# Relationship between daily mean Gage Height and Discharge on Boneyard Creek

Here in the first portion we establish and run our model and evaluate the significance of model coefficients and R^2 term.

Import the necessary packages and read the given CSV file

```
library("data.table")
library("moments")
t1unfiltered <-fread("6-daily_Boneyard_disharge_and_gage_height.csv", header=TRUE)
t1unfiltered <- setDT(t1unfiltered)
t1 <- subset(t1unfiltered, gage_height_feet_daily_mean >= 5 & gage_height_feet_daily_mean <= 12.5)
```

# Data Visualization

Convert the information from the CSV into a vector usable in R graphing functions:

```
discharge <- t1$discharge_cubic_feet_per #independent X variable
gageHeight <- t1$gage_height_feet_daily_mean #dependent Y variable
```

## Scatter plot of Discharge against Gage Height

```
plot (discharge, gageHeight, main='Scatterplot of Discharge against Gage Height', xlab = 'discharge (Cubic Ft.)',
ylab='Gage Height (Ft.)')
```
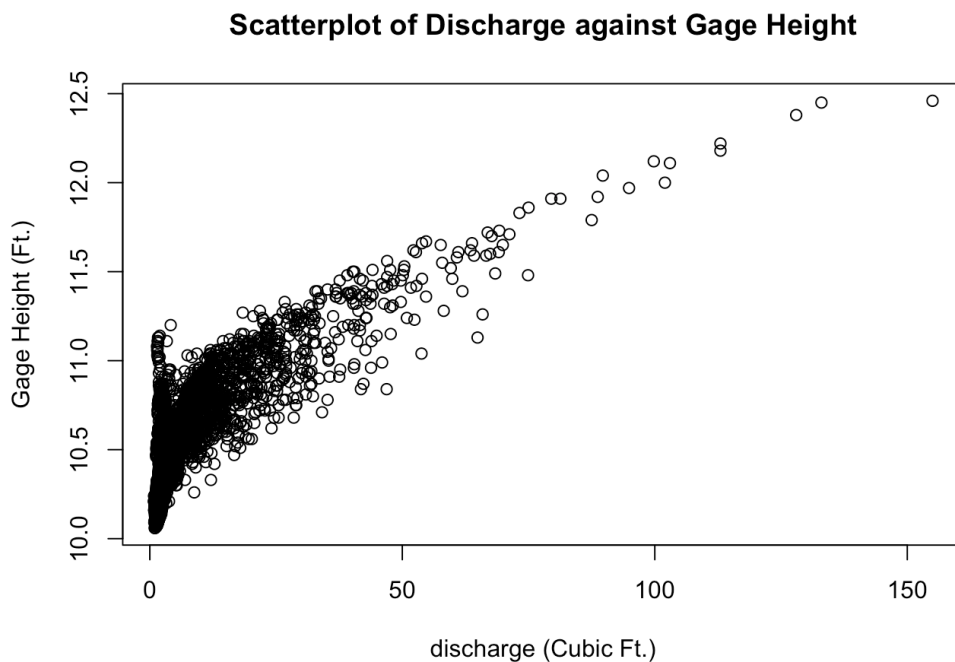


Calculating the correlation coefficient of the Discharge against Gage Height

```
correl=cor(discharge, gageHeight, use = "complete.obs")
print(correl)
```

```
## [1] 0.8031571
```

Linear Regression of Discharge against Gage Height and summary of data

```
lmodel = lm(gageHeight ~ discharge)

print(summary (lmodel))
```

```
##
## Call:
## lm(formula = gageHeight ~ discharge)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.63091 -0.10819 -0.02686  0.08522  0.79273
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.031e+01  2.248e-03  4584.1   <2e-16 ***
## discharge   2.442e-02  2.174e-04   112.3   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1589 on 6942 degrees of freedom
## Multiple R-squared:  0.6451, Adjusted R-squared:  0.645
## F-statistic: 1.262e+04 on 1 and 6942 DF,  p-value: < 2.2e-16
```

## Null and Alternate Hypothesis

Our Null Hypothesis H0 is that there is not a linear relationship between the Discharge and Gage Height.

Our Alternate Hypothesis H1 is that there is a linear relationship between the Discharge and Gage Height.

```
lmAttributes = attributes(lmodel)
lmHeading = lmodel$coefficients
lmConfidenceInterval = confint(lmodel, level=0.95) #for alpha = 0.05
correlationTest = cor.test(gageHeight, discharge, method = "pearson", conf.level = 0.95)
print(lmHeading)
```

```
## (Intercept)    discharge
## 10.30642757   0.02441603
```

```
print(lmConfidenceInterval)
```

```
##                  2.5 %       97.5 %
## (Intercept) 10.30202023 10.31083490
## discharge    0.02398991  0.02484216
```

```
print(correlationTest)
```

```
##
##  Pearson's product-moment correlation
##
## data:  gageHeight and discharge
## t = 112.32, df = 6942, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7946478 0.8113508
## sample estimates:
##       cor
## 0.8031571
```

Adjusted plot with line

```
plot (discharge, gageHeight, main='Scatterplot with Trend Line + 95% CI', xlab = 'discharge (Cubic Ft.)',
ylab='Gage Height (Ft.)')

abline(lmodel, col="red", lwd=2)
```

## Scatterplot with Trend Line + 95% CI
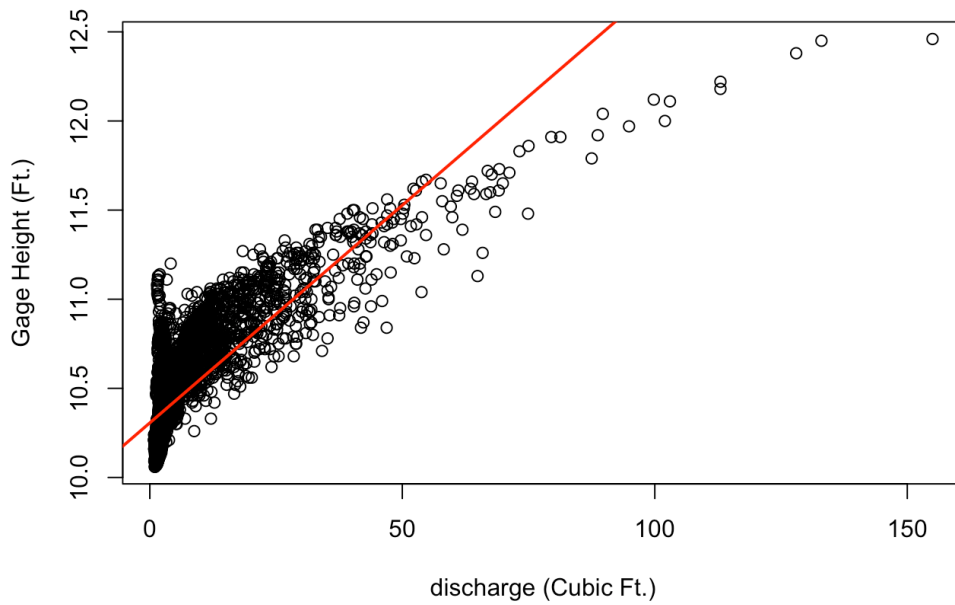


# Checking Our Assumptions

Here in the second portion we use R to test our assumptions by producing four plots that examine parameters of our data

```
par(mfrow=c(2,2))
plot(lmodel)
```