Doctor_Effectiveness.R

kenneywl

Thu Dec 27 13:24:34 2018

```r
#Objective: To maximize net profit by considering doctor effectiveness individually
#           and in pairs.

#Methods:1.Clean the data and put it in usable form.
#        2.Do power analysis to determine what we should expect.
#        3A.Determine the effectiveness of each doctor individually. (block by month and day)
#        3B.Determine if certain doctor pairs work better or worse together.
#        3C. Model adequcy checking.
#        4.Interpretation and discussion.
#        5. Final words.
################################################################################
################################################################################
#1) Clean the data.

#First we clean the data and put in all in a dataframe so that it is usable.
#This first part is parsing the data to get in a usable form.
#It is not important to follow this part.
sw <- read.csv("SW2018.csv",header=T)
swp <- sw[11:153,c(8,9,13,6,7,10,20,23:26)]
swp <- swp[swp[,"X..OF.DOCTORS"]!=0,]
#there is only one day with one doctor,
#this is not included in the analysis.
swp <- swp[-135,]

swp[,"Pediatric.Dentist.Day."] <- factor(swp[,"Pediatric.Dentist.Day."])
swp[,"TOTAL.PRODUCTION"] <- as.numeric(substring(gsub(",","",gsub(" ","",
                                                 as.character(swp[,"TOTAL.PRODUCTION"]))),2))
swp[,"X..OF.DOCTORS"] <- as.integer(as.character(swp[,"X..OF.DOCTORS"]))

#Now we have some missing values for total patients, we delete them
swp <- swp[-(136:139),]

#This will make the indicator matrix of each doctor:
library(ade4)
docsn <- swp[,8:11]
docsn[docsn==""] <- NA
docsn <- data.frame(sapply(docsn,function(x){as.factor(x)}))
docs <- acm.disjonctif(docsn)
#we can just add repeat docs together to consolidate
names(docs) <- sapply(names(docs),function(x){substring(x,13)})

for(i in 1:22){
  for(j in (i+1):23){
    if(names(docs)[i] == names(docs)[j]){
      docs[i] <- docs[i]+docs[j]
    }
  }
```

```r
}
#I nulled out the duplicates by hand to avoid a complicated for loop.
docs[,c(8,16,9,17,10,11,18,12,15,19,22,20,21,23)] <- NULL

#I changed a single day with 2 temps to 1 temp to smooth out the analysis
#it makes understanding the data easier and will have little
#effect on the final analysis.

docs[c(32,118),9] <- 1
docs <- data.frame(sapply(docs,function(x){y <- as.factor(x);y}))

#and combind docs with swp:

swp <- cbind(swp[,1:7],docs)
#I remove the factor pediatric here because the effect is included
#when we test for "pino", who is the only pediatric doctor.

names(swp) <- c("Total_Operative","Total_Hygiene","Total_Production",
                "Number_of_Doctors","Pediatric","Total_Patients",
                "Appt_Scheduled", names(swp)[8:16])

sx <- sw[11:153,c(3,4,6)]
sx <- sx[sx[,"X..OF.DOCTORS"]!=0,]
#there is only one day with one doctor,
#this is not included in the analysis.
sx <- sx[-135,]
sx <- sx[-(136:139),]

swp <- cbind(swp,sx[,c(1,2)])

#there is an extranous level in DAY

swp$DAY <- factor(swp$DAY)
swp$MONTH <- factor(swp$MONTH)

#Our data is now usable. The data frame is swp.
#The first few rows are:
head(swp,4)
```

```
##    Total_Operative Total_Hygiene Total_Production Number_of_Doctors
## 11              24            30          8042.12                 2
## 12              20            35          7774.81                 3
## 13              30            40          9680.57                 3
## 14               0            62          6583.58                 2
##    Pediatric Total_Patients Appt_Scheduled BRITTANY HEATHER KEVIN OLIMBI
## 11        NO             54            107        0       0     1      0
## 12        NO             55            102        1       0     1      0
## 13        NO             70            100        1       0     0      1
## 14        NO             62            109        0       0     1      0
##    PINO POUYAN ABBY SAMARA TEMP       DAY   MONTH
## 11    0      1    0      0    0   Tuesday January
## 12    0      0    0      0    1 Wednesday January
## 13    0      0    0      0    1  Thursday January
```

```
## 14    0      0    0      1    0    Friday January
```

```r
##########################################################
##########################################################
#2) Do power analysis. What should we expect with
#   different numbers of predictors?

#I included a power analysis because the number of predictors
#is high compared to the number of data points (p=46, n=135)

#We want to use each doctor and also each doctor pair as a predictor.
#and block by 8 months.
#Thats 9 doctors + (9 choose 2) pairs + 8 months + a constant

9+choose(9,2)+8+1
```

```
## [1] 54
```

```r
#54 predictors total. We use the library pwr.
#with an effect ratio of .15 and alpha level .05

library(pwr)
pwr.f2.test(u=54,v=135-54,f2=.15,sig.level=.05)
```

```
##
##      Multiple regression power calculation
##
##               u = 54
##               v = 81
##              f2 = 0.15
##       sig.level = 0.05
##           power = 0.3642497
```

```r
#our power is 36%, that means we have much worse than
#equal chance of finding significance. The standard
#is to want a power of 80%

#At best, assuming the month factor is insignificant we have
#9 + 9 choose 2 + 1 predictors

pwr.f2.test(u=46,v=135-46,f2=.15,sig.level=.05)
```

```
##
##      Multiple regression power calculation
##
##               u = 46
##               v = 89
##              f2 = 0.15
##       sig.level = 0.05
##           power = 0.4244568
```

```r
#A power of 42% that is too low.
#Effectivly we don't have enough data.

#If we just had the doctors and the 8 months

pwr.f2.test(u=18,v=135-18,f2=.15,sig.level=.05)
```

```
##
##      Multiple regression power calculation
##
##               u = 18
##               v = 117
##              f2 = 0.15
##       sig.level = 0.05
##           power = 0.7440315
```

```r
#That's not terrible. 74% is just about good enough.
#Unfortunately, the answer they want is whether
#doctor pairs make any difference.

#The answer is "Not enough data"

#It is determined that we do not have enough data to
#reasonably expect to get significance from each doctor pair.
#I'll run the models, but I don't expect to get anything.

########################################################################
#############################################################################
#3a) Detrmine the effectiveness of individual doctors.

#Lets build just the blocked factors.

contrasts(swp$DAY) <- "contr.sum" #Force factor to sum to zero.
contrasts(swp$MONTH) <- "contr.sum"

swlm <- lm(Total_Production ~ DAY + MONTH,data=swp)
summary(swlm)
```

```
##
## Call:
## lm(formula = Total_Production ~ DAY + MONTH, data = swp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3657.5  -702.2    53.4   711.1  3823.0
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10075.17     200.23  50.317  < 2e-16 ***
## DAY1        -2543.54     359.66  -7.072 1.04e-10 ***
## DAY2         1256.92     293.49   4.283 3.70e-05 ***
## DAY3          -89.59     859.40  -0.104  0.91714
```

```
## DAY4          241.03     290.42   0.830  0.40819
## DAY5          556.78     289.51   1.923  0.05678 .
## MONTH1         38.82     301.84   0.129  0.89787
## MONTH2       1090.59     380.29   2.868  0.00487 **
## MONTH3       -422.05     321.51  -1.313  0.19174
## MONTH4      -1074.96     389.15  -2.762  0.00663 **
## MONTH5        597.11     307.27   1.943  0.05429 .
## MONTH6        235.16     307.06   0.766  0.44524
## MONTH7       -246.82     322.78  -0.765  0.44593
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1408 on 122 degrees of freedom
## Multiple R-squared:  0.4414, Adjusted R-squared:  0.3864
## F-statistic: 8.033 on 12 and 122 DF,  p-value: 6.236e-11
```

```r
#adj R is .39
#Lets build the main effects.

docsnam <- paste0(c("(",paste0(names(docs),collapse="+"),")"),collapse="")

form <- formula(paste0(c("Total_Production ~  MONTH + DAY +",docsnam), collapse=""))
swlm1 <- lm(form,data=swp)

#Lets look at the model:
anova(swlm1)
```

```
## Analysis of Variance Table
##
## Response: Total_Production
##            Df    Sum Sq  Mean Sq F value    Pr(>F)
## MONTH       7  39155687  5593670  3.1656 0.0042864 **
## DAY         5 151922573 30384515 17.1954 1.269e-12 ***
## BRITTANY    1   1325897  1325897  0.7504 0.3881982
## HEATHER     1   6856670  6856670  3.8804 0.0513007 .
## KEVIN       1    672235   672235  0.3804 0.5386114
## OLIMBI      1    111817   111817  0.0633 0.8018415
## PINO        1  20566109 20566109 11.6389 0.0008969 ***
## POUYAN      1  11528687 11528687  6.5244 0.0119714 *
## ABBY        1     93299    93299  0.0528 0.8186755
## SAMARA      1    212346   212346  0.1202 0.7294928
## TEMP        1    791152   791152  0.4477 0.5047782
## Residuals 113 199672926  1767017
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(swlm1)
```

```
##
## Call:
## lm(formula = form, data = swp)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -3323.5  -717.8   -34.3   672.7  2997.7
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8696.54     895.61   9.710  < 2e-16 ***
## MONTH1         47.44     288.06   0.165  0.86949
## MONTH2       1457.04     415.93   3.503  0.00066 ***
## MONTH3       -670.73     328.21  -2.044  0.04332 *
## MONTH4       -763.22     398.09  -1.917  0.05773 .
## MONTH5        305.22     333.02   0.917  0.36135
## MONTH6         41.91     307.31   0.136  0.89177
## MONTH7       -124.91     361.88  -0.345  0.73061
## DAY1        -2062.44     956.84  -2.155  0.03325 *
## DAY2          395.04     511.10   0.773  0.44118
## DAY3         -193.17     958.34  -0.202  0.84061
## DAY4          788.13     671.50   1.174  0.24299
## DAY5          379.98     560.91   0.677  0.49952
## BRITTANY1    -215.98     698.09  -0.309  0.75760
## HEATHER1      738.06     504.95   1.462  0.14661
## KEVIN1        427.10     452.57   0.944  0.34733
## OLIMBI1       674.77     565.99   1.192  0.23568
## PINO1        1489.99     461.48   3.229  0.00163 **
## POUYAN1      1267.48     568.94   2.228  0.02787 *
## ABBY1          69.98     545.51   0.128  0.89815
## SAMARA1       325.97    1109.23   0.294  0.76939
## TEMP1        -223.37     333.82  -0.669  0.50478
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1329 on 113 degrees of freedom
## Multiple R-squared:  0.5388, Adjusted R-squared:  0.453
## F-statistic: 6.285 on 21 and 113 DF,  p-value: 3.002e-11
```

```r
#Adjusted R^2 is low: .45


###########################################################################
###########################################################################
#Lets run the doctor pairs.

#3b) Determine effectiveness of doctors individually and in pairs.
#We build a linear model.
#Net_Production ~ DAY + Month + (each doct)+(each doc pair interaction)

#These are the pairs for interaction term.
interact <- NULL
for(i in 1:8){
  for(j in (i+1):9){
    intname <- paste0(c(names(docs)[i],names(docs)[j]),collapse = "*")
    interact <- c(interact,intname)
  }
}
```

```r
#There should be 9 Choose 2 of them:

choose(9,2);length(interact)
```

```
## [1] 36
```

```
## [1] 36
```

```r
#Good. Now lets put it all together.
docsnam_all <- paste0(c("(",paste0(c(names(docs),interact),collapse="+"),")"),collapse="")

form <- formula(paste0(c("Total_Production ~ MONTH + DAY+",docsnam_all), collapse=""))
swlm2 <- lm(form,data=swp)
summary(swlm2)
```

```
##
## Call:
## lm(formula = form, data = swp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2949.2  -569.5     0.0   610.5  2790.9
##
## Coefficients: (10 not defined because of singularities)
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        12038.273   5149.013   2.338  0.02168 *
## MONTH1               390.351    325.719   1.198  0.23401
## MONTH2              1536.413    577.969   2.658  0.00935 **
## MONTH3              -984.018    362.187  -2.717  0.00795 **
## MONTH4              -824.287    461.756  -1.785  0.07773 .
## MONTH5               349.719    361.617   0.967  0.33618
## MONTH6               354.215    323.095   1.096  0.27597
## MONTH7              -624.748    437.951  -1.427  0.15730
## DAY1               -2408.912   2631.229  -0.916  0.36246
## DAY2                 960.019   1433.854   0.670  0.50493
## DAY3                -988.711   2820.057  -0.351  0.72674
## DAY4                 377.031   1173.620   0.321  0.74879
## DAY5                 990.465   1163.049   0.852  0.39677
## BRITTANY1           2076.941   3125.798   0.664  0.50816
## HEATHER1           -2250.602   5193.987  -0.433  0.66586
## KEVIN1             -3338.416   4748.989  -0.703  0.48395
## OLIMBI1            -1622.324   4475.137  -0.363  0.71784
## PINO1               3247.622   2285.800   1.421  0.15895
## POUYAN1            -2362.111   4518.736  -0.523  0.60249
## ABBY1                162.752   2880.732   0.056  0.95508
## SAMARA1              687.728   3102.980   0.222  0.82512
## TEMP1              -2380.519   3024.127  -0.787  0.43332
## BRITTANY1:HEATHER1 -1589.102   3387.480  -0.469  0.64016
## BRITTANY1:KEVIN1   -1908.608   2127.787  -0.897  0.37220
## BRITTANY1:OLIMBI1   -143.104   2186.230  -0.065  0.94796
## BRITTANY1:PINO1    -1784.775   2054.147  -0.869  0.38731
## BRITTANY1:POUYAN1      3.225   1471.836   0.002  0.99826
```

```
## BRITTANY1:ABBY1    -1041.043   1942.431  -0.536  0.59336
## BRITTANY1:SAMARA1         NA         NA      NA       NA
## BRITTANY1:TEMP1     -819.726   1302.425  -0.629  0.53075
## HEATHER1:KEVIN1     3986.581   2797.327   1.425  0.15769
## HEATHER1:OLIMBI1     470.870   3448.790   0.137  0.89172
## HEATHER1:PINO1     -1570.968   2008.048  -0.782  0.43614
## HEATHER1:POUYAN1    2480.802   5443.504   0.456  0.64972
## HEATHER1:ABBY1      2376.933   3660.742   0.649  0.51785
## HEATHER1:SAMARA1     522.425   3085.062   0.169  0.86592
## HEATHER1:TEMP1      2162.418   1697.789   1.274  0.20617
## KEVIN1:OLIMBI1      2567.263   3740.818   0.686  0.49436
## KEVIN1:PINO1       -1933.533   3083.603  -0.627  0.53228
## KEVIN1:POUYAN1      3147.764   4438.147   0.709  0.48006
## KEVIN1:ABBY1         204.675   2814.974   0.073  0.94220
## KEVIN1:SAMARA1            NA         NA      NA       NA
## KEVIN1:TEMP1        2486.975   1926.974   1.291  0.20026
## OLIMBI1:PINO1             NA         NA      NA       NA
## OLIMBI1:POUYAN1     2295.499   5460.032   0.420  0.67522
## OLIMBI1:ABBY1             NA         NA      NA       NA
## OLIMBI1:SAMARA1           NA         NA      NA       NA
## OLIMBI1:TEMP1        413.798   1567.182   0.264  0.79237
## PINO1:POUYAN1        259.291   2566.908   0.101  0.91977
## PINO1:ABBY1        -1905.191   1893.199  -1.006  0.31705
## PINO1:SAMARA1             NA         NA      NA       NA
## PINO1:TEMP1         -821.314   1616.391  -0.508  0.61266
## POUYAN1:ABBY1             NA         NA      NA       NA
## POUYAN1:SAMARA1           NA         NA      NA       NA
## POUYAN1:TEMP1        952.290   1297.067   0.734  0.46481
## ABBY1:SAMARA1             NA         NA      NA       NA
## ABBY1:TEMP1        -1382.947   1256.336  -1.101  0.27403
## SAMARA1:TEMP1             NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1250 on 87 degrees of freedom
## Multiple R-squared:  0.686,  Adjusted R-squared:  0.5163
## F-statistic: 4.044 on 47 and 87 DF,  p-value: 8.949e-09
```

```
anova(swlm2)
```

```
## Analysis of Variance Table
##
## Response: Total_Production
##                   Df    Sum Sq  Mean Sq F value    Pr(>F)
## MONTH              7  39155687  5593670  3.5798 0.0019729 **
## DAY                5 151922573 30384515 19.4451 5.965e-13 ***
## BRITTANY           1   1325897  1325897  0.8485 0.3595162
## HEATHER            1   6856670  6856670  4.3880 0.0390997 *
## KEVIN              1    672235   672235  0.4302 0.5136182
## OLIMBI             1    111817   111817  0.0716 0.7897131
## PINO               1  20566109 20566109 13.1616 0.0004813 ***
## POUYAN             1  11528687 11528687  7.3780 0.0079644 **
## ABBY               1     93299    93299  0.0597 0.8075329
## SAMARA             1    212346   212346  0.1359 0.7132930
```

```
## TEMP               1    791152   791152  0.5063 0.4786440
## BRITTANY:HEATHER   1    569140   569140  0.3642 0.5477364
## BRITTANY:KEVIN     1  14215675 14215675  9.0975 0.0033552 **
## BRITTANY:OLIMBI    1    387182   387182  0.2478 0.6198960
## BRITTANY:PINO      1   8741360  8741360  5.5942 0.0202436 *
## BRITTANY:POUYAN    1      2579     2579  0.0017 0.9676842
## BRITTANY:ABBY      1     62710    62710  0.0401 0.8416894
## BRITTANY:TEMP      1   6618671  6618671  4.2357 0.0425740 *
## HEATHER:KEVIN      1   6406779  6406779  4.1001 0.0459462 *
## HEATHER:OLIMBI     1    813746   813746  0.5208 0.4724485
## HEATHER:PINO       1    783535   783535  0.5014 0.4807628
## HEATHER:POUYAN     1    126585   126585  0.0810 0.7766093
## HEATHER:ABBY       1    639749   639749  0.4094 0.5239468
## HEATHER:SAMARA     1    976285   976285  0.6248 0.4314224
## HEATHER:TEMP       1      7583     7583  0.0049 0.9446239
## KEVIN:OLIMBI       1   1741903  1741903  1.1148 0.2939731
## KEVIN:PINO         1   4146718  4146718  2.6538 0.1069213
## KEVIN:POUYAN       1     58913    58913  0.0377 0.8464951
## KEVIN:ABBY         1     35330    35330  0.0226 0.8808238
## KEVIN:TEMP         1  12342277 12342277  7.8986 0.0061106 **
## OLIMBI:POUYAN      1    222283   222283  0.1423 0.7069690
## OLIMBI:TEMP        1    272456   272456  0.1744 0.6772924
## PINO:POUYAN        1    293090   293090  0.1876 0.6660209
## PINO:ABBY          1    577918   577918  0.3698 0.5446715
## PINO:TEMP          1    343279   343279  0.2197 0.6404512
## POUYAN:TEMP        1   1449057  1449057  0.9273 0.3382223
## ABBY:TEMP          1   1893399  1893399  1.2117 0.2740299
## Residuals         87 135944724  1562583
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Adj R squared is .53, not terrible, but not good.

#I'll fit one more model with just the significant terms of the above model.
docsnam_int_sig <- paste0(c("(",paste0(c(names(docs),interact[c(21,9,8,4,2)]),
                                       collapse="+"),")"),collapse="")

form <- formula(paste0(c("Total_Production ~ MONTH + DAY +",docsnam_int_sig), collapse=""))
swlm3 <- lm(form,data=swp)
summary(swlm3)
```

```
##
## Call:
## lm(formula = form, data = swp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2944.7  -657.4    55.7   731.0  3353.6
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       9886.23     901.95  10.961  < 2e-16 ***
## MONTH1             272.93     278.36   0.980  0.32904
## MONTH2            1316.12     398.75   3.301  0.00131 **
```

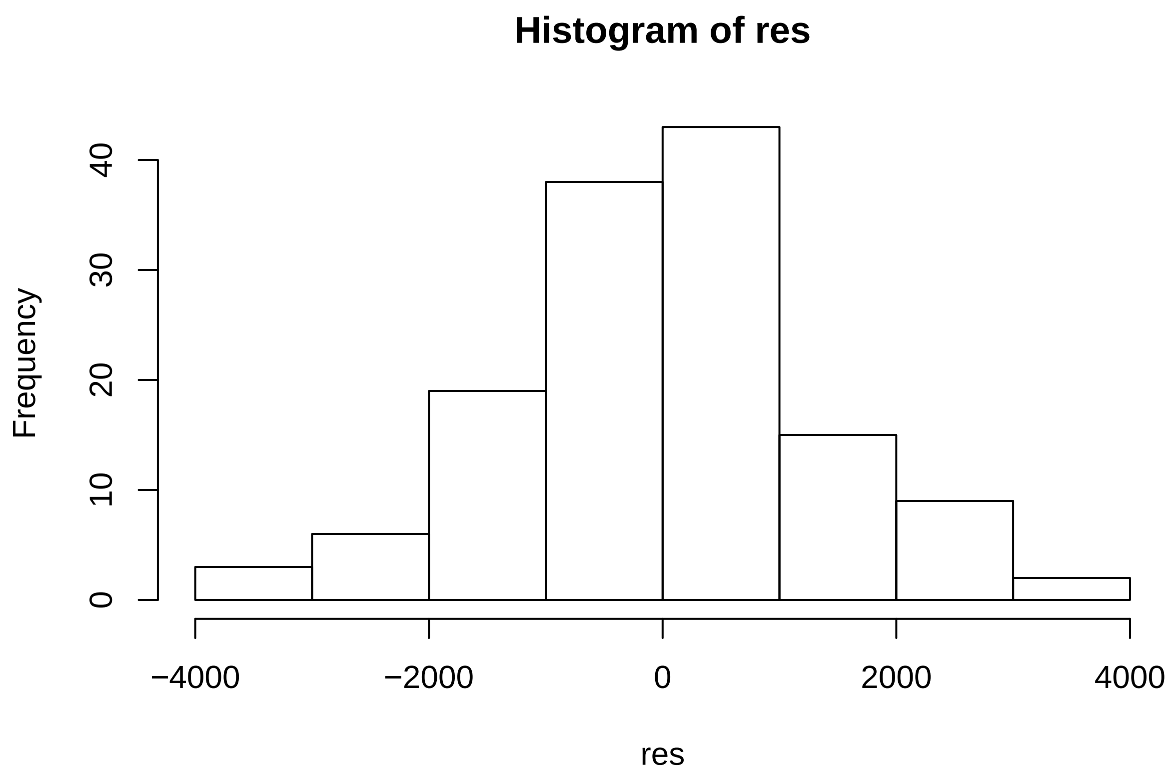
```
## MONTH3              -981.43     307.38  -3.193  0.00185 **
## MONTH4              -776.90     380.11  -2.044  0.04340 *
## MONTH5               492.33     305.49   1.612  0.10996
## MONTH6               380.38     289.60   1.313  0.19180
## MONTH7              -523.00     348.63  -1.500  0.13649
## DAY1               -3039.76     964.14  -3.153  0.00209 **
## DAY2                 851.49     495.57   1.718  0.08863 .
## DAY3                1474.88    1033.11   1.428  0.15629
## DAY4                 274.23     663.64   0.413  0.68026
## DAY5                  72.42     523.62   0.138  0.89026
## BRITTANY1           1056.16     970.06   1.089  0.27869
## HEATHER1            -484.58     677.17  -0.716  0.47578
## KEVIN1              -505.87     663.59  -0.762  0.44753
## OLIMBI1              699.51     563.57   1.241  0.21722
## PINO1                425.67     523.49   0.813  0.41792
## POUYAN1             1377.93     533.04   2.585  0.01107 *
## ABBY1               -192.73     516.67  -0.373  0.70986
## SAMARA1              801.77    1028.32   0.780  0.43728
## TEMP1              -1237.25     566.46  -2.184  0.03111 *
## KEVIN1:TEMP1        1536.44     630.98   2.435  0.01653 *
## HEATHER1:KEVIN1     2048.73     748.49   2.737  0.00725 **
## BRITTANY1:TEMP1     -920.79     578.17  -1.593  0.11417
## BRITTANY1:PINO1     1130.89     908.41   1.245  0.21586
## BRITTANY1:KEVIN1 -1381.62     744.68  -1.855  0.06628 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1206 on 108 degrees of freedom
## Multiple R-squared:  0.6372, Adjusted R-squared:  0.5498
## F-statistic: 7.295 on 26 and 108 DF,  p-value: 4.209e-14
```

```
#Adj R is .55

########################################################################################
########################################################################################
#3c) Model Adequecy


#We need to first look at correlation of the factors.

docs_n <- sapply(docs,FUN=as.numeric)
cor(docs_n)
```

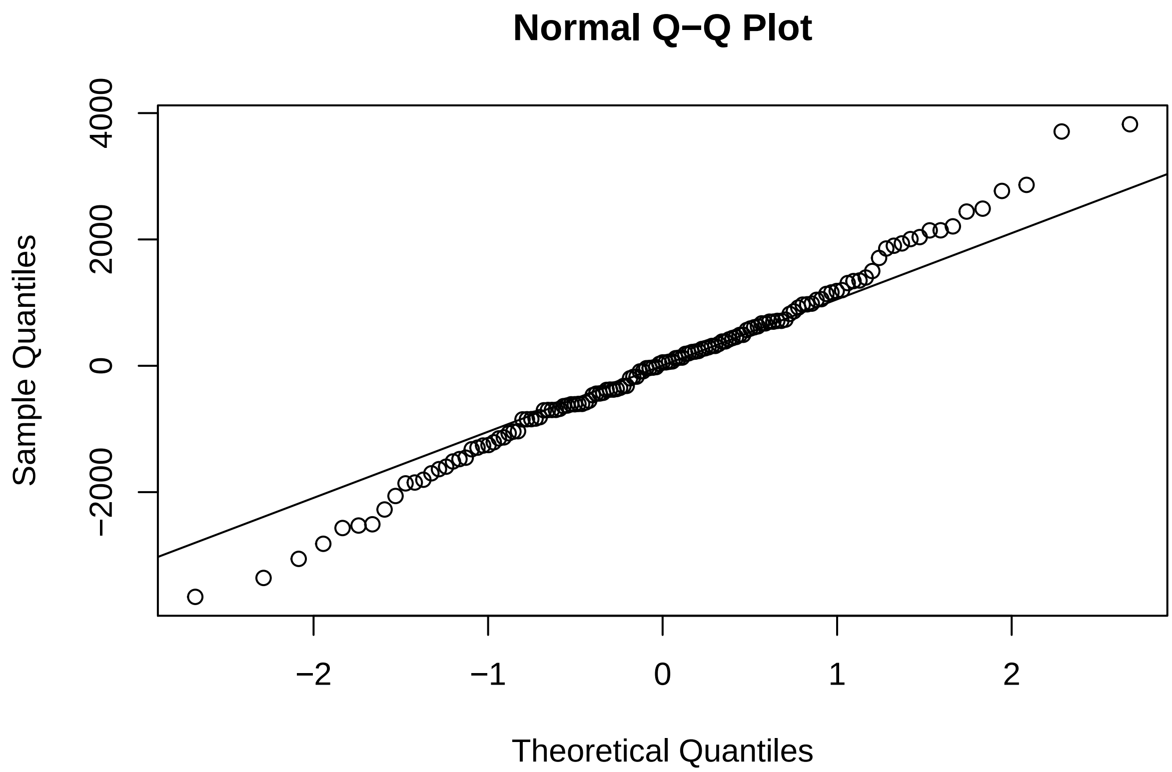
```
##              BRITTANY      HEATHER       KEVIN     OLIMBI       PINO
## BRITTANY  1.000000000 -0.361619132  0.06471087  0.1557398 -0.2455069
## HEATHER  -0.361619132  1.000000000 -0.39575357  0.1462642  0.5038487
## KEVIN     0.064710870 -0.395753572  1.00000000 -0.5994312 -0.4871723
## OLIMBI    0.155739825  0.146264181 -0.59943122  1.0000000  0.3033686
## PINO     -0.245506885  0.503848722 -0.48717227  0.3033686  1.0000000
## POUYAN    0.004929401 -0.395313285  0.32727069 -0.5468531 -0.3144680
## ABBY     -0.323974995 -0.217536034  0.10839166 -0.3124638 -0.2119115
## SAMARA   -0.262435613 -0.015854406  0.20431380 -0.2543532 -0.1856730
## TEMP     -0.109913438  0.005837572 -0.19951223  0.1361255  0.1632187
##                POUYAN         ABBY      SAMARA         TEMP
```

```
## BRITTANY  0.004929401 -0.32397500 -0.26243561 -0.109913438
## HEATHER  -0.395313285 -0.21753603 -0.01585441  0.005837572
## KEVIN     0.327270691  0.10839166  0.20431380 -0.199512232
## OLIMBI   -0.546853073 -0.31246379 -0.25435320  0.136125543
## PINO     -0.314468033 -0.21191154 -0.18567295  0.163218722
## POUYAN    1.000000000  0.36312704 -0.24246706 -0.185517066
## ABBY      0.363127038  1.00000000 -0.14792673 -0.083024467
## SAMARA   -0.242467064 -0.14792673  1.00000000 -0.238560949
## TEMP     -0.185517066 -0.08302447 -0.23856095  1.000000000
```

```r
#We have two stronger than .5, but not by much. .59 and .54

res <- resid(swlm)
res1 <- resid(swlm1)
res2 <- resid(swlm2)
res3 <- resid(swlm3)

hist(res)
```

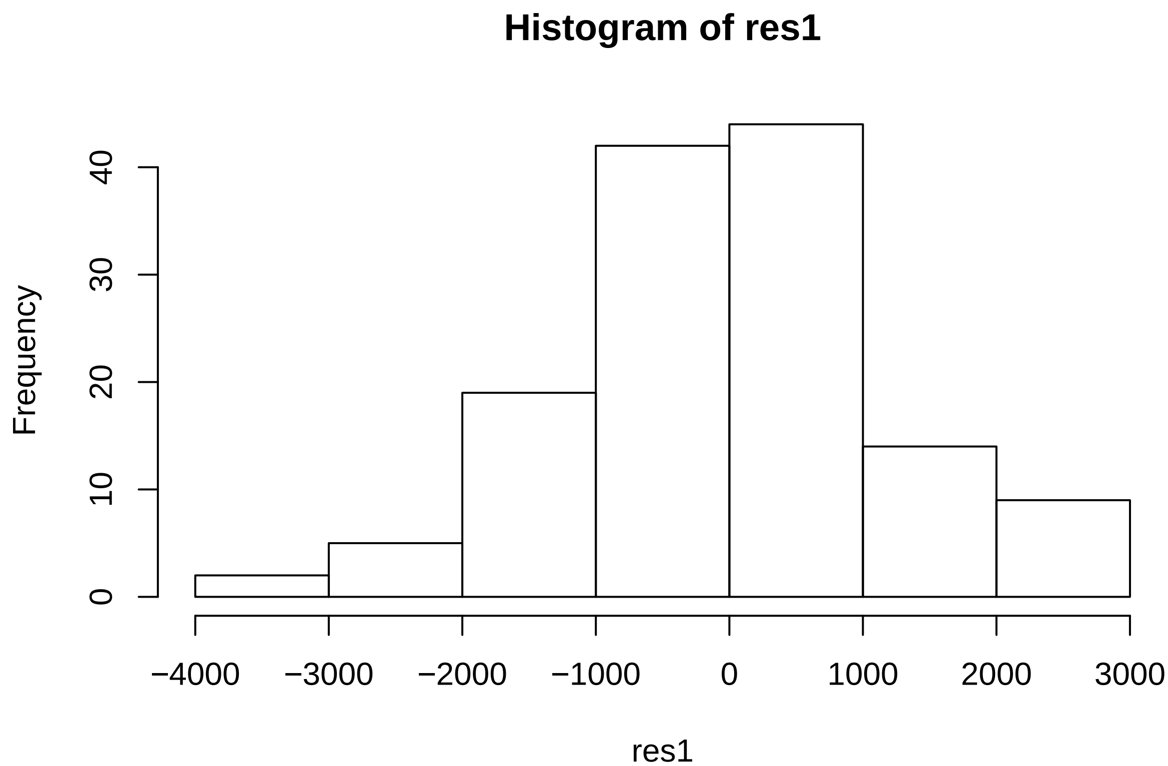



```r
qqnorm(res);qqline(res)
```

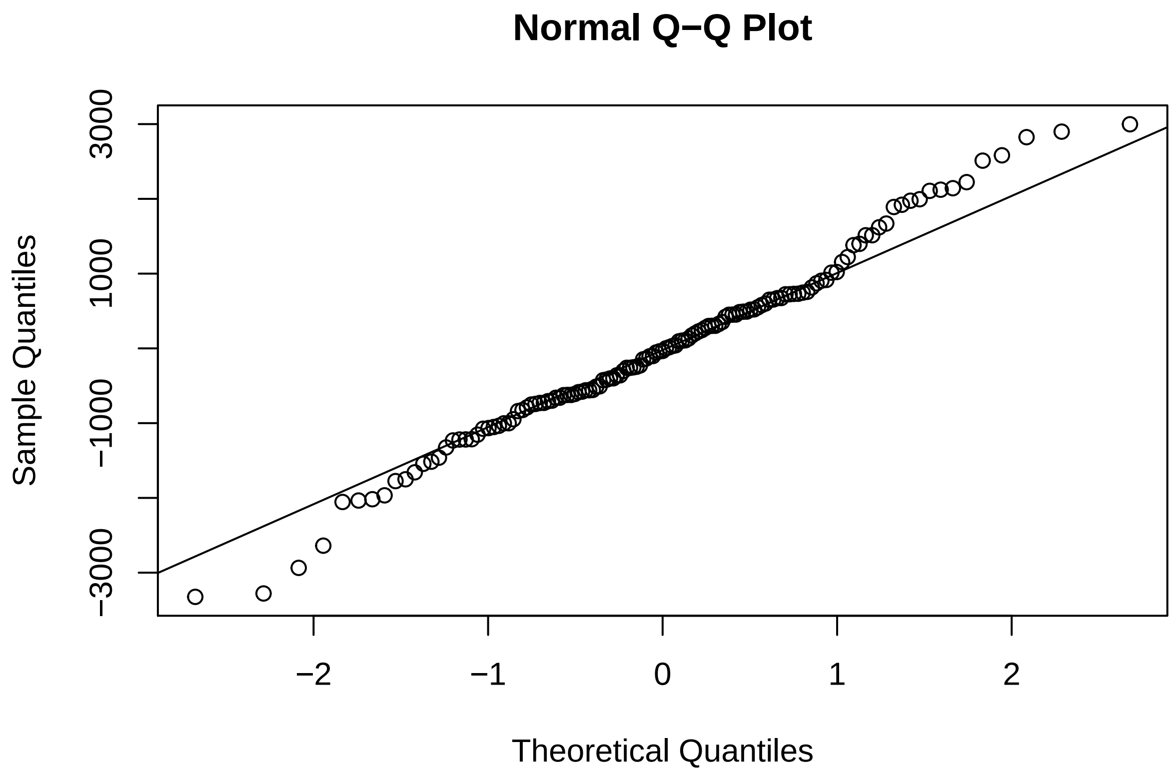


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
4000
2000
0
−2000
−2
−1
0
1
2

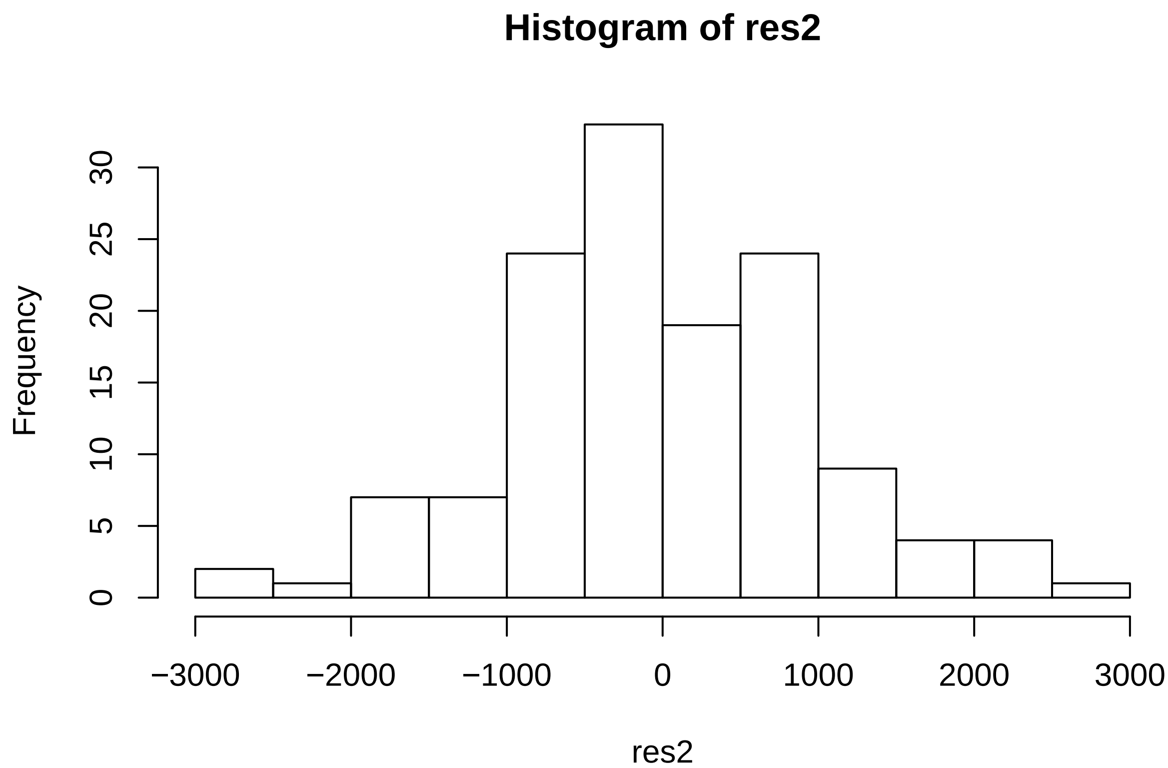

```r
hist(res1)
```



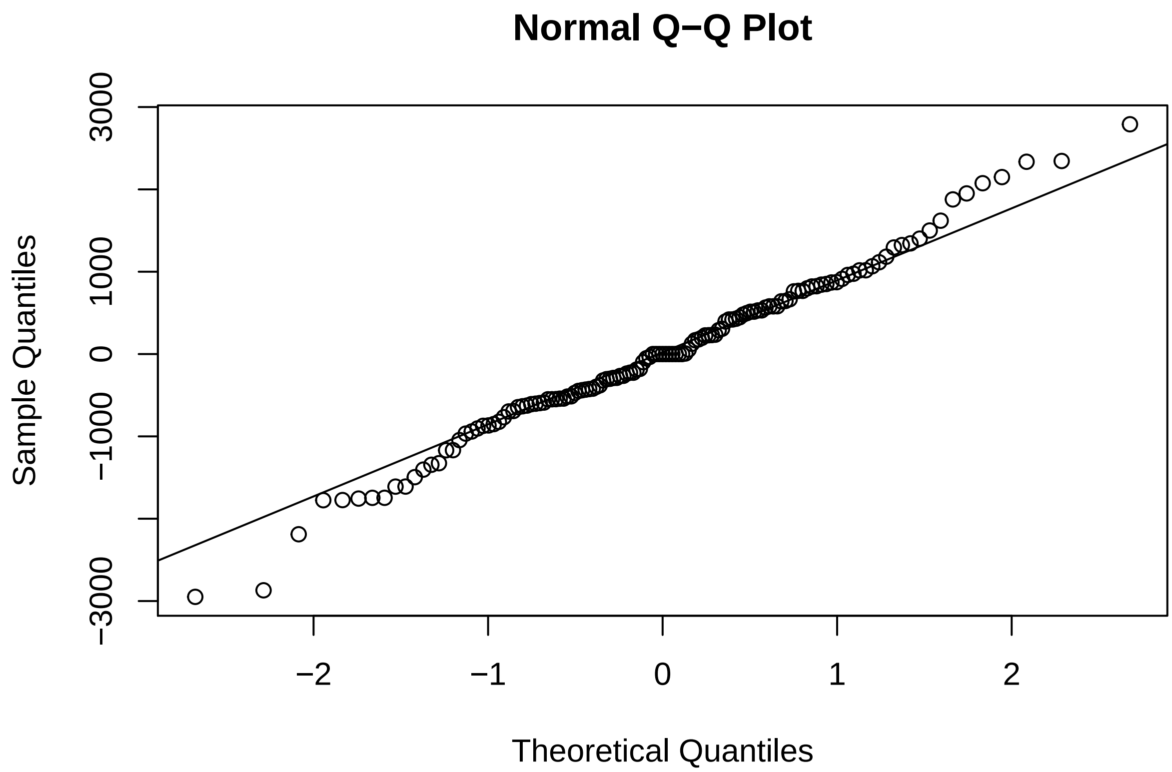

```r
qqnorm(res1);qqline(res1)
```


Normal Q−Q Plot
Sample Quantiles
−3000
−1000
1000
3000
−2
−1
0
1
2
Theoretical Quantiles

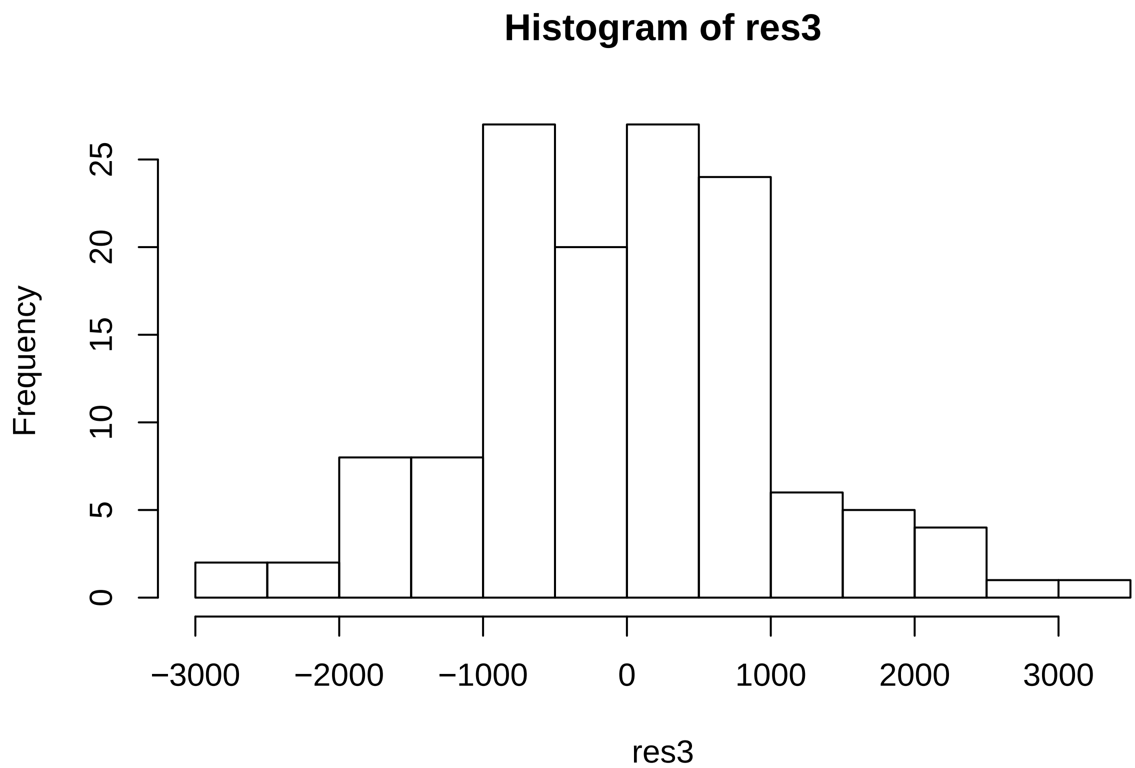
```r
hist(res2)
```



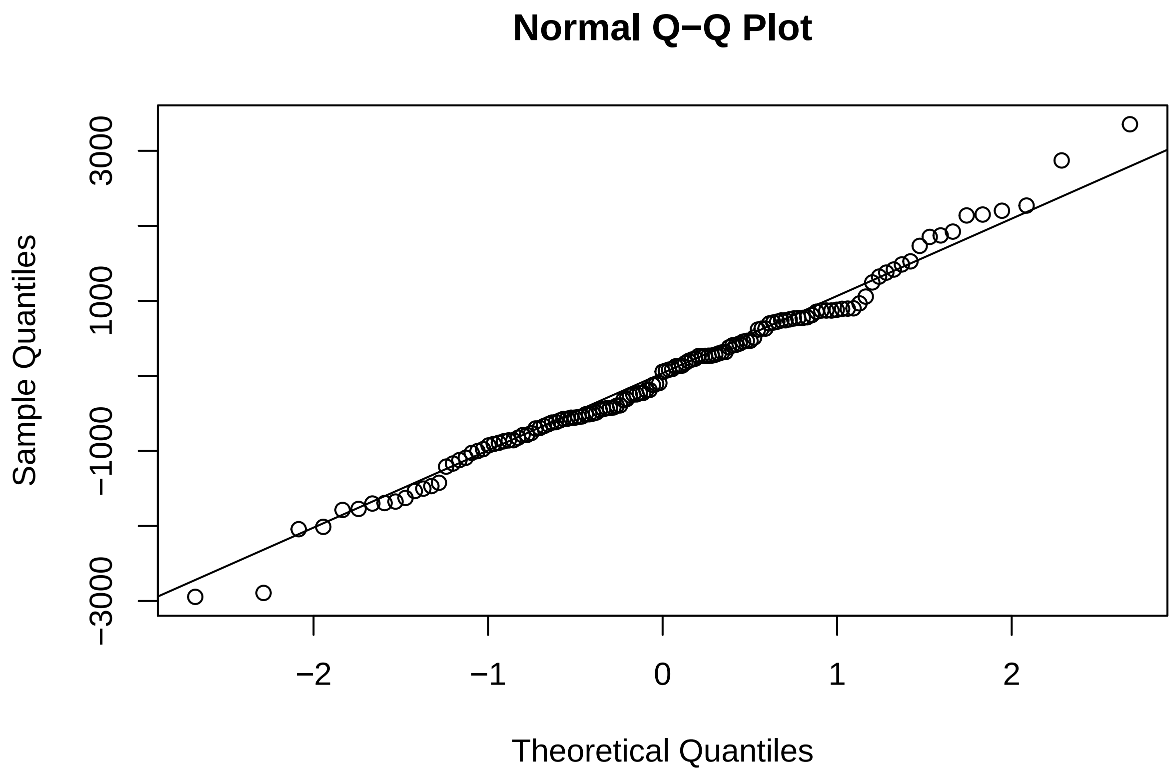

```
qqnorm(res2);qqline(res2)
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−3000
−1000
0
1000
3000
−2
−1
0
1
2

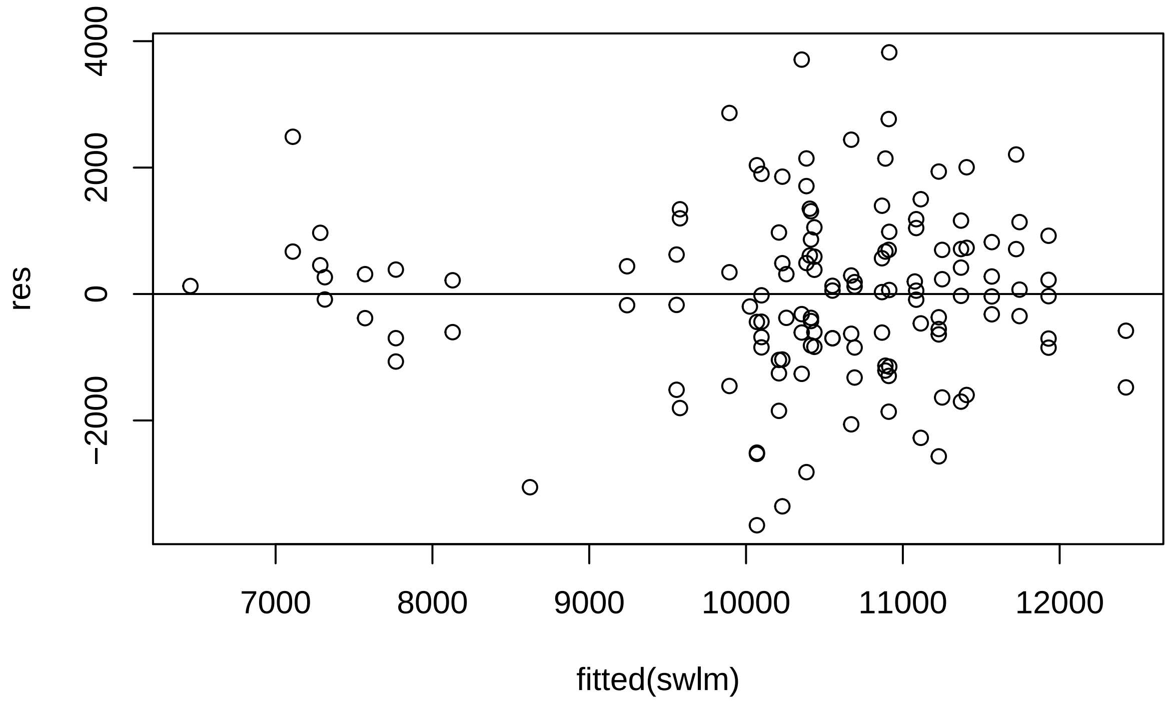
```r
hist(res3)
```



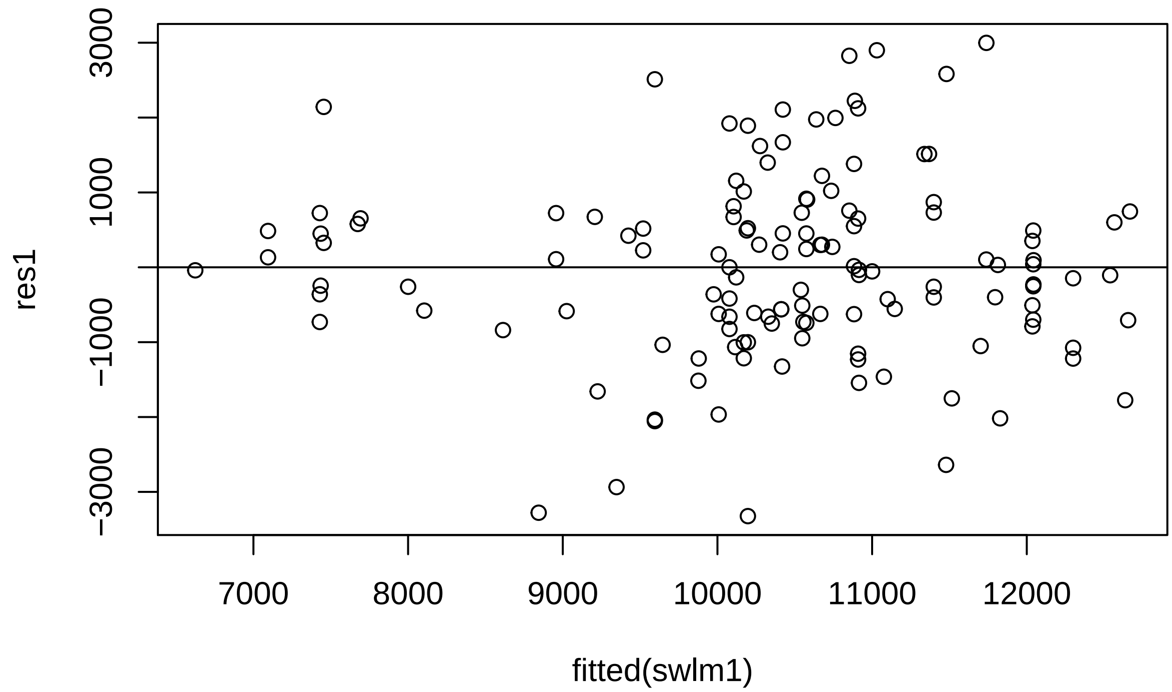

```r
qqnorm(res3);qqline(res3)
```



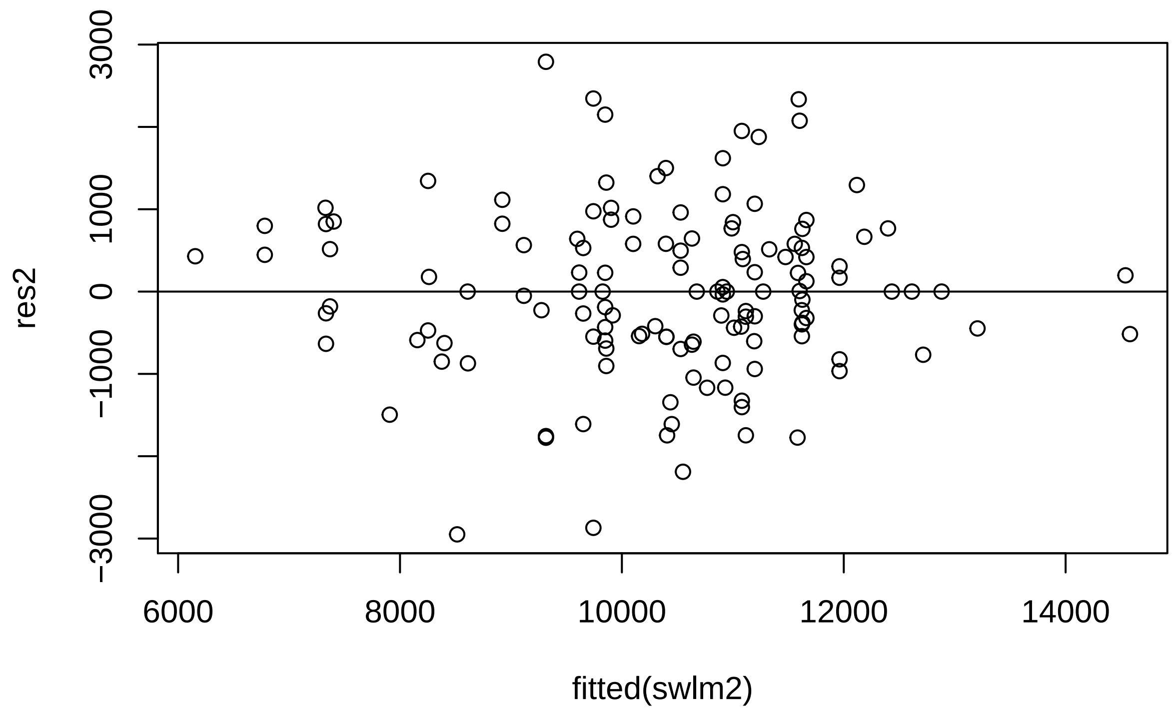

```r
#Residuals are normal across the board.

plot(res~fitted(swlm));abline(h=0)
```



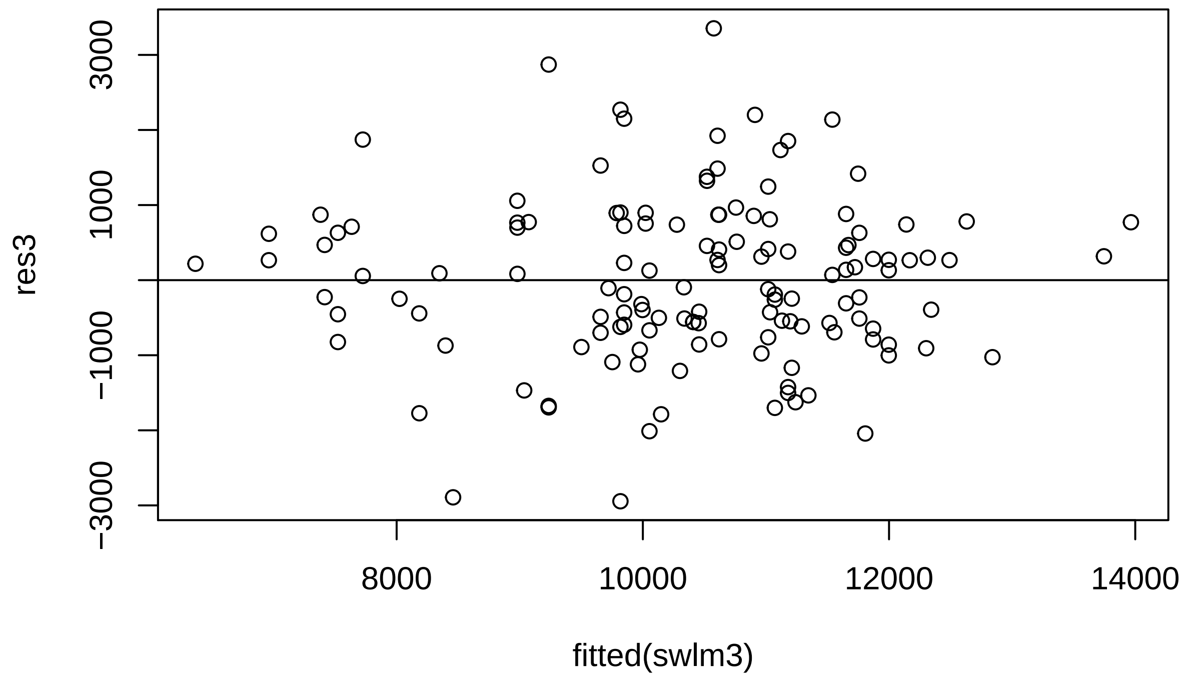

```
plot(res1~fitted(swlm1));abline(h=0)
```



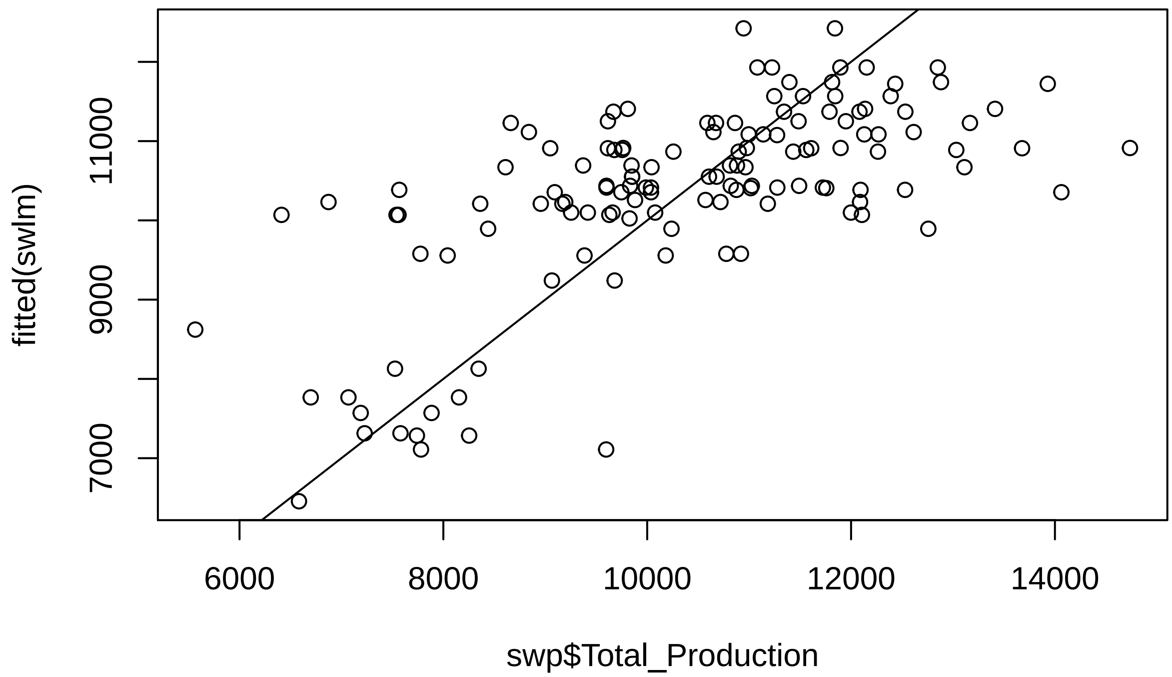

```
plot(res2~fitted(swlm2));abline(h=0)
```



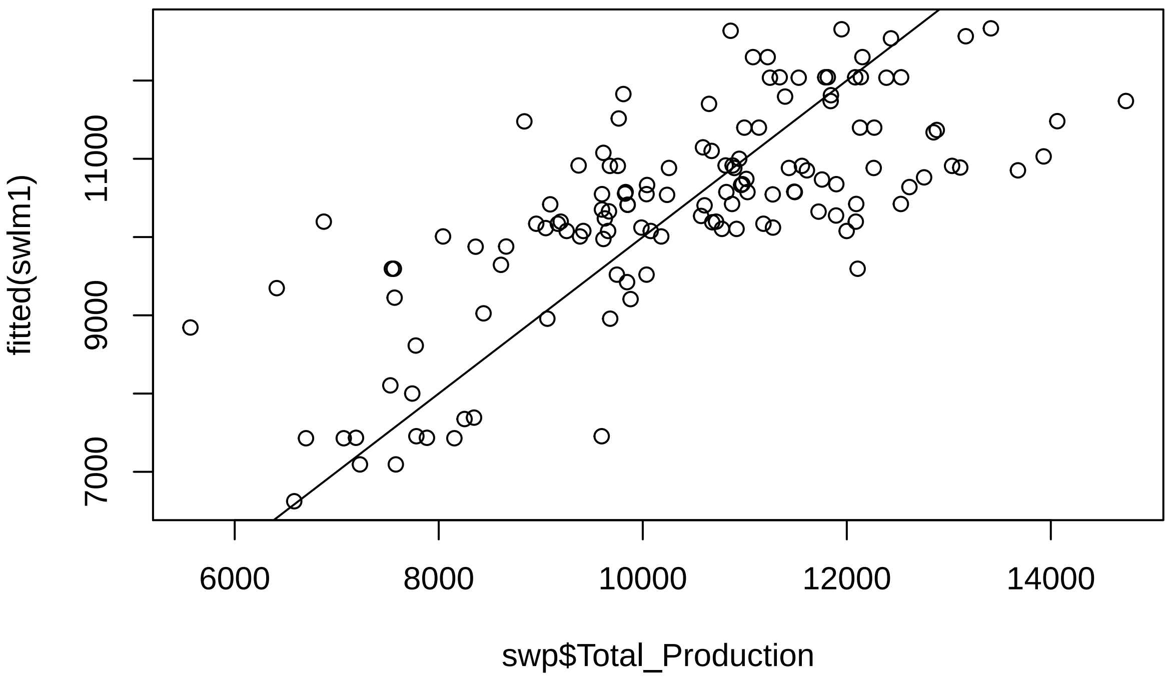

```
plot(res3~fitted(swlm3));abline(h=0)
```



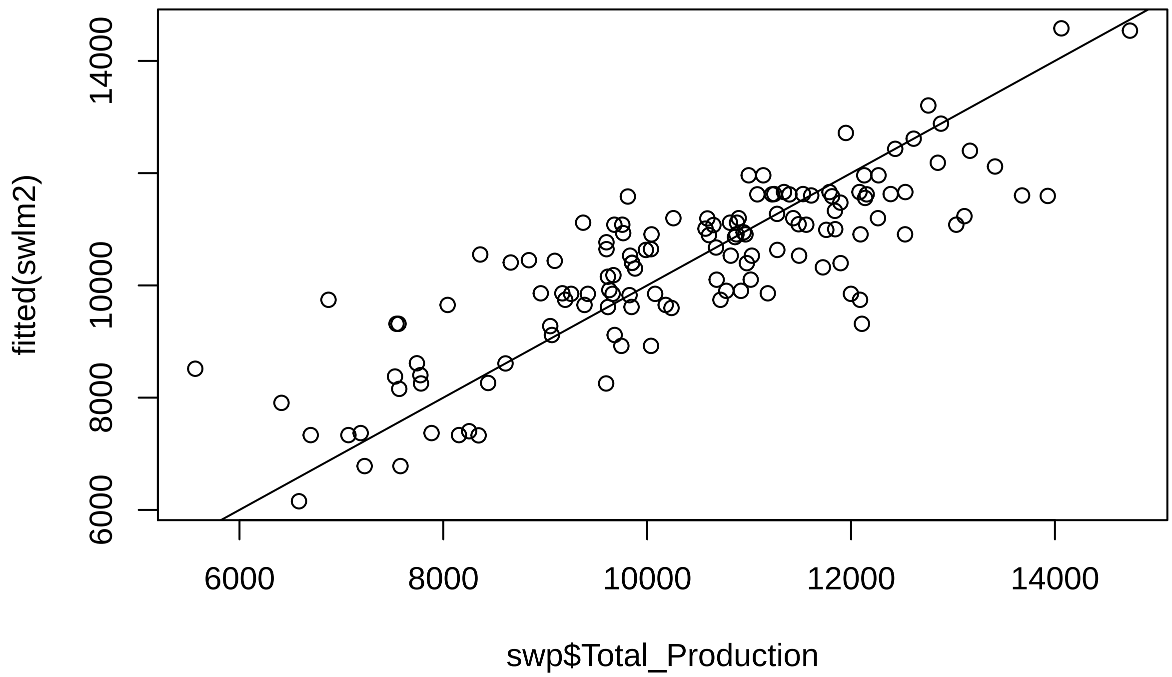

```r
#There is heteroscedasticity.
#There is definatly a trend.
#For some the middle the variance increases.
#it starts low gets high and then gets low.
#For some there is an increase in variance
#with an increase in level.

#This does not effect estimates
#but it does effect estimated variances (and hence p-values.)
#Our p-values may be lower than they should be.

plot(fitted(swlm)~swp$Total_Production);abline(c(0,1))
```



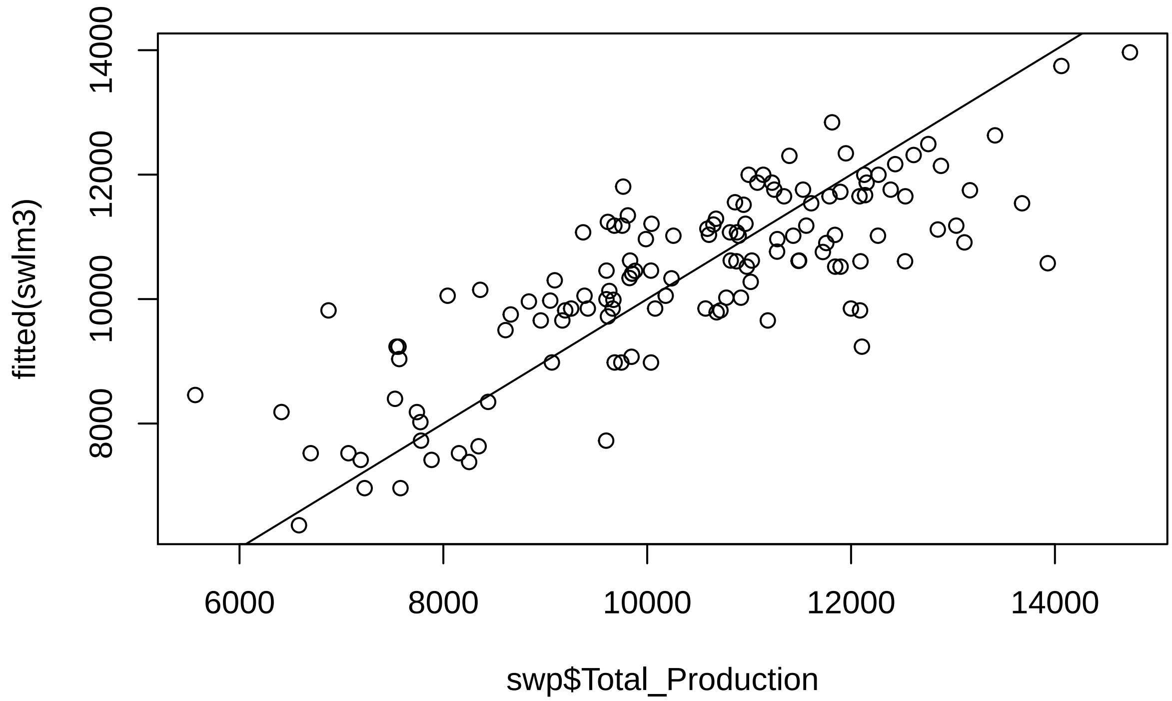

```
plot(fitted(swlm1)~swp$Total_Production);abline(c(0,1))
```



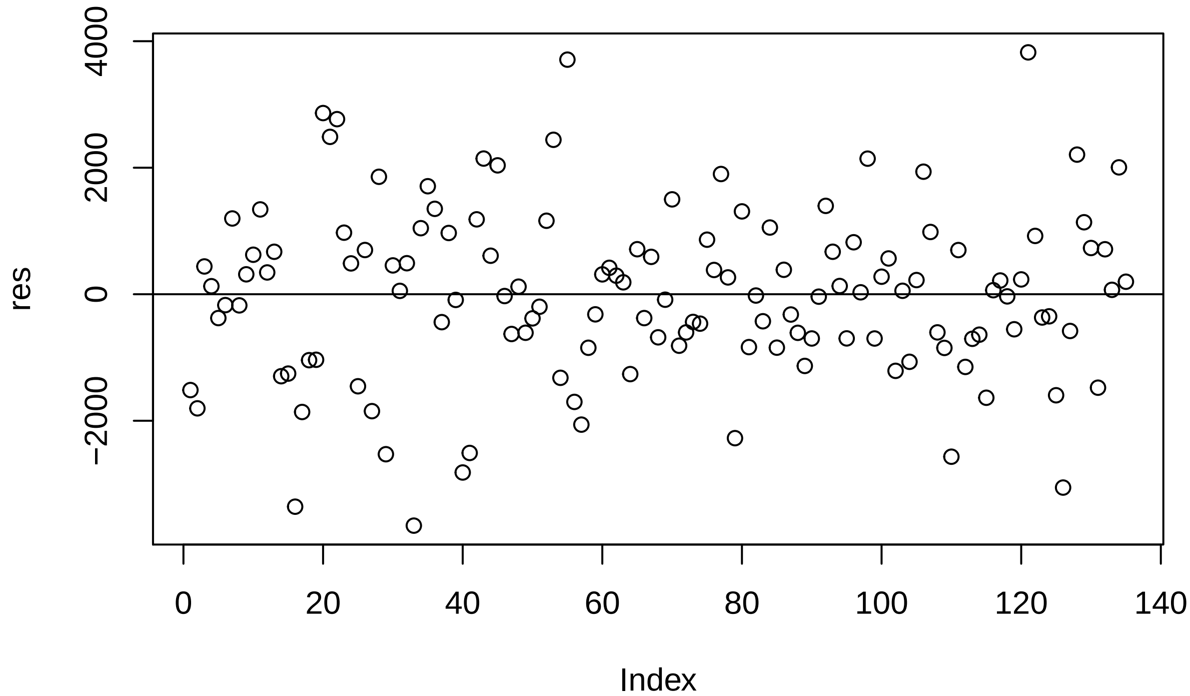

```
plot(fitted(swlm2)~swp$Total_Production);abline(c(0,1))
```



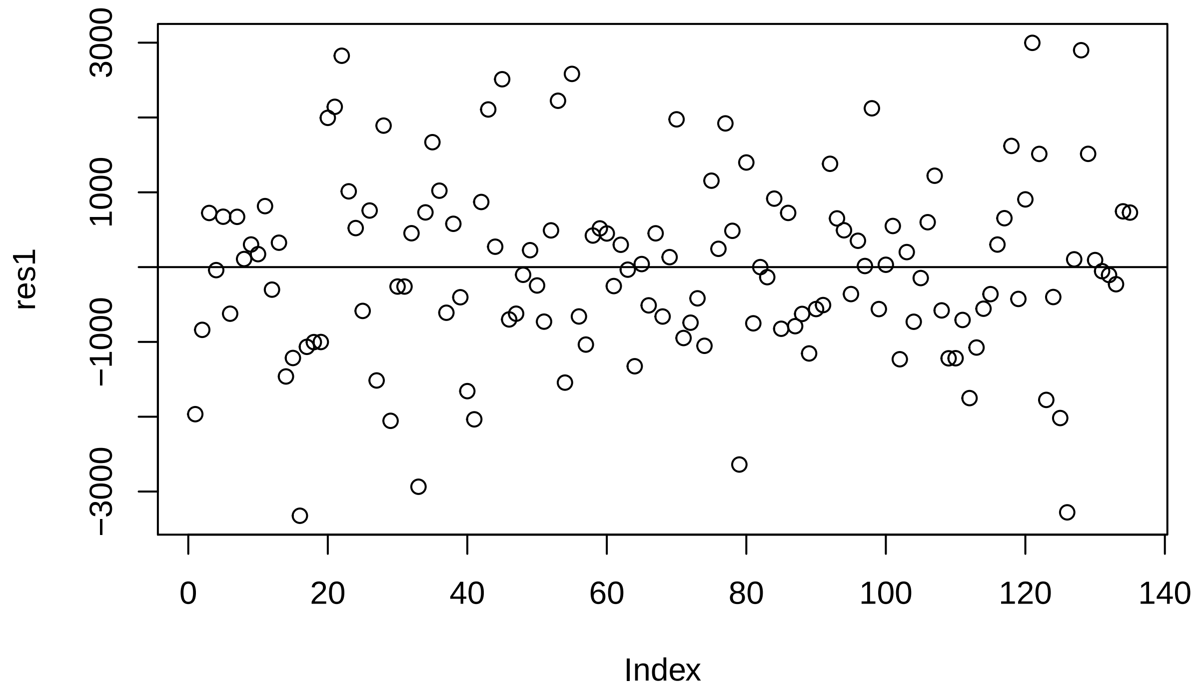

```
plot(fitted(swlm3)~swp$Total_Production);abline(c(0,1))
```



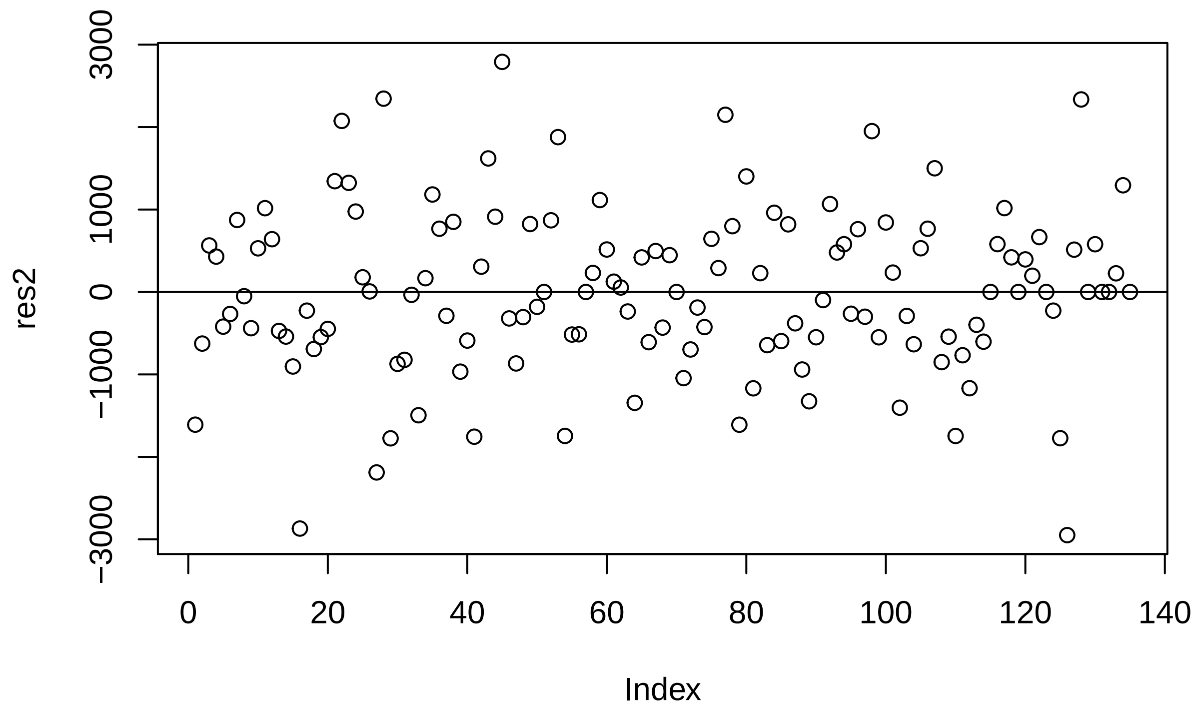

```r
#Actual vs fitted looks good, on the whole.
#there is some slight non-linearity.
#but it is small.

plot(res);abline(h=0)
```



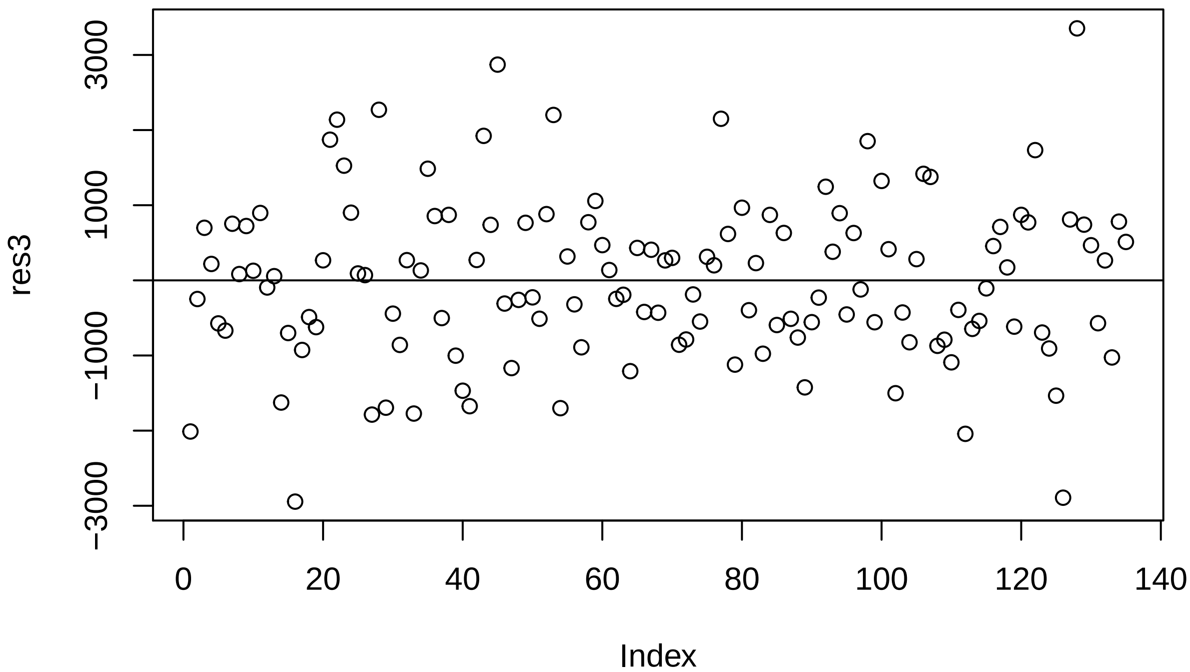

```
plot(res1);abline(h=0)
```



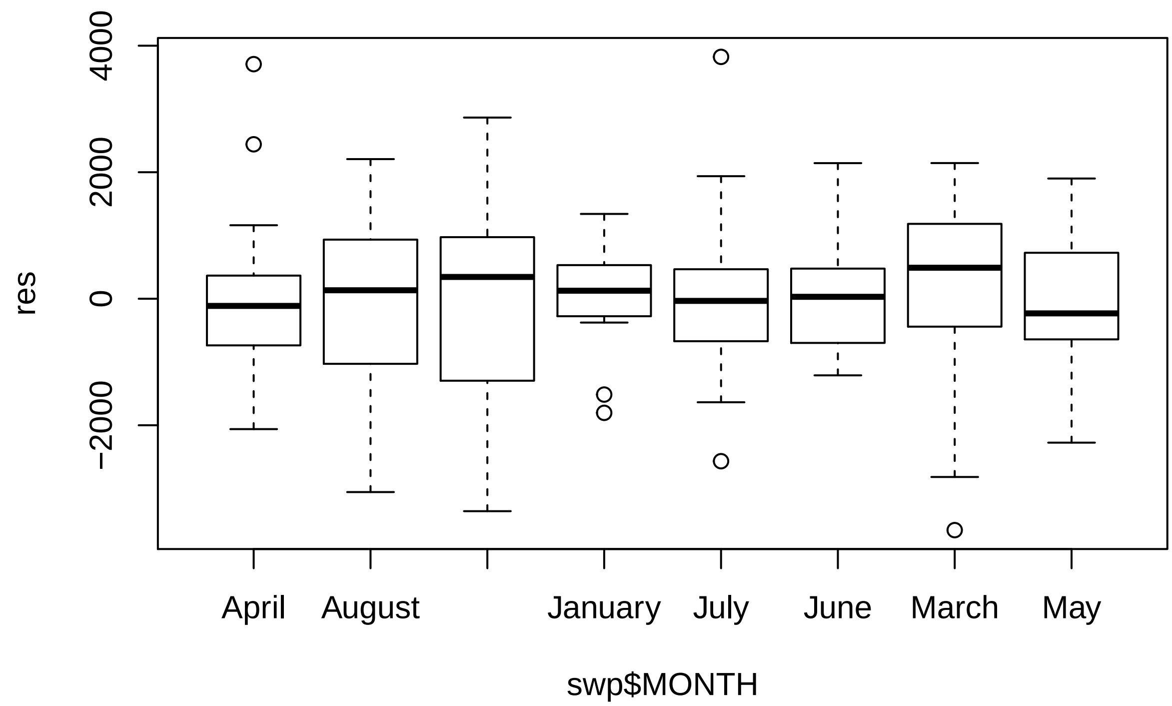

```
plot(res2);abline(h=0)
```



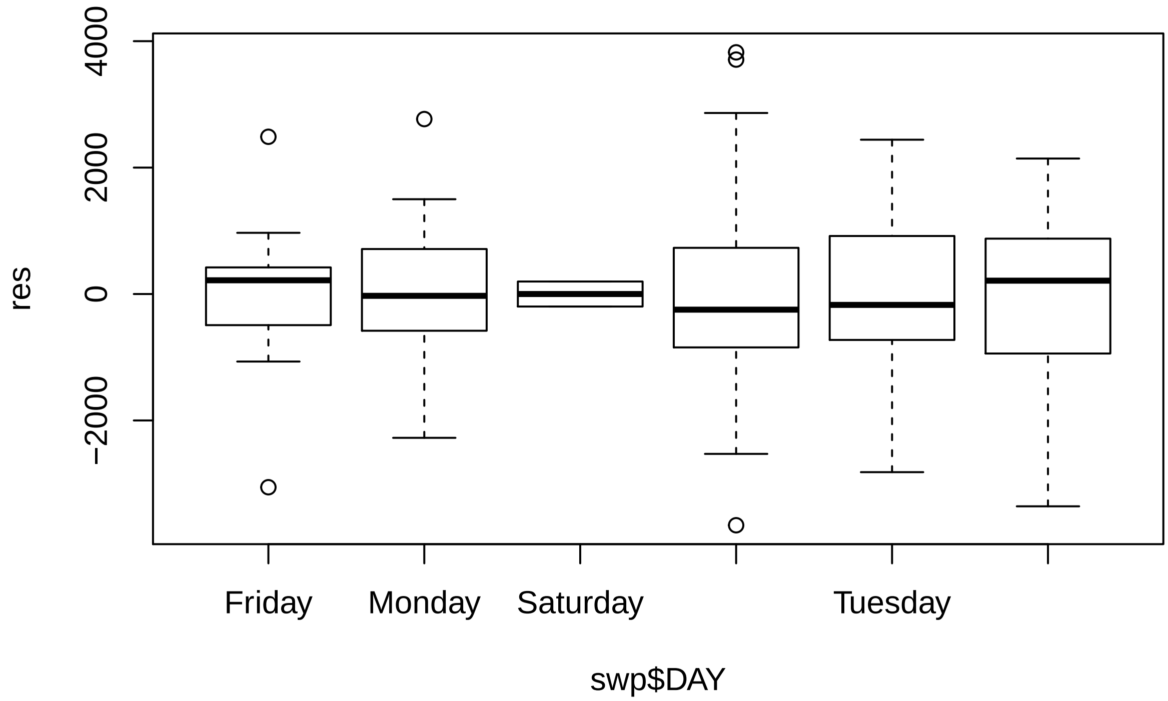

```r
plot(res3);abline(h=0)
```



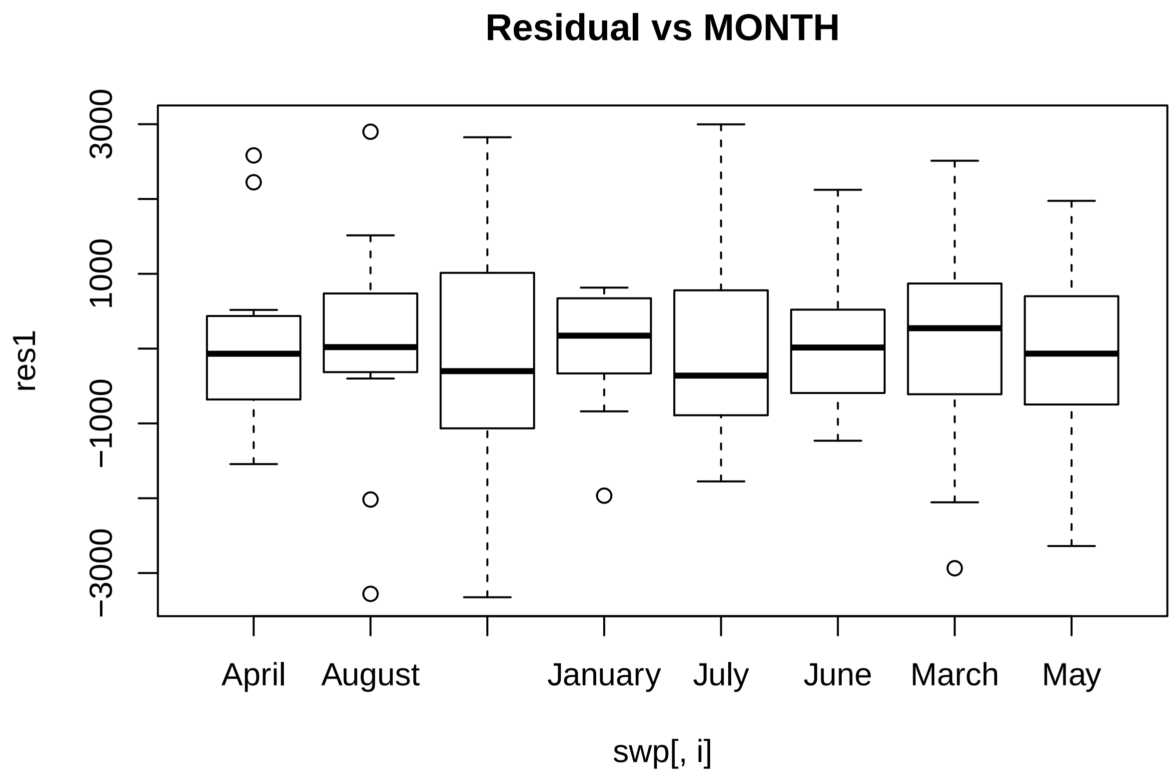

```r
#Not serially correlated.

#Lets check residuals vs each factor.

plot(res~swp$MONTH)
```

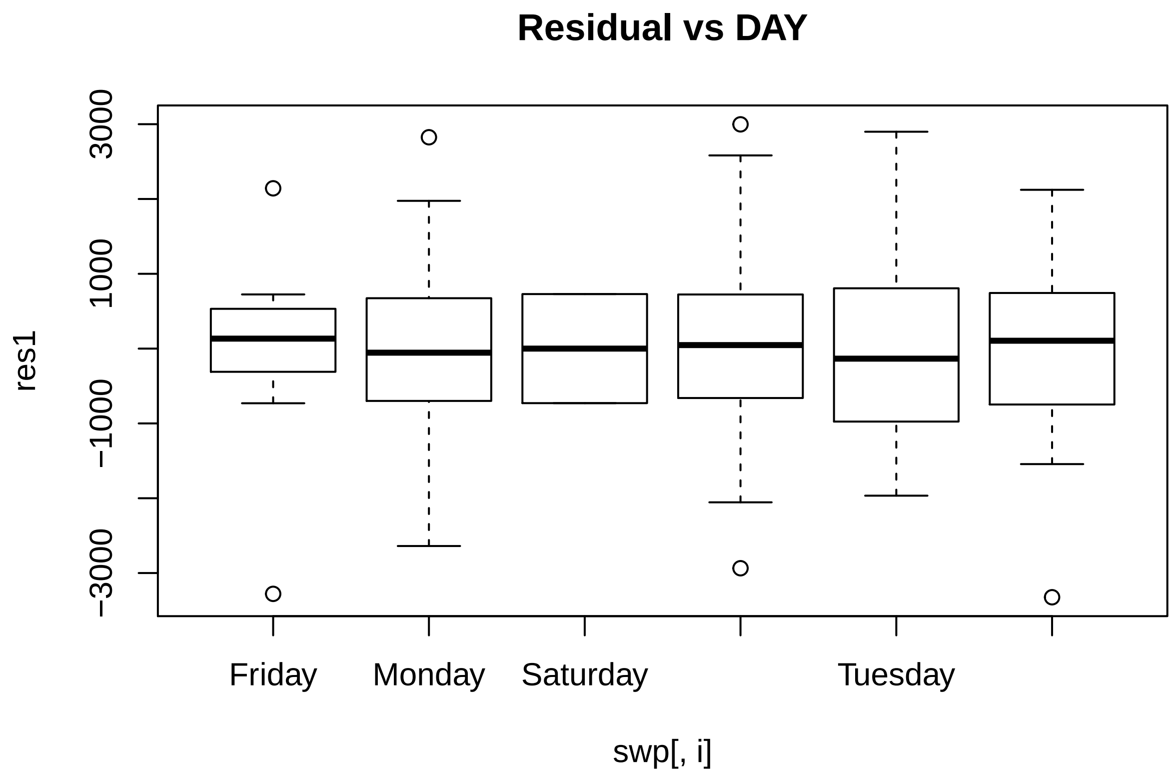



```
plot(res~swp$DAY)
```



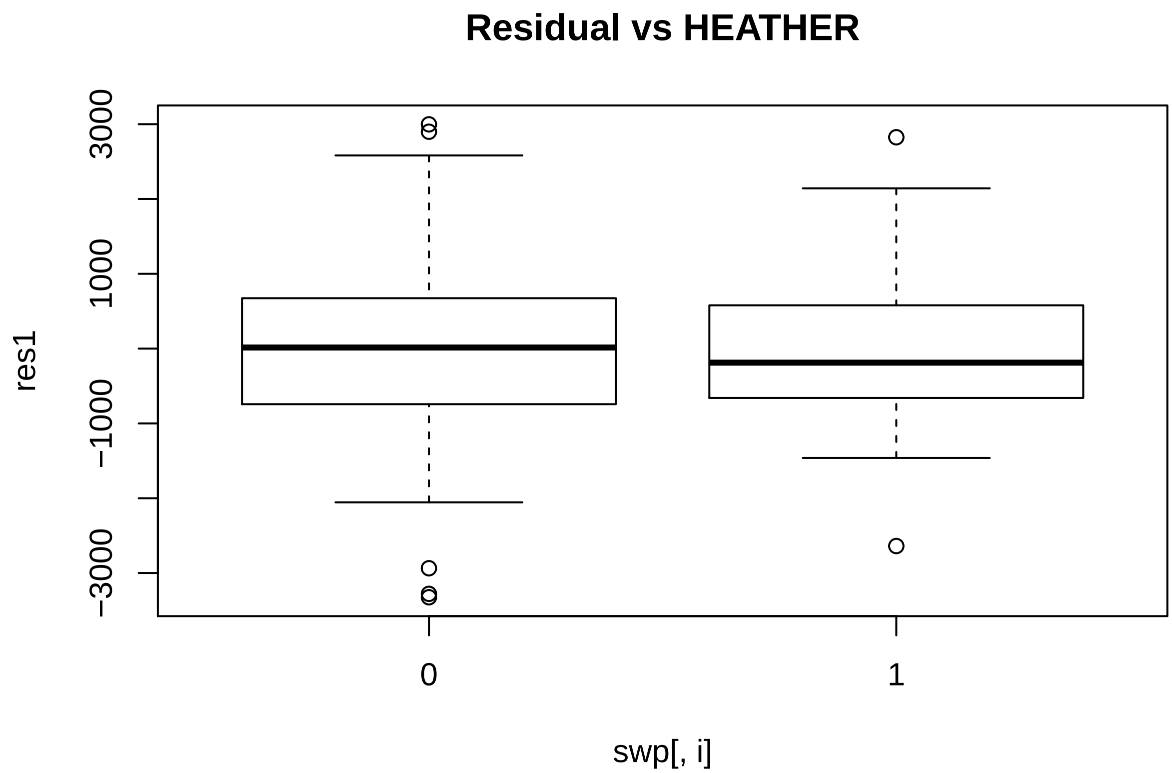

```
#Yep, heteroscedasticity.

for(i in attr(swlm1$terms,"term.labels")[1:11]){
  plot(res1~swp[,i],main=paste0("Residual vs ",i))
  readline("Press enter to continue.")
}
```

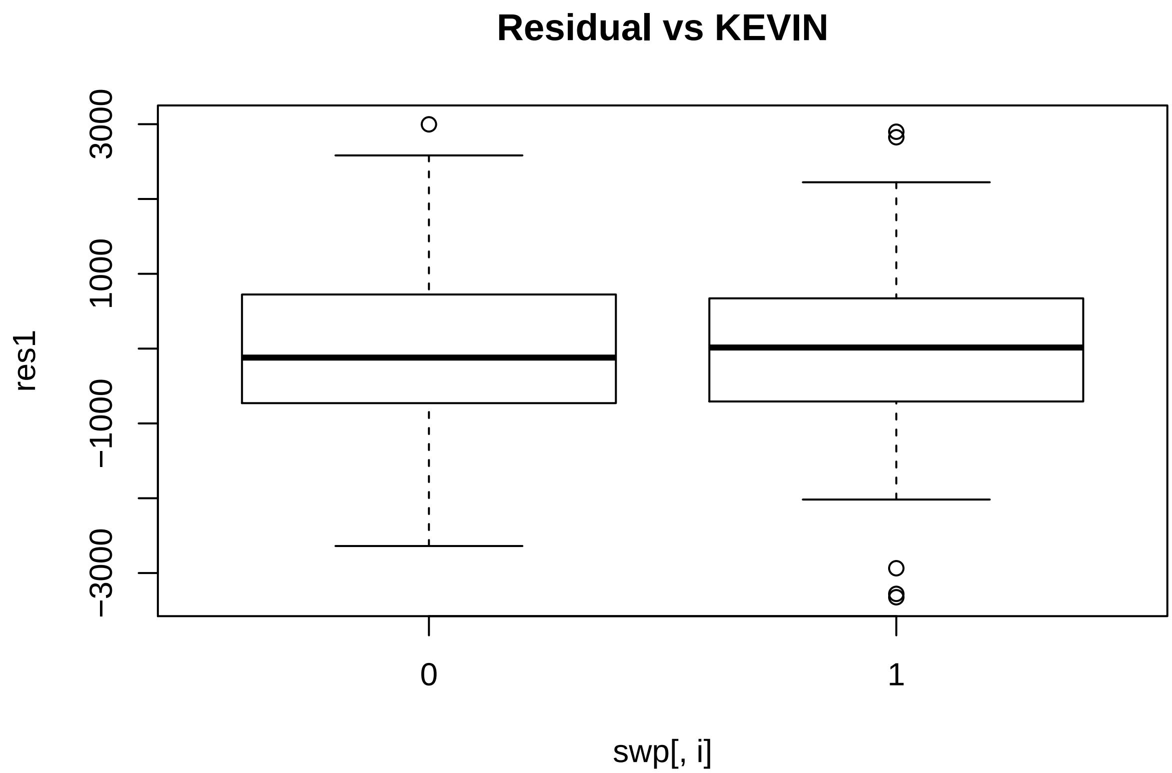

Residual vs MONTH
res1
3000
1000
−1000
−3000
April
August
January
July
June
March
May
swp[, i]



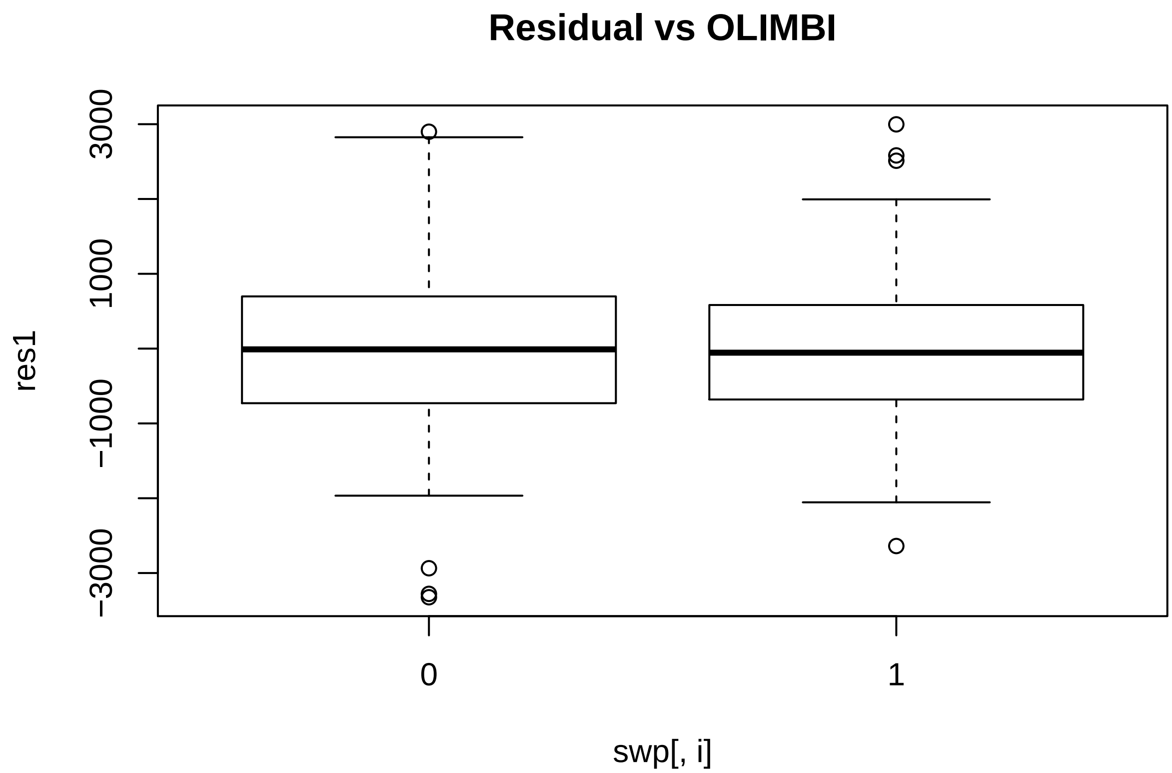
Residual vs DAY
3000
1000
−1000
−3000
res1
Friday
Monday
Saturday
Tuesday
swp[, i]

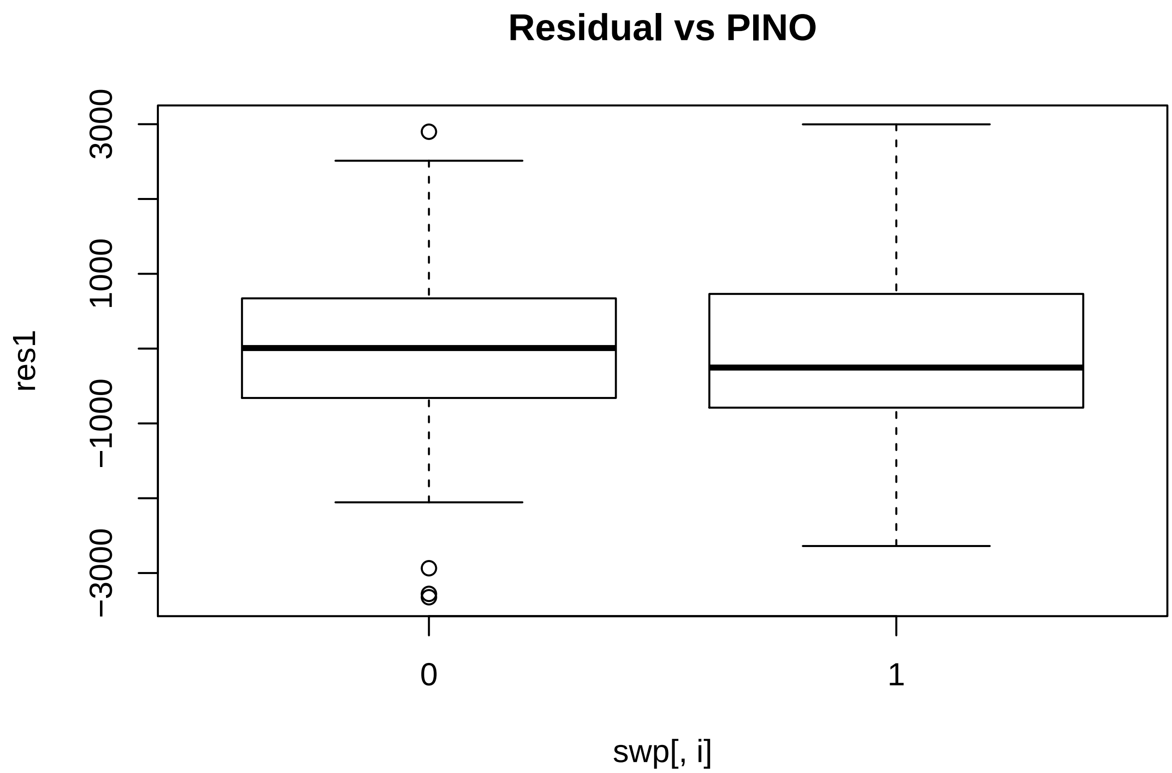
Residual vs BRITTANY
3000
1000
−1000
−3000
res1
0
1
swp[, i]

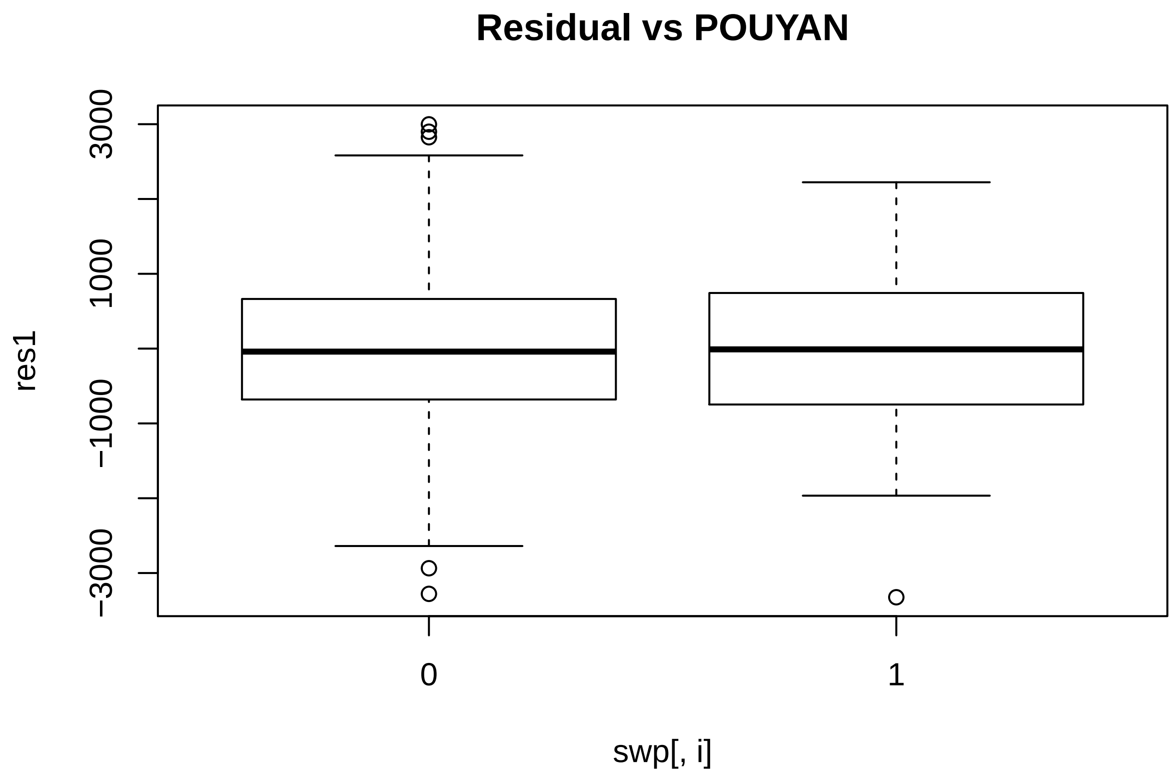

Residual vs HEATHER
res1
3000
1000
−1000
−3000
0
1
swp[, i]

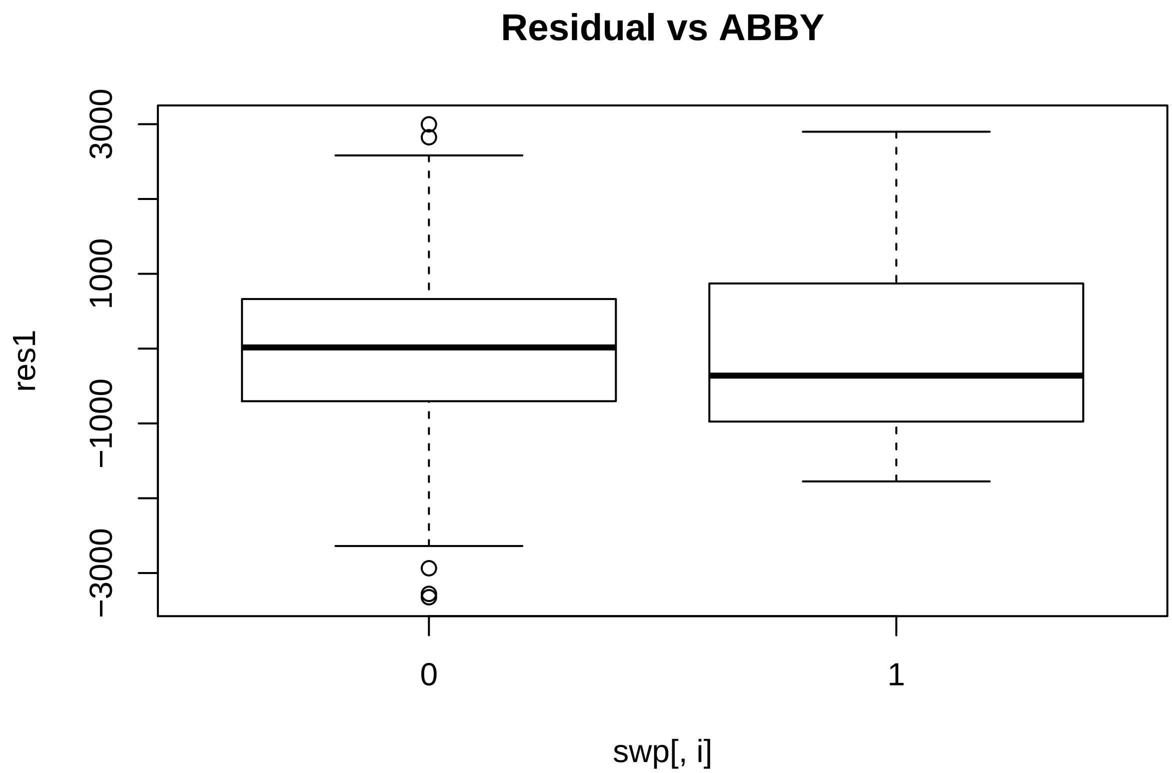


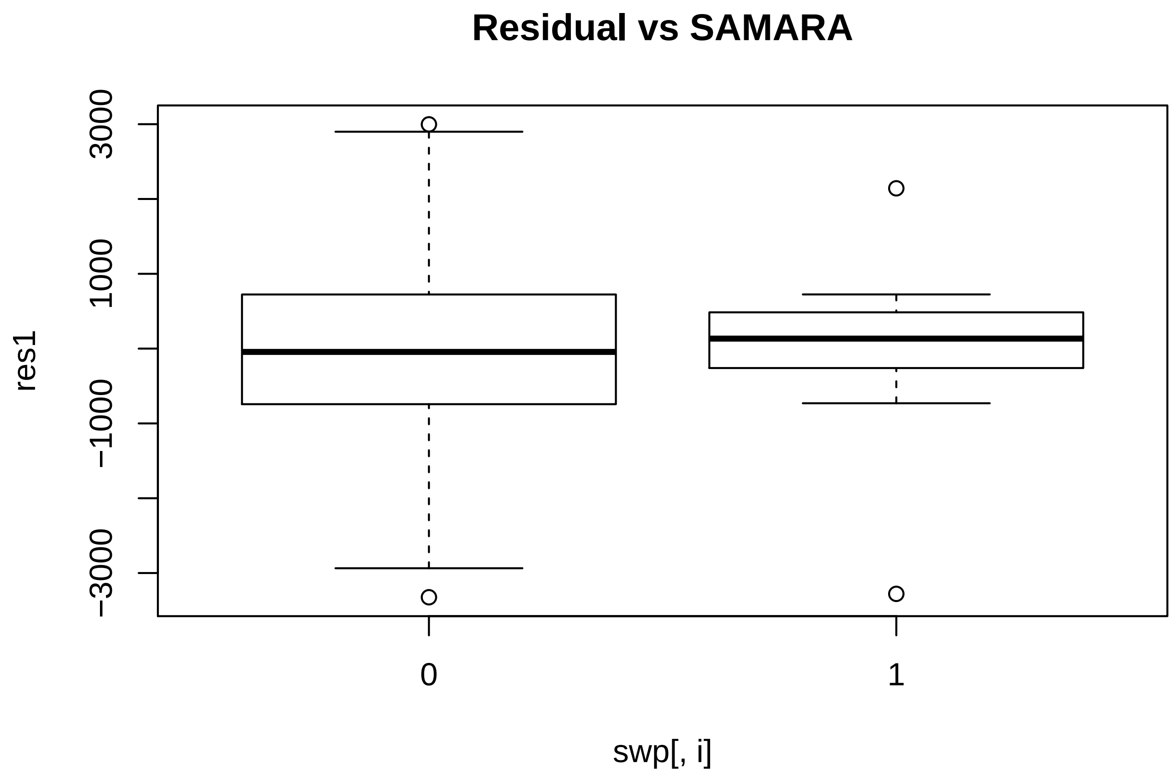
Residual vs KEVIN
3000
1000
−1000
−3000
res1
0
1
swp[, i]

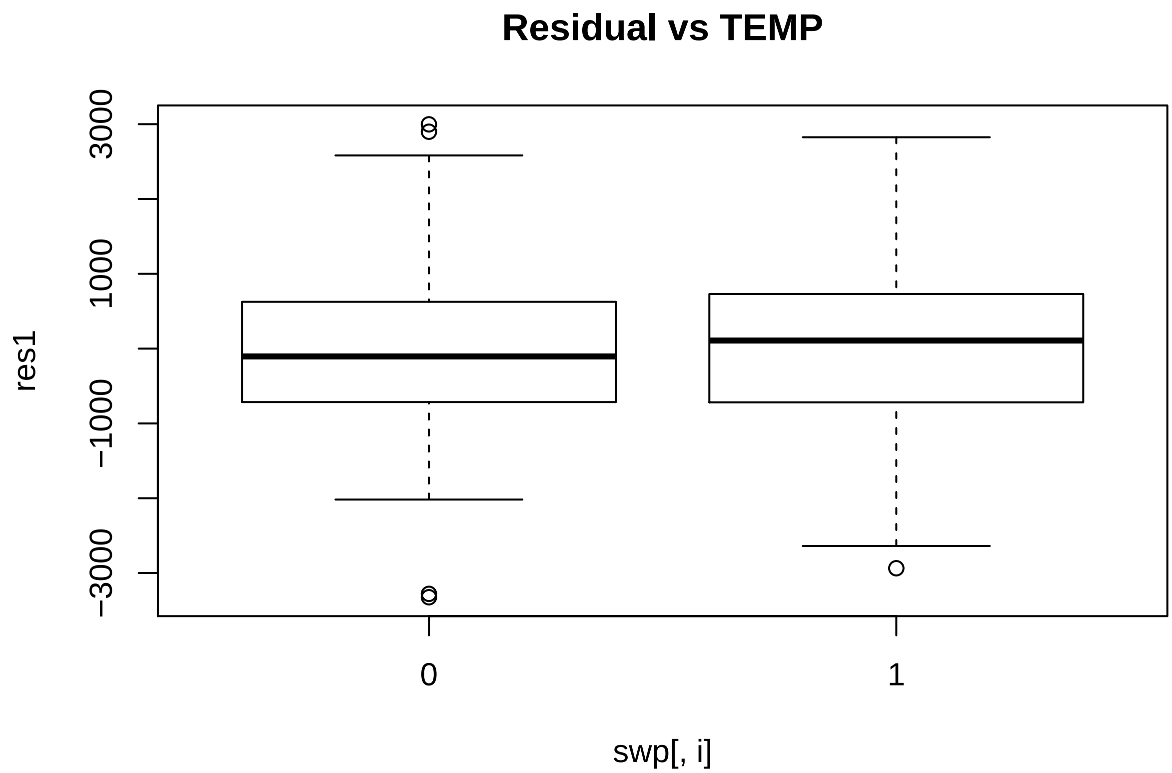

Residual vs OLIMBI
res1
3000
1000
−1000
−3000
0
1
swp[, i]


Residual vs PINO
res1
3000
1000
−1000
−3000
0
1
swp[, i]

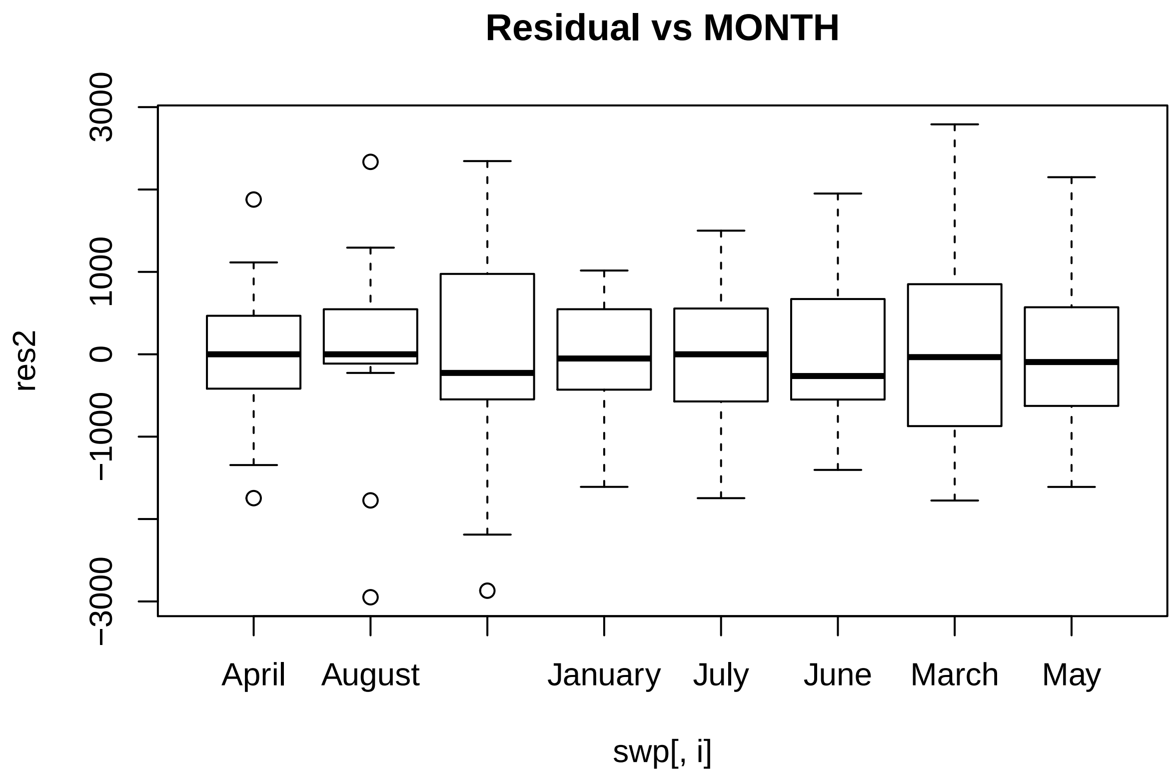

Residual vs POUYAN
res1
3000
1000
−1000
−3000
0
1
swp[, i]


Residual vs ABBY
res1
3000
1000
−1000
−3000
0
1
swp[, i]

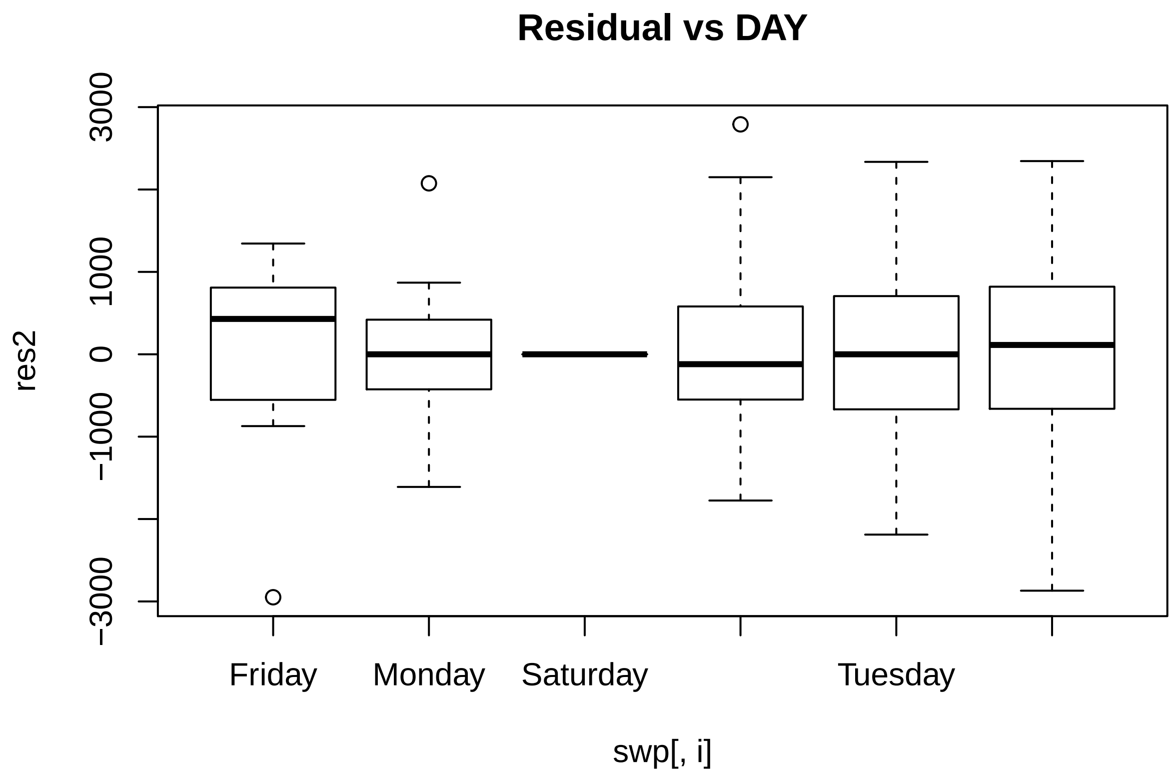

Residual vs SAMARA
res1
3000
1000
−1000
−3000
0
1
swp[, i]



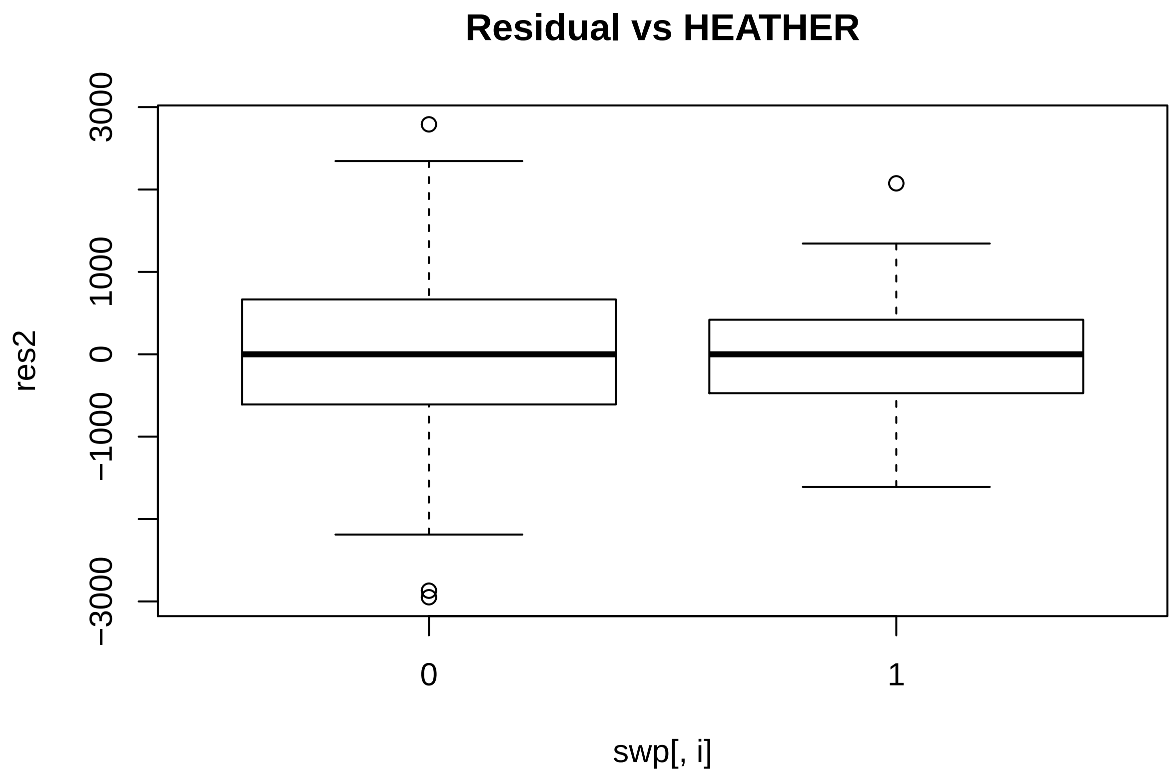

```
#Yep, heteroscedasticity.

for(i in attr(swlm2$terms,"term.labels")[1:11]){
  plot(res2~swp[,i],main=paste0("Residual vs ",i))
  readline("Press enter to continue.")
}
```

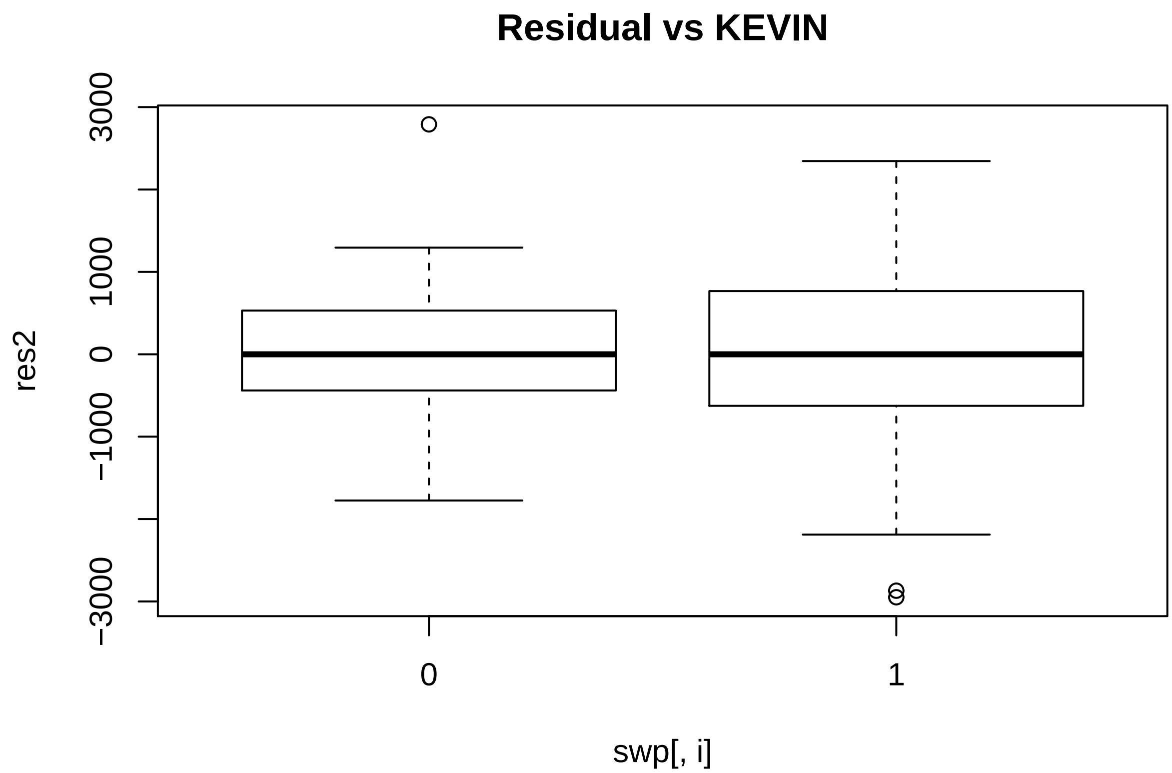


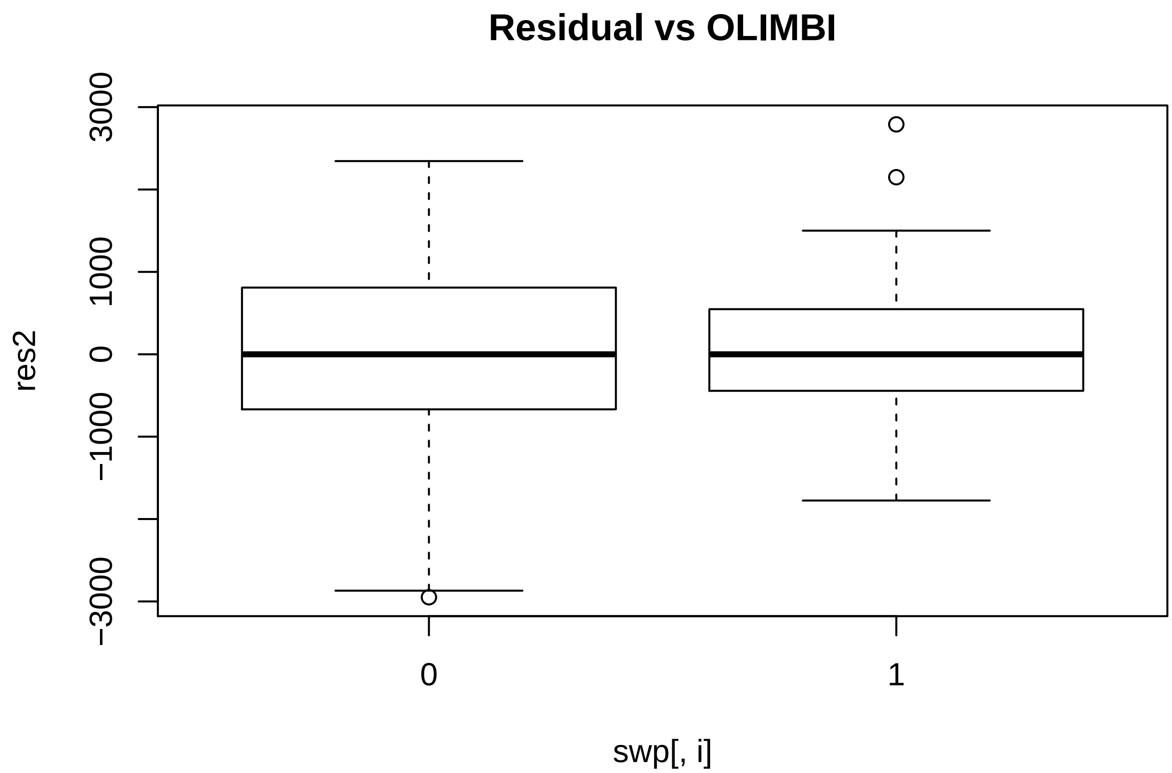
Residual vs MONTH
res2
3000
1000
0
−1000
−3000
April
August
January
July
June
March
May
swp[, i]


Residual vs DAY
3000
1000
0
−1000
−3000
res2
Friday
Monday
Saturday
Tuesday
swp[, i]

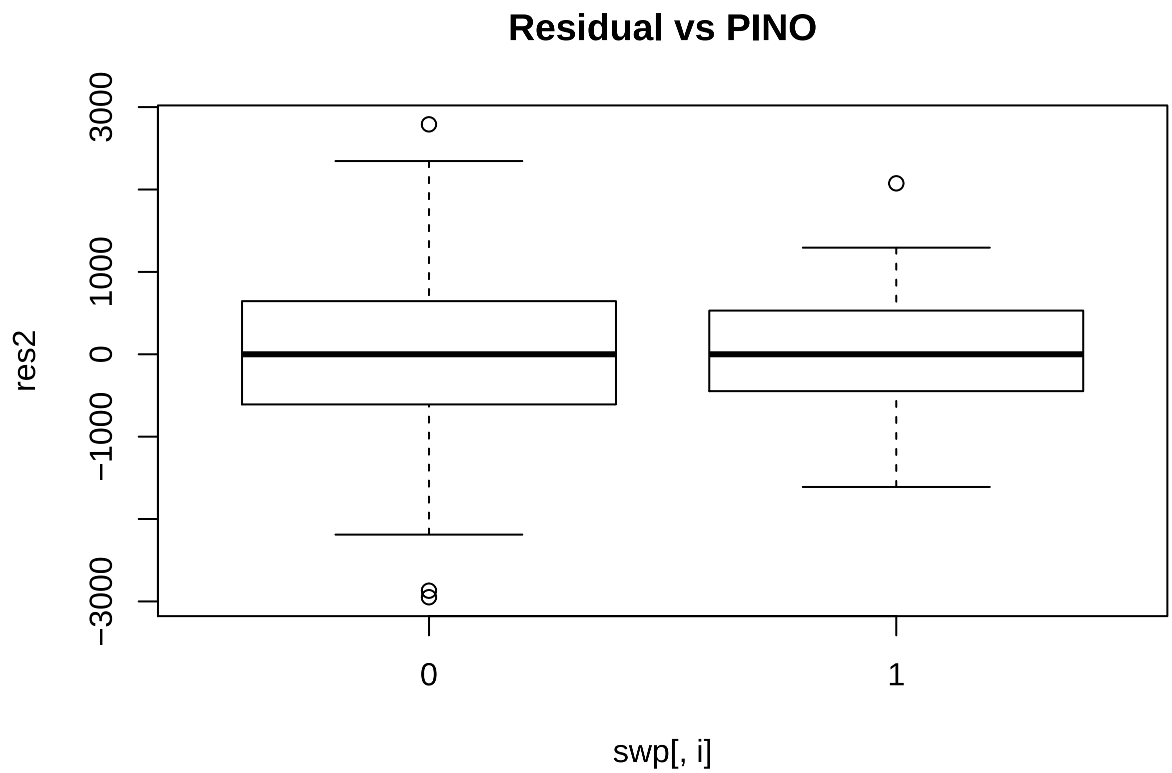

Residual vs BRITTANY
res2
3000
1000
0
−1000
−3000
0
1
swp[, i]

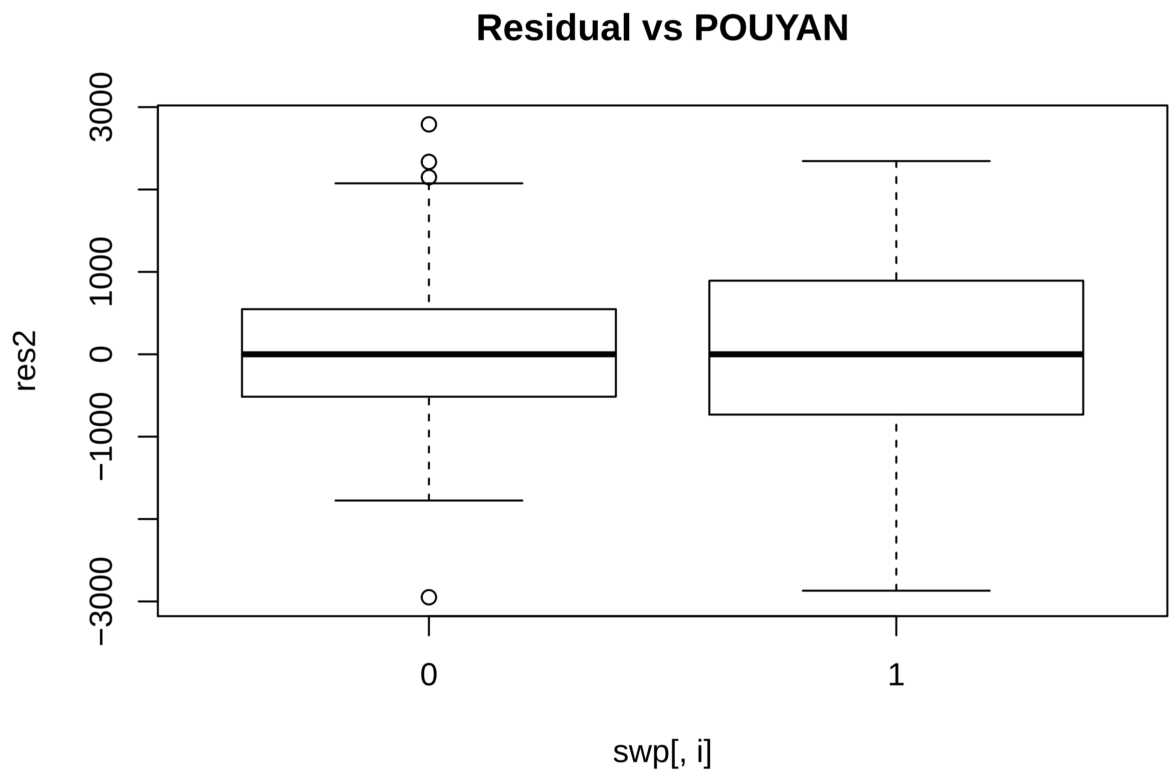

Residual vs HEATHER
res2
3000
1000
0
−1000
−3000
0
1
swp[, i]

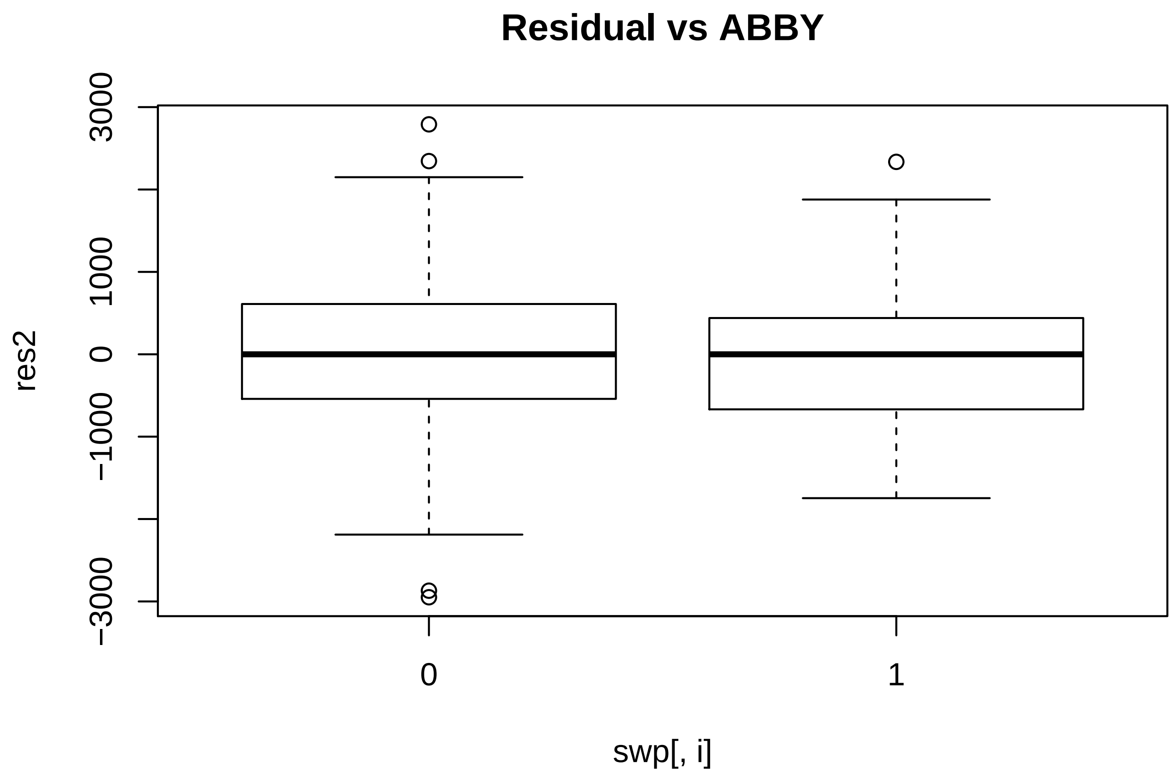

Residual vs KEVIN
res2
3000
1000
0
−1000
−3000
0
1
swp[, i]

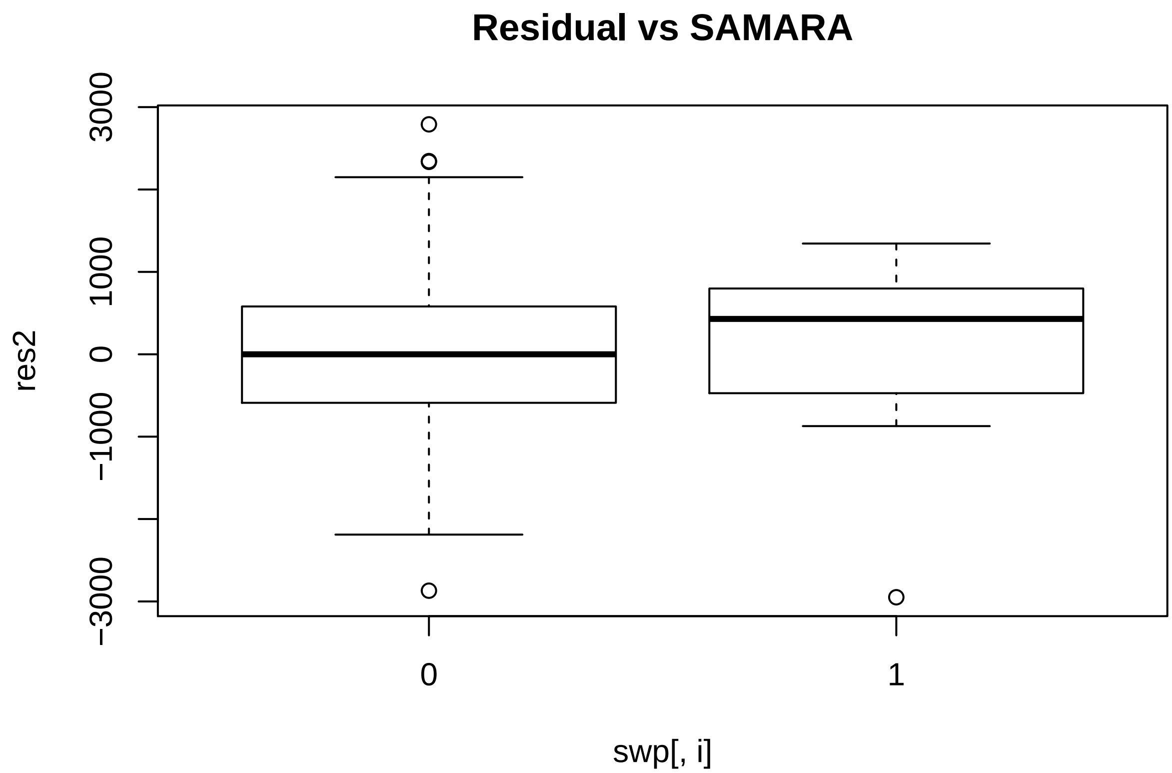


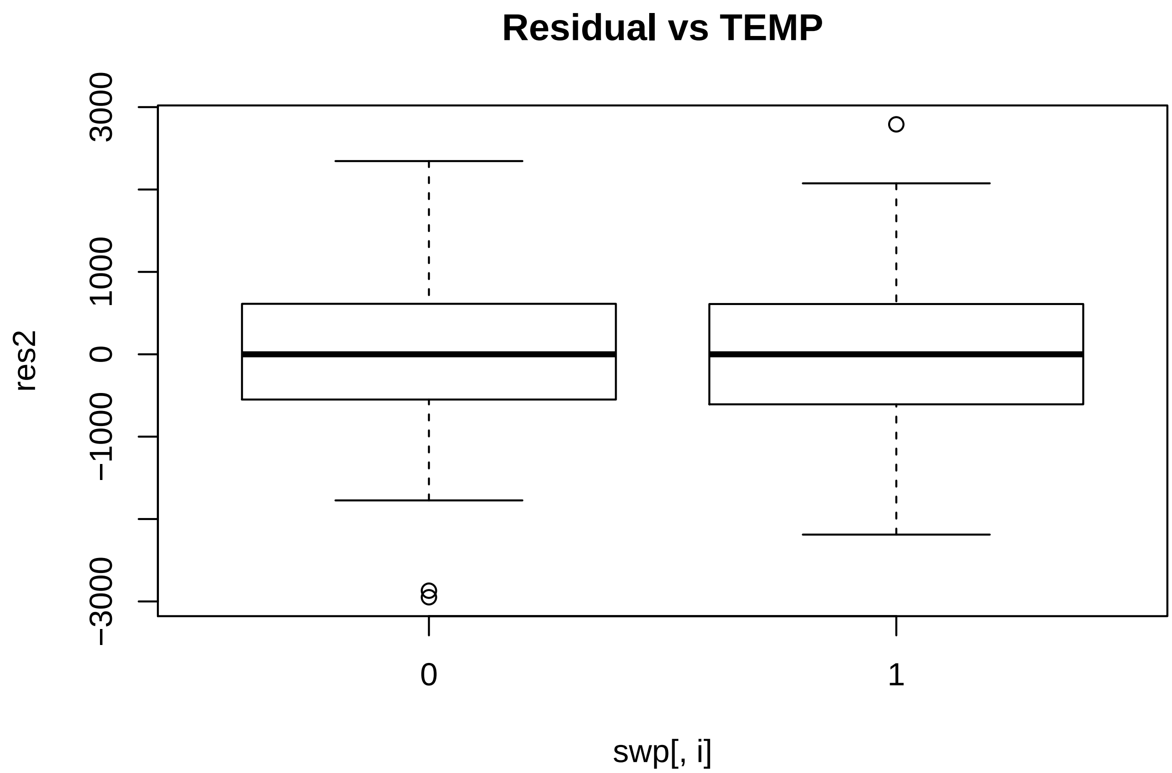
Residual vs OLIMBI
res2
3000
1000
0
−1000
−3000
0
1
swp[, i]


Residual vs PINO
res2
3000
1000
0
−1000
−3000
0
1
swp[, i]

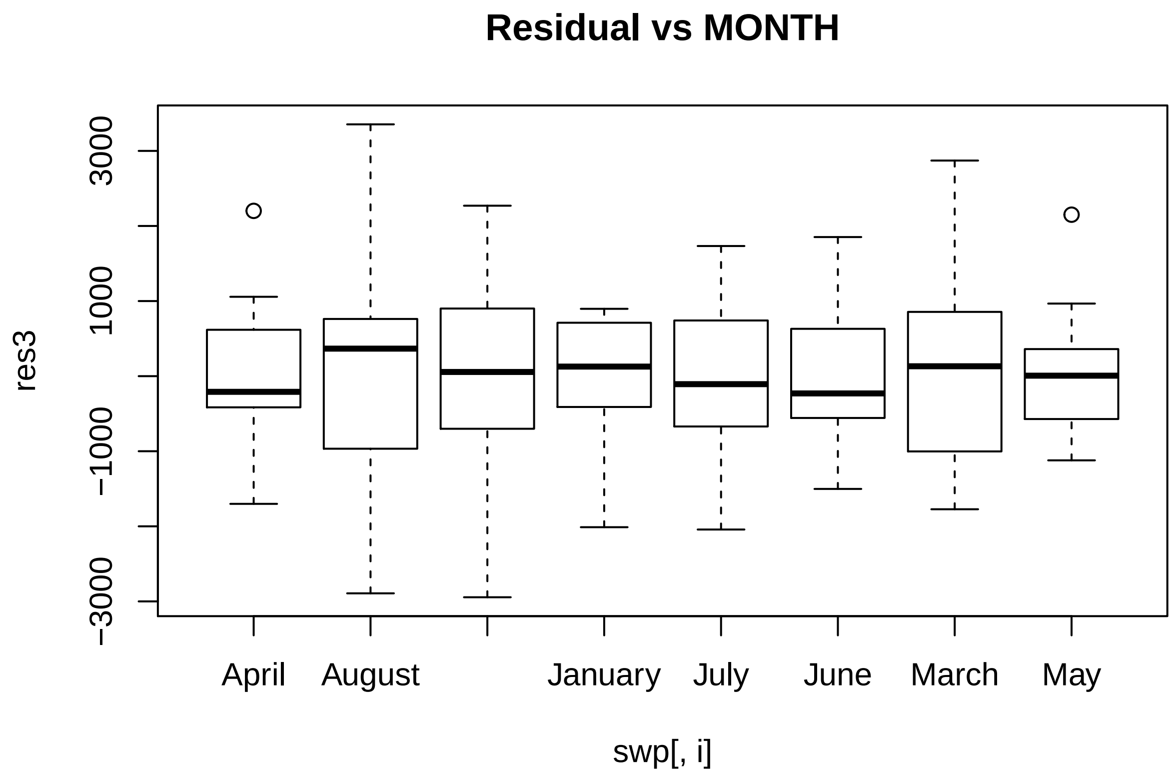

Residual vs POUYAN
res2
3000
1000
0
−1000
−3000
0
1
swp[, i]

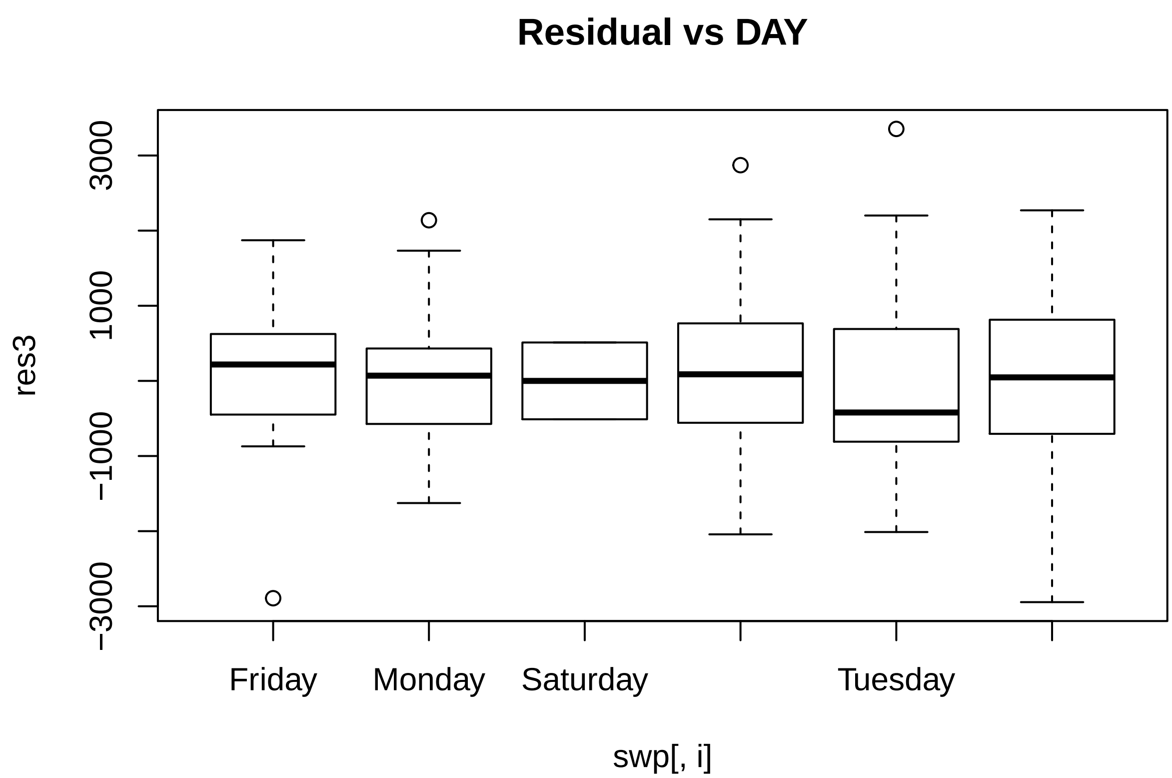

Residual vs ABBY
res2
3000
1000
0
−1000
−3000
0
1
swp[, i]


Residual vs SAMARA
res2
3000
1000
0
−1000
−3000
0
1
swp[, i]

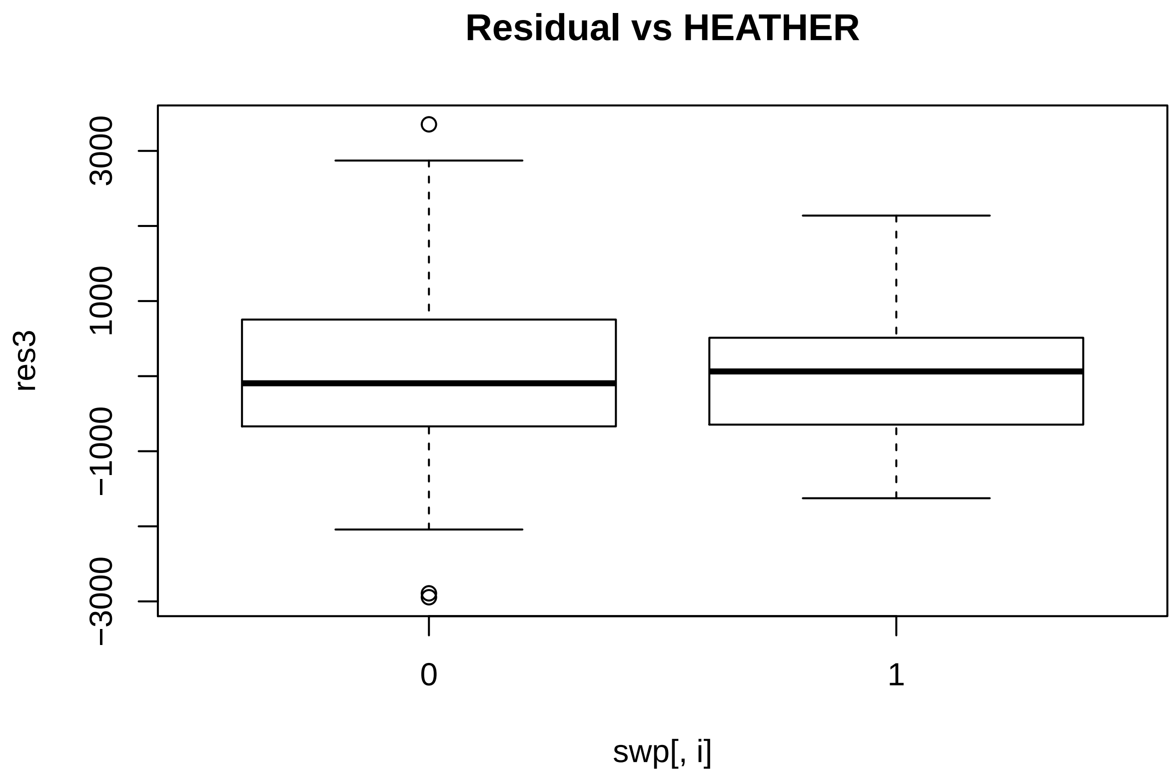


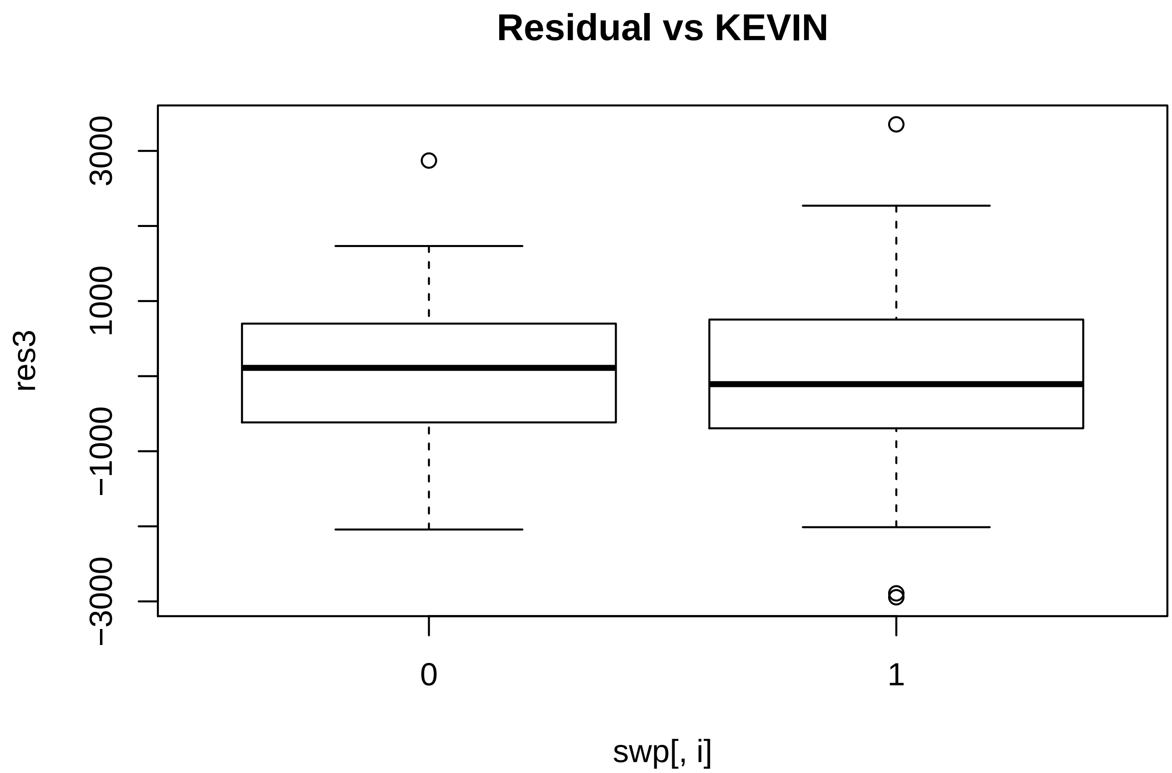

```
#Yep, heteroscedasticity.

for(i in attr(swlm3$terms,"term.labels")[1:11]){
  plot(res3~swp[,i],main=paste0("Residual vs ",i))
  readline("Press enter to continue.")
}
```


Residual vs MONTH
res3
3000
1000
−1000
−3000
April
August
January
July
June
March
May
swp[, i]

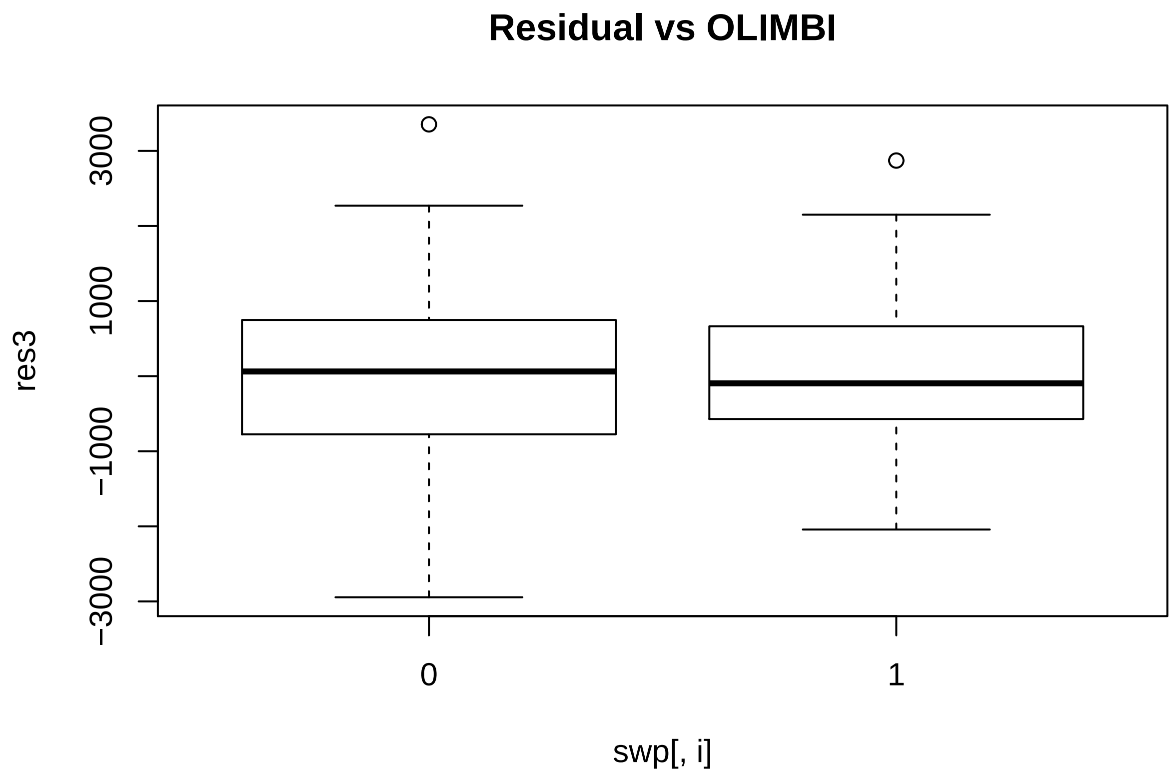

Residual vs DAY
3000
1000
−1000
−3000
res3
Friday
Monday
Saturday
Tuesday
swp[, i]

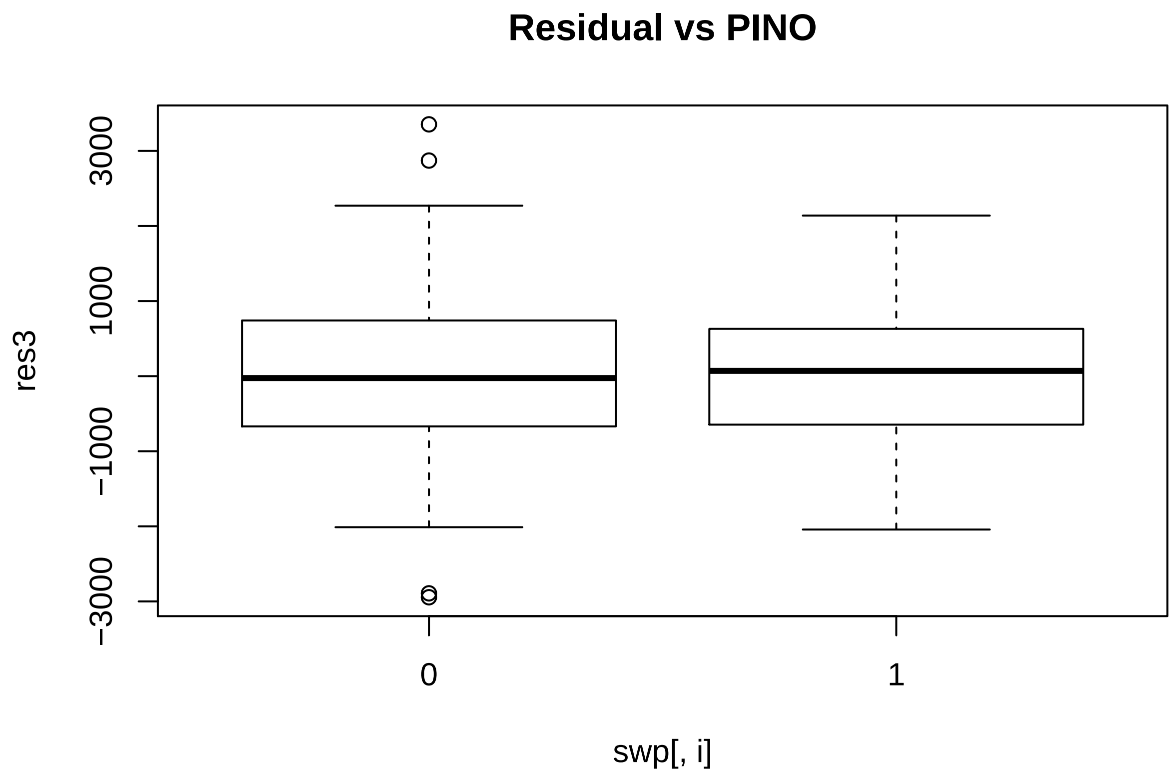

Residual vs BRITTANY
res3
3000
1000
−1000
−3000
0
1
swp[, i]

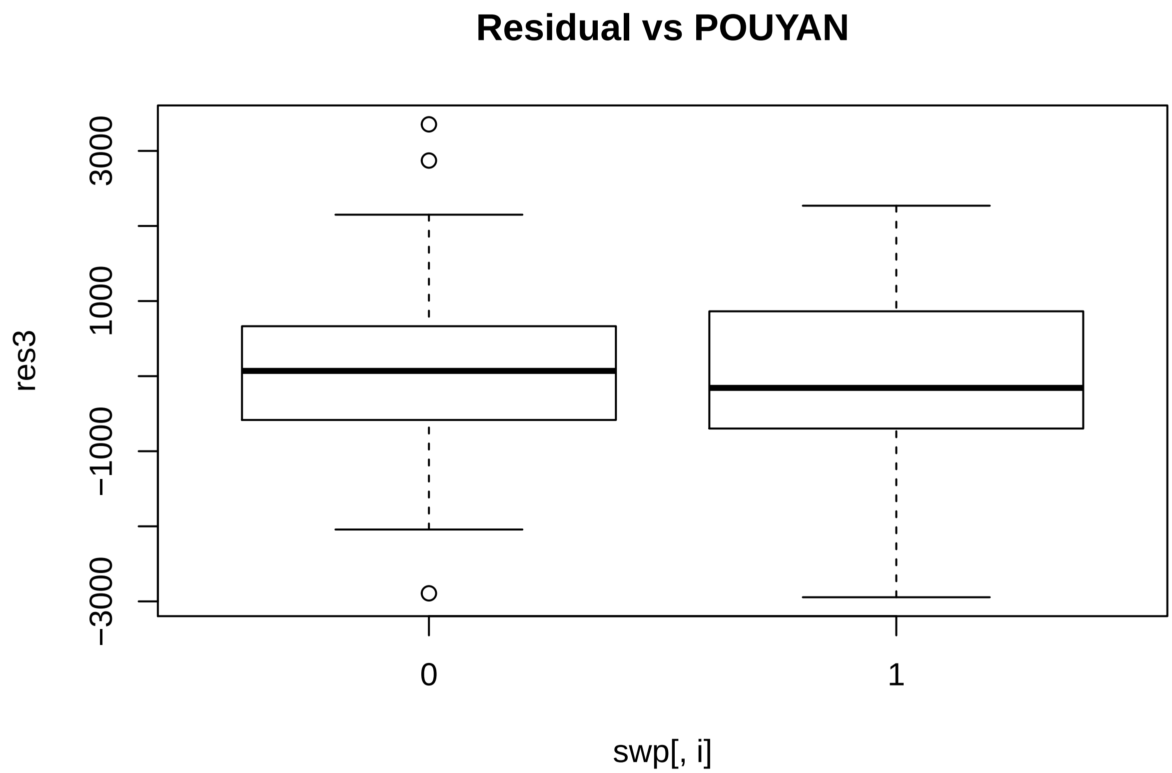

Residual vs HEATHER
res3
3000
1000
−1000
−3000
0
1
swp[, i]

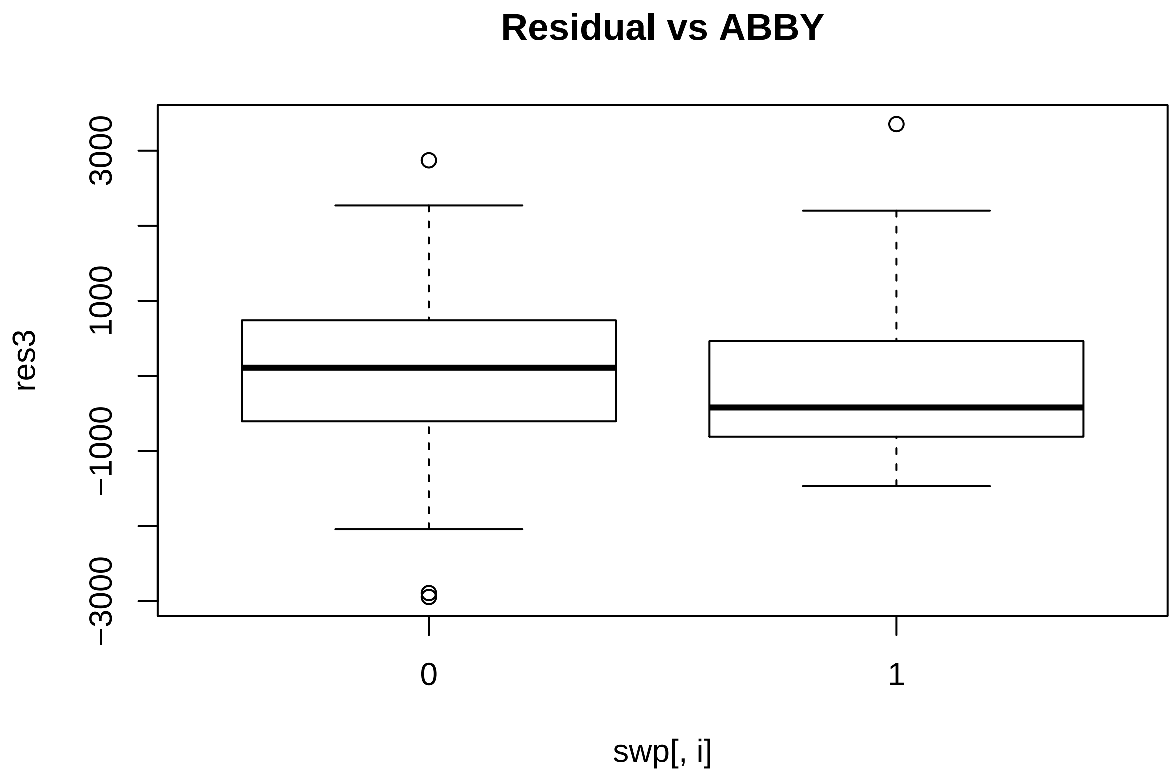


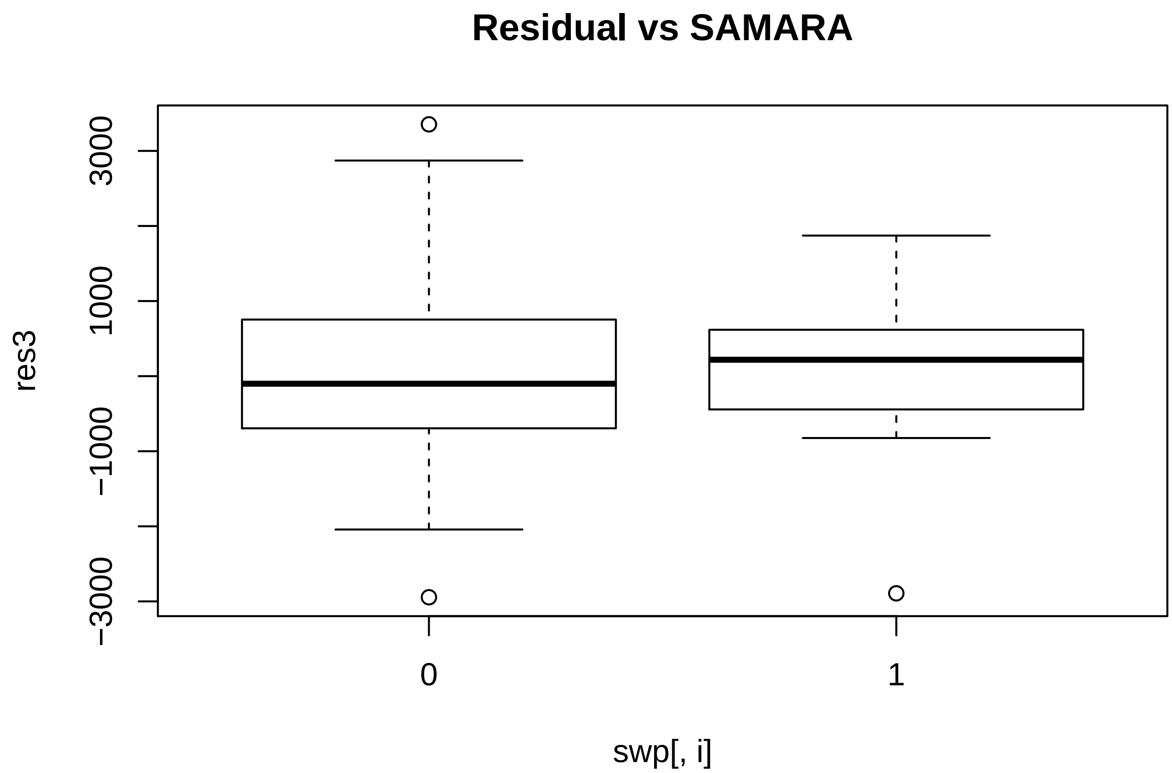
Residual vs KEVIN
res3
3000
1000
−1000
−3000
0
1
swp[, i]

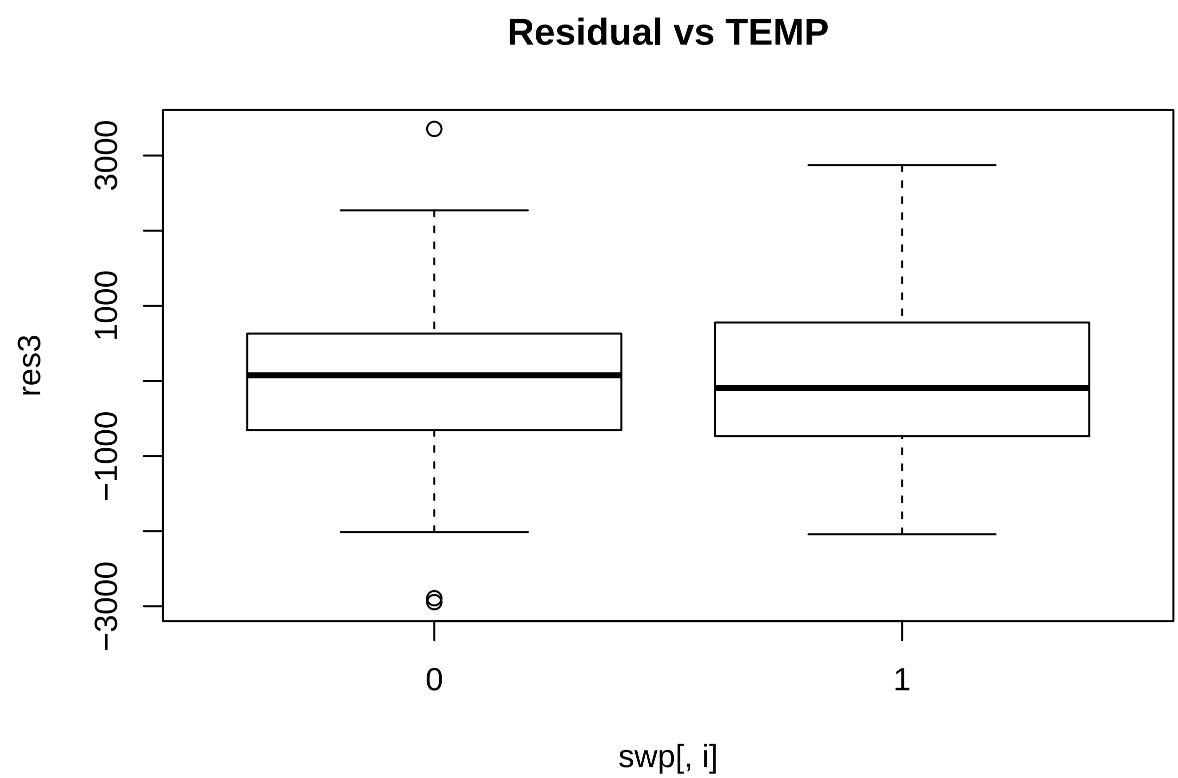

Residual vs OLIMBI
res3
3000
1000
−1000
−3000
0
1
swp[, i]


Residual vs PINO
res3
3000
1000
−1000
−3000
0
1
swp[, i]


Residual vs POUYAN
res3
3000
1000
−1000
−3000
0
1
swp[, i]


Residual vs ABBY
res3
3000
1000
−1000
−3000
0
1
swp[, i]


Residual vs SAMARA
res3
3000
1000
−1000
−3000
0
1
swp[, i]




```r
#Yep, heteroscedasticity.

#So there may be significance when there isn't any.
#We press forward with this in mind.
#################################################################
#################################################################
#4) Interpretation and Discussion.

#Lets look at all the models together.

anova(swlm,swlm1,swlm2,swlm3)
```

```
## Analysis of Variance Table
##
## Model 1: Total_Production ~ DAY + MONTH
## Model 2: Total_Production ~ MONTH + DAY + (BRITTANY + HEATHER + KEVIN +
##     OLIMBI + PINO + POUYAN + ABBY + SAMARA + TEMP)
## Model 3: Total_Production ~ MONTH + DAY + (BRITTANY + HEATHER + KEVIN +
##     OLIMBI + PINO + POUYAN + ABBY + SAMARA + TEMP + BRITTANY *
##     HEATHER + BRITTANY * KEVIN + BRITTANY * OLIMBI + BRITTANY *
##     PINO + BRITTANY * POUYAN + BRITTANY * ABBY + BRITTANY * SAMARA +
##     BRITTANY * TEMP + HEATHER * KEVIN + HEATHER * OLIMBI + HEATHER *
##     PINO + HEATHER * POUYAN + HEATHER * ABBY + HEATHER * SAMARA +
##     HEATHER * TEMP + KEVIN * OLIMBI + KEVIN * PINO + KEVIN *
##     POUYAN + KEVIN * ABBY + KEVIN * SAMARA + KEVIN * TEMP + OLIMBI *
##     PINO + OLIMBI * POUYAN + OLIMBI * ABBY + OLIMBI * SAMARA +
```

```
##     OLIMBI * TEMP + PINO * POUYAN + PINO * ABBY + PINO * SAMARA +
##     PINO * TEMP + POUYAN * ABBY + POUYAN * SAMARA + POUYAN *
##     TEMP + ABBY * SAMARA + ABBY * TEMP + SAMARA * TEMP)
## Model 4: Total_Production ~ MONTH + DAY + (BRITTANY + HEATHER + KEVIN +
##     OLIMBI + PINO + POUYAN + ABBY + SAMARA + TEMP + KEVIN * TEMP +
##     HEATHER * KEVIN + BRITTANY * TEMP + BRITTANY * PINO + BRITTANY *
##     KEVIN)
##   Res.Df       RSS  Df Sum of Sq      F   Pr(>F)
## 1    122 241831139
## 2    113 199672926   9  42158213 2.9978 0.003692 **
## 3     87 135944724  26  63728202 1.5686 0.063170 .
## 4    108 157073438 -21 -21128714 0.6439 0.873981
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#The absolutly most significant model compared to our
#blocking factors (Month and Day) is the model with
#the main effects of the doctors alone (with blocked factors.)

#The reduced F-Test shows that the interaction model
#is somewhat significant at .06 compared to blocked factors.
#All the interaction terms together are significant

anova(swlm2)
```

```
## Analysis of Variance Table
##
## Response: Total_Production
##                   Df    Sum Sq  Mean Sq F value    Pr(>F)
## MONTH              7  39155687  5593670  3.5798 0.0019729 **
## DAY                5 151922573 30384515 19.4451 5.965e-13 ***
## BRITTANY           1   1325897  1325897  0.8485 0.3595162
## HEATHER            1   6856670  6856670  4.3880 0.0390997 *
## KEVIN              1    672235   672235  0.4302 0.5136182
## OLIMBI             1    111817   111817  0.0716 0.7897131
## PINO               1  20566109 20566109 13.1616 0.0004813 ***
## POUYAN             1  11528687 11528687  7.3780 0.0079644 **
## ABBY               1     93299    93299  0.0597 0.8075329
## SAMARA             1    212346   212346  0.1359 0.7132930
## TEMP               1    791152   791152  0.5063 0.4786440
## BRITTANY:HEATHER   1    569140   569140  0.3642 0.5477364
## BRITTANY:KEVIN     1  14215675 14215675  9.0975 0.0033552 **
## BRITTANY:OLIMBI    1    387182   387182  0.2478 0.6198960
## BRITTANY:PINO      1   8741360  8741360  5.5942 0.0202436 *
## BRITTANY:POUYAN    1      2579     2579  0.0017 0.9676842
## BRITTANY:ABBY      1     62710    62710  0.0401 0.8416894
## BRITTANY:TEMP      1   6618671  6618671  4.2357 0.0425740 *
## HEATHER:KEVIN      1   6406779  6406779  4.1001 0.0459462 *
## HEATHER:OLIMBI     1    813746   813746  0.5208 0.4724485
## HEATHER:PINO       1    783535   783535  0.5014 0.4807628
## HEATHER:POUYAN     1    126585   126585  0.0810 0.7766093
## HEATHER:ABBY       1    639749   639749  0.4094 0.5239468
## HEATHER:SAMARA     1    976285   976285  0.6248 0.4314224
## HEATHER:TEMP       1      7583     7583  0.0049 0.9446239
```

```
## KEVIN:OLIMBI       1   1741903  1741903  1.1148 0.2939731
## KEVIN:PINO         1   4146718  4146718  2.6538 0.1069213
## KEVIN:POUYAN       1     58913    58913  0.0377 0.8464951
## KEVIN:ABBY         1     35330    35330  0.0226 0.8808238
## KEVIN:TEMP         1  12342277 12342277  7.8986 0.0061106 **
## OLIMBI:POUYAN      1    222283   222283  0.1423 0.7069690
## OLIMBI:TEMP        1    272456   272456  0.1744 0.6772924
## PINO:POUYAN        1    293090   293090  0.1876 0.6660209
## PINO:ABBY          1    577918   577918  0.3698 0.5446715
## PINO:TEMP          1    343279   343279  0.2197 0.6404512
## POUYAN:TEMP        1   1449057  1449057  0.9273 0.3382223
## ABBY:TEMP          1   1893399  1893399  1.2117 0.2740299
## Residuals         87 135944724  1562583
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#The last model is just the significant interaction terms
#and the main effects (and blocked factors)

#Lets look at the significant doctor pairs.
#Kevin and Temp
#Brittany and Temp
#Heather and Samara
#Brittany and Kevin

#This model look more closely at those

summary(swlm3)
```

```
##
## Call:
## lm(formula = form, data = swp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2944.7  -657.4    55.7   731.0  3353.6
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       9886.23     901.95  10.961  < 2e-16 ***
## MONTH1             272.93     278.36   0.980  0.32904
## MONTH2            1316.12     398.75   3.301  0.00131 **
## MONTH3            -981.43     307.38  -3.193  0.00185 **
## MONTH4            -776.90     380.11  -2.044  0.04340 *
## MONTH5             492.33     305.49   1.612  0.10996
## MONTH6             380.38     289.60   1.313  0.19180
## MONTH7            -523.00     348.63  -1.500  0.13649
## DAY1             -3039.76     964.14  -3.153  0.00209 **
## DAY2               851.49     495.57   1.718  0.08863 .
## DAY3              1474.88    1033.11   1.428  0.15629
## DAY4               274.23     663.64   0.413  0.68026
## DAY5                72.42     523.62   0.138  0.89026
## BRITTANY1         1056.16     970.06   1.089  0.27869
## HEATHER1          -484.58     677.17  -0.716  0.47578
```

```
## KEVIN1            -505.87     663.59  -0.762  0.44753
## OLIMBI1            699.51     563.57   1.241  0.21722
## PINO1              425.67     523.49   0.813  0.41792
## POUYAN1           1377.93     533.04   2.585  0.01107 *
## ABBY1             -192.73     516.67  -0.373  0.70986
## SAMARA1            801.77    1028.32   0.780  0.43728
## TEMP1            -1237.25     566.46  -2.184  0.03111 *
## KEVIN1:TEMP1      1536.44     630.98   2.435  0.01653 *
## HEATHER1:KEVIN1   2048.73     748.49   2.737  0.00725 **
## BRITTANY1:TEMP1   -920.79     578.17  -1.593  0.11417
## BRITTANY1:PINO1   1130.89     908.41   1.245  0.21586
## BRITTANY1:KEVIN1 -1381.62     744.68  -1.855  0.06628 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1206 on 108 degrees of freedom
## Multiple R-squared:  0.6372, Adjusted R-squared:  0.5498
## F-statistic: 7.295 on 26 and 108 DF,  p-value: 4.209e-14
```

```r
#Heather and Kevin appears to be a good team.
#But Brittany and kevin aren't too good.

#I believe that these interaction terms are not
#important enough to consider when scheduling doctors.
#Not when considering the main effects.

#Lets look at the individual doctor main effects.

summary(swlm1)
```

```
##
## Call:
## lm(formula = form, data = swp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3323.5  -717.8   -34.3   672.7  2997.7
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8696.54     895.61   9.710  < 2e-16 ***
## MONTH1         47.44     288.06   0.165  0.86949
## MONTH2       1457.04     415.93   3.503  0.00066 ***
## MONTH3       -670.73     328.21  -2.044  0.04332 *
## MONTH4       -763.22     398.09  -1.917  0.05773 .
## MONTH5        305.22     333.02   0.917  0.36135
## MONTH6         41.91     307.31   0.136  0.89177
## MONTH7       -124.91     361.88  -0.345  0.73061
## DAY1        -2062.44     956.84  -2.155  0.03325 *
## DAY2          395.04     511.10   0.773  0.44118
## DAY3         -193.17     958.34  -0.202  0.84061
## DAY4          788.13     671.50   1.174  0.24299
## DAY5          379.98     560.91   0.677  0.49952
## BRITTANY1    -215.98     698.09  -0.309  0.75760
```

```
## HEATHER1      738.06     504.95   1.462  0.14661
## KEVIN1        427.10     452.57   0.944  0.34733
## OLIMBI1       674.77     565.99   1.192  0.23568
## PINO1        1489.99     461.48   3.229  0.00163 **
## POUYAN1      1267.48     568.94   2.228  0.02787 *
## ABBY1          69.98     545.51   0.128  0.89815
## SAMARA1       325.97    1109.23   0.294  0.76939
## TEMP1        -223.37     333.82  -0.669  0.50478
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1329 on 113 degrees of freedom
## Multiple R-squared:  0.5388, Adjusted R-squared:  0.453
## F-statistic: 6.285 on 21 and 113 DF,  p-value: 3.002e-11
```

```r
#If we just look at significace:
#Pino, Pouyan are significant. But we know that Pino
#Is signifinant because he is a pediatric doctor and
#charges more. So that doesn't help.
#The others are then "average" and should be considered equal.

#The estimated impact on prodution of Pino and Pouyan is:
#Pouyan: +$1267
#Pino: +$1490

#But this doesn't help very much.

#Unfortinuatly there isn't much to help for scheduling.
#It turns out that there isn't much to get from the data.
#Except for Pouyan, which it is known that he produces more.

#Lets look at the base, blocked model.

summary(swlm)
```

```
##
## Call:
## lm(formula = Total_Production ~ DAY + MONTH, data = swp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3657.5  -702.2    53.4   711.1  3823.0
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10075.17     200.23  50.317  < 2e-16 ***
## DAY1        -2543.54     359.66  -7.072 1.04e-10 ***
## DAY2         1256.92     293.49   4.283 3.70e-05 ***
## DAY3          -89.59     859.40  -0.104  0.91714
## DAY4          241.03     290.42   0.830  0.40819
## DAY5          556.78     289.51   1.923  0.05678 .
## MONTH1         38.82     301.84   0.129  0.89787
## MONTH2       1090.59     380.29   2.868  0.00487 **
## MONTH3       -422.05     321.51  -1.313  0.19174
```

```
## MONTH4      -1074.96     389.15  -2.762  0.00663 **
## MONTH5        597.11     307.27   1.943  0.05429 .
## MONTH6        235.16     307.06   0.766  0.44524
## MONTH7       -246.82     322.78  -0.765  0.44593
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1408 on 122 degrees of freedom
## Multiple R-squared:  0.4414, Adjusted R-squared:  0.3864
## F-statistic: 8.033 on 12 and 122 DF,  p-value: 6.236e-11
```

```r
#It is likely that is best to just schedule based on the day.
#Lets make a quick model with just the day.

levels(swp$DAY)
```

```
## [1] "Friday"    "Monday"    "Saturday"  "Thursday"  "Tuesday"   "Wednesday"
```

```r
#And it appears that Friday is terrible and monday is great for
#for production, but this makes lots of sense.

levels(swp$MONTH)
```

```
## [1] "April"    "August"   "February" "January"  "July"     "June"
## [7] "March"    "May"
```

```r
#August is good and January isn't.
############################################################################
############################################################################
#5) Final words.

#We determined that the data doesn't really have much to help.
#The most signifince came from known points.

#Pouyan makes a lot of money, (he is a pediatric doctor.)
#Friday is terrible, monday is better. So the advice is to try to
#schedule more on friday to make more even, or schdule more
#doctors on monday.

#The doctor pairs don't have much significance. Heather and Kevin seem to be
#a good pair and Brittany and Kevin aren't too good.

#August makes a lot of money and January doesn't make a lot. This might
#help in determing the best days to take a vacation.

#There is some information that can be sussed out fromt this data set
#but not much of it is too helpful. The initial question was to determine
#doctor effectivness in order to better schedule.

#The final answer is that the data doesn't show much
#more than they already know.
```