Case Study 4: Snow Gauge

Justin Glommen
Peter Yao
Atharva Fulay
Alex Hsieh

3/7/2017

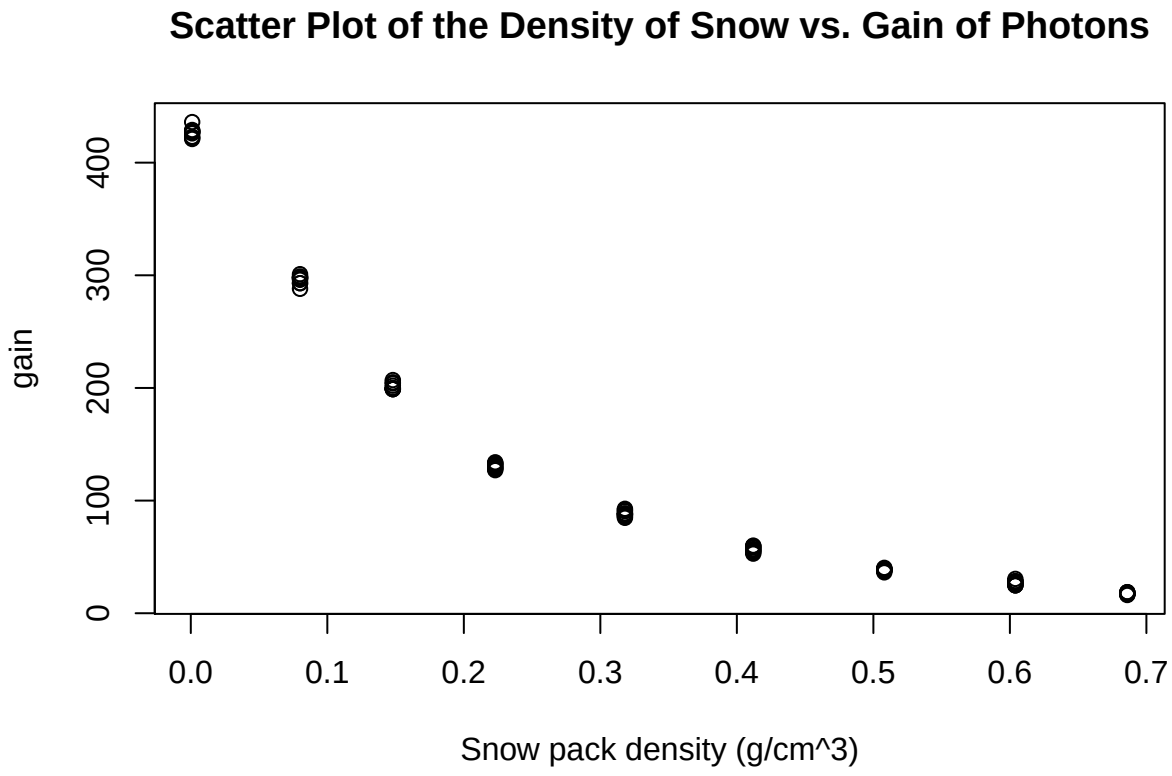
Setup

Here the data is loaded into its corresponding variable for analysis.

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

```
##   density gain
## 1   0.686 17.6
## 2   0.686 17.3
## 3   0.686 16.9
## 4   0.686 16.2
## 5   0.686 17.1
## 6   0.686 18.5
```

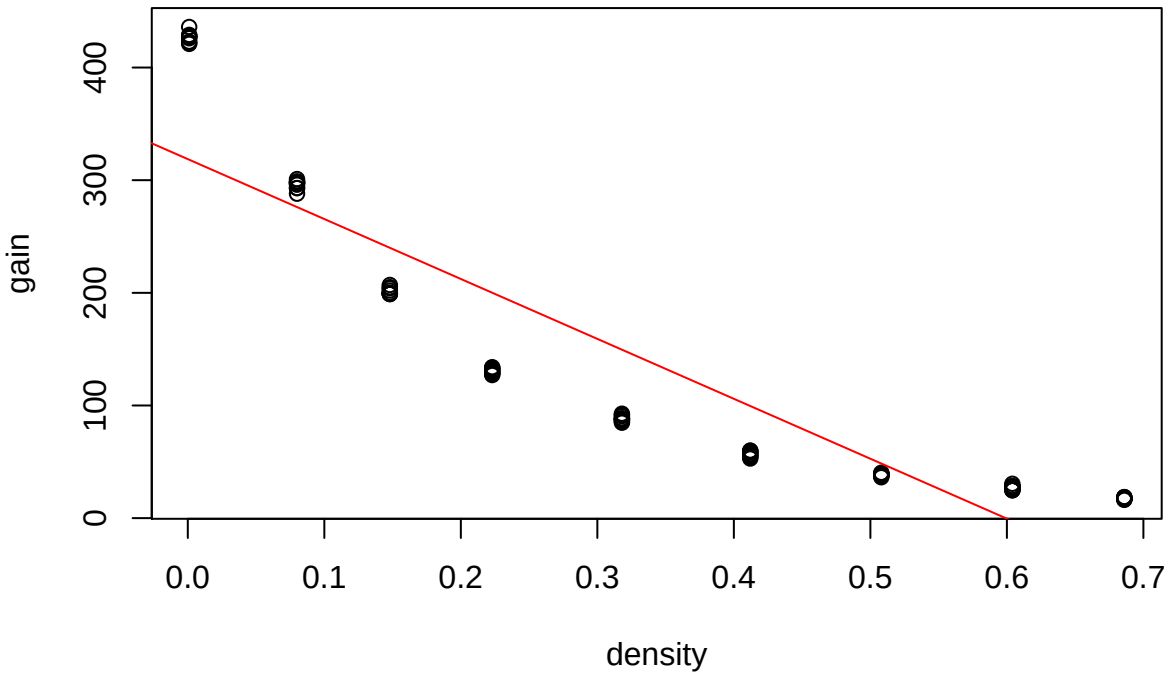
```
stringMain <- "Scatter Plot of the Density of Snow vs. Gain of Photons"
densityLabel <- "Snow pack density (g/cm^3)"
gainLabel <- "gain"
plot(gauge, xlab = densityLabel , ylab = gainLabel, main = stringMain)
```



Fitting

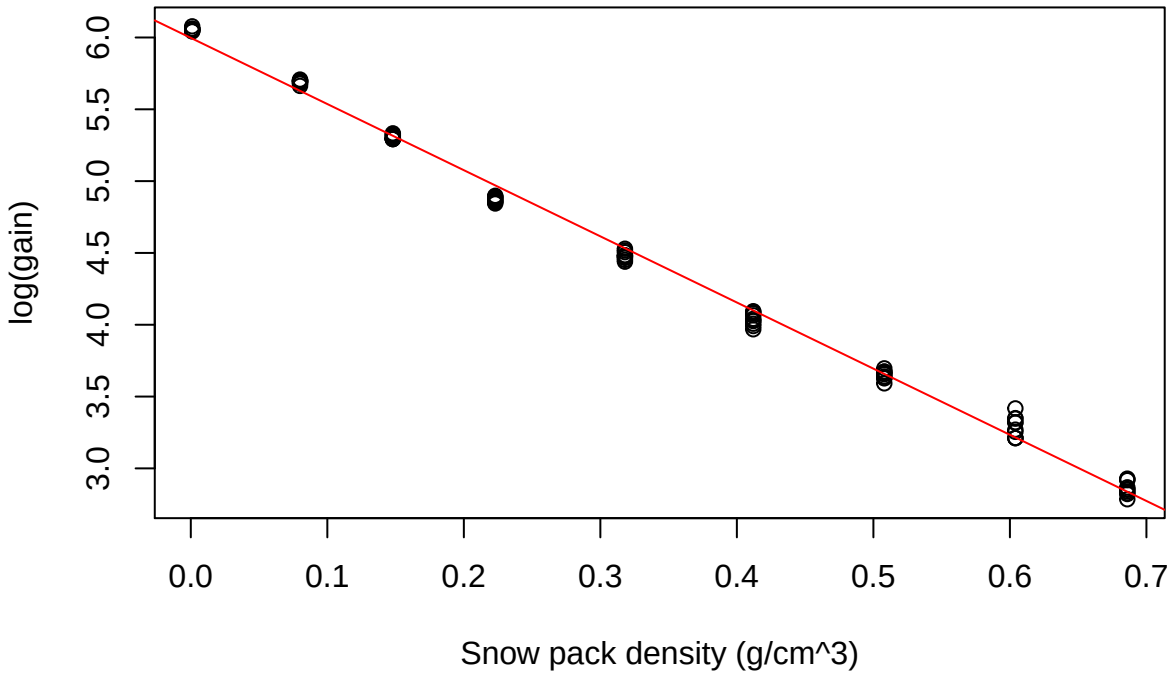
We now want to fit the lease squares line of the original data, such that we can plot and use it to help us make predictions about the snow-pack densities.

```
# Fit least squares line of orig data
fit <- lm(formula=gain~density, gauge)
plot(gauge)
abline(fit, col="red")
```



Here, we notice that the data follows a somewhat exponential-like pattern. Therefore, in order to enhance the regression line, we will take the log of the densities and display that instead.

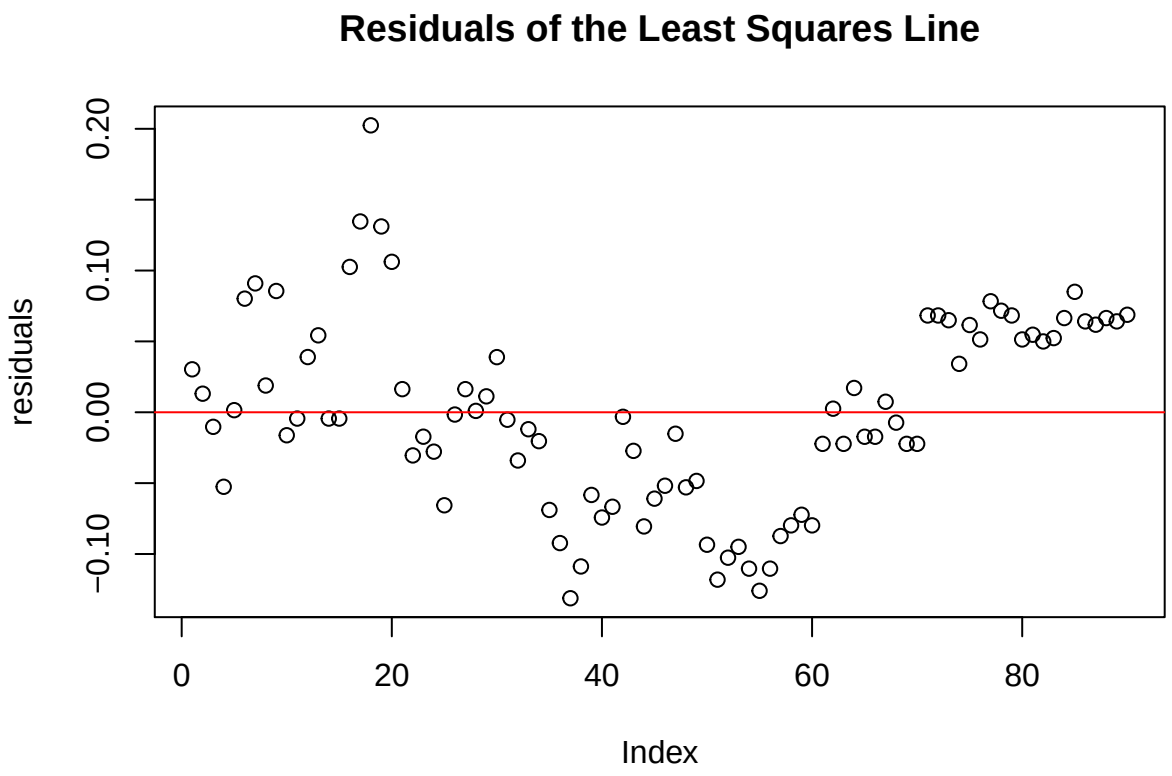
```
#transformed data log(gain)
gauge$gain <- log(gauge$gain)
#fit least squares line of trans data
fit <- lm(formula=gain~density, gauge)
plot(gauge, xlab = densityLabel, ylab = yLogGainLabel, main = transformedStringMain)
abline(fit, col="red")
```



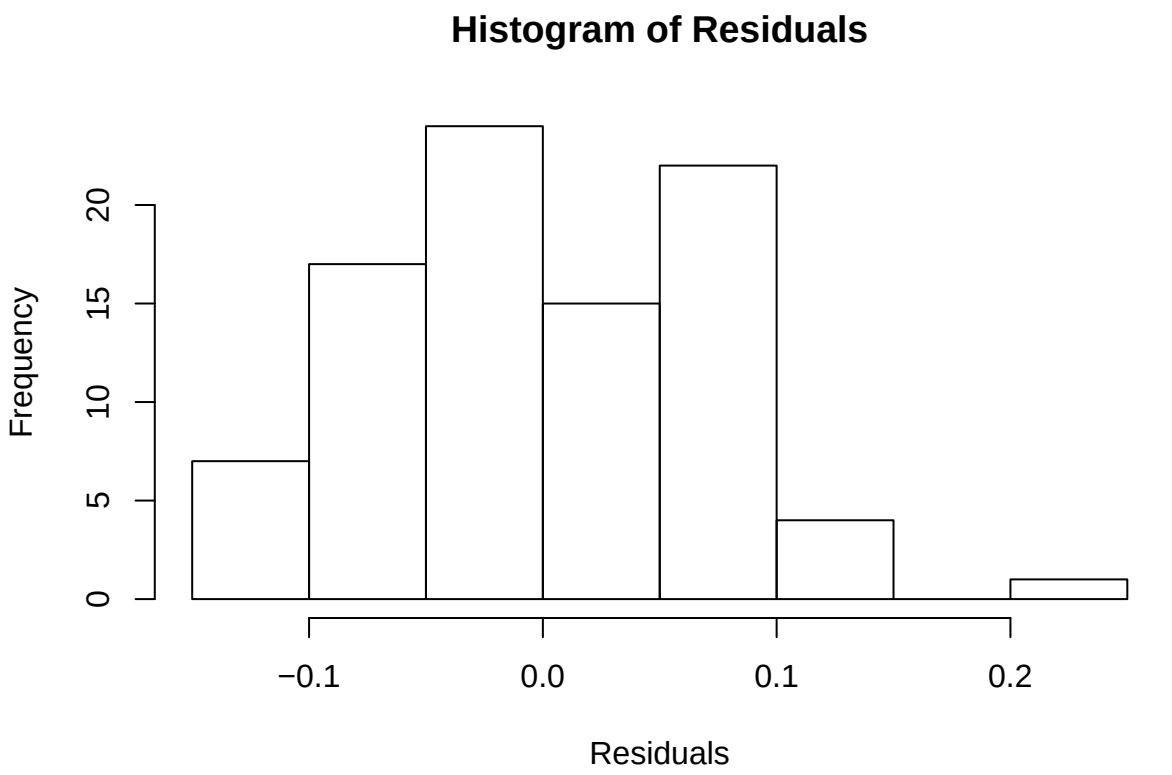

```
#find R^2 of trans data (.9958183)
R.square <- sum((fit$fitted.values-mean(gauge$gain))^2) / (sum((gauge$gain - mean(gauge$gain))^2))
R.square
```

```
## [1] 0.9958183
```

```
#residuals of trans data
plot(fit$residuals, ylab = "residuals", main = "Residuals of the Least Squares Line")
abline(0, 0, col="red")
```




```
hist(fit$residuals, xlab = "Residuals", main = "Histogram of Residuals")
```




```
library(e1071)
skewness(fit$residuals)
```

```
## [1] 0.1571663
```

```r
kurtosis(fit$residuals)
```

```
## [1] -0.35133
```

```r
x <- gauge[["gain"]]
quantile(x, probs = seq(0.1, 0.9, by = 0.2))
```

```
##      10%      30%      50%      70%      90%
## 2.927987 3.672239 4.480174 5.293305 6.042870
```

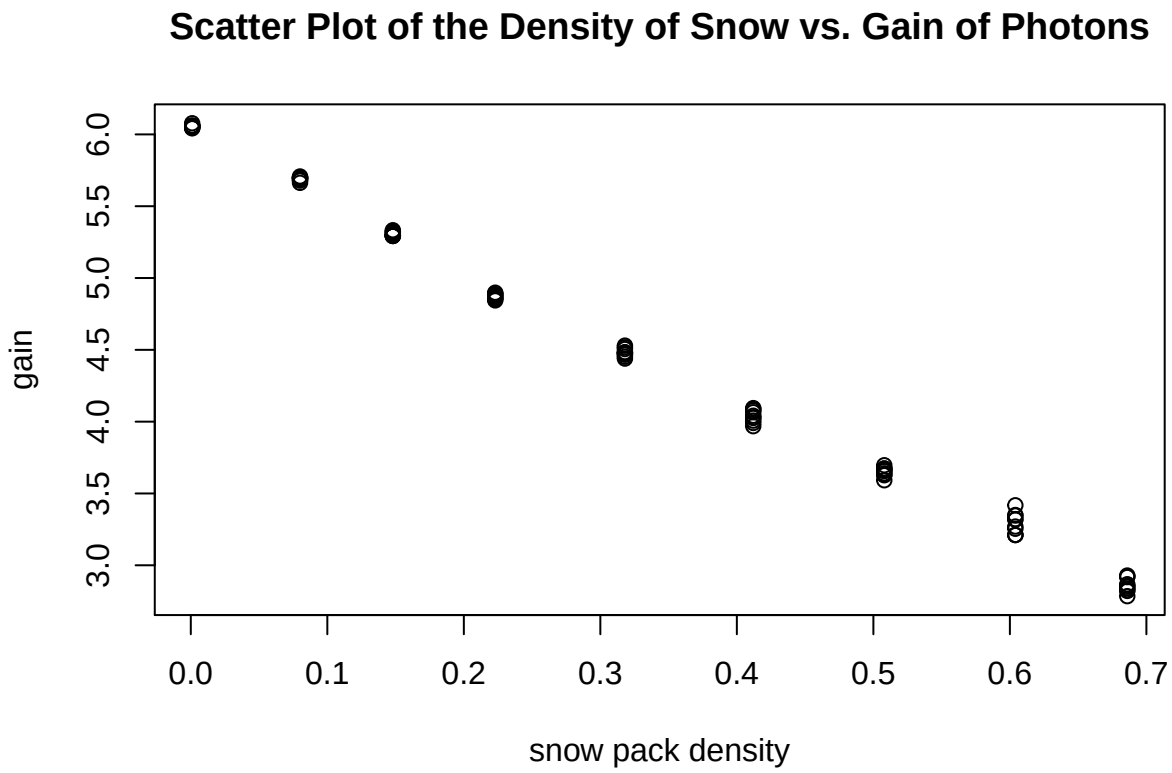
```r
#quantile regression
plot(gauge, xlab = "snow pack density", ylab = "gain", main = "Scatter Plot of the Density of Snow vs. G

library(quantreg)
```

```
## Loading required package: SparseM
```

```
##
## Attaching package: 'SparseM'
```

```
## The following object is masked from 'package:base':
##
##     backsolve
```




```r
x <- seq(0, 0.7, length.out = 90)
y <- x*gauge[["gain"]]
plot(x, y, pch = ".", ylim = c(-5, 5))
# median
fit1 <- rq(y ~ x, tau = 0.5)
abline(fit1, col = 2)

# true median
```

```r
true1 <- x
lines(x, true1, col = 2, lty = 3)

# 0.2 quantile
fit2 <- rq(y ~ x, tau = 0.2)
abline(fit2, col = 3)

# true 0.2 quantile
true2 <- qnorm(p = 0.2, mean = x, sd = x)
lines(x, true2, col = 3, lty = 3)

# 0.7 quantile
fit3 <- rq(y ~ x, tau = 0.7)
abline(fit3, col = 4)

# true 0.7 quantile
true3 <- qnorm(p = 0.7, mean = x, sd = x)
lines(x, true3, col = 4, lty = 3)

legend(x = 0, y = 5, legend = c(expression(paste("estimated", rho, "=", 0.2)),
                                expression(paste("estimated", rho, "=", 0.5)),
                                expression(paste("estimated", rho, "=", 0.7)),
                                expression(paste("true", rho, "=", 0.2)),
                                expression(paste("true", rho, "=", 0.5)),
                                expression(paste("true", rho, "=", 0.7))),
       lty = c(1,1,1,3,3,3), col = c(3,2,4,3,2,4))
```