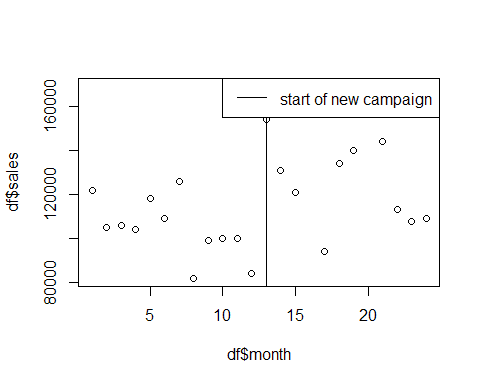
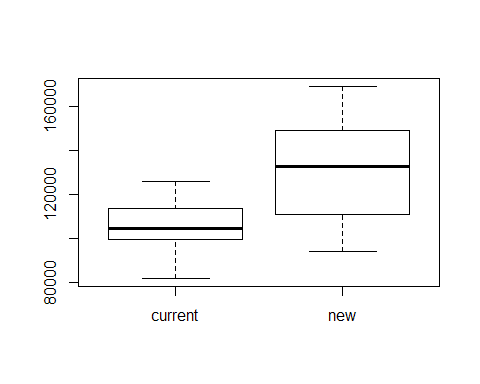
#1a.  
month <- seq(1,24)  
campaign <- append(rep("current", 12), rep("new", 12))  
sales <- scan('wk4.csv', nlines=1, sep=',', skip=0)  
df <- data.frame(month, campaign, sales)  
#1b.  
plot(df$month, df$sales)  
abline(v=13)  
legend('topright', c("start of new campaign"), lty=1)



#1c.  
boxplot(df$sales ~ df$campaign)



t.test(df$sales ~ df$campaign, alternative="less")

##   
## Welch Two Sample t-test  
##   
## data: df$sales by df$campaign  
## t = -3.478, df = 17.58, p-value = 0.001381  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
## -Inf -13479.23  
## sample estimates:  
## mean in group current mean in group new   
## 104583.3 131500.0

### 1c.

The new model has a larger mean and range. This seems to indicate the new model will lead to larger sales. The range is much greater in the new model. As such, the new model may be more volitile from month to month.

### 1d.

Yes. The new marketing campaign appears to have been successful. The mean is higher.

### 1e.

### 1f.

The 2-sample t-test is appropriate because it determines if two population means are equal. The two populations in question are the populations from before and after the new marketing campaign.

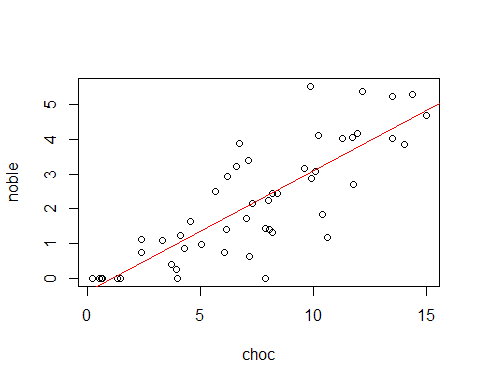
### 1g.

There is enough evidence in this data set to reject the null hypothesis. In other words, the new campaign does appear to increase mean sales.

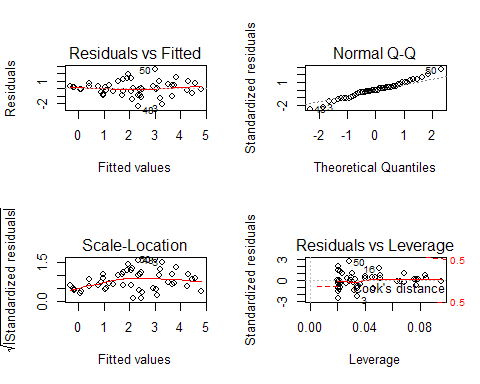
#2a.  
choc <- scan('wk4.csv', nlines=1, sep=',', skip=1)  
noble <- scan('wk4.csv', nlines=1, sep=',', skip=2)  
df <- data.frame(choc, noble)  
plot(df)  
#2b.  
model <- lm(noble ~ choc)   
model

##   
## Call:  
## lm(formula = noble ~ choc)  
##   
## Coefficients:  
## (Intercept) choc   
## -0.3827 0.3466

abline(model, col='red')



#2e.  
par(mfrow=c(2,2))  
plot(model)



### 2b.

The equation is y = -0.3827 + 0.3466 \* x. The model is how I would expect it to be. To generate the random numbers, we started with a perfect linear equation and added an e term which was a random number from a normal distribution. I can see that the data points appear to have equal variance above and below the mean.

### 2c.

H0: There is no correlation between the variables

Ha: There is a correlation between the variables

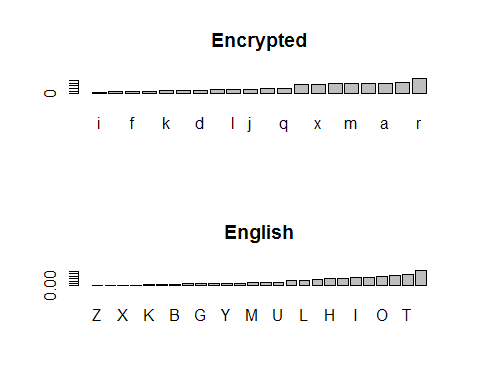
### 2d.

There is evidence to request the null hypothesis. In other words, as chocolate consumption goes up so do number of Nobel Prize winners.

### 2e.

The residuals vs. fitted graph shows an even dispersion around the line. The normal Q-Q plot shows a less than ideal plot around the lower and upper value ranges.

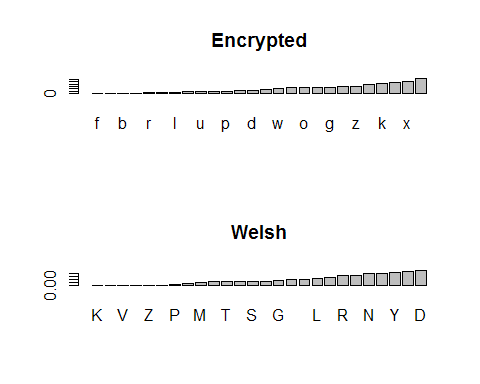
#3a.  
df\_encryptA <- read.csv('encryptedA.csv', header=FALSE)  
colnames(df\_encryptA) <- c('key', 'count')  
encryptA\_order <- order(df\_encryptA$count)  
#3b.  
df\_letterFreq <- read.csv('Letter Frequencies.csv')  
englishFreq\_order <- order(df\_letterFreq$English)  
#3c.  
par(mfrow=c(2,1))  
barplot(df\_encryptA$count[encryptA\_order],   
 names.arg=df\_encryptA$key[encryptA\_order],   
 main="Encrypted")  
barplot(df\_letterFreq$English[englishFreq\_order],   
 names.arg=df\_letterFreq$Letter[englishFreq\_order],   
 main="English")



#3e  
encryptA\_sort <- df\_encryptA$count[encryptA\_order]  
english\_sort <- df\_letterFreq$English[englishFreq\_order]  
#3f.  
encryptA\_sort <- c(rep(0,6), encryptA\_sort)  
#3i.  
expected\_countsA <- english\_sort \* sum(df\_encryptA$count)  
english\_sortCombined <- c(sum(english\_sort[1:4]), english\_sort[5:26])  
encryptA\_sortCombined <- c(sum(encryptA\_sort[1:4]), encryptA\_sort[5:26])  
#3j.  
chisq.test(encryptA\_sortCombined, p=english\_sortCombined)

##   
## Chi-squared test for given probabilities  
##   
## data: encryptA\_sortCombined  
## X-squared = 44.6421, df = 22, p-value = 0.002945

#3l.  
df\_encryptB <- read.csv('encryptedB.csv', header=FALSE)  
colnames(df\_encryptB) <- c('key', 'count')  
encryptB\_order <- order(df\_encryptB$count)  
df\_letterFreq <- read.csv('Letter Frequencies.csv')  
welshFreq\_order <- order(df\_letterFreq$Welsh)  
par(mfrow=c(2,1))  
barplot(df\_encryptB$count[encryptB\_order],   
 names.arg=df\_encryptB$key[encryptB\_order],   
 main="Encrypted")  
barplot(df\_letterFreq$Welsh[welshFreq\_order],   
 names.arg=df\_letterFreq$Letter[welshFreq\_order],   
 main="Welsh")



encryptB\_sort <- df\_encryptB$count[encryptB\_order]  
#encryptB\_sort <- c(rep(0,6), encryptB\_sort)  
welsh\_sort <- df\_letterFreq$Welsh[welshFreq\_order]  
expected\_countsB <- welsh\_sort \* sum(df\_encryptB$count)  
welsh\_sortCombined <- c(sum(welsh\_sort[1:7]), welsh\_sort[8:26])  
encryptB\_sortCombined <- c(sum(encryptB\_sort[1:7]), encryptB\_sort[8:26])  
chisq.test(encryptB\_sortCombined, p=welsh\_sortCombined)

##   
## Chi-squared test for given probabilities  
##   
## data: encryptB\_sortCombined  
## X-squared = 201.6661, df = 19, p-value < 2.2e-16

### 3d.

Yes. The shapes are very similar. The distributions appear to be close to the same.

### 3g.

H0: The distribution in the letter frequency table is a good fit for the encrypted data.

Ha: At least one proportion of the encrypted data is different from the porportion of the letter frequency table.

### 3k.

Based on the chi-squared Goodness of Fit, we have failed to reject the null hypothesis. This means its entirely possible the encrypted data is English.

### 3l.

Based on the hypothesis tests, I would claim that the encryptedA file is in English and the encryptedB file is in Welsh. I am highly confident in this assessment. This is because both chi-squared Goodness of Fit tests produced low p-values.