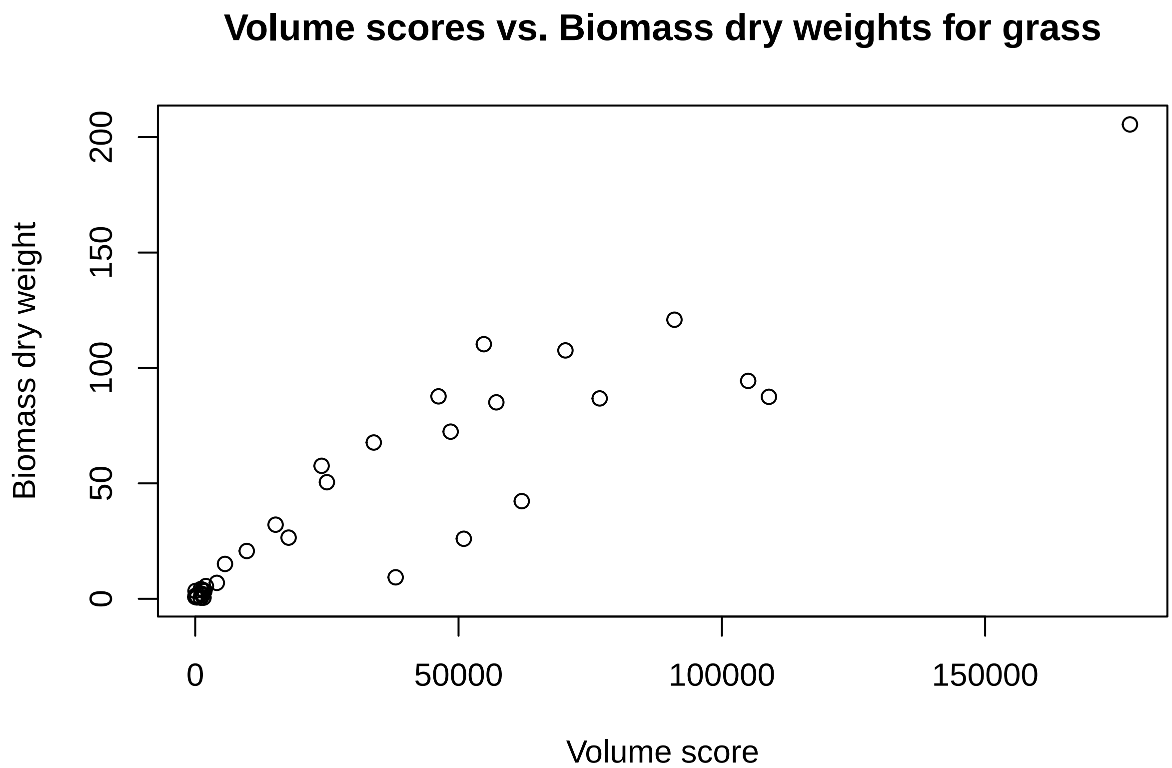
Initialization

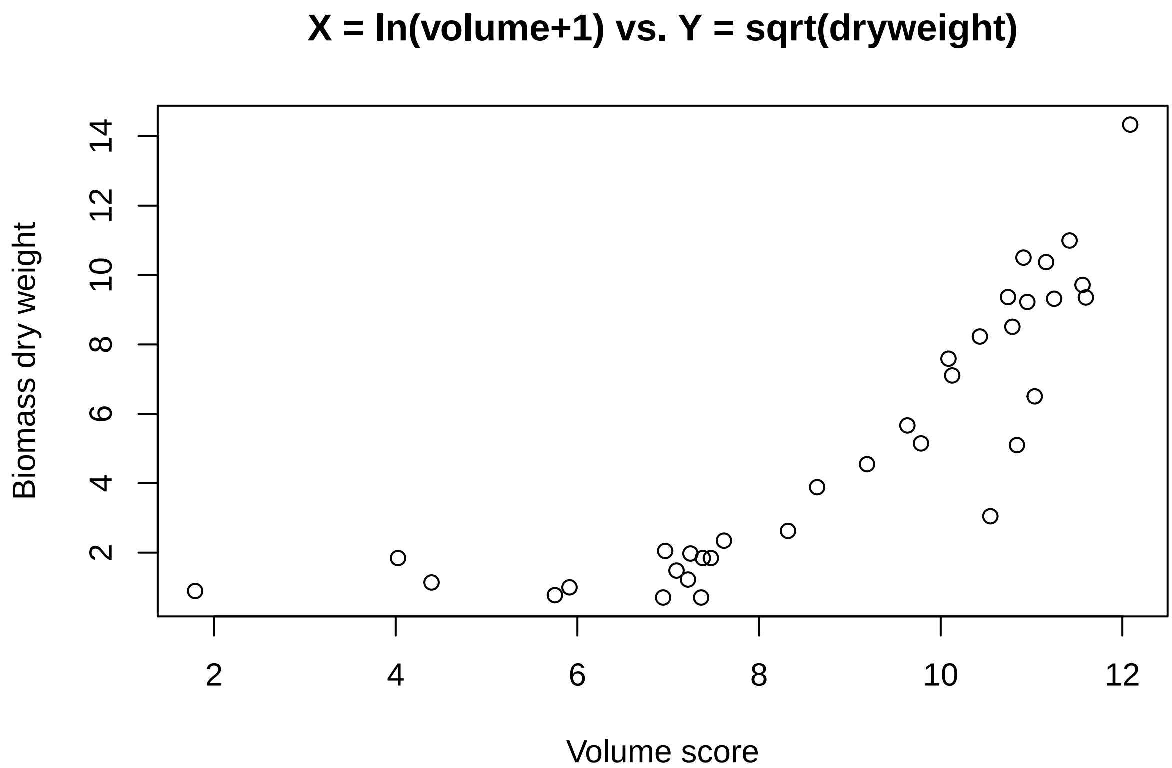
```
knitr::opts_chunk$set(echo = TRUE)

dw<-read.table("dryweight.txt",header=T)
attach(dw)
```

```
## The following objects are masked from dw (pos = 13):
##
##     dryweight, volume

## The following objects are masked from dw (pos = 14):
##
##     dryweight, volume
```

```
# the entire table
#dw

# the independent/dependent variables
#volume
#dryweight

# volume is independent/predictor, dryweight is dependent
# volume is X, dryweight is Y

# volume score is volume of space occupied by the plant (in this case grass)
# Biomass dry weight is biomass dry weights for grass

plot(volume, dryweight, xlab="Volume score", ylab="Biomass dry weight",main="Volume
↪  scores vs. Biomass dry weights for grass")
```




```
# ln(x) is log(x) in R

Y <- sqrt(dryweight)
X <- log(volume + 1)

plot(X, Y, xlab="Volume score", ylab="Biomass dry weight",main="X = ln(volume+1) vs. Y =
↪  sqrt(dryweight)")
```



Fitting a linear model

```
fitlinear<-lm(Y~ X)
summary(fitlinear)
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3918 -1.4560 -0.4075  1.1055  4.8836
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -6.3281     1.3226  -4.784 3.48e-05 ***
## X             1.3055     0.1446   9.028 1.96e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.109 on 33 degrees of freedom
## Multiple R-squared:  0.7118, Adjusted R-squared:  0.7031
## F-statistic:  81.5 on 1 and 33 DF,  p-value: 1.964e-10
```

```
step(fitlinear)
```

```
## Start:  AIC=54.18
## Y ~ X
##
##        Df Sum of Sq    RSS    AIC
## <none>              146.78 54.177
## - X     1    362.53 509.32 95.720
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Coefficients:
## (Intercept)            X
##      -6.328        1.306
```

Fitting a quadratic model

```
fitqr<-lm(Y~ X + I(X^2))
summary(fitqr)
```

```
##
## Call:
## lm(formula = Y ~ X + I(X^2))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5032 -0.5126  0.2398  0.7353  2.4362
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.12914    1.95059   2.630 0.013032 *
## X           -2.03267    0.51674  -3.934 0.000422 ***
## I(X^2)       0.21451    0.03263   6.574 2.08e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.397 on 32 degrees of freedom
## Multiple R-squared:  0.8774, Adjusted R-squared:  0.8697
## F-statistic: 114.5 on 2 and 32 DF,  p-value: 2.608e-15
```

```
step(fitqr)
```

```
## Start:  AIC=26.26
## Y ~ X + I(X^2)
##
##          Df Sum of Sq     RSS    AIC
```

```
## <none>                 62.446 26.263
## - X       1    30.196  92.642 38.069
## - I(X^2)  1    84.339 146.785 54.177
```

```
##
## Call:
## lm(formula = Y ~ X + I(X^2))
##
## Coefficients:
## (Intercept)            X       I(X^2)
##      5.1291      -2.0327       0.2145
```

Fitting a cubic model

```
fitcb<-lm(Y~ X + I(X^2)+I(X^3))
summary(fitcb)
```

```
##
## Call:
## lm(formula = Y ~ X + I(X^2) + I(X^3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2373 -0.2723  0.1035  0.7331  2.0993
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.31177    3.36006   0.093   0.9267
## X            0.80114    1.70856   0.469   0.6424
## I(X^2)      -0.23871    0.26315  -0.907   0.3713
## I(X^3)       0.02138    0.01232   1.735   0.0927 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.355 on 31 degrees of freedom
## Multiple R-squared:  0.8882, Adjusted R-squared:  0.8774
## F-statistic: 82.13 on 3 and 31 DF,  p-value: 7.621e-15
```

Backwards elimination for cubic model

```
#1. Fit the model with predictors (initial is all predictors)
#2. Find the predictor with the highest p-value
#3. If the predictor with the highest p-value has p-value > alpha, then remove that
↪  predictor
#4. Repeat to step 1. Or if all p-values are less than the alpha, then backwards
↪  elimination is complete.
```

```r
# Using significance level alpha = 0.05,
# Also, will not be removing the intercept for the backwards elimination for now

# removing predictor X
fitcb2<-lm(Y~ I(X^2)+I(X^3))
summary(fitcb2)
```

```
##
## Call:
## lm(formula = Y ~ I(X^2) + I(X^3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3041 -0.3511  0.2547  0.7649  2.0704
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.80585    1.05325   1.715   0.0961 .
## I(X^2)      -0.11733    0.04669  -2.513   0.0172 *
## I(X^3)       0.01585    0.00357   4.440   0.0001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.338 on 32 degrees of freedom
## Multiple R-squared:  0.8875, Adjusted R-squared:  0.8804
## F-statistic: 126.2 on 2 and 32 DF,  p-value: 6.629e-16
```

```r
# Backwards regression shows Y ~ I(X^2) + I(X^3) is best model
```

Stepwise regression for cubic model

```r
step(fitcb)
```

```
## Start:  AIC=25.02
## Y ~ X + I(X^2) + I(X^3)
##
##          Df Sum of Sq    RSS    AIC
## - X       1    0.4037 57.323 23.267
## - I(X^2)  1    1.5109 58.430 23.937
## <none>                56.919 25.020
## - I(X^3)  1    5.5267 62.446 26.263
##
## Step:  AIC=23.27
## Y ~ I(X^2) + I(X^3)
##
##          Df Sum of Sq    RSS    AIC
## <none>                57.323 23.267
## - I(X^2)  1    11.312 68.636 27.571
## - I(X^3)  1    35.319 92.642 38.069
```

```
##
## Call:
## lm(formula = Y ~ I(X^2) + I(X^3))
##
## Coefficients:
## (Intercept)       I(X^2)       I(X^3)
##     1.80585     -0.11733      0.01585
```

```
# Stepwise regression shows Y ~ I(X^2) + I(X^3) is best model
```

Fitting regression model with 4th degree polynomial

```
fit4d<-lm(Y ~ X + I(X^2) + I(X^3) + I(X^4))
summary(fit4d)
```

```
##
## Call:
## lm(formula = Y ~ X + I(X^2) + I(X^3) + I(X^4))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2659 -0.2809  0.0574  0.6924  2.0783
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.5233379  6.5245497  -0.080    0.937
## X            1.5063707  5.0058940   0.301    0.766
## I(X^2)      -0.4204591  1.2392220  -0.339    0.737
## I(X^3)       0.0398440  0.1235873   0.322    0.749
## I(X^4)      -0.0006484  0.0043170  -0.150    0.882
##
## Residual standard error: 1.377 on 30 degrees of freedom
## Multiple R-squared:  0.8883, Adjusted R-squared:  0.8734
## F-statistic: 59.66 on 4 and 30 DF,  p-value: 7.504e-14
```

```
step(fit4d)
```

```
## Start:  AIC=26.99
## Y ~ X + I(X^2) + I(X^3) + I(X^4)
##
##          Df Sum of Sq    RSS    AIC
## - I(X^4)  1  0.042773 56.919 25.020
## - X       1  0.171677 57.048 25.099
## - I(X^3)  1  0.197056 57.074 25.115
## - I(X^2)  1  0.218254 57.095 25.128
## <none>                56.877 26.994
##
## Step:  AIC=25.02
## Y ~ X + I(X^2) + I(X^3)
```

```
##
##           Df Sum of Sq    RSS    AIC
## - X        1    0.4037 57.323 23.267
## - I(X^2)  1    1.5109 58.430 23.937
## <none>                56.919 25.020
## - I(X^3)  1    5.5267 62.446 26.263
##
## Step:  AIC=23.27
## Y ~ I(X^2) + I(X^3)
##
##           Df Sum of Sq    RSS    AIC
## <none>                57.323 23.267
## - I(X^2)  1    11.312 68.636 27.571
## - I(X^3)  1    35.319 92.642 38.069
```

```
##
## Call:
## lm(formula = Y ~ I(X^2) + I(X^3))
##
## Coefficients:
## (Intercept)       I(X^2)       I(X^3)
##     1.80585     -0.11733      0.01585
```

```
# Stepwise regression still says Y ~ I(X^2) + I(X^3) is best model.

# Then, the best model is from the 3rd degree polynomial.
```

Forcing origin to be 0

```
fit_best_model_no_origin<-lm(Y ~ 0 + I(X^2) + I(X^3))
summary(fit_best_model_no_origin)
```

```
##
## Call:
## lm(formula = Y ~ 0 + I(X^2) + I(X^3))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4318 -0.4201  0.2784  0.9998  2.1299
##
## Coefficients:
##         Estimate Std. Error t value Pr(>|t|)
## I(X^2) -0.044332   0.019720  -2.248   0.0314 *
## I(X^3)  0.010580   0.001866   5.670 2.55e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.377 on 33 degrees of freedom
## Multiple R-squared:  0.9567, Adjusted R-squared:  0.9541
## F-statistic: 364.5 on 2 and 33 DF,  p-value: < 2.2e-16
```

```r
step(fit_best_model_no_origin)
```

```
## Start:  AIC=24.34
## Y ~ 0 + I(X^2) + I(X^3)
##
##          Df Sum of Sq     RSS    AIC
## <none>                 62.589 24.344
## - I(X^2)  1     9.586  72.175 27.331
## - I(X^3)  1    60.977 123.566 46.150
```

```
##
## Call:
## lm(formula = Y ~ 0 + I(X^2) + I(X^3))
##
## Coefficients:
##   I(X^2)    I(X^3)
## -0.04433   0.01058
```

```r
# It does not make sense to force the origin to be zero. Stepwise regression shows that
↪  the model with the intercept included has a better fit to the the data.

# However, if you decide to force the origin to be zero, the model is still a great
↪  fitting model. The fit to the data will definitely still be satisfactory if the
↪  origin is forced to be zero.
```