Regression Models - Morrison Bridge NAWQA

Mark Roes

June 3, 2017

```
data.tsi <- read.csv(series, stringsAsFactors = FALSE)
data.di <- read.csv(discrete, stringsAsFactors = FALSE)
data.all <- merge(data.tsi, data.di, by="DateTime", all.y=TRUE)
#data.all <- with(data.all, rm("DateTime", "Date", "date", "time", "sample_dt", "sample_tm"))
# manage dates
data.all$DateTime <- chron(dates=data.all$Date.chron, times=data.all$Time.chron,
                           format=c("m/d/y", "h:m:s"))
```

Some info of Metolachlor:

Metolachlor is an herbicide used in agriculture to control grasses. Its use is controversial, the chemical is approved in the US for use, but NOT in the EU. There are tolerance levels for this chemical in agriculture, but no maximum allowable level in drinking water – only an advisory level. Acoording to GHS (Globally Harmonized System for Hazard Communication), an international accord on chemical hazard, Metolachlor is:

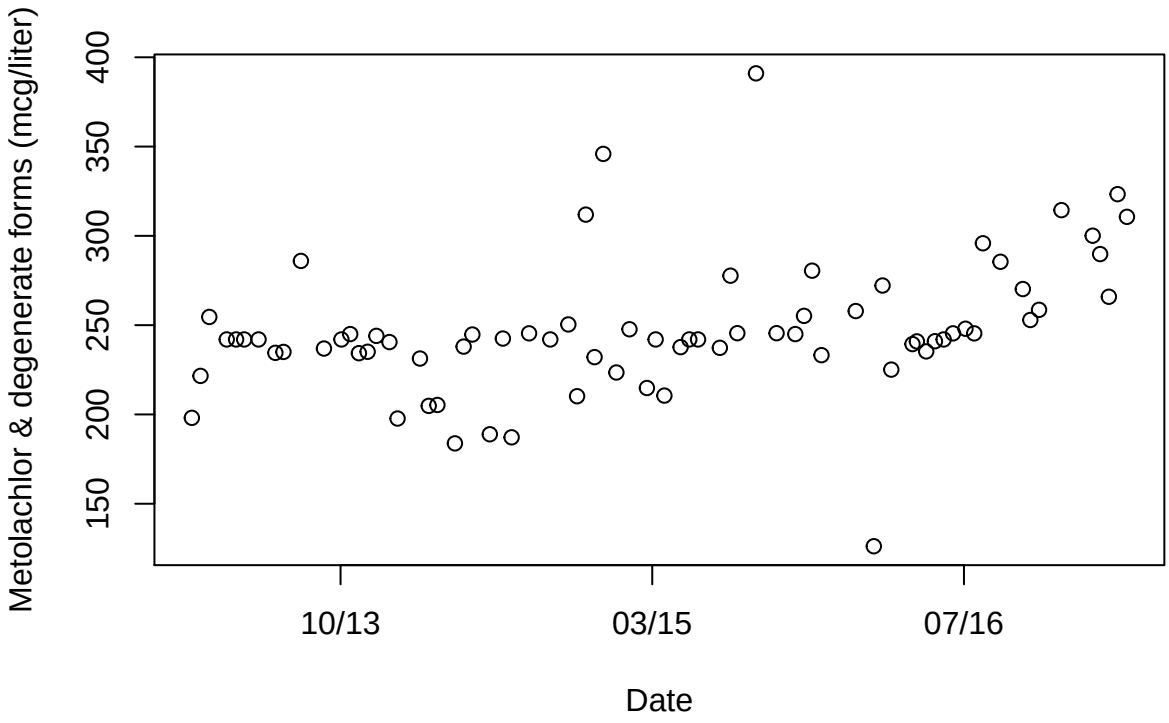
May cause an allergic skin reaction [Warning Sensitization, Skin - Category 1]

Fatal if inhaled [Danger Acute toxicity, inhalation - Category 1, 2]

Very toxic to aquatic life [Warning Hazardous to the aquatic environment, acute hazard - Category 1]

Very toxic to aquatic life with long lasting effects [Warning Hazardous to the aquatic environment, long-term hazard - Category 1]

```
plot(data.all$DateTime, data.all$metoforms, xlab = "Date", ylab = "Metolachlor & degenerate forms (mcg/
```




```
#lets try a simple general additive mixed model!
gam.meto<-gam(metoforms~Discharge+StreamVelocity+GageHeight+SensorDepth+WaterTemp+DissolvedO2+O2Saturati
summary(gam.meto)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## metoforms ~ Discharge + StreamVelocity + GageHeight + SensorDepth +
##     WaterTemp + DissolvedO2 + O2Saturation + pH + Conductance +
##     Turbidity + SecchiDepth + Chlorophyll + Phycocyanin + fDOM +
##     Nitrate
##
## Parametric coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -735.5961   453.0750  -1.624   0.1099
## Discharge        -0.5350     3.0007  -0.178   0.8591
## StreamVelocity   15.6134   116.5071   0.134   0.8939
## GageHeight       -5.0014     9.4019  -0.532   0.5968
## SensorDepth       5.7730     9.0040   0.641   0.5239
## WaterTemp        10.9407     6.9564   1.573   0.1212
## DissolvedO2      37.1763    28.6895   1.296   0.2002
## O2Saturation     -1.7410     2.6390  -0.660   0.5120
## pH               67.8281    52.0944   1.302   0.1981
## Conductance       1.5553     0.9301   1.672   0.0999 .
```

```
## Turbidity         1.7402     1.5456   1.126   0.2648
## SecchiDepth      -3.8917     4.4564  -0.873   0.3861
## Chlorophyll      -9.2617     4.1103  -2.253   0.0280 *
## Phycocyanin      37.4974    33.4809   1.120   0.2673
## fDOM             -0.4162     1.8511  -0.225   0.8229
## Nitrate         -26.3492    40.4253  -0.652   0.5171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) =  0.0658   Deviance explained = 25.8%
## GCV =   1710  Scale est. = 1340.3    n = 74
```

```
#reduced gam model, did a stepwise removal of factors with highest p-values
#If i remove any more than this, the % deviance explained drops drastically
reduced.gam.meto<-gam(metoforms~WaterTemp+DissolvedO2+pH+Conductance+Turbidity+Chlorophyll+Phycocyanin,
summary(reduced.gam.meto)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## metoforms ~ WaterTemp + DissolvedO2 + pH + Conductance + Turbidity +
##     Chlorophyll + Phycocyanin
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -590.2417   367.3369  -1.607   0.1129
## WaterTemp      7.3345     3.5943   2.041   0.0453 *
## DissolvedO2   20.6528    12.3621   1.671   0.0995 .
## pH            57.2329    39.2944   1.457   0.1500
## Conductance    1.1694     0.6158   1.899   0.0619 .
## Turbidity      1.4945     0.9118   1.639   0.1060
## Chlorophyll   -7.9208     3.5383  -2.239   0.0286 *
## Phycocyanin   31.1671    24.9213   1.251   0.2155
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) =  0.146   Deviance explained = 22.8%
## GCV = 1374.2  Scale est. = 1225.7    n = 74
```

```
#linear model
lm.meto<-lm(metoforms~Discharge+StreamVelocity+GageHeight+SensorDepth+WaterTemp+DissolvedO2+O2Saturatio
summary(lm.meto)
```

```
##
## Call:
## lm(formula = metoforms ~ Discharge + StreamVelocity + GageHeight +
##     SensorDepth + WaterTemp + DissolvedO2 + O2Saturation + pH +
##     Conductance + Turbidity + SecchiDepth + Chlorophyll + Phycocyanin +
##     fDOM + Nitrate, data = data.all)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

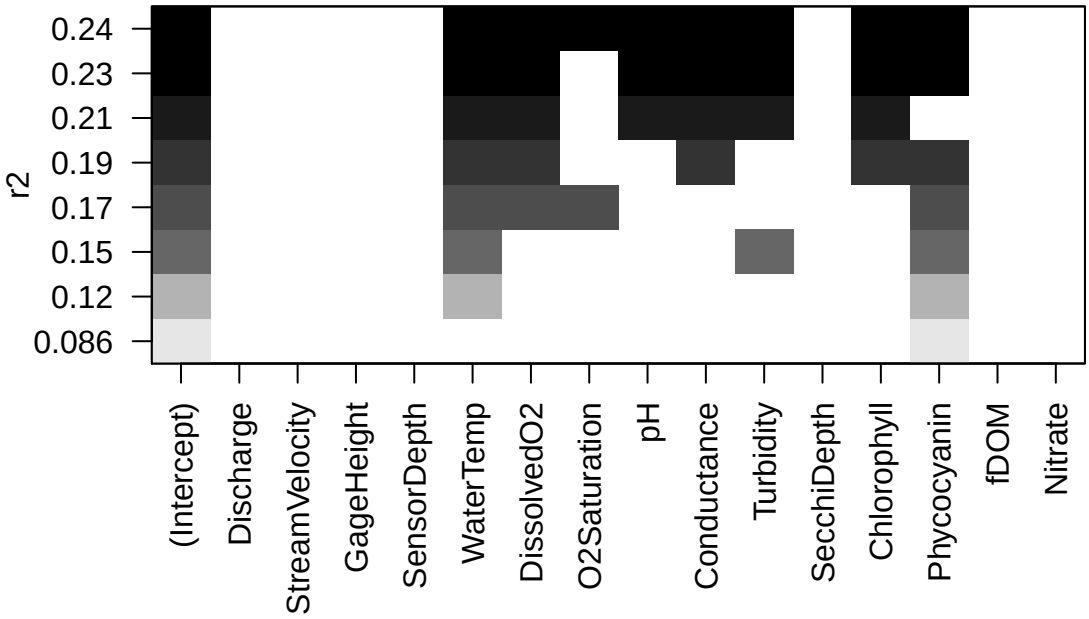
```
## -134.846  -14.244    2.444   10.521  108.683
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -735.5961   453.0750  -1.624   0.1099
## Discharge        -0.5350     3.0007  -0.178   0.8591
## StreamVelocity   15.6134   116.5071   0.134   0.8939
## GageHeight       -5.0014     9.4019  -0.532   0.5968
## SensorDepth       5.7730     9.0040   0.641   0.5239
## WaterTemp        10.9407     6.9564   1.573   0.1212
## DissolvedO2      37.1763    28.6895   1.296   0.2002
## O2Saturation     -1.7410     2.6390  -0.660   0.5120
## pH               67.8281    52.0944   1.302   0.1981
## Conductance       1.5553     0.9301   1.672   0.0999 .
## Turbidity         1.7402     1.5456   1.126   0.2648
## SecchiDepth      -3.8917     4.4564  -0.873   0.3861
## Chlorophyll      -9.2617     4.1103  -2.253   0.0280 *
## Phycocyanin      37.4974    33.4809   1.120   0.2673
## fDOM             -0.4162     1.8511  -0.225   0.8229
## Nitrate         -26.3492    40.4253  -0.652   0.5171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36.61 on 58 degrees of freedom
## Multiple R-squared:  0.2577, Adjusted R-squared:  0.06576
## F-statistic: 1.343 on 15 and 58 DF,  p-value: 0.2078
```

```
reduced.lm<-lm(metoforms~WaterTemp+DissolvedO2+pH+Conductance+Turbidity+Chlorophyll+Phycocyanin, data =
summary(reduced.lm)
```

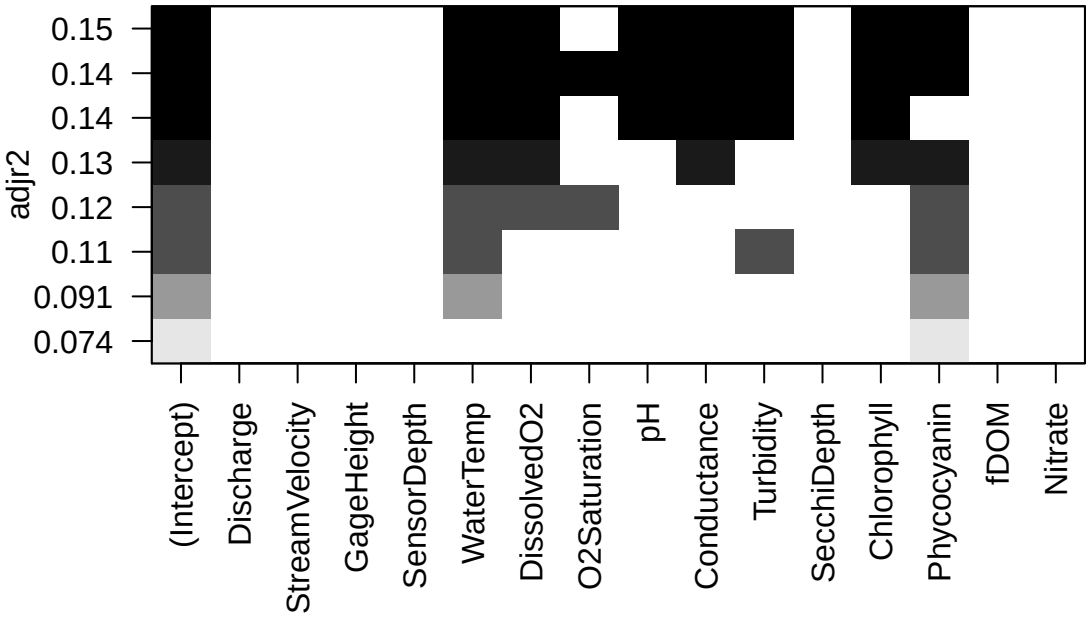
```
##
## Call:
## lm(formula = metoforms ~ WaterTemp + DissolvedO2 + pH + Conductance +
##     Turbidity + Chlorophyll + Phycocyanin, data = data.all)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -131.063  -12.714   -1.717   13.474  104.427
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -590.2417   367.3369  -1.607   0.1129
## WaterTemp      7.3345     3.5943   2.041   0.0453 *
## DissolvedO2   20.6528    12.3621   1.671   0.0995 .
## pH            57.2329    39.2944   1.457   0.1500
## Conductance    1.1694     0.6158   1.899   0.0619 .
## Turbidity      1.4945     0.9118   1.639   0.1060
## Chlorophyll   -7.9208     3.5383  -2.239   0.0286 *
## Phycocyanin   31.1671    24.9213   1.251   0.2155
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.01 on 66 degrees of freedom
## Multiple R-squared:  0.2276, Adjusted R-squared:  0.1456
## F-statistic: 2.778 on 7 and 66 DF,  p-value: 0.01357
```

The linear model with all factors has only 'chlorophyll' as significant. Using the leaps package to test the optimal set of factors.

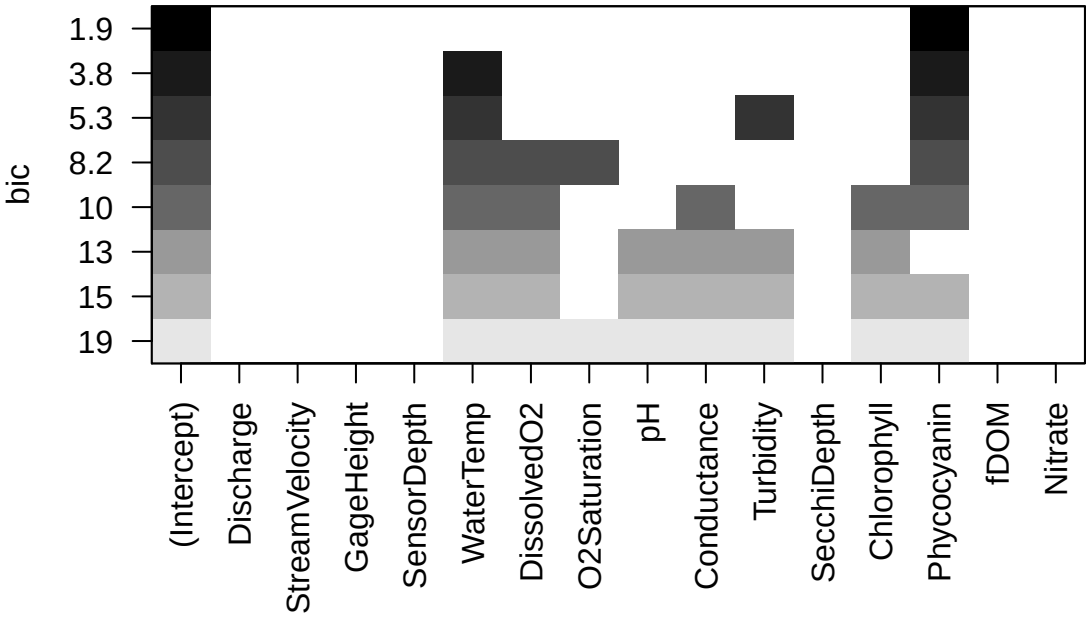
```
leaps.meto<-regsubsets(x = data.all[,4:18], y = data.all$metoforms)
plot(leaps.meto, scale = "r2")
```



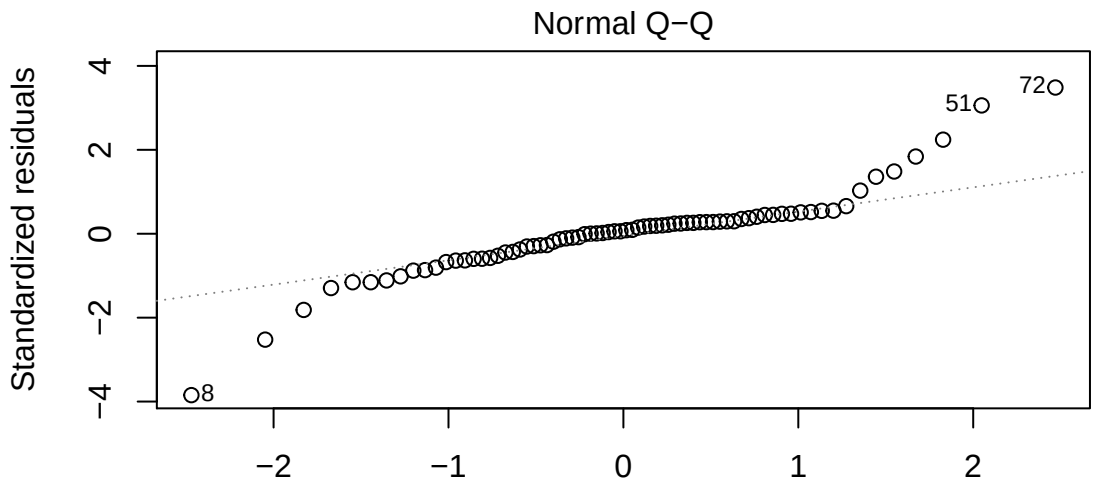

```
plot(leaps.meto, scale = "adjr2")
```



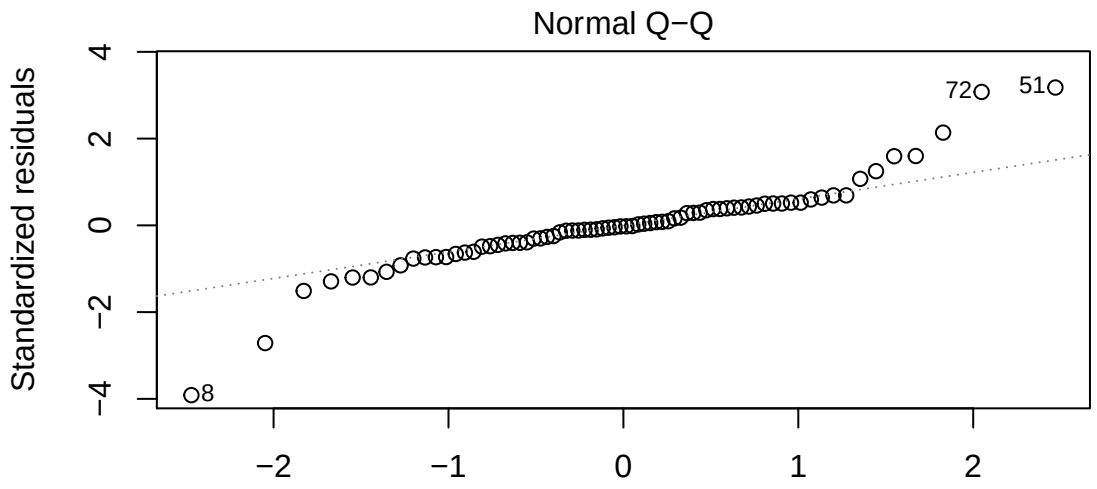

```
plot(leaps.meto, scale = "bic")
```



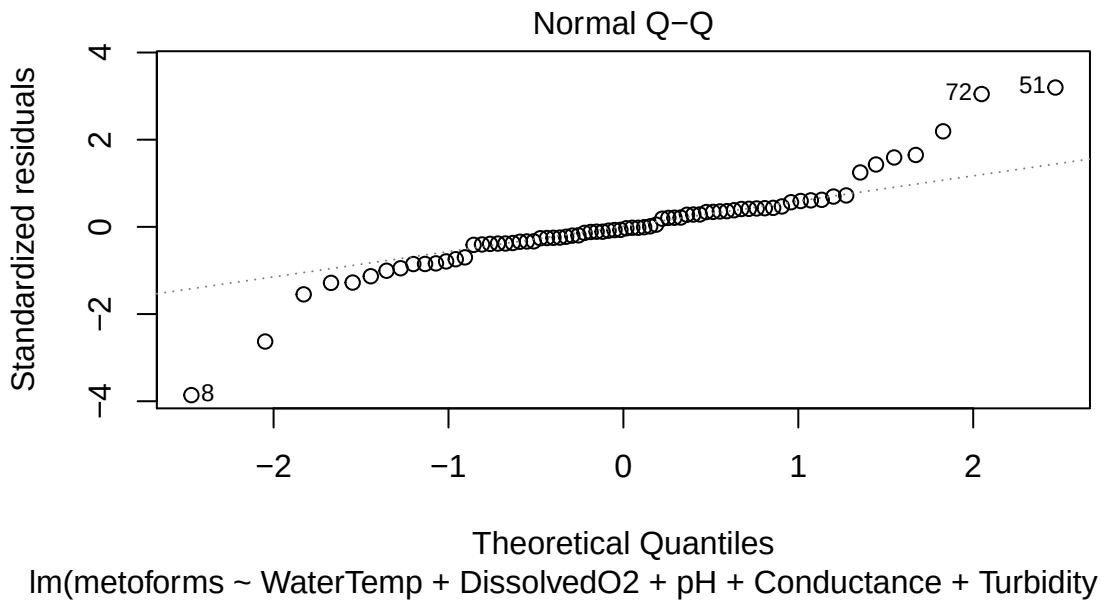

```
#Now test for fit with these 3 models created by leaps
#full model
plot(lm(metoforms~Discharge+StreamVelocity+GageHeight+SensorDepth+WaterTemp+DissolvedO2+O2Saturation+pH
```



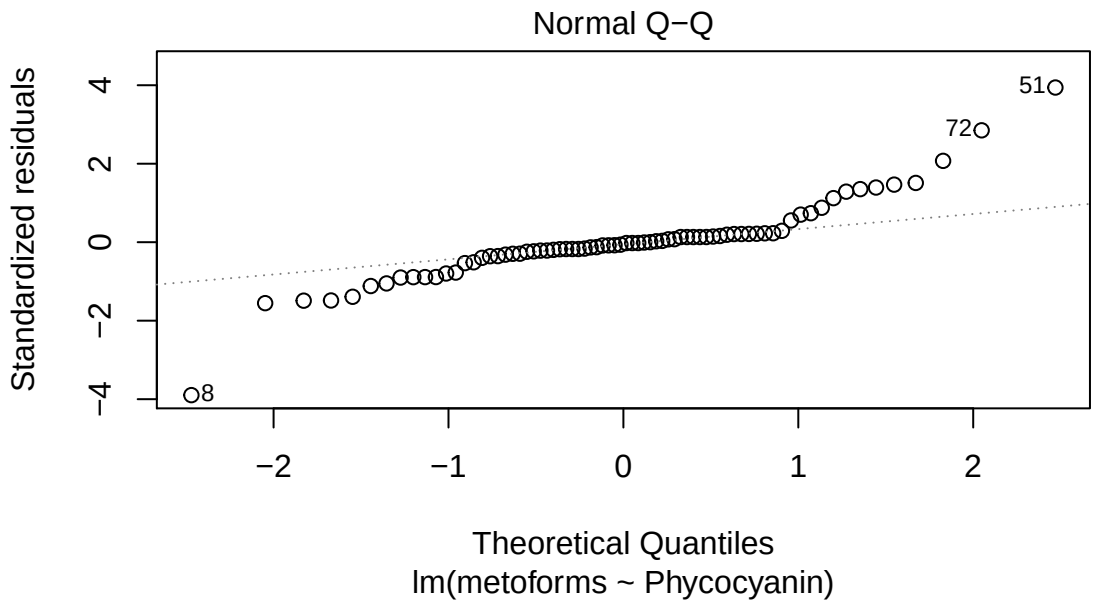

```
#highest r2
plot(lm(metoforms~WaterTemp+DissolvedO2+O2Saturation+pH+Conductance+Turbidity+Chlorophyll+Phycocyanin, d
```




```
#highest adjr2
plot(lm(metoforms~WaterTemp+DissolvedO2+pH+Conductance+Turbidity+Chlorophyll+Phycocyanin, data = data.al
```




```
#lowest BIC
plot(lm(metoforms~Phycocyanin, data = data.all), which = 2)
```




```
#quantile regression for meto with tau = .5
rq.meto<-rq(metoforms~WaterTemp+DissolvedO2+pH+Conductance+Turbidity+Chlorophyll+Phycocyanin, data = da
summary(rq.meto)
```

```
##
## Call: rq(formula = metoforms ~ WaterTemp + DissolvedO2 + pH + Conductance +
##     Turbidity + Chlorophyll + Phycocyanin, data = data.all)
##
## tau: [1] 0.5
##
## Coefficients:
##              coefficients lower bd   upper bd
## (Intercept) -285.12336   -983.40177  323.35978
## WaterTemp      1.33968     -2.17665    9.43410
## DissolvedO2    0.77693    -11.62171   20.75141
## pH            60.86486     -8.13841  110.82329
## Conductance    0.63455     -0.05009    1.03935
## Turbidity      1.70864      0.03051    2.38374
## Chlorophyll   -4.68017     -6.84272    0.48336
## Phycocyanin   35.00731      1.45508   85.75796
```

```
#Let's compare the AIC values for all of these models:
#AIC for quantile regression
AIC.rq(rq.meto)
```

```
## [1] 713.1188
## attr(,"edf")
```

```
## [1] 8
```

```
#AIC for full LM model
AIC(lm.meto)
```

```
## [1] 758.8205
```

```
#AIC for best linear model by BIC
AIC(lm(metoforms~Phycocyanin, data = data.all))
```

```
## [1] 746.2035
```

```
#AIC for best linear model by r2
AIC(lm(metoforms~WaterTemp+DissolvedO2+O2Saturation+pH+Conductance+Turbidity+Chlorophyll+Phycocyanin, da
```

```
## [1] 746.8268
```

```
#AIC for best linear model by adjr2
AIC(lm(metoforms~WaterTemp+DissolvedO2+pH+Conductance+Turbidity+Chlorophyll+Phycocyanin, data = data.al
```

```
## [1] 745.7691
```

```
#AIC for the two general additive mixed models
llgam<-logLik.gam(gam.meto)
AIC(llgam)
```

```
## [1] 758.8205
```

```
red.llgam<-logLik.gam(reduced.gam.meto)
AIC(red.llgam)
```

```
## [1] 745.7691
```

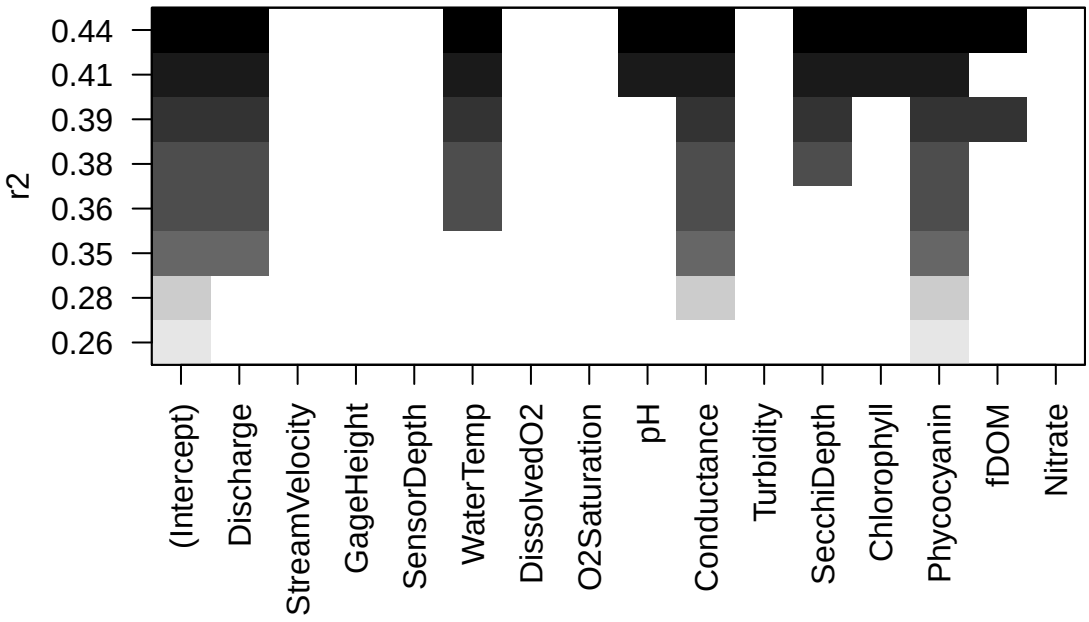
```
#nonlinear quantile regression estimate, havent got to work yet
#nlrq.meto<-nlrq(metoforms~WaterTemp+DissolvedO2+pH+Conductance+Turbidity+Chlorophyll+Phycocyanin, data
```

```
total<-data.all$diazforms+data.all$metoforms+data.all$hexaforms+data.all$fipforms
lm.all<-lm(total~Discharge+StreamVelocity+GageHeight+SensorDepth+WaterTemp+DissolvedO2+O2Saturation+pH+
summary(lm.all)
```

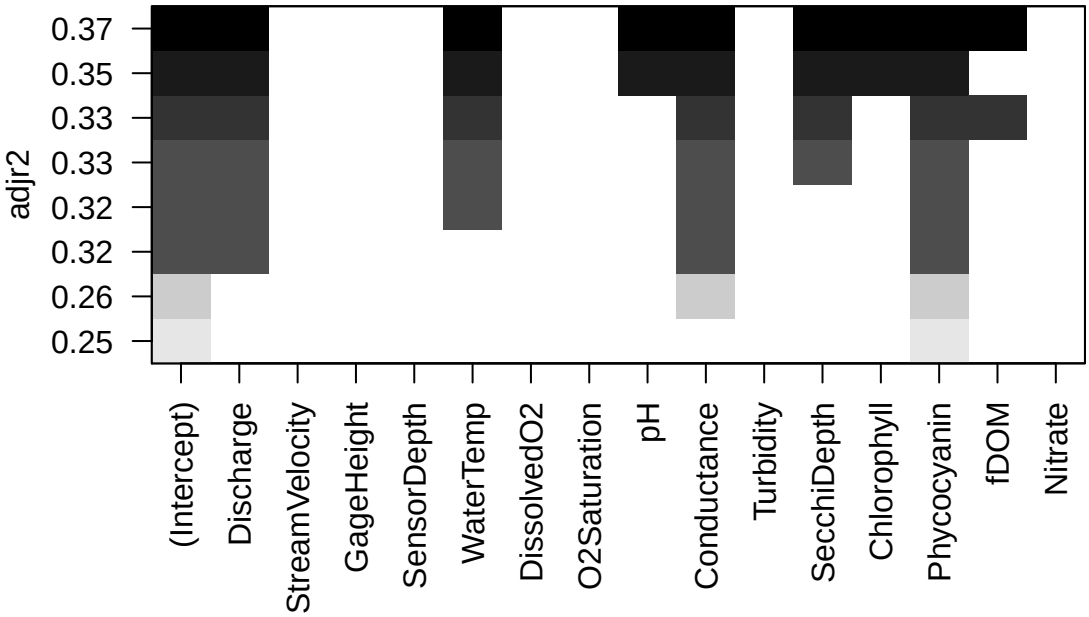
```
##
## Call:
## lm(formula = total ~ Discharge + StreamVelocity + GageHeight +
##     SensorDepth + WaterTemp + DissolvedO2 + O2Saturation + pH +
##     Conductance + Turbidity + SecchiDepth + Chlorophyll + Phycocyanin +
##     fDOM + Nitrate, data = data.all)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -136.00  -34.32  -10.10   27.45  246.89
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1853.5910   896.8906  -2.067   0.0432 *
## Discharge          2.9646     5.9401   0.499   0.6196
## StreamVelocity   -69.9011   230.6331  -0.303   0.7629
## GageHeight       -14.3647    18.6117  -0.772   0.4434
## SensorDepth       19.0579    17.8240   1.069   0.2894
## WaterTemp         16.6452    13.7707   1.209   0.2317
## DissolvedO2       45.9892    56.7927   0.810   0.4214
```

```
## O2Saturation      -3.7784     5.2240  -0.723   0.4724
## pH               258.2146   103.1242   2.504   0.0151 *
## Conductance        4.1235     1.8411   2.240   0.0290 *
## Turbidity         -0.9972     3.0596  -0.326   0.7456
## SecchiDepth      -18.4593     8.8217  -2.092   0.0408 *
## Chlorophyll      -18.6846     8.1367  -2.296   0.0253 *
## Phycocyanin      145.6791    66.2776   2.198   0.0320 *
## fDOM              -6.1385     3.6645  -1.675   0.0993 .
## Nitrate           89.3141    80.0245   1.116   0.2690
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 72.47 on 58 degrees of freedom
## Multiple R-squared:  0.483,  Adjusted R-squared:  0.3493
## F-statistic: 3.612 on 15 and 58 DF,  p-value: 0.0002078
```

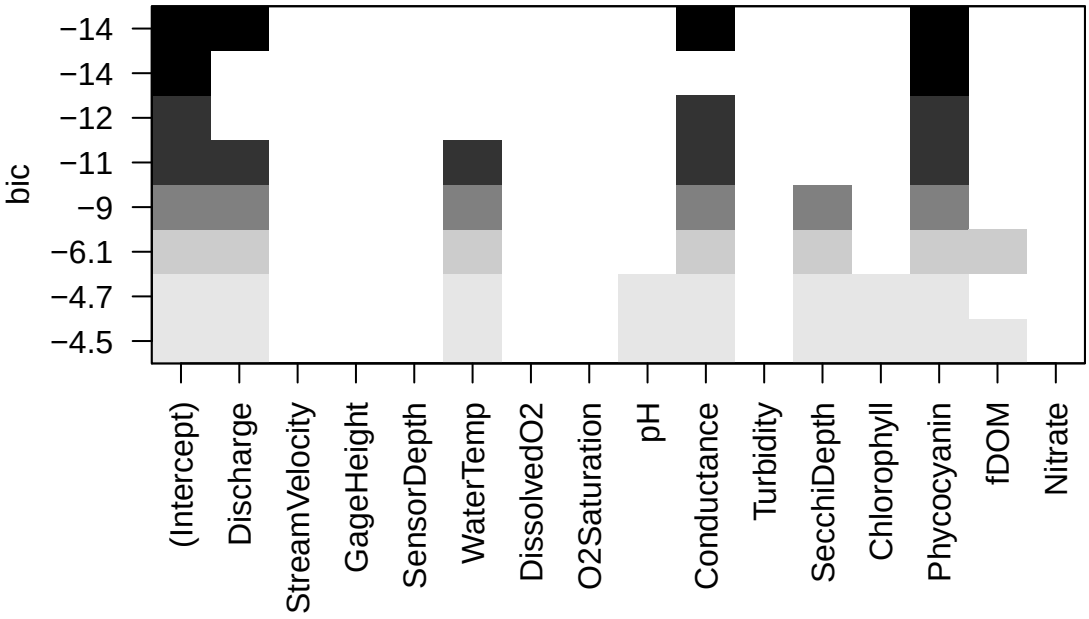
```
subset.all<-regsubsets(x = data.all[,4:18], y = total)
plot(subset.all, scale = "r2")
```




```
plot(subset.all, scale = "adjr2")
```




```
plot(subset.all, scale = "bic")
```


bic
−14
−14
−12
−11
−9
−6.1
−4.7
−4.5
(Intercept)
Discharge
StreamVelocity
GageHeight
SensorDepth
WaterTemp
DissolvedO2
O2Saturation
pH
Conductance
Turbidity
SecchiDepth
Chlorophyll
Phycocyanin
fDOM
Nitrate