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

flow<-read.table("flow.txt",header=TRUE)
flow
```

```
##      Qo   Qp   X1   X2 X3  X4
## 1    28   32 0.03  3.0 70 0.6
## 2   112  142 0.03  3.0 80 1.8
## 3   398  502 0.13  6.5 65 2.0
## 4   772  790 1.00 15.0 60 0.4
## 5  2294 3075 1.00 15.0 65 2.3
## 6  2484 3230 3.00  7.0 67 1.0
## 7  2586 3535 5.00  6.0 62 0.9
## 8  3024 4265 7.00  6.5 56 1.1
## 9  4179 6529 7.00  6.5 56 1.4
## 10  710  935 7.00  6.5 56 0.7
```

```r
class(flow)
```

```
## [1] "data.frame"
```

```r
str(flow)
```

```
## 'data.frame':    10 obs. of  6 variables:
##  $ Qo: int  28 112 398 772 2294 2484 2586 3024 4179 710
##  $ Qp: int  32 142 502 790 3075 3230 3535 4265 6529 935
##  $ X1: num  0.03 0.03 0.13 1 1 3 5 7 7 7
##  $ X2: num  3 3 6.5 15 15 7 6 6.5 6.5 6.5
##  $ X3: int  70 80 65 60 65 67 62 56 56 56
##  $ X4: num  0.6 1.8 2 0.4 2.3 1 0.9 1.1 1.4 0.7
```

```r
attach(flow)
```

```
## The following objects are masked from cp (pos = 5):
##
##     X1, X2, X3, X4

## The following objects are masked from demand (pos = 6):
##
##     X1, X2, X3

## The following objects are masked from cp (pos = 7):
##
##     X1, X2, X3, X4

## The following objects are masked from demand (pos = 8):
##
##     X1, X2, X3
```

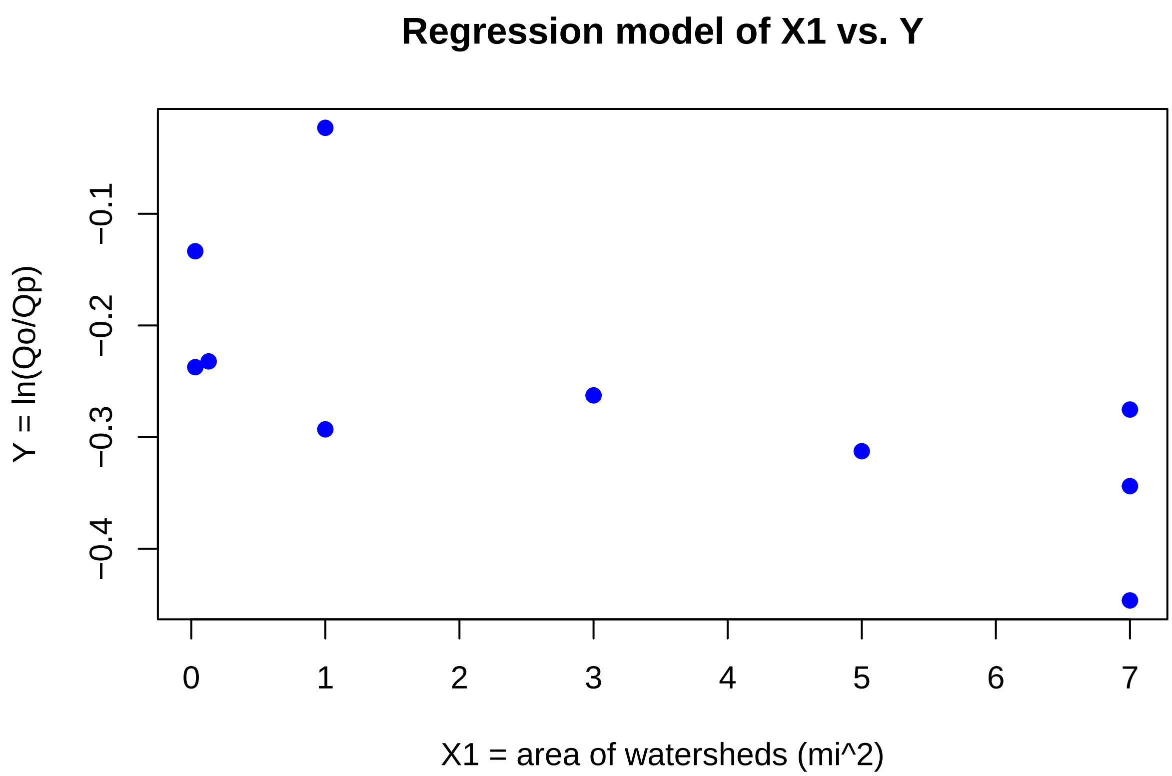
```
## The following objects are masked from demand (pos = 13):
##
##     X1, X2, X3

## The following objects are masked from demand (pos = 18):
##
##     X1, X2, X3

## The following objects are masked from flow (pos = 19):
##
##     Qo, Qp, X1, X2, X3, X4

## The following objects are masked from cp (pos = 20):
##
##     X1, X2, X3, X4

## The following objects are masked from cp (pos = 21):
##
##     X1, X2, X3, X4

## The following objects are masked from flow (pos = 22):
##
##     Qo, Qp, X1, X2, X3, X4

## The following objects are masked from flow (pos = 23):
##
##     Qo, Qp, X1, X2, X3, X4

## The following objects are masked from flow (pos = 24):
##
##     Qo, Qp, X1, X2, X3, X4
```

```r
#n<-length(qo)

dim(flow)
```

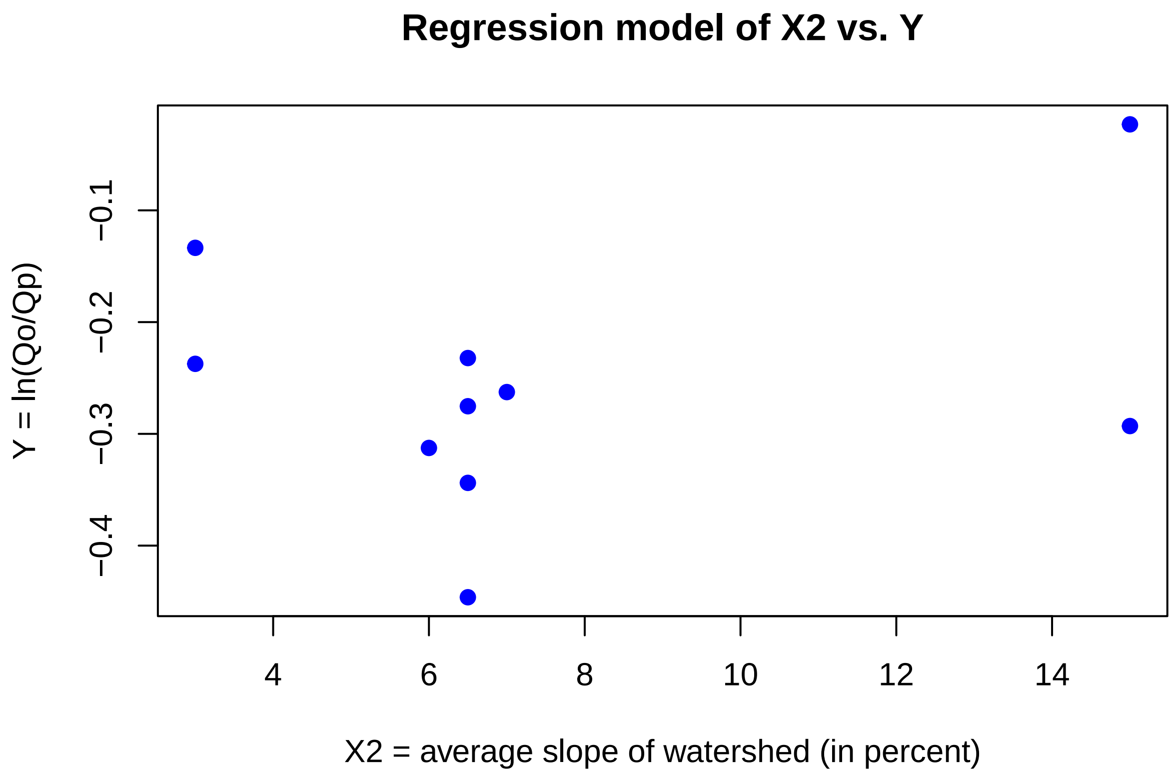
```
## [1] 10  6
```

```r
n<-dim(flow)[1]

#normally p is datapoints minus 1, but we have 2 variables for the response, so p is
↪  datapoints minus 2
p<-dim(flow)[2]-2

# sets Y = ln(Qo/Qp)
Y<-log(Qo/Qp)

fit<-lm(Y~X1+X2+X3+X4)

#summary(fit)

#fit$coef
```

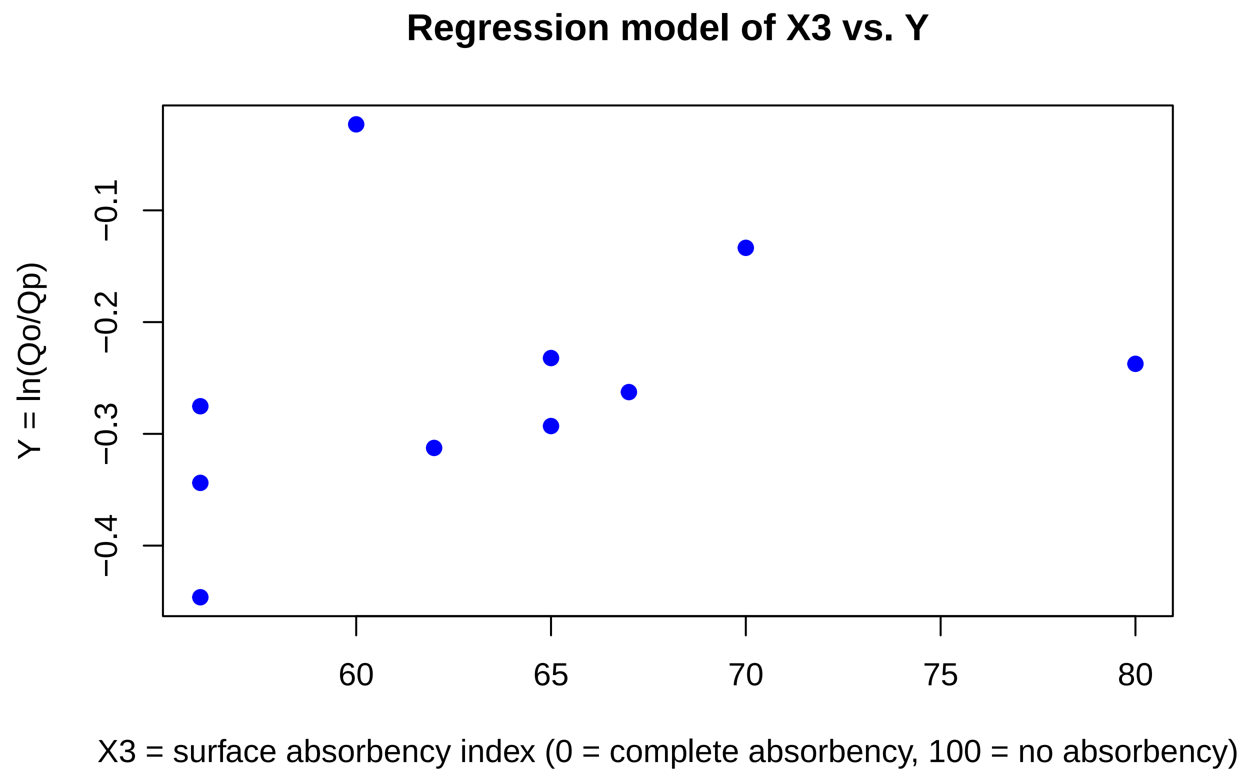
```r
# Visualization of correlation of X1 and Y
plot(X1,Y,pch=19,col="blue",xlab="X1 = area of watersheds (mi^2)",ylab="Y =
↪  ln(Qo/Qp)",main="Regression model of X1 vs. Y")
```



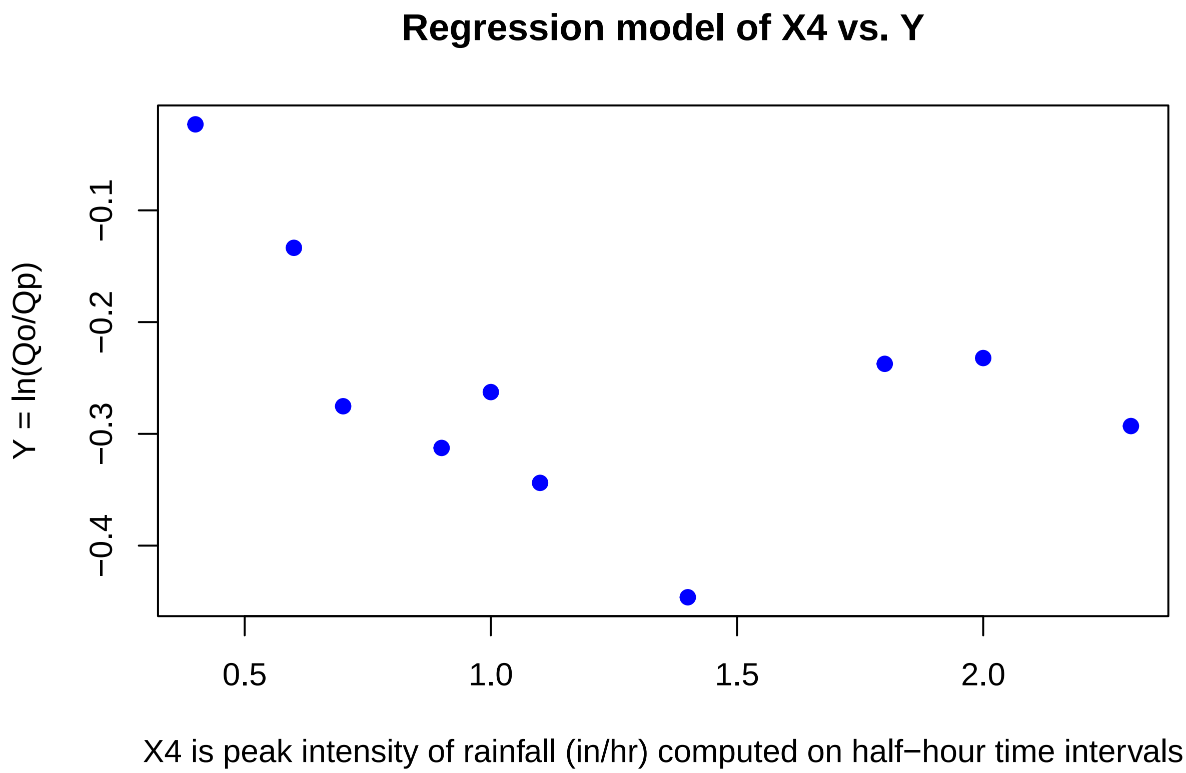

```r
# Visualization of correlation of X2 and Y
plot(X2,Y,pch=19,col="blue",xlab="X2 = average slope of watershed (in percent)",ylab="Y =
↪  ln(Qo/Qp)",main="Regression model of X2 vs. Y")
```




```
# Visualization of correlation of X3 and Y
plot(X3,Y,pch=19,col="blue",xlab="X3 = surface absorbency index (0 = complete absorbency,
↪  100 = no absorbency)",ylab="Y = ln(Qo/Qp)",main="Regression model of X3 vs. Y")
```




```
# Visualization of correlation of X4 and Y
plot(X4,Y,pch=19,col="blue",xlab="X4 is peak intensity of rainfall (in/hr) computed on
↪  half-hour time intervals",ylab="Y = ln(Qo/Qp)",main="Regression model of X4 vs. Y")
```




```r
# Overall, Y appears to have weak or some correlation with X1, X2, X3, X4 individually,
↪  but Y does not appear to have any strong correlation with X1, X2, X3, X4
↪  individually.
```

```r
# An intercept is included in the model
summary(fit)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3 + X4)
##
## Residuals:
##        1        2        3        4        5        6        7
## -0.02800  0.03143  0.01892  0.01694 -0.01530 -0.02860 -0.02467
##        8        9       10
##  0.02501 -0.04129  0.04556
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.087321   0.311931   0.280  0.79074
## X1          -0.035384   0.009336  -3.790  0.01276 *
## X2           0.004726   0.004765   0.992  0.36677
## X3          -0.001913   0.004189  -0.457  0.66702
## X4          -0.120080   0.023810  -5.043  0.00396 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04119 on 5 degrees of freedom
## Multiple R-squared:  0.9287, Adjusted R-squared:  0.8717
## F-statistic: 16.29 on 4 and 5 DF,  p-value: 0.004502
```

```
# Fitted values for the model with the intercept
fit$fit
```

```
##          1          2          3          4          5          6
## -0.1055298 -0.2687573 -0.2510728 -0.0399906 -0.2777083 -0.2340084
##          7          8          9         10
## -0.2879290 -0.3688713 -0.4048953 -0.3208393
```

```
# Fit of model with intercept removed (beta_0 = 0)
fit2<-lm(Y~0+X1+X2+X3+X4)
summary(fit2)
```

```
##
## Call:
## lm(formula = Y ~ 0 + X1 + X2 + X3 + X4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.04001 -0.02546 -0.00017  0.02573  0.04582
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## X1 -0.0330400  0.0037987  -8.698 0.000128 ***
## X2  0.0057518  0.0028025   2.052 0.085956 .
## X3 -0.0007525  0.0005484  -1.372 0.219124
## X4 -0.1215292  0.0213807  -5.684 0.001279 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0379 on 6 degrees of freedom
## Multiple R-squared:  0.9889, Adjusted R-squared:  0.9815
## F-statistic: 133.3 on 4 and 6 DF,  p-value: 5.468e-06
```

```
# Fitted values for the model with intercept removed
fit2$fit
```

```
##           1           2           3           4           5
## -0.10932768 -0.26268758 -0.25887880 -0.04052427 -0.27519213
##           6           7           8           9          10
## -0.23080341 -0.28671976 -0.36971468 -0.40617343 -0.32110301
```

```
# Compare model with intercept removed to model with intercept
summary(fit)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3 + X4)
##
## Residuals:
##        1        2        3        4        5        6        7
## -0.02800  0.03143  0.01892  0.01694 -0.01530 -0.02860 -0.02467
##        8        9       10
##  0.02501 -0.04129  0.04556
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.087321   0.311931   0.280  0.79074
## X1          -0.035384   0.009336  -3.790  0.01276 *
## X2           0.004726   0.004765   0.992  0.36677
## X3          -0.001913   0.004189  -0.457  0.66702
## X4          -0.120080   0.023810  -5.043  0.00396 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04119 on 5 degrees of freedom
## Multiple R-squared:  0.9287, Adjusted R-squared:  0.8717
## F-statistic: 16.29 on 4 and 5 DF,  p-value: 0.004502
```

```r
# Fit of model with intercept, X2, X3 removed
fit3<-lm(Y~0+X1+X4)
summary(fit3)
```

```
##
## Call:
## lm(formula = Y ~ 0 + X1 + X4)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.057972 -0.037806  0.004037  0.025781  0.059843
##
## Coefficients:
##    Estimate Std. Error t value Pr(>|t|)
## X1 -0.033182   0.003892  -8.525 2.76e-05 ***
## X4 -0.124274   0.012236 -10.156 7.56e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04342 on 8 degrees of freedom
## Multiple R-squared:  0.9805, Adjusted R-squared:  0.9756
## F-statistic: 201.3 on 2 and 8 DF,  p-value: 1.441e-07
```

```r
# Fitted values for the model with intercept, X2, X3 removed
fit3$fit
```

```
##           1           2           3           4           5
## -0.07555980 -0.22468848 -0.25286144 -0.08289136 -0.31901178
```

```
##          6           7           8           9          10
## -0.22381930 -0.27775551 -0.36897389 -0.40625606 -0.31926433
```

```r
# Compare to model with no intercept, but using X1, X2, X3, X4
summary(fit2)
```

```
##
## Call:
## lm(formula = Y ~ 0 + X1 + X2 + X3 + X4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.04001 -0.02546 -0.00017  0.02573  0.04582
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## X1 -0.0330400  0.0037987  -8.698 0.000128 ***
## X2  0.0057518  0.0028025   2.052 0.085956 .
## X3 -0.0007525  0.0005484  -1.372 0.219124
## X4 -0.1215292  0.0213807  -5.684 0.001279 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0379 on 6 degrees of freedom
## Multiple R-squared:  0.9889, Adjusted R-squared:  0.9815
## F-statistic: 133.3 on 4 and 6 DF,  p-value: 5.468e-06
```

The regression coefficients for X1 and X4 change across models. The reason for this change is because the independent variables X1, X2, X3, X4 are correlated. In order for there to be no change for the regression coefficients when independent variables (such as X2 and X3) are removed, the independent variables X1, X2, X3, X4 need to have zero correlation (be uncorrelated).

The standard error changes when the intercept is dropped because the data points are now in a completely different area of the graph when the intercept is dropped. This changes the regression line and makes small changes to the standard error of each regression coefficient.

Furthermore, the SE is reduced because removing the intercept makes the new model more constrainted to the regression line.

For when X2 and X3 are dropped from the model,

Consider beta_hat = (X^T X)^(-1) X^T Y

Also, consider SE of beta_hat = sqrt(diag(s^2 (X^T X)^(-1))) where s^2 is the variance of the residuals and diag() is the diagonal of a matrix

It is very obvious that removing X2 and X3 will change the (X^T X)^(-1), which in turn is used in the equation to calculate the SE of beta_hat. Obviously, if you modify a variable in the equation for SE of beta_hat, you will get a different value for SE of beta_hat.

Furthermore, the SE is reduced because removing X2 and X3 makes the new model more contrainted, and hence, a regression line is able to fit the model better.