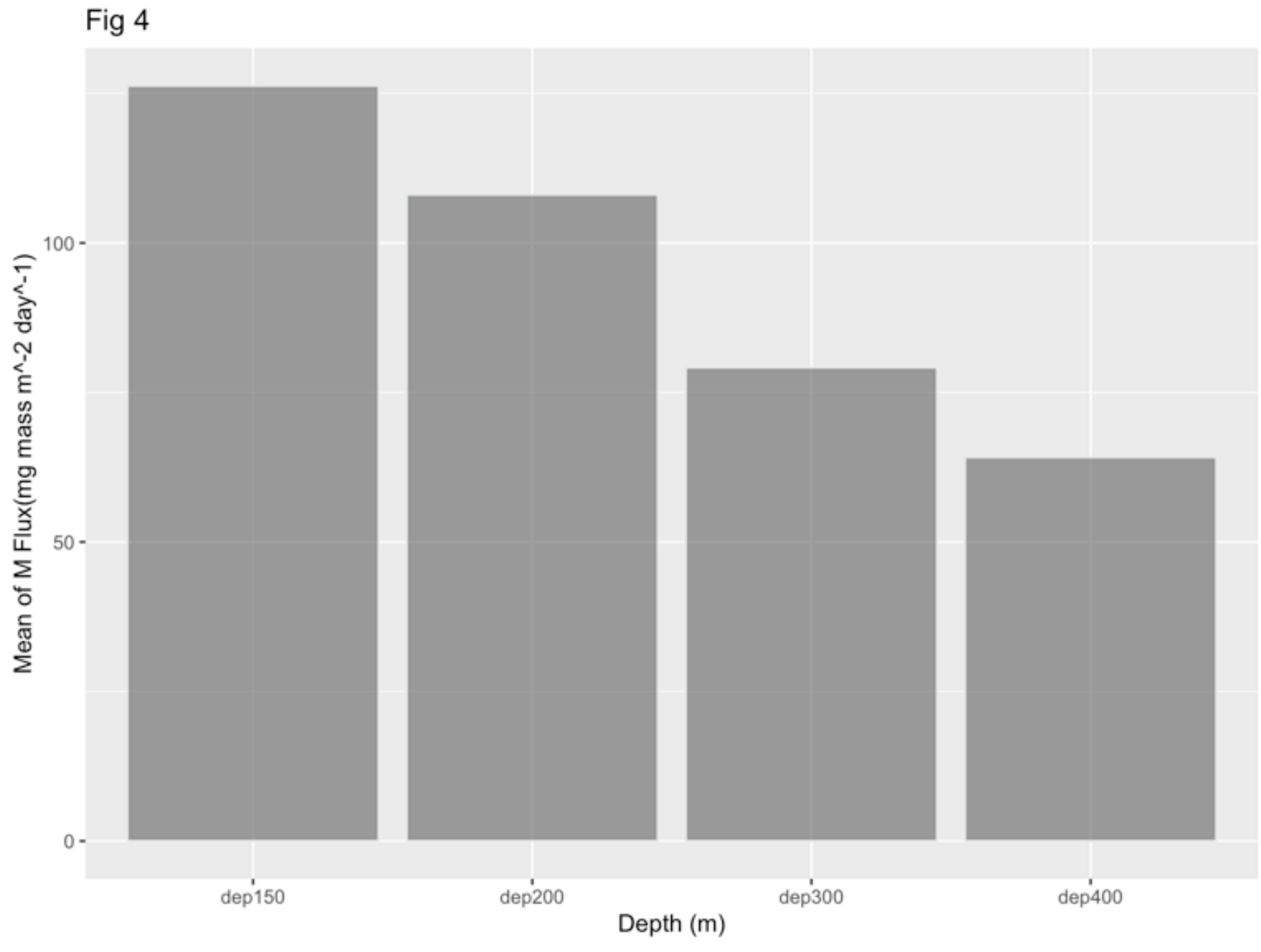
```
## [1] 17
```

Mean and SE of each depth of M-flux

Depth <fctr>	**M_flux.Mean** <dbl>	**M_flux.SE** <dbl>
dep150	126.28223	6.450246
dep200	108.07770	7.316204
dep300	79.17462	4.341316
dep400	64.18529	4.928044

4 rows

Graph of M-flux



M-flux decreasing with depth

Distributionand box plot of our data by depth

Checking for normality and outliers in flux distribution

Log transfromation was needed for all Fluxes

Fig 5 C Flux depth 150

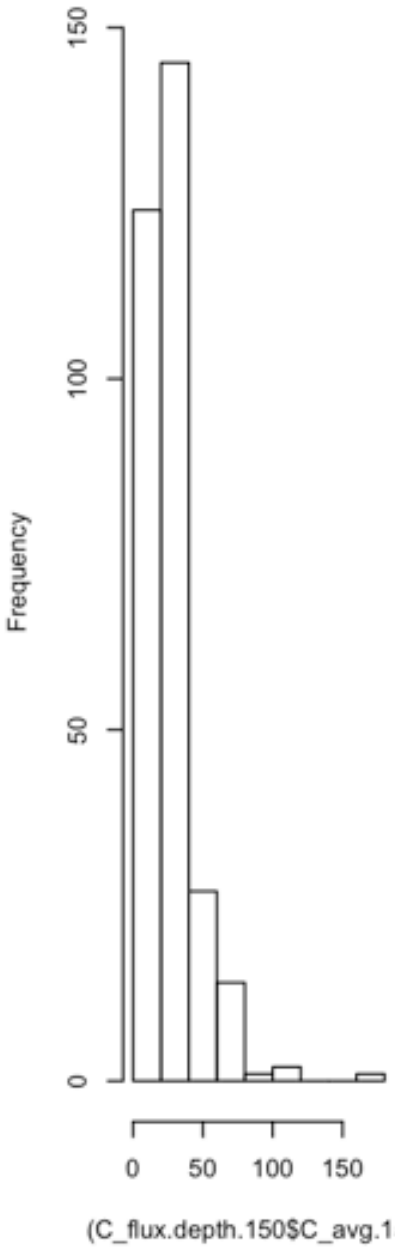


Fig 6 C Flux depth 200

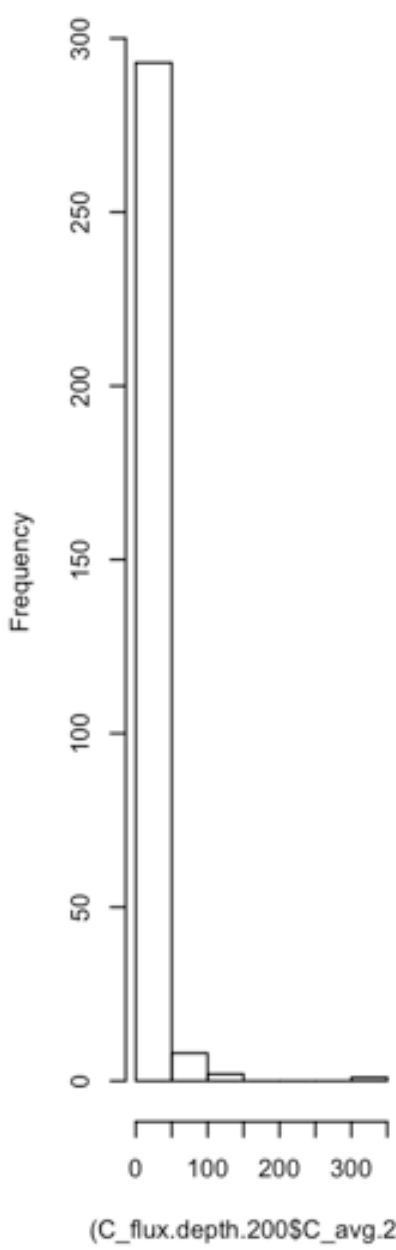


Fig 7 C Flux depth 300

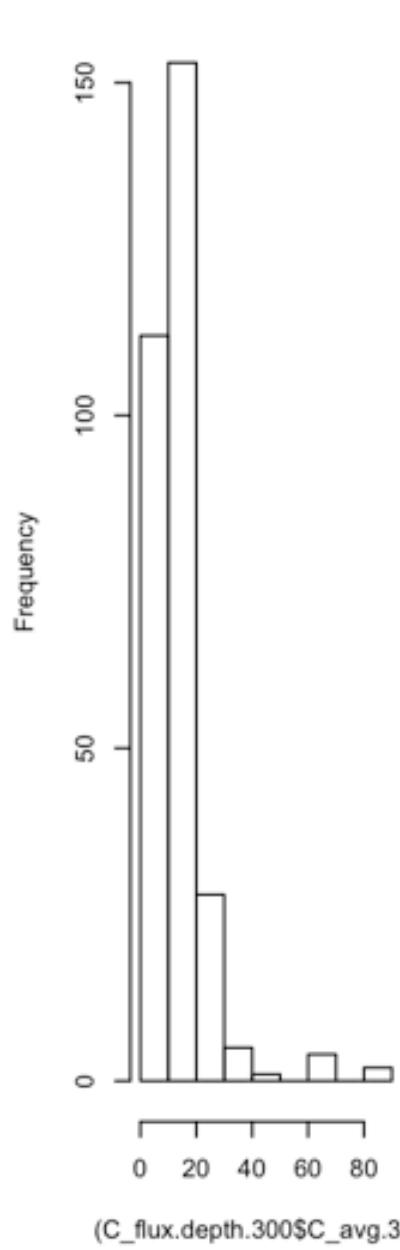


Fig 8 C Flux depth 400

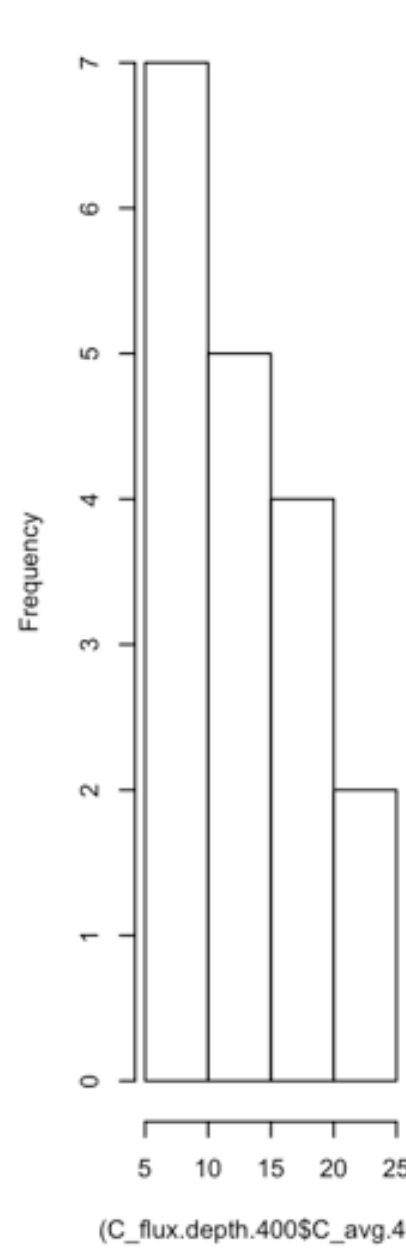


Fig 9 C Flux depth 150 (log)

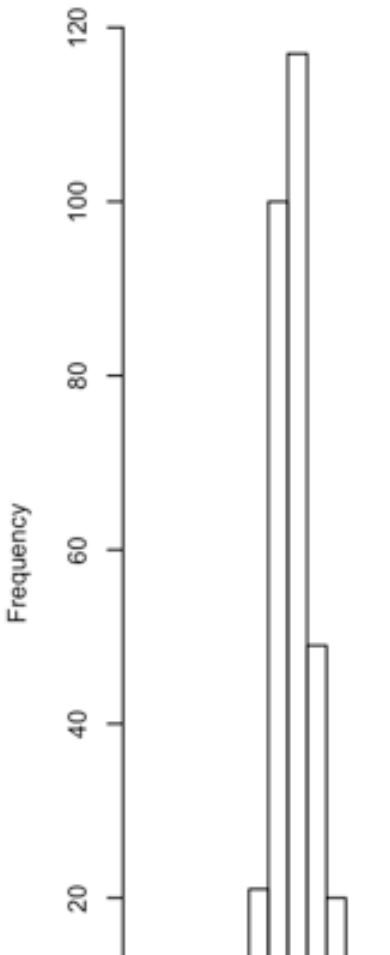


Fig 10 C Flux depth 200 (log

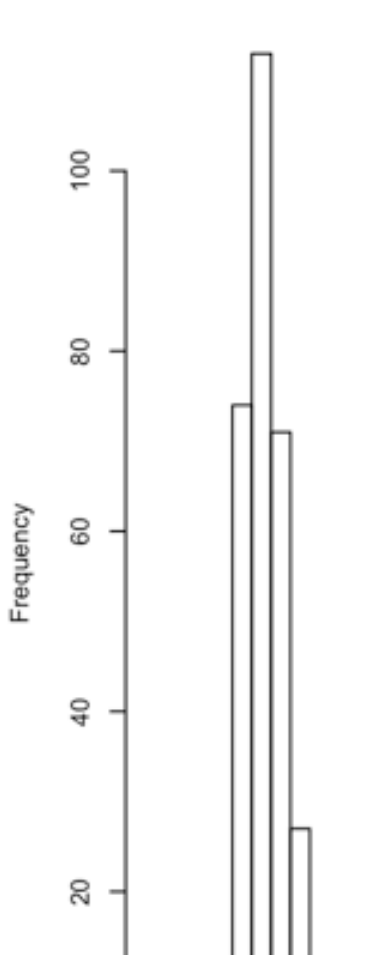


Fig 11 C Flux depth 300 (log

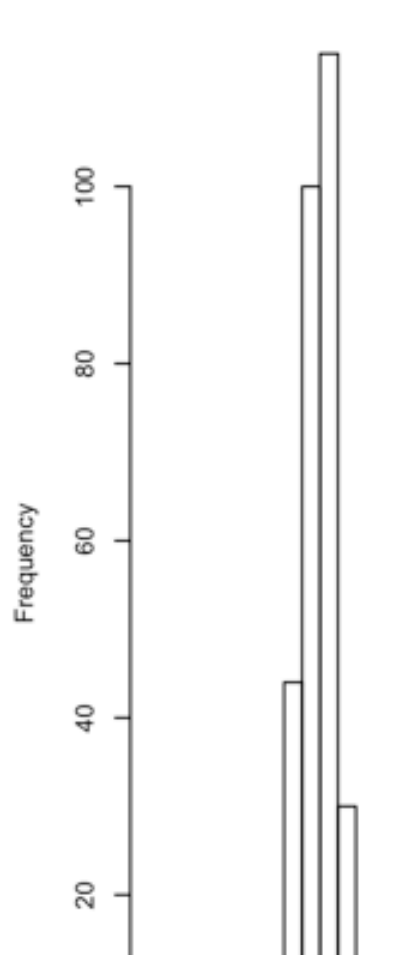
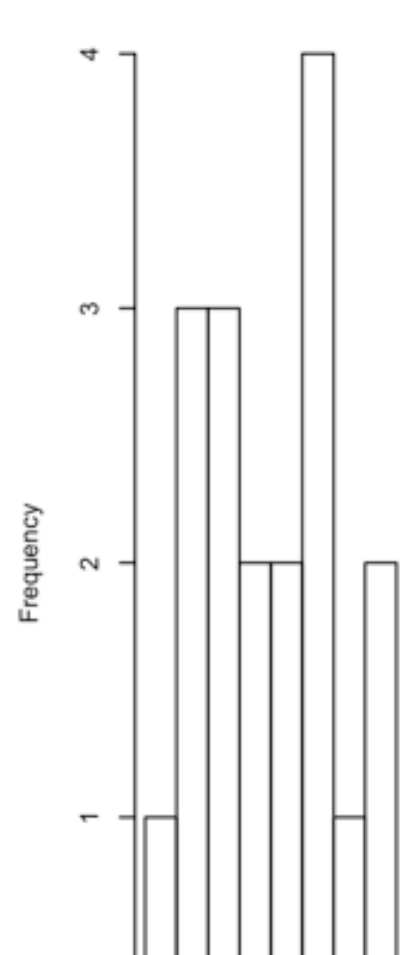
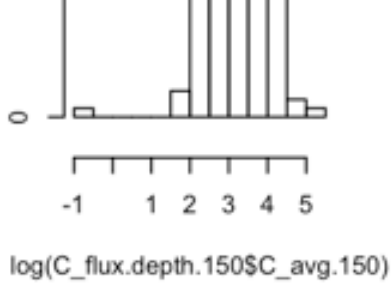
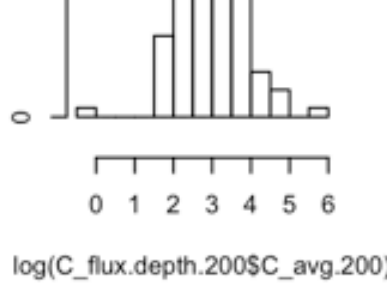
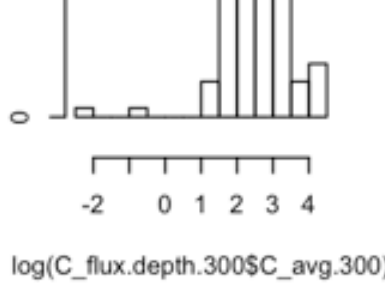


Fig 12 C Flux depth 400 (log









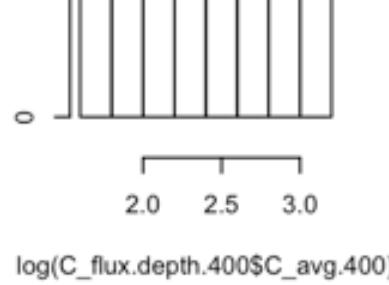


Fig 13 C Flux depth 150 (log

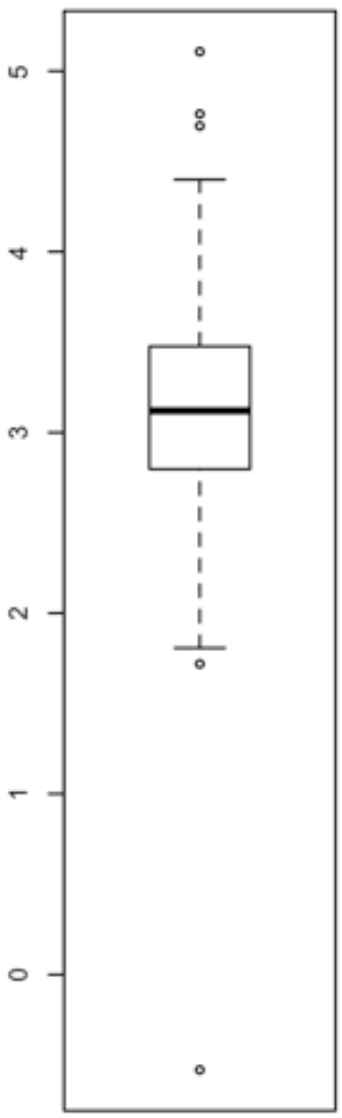


Fig 14 C Flux depth 200 (log

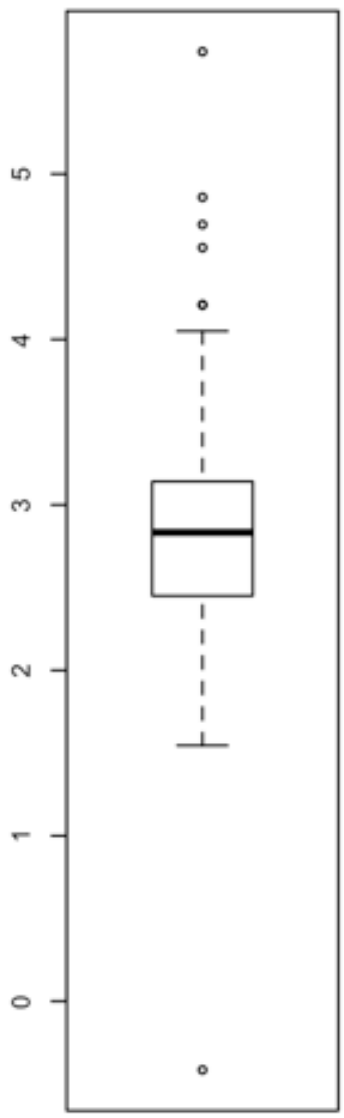


Fig 15 C Flux depth 300 (log

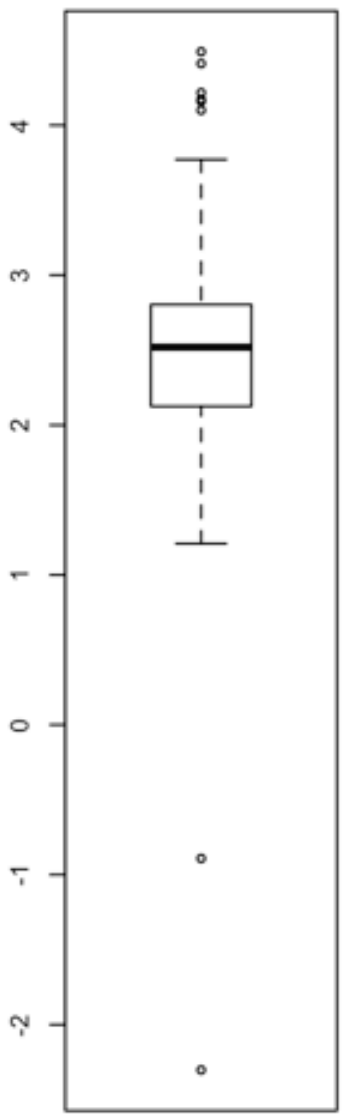
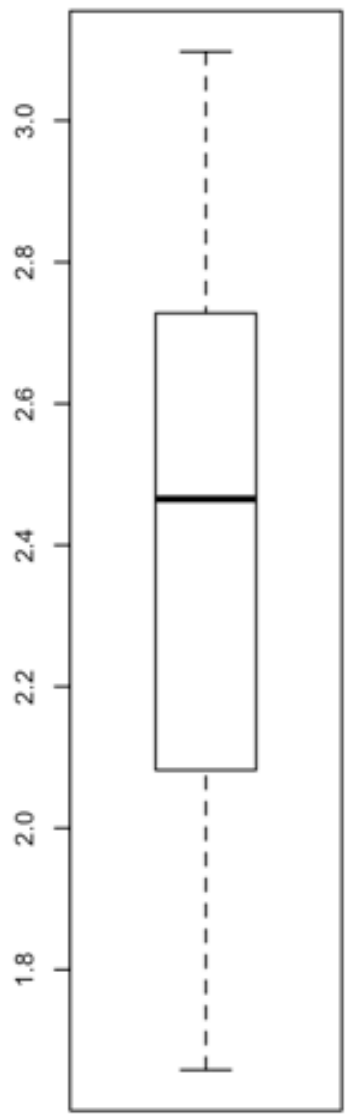


Fig 16 C Flux depth 400 (log



boxplot means do not overlap, but box sizes for 150, 200, and 300 seem to be similar in box size as well as whisker length which means that they could be good predictors of each other and do not break homoscedasticity. However, the 400 depth has a lot of overlap and a very different shape which could break homosedasticity

Fig 17 N Flux depth 150

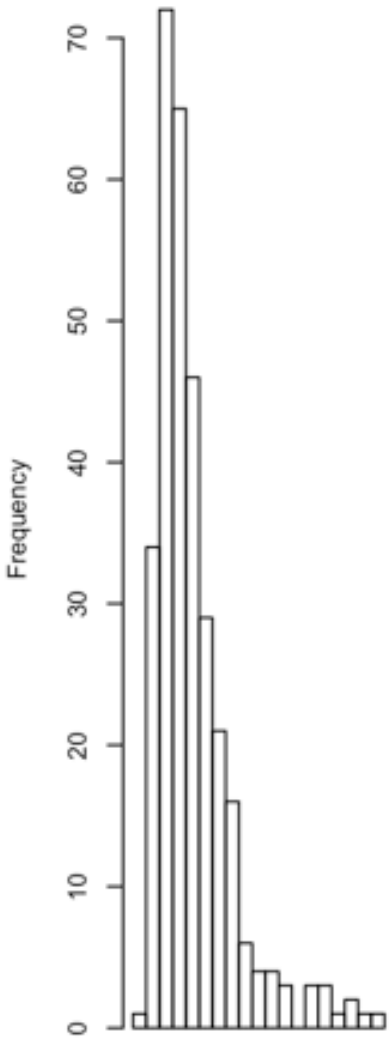


Fig 18 N Flux depth 200

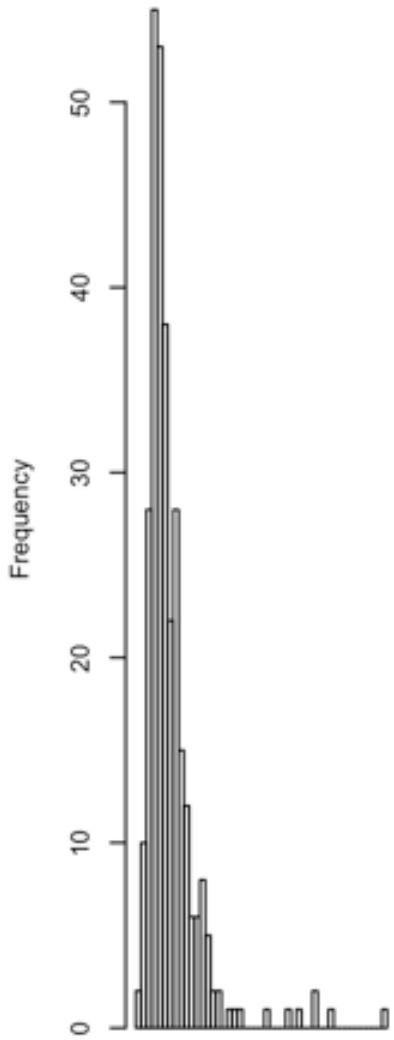


Fig 19 N Flux depth 300

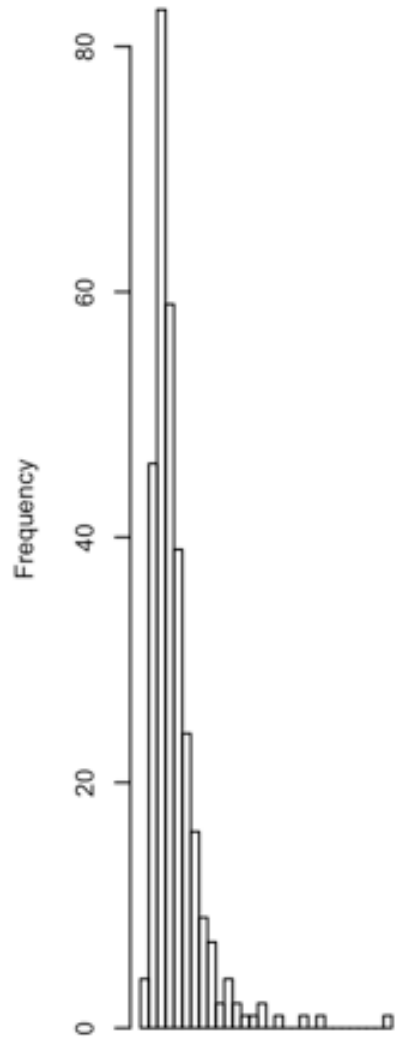


Fig 20 N Flux depth 400

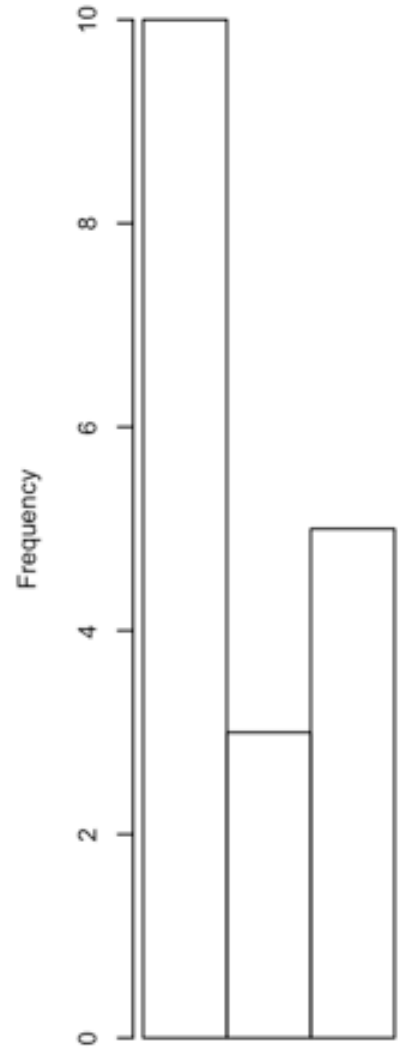


Fig 21 N Flux depth 150 (log

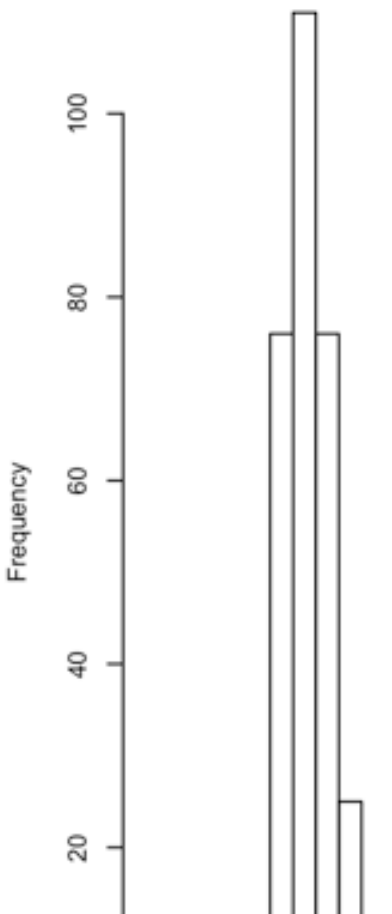


Fig 22 N Flux depth 200 (log

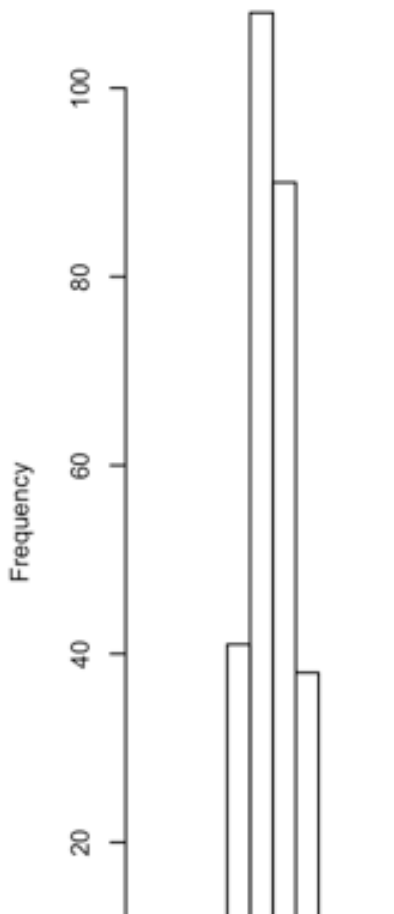


Fig 23 N Flux depth 300 (log

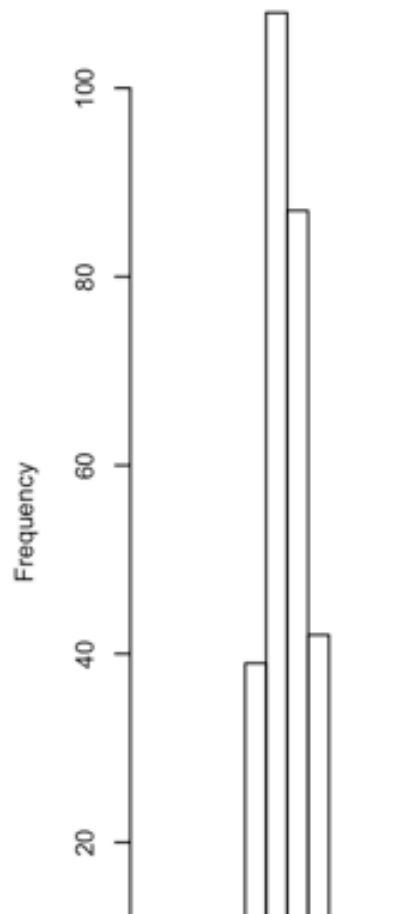
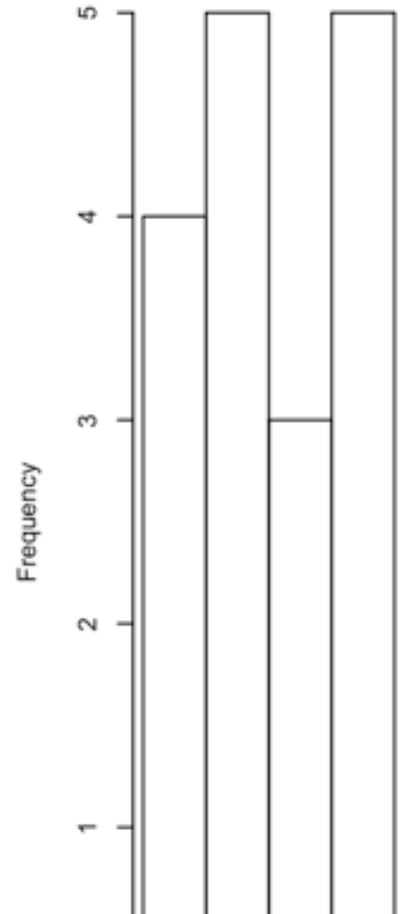
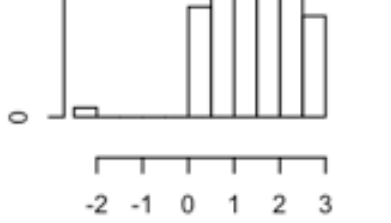
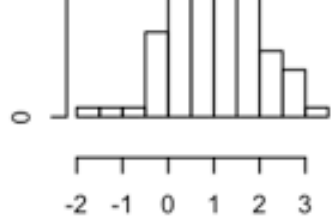


Fig 24 N Flux depth 400 (log

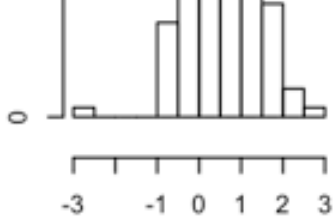




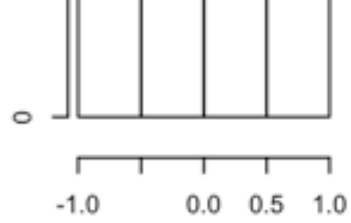
log(N_flux.depth.150$N_avg.150)



log(N_flux.depth.200$N_avg.200)



log(N_flux.depth.300$N_avg.300)



log(N_flux.depth.400$N_avg.400)

Fig 25 N Flux depth 150 (log

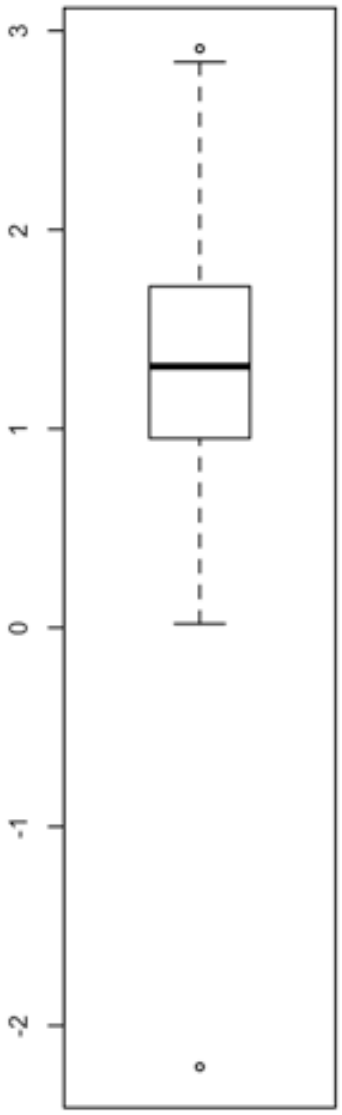


Fig 26 N Flux depth 200 (log

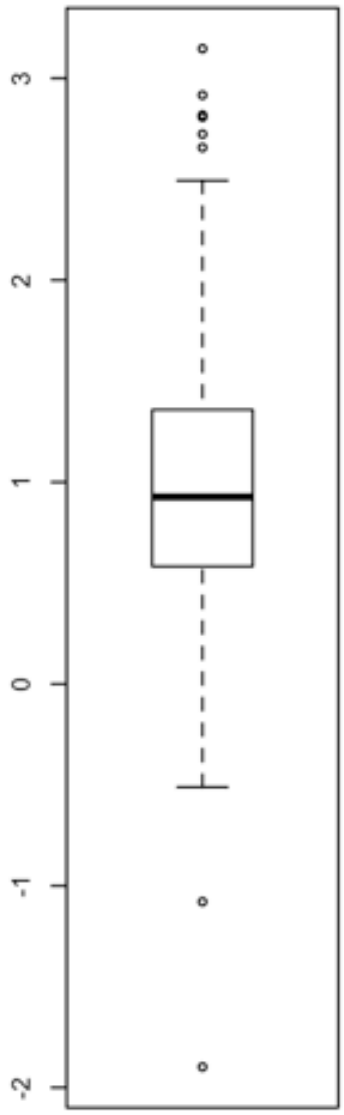


Fig 27 N Flux depth 300 (log

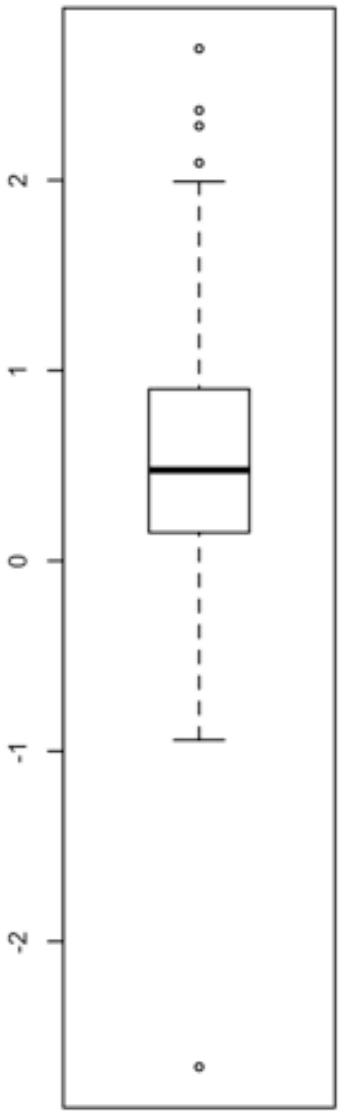
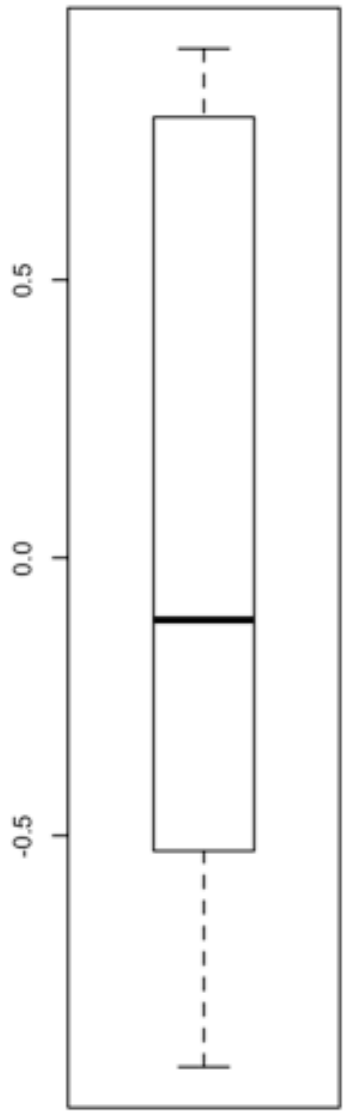


Fig 28 N Flux depth 400 (log



boxplot means do not too seem to overlab too much for 150 and 200, but ther is a bit of overlap in 200 and 300. 150, 200, and 300 seem to be similar in box size as well as whisker length which means that they could be good predictors of each other and do not break homoscedasticity. However, the 400 depth has a lot of overlap and a very different shape which could break homosedasticity

Fig 29 P Flux depth 150

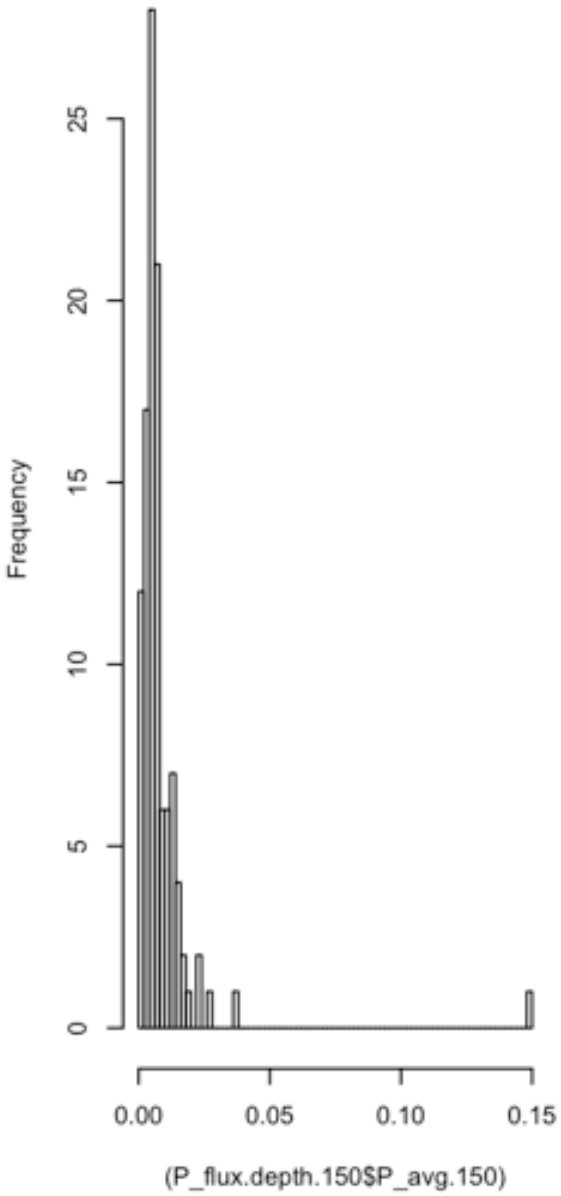


Fig 30 P Flux depth 200

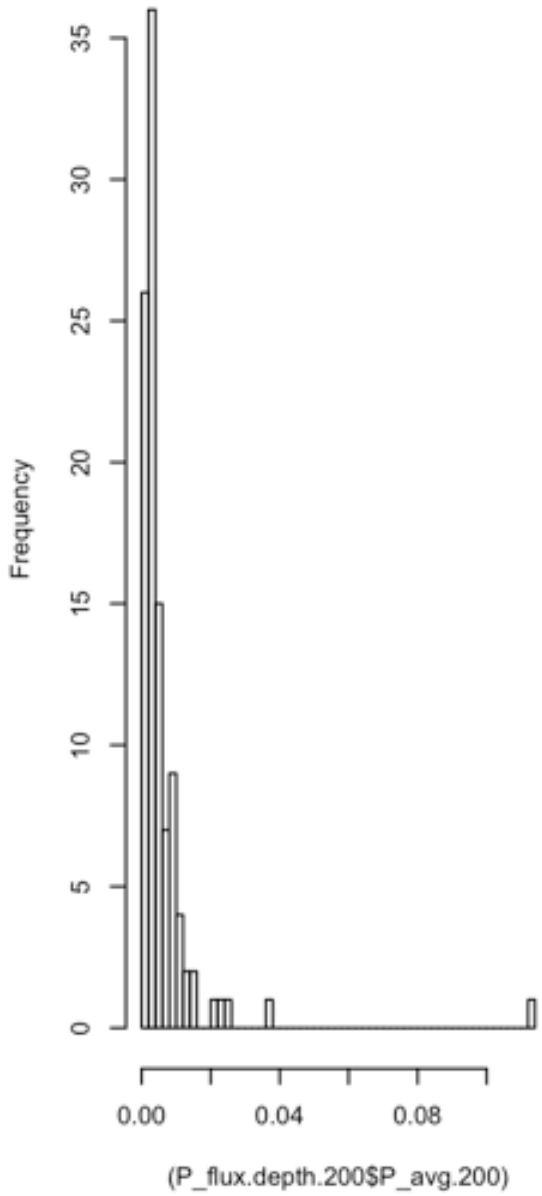


Fig 31 P Flux depth 300

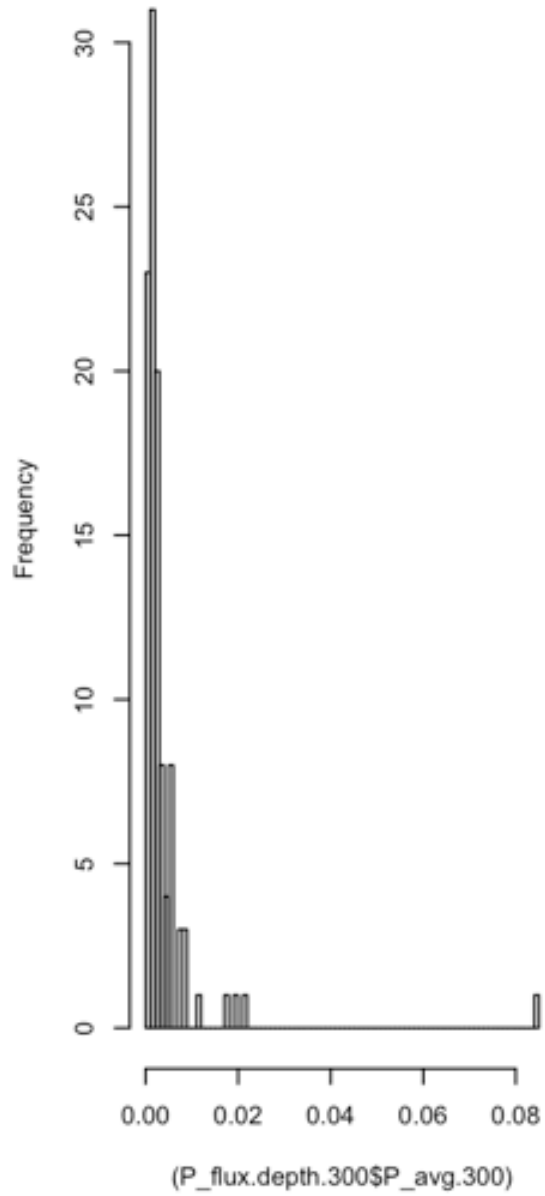


Fig 32 P Flux depth 150 (log)

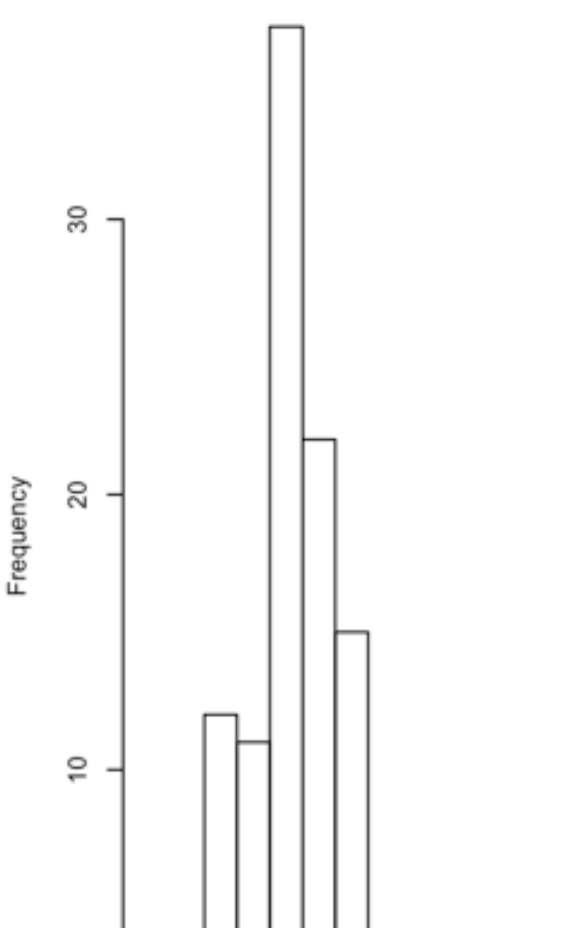


Fig 33 P Flux depth 200 (log)

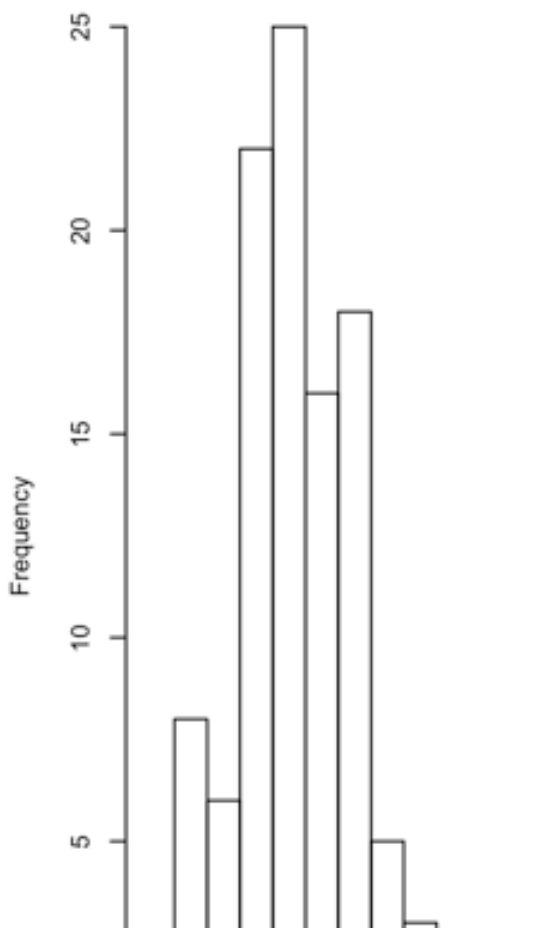
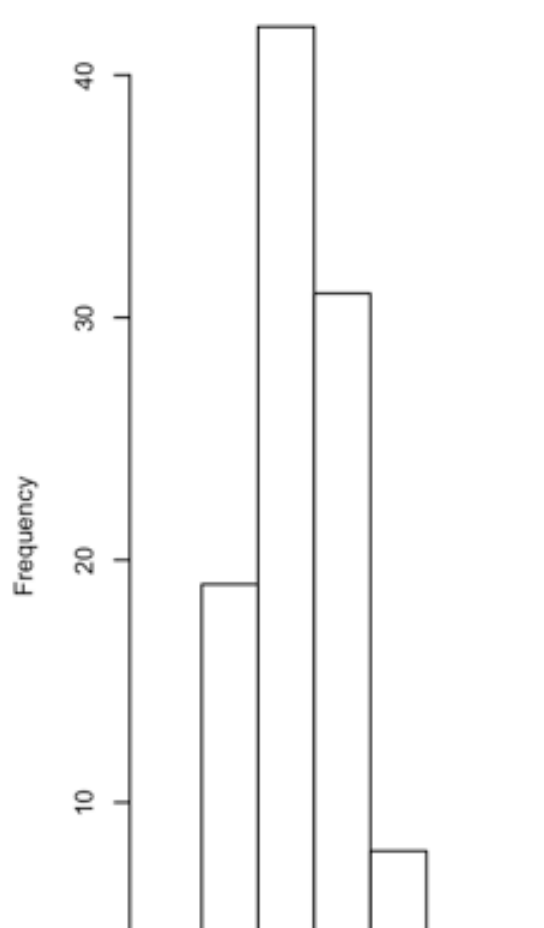
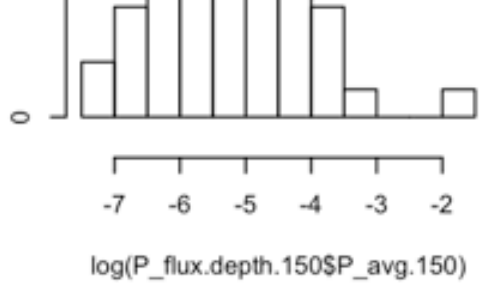
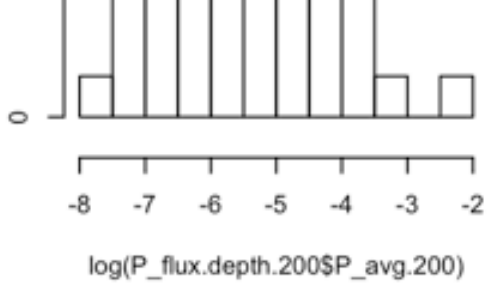


Fig 34 P Flux depth 300 (log)







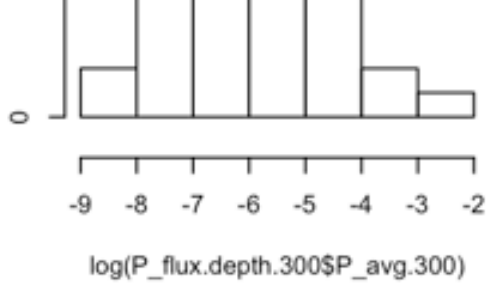
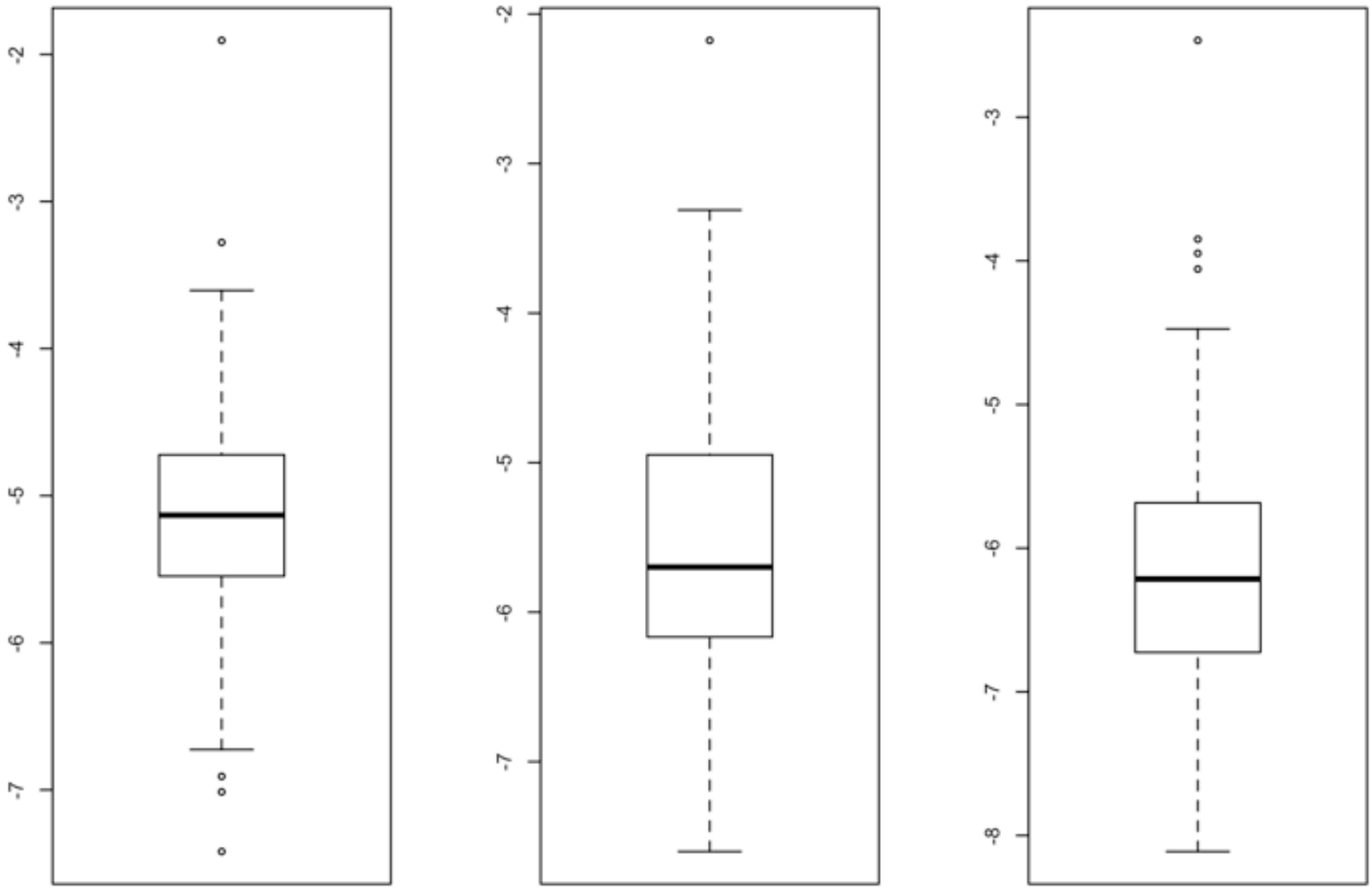


Fig 35 P Flux depth 150 (log)

Fig 36 P Flux depth 200 (log)

Fig 37 P Flux depth 300 (log)



boxplot means seem to have some overlap. They are similar in box size, but the concern is the overlap, making them not good predictors of one another and potentially breaking homosedasticity

Fig 38 M Flux depth 150

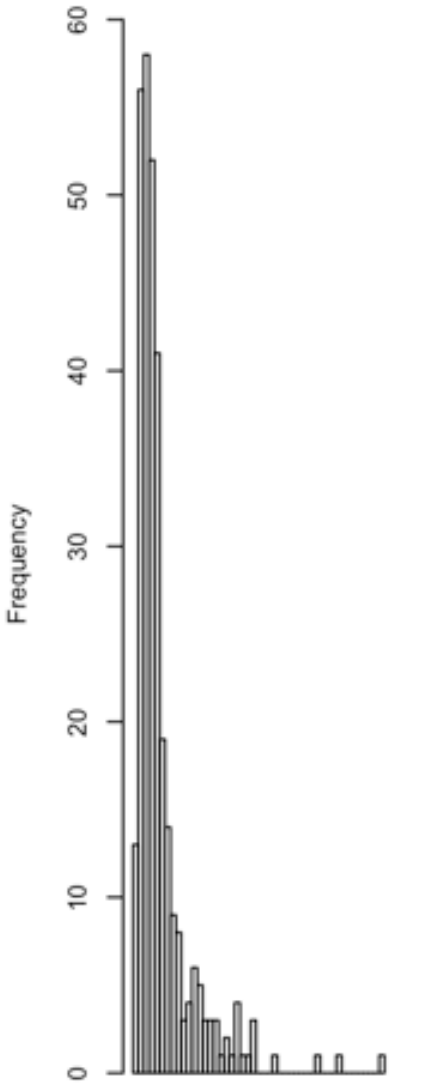


Fig 39 M Flux depth 200

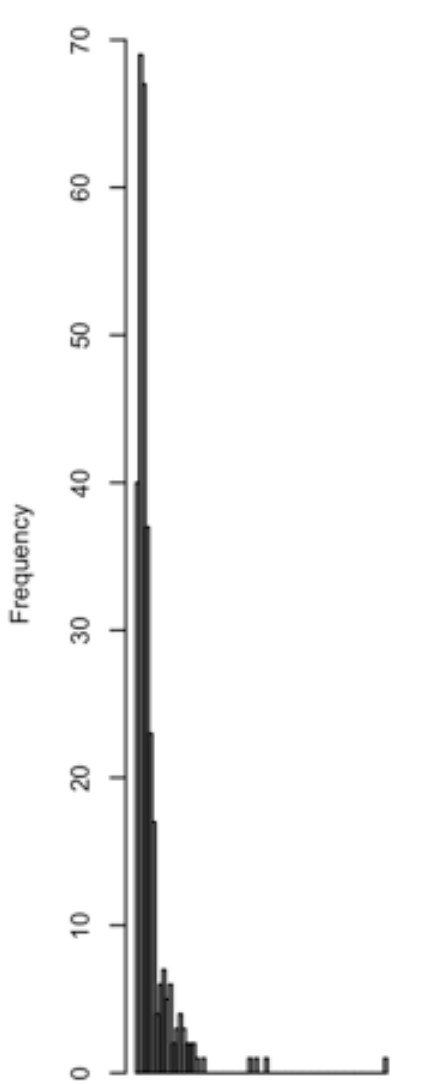


Fig 40 M Flux depth 300

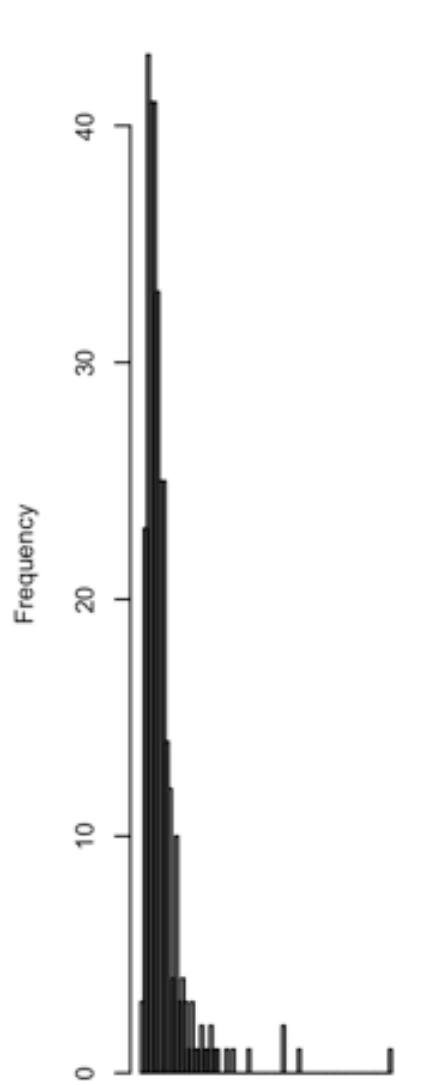


Fig 41 M Flux depth 400

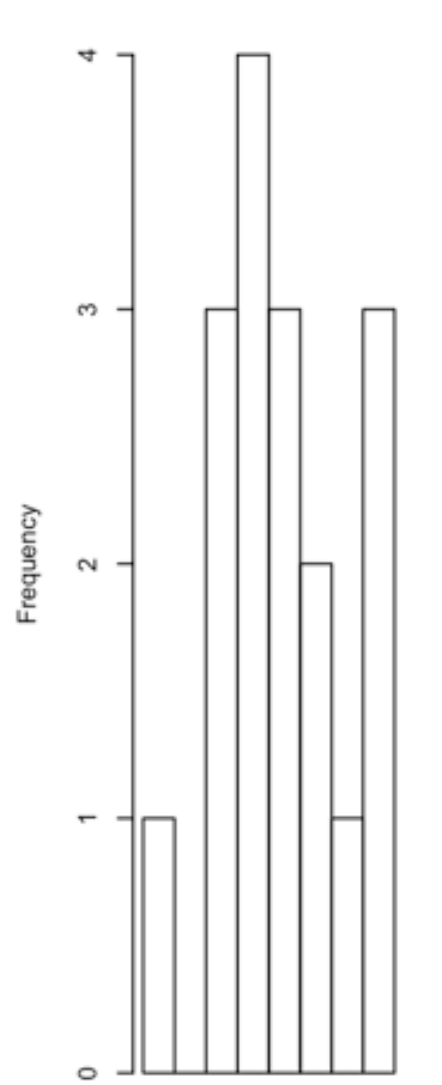


Fig 42 M Flux depth 150 (log

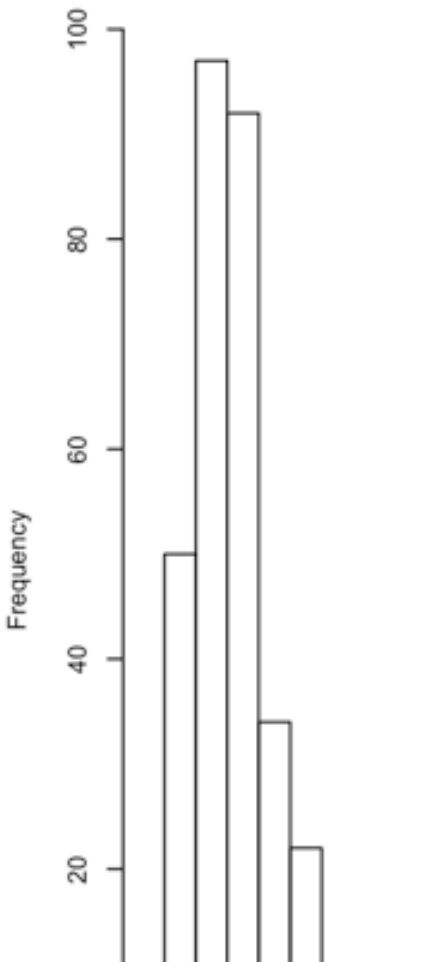


Fig 43 M Flux depth 200 (log

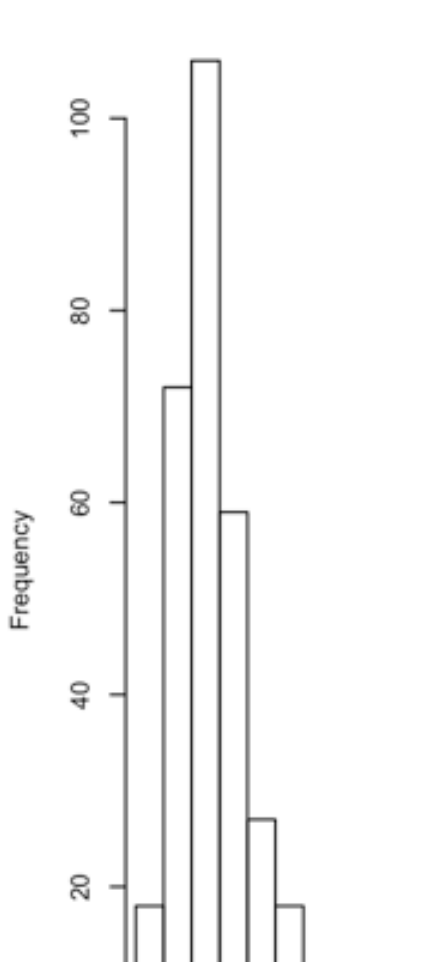


Fig 44 M Flux depth 300 (log

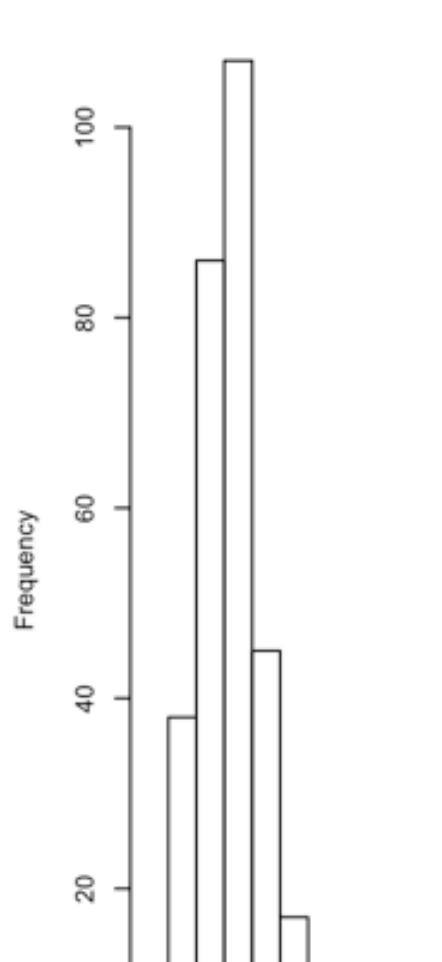
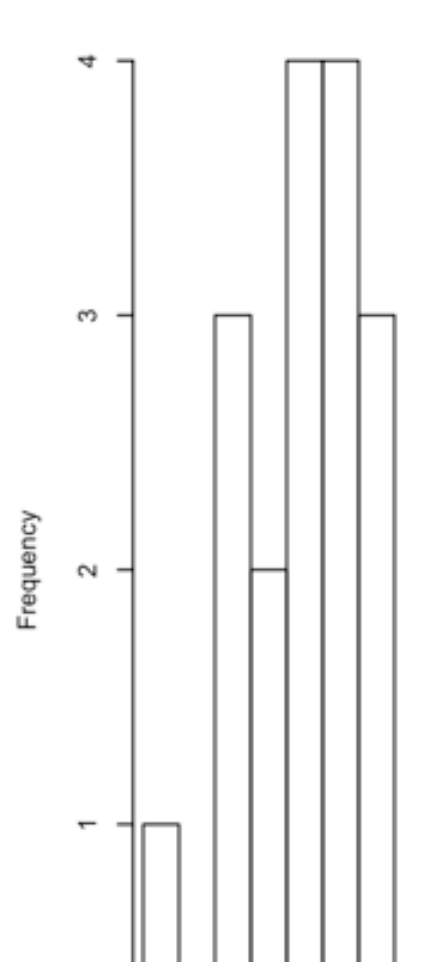
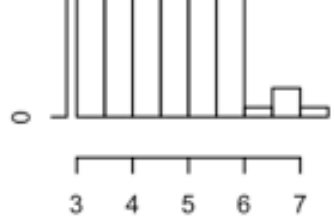


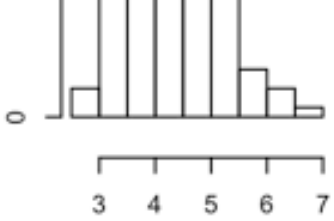
Fig 45 M Flux depth 400 (log



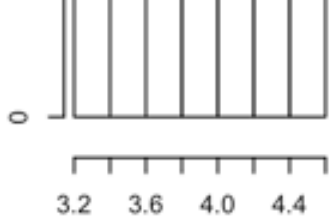
log(M_flux.depth.150$M_avg.150)



log(M_flux.depth.200$M_avg.200)



log(M_flux.depth.300$M_avg.300)



log(M_flux.depth.400$M_avg.400)

Fig 46 M Flux depth 150 (log

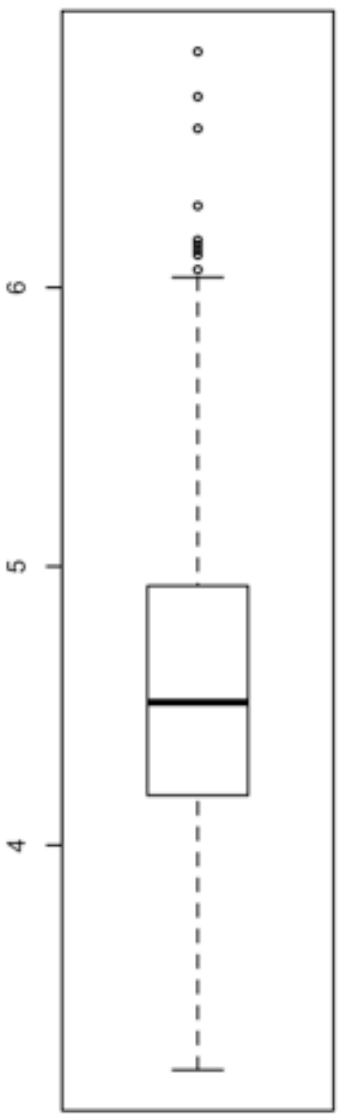


Fig 47 M Flux depth 200 (log

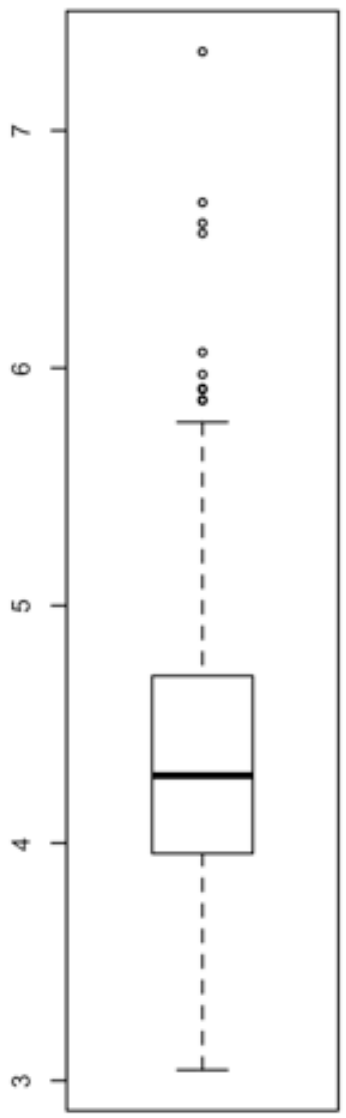


Fig 48 M Flux depth 300 (log

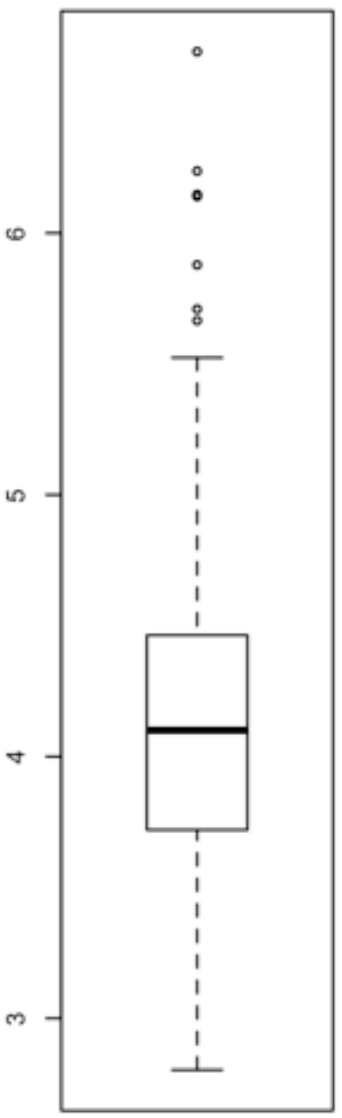
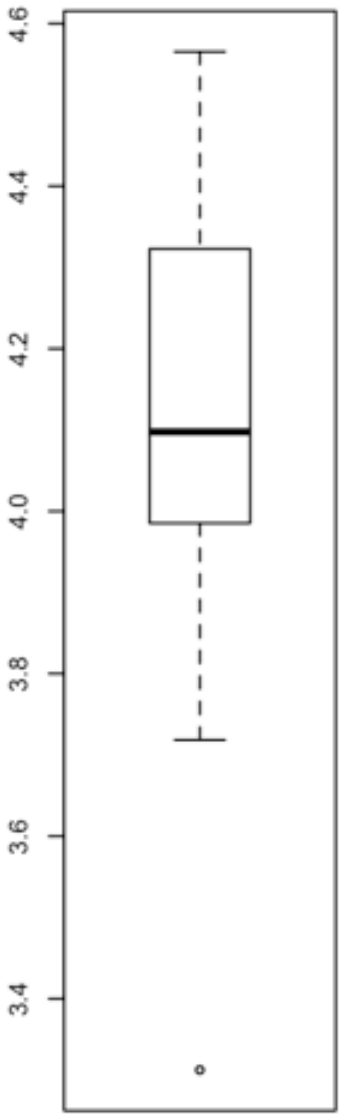


Fig 49 M Flux depth 400 (log



boxplot means seem to have some overlap. They are similar in box size, but the concern is the overlap, making them not good predictors of one another and potentially breaking homosedasticity

Since many of the 400 m depth distrubutions were not normal, even after transformation and have small n, they will not be used in the hypothesis testing

Paired sample t test for each depth and the depth below it

This is a paired test because samples were collected on the same cruise and at the same location. So for each flux value at one depth, there was a flux value collected at a different depth at the same location! Data frames were merged based on cruise, year, and location to have the same number of values.

For a paired t test to be preformed it must meet these assumptions:

$\overline{Y_1}$ - $\overline{Y_2}$ = $\overline{D}$

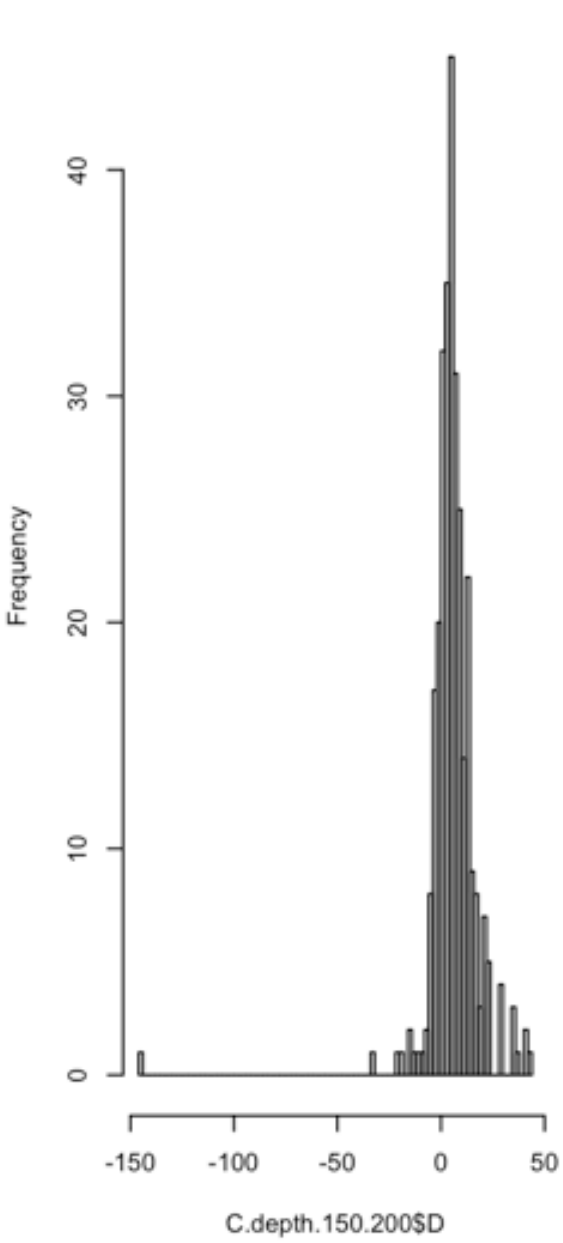
- The dependent variable must be continuous (interval/ratio). ✓
- The observations are independent of one another. ✓
- The dependent variable ($\overline{D}$) should be approximately normally distributed. ☐ ?
- The dependent variable ($\overline{D}$) should not contain any outliers. ☐ ?

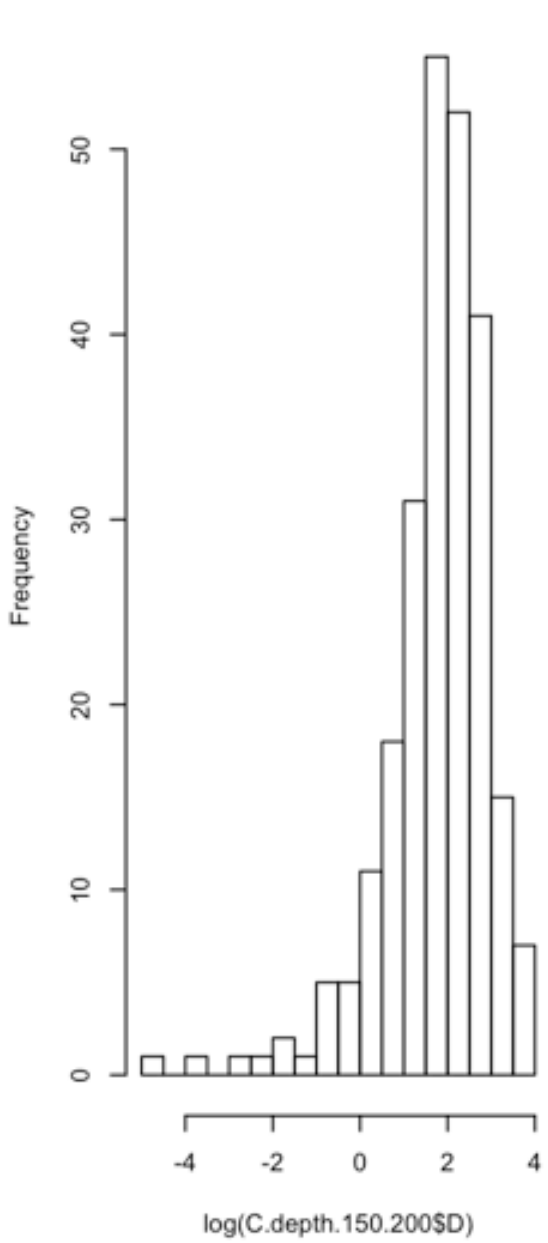
Merging data bases for hypothesis test and Creating $\overline{D}$ for each paired group

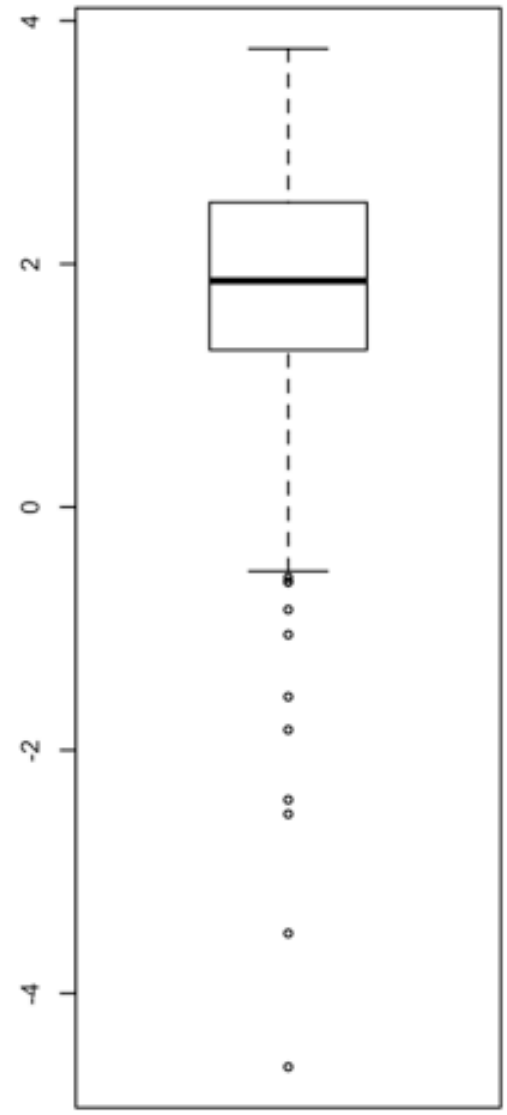
Depth.Diff <fctr>	**C_flux** <dbl>	**N_flux** <dbl>	**P_flux** <dbl>	**M_flux** <dbl>
dep150.200	6.188016	1.282117	0.00244215	18.20452
dep200.300	6.791873	1.210449	0.00233522	28.90308

2 rows

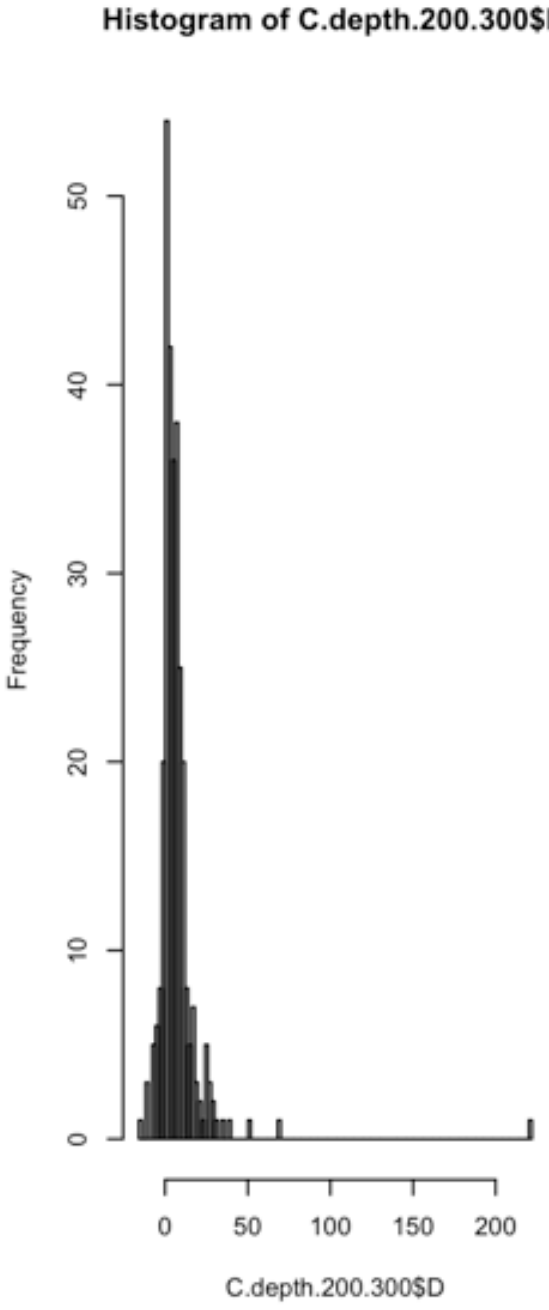
Checking the distribution of the $\overline{D}$ for each flux

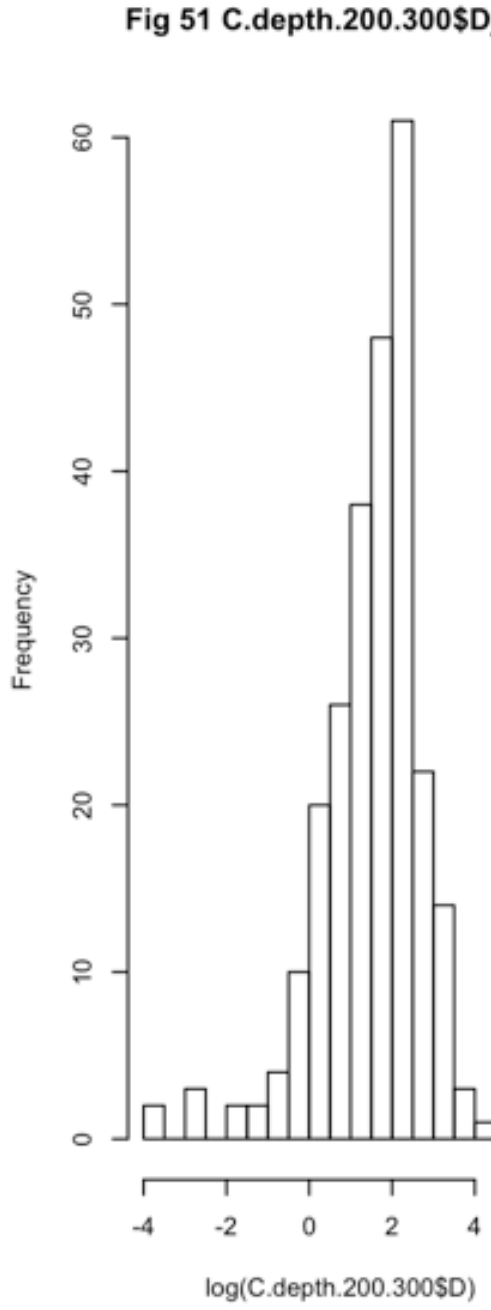


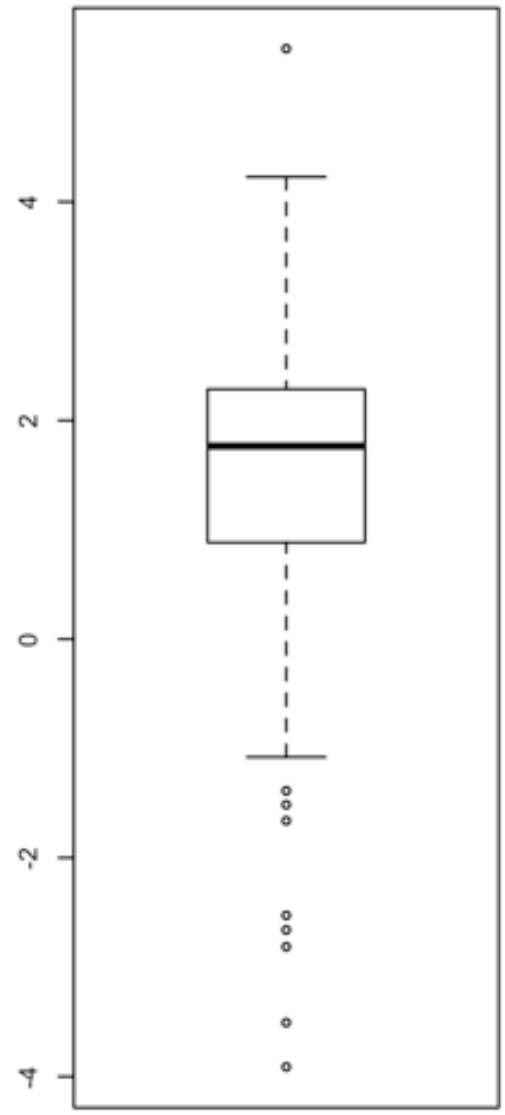




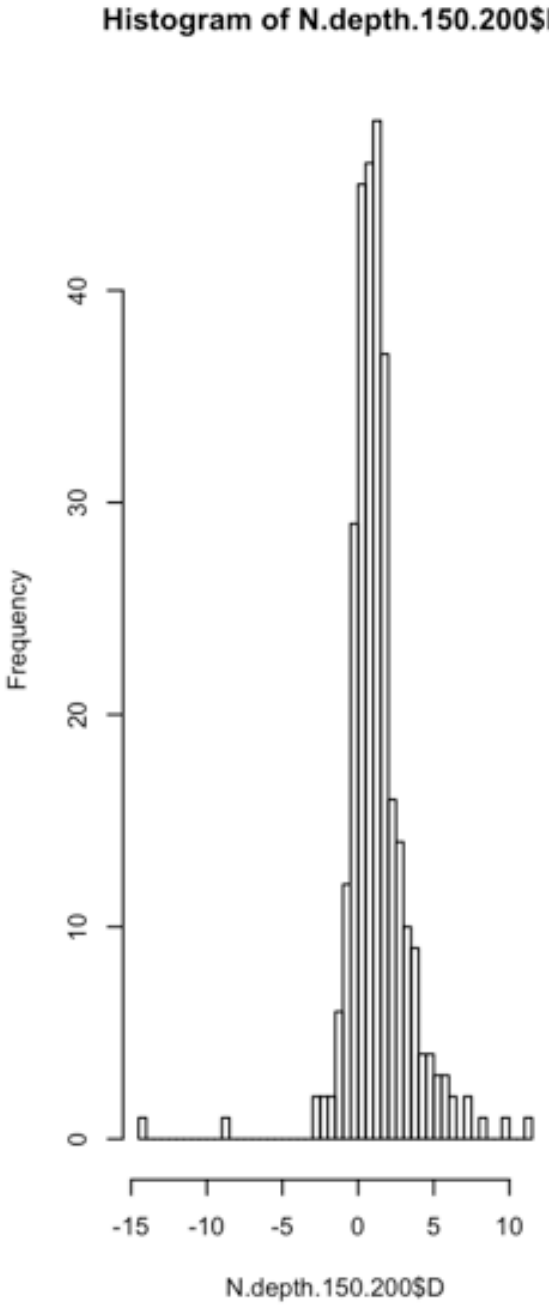
Cdepth (150.200) is not very normal, but log transformation increases normality. Outliers present in box plot

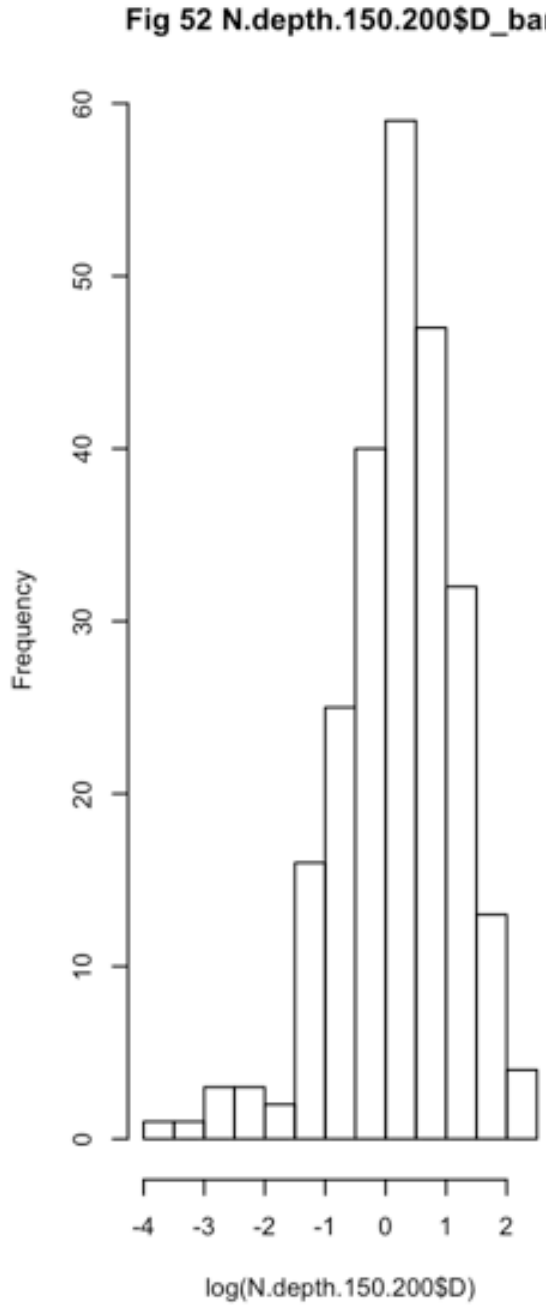


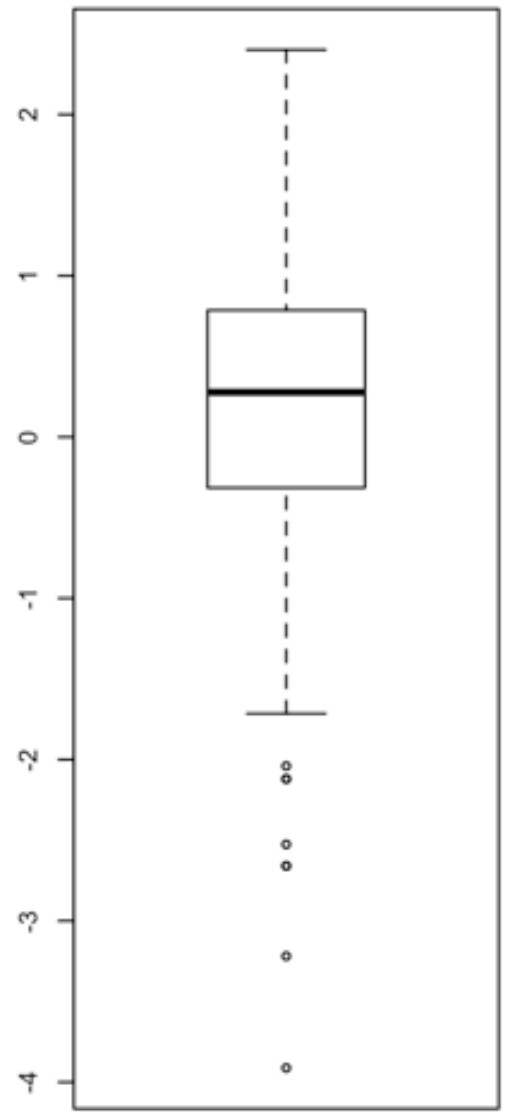




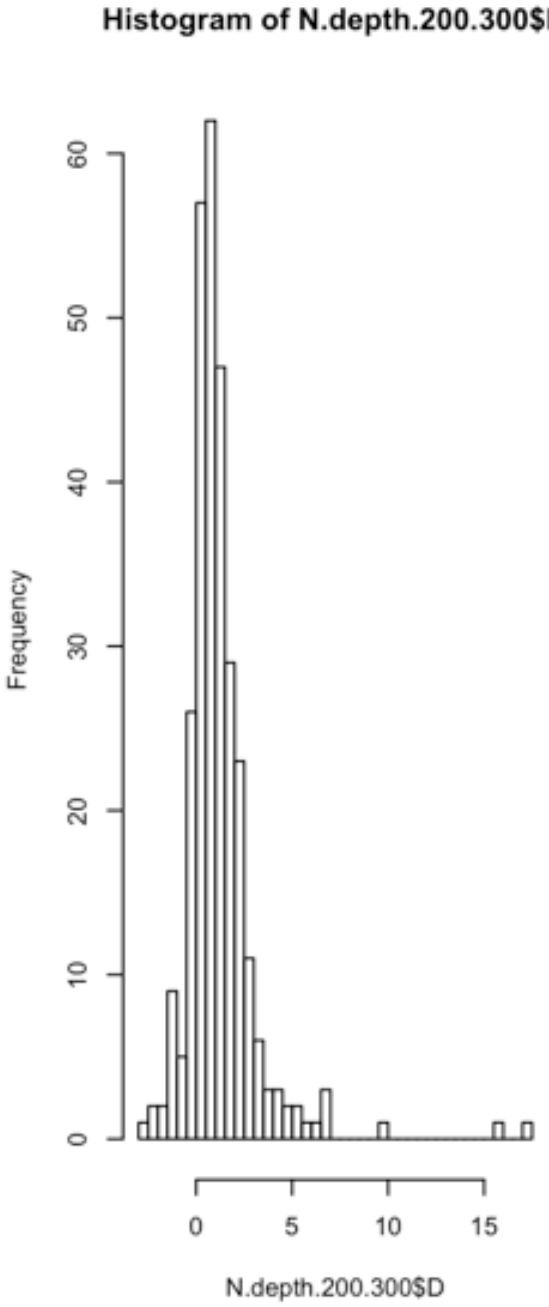
Cdepth (200.300) is not normal, but log transformation increases normality. Outliers present in box plot

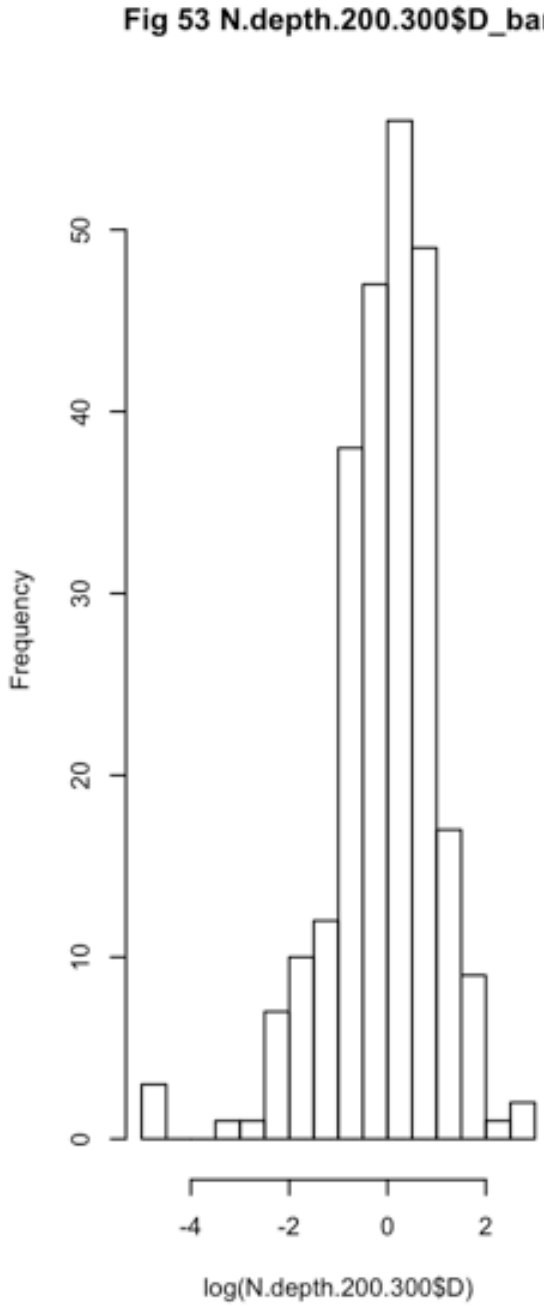


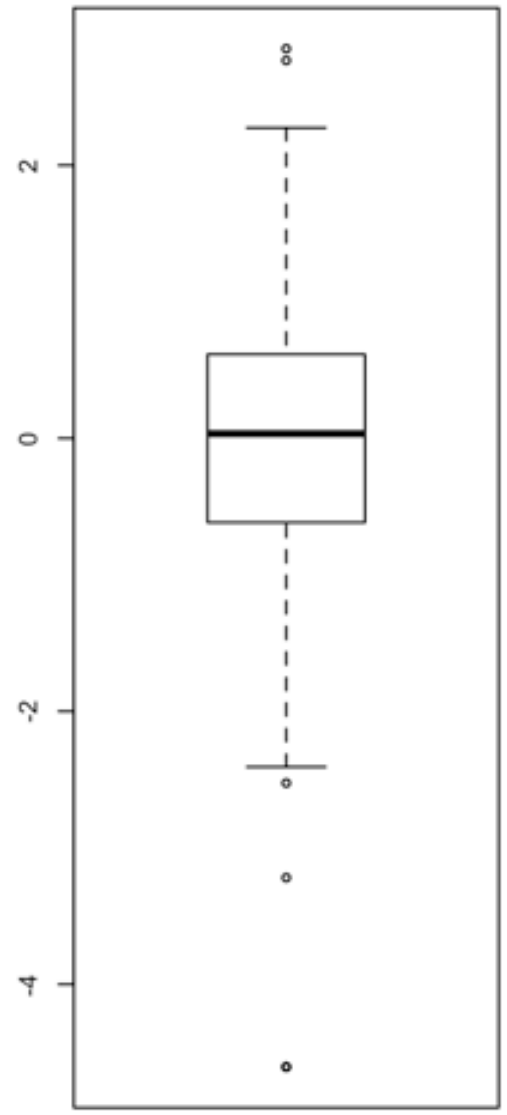




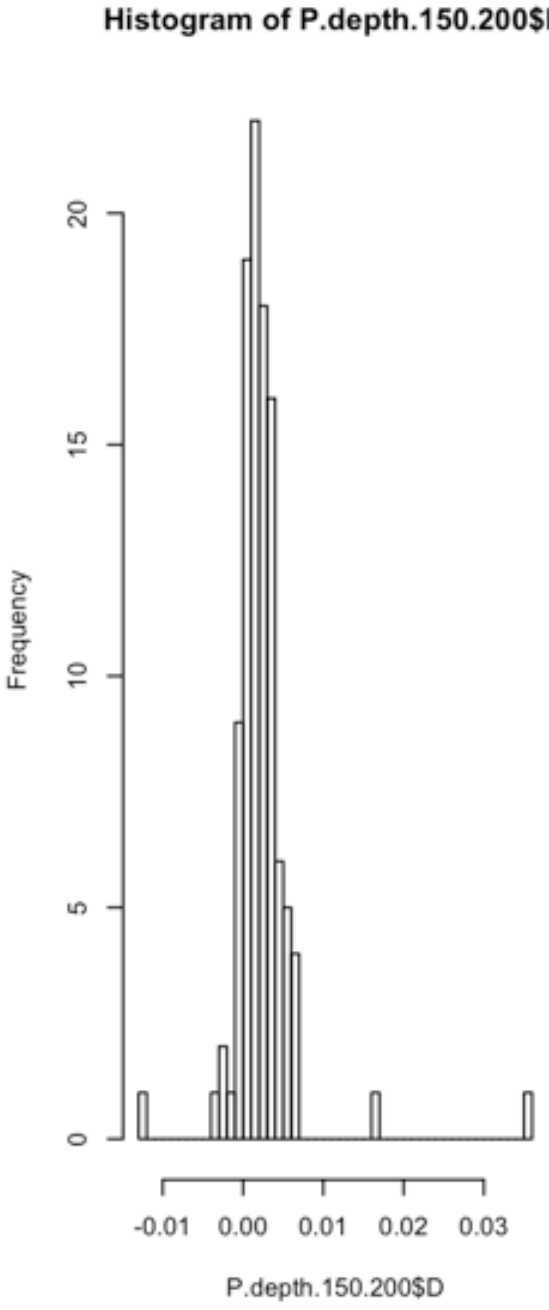
Ndepth (150.200) is not very normal, but log transformation increases normality. Outliers present in box plot

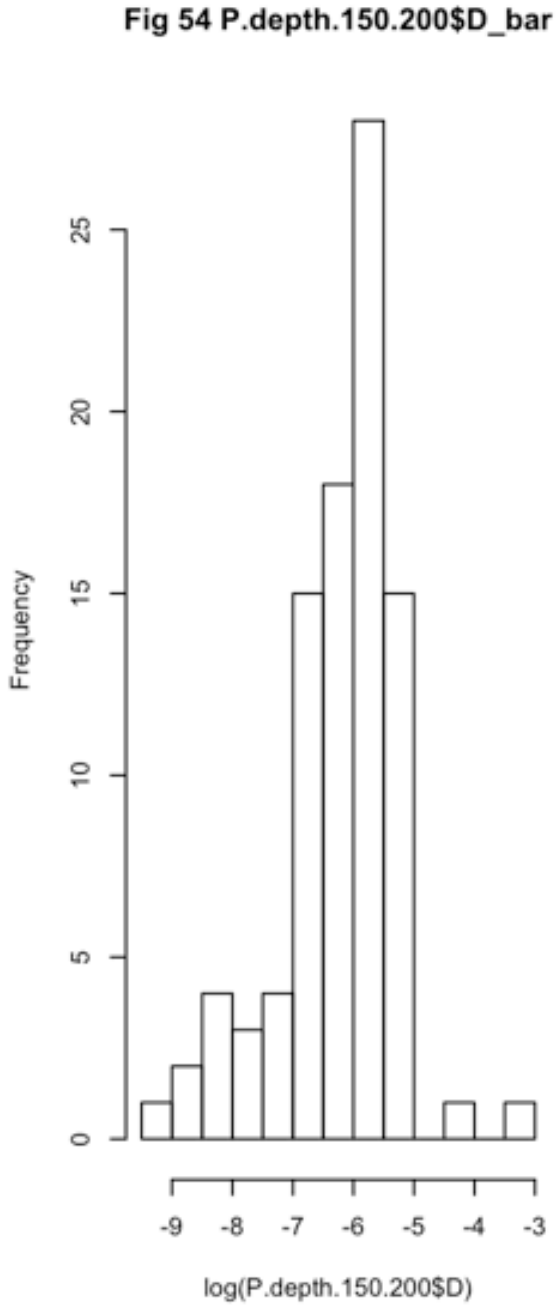


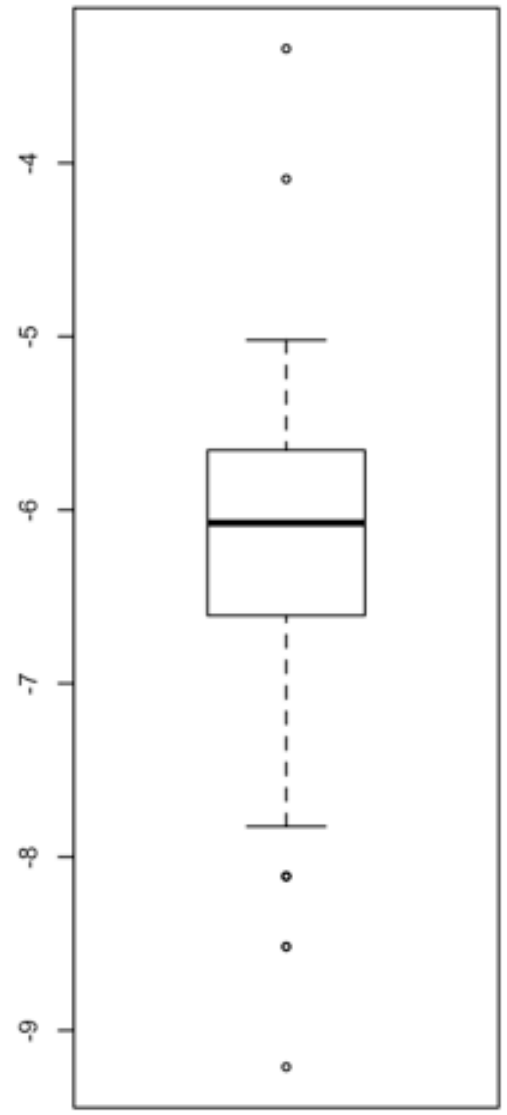




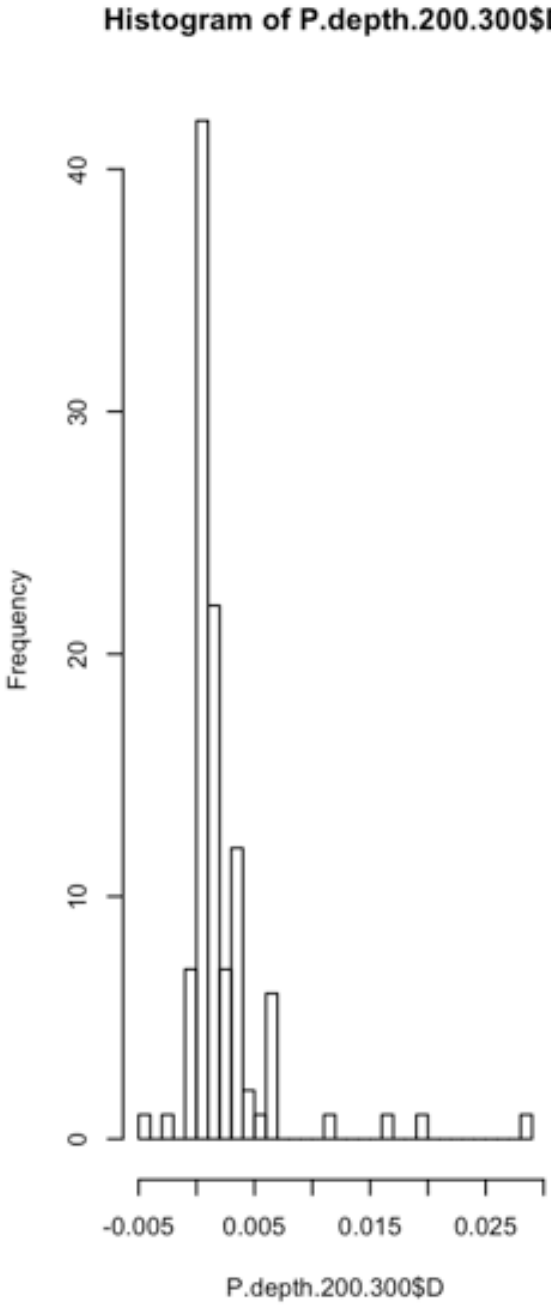
Ndepth (200.300) is skewed, but log transformation increases normality. Outliers present in box plot

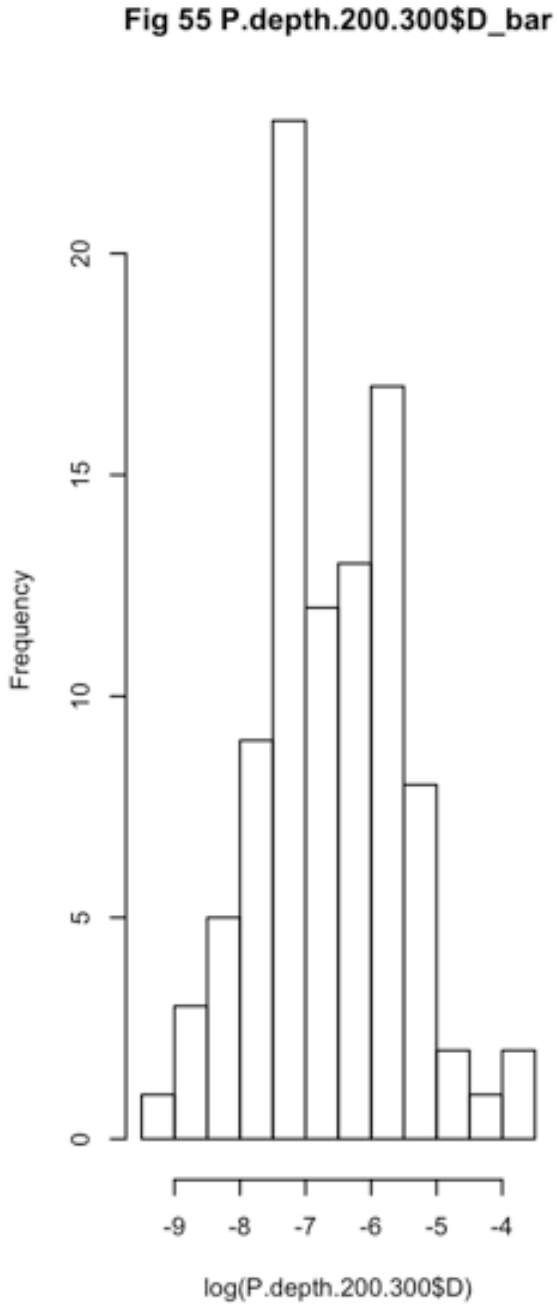


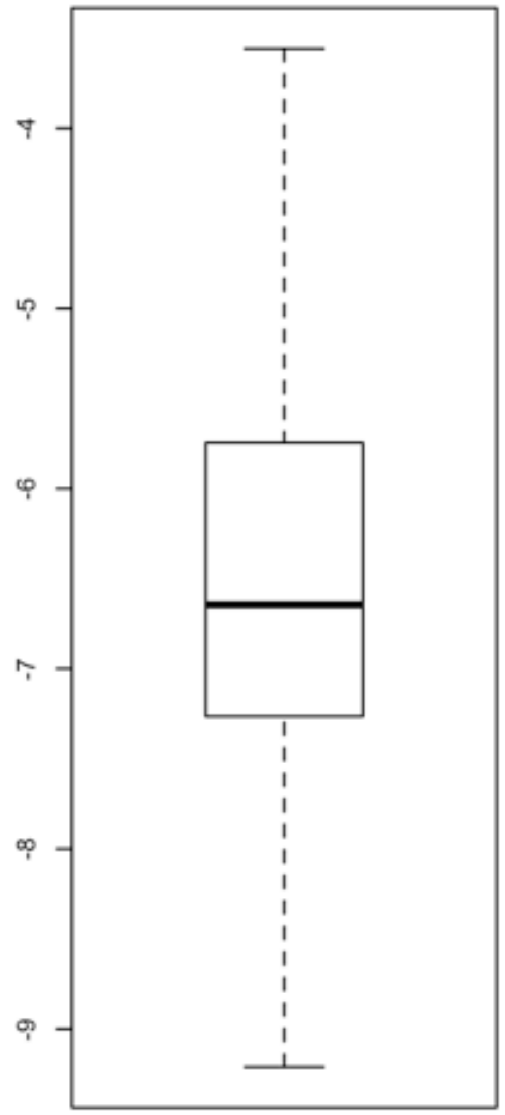




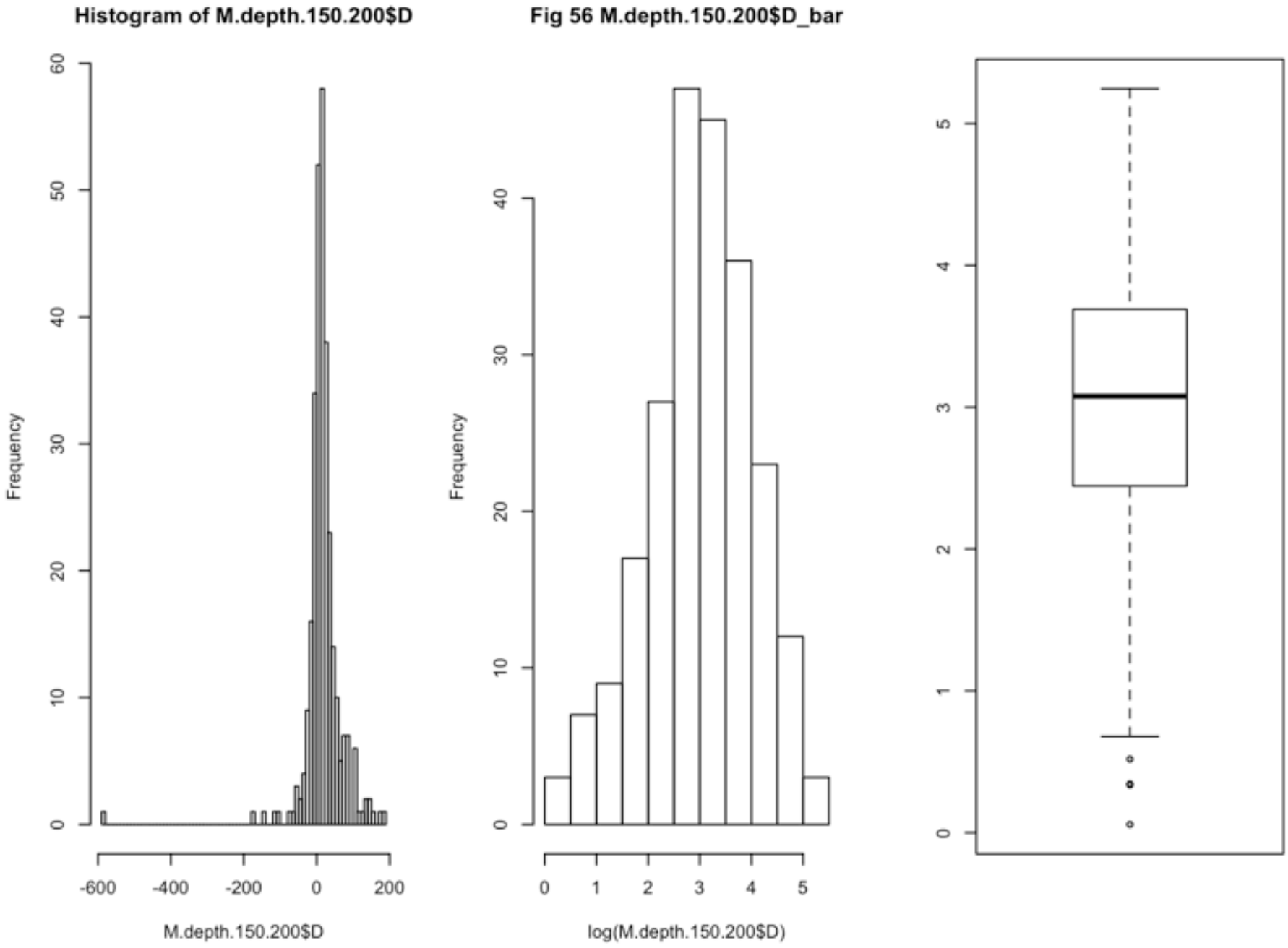
Pdepth (150.200) is not very normal, but log transformation increases normality. Outliers present in box plot



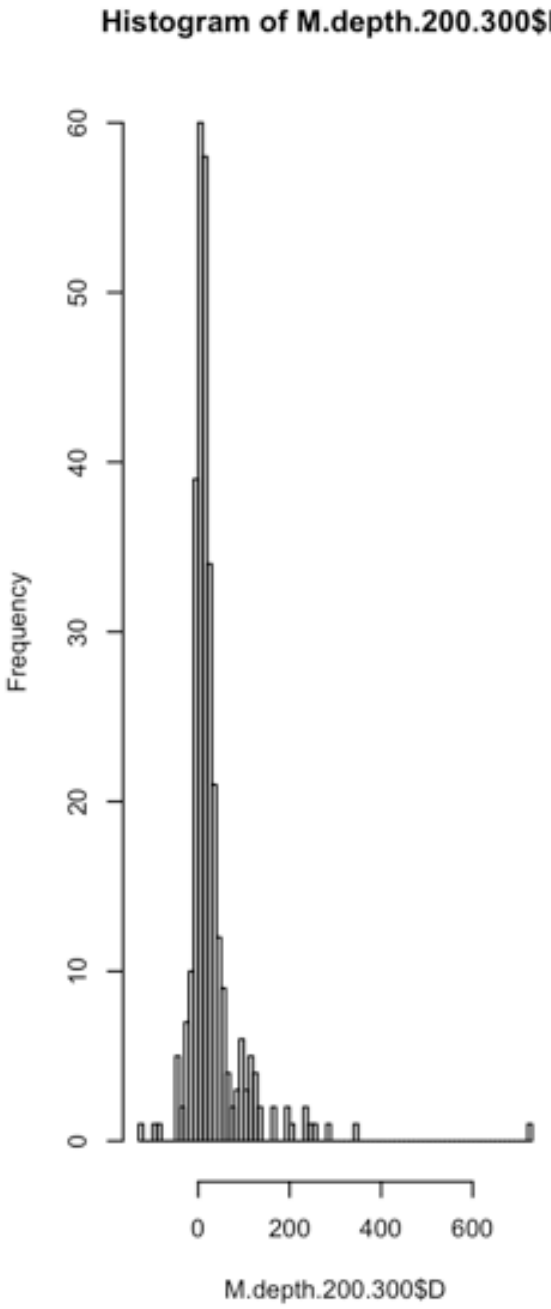


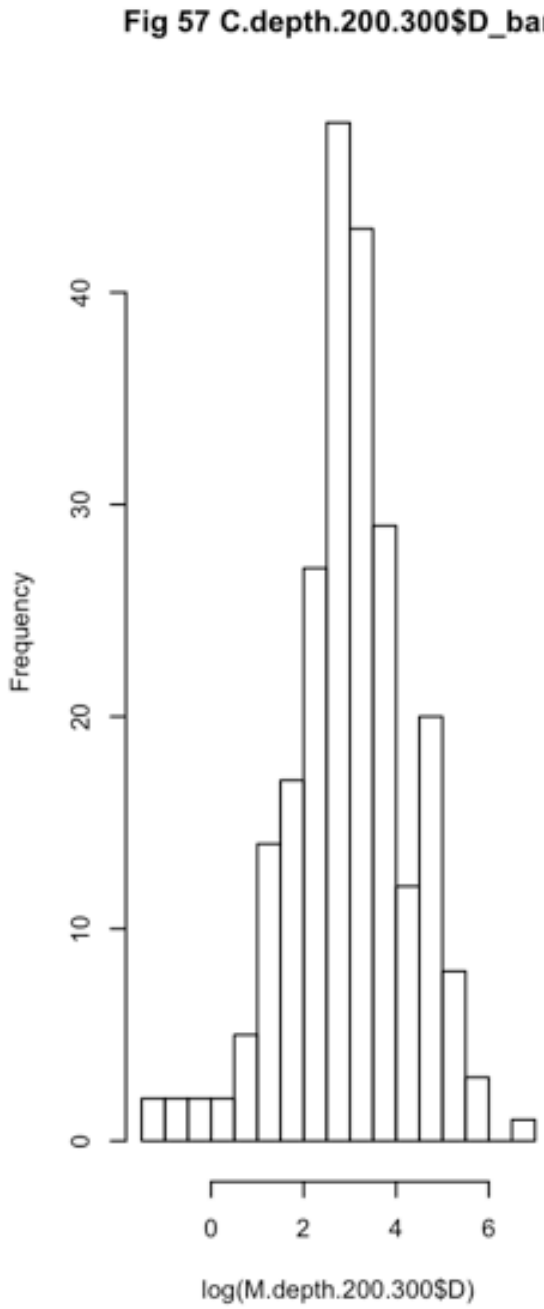


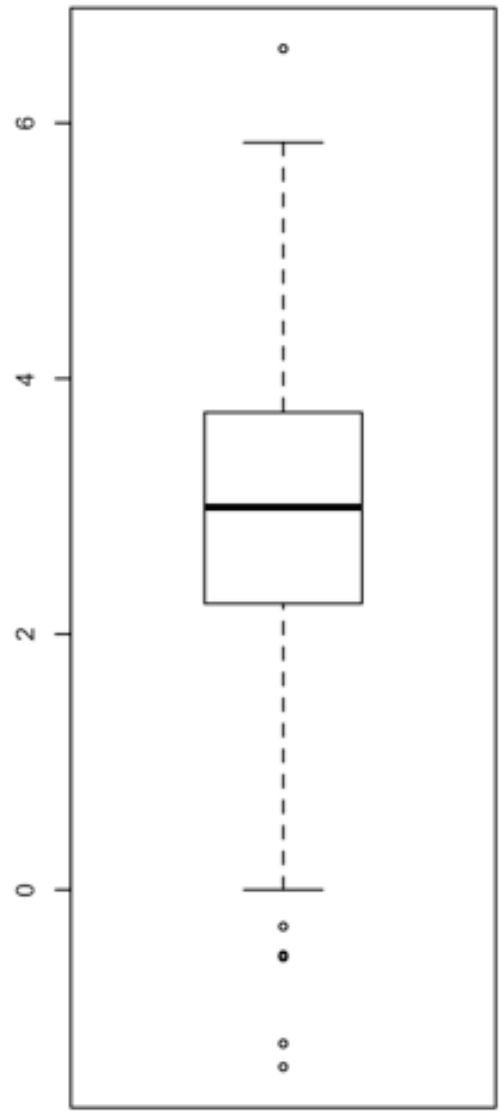
Pdepth (200.300) is not very normal, but log transformation increases normality. NO outliers present in box plot



Mdepth (150.200) is fairly normal, but log transformation increases normality. Outliers present in box plot







Mdepth (200.300) is not normal, but log transformation increases normality. Outliers present in box plot

Proceed with hypothesis test of log transformed data

Hyp 1

μ_1 = mean Carbon flux at depth 150 μ_2 = mean Carbon flux at depth 200

H_0 : Carbon flux does not change between 150 and 200 meter. $\mu_1 = \mu_2$ H_1 : Carbon flux is lower at 200 meter then at 150 meter. $\mu_1 > \mu_2$

```
##
##  Paired t-test
##
## data:  log(C.depth.150.200$C_avg.150) and log(C.depth.150.200$C_avg.200)
## t = 14.668, df = 302, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.2578256       Inf
## sample estimates:
## mean of the differences
##               0.2905015
```

super low p-value,significant, reject the null

Hyp 2

μ_1 = mean Carbon flux at depth 200 μ_2 = mean Carbon flux at depth 300

H_0 : Carbon flux does not change between 200 and 300 meter. $\mu_1 = \mu_2$ H_1 : Carbon flux is lower at 300 meter then at 200 meter. $\mu_1 > \mu_2$

```
##
##  Paired t-test
##
## data:  log(C.depth.200.300$C_avg.200) and log(C.depth.200.300$C_avg.300)
## t = 13.324, df = 299, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.3254642       Inf
## sample estimates:
## mean of the differences
##               0.3714623
```

very low pval. We reject the null, significant

Hyp 3

μ_1 = mean Nitrogen flux at depth 150 μ_2 = mean Nitrogen flux at depth 200

H_0 : Nitrogen flux does not change between 150 and 200 meter. $\mu_1 = \mu_2$ H_1 : Nitrogen flux is lower at 200 meter then at 150 meter. $\mu_1 > \mu_2$

```
##
##  Paired t-test
##
## data:  log(N.depth.150.200$N_avg.150) and log(N.depth.150.200$N_avg.200)
## t = 14.918, df = 300, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.3202794       Inf
## sample estimates:
## mean of the differences
##                0.360109
```

super low p-value,significant, reject the null

Hyp 4

μ_1 = mean Nitrogen flux at depth 200 μ_2 = mean Nitrogen flux at depth 300

H_0 : Nitrogen flux does not change between 200 and 300 meter. $\mu_1 = \mu_2$ H_1 : Nitrogen flux is lower at 300 meter then at 200 meter. $\mu_1 > \mu_2$

```
##
##  Paired t-test
##
## data:  log(N.depth.200.300$N_avg.200) and log(N.depth.200.300$N_avg.300)
## t = 15.938, df = 297, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4047791       Inf
## sample estimates:
## mean of the differences
##                0.4515234
```

super low p-value,significant, reject the null

Hyp 5

μ_1 = mean Phosporous flux at depth 150 μ_2 = mean Phosporous flux at depth 200

H_0 : Phosporous flux does not change between 150 and 200 meter. $\mu_1 = \mu_2$ H_1 : Phosporous flux is lower at 200 meter then at 150 meter. $\mu_1 > \mu_2$

```
##
##  Paired t-test
##
## data:  log(P.depth.150.200$P_avg.150) and log(P.depth.150.200$P_avg.200)
## t = 10.941, df = 105, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.3862101       Inf
## sample estimates:
## mean of the differences
##               0.4552613
```

super low p-value,significant, reject the null

Hyp 6

μ_1 = mean Phosporous flux at depth 200 μ_2 = mean Phosporous flux at depth 300

H_0 : Phosporous flux does not change between 200 and 300 meter. $\mu_1 = \mu_2$ H_1 : Phosporous flux is lower at 300 meter then at 200 meter. $\mu_1 > \mu_2$

```
##
##  Paired t-test
##
## data:  log(P.depth.200.300$P_avg.200) and log(P.depth.200.300$P_avg.300)
## t = 12.446, df = 104, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4504533       Inf
## sample estimates:
## mean of the differences
##               0.5197648
```

super low p-value,significant, reject the null

Hyp 7

μ_1 = mean Mass flux at depth 150 μ_2 = mean Mass flux at depth 200

H_0 : Mass flux does not change between 150 and 200 meter. $\mu_1 = \mu_2$ H_1 : Mass flux is lower at 200 meter then at 150 meter. $\mu_1 > \mu_2$

```
##
##  Paired t-test
##
## data:  log(M.depth.150.200$M_avg.150) and log(M.depth.150.200$M_avg.200)
## t = 11.049, df = 303, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.1613635       Inf
## sample estimates:
## mean of the differences
##               0.1896883
```

super low p-value,significant, reject the null

Hyp 8

μ_1 = mean Mass flux at depth 200 μ_2 = mean Mass flux at depth 300

H_0 : Mass flux does not change between 200 and 300 meter. $\mu_1 = \mu_2$ H_1 : Mass flux is lower at 300 meter then at 200 meter. $\mu_1 > \mu_2$

```
##
##  Paired t-test
##
## data:  log(M.depth.200.300$M_avg.200) and log(M.depth.200.300$M_avg.300)
## t = 12.054, df = 300, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.2143744       Inf
## sample estimates:
## mean of the differences
##               0.2483726
```

super low p-value,significant, reject the null

Simple Linear Regression via LS

How does each flux predict Mass Flux?

Regression Model: $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$

Best Fit line: $f(x) = \hat{\beta_0} + \hat{\beta_1} x$

```
##
## Call:
## lm(formula = log(M_avg) ~ log(C_avg), data = Data.noNA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.83440 -0.25212 -0.04393  0.18902  2.05452
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.00513    0.08106   24.74   <2e-16 ***
## log(C_avg)   0.83365    0.02835   29.41   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3784 on 303 degrees of freedom
## Multiple R-squared:  0.7405, Adjusted R-squared:  0.7396
## F-statistic: 864.6 on 1 and 303 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = log(M_avg) ~ log(N_avg), data = Data.noNA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96805 -0.31451 -0.06096  0.25532  1.77284
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.58148    0.03984   89.89   <2e-16 ***
## log(N_avg)   0.76636    0.03269   23.44   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4429 on 303 degrees of freedom
## Multiple R-squared:  0.6446, Adjusted R-squared:  0.6434
## F-statistic: 549.6 on 1 and 303 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = log(M_avg) ~ log(P_avg), data = Data.noNA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.47045 -0.37370 -0.05109  0.33124  2.41221
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.86337    0.19945   34.41   <2e-16 ***
## log(P_avg)   0.45385    0.03482   13.03   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5946 on 303 degrees of freedom
## Multiple R-squared:  0.3592, Adjusted R-squared:  0.3571
## F-statistic: 169.9 on 1 and 303 DF,  p-value: < 2.2e-16
```

Fig 58 M v. C

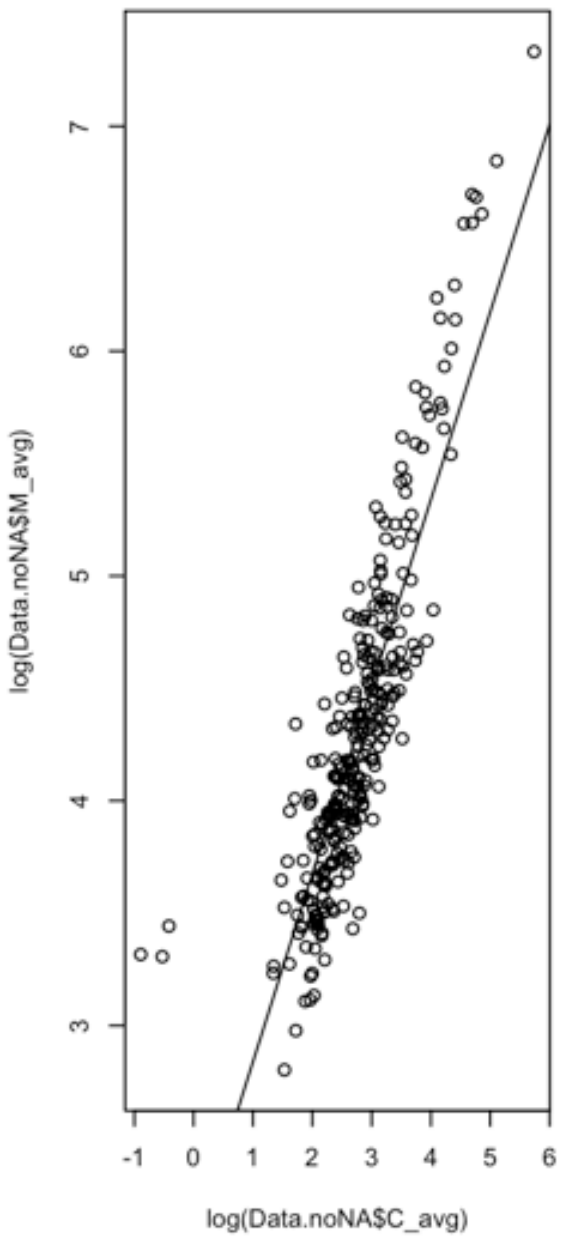


Fig 59 M v. N

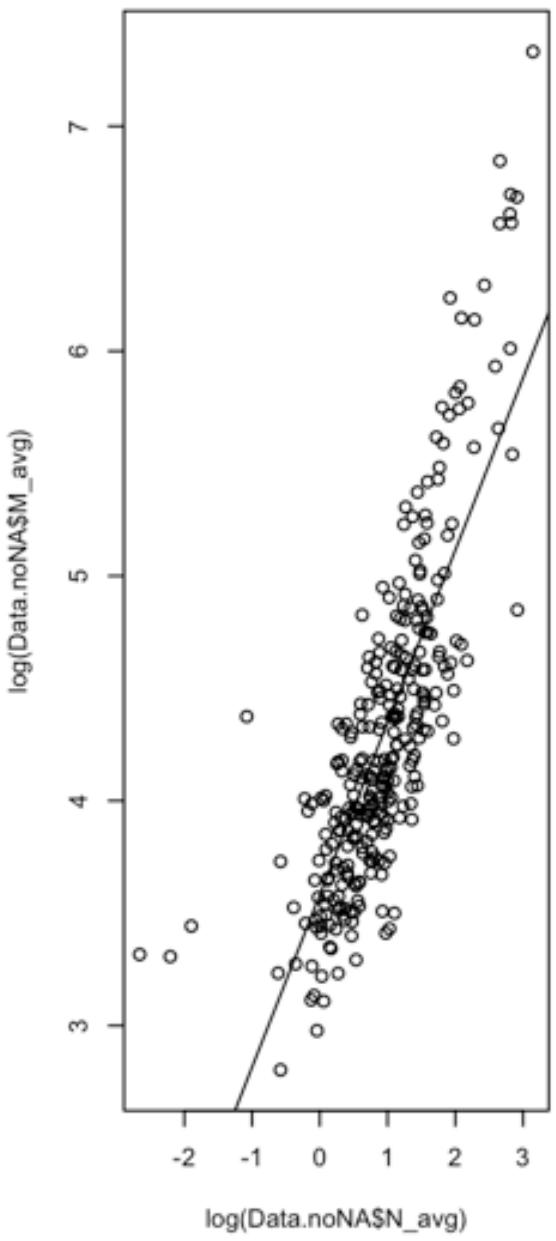
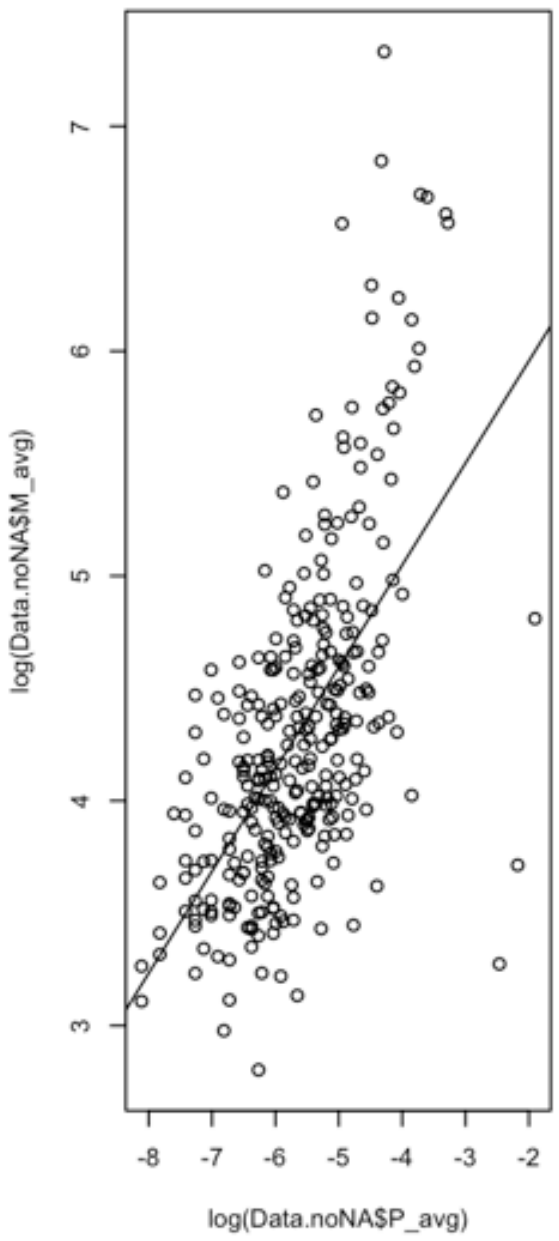


Fig 60 M v. P



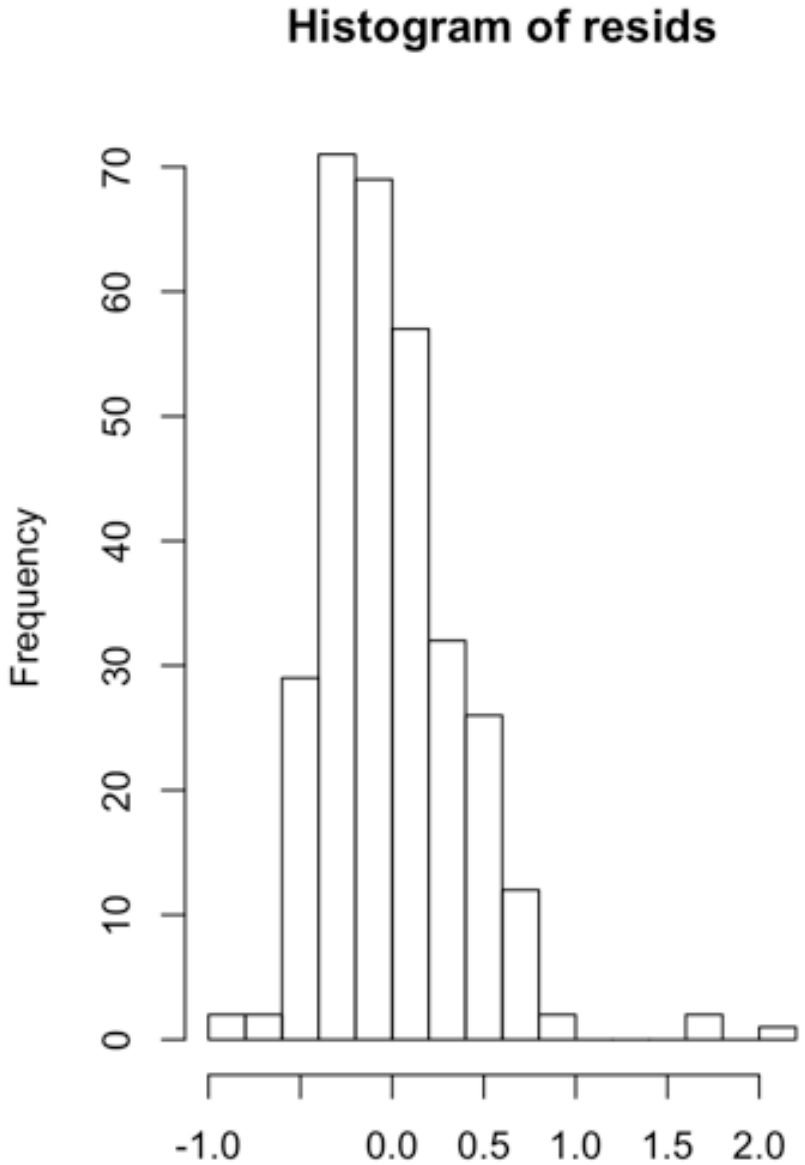
Both predictor and predictee had to be log transformed in order to make this data more linear with less clumping for all fluxes

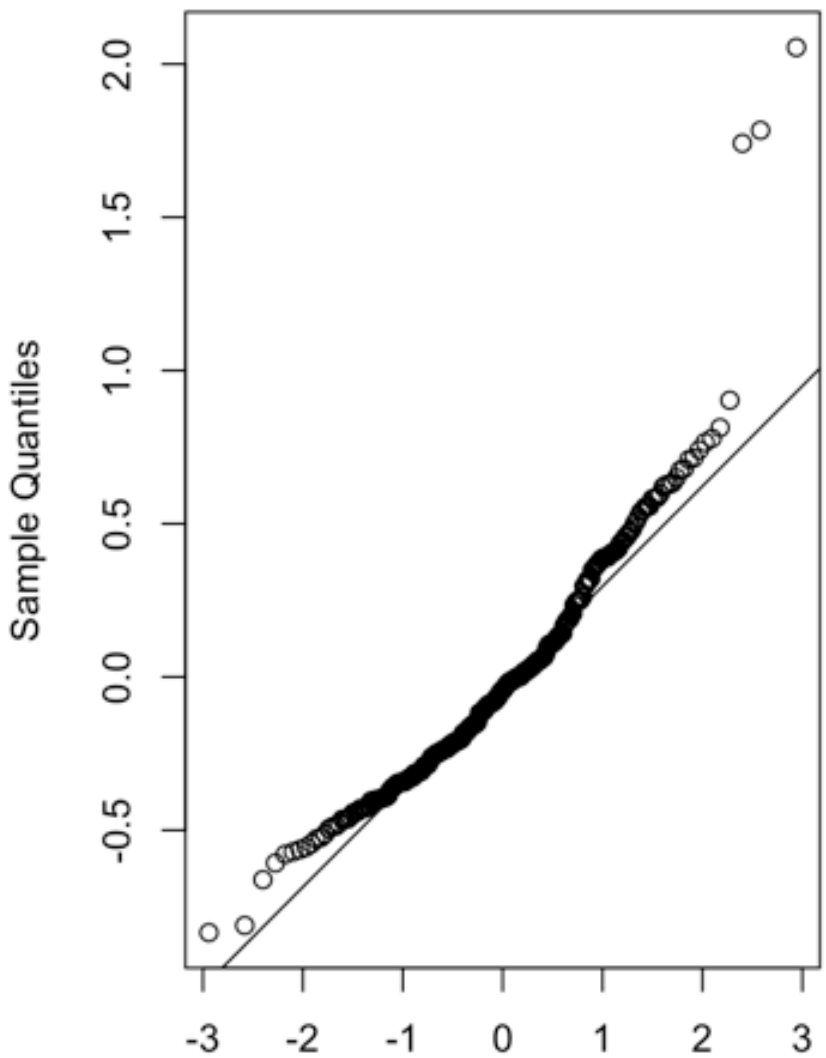
For C flux, multilple r squared is 0.7405, so pretty good fit. Some clumping still

For N flux, multiple r squared is 0.6446, so still pretty decent, just not as strong as C flux. Some clumping still

For P flux, multiple r squared is 0.3592, so very poor fit and a lot of clumping

Carbon flux seems to be the best predictor of Mass flux. Lets run some EDA.


Histogram of resids
Frequency
0
10
20
30
40
50
60
70
-1.0
0.0
0.5
1.0
1.5
2.0
resids


Normal Q-Q Plot
Sample Quantiles
-0.5
0.0
0.5
1.0
1.5
2.0
-3
-2
-1
0
1
2
3
Theoretical Quantiles

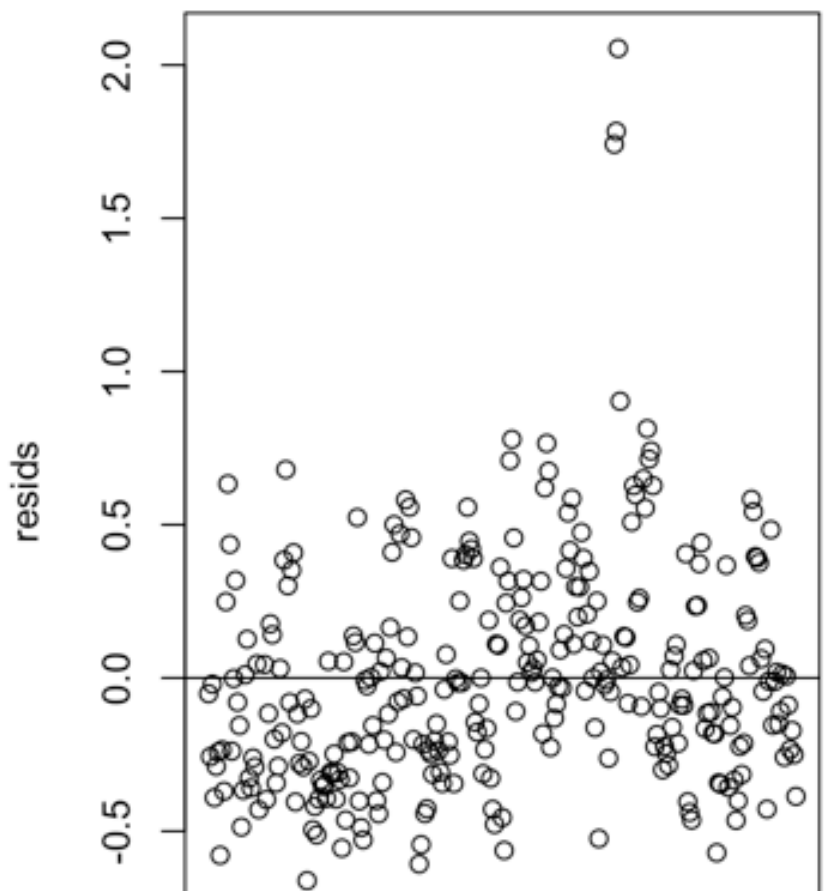


Fig 61 resid vs i
resids
-0.5
0.0
0.5
1.0
1.5
2.0

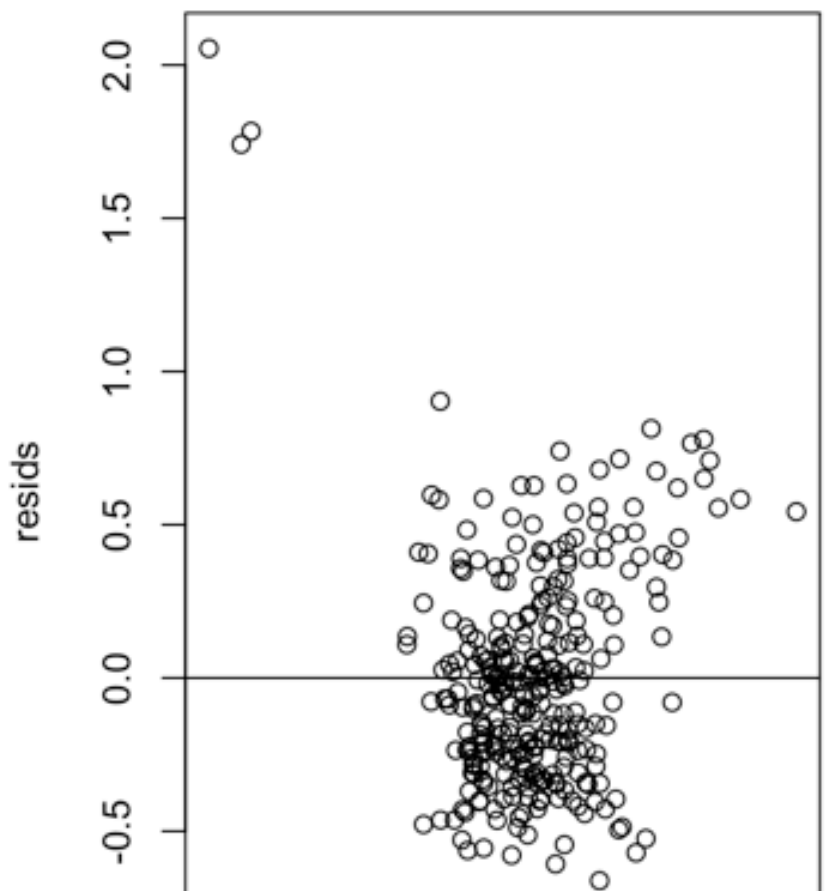
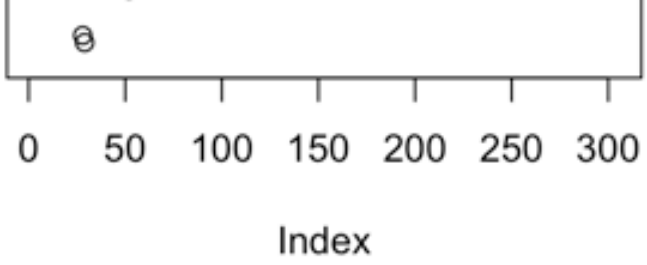
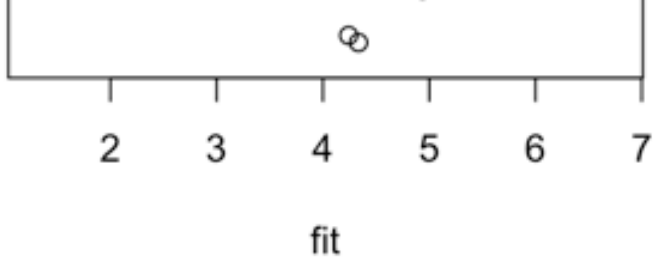


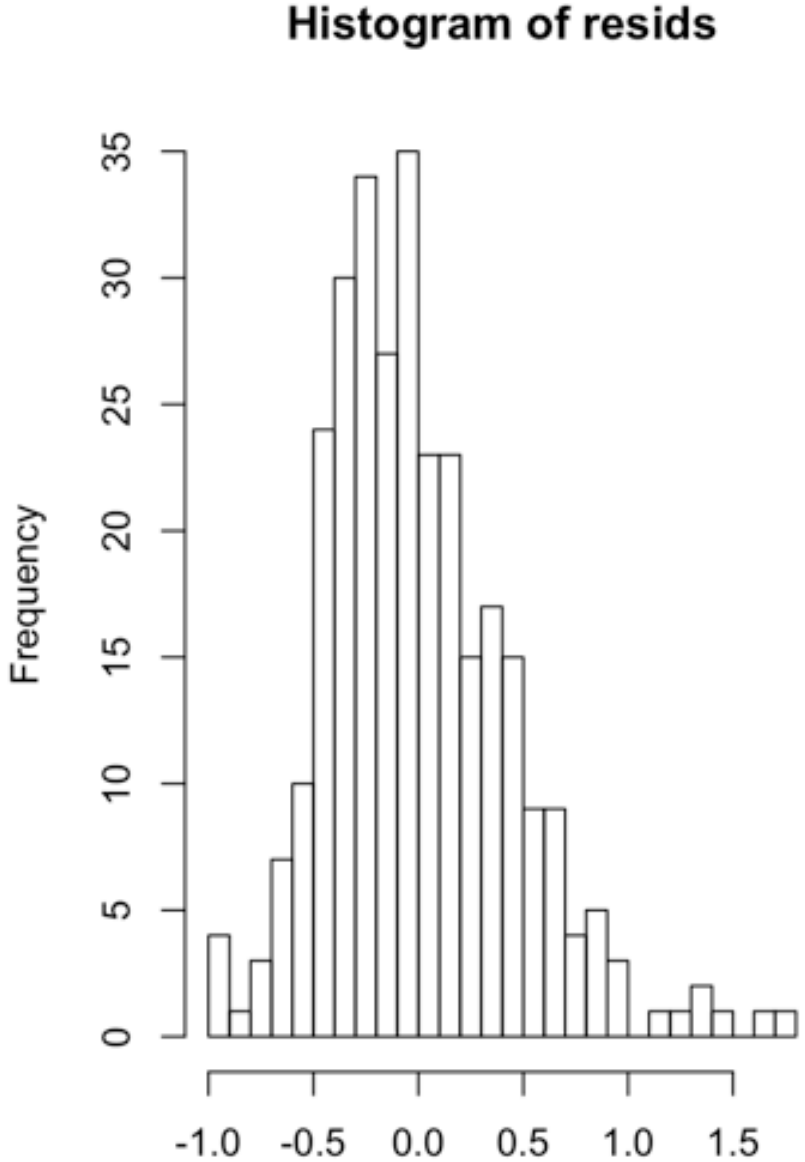
Fig 62 resid vs y_hat
resids
-0.5
0.0
0.5
1.0
1.5
2.0

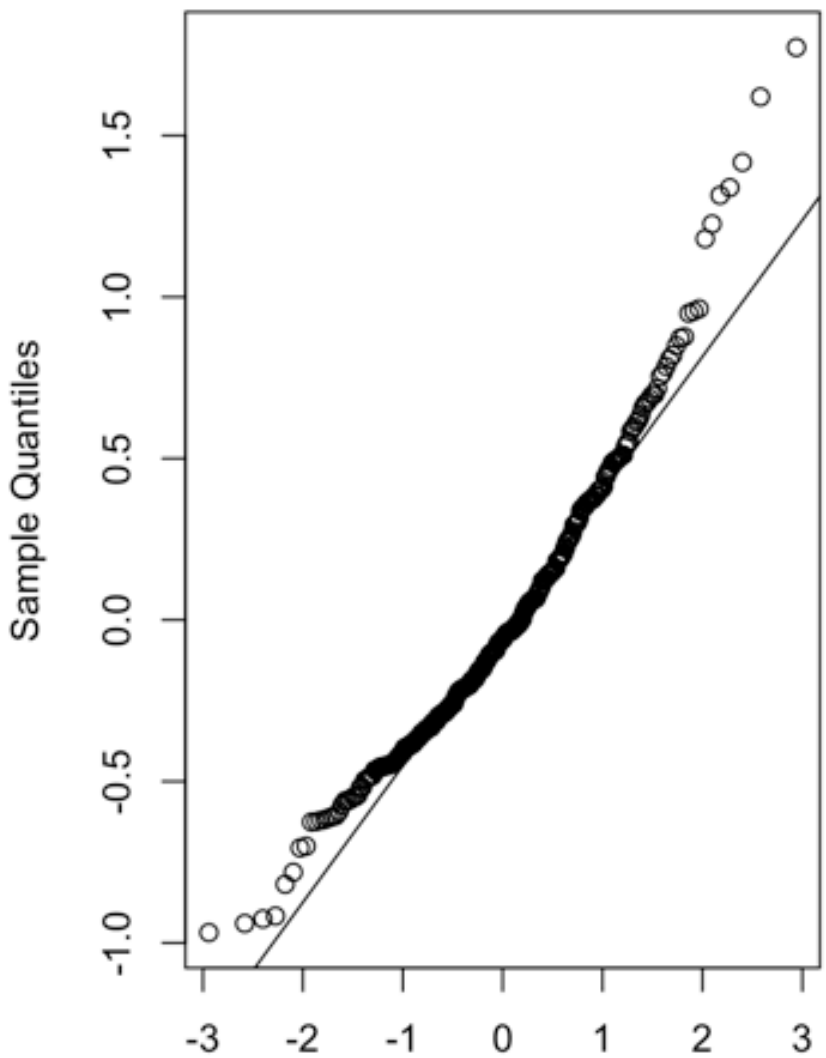




Not very normal, linearity is okay, but not great, both residual are not very evenly scattered and there is clumping around the line, so not heteroscedasicity, trifecta is broken

Lets compare these EDA's to that of Nitrogen Flux and Mass Flux. Nitrogen Flux was also a strong predictor of Mass Flux.


Histogram of resids
Frequency
0
5
10
15
20
25
30
35
-1.0
-0.5
0.0
0.5
1.0
1.5
resids


Normal Q-Q Plot
Sample Quantiles
-1.0
-0.5
0.0
0.5
1.0
1.5
-3
-2
-1
0
1
2
3
Theoretical Quantiles

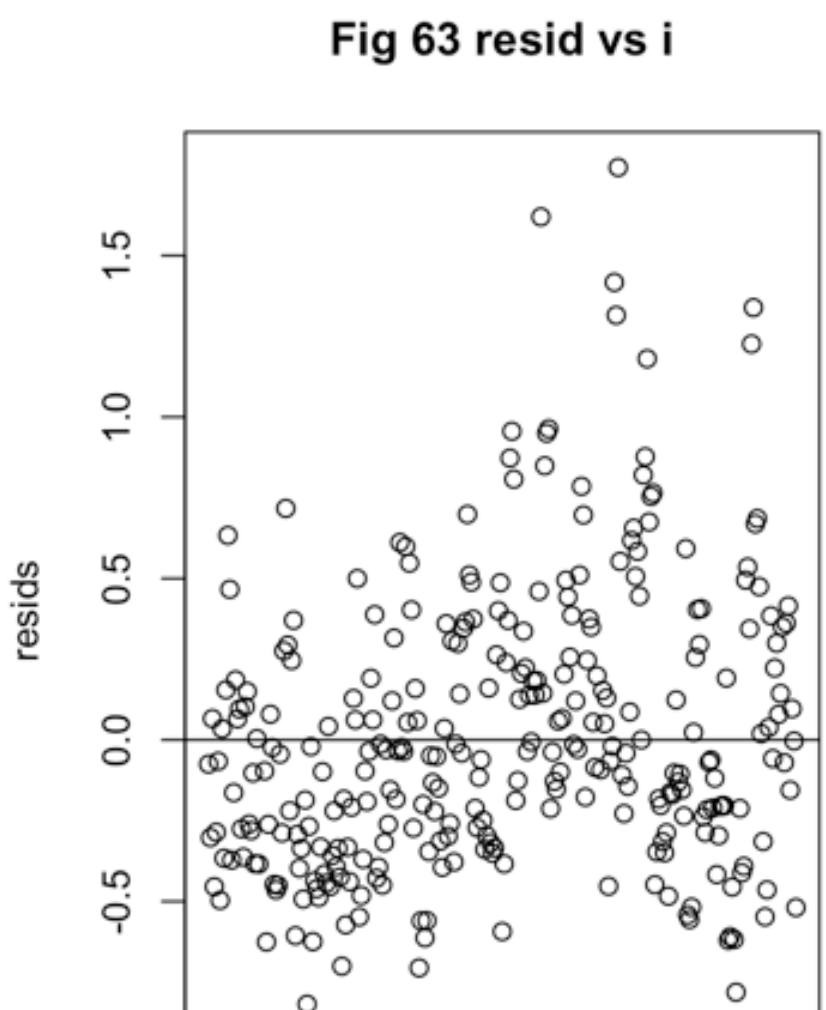


Fig 63 resid vs i
resids
-0.5
0.0
0.5
1.0
1.5

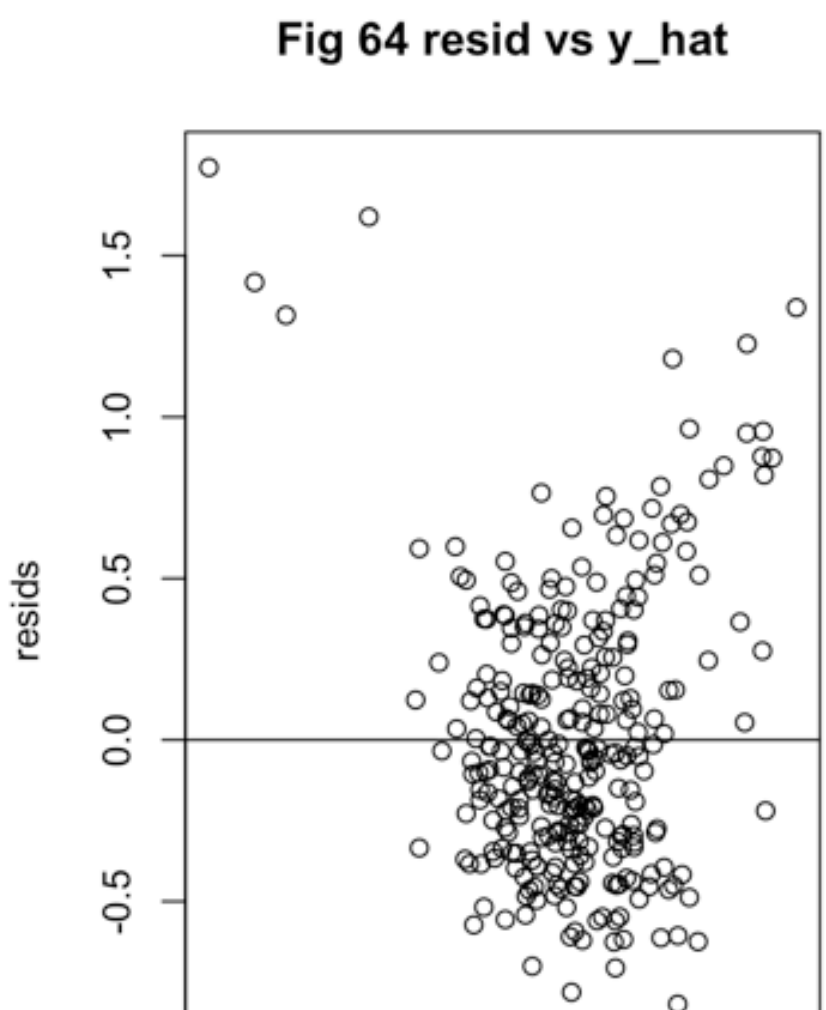
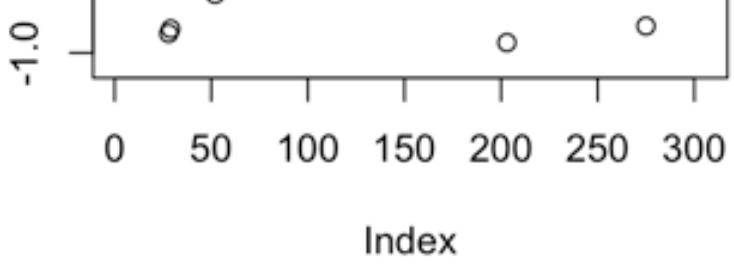
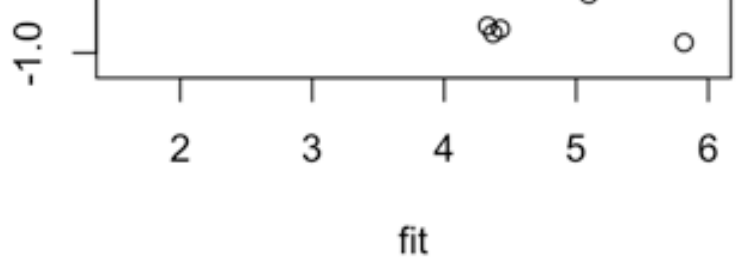


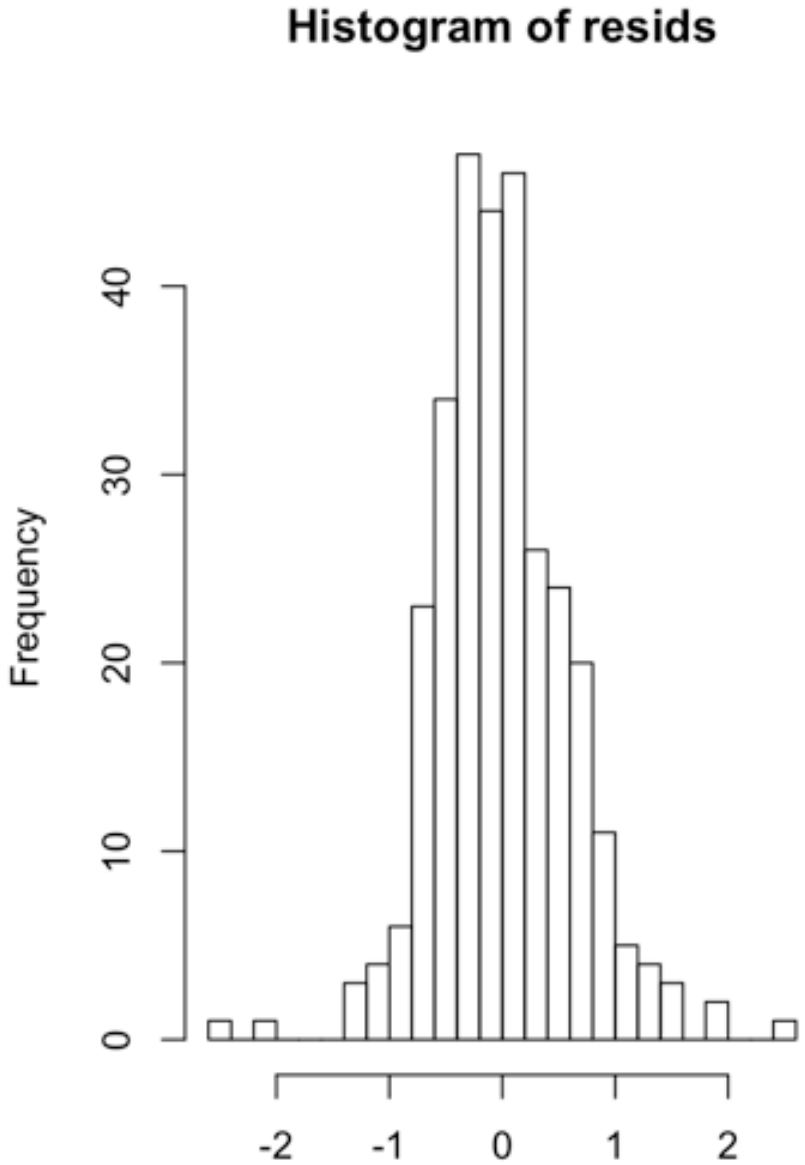
Fig 64 resid vs y_hat
resids
-0.5
0.0
0.5
1.0
1.5

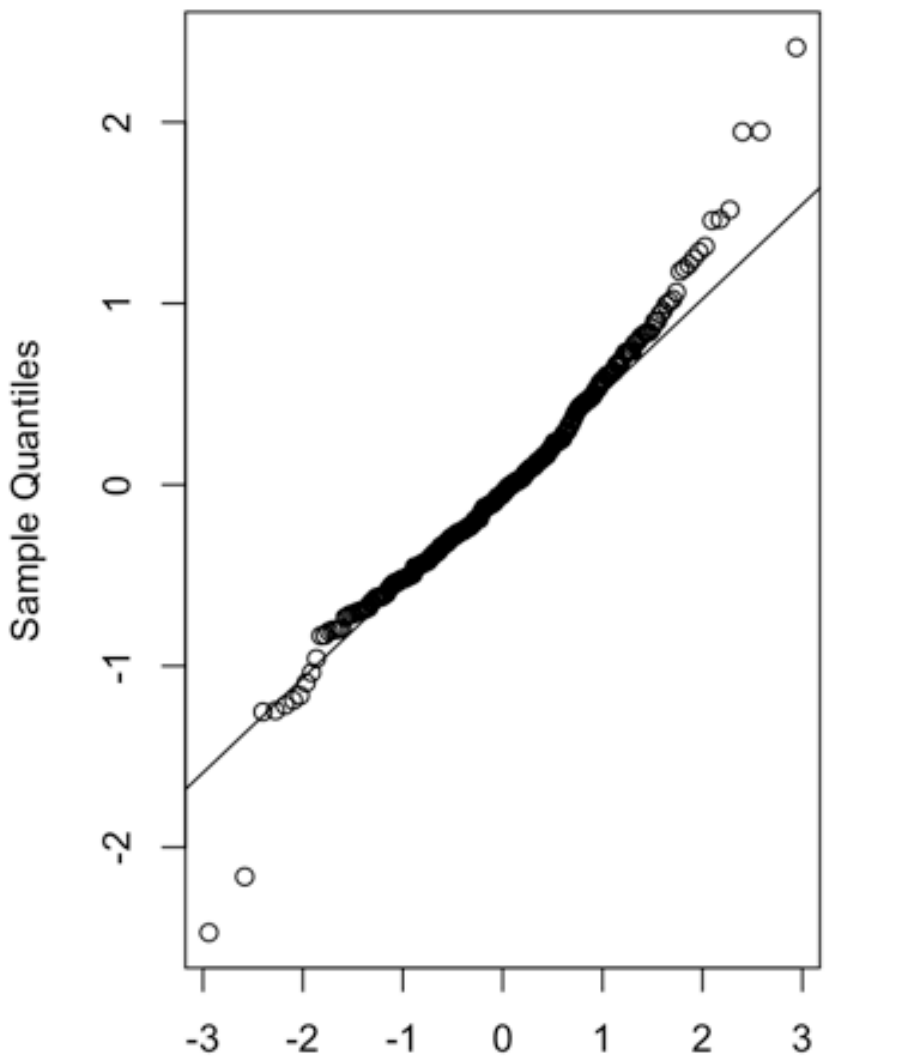




Not very normal, linearity is okay, but not great, both residual are not very evenly scattered and there is clumping around the line, so not heteroscedasicity, trifecta is broken about the same as C flux

Lets now compare these to Phosphorous Flux, a weak predcitor of Mass Flux.


Histogram of resids
Frequency
0
10
20
30
40
-2
-1
0
1
2
resids


Normal Q-Q Plot
Sample Quantiles
-2
-1
0
1
2
-3
-2
-1
0
1
2
3
Theoretical Quantiles

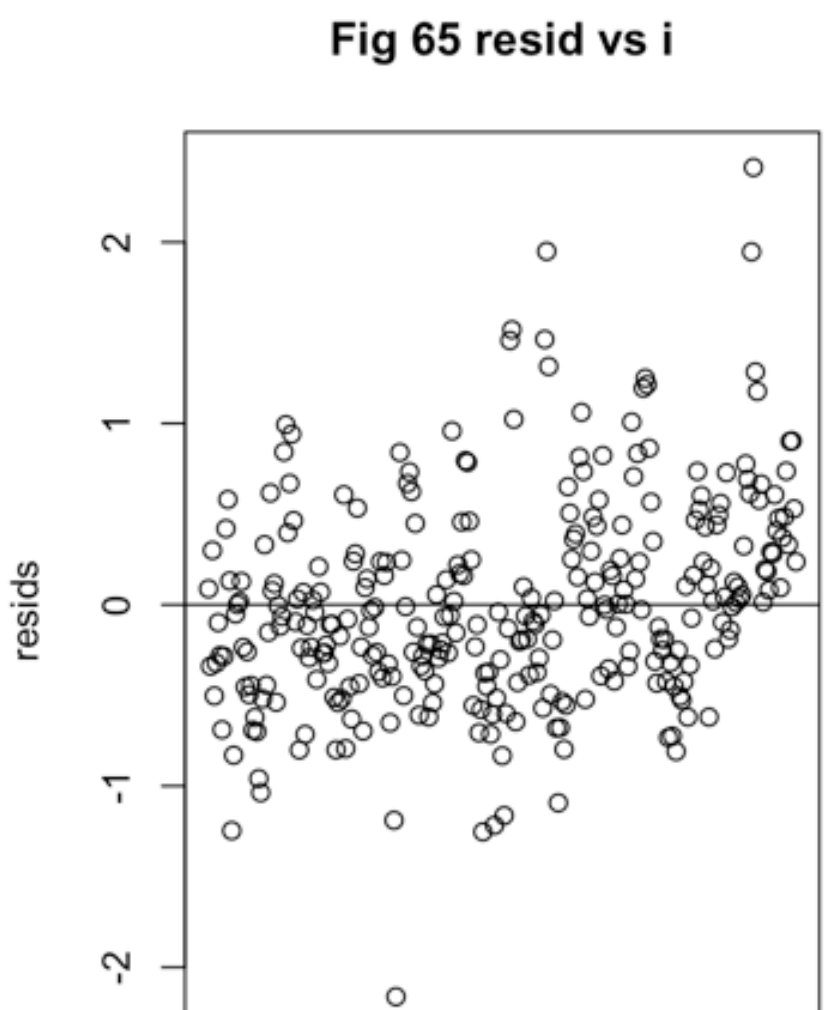


Fig 65 resid vs i
resids
-2
-1
0
1
2

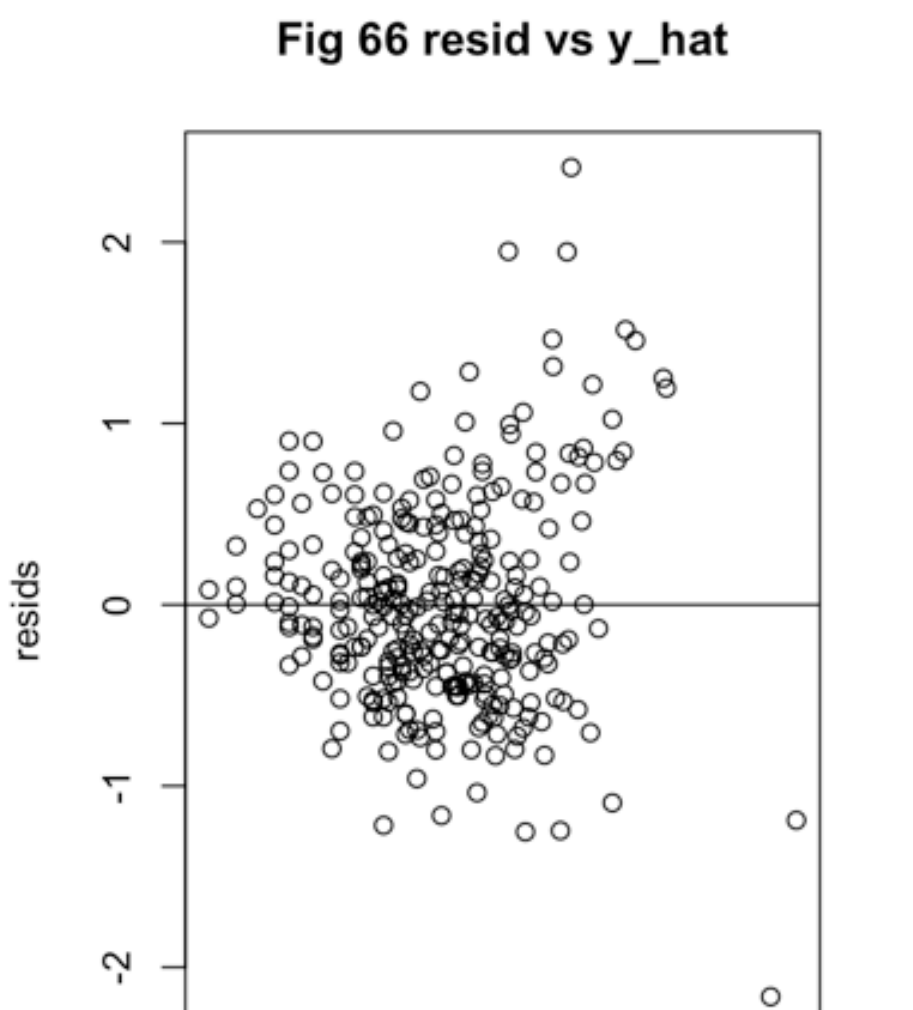
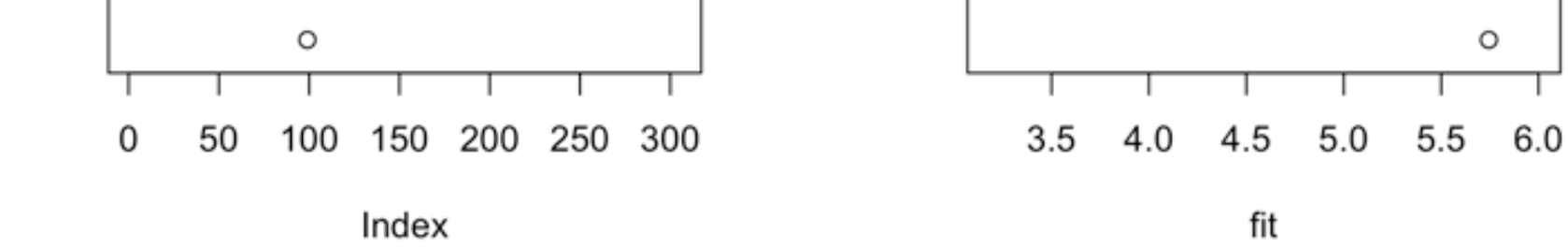


Fig 66 resid vs y_hat
resids
-2
-1
0
1
2

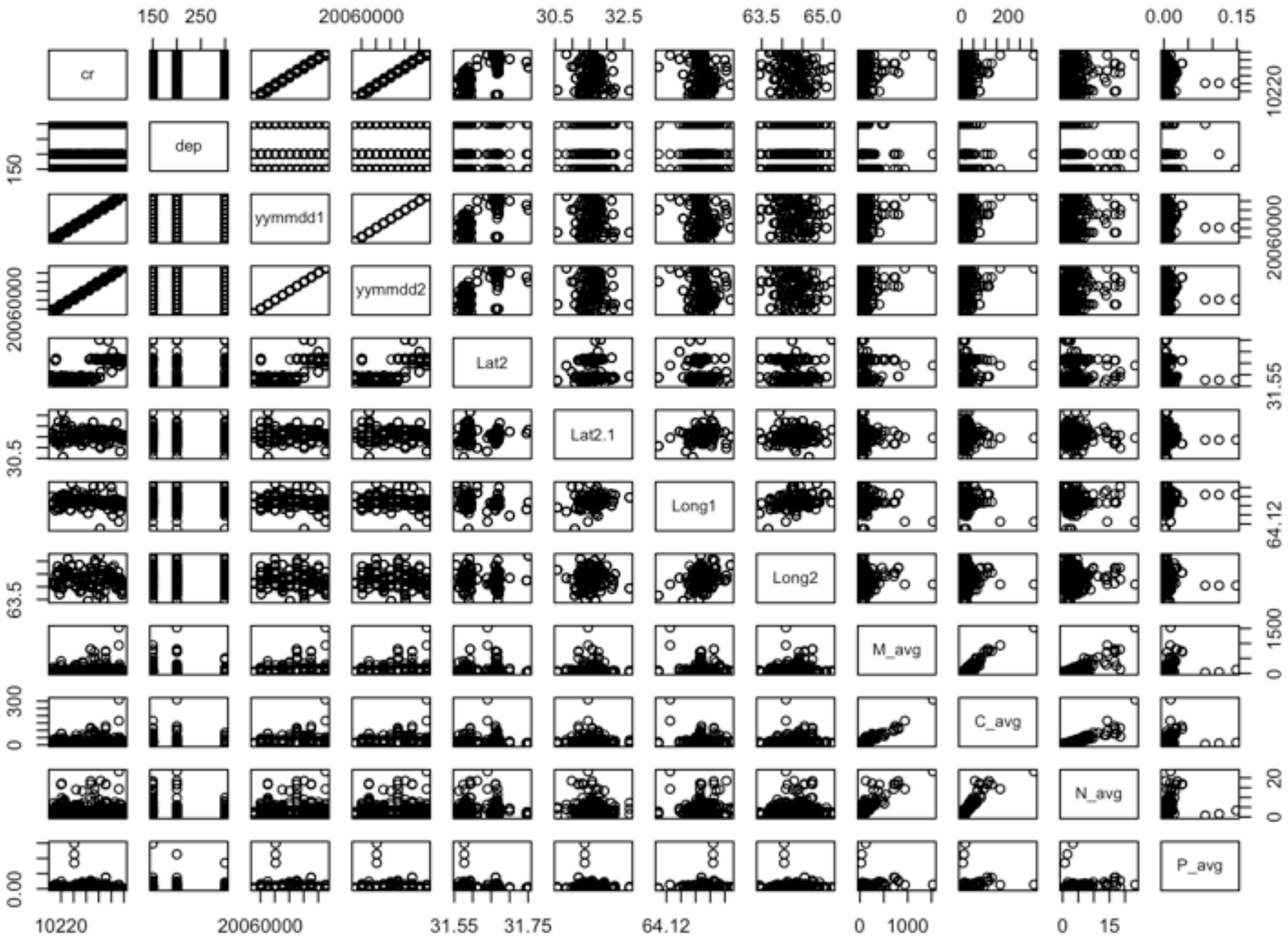


Somewhat normal, linearity is okay, but not great, both residual are not very evenly scattered and there is clumping around the line, so not heteroscedasicity, trifecta is broken

Do we see collinearity between the different fluxes that are predicting mass flux? Should we combine the fluxes?

Regression Model: $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$

Best Fit line: $f(x) = \hat{\beta}_0 + \hat{\beta}_1 x$



```
##
## Call:
## lm(formula = log(M_avg) ~ log(flux), data = Data.noNA)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8179 -0.2604 -0.0524  0.1949  2.0419
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.16923    0.07515   28.86   <2e-16 ***
## log(flux)    0.82943    0.02800   29.62   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3764 on 303 degrees of freedom
## Multiple R-squared:  0.7433, Adjusted R-squared:  0.7425
## F-statistic: 877.5 on 1 and 303 DF,  p-value: < 2.2e-16
```

Fig 67

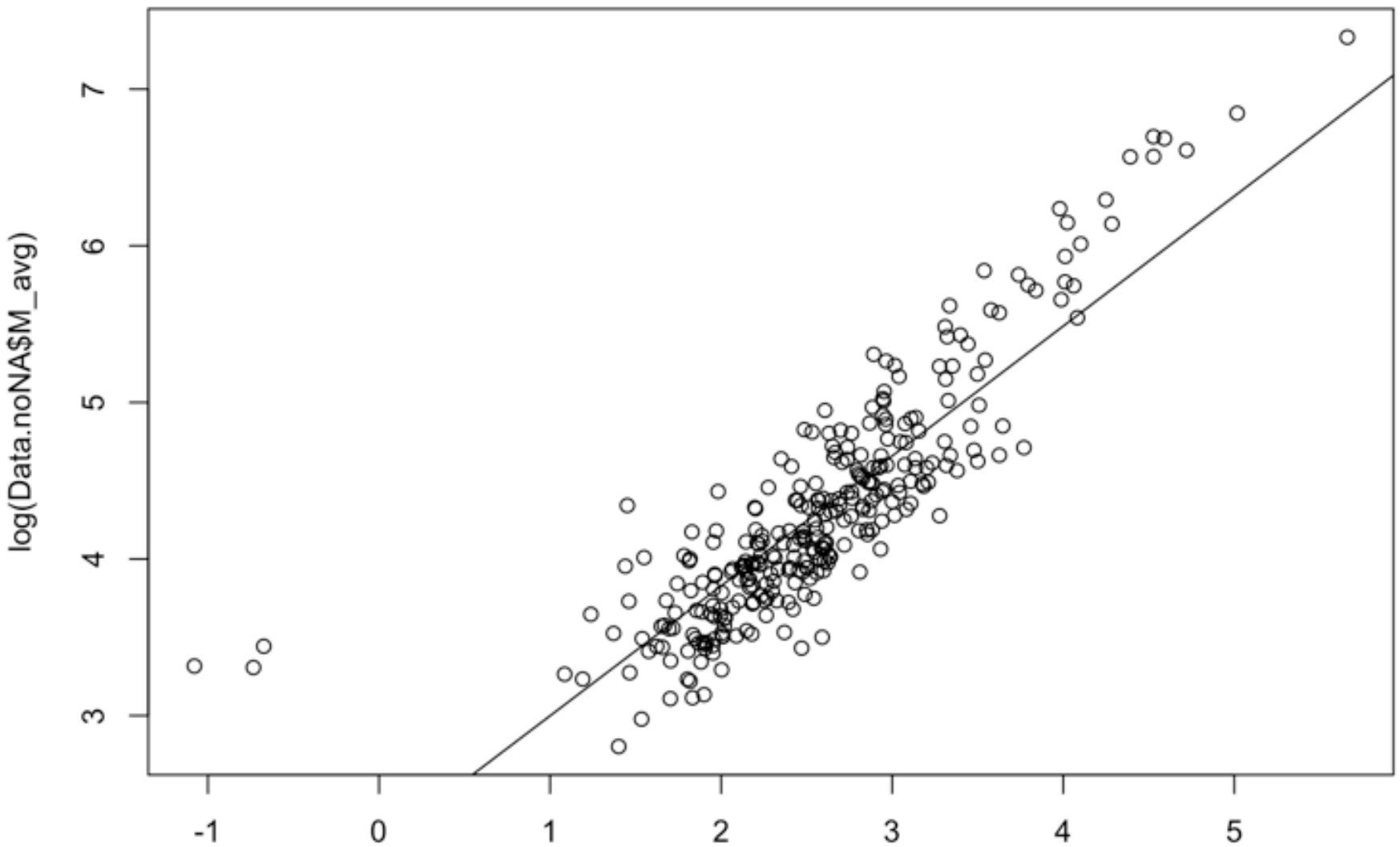
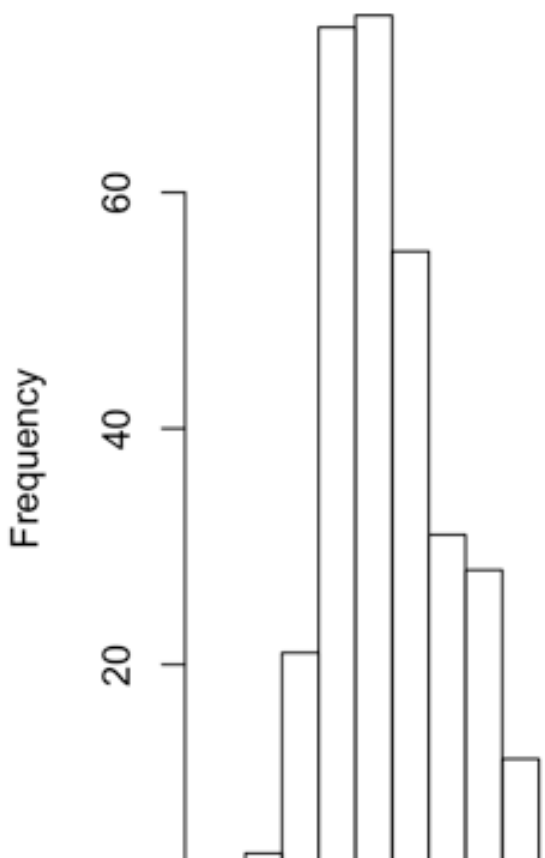
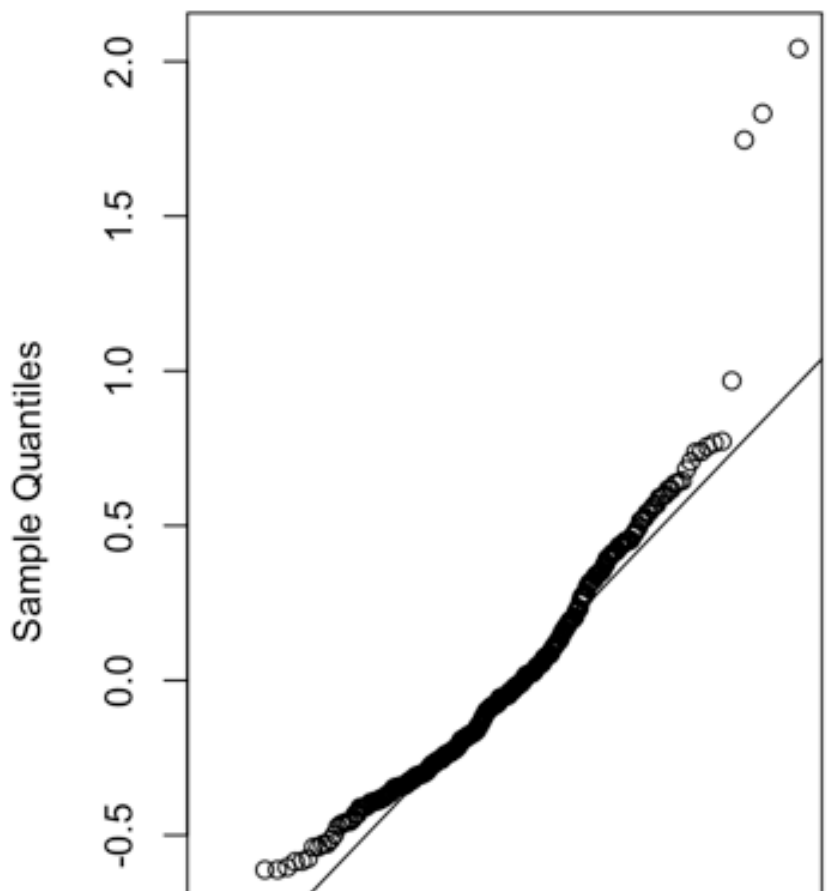
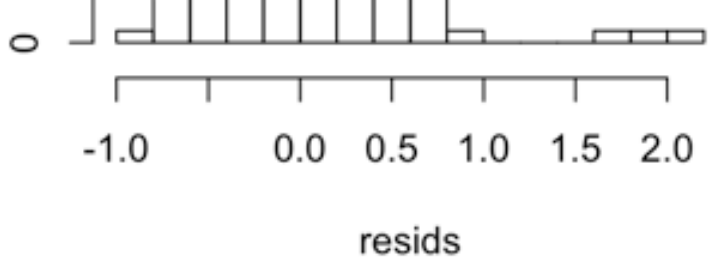


Fig 68



Normal Q-Q Plot





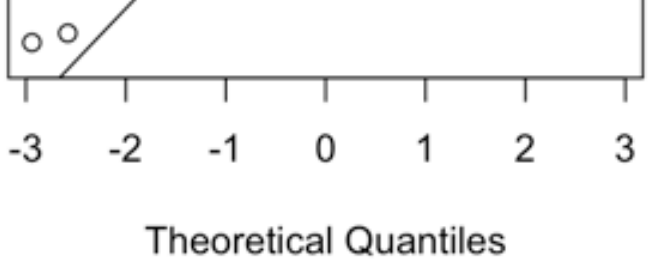


Fig 69 resid vs i

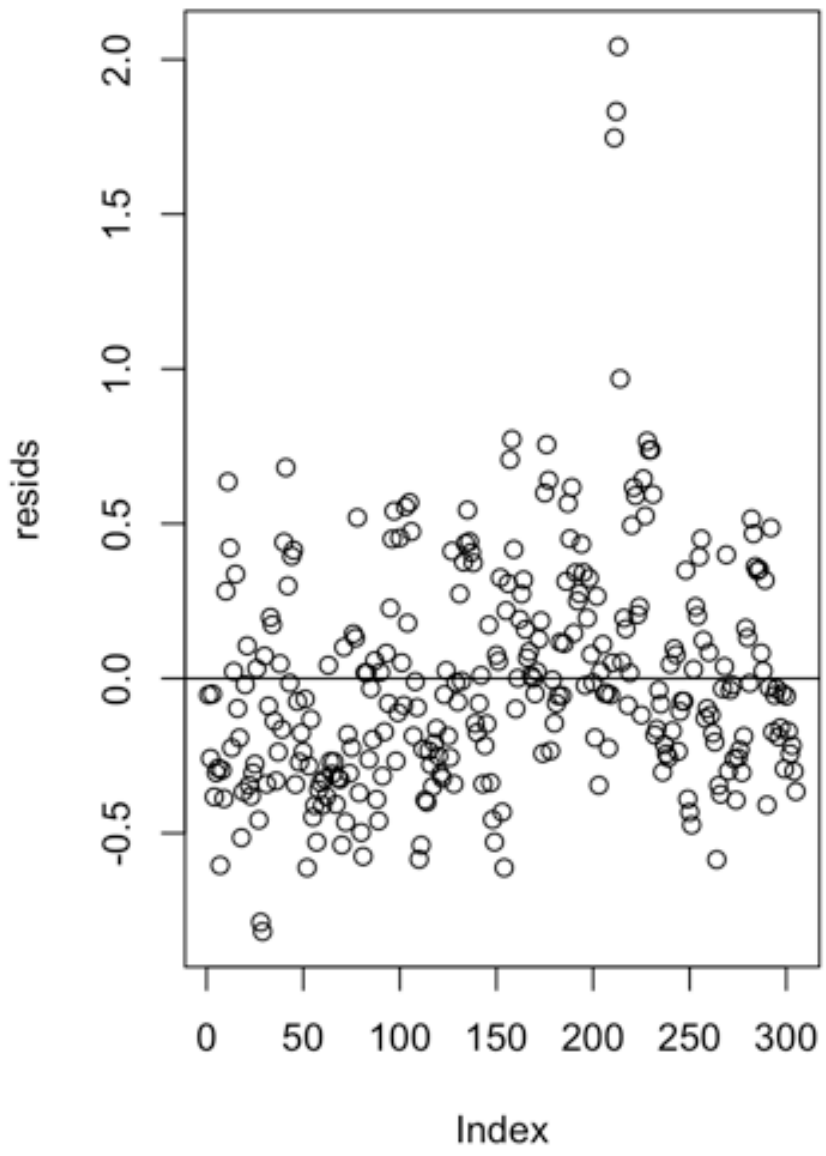
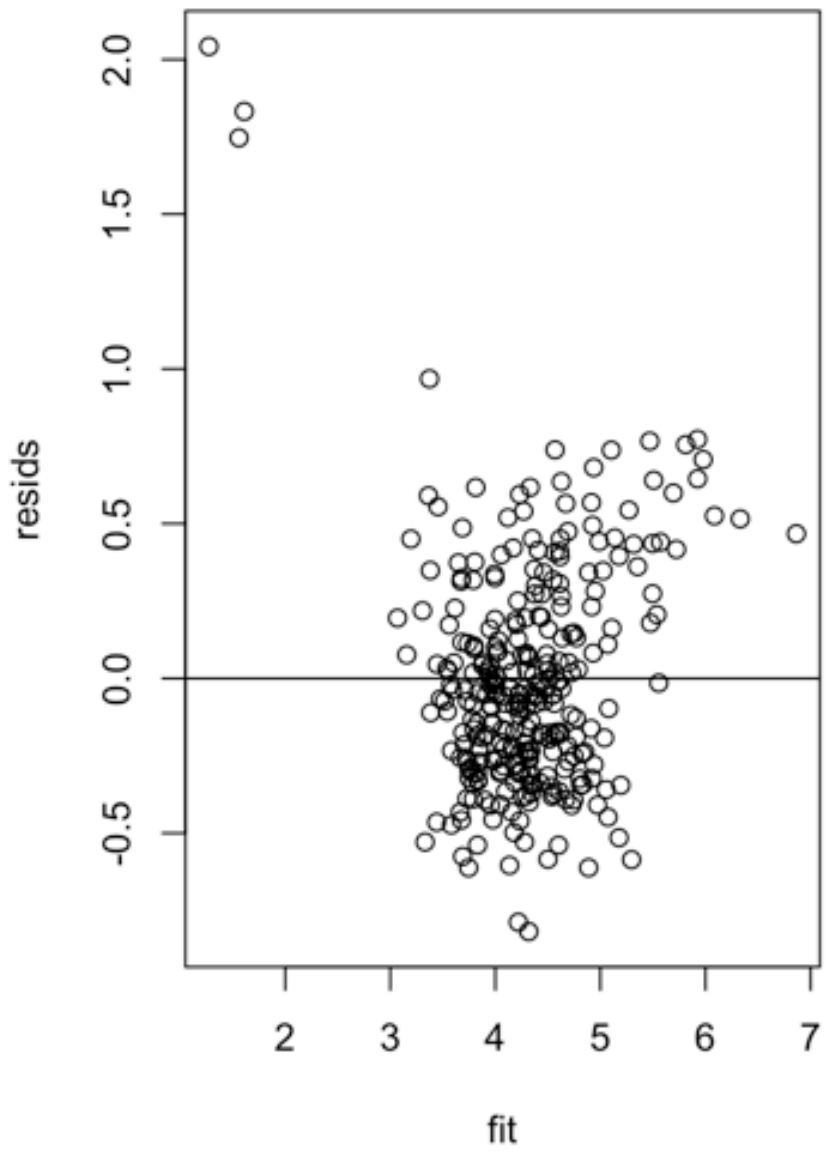


Fig 69 resid vs y_hat



We see collineary between Nitrogen and Carbon flux predictors, so they were combined into a sing flux categorie to predict mass flux

add comment of EDA ouput

How does depth affect flux?

Regression Model: $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$

Best Fit line: $f(x) = \hat{\beta}_0 + \hat{\beta}_1 x$

```
##
## Call:
## lm(formula = log(C_avg) ~ dep, data = Data.noNA)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6065 -0.4061 -0.0544  0.3198  2.9091
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.8268299  0.1447146  26.444  < 2e-16 ***
## dep         -0.0049867  0.0006469  -7.708 1.84e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7011 on 303 degrees of freedom
## Multiple R-squared:  0.164,  Adjusted R-squared:  0.1612
## F-statistic: 59.42 on 1 and 303 DF,  p-value: 1.839e-13
```

```
##
## Call:
## lm(formula = log(N_avg) ~ dep, data = Data.noNA)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5182 -0.3957 -0.0525  0.3702  2.1214
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.1679437  0.1428700  15.174   <2e-16 ***
## dep         -0.0057132  0.0006387  -8.945   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6922 on 303 degrees of freedom
## Multiple R-squared:  0.2089, Adjusted R-squared:  0.2063
## F-statistic: 80.02 on 1 and 303 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = log(P_avg) ~ dep, data = Data.noNA)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2015 -0.5223 -0.0115  0.5191  3.7380
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.2310162  0.1840231 -22.992  < 2e-16 ***
## dep         -0.0065736  0.0008226  -7.991 2.83e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8916 on 303 degrees of freedom
## Multiple R-squared:  0.1741, Adjusted R-squared:  0.1713
## F-statistic: 63.85 on 1 and 303 DF,  p-value: 2.83e-14
```

```
##
## Call:
## lm(formula = log(M_avg) ~ dep, data = Data.noNA)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2363 -0.4836 -0.1391  0.3049  2.9777
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.0517835  0.1466243  34.454  < 2e-16 ***
## dep         -0.0034891  0.0006555  -5.323 1.99e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7104 on 303 degrees of freedom
## Multiple R-squared:  0.08552,    Adjusted R-squared:  0.0825
## F-statistic: 28.34 on 1 and 303 DF,  p-value: 1.989e-07
```

Fig 70

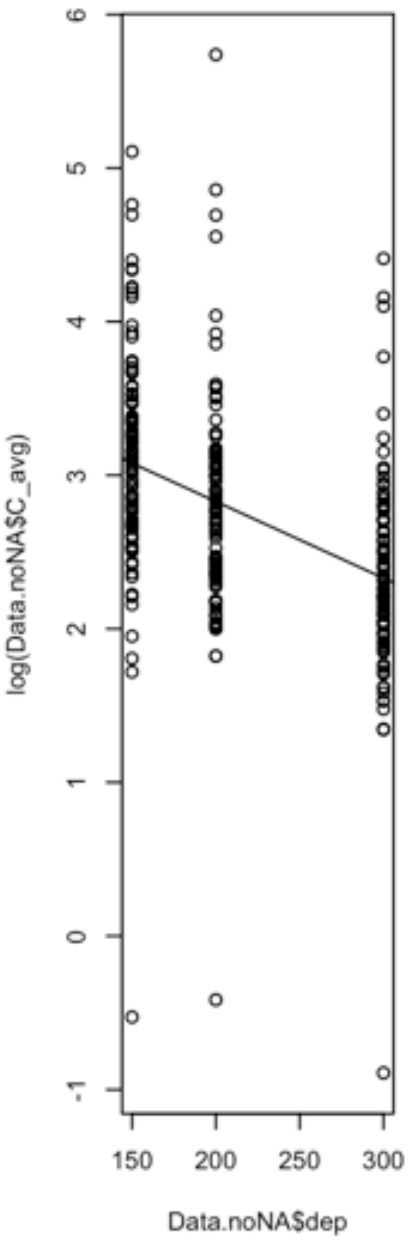


Fig 71

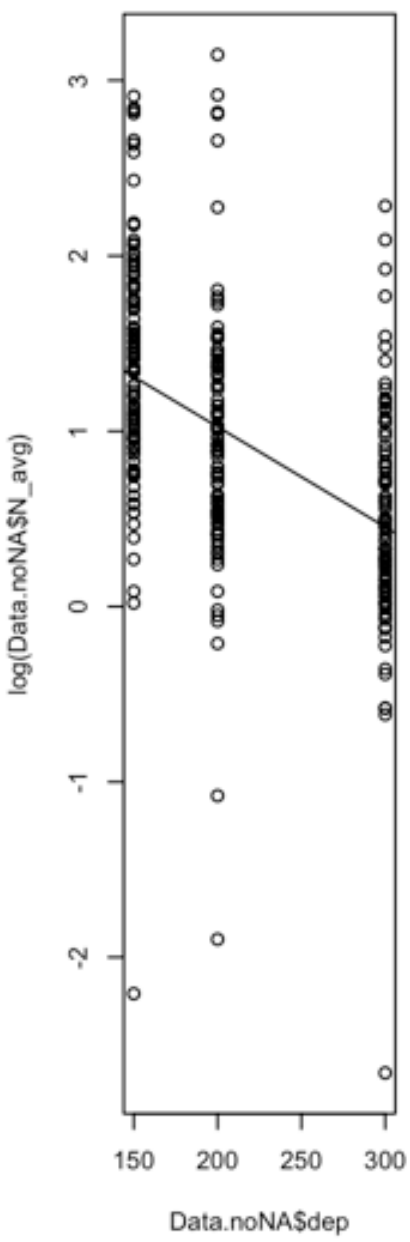


Fig 72

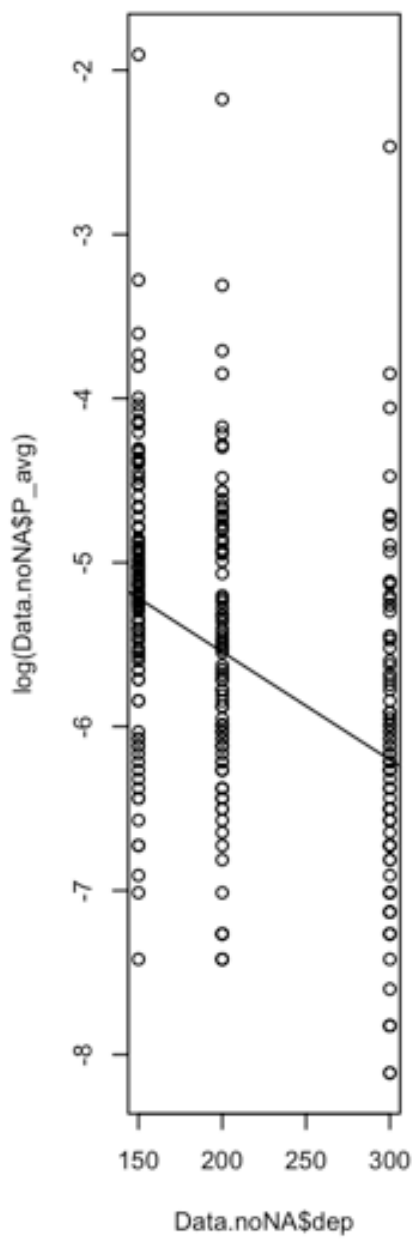
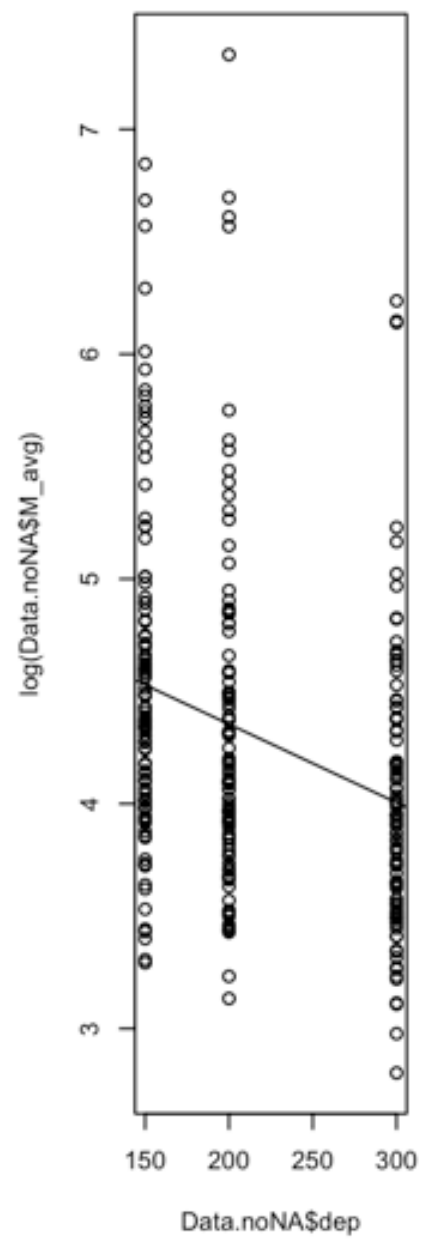


Fig 73



shows that as depth decreases so does flux, depth can be a predictor of flux

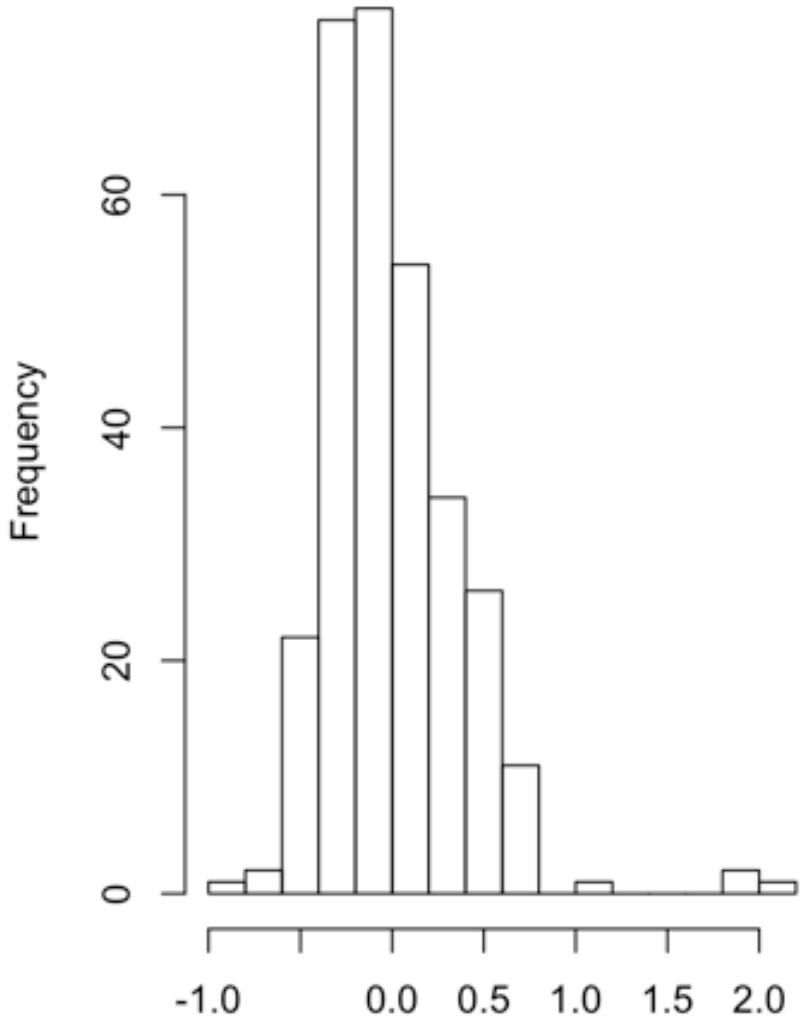
Multiple regression

Regression Model: $Y_i = \beta_0 + \beta_1 x_i + \ldots + \beta_{p-1} x_{i,p-1} + \epsilon_i = X_i' \beta + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$

Best Fit line: $f(x) = \hat{\beta}_0 + \hat{\beta}_1 x$

```
##
## Call:
## lm(formula = log(M_avg) ~ log(flux) + dep, data = Data.noNA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.80880 -0.25965 -0.04166  0.19793  2.06123
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.9800659  0.1342694  14.747   <2e-16 ***
## log(flux)   0.8496022  0.0303383  28.004   <2e-16 ***
## dep         0.0006387  0.0003763   1.698   0.0906 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3752 on 302 degrees of freedom
## Multiple R-squared:  0.7458, Adjusted R-squared:  0.7441
## F-statistic: 442.9 on 2 and 302 DF,  p-value: < 2.2e-16
```

Fig 74



Normal Q-Q Plot

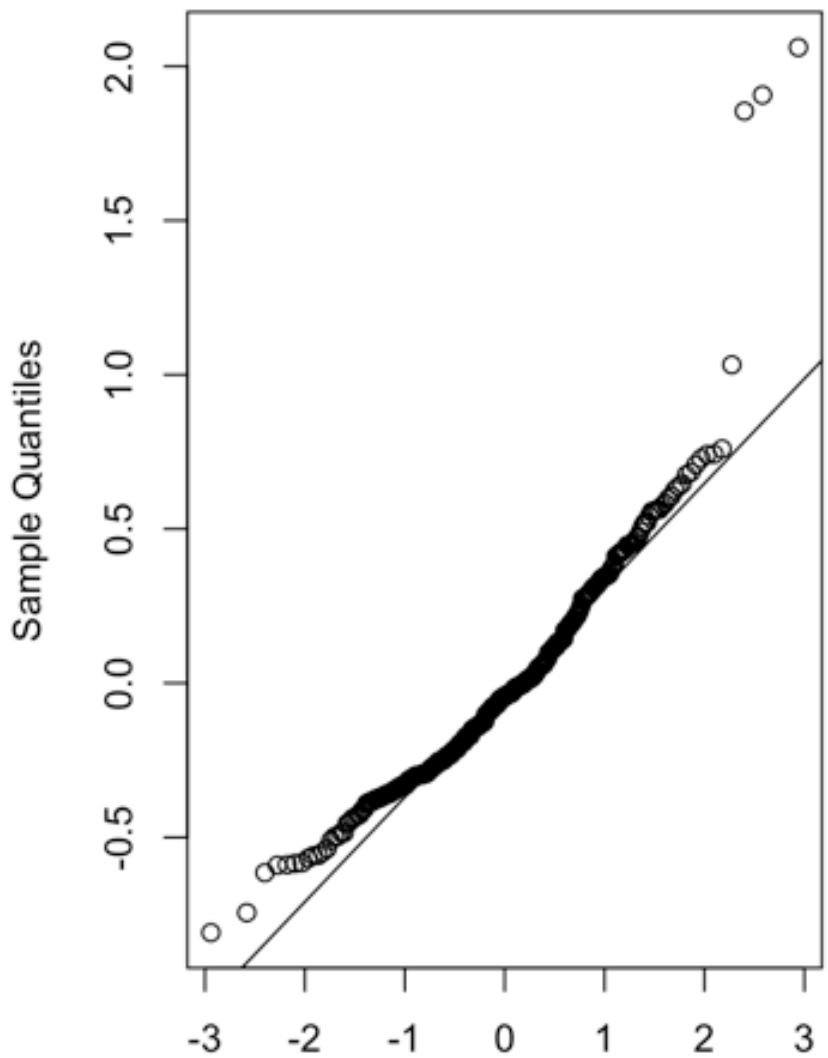


Fig 75 resid vs i

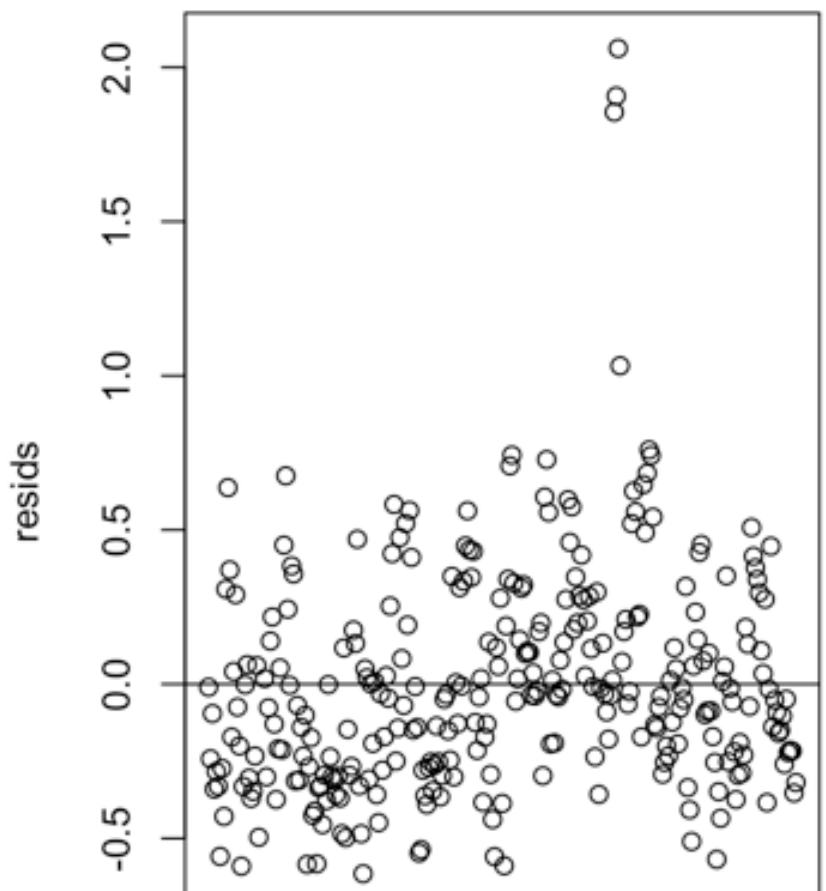
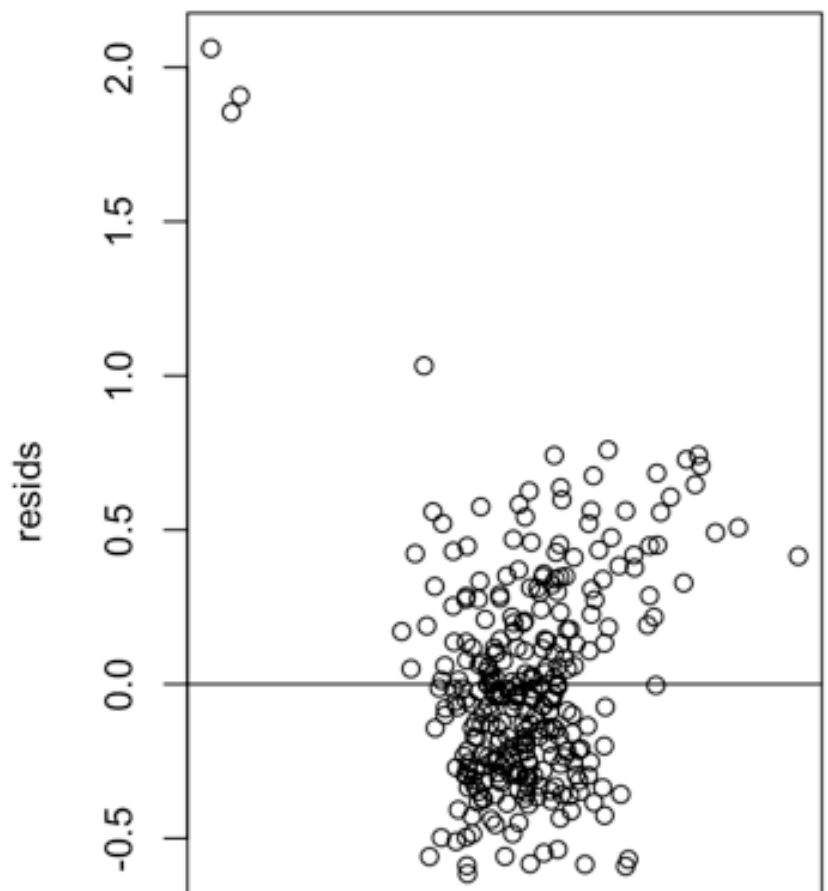
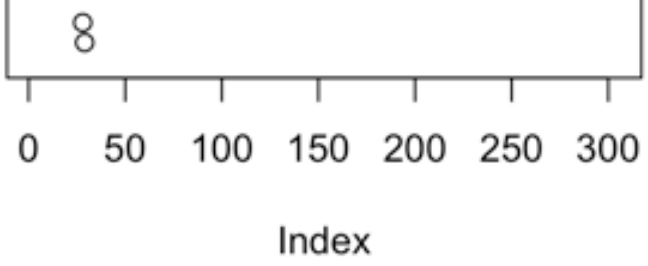
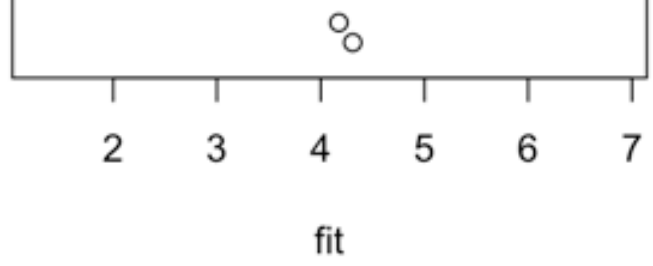


Fig 76 resid vs y_hat







Fairly normal and linear, but very clustered for residules

which one is better?

	Res.Df <dbl>	RSS <dbl>	Df <dbl>	Sum of Sq <dbl>	F <dbl>	Pr(>F) <dbl>
1	303	42.91741	*NA*	*NA*	*NA*	*NA*
2	302	42.51172	1	0.4056958	2.882032	0.09060306

2 rows

There is no significance difference in the first anova model, our p-value is to large to assume that lossing df due to an extra parameter is worthwhile

	Res.Df <dbl>	RSS <dbl>	Df <dbl>	Sum of Sq <dbl>	F <dbl>	Pr(>F) <dbl>
1	303	152.90688	*NA*	*NA*	*NA*	*NA*
2	302	42.51172	1	110.3952	784.2388	6.138939e-86

2 rows

However there our p-value is small in our second anova model indicates that losing df due to an extra paramter is worthwhile, or signiicantly improves the fit

ANCOVA

```
##       dep             M_avg              C_avg             N_avg
##  Min.   :150.0   Min.   :  16.49   Min.   :  0.41   Min.   : 0.070
##  1st Qu.:150.0   1st Qu.:  45.57   1st Qu.: 10.07   1st Qu.: 1.600
##  Median :200.0   Median :  64.90   Median : 15.09   Median : 2.510
##  Mean   :214.9   Mean   : 107.31   Mean   : 21.67   Mean   : 3.464
##  3rd Qu.:300.0   3rd Qu.: 101.87   3rd Qu.: 23.15   3rd Qu.: 4.100
##  Max.   :300.0   Max.   :1527.97   Max.   :310.63   Max.   :23.260
##      P_avg
##  Min.   :0.000300
##  1st Qu.:0.001900
##  Median :0.003500
##  Mean   :0.006229
##  3rd Qu.:0.006600
##  Max.   :0.148900
```

Fig 77

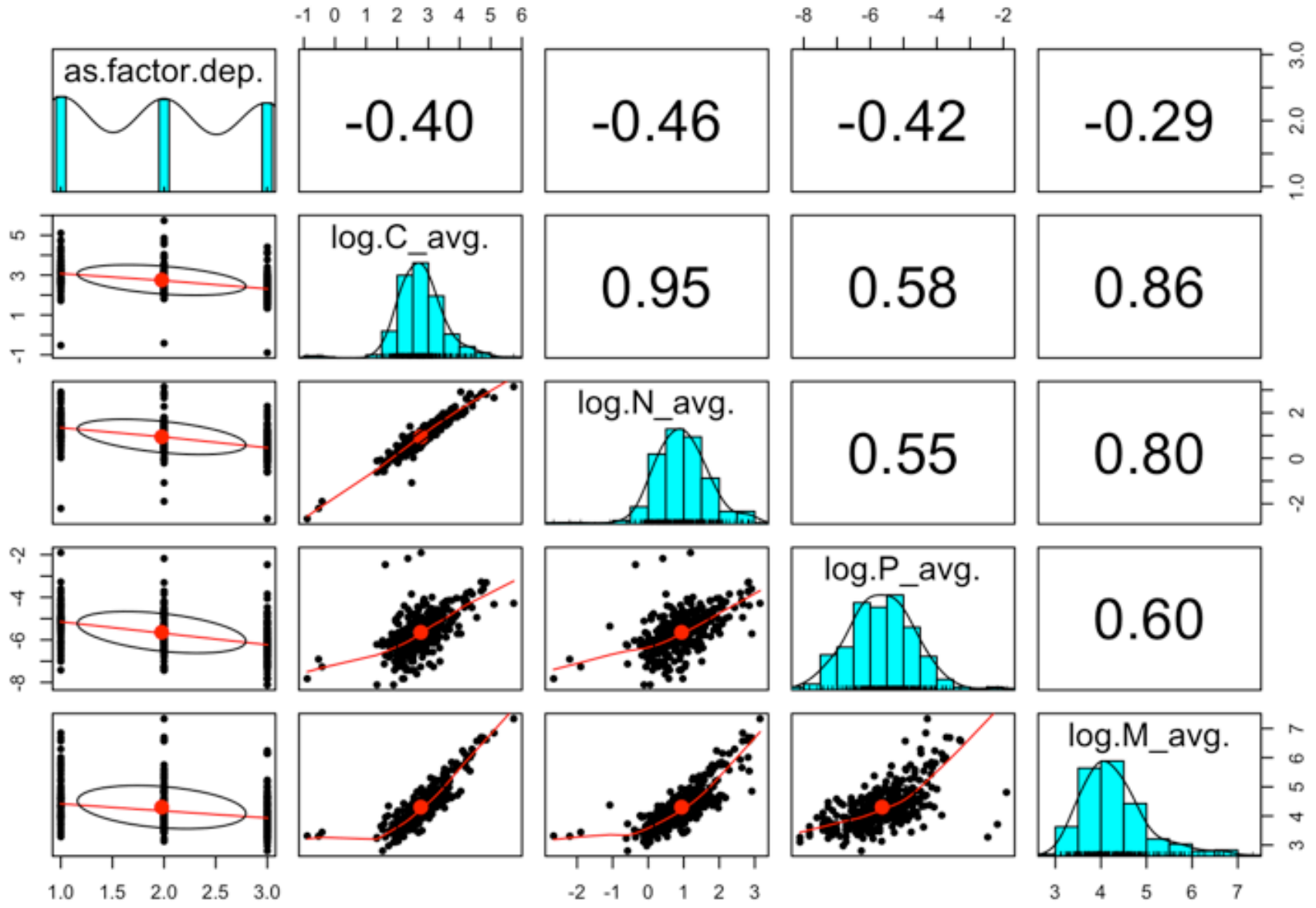
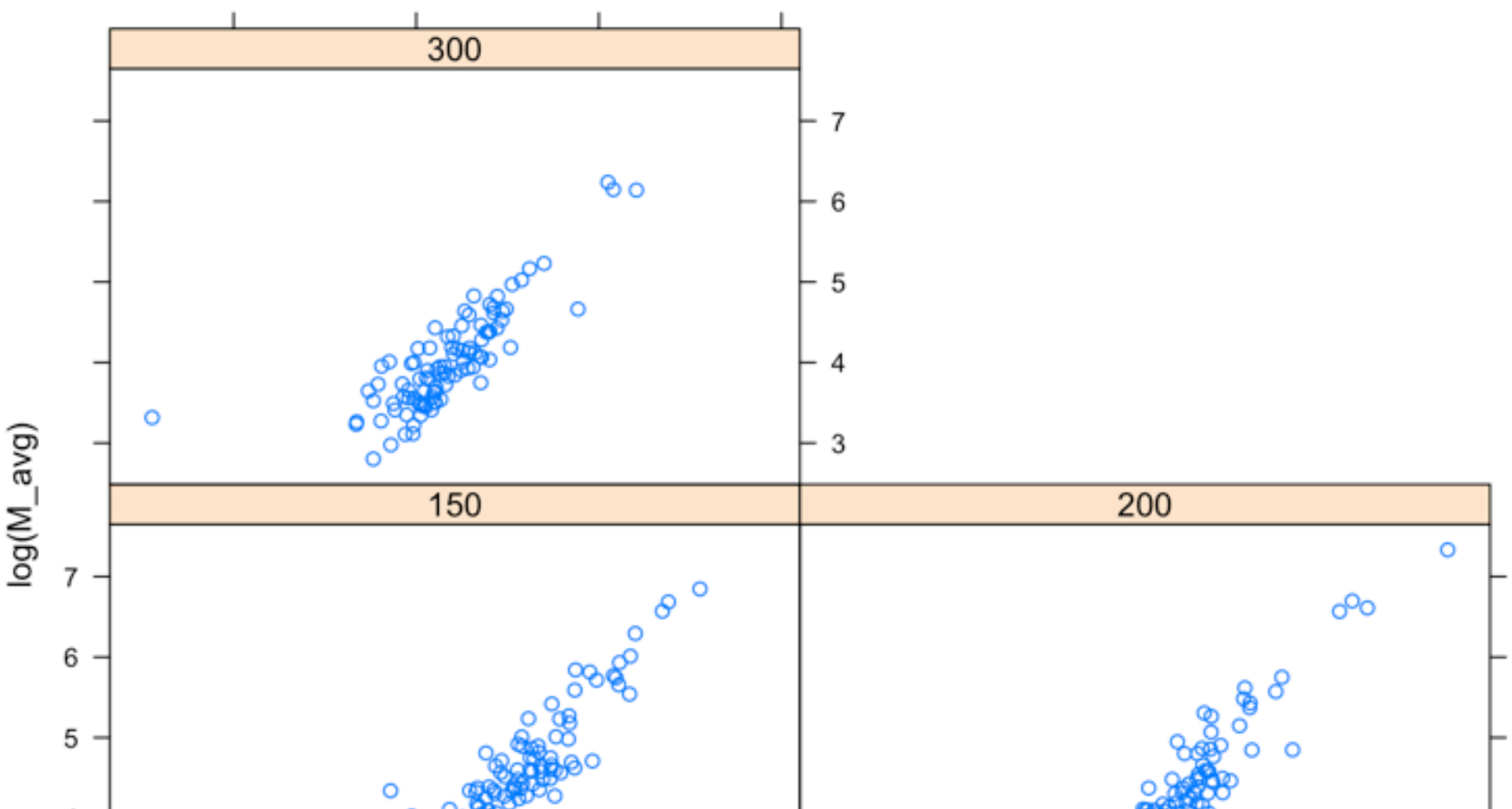


Fig 78: Activity level-specific scatterplots



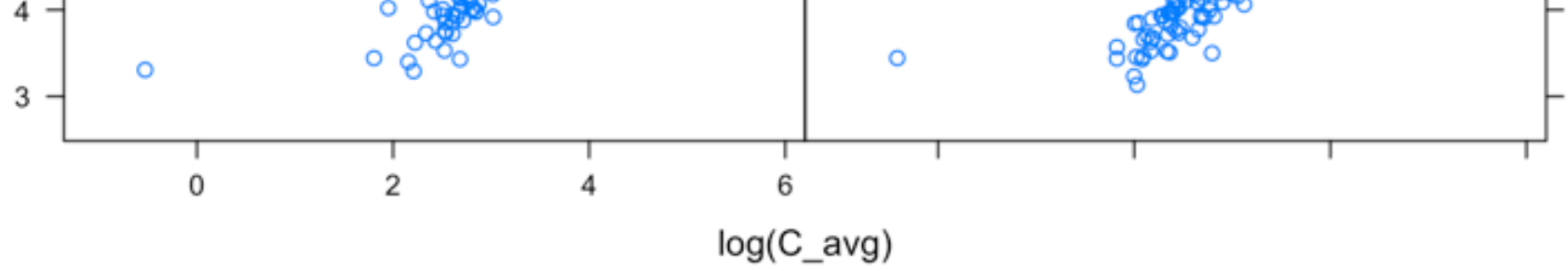


Fig 79: Scatterplot with color=group level

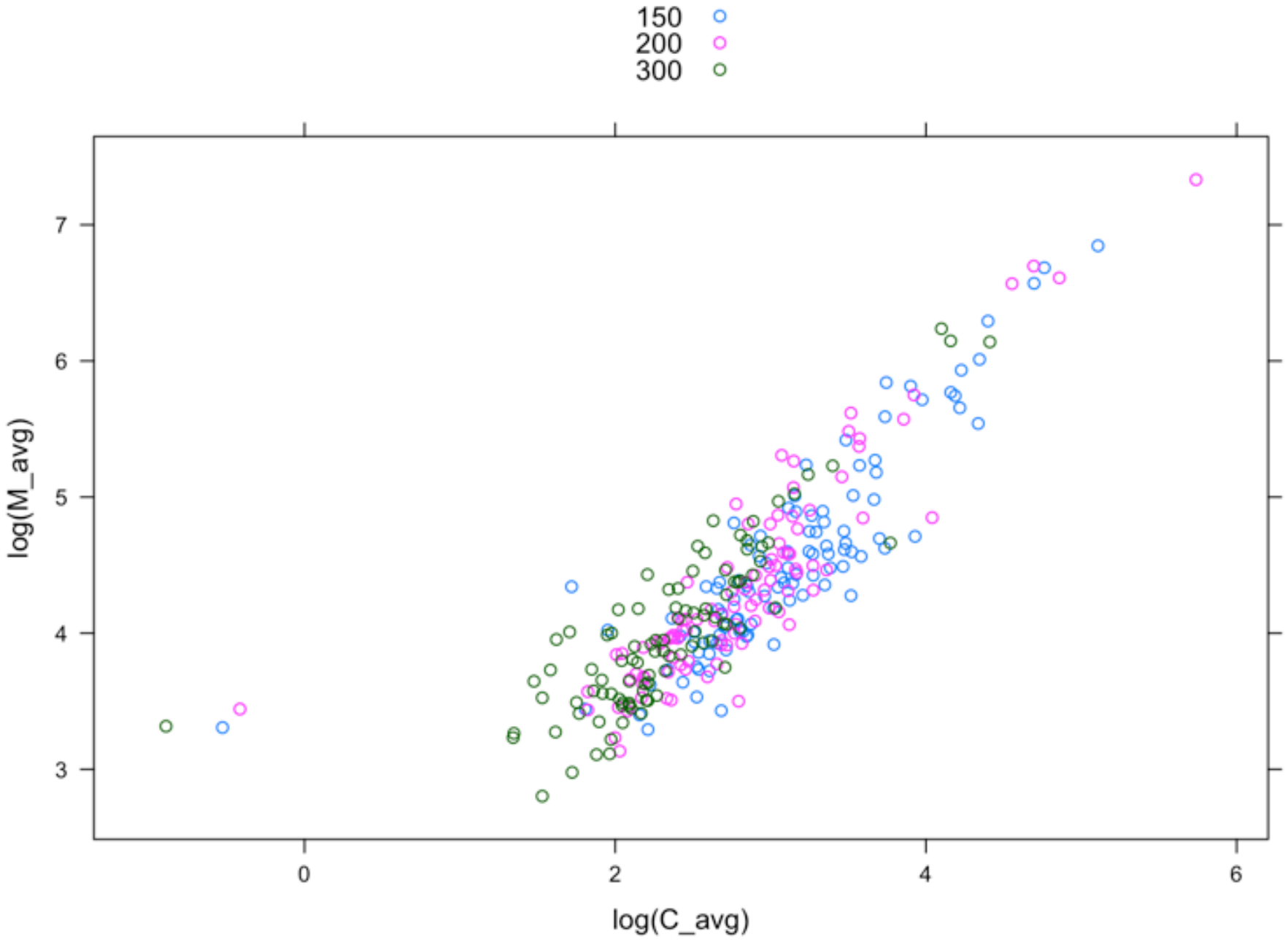
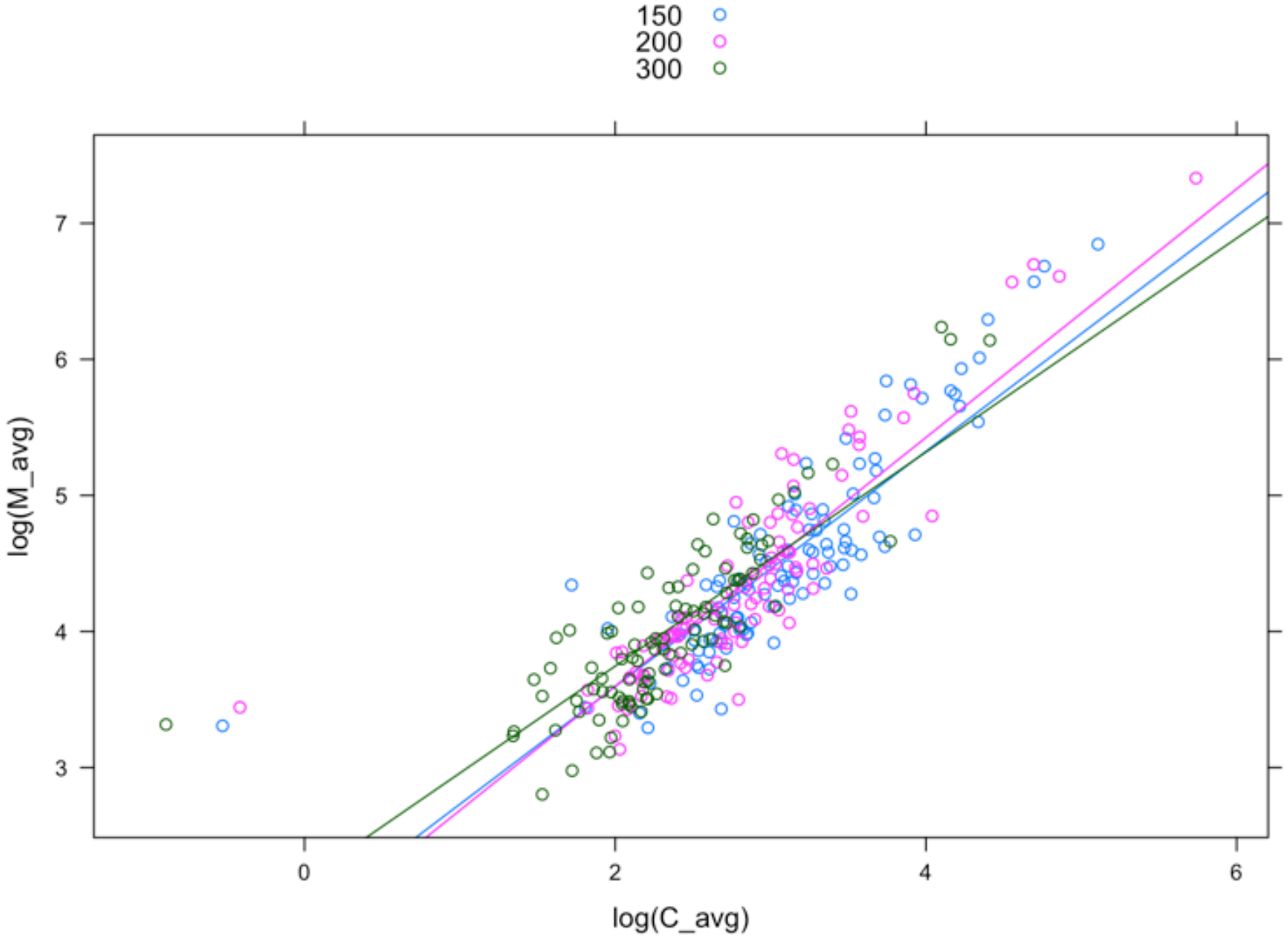


Fig 80: Three standalone depth level-specific `lm()` fits

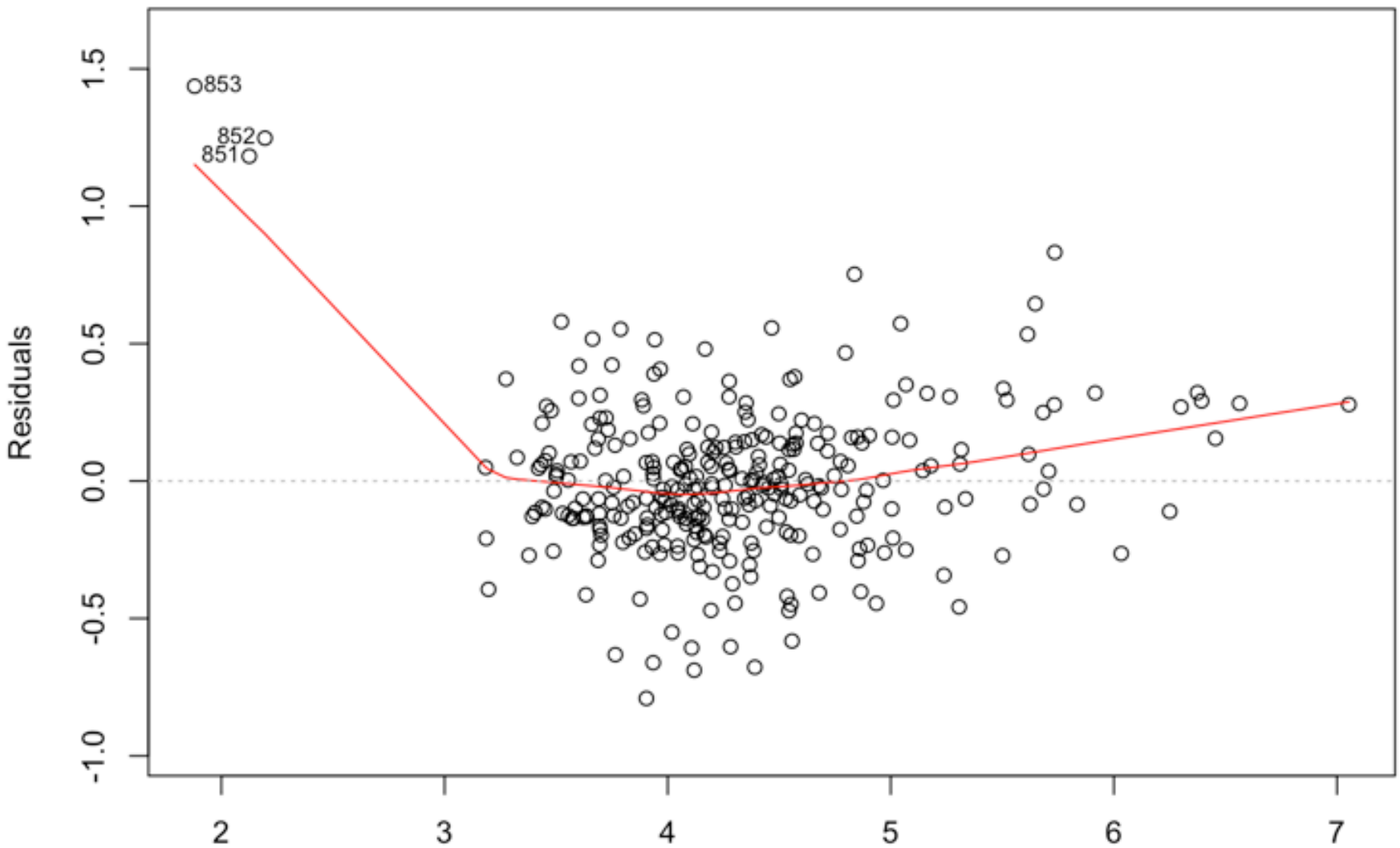


After log transforming data, C, N, P, and M distributions look much more noraml. Linearlity is looking better, especially for C and N. Activity level specific scatterplots highlight the interaction with C and M through depths and have somewhat of a linear reationship

```
##
## Call:
## lm(formula = log(M_avg) ~ log(C_avg) + log(N_avg) + log(P_avg) +
##     dep + as.factor(Year) + as.factor(month), data = ANOCOVA.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.79080 -0.13761 -0.01913  0.13800  1.43640
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.1133046  0.3096395  10.055  < 2e-16 ***
## log(C_avg)             0.7104828  0.0953704   7.450 1.18e-12 ***
## log(N_avg)            -0.0154473  0.0893634  -0.173 0.862888
## log(P_avg)             0.1342152  0.0280565   4.784 2.79e-06 ***
## dep                    0.0008642  0.0003428   2.521 0.012248 *
## as.factor(Year)2006    0.0308480  0.1219263   0.253 0.800451
## as.factor(Year)2007   -0.0660690  0.1188095  -0.556 0.578594
## as.factor(Year)2008    0.0231154  0.1239489   0.186 0.852195
## as.factor(Year)2009    0.0406626  0.1181832   0.344 0.731057
## as.factor(Year)2010    0.0780500  0.1244893   0.627 0.531197
## as.factor(Year)2011    0.2954345  0.1219114   2.423 0.016014 *
## as.factor(Year)2012    0.4328774  0.1200874   3.605 0.000370 ***
## as.factor(Year)2013    0.1907860  0.1180029   1.617 0.107053
## as.factor(Year)2014    0.1115483  0.1224068   0.911 0.362928
## as.factor(Year)2015    0.2994775  0.1207850   2.479 0.013750 *
## as.factor(month)Aug   -0.5090202  0.0904291  -5.629 4.41e-08 ***
## as.factor(month)Dec   -0.4301884  0.0830806  -5.178 4.30e-07 ***
## as.factor(month)Feb   -0.0916026  0.0795033  -1.152 0.250231
## as.factor(month)Jan   -0.4878071  0.1898550  -2.569 0.010708 *
## as.factor(month)Jul   -0.4706758  0.0759314  -6.199 2.04e-09 ***
## as.factor(month)Jun   -0.4504058  0.0921921  -4.886 1.74e-06 ***
## as.factor(month)Mar    0.0141537  0.0983848   0.144 0.885715
## as.factor(month)May   -0.4274599  0.0834863  -5.120 5.70e-07 ***
## as.factor(month)Nov   -0.2829892  0.0891804  -3.173 0.001676 **
## as.factor(month)Oct   -0.3367639  0.0923424  -3.647 0.000317 ***
## as.factor(month)Sep   -0.3769393  0.0804426  -4.686 4.36e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2917 on 279 degrees of freedom
## Multiple R-squared:  0.858,  Adjusted R-squared:  0.8453
## F-statistic: 67.46 on 25 and 279 DF,  p-value: < 2.2e-16
```

Fig 81

Residuals vs Fitted

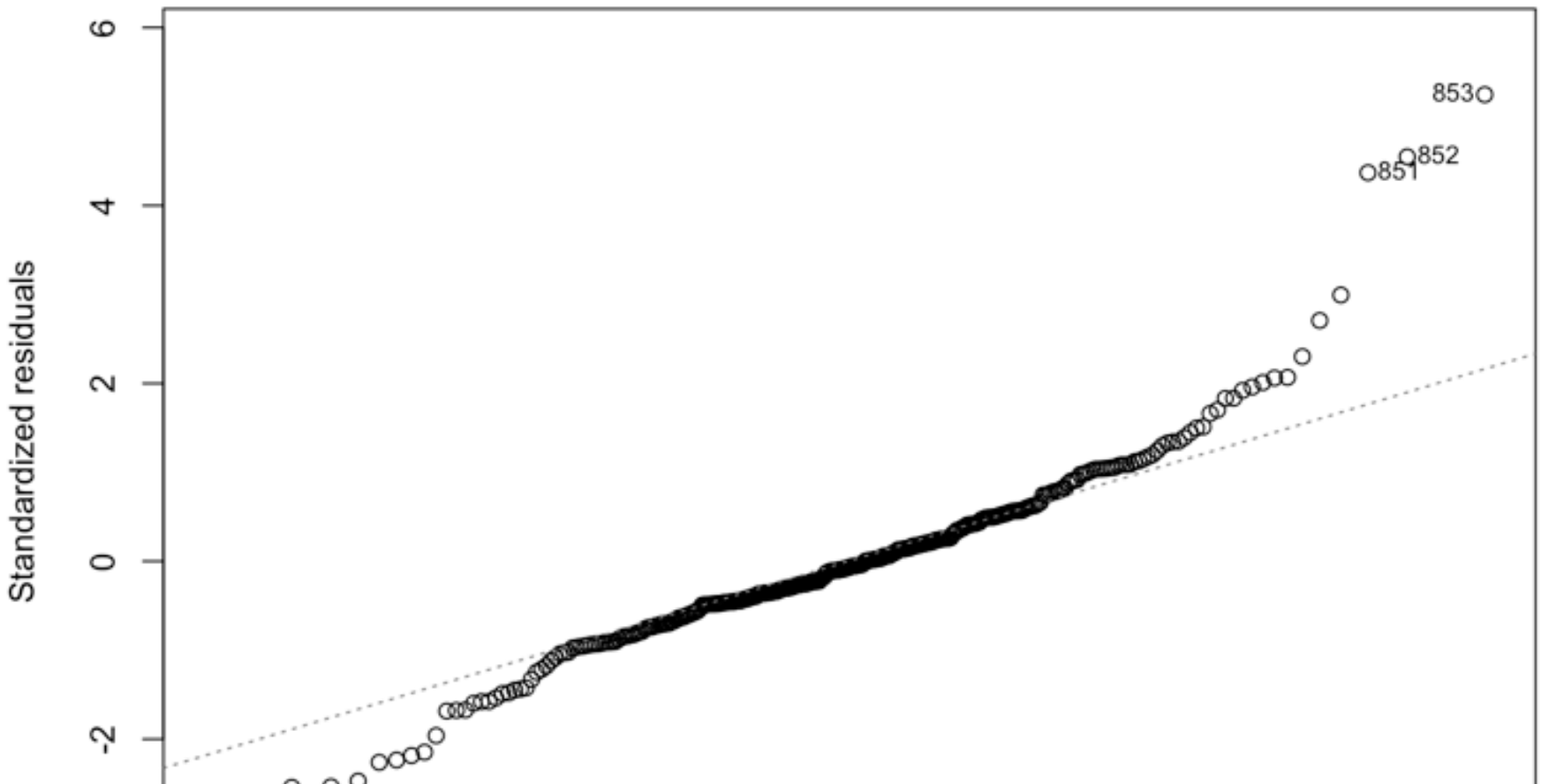


Fitted values

lm(log(M_avg) ~ log(C_avg) + log(N_avg) + log(P_avg) + dep + as.factor(Year ...

Fig 81

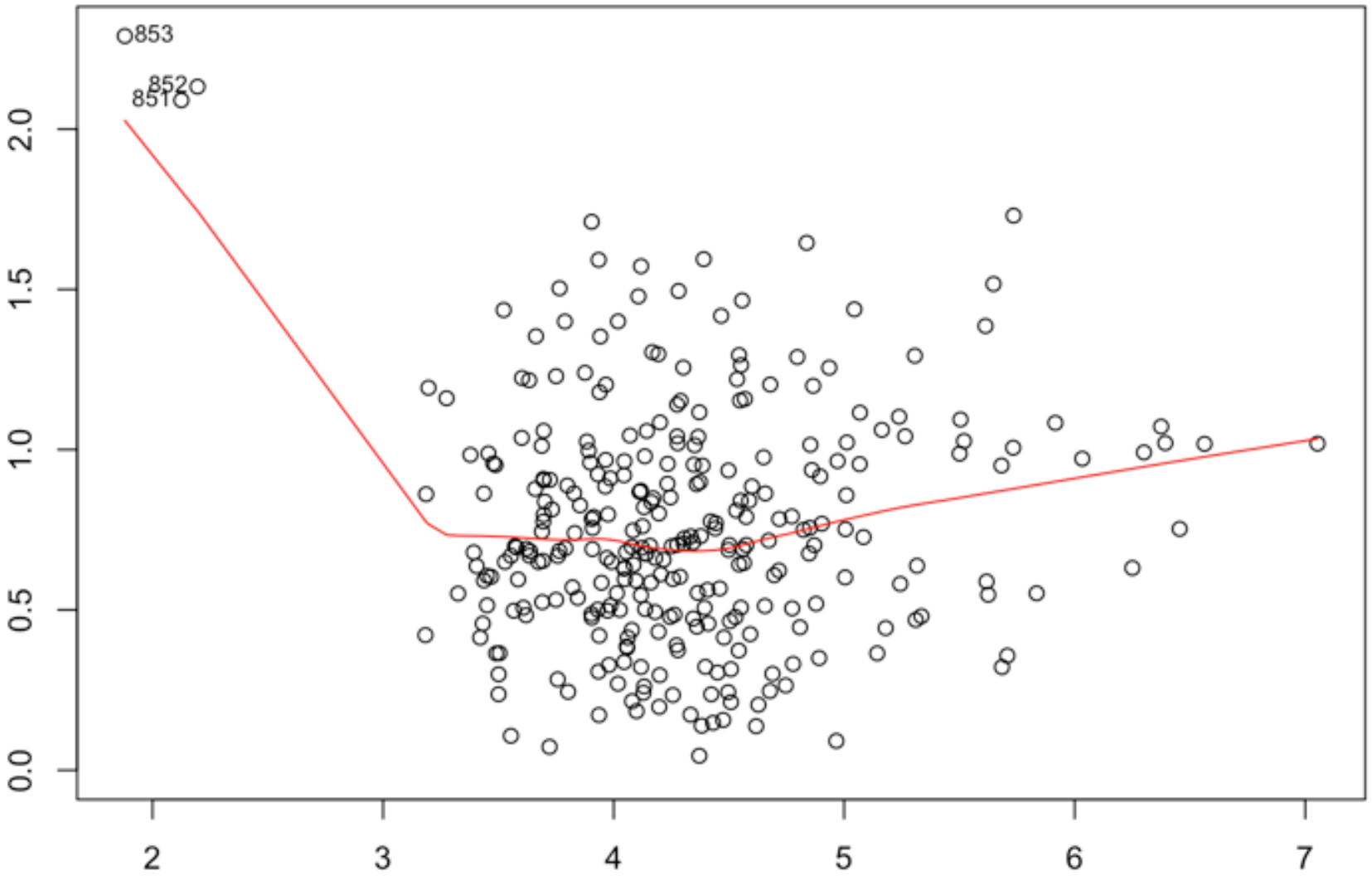
Normal Q-Q



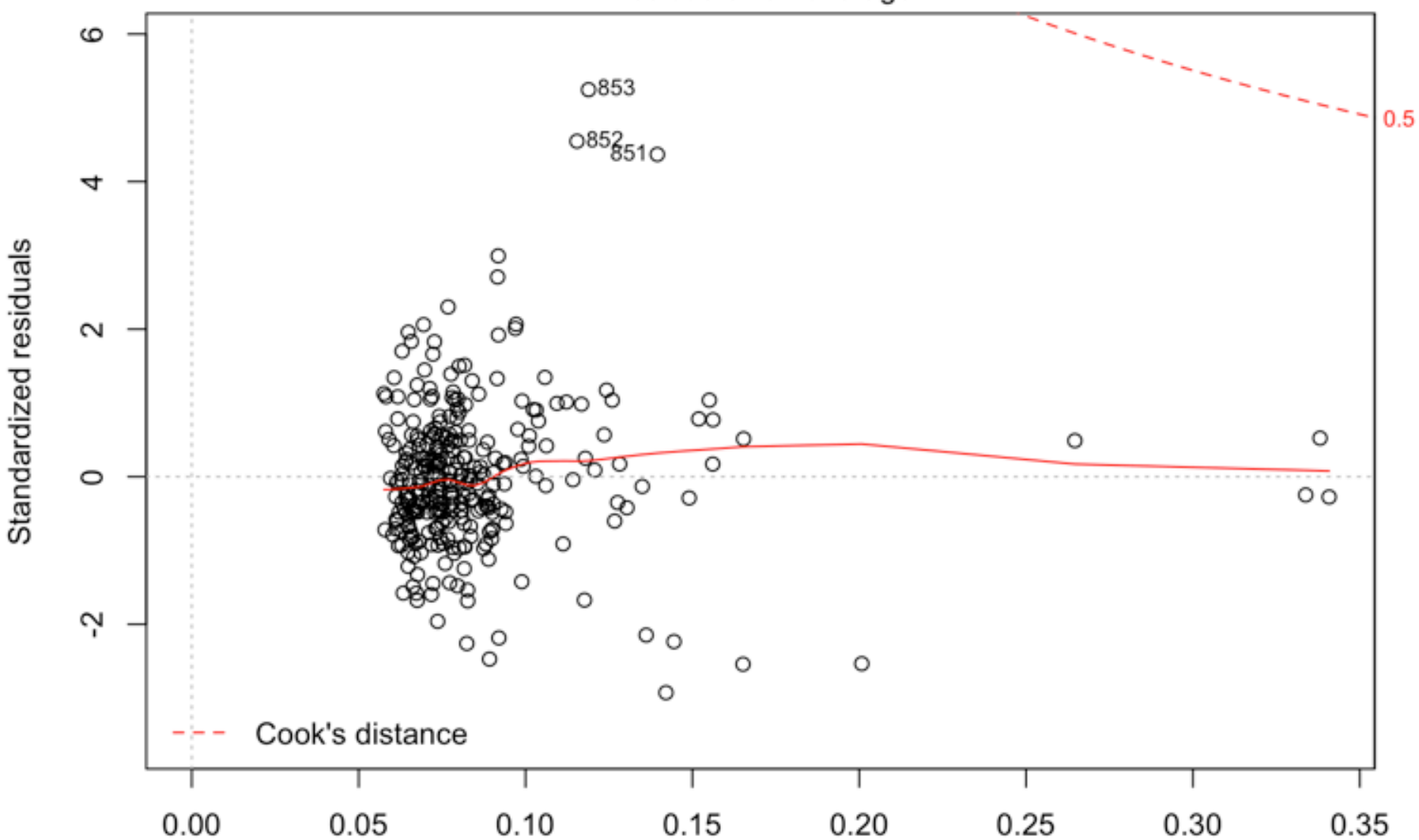
Theoretical Quantiles
lm(log(M_avg) ~ log(C_avg) + log(N_avg) + log(P_avg) + dep + as.factor(Year ...

Fig 81

Scale-Location



Fitted values
lm(log(M_avg) ~ log(C_avg) + log(N_avg) + log(P_avg) + dep + as.factor(Year ...



These do not present strong relationships

Results discussion:

Limitations:

References:

1. B. Szymczycha, Z. Klostowska, M. Lengier, L. Dzierzbicka-Glowacka, Significance of nutrient fluxes via submarine groundwater discharge in the Bay of Puck, southern Baltic Sea. Oceanologia 62, 117-125 (2020).
2. H. C. Zhao, L. Zhang, S. R. Wang, L. X. Jiao, Features and influencing factors of nitrogen and phosphorus diffusive fluxes at the sediment-water interface of Erhai Lake. Environmental Science and Pollution Research 25, 1933-1942 (2018).

3. K. Khan, C. W. Su, R. Tao, L. N. Hao, Urbanization and carbon emission: causality evidence from the new industrialized economies. Environment Development and Sustainability 22, 7193-7213 (2020).
4. C. A. Wynn-Edwards et al., Particle Fluxes at the Australian Southern Ocean Time Series (SOTS) Achieve Organic Carbon Sequestration at Rates Close to the Global Median, Are Dominated by Biogenic Carbonates, and Show No Temporal Trends Over 20-Years. Frontiers in Earth Science 8, 20 (2020).
5. BATS Methods. Chapter 20. Trap collected particle flux with surface tethered traps (2017)
6. BATS Methods. Chapter 1. Introduction (1997).

Appendix

```
rm(list=ls(all=TRUE))  #Housekeeping: clear out old files

options(warn = -1)
knitr::opts_chunk$set(echo = T, fig.height=6, fig.width=8, warning = F, message = F)
bats_flux <- read.csv("bats_flux.csv")

delete.na <- function(DF, n=0) {
  DF[rowSums(is.na(DF)) <= n,]
} #Function that takes rows out that contains NAs
bats_flux.noNA <- delete.na(bats_flux) # Take out rows containing NAs

# Too many observations taken out, lets subset a new data frame

Data <- subset(bats_flux,select = c("cr","dep","yymmdd1","yymmdd2","Lat2","Lat2.1","L
ong1","Long2","M_avg","C_avg","N_avg","P_avg"))
Data.noNA <- delete.na(Data)

# now we can use Data.noNA to compare each flux to one another, since no flux contain
s NAs

head(Data.noNA)
# separate each flux into its own dataframe
# delete rows that contain NAs
C_flux.data <- delete.na(subset(Data,select = c("cr","dep","yymmdd1","yymmdd2","Lat2"
,"Lat2.1","Long1","Long2","C_avg")))
N_flux.data <- delete.na(subset(Data,select = c("cr","dep","yymmdd1","yymmdd2","Lat2"
,"Lat2.1","Long1","Long2","N_avg")))
P_flux.data <- delete.na(subset(Data,select = c("cr","dep","yymmdd1","yymmdd2","Lat2"
,"Lat2.1","Long1","Long2","P_avg")))
M_flux.data <- delete.na(subset(Data,select = c("cr","dep","yymmdd1","yymmdd2","Lat2"
,"Lat2.1","Long1","Long2","M_avg")))
# Subset each depth and rename columns
library(tidyverse)
library(dplyr)
# use tideyverse instead
C_flux.depth.150 <- subset(C_flux.data, dep == 150)
```

```
C_flux.depth.150 <- rename(C_flux.depth.150,c("dep.150" = "dep"))
C_flux.depth.150 <- rename(C_flux.depth.150, c("C_avg.150" = "C_avg" ))
C_flux.depth.200 <- subset(C_flux.data, dep == 200)
C_flux.depth.200 <- rename(C_flux.depth.200, c("dep.200"="dep"))
C_flux.depth.200 <- rename(C_flux.depth.200, c("C_avg.200" ="C_avg"))
C_flux.depth.300 <- subset(C_flux.data, dep == 300)
C_flux.depth.300 <- rename(C_flux.depth.300, c("dep.300"="dep"))
C_flux.depth.300 <- rename(C_flux.depth.300, c("C_avg.300" ="C_avg"))
C_flux.depth.400 <- subset(C_flux.data, dep == 400)
C_flux.depth.400 <- rename(C_flux.depth.400, c("dep.400"="dep"))
C_flux.depth.400 <- rename(C_flux.depth.400, c("C_avg.400" ="C_avg"))
N_flux.depth.150 <- subset(N_flux.data, dep == 150)
N_flux.depth.150 <- rename(N_flux.depth.150, c("dep.150"="dep"))
N_flux.depth.150 <- rename(N_flux.depth.150, c("N_avg.150" ="N_avg"))
N_flux.depth.200 <- subset(N_flux.data, dep == 200)
N_flux.depth.200 <- rename(N_flux.depth.200, c("dep.200"="dep"))
N_flux.depth.200 <- rename(N_flux.depth.200, c("N_avg.200" ="N_avg"))
N_flux.depth.300 <- subset(N_flux.data, dep == 300)
N_flux.depth.300 <- rename(N_flux.depth.300, c("dep.300"="dep"))
N_flux.depth.300 <- rename(N_flux.depth.300, c("N_avg.300" ="N_avg"))
N_flux.depth.400 <- subset(N_flux.data, dep == 400)
N_flux.depth.400 <- rename(N_flux.depth.400, c("dep.400"="dep"))
N_flux.depth.400 <- rename(N_flux.depth.400, c("N_avg.400" ="N_avg"))
P_flux.depth.150 <- subset(P_flux.data, dep == 150)
P_flux.depth.150 <- rename(P_flux.depth.150, c("dep.150"="dep"))
P_flux.depth.150 <- rename(P_flux.depth.150, c("P_avg.150" ="P_avg"))
P_flux.depth.200 <- subset(P_flux.data, dep == 200)
P_flux.depth.200 <- rename(P_flux.depth.200, c("dep.200"="dep"))
P_flux.depth.200 <- rename(P_flux.depth.200, c("P_avg.200" ="P_avg"))
P_flux.depth.300 <- subset(P_flux.data, dep == 300)
P_flux.depth.300 <- rename(P_flux.depth.300, c("dep.300"="dep"))
P_flux.depth.300 <- rename(P_flux.depth.300, c("P_avg.300" ="P_avg"))
P_flux.depth.400 <- subset(P_flux.data, dep == 400)
P_flux.depth.400 <- rename(P_flux.depth.400, c("dep.400"="dep"))
P_flux.depth.400 <- rename(P_flux.depth.400, c("P_avg.400" ="P_avg"))
M_flux.depth.150 <- subset(M_flux.data, dep == 150)
M_flux.depth.150 <- rename(M_flux.depth.150, c("dep.150"="dep"))
M_flux.depth.150 <- rename(M_flux.depth.150, c("M_avg.150" ="M_avg"))
M_flux.depth.200 <- subset(M_flux.data, dep == 200)
M_flux.depth.200 <- rename(M_flux.depth.200, c("dep.200"="dep"))
M_flux.depth.200 <- rename(M_flux.depth.200, c("M_avg.200" ="M_avg"))
M_flux.depth.300 <- subset(M_flux.data, dep == 300)
M_flux.depth.300 <- rename(M_flux.depth.300, c("dep.300"="dep"))
M_flux.depth.300 <- rename(M_flux.depth.300, c("M_avg.300" ="M_avg"))
M_flux.depth.400 <- subset(M_flux.data, dep == 400)
M_flux.depth.400 <- rename(M_flux.depth.400, c("dep.400"="dep"))
M_flux.depth.400 <- rename(M_flux.depth.400, c("M_avg.400" ="M_avg"))
#n values for C-flux
length(C_flux.depth.150$C_avg.150)
length(C_flux.depth.200$C_avg.200)
```

```
length(C_flux.depth.300$C_avg.300)
length(C_flux.depth.400$C_avg.400)
muCflux150 <- mean(C_flux.depth.150$C_avg.150) #focusing on mean of C-avg at depth 15
0
s.e150c <- sd(C_flux.depth.150$C_avg.150) / sqrt(314)  #need to find s.e

muCflux200 <- mean(C_flux.depth.200$C_avg.200) #focusing on mean of C-avg at depth 20
0
s.e200c <- sd(C_flux.depth.200$C_avg.200) / sqrt(304)

muCflux300 <- mean(C_flux.depth.300$C_avg.300) #focusing on mean of C-avg at depth 30
0
s.e300c <- sd(C_flux.depth.300$C_avg.300) / sqrt(305)

muCflux400 <- mean(C_flux.depth.400$C_avg.400) #focusing on mean of C-avg at depth 40
0
s.e400c <- sd(C_flux.depth.400$C_avg.400) / sqrt(18)

bar.Cflux <- data.frame(c("dep150","dep200","dep300","dep400"), c(muCflux150,muCflux2
00,muCflux300,muCflux400), c(s.e150c,s.e200c,s.e300c,s.e400c))
colnames(bar.Cflux)[1] <- "Depth"
colnames(bar.Cflux)[2] <- "C_flux.Mean"
colnames(bar.Cflux)[3] <- "C_flux.SE"
bar.Cflux
library(ggplot2)
library(dplyr)

xval <- bar.Cflux$Depth #character objects need quotes
yval <- bar.Cflux$`C_flux.Mean`

ggplot()+geom_col(bar.Cflux, mapping=aes(x= Depth, y = C_flux.Mean), color="#e9ecef",
alpha=0.6,) +
xlab("Depth (m)") + ylab("Mean of C Flux(mg carbon m^-2 day^-1)") + ggtitle("Fig 1")
#n values for P-flux
length(P_flux.depth.150$P_avg.150)
length(P_flux.depth.200$P_avg.200)
length(P_flux.depth.300$P_avg.300)
length(P_flux.depth.400$P_avg.400)
muPflux150 <- mean(P_flux.depth.150$P_avg.150) #focusing on mean of P-avg at depth 15
0
s.e150p <- sd(P_flux.depth.150$P_avg.150) / sqrt(109)  #need to find s.e

muPflux200 <- mean(P_flux.depth.200$P_avg.200) #focusing on mean of P-avg at depth 20
0
s.e200p <- sd(P_flux.depth.200$P_avg.200) / sqrt(106)

muPflux300 <- mean(P_flux.depth.300$P_avg.300) #focusing on mean of P-avg at depth 30
0
s.e300p <- sd(P_flux.depth.300$P_avg.300) / sqrt(105)
```

```
bar.Pflux <- data.frame(c("dep150","dep200","dep300"), c(muPflux150,muPflux200,muPflu
x300), c(s.e150p,s.e200p,s.e300p))
colnames(bar.Pflux)[1] <- "Depth"
colnames(bar.Pflux)[2] <- "P_flux.Mean"
colnames(bar.Pflux)[3] <- "P_flux.SE"
bar.Pflux
library(ggplot2)
library(dplyr)

xval <- bar.Pflux$Depth #character objects need quotes
yval <- bar.Pflux$`P_flux.Mean`

ggplot()+geom_col(bar.Pflux, mapping=aes(x= Depth, y = P_flux.Mean), color="#e9ecef",
alpha=0.6,) +
xlab("Depth (m)") + ylab("Mean of P Flux(mg phospohorous m^-2 day^-1)") + ggtitle("Fi
g 2")
#n values for N-flux
length(N_flux.depth.150$N_avg.150)
length(N_flux.depth.200$N_avg.200)
length(N_flux.depth.300$N_avg.300)
length(N_flux.depth.400$N_avg.400)
muNflux150 <- mean(N_flux.depth.150$N_avg.150) #focusing on mean of P-avg at depth 15
0
s.e150n <- sd(N_flux.depth.150$N_avg.150) / sqrt(312)  #need to find s.e

muNflux200 <- mean(N_flux.depth.200$N_avg.200) #focusing on mean of P-avg at depth 20
0
s.e200n <- sd(N_flux.depth.200$N_avg.200) / sqrt(302)

muNflux300 <- mean(N_flux.depth.300$N_avg.300) #focusing on mean of P-avg at depth 30
0
s.e300n <- sd(N_flux.depth.300$N_avg.300) / sqrt(303)

muNflux400 <- mean(N_flux.depth.400$N_avg.400) #focusing on mean of C-avg at depth 40
0
s.e400n <- sd(N_flux.depth.400$N_avg.400) / sqrt(18)

bar.Nflux <- data.frame(c("dep150","dep200","dep300","dep400"), c(muNflux150,muNflux2
00,muNflux300,muNflux400), c(s.e150n,s.e200n,s.e300n,s.e400n))
colnames(bar.Nflux)[1] <- "Depth"
colnames(bar.Nflux)[2] <- "N_flux.Mean"
colnames(bar.Nflux)[3] <- "N_flux.SE"
bar.Nflux
library(ggplot2)
library(dplyr)

xval <- bar.Nflux$Depth #character objects need quotes
yval <- bar.Nflux$`N_flux.Mean`

ggplot()+geom_col(bar.Nflux, mapping=aes(x= Depth, y = N_flux.Mean), color="#e9ecef",
```

```
alpha=0.6,) +
xlab("Depth (m)") + ylab("Mean of N Flux(mg nitrogen m^-2 day^-1)") + ggtitle("Fig 3"
)
#n values for M-flux
length(M_flux.depth.150$M_avg.150)
length(M_flux.depth.200$M_avg.200)
length(M_flux.depth.300$M_avg.300)
length(M_flux.depth.400$M_avg.400)
muMflux150 <- mean(M_flux.depth.150$M_avg.150) #focusing on mean of P-avg at depth 15
0
s.e150m <- sd(M_flux.depth.150$M_avg.150) / sqrt(314)  #need to find s.e

muMflux200 <- mean(M_flux.depth.200$M_avg.200) #focusing on mean of P-avg at depth 20
0
s.e200m <- sd(M_flux.depth.200$M_avg.200) / sqrt(305)

muMflux300 <- mean(M_flux.depth.300$M_avg.300) #focusing on mean of P-avg at depth 30
0
s.e300m <- sd(M_flux.depth.300$M_avg.300) / sqrt(305)

muMflux400 <- mean(M_flux.depth.400$M_avg.400) #focusing on mean of C-avg at depth 40
0
s.e400m <- sd(M_flux.depth.400$M_avg.400) / sqrt(17)

bar.Mflux <- data.frame(c("dep150","dep200","dep300","dep400"), c(muMflux150,muMflux2
00,muMflux300,muMflux400), c(s.e150m,s.e200m,s.e300m,s.e400m))
colnames(bar.Mflux)[1] <- "Depth"
colnames(bar.Mflux)[2] <- "M_flux.Mean"
colnames(bar.Mflux)[3] <- "M_flux.SE"
bar.Mflux
library(ggplot2)
library(dplyr)


xval <- bar.Mflux$Depth #character objects need quotes
yval <- bar.Mflux$`M_flux.Mean`

ggplot()+geom_col(bar.Mflux, mapping=aes(x= Depth, y = M_flux.Mean), color="#e9ecef",
alpha=0.6,) +
xlab("Depth (m)") + ylab("Mean of M Flux(mg mass m^-2 day^-1)") + ggtitle("Fig 4")
# Checking for normaility and no outliers in Carbon Flux
par(mfrow=c(1,4))
hist((C_flux.depth.150$C_avg.150), main ="Fig 5 C Flux depth 150")
hist((C_flux.depth.200$C_avg.200), main ="Fig 6 C Flux depth 200")
hist((C_flux.depth.300$C_avg.300), main ="Fig 7 C Flux depth 300")
hist((C_flux.depth.400$C_avg.400), main ="Fig 8 C Flux depth 400")

# using natural log to Transform the data
par(mfrow=c(1,4))
hist(log(C_flux.depth.150$C_avg.150), main ="Fig 9 C Flux depth 150 (log)")
```

```
hist(log(C_flux.depth.200$C_avg.200), main ="Fig 10 C Flux depth 200 (log)")
hist(log(C_flux.depth.300$C_avg.300), main ="Fig 11 C Flux depth 300 (log)")
hist(log(C_flux.depth.400$C_avg.400), main ="Fig 12 C Flux depth 400 (log)")

# Checking for outliers
par(mfrow=c(1,4))
boxplot(log(C_flux.depth.150$C_avg.150), main ="Fig 13 C Flux depth 150 (log)")
boxplot(log(C_flux.depth.200$C_avg.200), main ="Fig 14 C Flux depth 200 (log)")
boxplot(log(C_flux.depth.300$C_avg.300), main ="Fig 15 C Flux depth 300 (log)")
boxplot(log(C_flux.depth.400$C_avg.400), main ="Fig 16 C Flux depth 400 (log)")
# Checking for normaility and no outliers in Nitrogen Flux
par(mfrow=c(1,4))
hist((N_flux.depth.150$N_avg.150), breaks = "FD", main ="Fig 17 N Flux depth 150")
hist((N_flux.depth.200$N_avg.200), breaks = "FD", main ="Fig 18 N Flux depth 200")
hist((N_flux.depth.300$N_avg.300), breaks = "FD", main ="Fig 19 N Flux depth 300")
hist((N_flux.depth.400$N_avg.400), breaks = "FD", main ="Fig 20 N Flux depth 400")

# using natural log to Transform the data
par(mfrow=c(1,4))
hist(log(N_flux.depth.150$N_avg.150), main ="Fig 21 N Flux depth 150 (log)")
hist(log(N_flux.depth.200$N_avg.200), main ="Fig 22 N Flux depth 200 (log)")
hist(log(N_flux.depth.300$N_avg.300), main ="Fig 23 N Flux depth 300 (log)")
hist(log(N_flux.depth.400$N_avg.400), main ="Fig 24 N Flux depth 400 (log)")

# Checking for outliers
par(mfrow=c(1,4))
boxplot(log(N_flux.depth.150$N_avg.150), main ="Fig 25 N Flux depth 150 (log)")
boxplot(log(N_flux.depth.200$N_avg.200), main ="Fig 26 N Flux depth 200 (log)")
boxplot(log(N_flux.depth.300$N_avg.300), main ="Fig 27 N Flux depth 300 (log)")
boxplot(log(N_flux.depth.400$N_avg.400), main ="Fig 28 N Flux depth 400 (log)")
# Checking for normaility and no outliers in Phosporous Flux
par(mfrow=c(1,3))
hist((P_flux.depth.150$P_avg.150), breaks = "FD", main ="Fig 29 P Flux depth 150")
hist((P_flux.depth.200$P_avg.200), breaks = "FD", main ="Fig 30 P Flux depth 200")
hist((P_flux.depth.300$P_avg.300), breaks = "FD", main ="Fig 31 P Flux depth 300")

# using natural log to Transform the data
par(mfrow=c(1,3))
hist(log(P_flux.depth.150$P_avg.150), main ="Fig 32 P Flux depth 150 (log)")
hist(log(P_flux.depth.200$P_avg.200), main ="Fig 33 P Flux depth 200 (log)")
hist(log(P_flux.depth.300$P_avg.300), main ="Fig 34 P Flux depth 300 (log)")

# Checking for outliers
par(mfrow=c(1,3))
boxplot(log(P_flux.depth.150$P_avg.150), main ="Fig 35 P Flux depth 150 (log)")
boxplot(log(P_flux.depth.200$P_avg.200), main ="Fig 36 P Flux depth 200 (log)")
boxplot(log(P_flux.depth.300$P_avg.300), main ="Fig 37 P Flux depth 300 (log)")
# Checking for normaility and no outliers in Mass Flux
par(mfrow=c(1,4))
hist((M_flux.depth.150$M_avg.150), breaks = "FD", main ="Fig 38 M Flux depth 150")
```

```
hist((M_flux.depth.200$M_avg.200), breaks = "FD", main ="Fig 39 M Flux depth 200")
hist((M_flux.depth.300$M_avg.300), breaks = "FD", main ="Fig 40 M Flux depth 300")
hist((M_flux.depth.400$M_avg.400), breaks = "FD", main ="Fig 41 M Flux depth 400")

# using natural log to Transform the data
par(mfrow=c(1,4))
hist(log(M_flux.depth.150$M_avg.150), main ="Fig 42 M Flux depth 150 (log)")
hist(log(M_flux.depth.200$M_avg.200), main ="Fig 43 M Flux depth 200 (log)")
hist(log(M_flux.depth.300$M_avg.300), main ="Fig 44 M Flux depth 300 (log)")
hist(log(M_flux.depth.400$M_avg.400), main ="Fig 45 M Flux depth 400 (log)")

# Checking for outliers
par(mfrow=c(1,4))
boxplot(log(M_flux.depth.150$M_avg.150), main ="Fig 46 M Flux depth 150 (log)")
boxplot(log(M_flux.depth.200$M_avg.200), main ="Fig 47 M Flux depth 200 (log)")
boxplot(log(M_flux.depth.300$M_avg.300), main ="Fig 48 M Flux depth 300 (log)")
boxplot(log(M_flux.depth.400$M_avg.400), main ="Fig 49 M Flux depth 400 (log)")
#create D-bar for each paired group

C.depth.150.200 <- merge(C_flux.depth.150,C_flux.depth.200,by = c("cr","yymmdd1","yym
mdd2","Lat2","Lat2.1","Long1","Long2"))
D_bar.C.150.200 <- muCflux150 - muCflux200
C.depth.150.200$D <- C.depth.150.200$C_avg.150 - C.depth.150.200$C_avg.200

C.depth.200.300 <- merge(C_flux.depth.200,C_flux.depth.300,by = c("cr","yymmdd1","yym
mdd2","Lat2","Lat2.1","Long1","Long2"))
D_bar.C.200.300 <- muCflux200 - muCflux300
C.depth.200.300$D <- C.depth.200.300$C_avg.200 - C.depth.200.300$C_avg.300

N.depth.150.200 <- merge(N_flux.depth.150,N_flux.depth.200,by = c("cr","yymmdd1","yym
mdd2","Lat2","Lat2.1","Long1","Long2"))
D_bar.N.150.200 <- muNflux150 - muNflux200
N.depth.150.200$D <- N.depth.150.200$N_avg.150 - N.depth.150.200$N_avg.200

N.depth.200.300 <- merge(N_flux.depth.200,N_flux.depth.300,by = c("cr","yymmdd1","yym
mdd2","Lat2","Lat2.1","Long1","Long2"))
D_bar.N.200.300 <- muNflux200 - muNflux300
N.depth.200.300$D <- N.depth.200.300$N_avg.200 - N.depth.200.300$N_avg.300

P.depth.150.200 <- merge(P_flux.depth.150,P_flux.depth.200,by = c("cr","yymmdd1","yym
mdd2","Lat2","Lat2.1","Long1","Long2"))
D_bar.P.150.200 <- muPflux150 - muPflux200
P.depth.150.200$D <- P.depth.150.200$P_avg.150 - P.depth.150.200$P_avg.200

P.depth.200.300 <- merge(P_flux.depth.200,P_flux.depth.300,by = c("cr","yymmdd1","yym
mdd2","Lat2","Lat2.1","Long1","Long2"))
D_bar.P.200.300 <- muPflux200 - muPflux300
P.depth.200.300$D <- P.depth.200.300$P_avg.200 - P.depth.200.300$P_avg.300

M.depth.150.200 <- merge(M_flux.depth.150,M_flux.depth.200,by = c("cr","yymmdd1","yym
```

```
mdd2","Lat2","Lat2.1","Long1","Long2"))
D_bar.M.150.200 <- muMflux150 - muMflux200
M.depth.150.200$D <- M.depth.150.200$M_avg.150 - M.depth.150.200$M_avg.200

M.depth.200.300 <- merge(M_flux.depth.200,M_flux.depth.300,by = c("cr","yymmdd1","yym
mdd2","Lat2","Lat2.1","Long1","Long2"))
D_bar.M.200.300 <- muMflux200 -muMflux300
M.depth.200.300$D <- M.depth.200.300$M_avg.200 - M.depth.200.300$M_avg.300

D_bar.mean <- data.frame(c("dep150.200","dep200.300"), c(D_bar.C.150.200, D_bar.C.200
.300), c(D_bar.N.150.200, D_bar.N.200.300), c(D_bar.P.150.200 ,D_bar.P.200.300 ), c(D
_bar.M.150.200 ,D_bar.M.200.300))
colnames(D_bar.mean)[1] <- "Depth.Diff"
colnames(D_bar.mean)[2] <- "C_flux"
colnames(D_bar.mean)[3] <- "N_flux"
colnames(D_bar.mean)[4] <- "P_flux"
colnames(D_bar.mean)[5] <- "M_flux"
D_bar.mean
#check distribution of D (difference between paried depth samples)

# Carbon
par(mfrow=c(1,3))
hist(C.depth.150.200$D, breaks = "FD")
hist(log(C.depth.150.200$D), breaks = "FD", main ="Fig 50 C.depth.150.200$D_bar") #tr
y lo transform
boxplot(log(C.depth.150.200$D)) #lots of outliers
# Carbon
par(mfrow=c(1,3))
hist(C.depth.200.300$D, breaks = "FD")
hist(log(C.depth.200.300$D), breaks = "FD", main ="Fig 51 C.depth.200.300$D_bar") #tr
y lo transform
boxplot(log(C.depth.200.300$D)) #lots of outliers
# Nitrogen
par(mfrow=c(1,3))
hist(N.depth.150.200$D, breaks = "FD")
hist(log(N.depth.150.200$D), breaks = "FD", main ="Fig 52 N.depth.150.200$D_bar") #tr
y lo transform
boxplot(log(N.depth.150.200$D)) #lots of outliers
# Nitrogen
par(mfrow=c(1,3))
hist(N.depth.200.300$D, breaks = "FD")
hist(log(N.depth.200.300$D), breaks = "FD", main ="Fig 53 N.depth.200.300$D_bar") #tr
y lo transform
boxplot(log(N.depth.200.300$D)) #lots of outliers
# Phosphorous
par(mfrow=c(1,3))
hist(P.depth.150.200$D, breaks = "FD")
hist(log(P.depth.150.200$D), breaks = "FD", main ="Fig 54 P.depth.150.200$D_bar") #tr
y lo transform
boxplot(log(P.depth.150.200$D)) #lots of outliers
```

```
# Phosporous
par(mfrow=c(1,3))
hist(P.depth.200.300$D, breaks = "FD")
hist(log(P.depth.200.300$D), breaks = "FD", main ="Fig 55 P.depth.200.300$D_bar") #tr
y lo transform
boxplot(log(P.depth.200.300$D)) #lots of outliers
# Mass
par(mfrow=c(1,3))
hist(M.depth.150.200$D, breaks = "FD")
hist(log(M.depth.150.200$D), breaks = "FD", main ="Fig 56 M.depth.150.200$D_bar") #tr
y lo transform
boxplot(log(M.depth.150.200$D)) #lots of outliers
# Mass
par(mfrow=c(1,3))
hist(M.depth.200.300$D, breaks = "FD")
hist(log(M.depth.200.300$D), breaks = "FD", main ="Fig 57 C.depth.200.300$D_bar") #tr
y lo transform
boxplot(log(M.depth.200.300$D)) #lots of outliers
# hyp test 1
t.test(log(C.depth.150.200$C_avg.150) ,log(C.depth.150.200$C_avg.200), alternative =
"greater", paired = TRUE, var.equal = TRUE)
# hyp test 2
t.test(log(C.depth.200.300$C_avg.200) ,log(C.depth.200.300$C_avg.300), alternative =
"greater", paired = TRUE, var.equal = TRUE)
# hyp test 3
t.test(log(N.depth.150.200$N_avg.150) ,log(N.depth.150.200$N_avg.200), alternative =
"greater", paired = TRUE, var.equal = TRUE)
# hyp test 4
t.test(log(N.depth.200.300$N_avg.200) ,log(N.depth.200.300$N_avg.300), alternative =
"greater", paired = TRUE, var.equal = TRUE)
# hyp test 5
t.test(log(P.depth.150.200$P_avg.150) ,log(P.depth.150.200$P_avg.200), alternative =
"greater", paired = TRUE, var.equal = TRUE)
# hyp test 6
t.test(log(P.depth.200.300$P_avg.200) ,log(P.depth.200.300$P_avg.300), alternative =
"greater", paired = TRUE, var.equal = TRUE)
# hyp test 7
t.test(log(M.depth.150.200$M_avg.150) ,log(M.depth.150.200$M_avg.200), alternative =
"greater", paired = TRUE, var.equal = TRUE)
# hyp test 8
t.test(log(M.depth.200.300$M_avg.200) ,log(M.depth.200.300$M_avg.300), alternative =
"greater", paired = TRUE, var.equal = TRUE)
# plotting each nutrient flux against mass flux

#plot(Data.noNA) # we see some relationships between mass flux and the rest of the nu
trient fluxes
par(mfrow=c(1,3))
plot(log(Data.noNA$C_avg),log(Data.noNA$M_avg), main ="Fig 58 M v. C") #good linearit
y, but clumping
lmfit_C.M <- lm(log(M_avg) ~ log(C_avg), data = Data.noNA)
```

```
summary(lmfit_C.M)
abline(lmfit_C.M)

plot(log(Data.noNA$N_avg),log(Data.noNA$M_avg), main ="Fig 59 M v. N") # alright line
arity, still clumping
lmfit_N.M <- lm(log(M_avg) ~ log(N_avg), data = Data.noNA)
summary(lmfit_N.M)
abline(lmfit_N.M)

plot(log(Data.noNA$P_avg),log(Data.noNA$M_avg), main ="Fig 60 M v. P") # worse, still
clumping
lmfit_P.M <- lm(log(M_avg) ~ log(P_avg), data = Data.noNA)
summary(lmfit_P.M)
abline(lmfit_P.M)

#Carbon EDA
resids <- resid( lmfit_C.M ) # extract epsilon_hats
fit <- fitted ( lmfit_C.M ) # extract y_hats

par(mfrow=c(1,2))

hist(resids, breaks=20)
qqnorm(resids)
qqline(resids) # add straight line from true normal

plot(resids, main="Fig 61 resid vs i")
abline(h=0) # mean of epsilon_hat

plot(x=fit, y=resids, main="Fig 62 resid vs y_hat")
abline(h=0)

#Nitrogen EDA
resids <- resid( lmfit_N.M ) # extract epsilon_hats
fit <- fitted ( lmfit_N.M ) # extract y_hats

par(mfrow=c(1,2))

hist(resids, breaks=20)
qqnorm(resids)
qqline(resids) # add straight line from true normal

plot(resids, main="Fig 63 resid vs i")
abline(h=0) # mean of epsilon_hat

plot(x=fit, y=resids, main="Fig 64 resid vs y_hat")
abline(h=0)

#Phosphorus EDA
resids <- resid( lmfit_P.M ) # extract epsilon_hats
fit <- fitted ( lmfit_P.M ) # extract y_hats
```

```
par(mfrow=c(1,2))

hist(resids, breaks=20)
qqnorm(resids)
qqline(resids) # add straight line from true normal

plot(resids, main="Fig 65 resid vs i")
abline(h=0) # mean of epsilon_hat

plot(x=fit, y=resids, main="Fig 66 resid vs y_hat")
abline(h=0)
#Collinearity??
plot(Data.noNA) # we see collinearity between Carbon and Nitrogen
Data.noNA$flux <- Data.noNA$C_avg - Data.noNA$N_avg  #combine the fluxes

lmfit.flux <- lm(log(M_avg) ~ log(flux),
         data=Data.noNA)
summary(lmfit.flux)
plot(log(Data.noNA$flux), log(Data.noNA$M_avg), main ="Fig 67")
abline(lmfit.flux)

resids <- resid( lmfit.flux ) # extract epsilon_hats
fit <- fitted ( lmfit.flux ) # extract y_hats
```

```
par(mfrow=c(1,2))

hist(resids, breaks=20, main ="Fig 68")
qqnorm(resids)
qqline(resids) # add straight line from true normal

plot(resids, main="Fig 69 resid vs i")
abline(h=0) # mean of epsilon_hat

plot(x=fit, y=resids, main="Fig 69 resid vs y_hat")
abline(h=0)

# dept v. C_avg

par(mfrow=c(1,4))

plot(Data.noNA$dep, log(Data.noNA$C_avg), main ="Fig 70")
lmfit_Dep.C <- lm(log(C_avg) ~ dep, data = Data.noNA)
summary(lmfit_Dep.C)
abline(lmfit_Dep.C)

plot(Data.noNA$dep, log(Data.noNA$N_avg), main= "Fig 71")
lmfit_Dep.N <- lm(log(N_avg) ~ dep, data = Data.noNA)
summary(lmfit_Dep.N)
abline(lmfit_Dep.N)
```

```
plot(Data.noNA$dep, log(Data.noNA$P_avg),  main= "Fig 72")
lmfit_Dep.P <- lm(log(P_avg) ~ dep, data = Data.noNA)
summary(lmfit_Dep.P)
abline(lmfit_Dep.P)

plot(Data.noNA$dep, log(Data.noNA$M_avg),  main= "Fig 73")
lmfit_Dep.M <- lm(log(M_avg) ~ dep, data = Data.noNA)
summary(lmfit_Dep.M)
abline(lmfit_Dep.M)
#multiple regression
lmfit.MR <- lm(log(M_avg) ~ log(flux) + dep, data = Data.noNA)
summary(lmfit.MR)

resids <- resid( lmfit.MR ) # extract epsilon_hats
fit <- fitted ( lmfit.MR ) # extract y_hats

par(mfrow=c(1,2))

hist(resids, breaks=20,  main= "Fig 74")
qqnorm(resids)
qqline(resids) # add straight line from true normal

plot(resids, main="Fig 75 resid vs i")
abline(h=0) # mean of epsilon_hat

plot(x=fit, y=resids, main="Fig 76 resid vs y_hat")
abline(h=0)
#anova 1
smaller <- lmfit.flux; larger <- lmfit.MR
anova(smaller, larger)
#anova 2
smaller.1 <- lmfit_Dep.M; larger <- lmfit.MR
anova(smaller.1, larger)
#we want to make predictions on the Mass average flux.

# we want to compare this against depth and C,N,P fluxes

# Depth is categorical
# N,P,C are numerical

library(psych) # for `pairs.panels()`
library(lattice)

ANOCOVA.data <- subset(Data.noNA, select = c("dep", "M_avg", "C_avg","N_avg","P_avg")
)
attach(ANOCOVA.data)

summary(ANOCOVA.data)
```

```
pairs.panels(data.frame(as.factor(dep),log(C_avg),log(N_avg),log(P_avg),log(M_avg)),
main = "Fig 77")

xyplot(log(M_avg) ~ log(C_avg) | as.factor(dep),
                  main="Fig 78: Activity level-specific scatterplots"
           )

    xyplot(log(M_avg) ~ log(C_avg), groups=as.factor(dep),
                  auto.key=TRUE,
                  main="Fig 79: Scatterplot with color=group level"
           )

    xyplot(log(M_avg) ~ log(C_avg), groups=as.factor(dep),
                  type=c("p","r"), # `p` for _points_, `r` for _regression line_
                                   # see https://stackoverflow.com/questions/12972039
/plotting-xyplot-with-regression-line-on-lattice-graphics for more
                  auto.key=TRUE,
                  main="Fig 80: Three standalone depth level-specific `lm()` fits"
           )
#reformat yymmdd so that r can better use it blocking factors

# only yymmdd1 was used to create new year and month columns since yymmdd1 and yymmdd
2 sampled were only a couple days apart and were averaged out in flux columns

Data.noNA$Date <- as.Date(paste(substr(Data.noNA$yymmdd1,1,4),
                                substr(Data.noNA$yymmdd1,5,6),
                                substr(Data.noNA$yymmdd1,7,8), sep = "-"),
                          format = '%Y-%m-%d')

Data.noNA$Year <- substr(Data.noNA$yymmdd1,1,4)

#install.packages("lubridate")
library(lubridate)
Data.noNA$month <- as.Date(paste(substr(Data.noNA$yymmdd1,5,6),
                                 substr(Data.noNA$yymmdd1,7,8), sep = "-"),
                           format = '%m-%d')
Data.noNA$month <- round_date(Data.noNA$month, unit = "month")
Data.noNA$month <- format(Data.noNA$month,format = "%Y-%b-%d")
Data.noNA$month <- substr(paste(Data.noNA$month),6,8)

#ANCOVA
ANOCOVA.data <- subset(Data.noNA, select = c("dep", "M_avg", "C_avg","N_avg","P_avg",
"Year", "month"))
ANOCOVA.data <- as.vector(ANOCOVA.data)
ANCOVA <- lm(log(M_avg) ~ log(C_avg) + log(N_avg) + log(P_avg) + dep + as.factor(Year
) + as.factor(month),data = ANOCOVA.data)
summary(ANCOVA)
plot(ANCOVA, main = "Fig 81") #challen
resids <- resid(ANCOVA)
fit <-fitted(ANCOVA)
```